Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 22–05

SUBSTITUTION PRINCIPLE AND SEMIDIRECT PRODUCTS

CÉLIA BORLIDO AND MAI GEHRKE

Abstract: In the classical theory of regular languages the concept of recognition by profinite monoids is an important tool. Beyond regularity, Boolean spaces with internal monoids (BiMs) were recently proposed as a generalization. On the other hand, fragments of logic defining regular languages can be studied inductively via the so-called "Substitution Principle". In this paper we make the logical underpinnings of this principle explicit and extend it to arbitrary languages using Stone duality. Subsequently we show how it can be used to obtain topo-algebraic recognizers for classes of languages defined by a wide class of first-order logic fragments. This naturally leads to a notion of semidirect product of BiMs extending the classical such construction for profinite monoids. Our main result is a generalization of Almeida and Weil's Decomposition Theorem for semidirect products from the profinite setting to that of BiMs. This is a crucial step in a program to extend the profinite methods of regular language theory to the setting of complexity theory.

1. Introduction

Profinite monoids have proved to be a powerful tool in the theory of regular languages. Eilenberg and Reiterman Theorems allow for the study of the so-called *varieties of regular languages* through the study of topo-algebraic properties of suitable profinite monoids. In 2008, Gehrke, Grigorieff and Pin [8] proposed a unified approach for the study of regular languages, using Stone duality. In particular, they realized that profinite monoids could be seen as the extended dual spaces of certain Boolean algebras equipped with a residuation structure. More generally, to a Boolean algebra of (possibly non-regular) languages closed under quotients, one can assign a Boolean space equipped with a biaction of a dense monoid [9]. A slight variant (a so-called BiM - *Boolean space with an internal monoid*) was identified in [10].

Received February 8, 2022.
This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No.670624). Célia Borlido was also partially supported by the Centre for Mathematics of the University of Coimbra - UIDB/00324/2020, funded by the Portuguese Government through FCT/MCTES.

Complexity theory and the theory of regular languages are intimately connected through logic. As with classes of regular languages, many computational complexity classes have been given characterizations as model classes of appropriate logic fragments on finite words [13]. For example, $\mathrm{AC}^0 = \mathbf{FO}[\mathrm{arb}]$, $\mathrm{ACC}^0 = (\mathbf{FO}+\mathbf{MOD})[\mathrm{arb}]$, and $\mathrm{TC}^0 = \mathbf{MAJ}[\mathrm{arb}]$ where arb is the set of all possible (numerical) predicates of all arities on the positions of a word, **FO** is usual first-order logic, and **MOD** and **MAJ** stand for the *modular* and *majority* quantifiers, respectively. On the one hand, the presence of *arbitrary predicates*, and on the other hand, the presence of the *majority quantifier* is what brings one far beyond the scope of the profinite algebraic theory of regular languages.

Most results in the field of complexity theory are proved using combinatorial and probabilistic, as well as algorithmic methods [23]. However, there are a few connections with the topo-algebraic tools of the theory of regular languages. A famous result of Barrington, Compton, Straubing, and Thérien [4] states that a regular language belongs to AC^0 if and only if its syntactic homomorphism is quasi-aperiodic. Although this result relies on [6] and no purely algebraic proof is known, being able to characterize the class of regular languages that are in AC^0 gives some hope that the non-regular classes might be amenable to treatment by the generalized topo-algebraic methods.

Indeed, the hope is that one can generalize the tools of algebraic automata theory and this paper is a contribution to this program. In the paper *Logic Meets Algebra: the case of regular languages* [20], Tesson and Thérien lay out the theory used to characterize logic classes in the setting of regular languages in terms of their recognizers. Here we add topology to the picture, using Stone duality, to obtain corresponding tools that apply beyond the setting of regular languages. In particular, we generalize the *Substitution Principle* and the associated semidirect product construction used in the study of logic on words for regular languages to the general setting. This is an extended version of a part of [5] where many proofs are either omitted or not written up in detail.

In the sequence of papers [10, 11, 12], a semidirect product construction for BiMs corresponding to the addition of a layer of quantifiers has been studied. While related, our results here differ in three main points. For one, those results require the lp-variety (that is, the input corresponding to the quantifier) to be regular, and only the predicates or formula algebra need not be regular. Secondly, the lp-variety needs to be generated by a finite *semiring*

– from which only the (not-necessarily-commutative) sum is required for our result. Finally, the approach is category theoretic and does not make the link with the substitution principle of [20].

The paper is organized as follows. In Section 2 we introduce the background needed and set up the main notation used throughout the paper. Important concepts here are those of an *lp-strain* of languages and of a *class of sentences*. Essentially, an *lp-strain* of languages is an assignment to any finite *alphabet* A (i.e., a finite set) of a Boolean algebra of *languages* (i.e., a Boolean subalgebra of the powerset $\mathcal{P}(A^*)$, where A^* denotes the A-generated free monoid), which is closed under taking preimages under *length-preserving homomorphisms*. A *class of sentences* is the corresponding notion in the setting of logic on words. Section 3 is devoted to languages over profinite alphabets. These are essential to the main results of this paper. Section 4 is an introduction to recognition of Boolean algebras of (not necessarily regular) languages that are *closed under quotients*. We define what is a *Boolean space with an internal monoid* (BiM) [9, 10], and we consider a slight generalization of this notion: that of a BiM-*stamp*. Informally, a BiM is to a *variety of languages* what a BiM-*stamp* is to an *lp-variety of languages* (i.e., an lp-strain in which we require *closure under quotients*). In Section 5 we generalize the *Substitution Principle* [20] beyond regularity. We start by considering the case of finite Boolean algebras in Section 5.1. The key idea in this section is to use letters of a finite alphabet to play the role of atoms of a finite Boolean algebra of formulas. Since any Boolean algebra is the direct limit of its finite subalgebras, its Stone dual space is the projective limit of the sets of atoms of its finite subalgebras. This is why *profinite alphabets* (i.e. Boolean spaces) come into play in a natural way in Section 5.2, where we study substitution with respect to arbitrary Boolean algebras, and why the concept of *lp-strain* (and *class of sentences*) is crucial in the definition of a language over a profinite alphabet. In [20] only formulas with at most one free variable are considered since the extension to finite sets of free variables, needed to study predicate logic, is straightforward. For completeness, we prove this in Section 5.3. Finally, in Section 5.4, we illustrate the *Substitution Principle* in the setting of logic on words, by showing how it can be used to understand the effect of applying a layer of quantifiers to a given Boolean algebra of formulas. The main result of the paper, Theorem 6.7, is presented in Section 6. In this section we get out of the context of logic, and we study the closure under quotients of the

operation on languages derived in Section 5 via duality. Theorem 6.7 is a generalization of the classical result by Almeida and Weil [1].

2. Preliminaries

In this section we briefly present the background needed in the rest of the paper. For further reading on duality we refer to [14], and for formal language theory and logic on words to [18].

2.1. Discrete duality. This is the most basic duality we use and it provides a correspondence between powerset Boolean algebras (these are the complete and atomic Boolean algebras) and sets. Given such a Boolean algebra $\mathcal{B}$, its dual is its set of atoms, denoted $\mathrm{At}(\mathcal{B})$ and, given a set X, its dual is the Boolean algebra $\mathcal{P}(X)$. Clearly going back and forth yields isomorphic objects. If $h\colon \mathcal{B} \to \mathcal{A}$ preserves arbitrary meets and joins, the dual of h, denoted $\mathrm{At}(h)\colon \mathrm{At}(\mathcal{A}) \to \mathrm{At}(\mathcal{B})$, is given by the adjunction:

$$\forall\ a \in \mathcal{A},\ x \in \mathrm{At}(\mathcal{B}) \quad (\mathrm{At}(h)(a) \leq x \iff a \leq h(x))\,.$$

For example, if $\iota\colon \mathcal{B} \rightarrowtail \mathcal{P}(X)$ is the inclusion of a finite Boolean subalgebra of a powerset, then $\mathrm{At}(\iota)\colon X \twoheadrightarrow \mathrm{At}(\mathcal{B})$ is the quotient map corresponding to the finite partition of X given by the atoms of $\mathcal{B}$. Conversely, given a function $f\colon X \to Y$, the dual is just $f^{-1}\colon \mathcal{P}(Y) \to \mathcal{P}(X)$.

2.2. Stone duality. Generally Boolean algebras do not have enough atoms, and we have to consider *ultrafilters* instead (which may be seen as 'searches downwards' for atoms). Given an arbitrary Boolean algebra $\mathcal{B}$, an ultrafilter of $\mathcal{B}$ is a subset γ of $\mathcal{B}$ satisfying:

- γ is an upset, i.e., $a \in \gamma$ and $a \leq b$ implies $b \in \gamma$;
- γ is closed under finite meets, i.e., $a, b \in \gamma$ implies $a \wedge b \in \gamma$;
- for all $a \in \mathcal{B}$ exactly one of a and $\neg a$ is in γ.

We will denote the set of ultrafilters of $\mathcal{B}$ by $X_{\mathcal{B}}$, and we will consider it as a topological space equipped with the topology generated by the sets $\widehat{a} = \{\gamma \in X_{\mathcal{B}} \mid a \in \gamma\}$ for $a \in \mathcal{B}$. The last property in the definition of ultrafilters implies that these basic open sets are also closed (and thus *clopen*). The resulting spaces are compact, Hausdorff, and have a basis of clopen sets. Such spaces are called *Boolean spaces*. Conversely, given a Boolean space, its clopen subsets form a Boolean algebra and one can show that going back and forth results in isomorphic objects. Given a homomorphism

E-mail address: mai.gehrke@unice.fr
URL: http://math.unice.fr/∼mgehrke

[7] M. Gehrke. Stone duality, topological algebra, and recognition. *J. Pure Appl. Algebra*, 220(7):2711–2747, 2016.
[8] M. Gehrke, S. Grigorieff, and J.-É. Pin. Duality and equational theory of regular languages. In L. Aceto and al., editors, *ICALP 2008, Part II*, volume 5126 of *Lect. Notes Comp. Sci.*, pages 246–257, Berlin, 2008. Springer.
[9] M. Gehrke, S. Grigorieff, and J.-É. Pin. A topological approach to recognition. In *Automata, languages and programming. Part II*, volume 6199 of *Lecture Notes in Comput. Sci.*, pages 151–162. Springer, Berlin, 2010.
[10] M. Gehrke, D. Petrişan, and L. Reggio. The Schützenberger product for syntactic spaces. In *43rd International Colloquium on Automata, Languages, and Programming*, volume 55 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 112, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.
[11] M. Gehrke, D. Petrişan, and L. Reggio. Quantifiers on languages and codensity monads. In *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, page 12. IEEE, [Piscataway], NJ, 2017.
[12] M. Gehrke, D. Petrişan, and L. Reggio. Quantifiers on languages and codensity monads. *Mathematical Structures in Computer Science*, pages 1–35, 2021.
[13] N. Immerman. *Descriptive complexity*. Graduate Texts in Computer Science. Springer-Verlag, New York, 1999.
[14] P. T. Johnstone. *Stone spaces*, volume 3 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1986. Reprint of the 1982 edition.
[15] J.-É. Pin and H. Straubing. Some results on $\mathcal{C}$-varieties. *Theoret. Informatics Appl.*, 39:239–262, 2005.
[16] N. Pippenger. Regular languages and Stone duality. *Theory Comput. Syst.*, 30(2):121–134, 1997.
[17] J. R. Porter and R. G. Woods. *Extensions and absolutes of Hausdorff spaces*. Springer-Verlag, New York, 1988.
[18] H. Straubing. *Finite automata, formal logic, and circuit complexity*. Progress in Theoretical Computer Science. Birkhäuser Boston, Inc., Boston, MA, 1994.
[19] H. Straubing. On logical descriptions of regular languages. In *LATIN 2002*, number 2286 in Lect. Notes Comp. Sci., pages 528–538, Berlin, 2002. Springer.
[20] P. Tesson and D. Thérien. Logic meets algebra: the case of regular languages. *Log. Methods Comput. Sci.*, 3(1):1:4, 37, 2007.
[21] P. Weil. On varieties of languages closed under products with counter. In *Mathematical foundations of computer science 1989 (Porabka-Kozubnik, 1989)*, volume 379 of *Lecture Notes in Comput. Sci.*, pages 534–544. Springer, Berlin, 1989.
[22] S. Willard. *General topology*. Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1970 original.
[23] R. Williams. Nonuniform ACC circuit lower bounds. *J. ACM*, 61(1):Art. 2, 32, 2014.

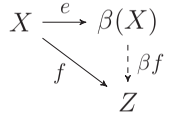
Célia Borlido
CMUC, University of Coimbra, Department of Mathematics, 3001-501 Coimbra, Portugal
E-mail address: cborlido@mat.uc.pt
URL: http://mat.uc.pt/∼cborlido

Mai Gehrke
Laboratoire J.A. Dieudonné UMR CNRS 7351, Université Côte d'Azur, Parc Valrose, 06108 Nice Cedex 02 France

$h\colon \mathcal{A} \to \mathcal{B}$ between Boolean algebras, the inverse image of an ultrafilter is an ultrafilter and thus h^{-1} induces a continuous map $X_{\mathcal{B}} \to X_{\mathcal{A}}$. Finally, given a continuous function $f\colon X \to Y$, the inverse map restricts to clopens and yields a homomorphism of Boolean algebras.

2.3. Compactification of topological spaces. We consider two compactifications of topological spaces: the *Čech-Stone* and the *Banaschewski* compactification, being the latter only defined for zero-dimensional spaces. We also relate them with the dual spaces of suitable Boolean algebras.

Let X be a topological space. Then, the *Čech-Stone compactification* of X is the unique (up to homeomorphism) compact Hausdorff space $\beta(X)$ together with a continuous function $e : X \to \beta(X)$ satisfying the following universal property: every continuous function $f : X \to Z$ into a compact Hausdorff space Z uniquely determines a continuous function $\beta f : \beta(X) \to Z$ making the following diagram commute:



In the case where X is a completely regular T_1 space, the map e is in fact an embedding [22, Section 19]. This is for instance the case of discrete and of Boolean spaces. It may be proved that, for a set S, the dual space of $\mathcal{P}(S)$ is the *Čech-Stone compactification* of the discrete space S. This fact has been extensively used in the topo-algebraic approach to non-regular languages over finite alphabets (see e.g. [9, 10]). However, as already mentioned, in the present work we also consider languages over profinite alphabets. For that reason, we will be handling dual spaces of Boolean algebras of the form $\mathsf{Clopen}(X)$, for spaces X that are not discrete. In the case where X is zero-dimensional, the dual space of $\mathsf{Clopen}(X)$ is known as the *Banaschewski compactification* [2, 3] of X and it is denoted by $\beta_0(X)$. At the level of morphisms, β_0 assigns to each continuous map the dual of the Boolean algebra homomorphism taking inverse image on clopens. Moreover, we have a topological embedding $e_X : X \rightarrowtail \beta_0(X)$ with dense image defined by $e_X(x) = \{K \in \mathsf{Clopen}(X) \mid x \in K\}$ (see e.g. [17, Section 4.7, Proposition (b)]) and it is easy to see that for every continuous function $f : X \to Y$ between zero-dimensional spaces the following diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{e_X} & \beta_0(X) \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle \beta_0 f} \\ Y & \xrightarrow[e_Y]{} & \beta_0(Y) \end{array}$$

There are some cases where the Čech-Stone and the Banaschewski compactifications coincide, e.g. for discrete spaces as mentioned above. More generally, we have the following:

A space is said to have *Čech-dimension zero* provided that each finite open cover admits a finite clopen refinement. In particular, every compact zero-dimensional space, and so, every Boolean space, has Čech-dimension zero.

Theorem 2.1 ([2, Theorem 4])**.** *A zero-dimensional space X is of Čech-dimension zero if and only if it is normal and the Čech-Stone and Banaschewski compactifications of X coincide.*

This is of interest to us as it shows that β is given by duality not only for discrete spaces but more generally for zero dimensional spaces such as Y^* for Y a profinite alphabet (cf. Theorem 3.5).

2.4. Projective and direct limits. A *projective limit system* (also known as an *inverse limit system*, or a *cofiltered diagram*) $\mathcal{F}$ of sets assigns to each element i of a *directed* partially ordered set I, a set S_i, and to each ordered pair $i \geq j$ in I, a map $f_{i,j}\colon S_i \to S_j$ so that for all $i, j, k \in I$ with $i \geq j \geq k$ we have $f_{i,i} = id_{S_i}$ and $f_{j,k} \circ f_{i,j} = f_{i,k}$. The *projective limit* (or *inverse limit* or *cofiltered limit*) of $\mathcal{F}$, denoted $\varprojlim \mathcal{F}$, comes equipped with projection maps $\pi_i : \varprojlim \mathcal{F} \to S_i$ compatible with the system. That is, for $i, j \in I$ with $i \geq j$, $f_{i,j} \circ \pi_i = \pi_j$. Further, it satisfies the following *universal property*: whenever $\{\pi'_i : S' \to S_i\}_{i\in I}$ is a family of maps satisfying $f_{i,j} \circ \pi'_i = \pi'_j$ for all $i \geq j$, there exists a unique map $g : S' \to \varprojlim \mathcal{F}$ satisfying $\pi'_i = \pi_i \circ g$, for all $i \in I$.

The notion dual to projective limit, obtained by reversing the directions of the maps, is that of *direct limit* (also known as an *injective limit*, *inductive limit* or *filtered colimit*).

and

$$\lambda^*_{\pi(w)}(\tau_{\mathcal{D}_1}(a)) = \lambda_{\pi(w)}(\pi(a,1)) = \pi(w(a,1)) = \pi(wa, |w|+1).$$

Therefore,

$$\begin{aligned} h(wa) &= (\eta(\pi(wa,1)\dots\pi(wa,|w|)\pi(wa,|w|+1)), \pi(wa)) \\ &= (\eta \circ \tau_{\mathcal{D}_1}(wa), \pi(wa)). \end{aligned}$$

Now, the languages recognized by $(\eta[T^*]{*}{*}M \rightarrowtail X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0})$ via h are precisely the finite unions of languages of the form

$$h^{-1}(\widehat{K} \times \widehat{L}) = \tau^{-1}_{\mathcal{D}_1} \circ \eta^{-1}(\widehat{K}) \cap \pi^{-1}(\widehat{L})$$

for some $K \in \mathcal{V}(X_{\mathcal{D}_1})$ and $L \in \mathcal{D}_0$. Since η and π are, respectively, dual to the embeddings $\mathcal{V}(X_{\mathcal{D}_1}) \rightarrowtail \mathsf{Clopen}(X^*_{\mathcal{D}_1})$ and $\mathcal{D} \rightarrowtail \mathcal{P}((A \times 2)^*)$, it follows that

$$h^{-1}(\widehat{K} \times \widehat{L}) = \tau^{-1}_{\mathcal{D}_1}(K) \cap L.$$

Thus, by Definition 6.1 of $\mathcal{V}(X_{\mathcal{D}_1}) \odot \mathcal{D}_1$, the languages recognized by h are precisely the lattice combinations of languages of $\mathcal{V}(X_{\mathcal{D}_1}) \odot \mathcal{D}_1$ and of $\mathcal{D}_0$. ■

In particular, since the set of languages recognized by a BiM via a fixed morphism is a Boolean algebra closed under quotients (cf. Section 4.1), it follows that the lattice generated by $(\mathcal{V}(X_{\mathcal{D}_1}) \odot \mathcal{D}_1) \cup \mathcal{D}_0$ is already a Boolean algebra closed under quotients.

Corollary 6.8. *Let $\mathcal{V}$ be an lp-variety of languages and $\mathcal{D} \subseteq \mathcal{P}((A \times 2)^*)$ be a Boolean subalgebra closed under quotients. Then, the lattice generated by $(\mathcal{V}(X_{\mathcal{D}_1}) \odot \mathcal{D}_1) \cup \mathcal{D}_0$ is a Boolean algebra closed under quotients.*

References

[1] J. Almeida and P. Weil. Free profinite semigroups over semidirect products. *Russian Math. (Iz. VUZ)*, 39(1):1–27, 1995.

[2] B. Banaschewski. Über nulldimensionale Räume. *Math. Nachr.*, 13:129–140, 1955.

[3] B. Banaschewski. Extensions of topological spaces. *Canad. Math. Bull.*, 7:1–22, 1964.

[4] D. A. M. Barrington, K. Compton, H. Straubing, and D. Thérien. Regular languages in NC^1. *J. Comput. System Sci.*, 44(3):478–499, 1992.

[5] C. Borlido, S. Czarnetzki, M. Gehrke, and A. Krebs. Stone duality and the substitution principle. In *Computer science logic 2017*, volume 82 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 13, 20. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.

[6] M. Furst, J. B. Saxe, and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. *Math. Systems Theory*, 17(1):13–27, 1984.

Section 2.5). To conclude that $\eta \times id$ is a monoid quotient, it suffices to observe that (26) may be rewritten as

$$(\underline{x}, m)(\underline{x}', m') = (\eta \times id)(\rho^*_{m'}(\underline{x})\lambda^*_m(\underline{x}'), mm') = (\eta \times id)((\underline{x}, m)(\underline{x}', m')). \quad \blacksquare$$

We just finish proving the following:

Proposition 6.6. *Let $\mathcal{V}$ be an lp-variety of languages and $\mathcal{D} \subseteq \mathcal{P}((A \times 2)^*)$ be a Boolean subalgebra closed under quotients. Then, we have a quotient of* BiM*s*

$$\begin{array}{ccc} T^{*}{*}{*}M & \rightarrowtail & \beta(X^*_{\mathcal{D}_1}) \times X_{\mathcal{D}_0} \\ \downarrow & & \downarrow{\scriptstyle \eta \times id} \\ \eta[T^*]{*}{*}M & \rightarrowtail & X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0} \end{array}$$

We are ready to prove the main result of this section.

Theorem 6.7. *Let $\mathcal{V}$ be an lp-variety of languages and $\mathcal{D} \subseteq \mathcal{P}((A \times 2)^*)$ be a Boolean subalgebra closed under quotients. Denote by $\pi : (A \times 2)^* \to X_{\mathcal{D}}$ the (restriction of the) map dual to the embedding $\mathcal{D} \rightarrowtail \mathcal{P}((A \times 2)^*)$, and let $h : A^* \to \eta[T^*]{*}{*}M$ be the homomorphism defined by*

$$h(a) = (\eta \circ \pi(a, 1), \pi(a)).$$

Then, a language is recognized by the BiM *$(\eta[T^*]{*}{*}M \rightarrowtail X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0})$ via h if and only if it is a lattice combination of languages of $\mathcal{V}(X_{\mathcal{D}_1}) \odot \mathcal{D}_1$ and languages of $\mathcal{D}_0$.*

Proof: We first show, by induction on the length of words, that $h(w) = (\eta \circ \tau_{\mathcal{D}_1}(w), \pi(w))$, for every $w \in A^*$. This is trivially the case if w is the empty word. Let $w \in A^*$ and $a \in A$ be a letter. Then, by induction hypothesis, we have

$$h(wa) = h(w)h(a) = (\eta \circ \tau_{\mathcal{D}_1}(w), \pi(w))(\eta \circ \tau_{\mathcal{D}_1}(a), \pi(a)),$$

and by definition of the multiplication on $(\eta[T^*]{*}{*}M)$ (cf. (26)), it follows that

$$h(wa) = (\eta(\rho^*_{\pi(a)}(\tau_{\mathcal{D}_1}(w))\lambda^*_{\pi(w)}(\tau_{\mathcal{D}_1}(a))), \pi(wa)).$$

Now, since $\tau_{\mathcal{D}_1}(w) = \pi(w, 1) \dots \pi(w, |w|)$ and $\tau_{\mathcal{D}_1}(a) = \pi(a, 1)$, we have

$$\rho^*_{\pi(a)}(\tau_{\mathcal{D}_1}(w)) = \rho_{\pi(a)} \circ \pi(w, 1) \dots \rho_{\pi(a)} \circ \pi(w, |w|) = \pi(wa, 1) \dots \pi(wa, |w|)$$

There are several things worth noting about these notions. First, the projective limit of $\mathcal{F}$ may be constructed as follows:

$$\varprojlim \mathcal{F} = \left\{ (s_i)_{i \in I} \in \prod_{i \in I} S_i \mid f_{i,j}(s_i) = s_j \text{ whenever } i \geq j \right\}. \tag{1}$$

Second, projective limits of finite sets, called *profinite sets*, are equivalent to Boolean spaces. If each S_i is finite, then it is a Boolean space in the discrete topology, and the projective limit is a closed subspace of the product and thus again a Boolean space. Conversely, a Boolean space is the projective limit of the projective system of its finite continuous quotients, corresponding via duality to the fact that Boolean algebras are locally finite and thus, direct limits of their finite subalgebras.

Third, one also has projective systems and projective limits of richer structures than sets, such as algebras, topological spaces, maps between sets, etc. In these enriched settings the connecting maps are then required to be morphisms of the appropriate kind. A very useful fact, which is used throughout this work, is that very often, and in all the settings we are interested in here, projective limits are given as for sets with the obvious enriched structure.

2.5. Biactions and semidirect products of monoids. Let S be a set and M a monoid. We denote the identity of M by 1. A *biaction* of M on S is a family of functions

$$\lambda_m : S \to S, \qquad \text{and} \qquad \rho_m : S \to S,$$

for each $m \in M$, that satisfy the following conditions:

- $\{\lambda_m\}_{m \in M}$ induces a left action of M on S, that is, λ_1 the identity function on S, and for every $m, m' \in M$, the equality $\lambda_m \circ \lambda_{m'} = \lambda_{mm'}$ holds;
- $\{\rho_m\}_{m \in M}$ induces a left action of M on S, that is, ρ_1 the identity function on S, and for every $m, m' \in M$, the equality $\rho_{m'} \circ \rho_m = \rho_{mm'}$ holds;
- $\{\lambda_m\}_{m \in M}$ and $\{\rho_m\}_{m \in M}$ are compatible, that is, for every $m, m' \in M$, we have $\lambda_m \circ \rho_{m'} = \rho_{m'} \circ \lambda_m$.

Notice that every biaction of M on S defines a function $M \times S \times M \to S$.

In the case where S is also a monoid, we have the notion of a *monoid biaction*. In order to improve readability, we shall denote the operation on S additively, although S is not assumed to be commutative. A *monoid biaction of M on S* is a biaction of M on the underlying set of S, satisfying the following additional properties:

- $m_1 \cdot (s_1 + s_2) \cdot m_2 = m_1 \cdot s_1 \cdot m_2 + m_1 \cdot s_2 \cdot m_2$, for all $m_1, m_2 \in M$, and $s_1, s_2 \in S$;
- $m_1 \cdot 0 \cdot m_2 = 0$, for all $m_1, m_2 \in M$.

Finally, given such a biaction, we may define a new monoid, called the *(two-sided) semidirect product* of S and M, usually denoted by $S{*}{*}M$. The monoid $S{*}{*}M$ has underlying set $S \times M$, and its binary operation is defined by

$$(s_1, m_1)(s_2, m_2) = (s_1 \cdot m_2 + m_1 \cdot s_2, m_1 m_2).$$

2.6. Formal languages. Let A be a set, which we call an *alphabet*. A *word over* A is an element of the A-generated free monoid A^*, and a *language* over A is a set $L \subseteq A^*$ of words, that is, an element of the powerset $\mathcal{P}(A^*)$. Whenever we write $w = a_1 \dots a_n$ for a word, we are assuming that each of the a_i's is a letter. If $w = a_1 \dots a_n$ is a word, then we say that w has *length* n, denoted $|w| = n$. The fact that A^* is a monoid means in particular that A^* is equipped with a biaction of itself. By discrete duality, it follows that $\mathcal{P}(A^*)$ also is equipped with a biaction of A^* given as follows. The *left* (respectively, *right*) *quotient* of a language L by a word w is the language $w^{-1}L = \{u \in A^* \mid wu \in L\}$ (respectively, $Lw^{-1} = \{u \in A^* \mid uw \in L\}$). We say that a set of languages is *closed under quotients* if it is invariant under this biaction.

In the discrete duality between complete and atomic Boolean algebras and sets, the complete Boolean subalgebras closed under quotients of the power set $\mathcal{P}(M)$, where M is a monoid, correspond to the monoid quotients of M [7]. In particular, given a language $L \subseteq A^*$, the discrete dual of $\mathcal{B}_L$, the complete Boolean subalgebra closed under quotients generated by L, is a monoid quotient $\mu_L : A^* \twoheadrightarrow M_L$. The monoid M_L is called the *syntactic monoid of* L and μ_L is the *syntactic homomorphism* of L. The quotient map μ_L is given by the corresponding congruence relation, known as the *syntactic congruence of* L, which, by the definition of μ_L, is given by

$$u \sim_L v \iff (\forall x, y \in A^*,\ xuy \in L \iff xvy \in L).$$

Notice that it follows that $L = \mu_L^{-1}(\mu_L[L])$, that is, L *is recognized by* M_L *via* μ_L, see below. More generally, for a set of languages $\mathcal{S}$, the *syntactic monoid, homomorphism, and congruence of* $\mathcal{S}$ are defined via discrete duality for the complete Boolean subalgebra closed under quotients generated by $\mathcal{S}$.

Given a monoid M, a language $L \subseteq A^*$ is said to be *recognized* by M provided there exists a homomorphism $\mu\colon A^* \to M$ and a subset $P \subseteq M$

and the restriction and co-restriction of η to a map $X^*_{\mathcal{D}_1} \twoheadrightarrow \eta[X^*_{\mathcal{D}_1}]$ is a monoid quotient. Since T^* is a submonoid of $X^*_{\mathcal{D}_1}$, we have that $\eta[T^*]$ is also a monoid. We will show that M biacts on $\eta[T^*]$ and that N is the semidirect product $\eta[T^*]{*}{*}M$ defined by this biaction (Lemma 6.5). In fact, we show the following slightly more general fact: M biacts on $\eta[X^*_{\mathcal{D}_1}]$ and the ensuing semidirect product $\eta[X^*_{\mathcal{D}_1}]{*}{*}M$ is a monoid quotient of $X_{\mathcal{D}_1}{*}{*}M$ (recall that M biacts on the free monoid $X^*_{\mathcal{D}_1}$ so that we have a semidirect product $X^*_{\mathcal{D}_1}{*}{*}M$ of which $T^*{*}{*}M$ is a submonoid, cf. (19)).

Lemma 6.5. *For every $m \in M$, the assignments*

$$\ell_m : \eta[X^*_{\mathcal{D}_1}] \to \eta[X^*_{\mathcal{D}_1}], \qquad \eta(\underline{x}) \mapsto \eta(\lambda^*_m(\underline{x}))$$

and

$$r_m : \eta[X^*_{\mathcal{D}_1}] \to \eta[X^*_{\mathcal{D}_1}], \qquad \eta(\underline{x}) \mapsto \eta(\rho^*_m(\underline{x}))$$

are well-defined functions, which define a monoid biaction of M on $\eta[X^*_{\mathcal{D}_1}]$. In particular, there is a semidirect product $\eta[X^*_{\mathcal{D}_1}]{*}{*}M$ whose multiplication is defined by

$$(\eta(\underline{x}), m)(\eta(\underline{x}'), m') = (\eta(\rho^*_{m'}(\underline{x})\lambda^*_m(\underline{x}')), mm') \tag{26}$$

for every $(\underline{x}, m), (\underline{x}', m') \in X^*_{\mathcal{D}_1}{*}{*}M$, and $(\eta \times id) : X^*_{\mathcal{D}_1}{*}{*}M \twoheadrightarrow \eta[X^*_{\mathcal{D}_1}]{*}{*}M$ is a monoid quotient.

Proof: We first show that ℓ_m is well-defined. Let $\underline{x}, \underline{x}' \in X^*_{\mathcal{D}_1}$ be such that $\eta(\underline{x}) = \eta(\underline{x}')$. We need to show that $\eta(\lambda^*_m(\underline{x})) = \eta(\lambda^*_m(\underline{x}'))$. By definition of dual map, having $\eta(\underline{x}) = \eta(\underline{x}')$ is equivalent to having that, for every $K \in \mathcal{V}(X_{\mathcal{D}_1})$,

$$\underline{x} \in K \iff \underline{x}' \in K.$$

Since λ_m is a continuous function and $\mathcal{V}$ is an lp-strain of languages, by Lemma 3.6, we then have that, for every $K \in \mathcal{V}(X_{\mathcal{D}_1})$,

$$\underline{x} \in (\lambda^*_m)^{-1}(K) \iff \underline{x}' \in (\lambda^*_m)^{-1}(K),$$

and this is equivalent to having $\eta(\lambda^*_m(\underline{x})) = \eta(\lambda^*_m(\underline{x}'))$ as required.

Similarly, one can show that r_m is well-defined. The fact that $\{\ell_m, r_m\}_{m \in M}$ defines a monoid biaction is inherited from the fact that $\{\lambda_m, \rho_m\}_{m \in M}$ defines a monoid biaction.

Finally, the fact that the multiplication on $\eta[X^*_{\mathcal{D}_1}]{*}{*}M$ is given by (26) is a straighforward consequence of the definition of semidirect products (cf.

subset of $X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0}$. In the proof of Proposition 6.3 we have already seen that, if $\widetilde{\lambda}^{-1}_{(\underline{t},m)}(K)$ is nonempty, then it is given by

$$(\widetilde{\lambda}_{(\underline{t},m)})^{-1}(K) = (\lambda_m^{-1}(\widehat{L}_{k+1}) \times \cdots \times \lambda_m^{-1}(\widehat{L}_n)) \times (\lambda_m^{-1}(\widehat{L}_0) \cap \bigcap_{i=1}^{k} \lambda_{t_i}^{-1}(\widehat{L}_i)),$$

and we argued that $(\lambda_m^{-1}(\widehat{L}_0) \cap \bigcap_{i=1}^k \lambda_{t_i}^{-1}(\widehat{L}_i))$ is a clopen subset of $X_{\mathcal{D}_0}$ (see (24)). We now need to show that

$$L := \lambda_m^{-1}(\widehat{L}_{k+1}) \times \cdots \times \lambda_m^{-1}(\widehat{L}_n)$$

is a clopen subset of $X_{\mathcal{V}(X_{\mathcal{D}_1})}$. First note that L may be rewritten as

$$L = (\lambda_m^*)^{-1}(\widehat{L}_{k+1} \times \cdots \times \widehat{L}_n) = (\lambda_m^*)^{-1}(\widehat{L'}),$$

where $L' := L_{k+1} \times \cdots \times L_n$. We first show that L' belongs to $\mathcal{V}(X_{\mathcal{D}_1})$. For this, pick any word $w \in L_1 \times \cdots \times L_k$. Then, since $L_1 \times \cdots \times L_n$ belongs to $\mathcal{V}(X_{\mathcal{D}_1})$, by Lemma 3.1, the quotient $w^{-1}(L_1 \times \cdots \times L_n)$ also does. But this quotient is precisely L'. Indeed, for every word v, we have

$$v \in w^{-1}(L_1 \times \cdots \times L_n) \iff wv \in L_1 \times \cdots \times L_n \iff v \in L_{k+1} \times \cdots \times L_n,$$

where the last equivalence follows from the choice of w. Finally, by Corollary 4.6, we have a continuous function $\lambda_m : X_{\mathcal{D}_1} \to X_{\mathcal{D}_1}$ and, by Lemma 3.6, we may conclude that $(\lambda_m^*)^{-1}(\widehat{L'})$ is a clopen subset of $X_{\mathcal{V}(X_{\mathcal{D}_1})}$, as required.

Similarly, one can show the existence of a continuous function $\rho_{(\underline{t},m)}$ making the right-hand side diagram of (25) commute. The fact that $\{\lambda_{(\underline{t},m)}, \rho_{(\underline{t},m)}\}_{(\underline{t},m)}$ defines a biaction of $T^{*}{*}{*}M$ on $X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0}$ follows from having that $\{\widetilde{\lambda}_{(\underline{t},m)}, \widetilde{\rho}_{(\underline{t},m)}\}_{(\underline{t},m)}$ defines a biaction of $T^{*}{*}{*}M$ on $X^*_{\mathcal{D}_1} \times X_{\mathcal{D}_0}$ (cf. Lemma 6.2) and that each homomorphism $\lambda^{-1}_{(\underline{t},m)}$ (respectively, $\rho^{-1}_{(\underline{t},m)}$) is a suitable restriction and co-restriction of $\widetilde{\lambda}^{-1}_{(\underline{t},m)}$ (respectively, $\widetilde{\rho}^{-1}_{(\underline{t},m)}$). ∎

By Lemma 6.4, we have that the restriction of $(\eta \times id) : \beta(X^*_{\mathcal{D}_1}) \times X_{\mathcal{D}_0} \to X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0}$ to $T^{*}{*}{*}M$ is a morphism of sets with $(T^{*}{*}{*}M)$-biactions. Therefore, $N := (\eta \times id)[T^{*}{*}{*}M]$ comes equipped with a monoid structure induced by the monoid structure of $T^{*}{*}{*}M$ and we have a BiM $(N \rightarrowtail X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0})$ which is a quotient of $(T^{*}{*}{*}M \rightarrowtail X^*_{\mathcal{D}_1} \times X_{\mathcal{D}_0})$ as in (21). We will now give a precise description of the monoid N. Note that the underlying set of N is $\eta[T^*] \times M$. Moreover, since $\mathcal{V}$ is an lp-variety, by Lemma 3.1, $\mathcal{V}(X^*_{\mathcal{D}_1})$ is closed under quotients and thus, $\eta[X^*_{\mathcal{D}_1}]$ is a monoid

satisfying $L = \mu^{-1}(P)$. By duality, the set of languages recognized by μ is a complete Boolean subalgebra closed under quotients which contains L. Since it must contain $\mathcal{B}_L$, the syntactic homomorphism of L factors through any homomorphism μ which recognizes L. That is, the syntactic monoid and homomorphism of L can be thought as the "optimal" recognizer of L . Finally, *regular languages* are those recognized by *finite* monoids. It is not hard to see that regular languages over an alphabet A form a Boolean algebra closed under quotients. Beyond regularity, the discrete notion of recognition introduced here is not adequate. In particular, any infinite monoid recognizes uncountably many languages. Recognition of Boolean algebras of (not-necessarily regular) languages closed under quotients require the introduction of topology. This will be addressed in Section 4.

2.7. lp-strains and lp-varieties of languages. A homomorphism between free monoids is said to be *length-preserving* (also called an *lp-morphism*) provided it maps generators to generators. Therefore, lp-morphisms $A^* \to B^*$ are in a bijection with functions $A \to B$. Given a map $h : A \to B$, we use h^* to denote the unique lp-morphism $h^* : A^* \to B^*$ whose restriction to A is h. As we will see in the paragraph on "classes of sentences", lp-morphisms are important in the treatment of fragments of logic on words.

Combining Pippenger [16] and Straubing's [19] terminology, we call *lp-strain of languages* an assignment, for each finite alphabet A, of a Boolean algebra $\mathcal{V}(A)$ of languages over A such that, for every lp-morphism $h^* : B^* \to A^*$, if $L \in \mathcal{V}(A)$ then $(h^*)^{-1}(L) \in \mathcal{V}(B)$. Since $(h \circ g)^* = h^* \circ g^*$ whenever f anf g are composable set functions, each *lp*-strain of languages defines a presheaf $\mathcal{V} : \mathbf{Set}^{\mathrm{op}}_{fin} \to \mathbf{BA}$ over Boolean algebras. This notion allows us to extend languages to profinite alphabets (cf. Section 3), which is crucial when handling infinite Boolean algebras of formulas (cf. Section 5.2). Notice that, if $\mathcal{V}$ is an lp-strain of languages, then every function $h : B \to A$ induces dually a continuous function $\widehat{h} : X_{\mathcal{V}(B)} \to X_{\mathcal{V}(A)}$ making the following diagram commute:

$$\begin{array}{ccc} B^* & \longrightarrow & X_{\mathcal{V}(B)} \\ {\scriptstyle h^*}\downarrow & & \vdots{\scriptstyle \widehat{h}} \\ A^* & \longrightarrow & X_{\mathcal{V}(A)} \end{array}$$

where $A^* \to X_{\mathcal{V}(A)}$ and $B^* \to X_{\mathcal{V}(B)}$ are the maps with dense image obtained by dualizing and restricting the embeddings $\mathcal{V}(A) \rightarrowtail \mathcal{P}(A^*)$ and $\mathcal{V}(B) \rightarrowtail$

$\mathcal{P}(B^*)$, respectively. This is a particular case of Lemma 3.6, for which we will provide a detailed proof.

In the case where $\mathcal{V}(A)$ is closed under quotients for every alphabet A, we call $\mathcal{V}$ an *lp-variety of languages*.

So far, we did not require A to be a finite set. In this paper, we will consider languages over profinite alphabets, where the topology of the alphabet will be taken into account (finite alphabets are simply viewed as discrete spaces). Languages over profinite alphabets will be the subject of Section 3. Unless specified otherwise, we will use the letters $A, B, \dots$ to denote finite alphabets, and $X, Y, \dots$ for profinite (not necessarily finite) ones.

2.8. Logic on words. We shall consider classes of languages that are definable in fragments of *first-order logic*. *Variables* are denoted by $x, y, z, \dots$, and we fix a finite alphabet A. Our logic signature consists of:

- (unary) *letter predicates* P_a, one for each letter $a \in A$;
- a set $\mathcal{N}$ of *numerical predicates* R, each of finite arity. A *k-ary numerical predicate* R is given by a subset $R \subseteq \mathbb{N}^k$. For instance, the usual (binary) numerical predicate $<$ formally corresponds to the predicate $R = \{(i,j) \mid i,j \in \mathbb{N},\ i < j\}$;
- a set $\mathcal{Q}$ of (unary) *quantifiers* Q, each one being given by a function $Q : \{0,1\}^* \to \{0,1\}$. For instance, the existential quantifier $\exists$ corresponds to the function mapping $\varepsilon_1 \dots \varepsilon_n \in \{0,1\}^*$ to 1 exactly when there exists an i so that $\varepsilon_i = 1$.

Then, *(first-order) formulas* are recursively built as follows:

- for each *letter predicate* P_a and variable x, there is an atomic formula $\mathsf{P}_a(x)$;
- if $R \in \mathcal{N}$ is a k-ary numerical predicate and $x_1, \dots, x_k$ are (possibly non-distinct) variables, then $R(x_1, \dots, x_k)$ is an atomic formula;
- *Boolean combinations* of formulas are formulas, that is, if ϕ and ψ are formulas, then so are $\phi \wedge \psi$, $\phi \vee \psi$, and $\neg\phi$;
- if Q is a quantifier, x a variable and ϕ a formula, then $Qx\ \phi$ is also a formula.

An occurrence of a variable x in a formula is said to be *free* provided it appears outside of the scope of a quantifier Qx. A *sentence* is a formula with no free occurrences of variables.

A *context* $\mathbf{x}$ is a finite set of distinct variables. Given disjoint contexts $\mathbf{x}$ and $\mathbf{y}$, we denote by $\mathbf{xy}$ the union of $\mathbf{x}$ and $\mathbf{y}$. Whenever we mention union

Therefore, we may conclude that either $(\widetilde{\lambda}_{(\underline{t},m)})^{-1}(K) = \emptyset$ (in the case where $n < k$) or else we have

$$(\widetilde{\lambda}_{(\underline{t},m)})^{-1}(K) = (\lambda_m^{-1}(\widehat{L}_{k+1}) \times \cdots \times \lambda_m^{-1}(\widehat{L}_n)) \times (\lambda_m^{-1}(\widehat{L}_0) \cap \bigcap_{i=1}^{k} \lambda_{t_i}^{-1}(\widehat{L}_i)). \quad (24)$$

Here, we follow the convention that, if $n = k$, then $(\lambda_m^{-1}(\widehat{L}_{k+1}) \times \cdots \times \lambda_m^{-1}(\widehat{L}_n))$ is the singleton consisting of the empty word $\varepsilon \in X_{\mathcal{D}_1}^*$. Finally, by Corollary 4.6, we have that $\lambda_m^{-1}(\widehat{L}_{k+j}) \in \mathsf{Clopen}(X_{\mathcal{D}_1})$ $(j \in \{1, \dots, n-k\})$, $\lambda_m^{-1}(\widehat{L}_0) \in \mathsf{Clopen}(X_{\mathcal{D}_0})$, and $\lambda_{t_i}^{-1}(\widehat{L}_i) \in \mathsf{Clopen}(X_{\mathcal{D}_0})$ $(i \in \{1, \dots, k\})$. Therefore, $(\widetilde{\lambda}_{(\underline{t},m)})^{-1}(K)$ is indeed a clopen subset of $X_{\mathcal{D}_1}^* \times X_{\mathcal{D}_0}$. Likewise, one can show that $(\widetilde{\rho}_{(\underline{t},m)})^{-1}(K)$ is also clopen. ∎

As promised, we now prove that the continuous quotient $(\eta \times id) : \beta(X_{\mathcal{D}_1}^*) \times X_{\mathcal{D}_0} \twoheadrightarrow X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0}$ induces a BiM quotient as in (21). This will be a consequence of the following lemma:

Lemma 6.4. *Let $\mathcal{V}$ be an lp-variety of languages and $\mathcal{D} \subseteq \mathcal{P}((A \times 2)^*)$ be a Boolean subalgebra closed under quotients. Then, for every $(\underline{t}, m) \in T^{*}{*}{*}M$, there are continuous function $\lambda_{(\underline{t},m)}$ and $\rho_{(\underline{t},m)}$ making the following diagrams commute:*

$$\begin{array}{ccc} \beta(X_{\mathcal{D}_1}^*) \times X_{\mathcal{D}_0} & \xrightarrow{\eta \times id} & X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0} \\ \downarrow{\scriptstyle \widetilde{\lambda}_{(\underline{t},m)}} & & \vdots{\scriptstyle \lambda_{(\underline{t},m)}} \\ \beta(X_{\mathcal{D}_1}^*) \times X_{\mathcal{D}_0} & \xrightarrow[\eta \times id]{} & X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0} \end{array} \qquad \begin{array}{ccc} \beta(X_{\mathcal{D}_1}^*) \times X_{\mathcal{D}_0} & \xrightarrow{\eta \times id} & X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0} \\ \downarrow{\scriptstyle \widetilde{\rho}_{(\underline{t},m)}} & & \vdots{\scriptstyle \rho_{(\underline{t},m)}} \\ \beta(X_{\mathcal{D}_1}^*) \times X_{\mathcal{D}_0} & \xrightarrow[\eta \times id]{} & X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0} \end{array} \quad (25)$$

where $\widetilde{\lambda}_{(\underline{t},m)}$ and $\widetilde{\rho}_{(\underline{t},m)}$ are, respectively, the left and right components at $(\underline{t}, m)$ of the biaction of $T^{*}{*}{*}M$ on $\beta(X_{\mathcal{D}_1}^*) \times X_{\mathcal{D}_0}$ (cf. Proposition 6.3). Moreover, the family $\{\lambda_{(\underline{t},m)}, \rho_{(\underline{t},m)}\}_{(\underline{t},m)}$ defines a biaction of $T^{*}{*}{*}M$ on $X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0}$.

Proof: By considering the dual of the left-hand side diagram of (25), in order to show that a continuous function $\lambda_{(\underline{t},m)}$ exists, it suffices to show that $\widetilde{\lambda}^{-1}_{(\underline{t},m)}$ restricts and co-restricts to $\mathsf{Clopen}(X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0})$. Let $K := \widehat{L}_1 \times \dots \widehat{L}_n \times \widehat{L}_0$ be a clopen subset of $X_{\mathcal{D}_1}^* \times X_{\mathcal{D}_0}$, for some $L_1, \dots, L_n \in \mathcal{D}_1$ and $L_0 \in \mathcal{D}_0$, and assume that $L_1 \times \cdots \times L_n$ belongs to $\mathcal{V}(X_{\mathcal{D}_1})$, so that K is actually a clopen

Here, every equality follows from the appropriate definitions, except for the equality marked with $(*)$, which uses the fact that $\{\lambda_m\}_{m\in M_{\mathcal{D}}}$ is itself a left action. ∎

We now see that the biaction defined in Lemma 6.2 further extends to a biaction of $T^{*}{*}{*}M$ on $\beta(X^{*}_{\mathcal{D}_1}) \times X_{\mathcal{D}_0}$ with continuous components, thereby defining a BiM

$$T^{*}{*}{*}M \rightarrowtail \beta(X^{*}_{\mathcal{D}_1}) \times X_{\mathcal{D}_0}.$$

Once again, the main ingredient for showing this is duality – since $X^{*}_{\mathcal{D}_1} \times X_{\mathcal{D}_0}$ and $\beta(X^{*}_{\mathcal{D}_1}) \times X_{\mathcal{D}_0}$ have the same clopen subsets (cf. Theorem 3.5), to show that $\widetilde{\lambda}_{(\underline{t},m)}$ and $\widetilde{\rho}_{(\underline{t},m)}$ continuously extend to $\beta(X^{*}_{\mathcal{D}_1}) \times X_{\mathcal{D}_0}$ it suffices to show that, for every $(\underline{t}, m) \in T^{*}{*}{*}M$, the maps $(\widetilde{\lambda}_{(\underline{t},m)})^{-1}$ and $(\widetilde{\rho}_{(\underline{t},m)})^{-1}$ define endofunctions of $\mathsf{Clopen}(X^{*}_{\mathcal{D}_1} \times X_{\mathcal{D}_0})$. Then, the desired extension will be given by the dual maps of $(\widetilde{\lambda}_{(\underline{t},m)})^{-1}$ and of $(\widetilde{\rho}_{(\underline{t},m)})^{-1}$.

Proposition 6.3. *The biaction of $T^{*}{*}{*}M$ on $X^{*}_{\mathcal{D}_1} \times X_{\mathcal{D}_0}$ extends to a biaction of $T^{*}{*}{*}M$ on $\beta(X^{*}_{\mathcal{D}_1}) \times X_{\mathcal{D}_0}$ with continuous components, and the inclusion $T^{*}{*}{*}M \rightarrowtail \beta(X^{*}_{\mathcal{D}_1}) \times X_{\mathcal{D}_0}$ admits a BiM structure with respect to such biaction.*

Proof: We first observe that the embedding $T^{*}{*}{*}M \rightarrowtail \beta(X^{*}_{\mathcal{D}_1}) \times X_{\mathcal{D}_0}$ has dense image. Indeed, this is simply because T and M are, respectively, dense subsets of $X_{\mathcal{D}_1}$ and of $X_{\mathcal{D}_0}$. Let us show that that there is a biaction of $T^{*}{*}{*}M$ on $\beta(X^{*}_{\mathcal{D}_1}) \times X_{\mathcal{D}_0}$ with continuous components. As already explained, this is the case provided the preimage under $\widetilde{\lambda}_{(\underline{t},m)}$ and under $\widetilde{\rho}_{(\underline{t},m)}$ of every clopen subset of $X^{*}_{\mathcal{D}_1} \times X_{\mathcal{D}_0}$ is again a clopen subset. Since the topological space $X^{*}_{\mathcal{D}_1}$ is the union over $n \geq 0$ of the product spaces $X^{n}_{\mathcal{D}_1}$, it suffices to consider clopen subsets of the form $K := \widehat{L}_1 \times \cdots \times \widehat{L}_n \times \widehat{L}_0$ for some $L_1, \ldots, L_n \in \mathcal{D}_1$ and $L_0 \in \mathcal{D}_0$. We let $(x_1 \ldots x_\ell, x_0) \in X^{*}_{\mathcal{D}_1} \times X_{\mathcal{D}_0}$, $\underline{t} = t_1 \ldots t_k \in T^{*}$, and $m \in M$. Then, we have $x_1 \ldots x_\ell, x_0) \in (\widetilde{\lambda}_{(\underline{t},m)})^{-1}(K)$ if and only if

$$(\lambda_{t_1}(x_0) \ldots \lambda_{t_k}(x_0)\lambda_m(x_1) \ldots \lambda_m(x_\ell),\, \lambda_m(x_0)) \in K.$$

Using the definition of K, we have that $(\underline{x}, x_0) \in (\widetilde{\lambda}_{(\underline{t},m)})^{-1}(K)$ if and only if $k + \ell = n$, and the following conditions hold:

- $\lambda_{t_i}(x_0) \in \widehat{L}_i$, for every $i \in \{1, \ldots, k\}$;
- $\lambda_m(x_j) \in \widehat{L}_{k+j}$, for every $j \in \{1, \ldots, \ell\}$;
- $\lambda_m(x_0) \in \widehat{L}_0$.

of contexts, we will assume they are disjoint without further mention. Also, we simply write x to refer to the context $\{x\}$, so that by $\mathbf{x}x$ we mean the context $\mathbf{x} \cup \{x\}$. We denote by $|\mathbf{x}|$ the cardinality of $\mathbf{x}$. The models of a formula will always be considered in a fixed context. We say that ϕ *is in context* $\mathbf{x}$ provided all the free variables of ϕ belong to $\mathbf{x}$. Notice that, if ϕ is in context $\mathbf{x}$, then it is also in every other context containing $\mathbf{x}$.

Models of sentences in the empty context are words over A. More generally, models of formulas in the context $\mathbf{x}$ are given by $|\mathbf{x}|$-*marked words*, that is, words $w \in A^*$ equipped with an interpretation $\mathbf{i} \in \{1, \ldots, |w|\}^{\mathbf{x}}$ of the context $\mathbf{x}$. We identify maps from $\mathbf{x}$ to $\{1, \ldots, |w|\}$ with $|\mathbf{x}|$-tuples $\mathbf{i} = (i_1, \ldots, i_{|\mathbf{x}|}) \in \{1, \ldots, |w|\}^{|\mathbf{x}|} \subseteq \mathbb{N}^{|\mathbf{x}|}$. We denote the set of all $|\mathbf{x}|$-marked words by $A^* \otimes \mathbb{N}^{|\mathbf{x}|}$. Given a word $w \in A^*$ and a vector $\mathbf{i} = (i_1, \ldots, i_{|\mathbf{x}|}) \in \{1, \ldots, |w|\}^{|\mathbf{x}|}$, we denote by $(w, \mathbf{i})$ the marked word based on w equipped with the map given by $\mathbf{i}$. Moreover, if $\mathbf{x}$ and $\mathbf{y}$ are disjoint contexts, $\mathbf{i} = (i_1, \ldots, i_{|\mathbf{x}|}) \in \{1, \ldots, |w|\}^{|\mathbf{x}|}$ and $\mathbf{j} = (j_1, \ldots, j_{|\mathbf{y}|}) \in \{1, \ldots, |w|\}^{|\mathbf{y}|}$, then $(w, \mathbf{i}, \mathbf{j})$ denotes the $|\mathbf{z}|$-marked word $(w, \mathbf{k})$, where $\mathbf{z} = \mathbf{x}\mathbf{y}$ and $\mathbf{k} = (i_1, \ldots, i_{|\mathbf{x}|}, j_1, \ldots, j_{|\mathbf{y}|})$.

The semantics of a formula in context $\mathbf{x} = \{x_1, \ldots, x_k\}$ is defined inductively as follows. We let $w \in A^*$ be a word and $\mathbf{i} = (i_1, \ldots, i_k) \in \{1, \ldots, |w|\}^{|\mathbf{x}|}$. Then,

- the marked word $(w, \mathbf{i})$ satisfies $\mathsf{P}_a(x_j)$ if and only if its i_j-th letter is an a. As a consequence, for all $j \in \{1, \ldots, k\}$, there exists exactly one letter $a \in A$ such that $(w, \mathbf{i}) \models \mathsf{P}_a(x_j)$;
- the marked word $(w, \mathbf{i})$ satisfies the formula $R(x_{j_1}, \ldots, x_{j_\ell})$ given by a numerical predicate R if and only if $(i_{j_1}, \ldots, i_{j_\ell}) \in R$;
- the Boolean connectives *and* $\wedge$, *or* $\vee$, and *not* $\neg$ are interpreted classically;
- the marked word $(w, \mathbf{i})$ satisfies the formula $Qx\ \phi$, for a *quantifier* $Q : \{0,1\}^* \to \{0,1\}$, if and only if $Q((w, \mathbf{i}, 1) \models \phi, \ldots, (w, \mathbf{i}, |w|) \models \phi) = 1$, where "$(w, \mathbf{i}, i) \models \phi$" denotes the truth-value of "$(w, \mathbf{i}, i)$ satisfies ϕ".

Important examples of quantifiers are given by the *existential quantifier* $\exists$ mentioned above and its variant $\exists!$ (*there exists a unique*); *modular quantifiers* $\exists^r_q$, for each $q \in \mathbb{N}$ and $0 \leq r < q$, mapping a word of $\{0,1\}^*$ to 1 if and only if its number of 1's is congruent to r modulo q; and the *majority quantifier* Maj that sends an element of $\{0,1\}^*$ to 1 if and only if it has strictly more occurrences of 1 than of 0. Finally, given a formula ϕ in context $\mathbf{x}$, we denote by L_ϕ the set of marked words that are models of ϕ.

Given a finite alphabet A, a context $\mathbf{x}$, a set of numerical predicates $\mathcal{N}$, and a set of quantifiers $\mathcal{Q}$, we denote by $\mathcal{Q}_{A,\mathbf{x}}[\mathcal{N}]$ the corresponding set of first-order formulas in context $\mathbf{x}$, up to semantic equivalence. The set of sentences $\mathcal{Q}_{A,\emptyset}[\mathcal{N}]$ in the empty context is simply denoted by $\mathcal{Q}_A[\mathcal{N}]$. Notice that, as long as $\mathcal{Q}_{A,\mathbf{x}}[\mathcal{N}]$ is non-empty, say it contains a formula ϕ, it will also contain the *always-true* formula $\mathbf{1}$, which is semantically equivalent to $\phi \vee \neg\phi$, and the *always-false* formula $\mathbf{0}$, which is semantically equivalent to $\phi \wedge \neg\phi$. In particular, we have $L_{\mathbf{1}} = A^* \otimes \mathbb{N}^{|\mathbf{x}|}$ and $L_{\mathbf{0}} = \emptyset$, and taking the set of models of a given formula defines an embedding of Boolean algebras

$$\mathcal{Q}_{A,\mathbf{x}}[\mathcal{N}] \rightarrowtail \mathcal{P}(A^* \otimes \mathbb{N}^{|\mathbf{x}|}).$$

In formal language theory one usually considers languages that are subsets of a free monoid, as presented in the previous paragraph. However, formulas in a non-empty context $\mathbf{x} = \{x_1, \ldots, x_k\}$ define languages of marked words. Nevertheless, there is a natural injection $A^* \otimes \mathbb{N}^{|\mathbf{x}|} \rightarrowtail (A \times 2^{\mathbf{y}})^*$ whenever $\mathbf{y}$ is a context containing $\mathbf{x}$. Here, we make a slightly abuse of notation by writing $2^{\mathbf{y}}$ to actually mean the set $\mathcal{P}(\mathbf{y})$. Indeed, let $(w, \mathbf{i})$ be a marked word, where $w = a_1 \ldots a_n$ and $\mathbf{i} = (i_1, \ldots, i_k)$. Then, $(w, \mathbf{i})$ is completely determined by the word $(a_1, S_1) \ldots (a_n, S_n)$ over $(A \times 2^{\mathbf{y}})$, where $S_i = \{x_j \mid i_j = i\}$. Thus, we will often regard the language defined by a formula in a context $\mathbf{x}$ as a language over an extended alphabet $A \times 2^{\mathbf{y}}$ for some $\mathbf{y} \supseteq \mathbf{x}$.

2.9. Marked words. For simplicity we consider here the case where $\mathbf{x}$ consists of a single variable x, though a similar treatment is possible in general. In particular,

$$A^* \otimes \mathbb{N} = \{(w, i) \mid w \in A^*,\ i \in \{1, \ldots, |w|\}\}.$$

and we have an embedding $A^* \otimes \mathbb{N} \rightarrowtail (A \times 2)^*$ defined by

$$(w, i) \mapsto (a_1, 0) \ldots (a_{i-1}, 0)(a_i, 1)(a_{i+1}, 0) \ldots (a_n, 0),$$

for every word $w = a_1 \ldots a_n$ and $i \in \{1, \ldots, n\}$.

Identifying the elements of $A^* \otimes \mathbb{N}$ with this image in $(A \times 2)^*$, we may compute the Boolean algebra closed under quotients generated by this language. It is not difficult to see that $A^* \otimes \mathbb{N}$ is a regular language in $(A \times 2)^*$ and that its syntactic monoid is the three-element commutative monoid $\{e, m, z\}$ satisfying $m^2 = z$, with z acting as zero and e its identity element. The syntactic morphism of the language $A^* \otimes \mathbb{N}$ is $\mu : (A \times 2)^* \twoheadrightarrow \{e, m, z\}$ given

(20) may be rewritten as follows:

$$(\underline{t}, m)(\underline{t}', m') = (\lambda_{t_1}(m') \ldots \lambda_{t_k}(m')\lambda_m(t'_1) \ldots \lambda_m(t'_\ell),\ mm').$$

This provides a natural way of defining an element $\lambda_{(\underline{t},m)}(\underline{x}, x_0)$ when $(\underline{x}, x_0)$ belongs to $X^*_{\mathcal{D}_1} \times X_{\mathcal{D}_0}$, namely,

$$\widetilde{\lambda}_{(\underline{t},m)}(\underline{x}, x_0) := (\lambda_{t_1}(x_0) \ldots \lambda_{t_k}(x_0)\lambda_m(x_1) \ldots \lambda_m(x_\ell),\ \lambda_m(x_0)), \tag{22}$$

where $\underline{x} = x_1 \ldots x_\ell$. Similarly, we define

$$\widetilde{\rho}_{(\underline{t},m)}(\underline{x}, x_0) := (\rho_m(x_1) \ldots \rho_m(x_\ell)\rho_{t_1}(x_0) \ldots \rho_{t_k}(x_0),\ \rho_m(x_0)). \tag{23}$$

Lemma 6.2. *The functions*

$$\begin{aligned}\widetilde{\lambda}_{(\underline{t},m)} &: X^*_{\mathcal{D}_1} \times X_{\mathcal{D}_0} \to X^*_{\mathcal{D}_1} \times X_{\mathcal{D}_0} \\ &(x_1 \ldots x_\ell, x_0) \mapsto (\lambda_{t_1}(x_0) \ldots \lambda_{t_k}(x_0)\lambda_m(x_1) \ldots \lambda_m(x_\ell),\ \lambda_m(x_0)),\end{aligned}$$

and

$$\begin{aligned}\widetilde{\rho}_{(\underline{t},m)} &: X^*_{\mathcal{D}_1} \times X_{\mathcal{D}_0} \to X^*_{\mathcal{D}_1} \times X_{\mathcal{D}_0} \\ &(x_1 \ldots x_\ell, x_0) \mapsto (\rho_m(x_1) \ldots \rho_m(x_\ell)\rho_{t_1}(x_0) \ldots \rho_{t_k}(x_0),\ \rho_m(x_0)),\end{aligned}$$

define a biaction of T^**_*M on on the space $X^*_{\mathcal{D}_1} \times X_{\mathcal{D}_0}$.

Proof: This is a consequence of the fact that the family $\{\lambda_m, \rho_m\}_{m \in M_{\mathcal{D}}}$ defines a biaction of $M_{\mathcal{D}}$ on $X_{\mathcal{D}}$. We illustrate the computations involved by showing that $\widetilde{\lambda}_{(\underline{t},m)} \circ \widetilde{\lambda}_{(\underline{t}',m')} = \widetilde{\lambda}_{(\underline{t},m)(\underline{t}',m')}$ for every $(\underline{t}, m), (\underline{t}', m') \in T^**_*M$, and leave the rest for the reader. Let us write $\underline{t} = t_1 \ldots t_k$ and $\underline{t}' = t'_1 \ldots t'_\ell$, and pick $(\underline{x}, x_0) \in X^*_{\mathcal{D}_1} \times X_{\mathcal{D}_0}$. We will use the following notation: if $f_1, \ldots, f_i$ are functions $P \to Q$ then, for every $p \in P$, we denote by $\langle f_1, \ldots, f_i \rangle(p)$ the word $f_1(p) \ldots f_i(p)$ over Q. Then, we may compute

$$\begin{aligned}&\widetilde{\lambda}_{(\underline{t},m)} \circ \widetilde{\lambda}_{(\underline{t}',m')}(\underline{x}, x_0) \\ &= \widetilde{\lambda}_{(\underline{t},m)}(\langle \lambda_{t'_1}, \ldots, \lambda_{t'_\ell} \rangle(x_0)\, \lambda^*_{m'}(\underline{x}), \lambda_{m'}(x_0)) \\ &= (\langle \lambda_{t_1}, \ldots \lambda_{t_k} \rangle(\lambda_{m'}(x_0))\lambda^*_m(\langle \lambda_{t'_1}, \ldots, \lambda_{t'_\ell} \rangle(x_0)\, \lambda^*_{m'}(\underline{x})), \lambda_m(\lambda_{m'}(x_0))) \\ &= (\langle \lambda_{t_1}\lambda_{m'}, \ldots \lambda_{t_k}\lambda_{m'}, \lambda_m\lambda_{t'_1}, \ldots, \lambda_m\lambda_{t'_\ell} \rangle(x_0)\, \lambda^*_m(\lambda^*_{m'}(\underline{x})), \lambda_m(\lambda_{m'}(x_0))) \\ &\overset{(*)}{=} (\langle \lambda_{t_1 m'}, \ldots \lambda_{t_k m'}, \lambda_{m t'_1}, \ldots, \lambda_{m t'_\ell} \rangle(x_0)\, \lambda^*_{mm'}(\underline{x}), \lambda_{mm'}(x_0)) \\ &= \widetilde{\lambda}_{(\underline{t},m)(\underline{t}',m')}(\underline{x}, x_0).\end{aligned}$$

Now, for each $m \in M_{\mathcal{D}}$, we let λ_m and ρ_m denote, respectively, the left and right components of the biaction of $M_{\mathcal{D}}$ on $X_{\mathcal{D}}$. Then, there is also a biaction of M on $X^*_{\mathcal{D}_1}$ with continuous components given by

$$\lambda^*_m(x_1 \dots x_k) = \lambda_m(x_1) \dots \lambda_m(x_k) \quad \text{and} \quad \rho^*_m(x_1 \dots x_k) = \rho_m(x_1) \dots \rho_m(x_k) \tag{19}$$

for every $m \in M$ and $x_1 \dots x_k \in X^*_{\mathcal{D}_1}$. Moreover, since T is invariant under the biaction of M on $X_{\mathcal{D}_1}$, the T-generated free monoid T^* is invariant under the above biaction of M on $X^*_{\mathcal{D}_1}$, thereby yielding a monoid biaction of M on T^*. Thus, we have a well-defined semidirect product $T^* ** M$ given by this monoid biaction, cf. Section 2.5. Explicitly, the multiplication on $T^* ** M$ is given by

$$(\underline{t}, m)(\underline{t}', m') = (\rho_{m'}(t_1) \dots \rho_{m'}(t_k)\lambda_m(t'_1) \dots \lambda_m(t'_\ell),\, mm'), \tag{20}$$

for every $(\underline{t}, m) = (t_1 \dots t_k, m)$ and $(\underline{t}', m') = (t'_1 \dots t'_\ell, m')$ in $T^* ** M$.

Our next goal is to show that $T^* ** M$ has a monoid quotient that is part of a BiM having $X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0}$ as space component. This BiM is relevant because it exactly recognizes the Boolean combinations of languages in $\mathcal{V}(X_{\mathcal{D}_1}) \odot \mathcal{D}_1$ and in $\mathcal{D}_0$ (cf. Theorem 6.7). We proceed in two steps. First we show that the multiplication on $T^* ** M$ naturally extends to a biaction of $T^* ** M$ on $\beta(X^*_{\mathcal{D}_1}) \times X_{\mathcal{D}_0}$ with continuous components, so that the inclusion

$$T^* ** M \rightarrowtail \beta(X^*_{\mathcal{D}_1}) \times X_{\mathcal{D}_0}$$

has a BiM structure (cf. Proposition 6.3). Then, for every Boolean subalgebra $\mathcal{W}$ of $\mathsf{Clopen}(X^*_{\mathcal{D}_1})$, we have a continuous quotient $\eta : \beta(X^*_{\mathcal{D}_1}) \twoheadrightarrow X_{\mathcal{W}}$, and hence, a continuous quotient $(\eta \times id) : \beta(X^*_{\mathcal{D}_1}) \times X_{\mathcal{D}_0} \twoheadrightarrow X_{\mathcal{W}} \times X_{\mathcal{D}_0}$. The second step is then to show that, if $\mathcal{W} = \mathcal{V}(X_{\mathcal{D}_1})$ for some lp-variety $\mathcal{V}$, then $(\eta \times id)$ defines a BiM quotient

$$\begin{array}{ccc} T^* ** M & \rightarrowtail & \beta(X^*_{\mathcal{D}_1}) \times X_{\mathcal{D}_0} \\ \downarrow & & \downarrow \eta \times id \\ N & \rightarrowtail & X_{\mathcal{V}(X_{\mathcal{D}_1})} \times X_{\mathcal{D}_0} \end{array} \tag{21}$$

where N denotes the image of $T^* ** M$ under $(\eta \times id)$, that is, $N = (\eta \times id)[T^* ** M]$. This is the content of Proposition 6.6.

For the first part, we first observe that, using the equality $\rho_{m'}(t) = \lambda_t(m')$ valid for every $t, m' \in M_{\mathcal{D}}$ (and in particular, for every $t \in T$ and $m' \in M$),

by $\mu(a, 0) = e$ and $\mu(a, 1) = m$, and we have

$$A^* = \mu^{-1}(e), \quad A^* \otimes \mathbb{N} = \mu^{-1}(m), \quad \text{and} \quad A_z = \mu^{-1}(z),$$

where we identify A^* with $(A \times \{0\})^*$ and

$$A_z := (A \times 2)^* - (A^* \cup A^* \otimes \mathbb{N}).$$

2.10. Classes of sentences. Given a function $\zeta : A \to B$, we may define a map $\zeta^*_{\mathbf{x}} : A^* \otimes \mathbb{N}^{|\mathbf{x}|} \to B^* \otimes \mathbf{x}$ by setting $\zeta^*_{\mathbf{x}}(w, \mathbf{i}) = (\zeta^*(w), \mathbf{i})$, for every $w \in A^*$ and $\mathbf{i} \in \{1, \dots, |w|\}^{\mathbf{x}}$. By discrete duality, $\zeta^*_{\mathbf{x}}$ yields a homomorphism of Boolean algebras $(\zeta^*_{\mathbf{x}})^{-1} : \mathcal{P}(B^* \otimes \mathbb{N}^{|\mathbf{x}|}) \to \mathcal{P}(A^* \otimes \mathbb{N}^{|\mathbf{x}|})$. By identifying fragments of logic with the set of languages they define, we may show that, for every set of quantifiers $\mathcal{Q}$ and every set of numerical predicates $\mathcal{N}$, the homomorphism $(\zeta^*_{\mathbf{x}})^{-1}$ restricts and co-restricts to a homomorphism $\mathcal{Q}_{B,\mathbf{x}}[\mathcal{N}] \to \mathcal{Q}_{A,\mathbf{x}}[\mathcal{N}]$. Indeed, for $\mathbf{x}$ consisting of a single variable x, we have

$$\zeta_x(w, i) \models \mathsf{P}_b(x) \iff (w, i) \models \bigvee_{\zeta(a)=b} \mathsf{P}_a(x),$$

for every $w \in A^*$ and $i \in \{1, \dots, |w|\}$. With a routine structural induction on the construction of formulas we can then show that $(\zeta^*_{\mathbf{x}})^{-1}$ sends the language defined by a formula $\phi \in \mathcal{Q}_{B,\mathbf{x}}[\mathcal{N}]$ to the language defined by the formula obtained from ϕ by substituting, for every occurrence of a predicate $\mathsf{P}_b(x)$ with $b \in B$, the formula $\bigvee_{\zeta(a)=b} \mathsf{P}_a(x)$. Note that if b is not in the image of ζ, then $\mathsf{P}_b(x)$ is replaced by the empty join, which is logically equivalent to the *always-false* proposition. We denote by $\zeta_{\mathcal{Q}_{\mathbf{x}}[\mathcal{N}]}$ this restriction and co-restriction of $(\zeta^*_{\mathbf{x}})^{-1}$, so that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{Q}_{B,\mathbf{x}}[\mathcal{N}] & \rightarrowtail & \mathcal{P}(B^* \otimes \mathbb{N}^{|\mathbf{x}|}) \\ \downarrow \zeta_{\mathcal{Q}_{\mathbf{x}}[\mathcal{N}]} & & \downarrow (\zeta^*_{\mathbf{x}})^{-1} \\ \mathcal{Q}_{A,\mathbf{x}}[\mathcal{N}] & \rightarrowtail & \mathcal{P}(A^* \otimes \mathbb{N}^{|\mathbf{x}|}) \end{array} \tag{2}$$

In particular, $\mathcal{Q}_{_,\mathbf{x}}[\mathcal{N}]$ defines a presheaf $\mathcal{Q}_{_,\mathbf{x}}[\mathcal{N}] : \mathbf{Set}^{\mathrm{op}}_{fin} \to \mathsf{BA}$. It is easy to see that $\mathcal{Q}_{_,\mathbf{x}}[\mathcal{N}]$ takes right (respectively, left) inverses to left (respectively, right) inverses and thus surjections (respectively, injections) on finite sets to embeddings (respectively, quotients) in Boolean algebras.

A *class of sentences* is, intuitively speaking, the notion corresponding to that of an lp-strain of languages in the setting of logic on words. This notion

models what is usually referred to as a *fragment of logic*. Formally, a *class of sentences* Γ is a map that associates to each finite alphabet A a set of sentences $\Gamma(A) \subseteq \mathcal{Q}_A[\mathcal{N}]$ which satisfies the following properties:

(LC.1) Each set $\Gamma(A)$ is closed under Boolean connectives $\wedge$ and $\neg$ (and thus, under $\vee$);

(LC.2) For each map $\zeta\colon A \to B$ between finite alphabets, the homomorphism $\zeta_{\mathcal{Q}[\mathcal{N}]} : \mathcal{Q}_B[\mathcal{N}] \to \mathcal{Q}_A[\mathcal{N}]$ restricts and co-restricts to a homomorphism $\zeta_\Gamma : \Gamma(B) \to \Gamma(A)$.

Since we consider sentences up to semantic equivalence, by (LC.1), each $\Gamma(A)$ is a Boolean algebra, and (LC.2) simply says that the assignment $A \mapsto \mathcal{V}_\Gamma(A) := \{L_\phi \mid \phi \in \Gamma(A)\}$ defines an lp-strain of languages.

3. Languages over profinite alphabets

As in the regular setting, it is sometimes useful to consider languages over profinite alphabets. Indeed, we will see in Section 5 that these appear naturally when extending *substitution* to arbitrary Boolean algebras. In this section we see how any lp-strain of languages may be naturally extended to profinite alphabets. Since lp-strains of languages are the formal language counterpart of classes of sentences, such extension will be another key ingredient in the subsequent sections (cf. Corollary 5.11, but also Theorem 6.7).

Let $\mathcal{V}$ be an lp-strain of languages. We extend the assignment $A \mapsto \mathcal{V}(A)$ for A finite to profinite alphabets as follows. Let Y be a profinite alphabet. By definition, Y is the projective limit of all its finite continuous quotients, say $\{h_i : Y \twoheadrightarrow Y_i\}_{i\in I}$, the connecting morphisms being all the surjective maps $h_{i,j} : Y_i \twoheadrightarrow Y_j$ satisfying $h_j = h_{i,j} \circ h_i$. Note that I is a directed set ordered by $i \geq j$ if and only if $h_{i,j}$ is defined. In turn, each of the maps $h_{i,j}$ uniquely defines an lp-morphism $h_{i,j}^* : Y_i^* \twoheadrightarrow Y_j^*$. Since $\mathcal{V}$ is an lp-strain, for every $i \geq j$, there is a well-defined embedding of Boolean algebras $\theta_{i,j} : \mathcal{V}(Y_j) \rightarrowtail \mathcal{V}(Y_i)$ sending $L \in \mathcal{V}(Y_j)$ to its preimage $(h_{i,j}^*)^{-1}(L) \in \mathcal{V}(Y_j)$. A similar phenomenon happens with respect to each of the continuous quotients $h_i : Y \twoheadrightarrow Y_i$, when Y^* is viewed as the topological space which is the union over $n \geq 0$ of the product spaces Y^n. Indeed, since the clopen subsets of Y^* are the unions of the form $\bigcup_{n\geq 0} C_n$, where $C_n \subseteq Y^n$ is a clopen subset of Y^n, we have that h_i^* is continuous and thus, it defines an embedding of Boolean algebras $\iota_i : \mathcal{V}(Y_i) \rightarrowtail \mathsf{Clopen}(Y^*)$. Clearly, the family $\{\mathcal{V}(Y_i) \mid i \in I\}$ forms a direct limit system with connecting morphisms $\{\theta_{i,j} \mid i \leq j\}$ satisfying $\iota_j = \iota_i \circ \theta_{i,j}$.

Definition 6.1. Let $\mathcal{C} \subseteq \mathcal{P}(A^* \otimes \mathbb{N})$ be a Boolean subalgebra and $\mathcal{W} \subseteq \mathsf{Clopen}(X_\mathcal{C}^*)$ be a Boolean subalgebra of languages over $X_\mathcal{C}$, the dual space of $\mathcal{C}$, viewed as a profinite alphabet. Also, let $\pi : A^* \otimes \mathbb{N} \to X_\mathcal{C}$ be the (restriction of the) dual of the embedding $\mathcal{C} \hookrightarrow \mathcal{P}(A^* \otimes \mathbb{N})$. Define the map $\tau_\mathcal{C} : A^* \to X_\mathcal{C}^*$ by

$$\tau_\mathcal{C}(w) = \pi(w, 1) \dots \pi(w, |w|),$$

and let

$$\mathcal{W} \odot \mathcal{C} := \{\tau_\mathcal{C}^{-1}(K) \mid K \in \mathcal{W}\}.$$

Clearly, $\mathcal{W} \odot \mathcal{C}$ is a Boolean subalgebra of $\mathcal{P}(A^*)$ and, by duality, we have that the dual space of $\mathcal{W} \odot \mathcal{C}$ is a closed subspace of $X_\mathcal{W}$ given by the dual of the map

$$\tau_\mathcal{C}^{-1}\colon \mathcal{W} \to \mathcal{P}(A^*).$$

First, note that by Corollary 5.11, as discussed above, if $\mathcal{C} = \mathcal{D}_\Delta$ and $\mathcal{W} = \mathcal{V}_\Gamma(X_\Delta)$, then $\mathcal{W} \odot \mathcal{C}$ is the Boolean algebra of languages over A given by the logic fragment $\Gamma \odot \Delta$ so that Definition 6.1 is indeed an abstraction of the construction in Section 5.2. Apart from this logic application, which is our focus here, note that this definition is useful more widely in the theory of formal languages, see [21] for some examples.

Our purpose now is to obtain a description of the BiM dual to the Boolean algebra closed under quotients generated by a Boolean algebra of the form $\mathcal{V}(X_\mathcal{C}) \odot \mathcal{C}$, for an lp-variety of languages $\mathcal{V}$ and $\mathcal{C}$ coming from a Boolean subalgebra closed under quotients of $\mathcal{P}((A \times 2)^*)$ as studied in Section 4.2.

Thus, we let $\mathcal{D}$ be a Boolean subalgebra of $\mathcal{P}((A \times 2)^*)$ which is closed under the quotient operations and is generated by the union of its retracts $\mathcal{D}_0 \subseteq \mathcal{P}(A^*)$ and $\mathcal{D}_1 \subseteq \mathcal{P}(A^* \otimes \mathbb{N})$ as considered in Lemma 4.4. Recall that $\mathcal{D}_0$ and $\mathcal{D}_1$ are, respectively, the following quotients of $\mathcal{D}$:

$$\mathcal{D}_0 = \{L \cap A^* \mid L \in \mathcal{D}\} \qquad \text{and} \qquad \mathcal{D}_1 = \{L \cap (A^* \otimes \mathbb{N}) \mid L \in \mathcal{D}\}.$$

Let $\pi : (A \times 2)^* \to X_\mathcal{D}$ be the restriction to $(A \times 2)^*$ of the map dual to the embedding $\mathcal{D} \rightarrowtail \mathcal{P}((A \times 2)^*)$. As in Section 4.2, we denote

$$M := \pi[A^*] \qquad \text{and} \qquad T := \pi[A^* \otimes \mathbb{N}].$$

These are, respectively, a dense monoid in $X_{\mathcal{D}_0}$ and a dense subset of $X_{\mathcal{D}_1}$. By Corollary 4.6(a), the natural biaction of the syntactic monoid $M_\mathcal{D}$ of $\mathcal{D}$ on $X_\mathcal{D}$ restricts and co-restricts to a biaction of M on $X_{\mathcal{D}_0}$ and to a biaction of M on $X_{\mathcal{D}_1}$. In particular, M is a monoid quotient of A^*, and T comes equipped with a biaction of M.

In particular, we have

$$\mathcal{Q}_A[\mathcal{N}] = \bigcup_{n\in\mathbb{N}} \underbrace{\Gamma_{\mathcal{Q}} \odot (\dots \odot (\Gamma_{\mathcal{Q}}}_{n \text{ times}} \odot \Delta_n)\dots).$$

Proof: Let n be a positive integer. For each $k \geq 0$, we denote

$$\Gamma^k_{\mathcal{Q}} \odot \Delta_n := \underbrace{\Gamma_{\mathcal{Q}} \odot (\dots \odot (\Gamma_{\mathcal{Q}}}_{k \text{ times}} \odot \Delta_n)\dots).$$

We prove by induction on $k \geq 0$ that $\Gamma^k_{\mathcal{Q}} \odot \Delta_n$ consists of the formulas of quantifier-depth k in the context $\{x_{k+1}, \dots, x_n\}$. The case $k = 0$ follows from the definition of Δ_n. Suppose the claim is valid for k, and let $\mathbf{x} = \{x_{k+2}, \dots, x_n\}$. By induction hypothesis, we have $\Gamma^k_{\mathcal{Q}} \odot \Delta_n \subseteq \mathcal{Q}_{A,\mathbf{x}x_{k+1}}[\mathcal{N}]$. Thus, by Corollary 5.19, the Boolean algebra $\Delta := \langle \varepsilon_{\mathbf{x},x_{k+1}}[\Gamma^k_{\mathcal{Q}} \odot \Delta_n]\rangle_{\mathsf{BA}}$ is such that

$$\Gamma^{k+1}_{\mathcal{Q}} \odot \Delta_n = \Gamma_{\mathcal{Q}} \odot (\Gamma^k_{\mathcal{Q}} \odot \Delta_n) = \delta_{\mathbf{x},\emptyset}\ [\Gamma_{\mathcal{Q}} \odot \Delta].$$

Since $\delta_{\mathbf{x},\emptyset}$ is a homomorphism of Boolean algebras, by Lemma 5.21, it follows that $\Gamma^{k+1}_{\mathcal{Q}} \odot \Delta_n$ is the Boolean algebra generated by the formulas of the form

$$\delta_{\mathbf{x},\emptyset}(Qx_{k+1}\ \psi) = Qx_{x+1}\ \delta_{\mathbf{x},\emptyset}(\psi), \tag{18}$$

where ψ belongs to Δ. In turn, the elements of Δ are Boolean combinations of formulas of the form $\varepsilon_{\mathbf{x},x_{k+1}}(\phi)$, for some $\phi \in \Gamma^k_{\mathcal{Q}} \odot \Delta_n$. Finally, by induction hypothesis, $\Gamma^k_{\mathcal{Q}} \odot \Delta_n$ consists of the formulas of quantifier-depth k in the context $\{x_{k+1}, \dots, x_n\} = \mathbf{x}x_{k+1}$. Therefore, by Corollary 5.17, each $\delta_{\mathbf{x},\emptyset}(\psi)$ is a Boolean combination of formulas of quantifier-depth k in the context $\{x_{k+1}, \dots, x_n\} = \mathbf{x}x_{k+1}$, and thus (18) is a quantifier-depth $k+1$ formula in the context $\{x_{k+2}, \dots, x_n\} = \mathbf{x}$ as we intended to show. ∎

6. Semidirect products

Let Γ be a class of sentences and $\Delta \subseteq \mathcal{Q}_{A,x}[\mathcal{N}]$ a Boolean subalgebra of formulas in one free variable. By Corollary 5.11, the languages definable by a formula of $\Gamma \odot \Delta$ may be described using the Boolean algebra of languages $\mathcal{V}_\Gamma(X_\Delta)$, where X_Δ is the dual space of Δ, and the map $\tau_\Delta : A^* \to X^*_\Delta$ as defined in Corollary 5.9. On the other hand, the map τ_Δ depends only on the Boolean subalgebra $\mathcal{D}_\Delta := \{L_\phi \mid \phi \in \Delta\} \subseteq \mathcal{P}(A^* \otimes \mathbb{N})$ of the languages definable by a formula in Δ. The following definition is an abstract version of this construction.

Thus, we may define

$$\mathcal{V}(Y) = \varinjlim \{\mathcal{V}(A) \mid A \text{ is a finite continuous quotient of } Y\}. \tag{3}$$

Note that, by identifying $\mathcal{V}(Y_i)$ with $\iota_i[\mathcal{V}(Y_i)] \subseteq \mathsf{Clopen}(Y^*)$ for each $i \in I$, we may regard $\mathcal{V}(Y)$ as a Boolean subalgebra of $\mathsf{Clopen}(Y^*) \subseteq \mathcal{P}(Y^*)$. More precisely, we have

$$\mathcal{V}(Y) = \bigcup_{i\in I} \mathcal{V}(Y_i),$$

and each $\mathcal{V}(Y_i)$ is a Boolean subalgebra of $\mathcal{V}(Y)$ (which dually means that $X_{\mathcal{V}(Y_i)}$ is a quotient of $X_{\mathcal{V}(Y)}$). Moreover, given a map $h : Y \to A$, words $u, v \in Y^*$, and a language $K \subseteq A^*$, a routine computation shows that

$$u^{-1}[(h^*)^{-1}(K)]v^{-1} = (h^*)^{-1}(s^{-1}Kt^{-1}),$$

where $s = h^*(u)$ and $t = h^*(v)$. Thus, if $\mathcal{V}(A)$ is closed under quotients, then so is $(h^*)^{-1}(\mathcal{V}(A))$. Therefore, we have:

Lemma 3.1. *A language $L \subseteq Y^*$ belongs to $\mathcal{V}(Y)$ if and only if there is a finite continuous quotient $h : Y \twoheadrightarrow A$ and a language $K \in \mathcal{V}(A)$ such that $L = (h^*)^{-1}(K)$. Further, if $\mathcal{V}$ is an lp-variety, then $\mathcal{V}(Y)$ is a Boolean algebra closed under quotients by words of Y^*.*

Remark 3.2. As we have seen above, every language over a profinite alphabet may be seen as a clopen subset of Y^*. However, even when $\mathcal{V}$ is the full variety of languages, meaning that $\mathcal{V}(A) = \mathcal{P}(A^*)$ for every finite alphabet A, the equality $\mathcal{V}(Y) = \mathsf{Clopen}(Y^*)$ does not hold in general. For instance, if $Y = \mathbb{N} \cup \{\infty\}$ is the one-point compactification of $\mathbb{N}$, we have $Y = \varprojlim \{h_i : \mathbb{N} \twoheadrightarrow \{0, 1, \dots, i\}\}_{i\in\mathbb{N}}$, where $h_i(j) = j$ for $j \in \{0, 1, \dots, j\}$, and $h_i(j) = i$ for $j \in \mathbb{N}_{>i} \cup \{\infty\}$. Moreover, the set $L = \{012\dots i \mid i \in \mathbb{N}\}$ is a clopen subset of Y^* which does not belong to $\mathcal{V}(Y)$. Indeed, by Lemma 3.1, this is due to the fact that there is no finite continuous quotient $h : Y \twoheadrightarrow A$ of Y so that L belongs to $(h^*)^{-1}(\mathcal{P}(A^*))$, as the preimage under h of every language over A necessarily contains some word with the letter ∞.

An interesting open question is then to provide a nice description of the dual space of $\mathcal{V}(Y)$ for the full variety $\mathcal{V}$. This would provide the "best" ambiant space for studying languages over a profinite alphabet. Even though, we define a *language over a profinite alphabet* Y to be a clopen subset of Y^*. Note that we have the following:

Proposition 3.3. *Let Y be a profinite set and $L \subseteq Y^*$. Then the following conditions are equivalent:*

(a) *There exists $q\colon Y \twoheadrightarrow A$ finite continuous quotient with $L = (q^*)^{-1}(q^*[L])$;*
(b) *There exist $V_1, \dots, V_n$ a clopen partition of Y such that, if $w \in L$, then there is $f : |w| \to n$ with*

$$w \in V_{f(1)} \dots V_{f(|w|)} \subseteq L;$$

(c) *There exist $V_1, \dots, V_n$ clopen subsets of Y such that, if $w \in L$, then there is $f : |w| \to n$ with*

$$w \in V_{f(1)} \dots V_{f(|w|)} \subseteq L.$$

Proof: (a)⇒(b): Fix an enumeration of $A = \{a_1, \dots, a_n\}$ and take $V_i = q^{-1}(a_i)$. Then $V_1, \dots, V_n$ is a clopen partition of Y. Also, for any $w \in L$

$$w \in (q^*)^{-1}(q^*(w)) = q^{-1}(q(w_1)) \dots q^{-1}(q(w_{|w|})) \subseteq (q^*)^{-1}(q^*[L]) = L,$$

where the last equality holds by hypothesis. Now (b) follows as

$$q^{-1}(q(w_1)) \dots q^{-1}(q(w_{|w|})) = V_{f(1)} \dots V_{f(|w|)},$$

where $f(j) = i$ if and only if $w_j = a_i$. Clearly (b)⇒(c).

For (c)⇒(a): Let $W_1, \dots, W_m$ be the atoms of the finite Boolean algebra of clopens generated by the clopen subsets $V_1, \dots, V_n$ stipulated in (c) and let $q\colon Y \twoheadrightarrow m$ be the corresponding quotient map. Then q is continuous and, by (c), for each $w \in L$ we have $f_w : |w| \to n$ with $w \in V_{f_w(1)} \dots V_{f_w(|w|)} \subseteq L$. Since the W_i are atoms, we have $w_j \in q^{-1}(q(w_j)) = W_{q(w_j)} \subseteq V_{f_w(j)}$, and it follows that

$$L \subseteq (q^*)^{-1}(q^*[L]) = \bigcup_{w \in L} (q^*)^{-1}(q^*(w)) = \bigcup_{w \in L} W_{q(w_1)} \dots W_{q(w_{|w|})}$$
$$\subseteq \bigcup_{w \in L} V_{f_w(1)} \dots V_{f_w(|w|)} \subseteq L. \quad \blacksquare$$

Notice that each of the spaces Y^n is a Boolean space, but Y^* is not. Nevertheless, since it is 0-dimensional, it may be naturally embedded in a Boolean space, namely in the Banaschewski compactification of Y^*, cf. Section 2.3. Now, we show that this compactification coincides with the Čech-Stone compactification of Y^*.

Lemma 3.4. *Let Y be a Boolean space and Y^* be the topological space which is the union over $n \geq 0$ of the product spaces Y^n. Then, Y^* has Čech-dimension zero.*

Corollary 5.19. *Let Γ be a class of sentences, and let $\Delta \subseteq \mathcal{Q}_{A,\mathbf{x}x}[\mathcal{N}]$ be a Boolean subalgebra. Then, there exists a Boolean subalgebra $\Delta' \subseteq \mathcal{Q}_{A\times 2^{\mathbf{x}},x}[\mathcal{N}]$, namely $\Delta' = \langle \varepsilon_{\mathbf{x},x}[\Delta] \rangle_{\mathsf{BA}}$, such that*

$$\Gamma \odot \Delta = \delta_{\mathbf{x},\emptyset}[\Gamma \odot \Delta'].$$

5.4. Applications to logic on words. In this section we show the utility of the Substitution Principle in handling the application of one layer of quantifiers to a Boolean algebra of formulas. As we will see, this allows for a systematic study of fragments of logic via smaller subfragments (cf. Proposition 5.22 below). Before stating it, we show how we can assign to a given set of quantifiers a class of sentences.

Lemma 5.20. *Let $\mathcal{Q}$ be a set of quantifiers. The assignment*

$$\Gamma_{\mathcal{Q}} : A \mapsto \mathcal{Q}_A[\emptyset] = \left\langle \left\{ Qx \bigvee_{a \in B} \mathsf{P}_a(x) \mid Q \in \mathcal{Q},\ B \subseteq A \right\} \right\rangle_{\mathsf{BA}}$$

defines a class of sentences.

Proof: By definition, $\Gamma_{\mathcal{Q}}(A)$ is a Boolean algebra, so Property (LC.1) holds. To check (LC.2) it is enough to observe that replacing a letter predicate by a disjunction of letter predicates yields formulas without numerical predicates. Thus, every map $\zeta : A \to B$ defines a homomorphism $\zeta_\Gamma : \mathcal{Q}_B[\emptyset] \to \mathcal{Q}_A[\emptyset]$. $\blacksquare$

The following is an immediate consequence of the definitions of $\Gamma_{\mathcal{Q}}$ and of $(_ \odot _)$, and it shows that $\Gamma_{\mathcal{Q}}$-substitutions model the application of a layer of quantifiers to a given Boolean algebra.

Lemma 5.21. *Let A be a finite alphabet, $\mathcal{Q}$ a set of quantifiers, $\mathcal{N}$ a set of numerical predicates, and $\Delta \subseteq \mathcal{Q}_{A,x}[\mathcal{N}]$ a Boolean subalgebra. Then,*

$$\Gamma_{\mathcal{Q}} \odot \Delta = \langle \{ Qx\ \phi \mid \phi \in \Delta \} \rangle_{\mathsf{BA}}.$$

Finally, an iteration of Lemma 5.21 yields the following:

Proposition 5.22. *Let A be a finite alphabet, $\mathcal{Q}$ a set of quantifiers, and $\mathcal{N}$ a set of numerical predicates. For each $n > 0$, let Δ_n be the Boolean algebra of quantifier-free formulas in the context $\mathbf{x} = \{x_1, \dots, x_n\}$. Then, a sentence of $\mathcal{Q}_A[\mathcal{N}]$ has quantifier-depth n if and only if it belongs to*

$$\underbrace{\Gamma_{\mathcal{Q}} \odot (\dots \odot (\Gamma_{\mathcal{Q}}}_{n \text{ times}} \odot\ \Delta_n) \dots).$$

Proof: Recall that we have a lattice embedding $\varepsilon_{\mathbf{x},x} : \mathcal{Q}_{A,\mathbf{x}x}[\mathcal{N}] \rightarrowtail \mathcal{Q}_{A\times 2^{\mathbf{x}},x}[\mathcal{N}]$. Let Δ' be the (finite) Boolean subalgebra of $\mathcal{Q}_{A\times 2^{\mathbf{x}},x}[\mathcal{N}]$ generated by $\varepsilon_{\mathbf{x},x}[\Delta]$. It is not hard to see that Δ' is the lattice generated by $\varepsilon_{\mathbf{x},x}[\Delta] \cup \{\neg\varepsilon_{\mathbf{x},x}(\mathbf{1})\}$, where $\mathbf{1} \in \mathcal{Q}_{A\times 2^{\mathbf{x}},x}[\mathcal{N}]$ denotes the *always-true* formula. Clearly, $\neg\varepsilon_{\mathbf{x},x}(\mathbf{1})$ is an atom of Δ'. We claim that, for every atom $\phi \in \Delta$, the formula $\varepsilon_{\mathbf{x},x}(\phi)$ is also an atom of Δ'. Let $\psi \in \mathcal{Q}_{A\times 2^{\mathbf{x}},x}[\mathcal{N}]$ be such that $\mathbf{0} < \psi \leq \varepsilon_{\mathbf{x},x}(\phi)$. Since $\delta_{\mathbf{x},x}$ is a homomorphism of Boolean algebras, we have $\mathbf{0} < \delta_{\mathbf{x},x}(\psi) \leq \delta_{\mathbf{x},x}\varepsilon_{\mathbf{x},x}(\phi)$. Since formulas are ordered by semantic implication, by Proposition 5.16, this implies $0 < \delta_{\mathbf{x},x}(\psi) \leq \phi$. Thus, the fact that ϕ is an atom yields that ϕ and $\delta_{\mathbf{x},x}(\psi)$ have the same models. On the other hand, since $\psi \leq \varepsilon_{\mathbf{x},x}(\phi)$, all models of ψ belong to $\mathsf{Im}(\iota_{\mathbf{x},x})$. Therefore, again by Proposition 5.16, we may conclude that $\varepsilon_{\mathbf{x},x}\delta_{\mathbf{x},x}(\psi)$ and ψ are semantically equivalent, thus, so are $\varepsilon_{\mathbf{x},x}(\phi)$ and ψ. This shows that $\varepsilon_{\mathbf{x},x}(\phi)$ is an atom, as required. Now, since $\varepsilon_{\mathbf{x},x}$ is a lattice embedding, we have

$$\bigvee \varepsilon_{\mathbf{x},x}[\mathrm{At}(\Delta)] = \varepsilon_{\mathbf{x},x}(\bigvee \mathrm{At}(\Delta)) = \varepsilon_{\mathbf{x},x}(\mathbf{1}).$$

Therefore, $\varepsilon_{\mathbf{x},x}[\mathrm{At}(\Delta)] \cup \{\neg\varepsilon_{\mathbf{x},x}(\mathbf{1})\}$ is a partition of $\mathcal{Q}_{A\times 2^{\mathbf{x}},x}[\mathcal{N}]$. Hence, we may conclude that

$$\mathrm{At}(\Delta') = \varepsilon_{\mathbf{x},x}[\mathrm{At}(\Delta)] \cup \{\neg\varepsilon_{\mathbf{x},x}(\mathbf{1})\}.$$

We let $\zeta : \mathrm{At}(\Delta) \rightarrowtail \mathrm{At}(\Delta')$ be the embedding defined by $\zeta(\phi) = \varepsilon_{\mathbf{x},x}(\phi)$, for every $\phi \in \mathrm{At}(\Delta)$. Let us see that the diagram of the statement commutes. Fix a formula $\theta \in \Gamma(C_{\Delta'})$. Then, $\sigma_{\Delta'}(\theta)$ is obtained from θ by replacing each occurrence of a letter predicate $\mathsf{P}_c(z)$, with $c \in C_{\Delta'} = \mathrm{At}(\Delta')$ by

- $\varepsilon_{\mathbf{x},x}(\phi)$ if c corresponds to the atom $\varepsilon_{\mathbf{x},x}(\phi)$ with $\phi \in \mathrm{At}(\Delta)$;
- $\neg\varepsilon_{\mathbf{x},x}(\mathbf{1})$ if c corresponds to the atom $\neg\varepsilon_{\mathbf{x},x}(\mathbf{1})$.

Therefore, since $\delta_{\mathbf{x},\emptyset}$ is a homomorphism of Boolean algebras which is a restriction and co-restriction of $\delta_{\mathbf{x},x}$, by Proposition 5.16, $\delta_{\mathbf{x},\emptyset} \circ \sigma_{\Delta'}(\theta)$ is obtained from θ by replacing each occurrence of a letter predicate $\mathsf{P}_c(z)$, with $c \in C_{\Delta'} = \mathrm{At}(\Delta')$ by a formula which is semantically equivalent to

- ϕ if c corresponds to the atom $\varepsilon_{\mathbf{x},x}(\phi)$ with $\phi \in \mathrm{At}(\Delta)$;
- $\neg\mathbf{1} \equiv \mathbf{0}$ if c corresponds to the atom $\neg\varepsilon_{\mathbf{x},x}(\mathbf{1})$.

But, by definition of σ_Δ and of ζ_Γ, this is precisely the result of $\sigma_\Delta \circ \zeta_\Gamma(\theta)$. Thus, we have $\delta_{\mathbf{x},\emptyset} \circ \sigma_{\Delta'}(\theta) \equiv \sigma_\Delta \circ \zeta_\Gamma(\theta)$ as required. ∎

Using the definition of the operator $(_\odot_)$, we may easily derive the following result from Proposition 5.18:

Proof: Let $Y^* = U_1 \cup \cdots \cup U_k$ be a finite open cover of Y^*. Then, for every $n \in \mathbb{N}$, $Y^n = \bigcup_{i=1}^k (U_i \cap Y^n)$ is a finite open cover of Y^n. Since Y^n is a Boolean space, this open cover admits a finite clopen refinement, say $Y_n = \bigcup_{j\in F_n} V_{j,n}$, where F_n is a finite set and each $V_{j,n}$ is a clopen subset of Y_n contained in some $U_i \cap Y^n$. For each $i = 1, \ldots, k$, we set

$$V_i = \bigcup \{V_{j,n} \mid n \in \mathbb{N},\ j \in F_n,\ V_{j,n} \subseteq U_i \cap Y^n\}.$$

Then, V_i is a clopen subset of Y^* contained in U_i and $Y^* = \bigcup_{i=1}^k V_i$. Thus, $\{V_i\}_{i=1}^k$ is the desired clopen finite refinement of the given finite open cover of Y^*. ∎

By a straightforward application of Theorem 2.1, we may conclude that the Banaschewski and Čech-Stone compactifications of Y^* indeed coincide.

Theorem 3.5. *Let Y be a Boolean space and Y^* be the topological space which is the union over $n \geq 0$ of the product spaces Y^n. Then, the dual space of $\mathsf{Clopen}(Y^*)$ is $\beta(Y^*)$, the Čech-Stone compatification of Y^*. In particular, there is a continuous embedding $Y^* \rightarrowtail \beta(Y^*)$ with dense image.*

A consequence of Lemma 3.1 is that the extension to profinite alphabets of an lp-strain of languages is closed under pre-images of continuous lp-morphisms between profinite alphabets. Dually, this may be phrased as follows:

Lemma 3.6. *Let $\alpha : Z \to Y$ be a continuous function and $\mathcal{V}$ an lp-strain of languages. Then, there exists a continuous map $\widehat{\alpha} : X_{\mathcal{V}(Y)} \to X_{\mathcal{V}(Z)}$ making the following diagram commute:*

$$\begin{array}{ccc} Z^* & \longrightarrow & X_{\mathcal{V}(Z)} \\ {\scriptstyle \alpha^*}\downarrow & & \downarrow{\scriptstyle \widehat{\alpha}} \\ Y^* & \longrightarrow & X_{\mathcal{V}(Y)} \end{array}$$

where $Y^* \to X_{\mathcal{V}(Y)}$ and $Z^* \to X_{\mathcal{V}(Z)}$ are the continuous functions with dense image obtained by dualizing and restricting the embeddings $\mathcal{V}(Y) \rightarrowtail \mathsf{Clopen}(Y^*)$ and $\mathcal{V}(Z) \rightarrowtail \mathsf{Clopen}(Z^*)$, respectively.

Proof: We only need to prove that the diagram below restricts correctly:

$$\begin{array}{ccc} \mathcal{V}(Y) & \rightarrowtail & \mathsf{Clopen}(Y^*) \\ \vdots & & \downarrow (\alpha^*)^{-1} \\ \mathcal{V}(Z) & \rightarrowtail & \mathsf{Clopen}(Y) \end{array}$$

Then, the map $\widehat{\alpha}$ is the dual of the homomorphism $(\alpha^*)^{-1} : \mathcal{V}(Y) \to \mathcal{V}(Z)$.

Let $L \in \mathcal{V}(Y)$. By Lemma 3.1, there is a finite continuous quotient $h : Y \twoheadrightarrow A$ and a language $K \in \mathcal{V}(A)$ such that $L = (h^*)^{-1}(K)$. In turn, $B = \mathsf{Im}(h \circ \alpha)$ is a finite continuous quotient of Z which embeds in A, say $e : B \rightarrowtail A$. Since $\mathcal{V}$ is an lp-strain of languages, and thus closed under pre-images of lp-morphisms, the language $(e^*)^{-1}(K)$ belongs to $\mathcal{V}(B)$. Finally, using again Lemma 3.1 we may conclude that $(\alpha^*)^{-1}(L)$ belongs to $\mathcal{V}(Z)$ as intended. ∎

4. Recognition of Boolean algebras closed under quotients

Quotients by words are natural operations on languages. In the regular setting they dually correspond to the multiplication in profinite monoids [8], and one is often interested in studying Boolean algebras that are closed under quotients. Beyond regularity, the closure under quotients will bring additional algebraic structure, and thus computational power, to the dual space of the Boolean algebra considered. We may thus think of the dual space of a Boolean algebra closed under quotients as having "almost" a monoid structure. This intuitive idea is captured by the *Boolean spaces with internal monoids*, which were introduced in [10] as an alternative to the semiuniform monoids of [9]. In Section 4.1 we will introduce *Boolean spaces with an internal monoid* (BiM's) and define its variant of a BiM-stamp.

Then, in Section 4.2, we will discuss recognition of languages that are defined by some formula with a free variable, that is, languages of marked words. We already saw in Section 2.9 that the set $A^* \otimes \mathbb{N}$ of all marked words is not equipped with a monoid structure. For this reason, BiM's will not appear as a natural notion of recognizer. Nevertheless, the fact that $A^* \otimes \mathbb{N}$ embeds in the free monoid $(A \times 2)^*$ allows us to identify some monoid actions that will be crucial when defining semidirect products in Section 6.

4.1. Languages over profinite alphabets. Let Y be a profinite alphabet, and $\mathcal{B} \subseteq \mathsf{Clopen}(Y^*)$ be a Boolean algebra of languages. If $\mathcal{B}$ is closed under

consists of a single variable x, the functions $\varepsilon_{x,\mathbf{y}}$ and $\delta_{x,\mathbf{y}}$ are precisely those defined in Definitions 5.12 and 5.14, respectively.

By Lemmas 5.13 and 5.15, we have:

Proposition 5.16. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be disjoint contexts. Then, the following diagrams commute:*

$$\begin{array}{ccc} \mathcal{Q}_{A,\mathbf{xy}}[\mathcal{N}] & \longrightarrow & \mathcal{P}(A^* \otimes \mathbb{N}^{|\mathbf{xy}|}) \\ {\scriptstyle \varepsilon_{\mathbf{x},\mathbf{y}}}\downarrow & & \downarrow{\scriptstyle \iota_{\mathbf{x},\mathbf{y}}[_]} \\ \mathcal{Q}_{A\times 2^{\mathbf{x}},\mathbf{y}}[\mathcal{N}] & \twoheadrightarrow & \mathcal{P}((A \times 2^{\mathbf{x}})^* \otimes \mathbb{N}^{|\mathbf{y}|}) \end{array} \qquad \begin{array}{ccc} \mathcal{Q}_{A\times 2^{\mathbf{x}},\mathbf{y}}[\mathcal{N}] & \twoheadrightarrow & \mathcal{P}((A \times 2^{\mathbf{x}})^* \otimes \mathbb{N}^{|\mathbf{y}|}) \\ {\scriptstyle \delta_{\mathbf{x},\mathbf{y}}}\downarrow & & \downarrow{\scriptstyle \iota^{-1}_{\mathbf{x},\mathbf{y}}} \\ \mathcal{Q}_{A,\mathbf{xy}}[\mathcal{N}] & \longrightarrow & \mathcal{P}(A^* \otimes \mathbb{N}^{|\mathbf{xy}|}) \end{array}$$

An immediate, but important, consequence of Proposition 5.16 is the following:

Corollary 5.17. *For every* $\phi \in \mathcal{Q}_{A,\mathbf{xy}}[\mathcal{N}]$ *and* $\psi \in \mathcal{Q}_{A\times 2^{\mathbf{x}},\mathbf{y}}[\mathcal{N}]$*, we have the following:*

(a) ϕ *is semantically equivalent to* $\delta_{\mathbf{x},\mathbf{y}}\varepsilon_{\mathbf{x},\mathbf{y}}(\phi)$,
(b) *the models of* ψ *that belong to* $\mathsf{Im}(\iota_{\mathbf{x},\mathbf{y}})$ *are precisely the models of* $\varepsilon_{\mathbf{x},\mathbf{y}}\delta_{\mathbf{x},\mathbf{y}}(\psi)$.

In particular, $\varepsilon_{\mathbf{x},\mathbf{y}}$ *is a lattice embedding, and* $\delta_{\mathbf{x},\mathbf{y}}$ *is a quotient of Boolean algebras.*

We finally show that it suffices to consider substitution with respect to a single variable.

Proposition 5.18. *Let* Γ *be a class of sentences, and let* $\Delta \subseteq \mathcal{Q}_{A,\mathbf{x}x}[\mathcal{N}]$ *be a finite Boolean subalgebra. Then, there exists a finite Boolean subalgebra* $\Delta' \subseteq \mathcal{Q}_{A\times 2^{\mathbf{x}},x}[\mathcal{N}]$*, namely* $\Delta' = \langle \varepsilon_{\mathbf{x},x}[\Delta] \rangle_{\mathsf{BA}}$*, and an embedding* $\zeta : \mathrm{At}(\Delta) \rightarrowtail \mathrm{At}(\Delta')$*, for which the following diagram commutes, up to semantic equivalence:*

$$\begin{array}{ccc} \Gamma(C_{\Delta'}) & \xrightarrow{\sigma_{\Delta'}} & \mathcal{Q}_{A\times 2^{\mathbf{x}}}[\mathcal{N}] \\ {\scriptstyle \zeta_\Gamma}\downarrow & & \downarrow{\scriptstyle \delta_{\mathbf{x},\emptyset}} \\ \Gamma(C_\Delta) & \xrightarrow[\sigma_\Delta]{} & \mathcal{Q}_{A,\mathbf{x}}[\mathcal{N}] \end{array}$$

where $C_\Delta = \mathrm{At}(\Delta)$ *and* $C_{\Delta'} = \mathrm{At}(\Delta')$*. In particular, since* ζ_Γ *is surjective, we have:*

$$\Gamma \odot \Delta = \delta_{\mathbf{x},\emptyset}[\Gamma \odot \Delta'].$$

Conversely, a formula ψ over $(A \times 2)$ in context $\mathbf{y}$ may be regarded as a formula over A in context $x\mathbf{y}$ as follows:

Definition 5.14. The *decoding function* (with respect to $x, \mathbf{y}$) is the map $\delta_{x,\mathbf{y}} : \mathcal{Q}_{A\times 2,\mathbf{y}}[\mathcal{N}] \to \mathcal{Q}_{A,x\mathbf{y}}[\mathcal{N}]$ sending a formula ψ to the formula obtained from ψ by replacing:

- each predicate $\mathsf{P}_{(a,1)}(z)$ by $\mathsf{P}_a(z) \wedge (x = z)$,
- each predicated $\mathsf{P}_{(a,0)}(z)$ by $\mathsf{P}_a(z) \wedge (x \neq z)$.

Of course, we are assuming that things have been arranged so that the variable x does not occur in ψ. It is easy to see that $\delta_{x,\mathbf{y}}$ is a well-defined homomorphism of Boolean algebras. Similarly to what happens for $\varepsilon_{x,\mathbf{y}}$, if $\mathbf{y}'$ is a context containing $\mathbf{y}$, then $\delta_{x,\mathbf{y}}(\psi)$ is the restriction and co-restriction of $\delta_{x,\mathbf{y}'}(\psi)$.

Again, a structural induction on construction of formulas shows the following:

Lemma 5.15. *The following diagram commutes:*

$$\begin{array}{ccc} \mathcal{Q}_{A\times 2,\mathbf{y}}[\mathcal{N}] & \rightarrowtail & \mathcal{P}((A\times 2)^* \otimes \mathbb{N}^{|\mathbf{y}|}) \\ {\scriptstyle \delta_{x,\mathbf{y}}}\downarrow & & \downarrow{\scriptstyle \iota^{-1}_{x,\mathbf{y}}} \\ \mathcal{Q}_{A,x\mathbf{y}}[\mathcal{N}] & \rightarrowtail & \mathcal{P}(A^* \otimes \mathbb{N}^{|x\mathbf{y}|}) \end{array}$$

That is, for every $\psi \in \mathcal{Q}_{A\times 2,\mathbf{y}}[\mathcal{N}]$, we have $L_{\delta_{x,\mathbf{y}}(\phi)} = \iota^{-1}_{x,\mathbf{y}}(L_\phi)$.

Let us now define $\varepsilon_{\mathbf{x},\mathbf{y}}$ and $\delta_{\mathbf{x},\mathbf{y}}$, for every contexts $\mathbf{x}$ and $\mathbf{y}$. We write $\mathbf{x} = \{x_1, \ldots, x_k\}$ and, for each $i \in \{0, \ldots, k\}$, we let $\mathbf{x}_i$ denote the context $\{x_1, \ldots, x_i\}$ (in particular, $\mathbf{x}_0$ denotes the empty context). Then, for each $i \in \{1, \ldots, k\}$, we have defined

$$\varepsilon_{x_i,\mathbf{x}_{i-1}\mathbf{y}} : \mathcal{Q}_{A\times 2^{k-i},\mathbf{x}_i\mathbf{y}}[\mathcal{N}] \to \mathcal{Q}_{A\times 2^{k-i+1},\mathbf{x}_{i-1}\mathbf{y}}[\mathcal{N}]$$

and

$$\delta_{x_i,\mathbf{x}_{i-1}\mathbf{y}} : \mathcal{Q}_{A\times 2^{k-i+1},\mathbf{x}_{i-1}\mathbf{y}}[\mathcal{N}] \to \mathcal{Q}_{A\times 2^{k-i},\mathbf{x}_i\mathbf{y}}[\mathcal{N}].$$

An iteration of these maps yields two functions

$$\varepsilon_{\mathbf{x},\mathbf{y}} : \mathcal{Q}_{A,\mathbf{x}\mathbf{y}}[\mathcal{N}] \to \mathcal{Q}_{A\times 2^{\mathbf{x}},\mathbf{y}}[\mathcal{N}] \quad \text{and} \quad \delta_{\mathbf{x},\mathbf{y}} : \mathcal{Q}_{A\times 2^{\mathbf{x}},\mathbf{y}}[\mathcal{N}] \to \mathcal{Q}_{A,\mathbf{x}\mathbf{y}}[\mathcal{N}],$$

which are given, respectively, by $\varepsilon_{\mathbf{x},\mathbf{y}} = \varepsilon_{x_1,\mathbf{x}_0\mathbf{y}} \circ \cdots \circ \varepsilon_{x_k,\mathbf{x}_{k-1}\mathbf{y}}$ and by $\delta_{\mathbf{x},\mathbf{y}} = \delta_{x_k,\mathbf{x}_{k-1}\mathbf{y}} \circ \cdots \circ \delta_{x_1,\mathbf{x}_0\mathbf{y}}$. In particular, $\varepsilon_{\mathbf{x},\mathbf{y}}$ is a homomorphism of bounded lattices, while $\delta_{\mathbf{x},\mathbf{y}}$ is a homomorphism of Boolean algebras. Clearly, when $\mathbf{x}$

quotients, then for every word w over Y, the homomorphisms

$$\ell_w : \mathsf{Clopen}(Y^*) \to \mathsf{Clopen}(Y^*), \qquad L \mapsto w^{-1}L$$

and

$$r_w : \mathsf{Clopen}(Y^*) \to \mathsf{Clopen}(Y^*), \qquad L \mapsto Lw^{-1}$$

restrict and co-restrict to endomorphisms of $\mathcal{B}$. Thus, we have the two following commutative diagrams:

$$\begin{array}{ccc} \mathsf{Clopen}(Y^*) & \xrightarrow{\ell_w} & \mathsf{Clopen}(Y^*) \\ \updownarrow & & \updownarrow \\ \mathcal{B} & \xrightarrow[\ell_w]{} & \mathcal{B} \end{array} \qquad \begin{array}{ccc} \mathsf{Clopen}(Y^*) & \xrightarrow{r_w} & \mathsf{Clopen}(Y^*) \\ \updownarrow & & \updownarrow \\ \mathcal{B} & \xrightarrow[r_w]{} & \mathcal{B} \end{array}$$

Dually, we have the following commutative diagrams of continuous functions

$$\begin{array}{ccc} Y^* & \xrightarrow{\widetilde{\ell}_w} & Y^* \\ {\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle \pi} \\ X_{\mathcal{B}} & \xrightarrow[\widetilde{\ell}_w]{} & X_{\mathcal{B}} \end{array} \qquad \begin{array}{ccc} Y^* & \xrightarrow{\widetilde{r}_w} & Y^* \\ {\scriptstyle \pi}\downarrow & & \downarrow{\scriptstyle \pi} \\ X_{\mathcal{B}} & \xrightarrow[\widetilde{r}_w]{} & X_{\mathcal{B}} \end{array} \tag{4}$$

We make a few remarks about these diagrams. First, notice that concatenation of words over Y turns Y^* into a topological monoid. Moreover, since for every $u, v \in Y^*$ we have $\ell_u \circ r_v = r_v \circ \ell_u$, we also have $\widetilde{\ell}_u \circ \widetilde{r}_v = \widetilde{r}_v \circ \widetilde{\ell}_u$. Therefore, the monoid structure on Y^* induces a biaction with continuous components of Y^* on $X_{\mathcal{B}}$ which is given by

$$Y^* \times X_{\mathcal{B}} \times Y^* \to X_{\mathcal{B}}, \qquad (u, x, v) \mapsto \widetilde{\ell}_u(\widetilde{r}_v(x)) = \widetilde{r}_v(\widetilde{\ell}_u(x)). \tag{5}$$

However, $X_{\mathcal{B}}$ itself does not necessarily inherit a monoid structure. Indeed, as it was shown in [9, 7], when A is a finite alphabet, this is case if and only if $\mathcal{B}$ consists of regular languages. Nevertheless, (5) induces a monoid structure on the dense subspace $M := \pi[Y^*]$ of $X_{\mathcal{B}}$, which is given by

$$\pi(u) \cdot \pi(v) := \pi(uv), \tag{6}$$

for every $u, v \in Y^*$. Indeed, with a routine computation we may derive the following equalities:

$$\widetilde{\ell}_u(\pi(v)) = \pi(uv) = \widetilde{r}_v(\pi(u)). \tag{7}$$

These not only show that (6) is well-defined in the sense that $\pi(uv) = \pi(u'v')$ whenever $\pi(u) = \pi(u')$ and $\pi(v) = \pi(v')$, but also that the monoid structure on M is indeed inherited from (5). In fact, it is not hard to see that $\pi : Y^* \twoheadrightarrow M$ is precisely the syntactic homomorphism of $\mathcal{B}$ as defined in Section 2.6. Moreover, since M is dense in $X_{\mathcal{B}}$, (7) also shows that whenever $\pi(u) = \pi(u')$ (respectively, $\pi(v) = \pi(v')$), the continuous functions $\widetilde{\ell}_u$ and $\widetilde{\ell}_{u'}$ (respectively, $\widetilde{r}_v$ and $\widetilde{r}_{v'}$) coincide on all of $X_{\mathcal{B}}$. Therefore, the natural biaction of M on itself extends to a biaction of M on $X_{\mathcal{B}}$ given by

$$M \times X_{\mathcal{B}} \to X_{\mathcal{B}}, \quad (m, x) \mapsto \lambda_m(x) \quad \text{and} \quad X_{\mathcal{B}} \times M \to X_{\mathcal{B}}, \quad (x, m) \mapsto \rho_m(x),$$

with continuous components at each $m \in M$. Here, for each $m \in M$, λ_m (respectively, ρ_m) denotes the continuous function $\widetilde{\ell}_u$ (respectively, $\widetilde{r}_u$) where $u \in Y^*$ is any word satisfying $\pi(u) = m$.

The following definition, which captures duality for Boolean subalgebras of monoids closed under the quotient operations, originates in [10], where it was used for recognition over finite alphabets. The above considerations verify that it remains the appropriate notion for recognition over profinite alphabets.

Definition 4.1. A *Boolean space with an internal monoid* (BiM) is a triple (M, p, X) where X is a Boolean space equipped with a biaction of a monoid M whose right and left components at each $m \in M$ are continuous and an injective function $p\colon M \rightarrowtail X$ which has dense image and is a morphism of sets with M-biactions. That is, for each $m \in M$, the following diagrams commute:

$$\begin{array}{ccc} M & \xrightarrow{p} & X \\ {\scriptstyle m\cdot(_)}\downarrow & & \downarrow{\scriptstyle \lambda_m} \\ M & \xrightarrow[p]{} & X \end{array} \qquad \begin{array}{ccc} M & \xrightarrow{p} & X \\ {\scriptstyle (_)\cdot m}\downarrow & & \downarrow{\scriptstyle \rho_m} \\ M & \xrightarrow[p]{} & X \end{array}$$

Given a profinite alphabet Y, we say that a language $L \subseteq \mathsf{Clopen}(Y^*)$ is *recognized* by the BiM (M, p, X) if there exists a homomorphism $\mu : Y^* \to M$ such that $p \circ \mu$ is continuous, and a clopen subset $C \subseteq X$ satisfying $L = (p \circ \mu)^{-1}(C)$. Notice that, since X is a Boolean space, $p \circ \mu$ may be uniquely extended to a continuous function $\beta(Y^*) \to X$; and the dual of this function is the homomorphism of Boolean algebras $(p \circ \mu)^{-1} : \mathsf{Clopen}(X) \to \mathsf{Clopen}(\beta(Y^*)) = \mathsf{Clopen}(Y^*)$ whose image consists of the languages recognized by (M, p, X) via μ. Moreover, since p is a morphism of sets with M-biactions

of all marked words $A^* \otimes \mathbb{N}$ as a language over $(A \times 2)$. Indeed, one can easily check that the models of the sentence

$$\Phi := \exists! z \bigvee_{a \in A} \mathsf{P}_{(a,1)}(z)$$

are precisely the words of $(A \times 2)^*$ having exactly one letter of the form $(a, 1)$, that is, those that belong to the image of $A^* \otimes \mathbb{N} \rightarrowtail (A \times 2)^*$.

In order to keep the notation simple, we start by defining $\varepsilon_{\mathbf{x},\mathbf{y}}$ and $\delta_{\mathbf{x},\mathbf{y}}$ in the cases where $\mathbf{x}$ consists of a single variable x.

Definition 5.12. Let ϕ be a formula, x a variable, and assume that all occurrence of x in ϕ are free. We denote by $\widetilde{\phi}^x$ the formula obtained from ϕ by:

- replacing each predicate $\mathsf{P}_a(x)$ by $\exists z\, \mathsf{P}_{(a,1)}(z)$,
- replacing each predicate $\mathsf{P}_a(z)$ by $\mathsf{P}_{(a,1)}(z) \vee \mathsf{P}_{(a,0)}(z)$, if $z \neq x$,
- replacing each predicate $R(_, x, _)$ by $\exists z\, (\bigvee_{a \in A} \mathsf{P}_{(a,1)}(z) \wedge R(_, z, _))$.

We define the *encoding function* (with respect to $x, \mathbf{y}$) by

$$\varepsilon_{x,\mathbf{y}} : \mathcal{Q}_{A,x\mathbf{y}}[\mathcal{N}] \to \mathcal{Q}_{A\times 2,\mathbf{y}}[\mathcal{N}], \qquad \phi \mapsto \widetilde{\phi}^x \wedge \Phi.$$

Clearly, this is a well-defined lattice homomorphism. Also, if $\mathbf{y}'$ is a context containing $\mathbf{y}$ (so that $\mathcal{Q}_{A,x\mathbf{y}}[\mathcal{N}] \subseteq \mathcal{Q}_{A,x\mathbf{y}'}[\mathcal{N}]$ and $\mathcal{Q}_{A\times 2,\mathbf{y}}[\mathcal{N}] \subseteq \mathcal{Q}_{A\times 2,\mathbf{y}'}[\mathcal{N}]$) then $\varepsilon_{x,\mathbf{y}}$ is the restriction and co-restriction of $\varepsilon_{x,\mathbf{y}'}$.

Given contexts $\mathbf{x}$ and $\mathbf{y}$, we let

$$\iota_{\mathbf{x},\mathbf{y}} : A^* \otimes \mathbb{N}^{|\mathbf{xy}|} \rightarrowtail (A \times 2^{\mathbf{x}})^* \otimes \mathbb{N}^{|\mathbf{y}|}$$

denote the natural embedding extending the embedding $A^* \otimes \mathbb{N}^{|\mathbf{x}|} \rightarrowtail (A \times 2^{\mathbf{x}})^*$. By a routine structural induction on the construction of a formula, we can show the following:

Lemma 5.13. *The following diagram commutes:*

$$\begin{array}{ccc} \mathcal{Q}_{A,x\mathbf{y}}[\mathcal{N}] & \rightarrowtail & \mathcal{P}(A^* \otimes \mathbb{N}^{|x\mathbf{y}|}) \\ {\scriptstyle \varepsilon_{x,\mathbf{y}}}\downarrow & & \downarrow{\scriptstyle \iota_{x,\mathbf{y}}[_]} \\ \mathcal{Q}_{A,\mathbf{y}}[\mathcal{N}] & \rightarrowtail & \mathcal{P}((A \times 2)^* \otimes \mathbb{N}^{|\mathbf{y}|}) \end{array}$$

That is, for every $\phi \in \mathcal{Q}_{A,x\mathbf{y}}[\mathcal{N}]$, we have $L_{\varepsilon_{x,\mathbf{y}}(\phi)} = \iota_{x,\mathbf{y}}[L_\phi]$.

Proof: We prove that the left-hand side diagram commutes. Let L be a language of $\Gamma \odot \Delta$. By Definition 5.8 of $\Gamma \odot \Delta$, this means that there exists a finite Boolean subalgebra $\Delta' \subseteq \Delta$ such that L belongs to $\Gamma \odot \Delta'$. Denote by $C_{\Delta'}$ the alphabet consisting of the atoms of Δ'. Using the local version of the Substitution Principle (cf. Corollary 5.3), this is equivalent to the existence of some language $K \in \mathcal{V}_\Gamma(C_{\Delta'})$ such that $L = \tau_{\Delta'}^{-1}(K)$. But by Lemma 3.1 and Corollary 5.9, this amounts to having that L belongs to $\tau_\Delta^{-1}(\mathcal{V}_\Gamma(X_\Delta))$. So, we just proved that the image of the following composition

$$\mathcal{V}_\Gamma(X_\Delta) \rightarrowtail \mathsf{Clopen}(X_\Delta^*) \xrightarrow{\tau_\Delta^{-1}} \mathcal{P}(A^*)$$

is precisely $\Gamma \odot \Delta$. That is, we have a commutative diagram as in the left-hand side of the statement. ∎

We can now state the already announced global version of the Substitution Principle, which is simply a rephrasing of the commutativity of the left-hand side diagram of Corollary 5.10.

Corollary 5.11 (Substitution Principle - global version)**.** *Let Γ be a class of sentences and $\Delta \subseteq \mathcal{Q}_{A,x}[\mathcal{N}]$ be a Boolean subalgebra. Then, the languages (over A) definable in $\Gamma \odot \Delta$ are precisely the languages of the form $\tau_\Delta^{-1}(K)$, where $K \in \mathcal{V}_\Gamma(X_\Delta)$.*

5.3. Using an alphabet to encode free variables. We now explain how free variables may be encoded in an alphabet, provided the logic at hand is expressive enough. The idea is that a formula $\phi \in \mathcal{Q}_{A,x}[\mathcal{N}]$ may be regarded as a sentence over the extended alphabet $(A \times 2)$, in the same way we may regard marked words as words over $(A \times 2)$ via the embedding $A^* \otimes \mathbb{N} \rightarrowtail (A \times 2)^*$. More generally, given two disjoint contexts $\mathbf{x}$ and $\mathbf{y}$, we will define two functions

$$\varepsilon_{\mathbf{x},\mathbf{y}} : \mathcal{Q}_{A,\mathbf{xy}}[\mathcal{N}] \to \mathcal{Q}_{A\times 2^{\mathbf{x}},\mathbf{y}}[\mathcal{N}] \quad \text{and} \quad \delta_{\mathbf{x},\mathbf{y}} : \mathcal{Q}_{A\times 2^{\mathbf{x}},\mathbf{y}}[\mathcal{N}] \to \mathcal{Q}_{A,\mathbf{xy}}[\mathcal{N}],$$

which we call, respectively, the *encoding* and the *decoding* function. As the name suggests, $\varepsilon_{\mathbf{xy}}$ "encodes" the variables from $\mathbf{x}$ in the extended alphabet $A \times 2^{\mathbf{x}}$, while $\delta_{\mathbf{x},\mathbf{y}}$ reverts this process (see Proposition 5.16 for a precise statement).

In what follows we assume that the quantifier $\exists!$ (*there exists a unique*) and the numerical predicate $=$ (formally given by $\{(n,n) \mid n \in \mathbb{N}\}$) are expressible in our logic. This assumption is needed so that we are able to define the set

and the biaction of M on X has continuous components, the Boolean algebra of languages recognized by (M,p,X) via μ is closed under quotients. Indeed, if $L = (p\circ\mu)^{-1}(C)$ for some clopen subset $C \subseteq X$, and $u,v,w \in Y^*$, then we have

$$\begin{aligned} w \in u^{-1}Lv^{-1} &\iff p\circ\mu(uwv) \in C \iff p(\mu(u)\,\mu(w)\,\mu(v)) \in C \\ &\iff \lambda_{\mu(u)} \circ \rho_{\mu(v)} \circ p \circ \mu(w) \in C \\ &\iff w \in (p\circ\mu)^{-1}((\lambda_{\mu(u)} \circ \rho_{\mu(v)})^{-1}(C)). \end{aligned}$$

Thus, $u^{-1}Lv^{-1} = (p\circ\mu)^{-1}((\lambda_{\mu(u)} \circ \rho_{\mu(v)})^{-1}(C))$ is recognized by the clopen $(\lambda_{\mu(u)} \circ \rho_{\mu(v)})^{-1}(C) \subseteq X$.

In order to handle lp-varieties (that is, lp-strains closed under quotients), it is useful to refine the notion of BiM. In the regular setting, this corresponds to the notion of *stamp* [15]. The idea is to consider BiMs with a chosen profinite set of generators for the monoid component, and constrain the set of languages we recognize accordingly. Formally, a BiM-*stamp* (called BiM *presentation* in [5]) is a tuple $\mathsf{R} = (Y,\mu,M,p,X)$, where $\mu : Y^* \twoheadrightarrow M$ is a monoid quotient, (M,p,X) is a BiM, and $p\circ\mu$ is a continuous function. We remark that each Boolean algebra closed under quotients $\mathcal{B} \subseteq \mathsf{Clopen}(Y^*)$ naturally defines a BiM-stamp

$$Y^* \twoheadrightarrow M = \pi[Y^*] \rightarrowtail X_{\mathcal{B}},$$

as described in the beginning of this section. This is called the *syntactic* BiM-*stamp of* $\mathcal{B}$ and it is denoted by $\mathsf{Synt}(\mathcal{B}) = (Y,\mu_{\mathcal{B}},M_{\mathcal{B}},p_{\mathcal{B}},X_{\mathcal{B}})$.

BiM-stamps encode two types of behavior: algebraic behavior given by the the triple (Y,μ,M) and topological behavior given by the continuous function $p\circ\mu : Y^* \to X$. The interplay between these two plays a central role in this paper. We say that a language L is *recognized* by the BiM-stamp $\mathsf{R} = (Y,\mu,M,p,X)$ provided L is recognized by the BiM (M,p,X) via μ. As observed above, the set of all languages recognized by a BiM-stamp forms a Boolean algebra closed under quotients. Similarly to what happens in the regular setting, the syntactic BiM-stamp of a given Boolean algebra closed under quotients is the "optimal" BiM-stamp recognizing that Boolean algebra, in the following sense:

Proposition 4.2. *A* BiM-*stamp* $\mathsf{R} = (Y,\mu,M,p,X)$ *recognizes a Boolean algebra closed under quotients $\mathcal{B}$ if and only if the syntactic* BiM-*stamp of $\mathcal{B}$ factors through* R*, that is, there is a commutative diagram:*

$$\begin{array}{ccccc} Y^* & \xrightarrow{\mu} & M & \xrightarrow{p} & X \\ & \searrow^{\mu_{\mathcal{B}}} & \downarrow g & & \downarrow \varphi \\ & & M_{\mathcal{B}} & \xrightarrow[p_{\mathcal{B}}]{} & X_{\mathcal{B}} \end{array}$$

where g is a homomorphism and φ a continuous function.

Proof: It is clear that if $\mathsf{Synt}(\mathcal{B})$ factors through R, then every language recognized by $\mathsf{Synt}(\mathcal{B})$ is also recognized by R and, in particular, $\mathcal{B}$ is recognized by R. Conversely, suppose that $\mathcal{B}$ is recognized by R. Then, the kernel of μ is contained in the syntactic congruence $\sim_{\mathcal{B}}$, and therefore, the syntactic morphism $\mu_{\mathcal{B}}$ factors through μ, say $g \circ \mu = \mu_{\mathcal{B}}$. On the other hand, since $p \circ \mu$ has dense image, there are embeddings $\mathcal{B} \rightarrowtail \mathsf{Clopen}(X)$ and $\mathsf{Clopen}(X) \rightarrowtail \mathsf{Clopen}(Y^*)$, or dually, continuous quotients $\pi : \beta(Y^*) \twoheadrightarrow X$ and $\varphi : X \twoheadrightarrow X_{\mathcal{B}}$, where the restriction of π to Y^* is $p \circ \mu$, and $p_{\mathcal{B}} = \varphi \circ p$. Finally, since μ is surjective and $\mu_{\mathcal{B}} = g \circ \mu$, it follows that $\varphi \circ p = p_{\mathcal{B}} \circ g$ as intended. ∎

A morphism between BiM-stamps $\mathsf{R} = (Y, \mu, M, p, X)$ and $\mathsf{S} = (Z, \nu, N, q, W)$ is a triple $\Phi = (h, g, \varphi)$, where $h : Y^* \to Z^*$ is a continuous homomorphism and (g, φ) is a morphism between the corresponding BiM components of R and S, so that the following diagram commutes:

$$\begin{array}{ccccc} Y^* & \xrightarrow{\mu} & M & \xrightarrow{p} & X \\ {\scriptstyle h}\downarrow & & \downarrow g & & \downarrow \varphi \\ Z^* & \xrightarrow[\nu]{} & N & \xrightarrow[q]{} & W \end{array}$$

When h is an lp-morphism, we say that Φ is an *lp-morphism of* BiM*-stamps*. Note that when h is onto, then so are g and φ. If moreover h is length-preserving, then we will call Φ an *lp-quotient*. It is worth mentioning that every lp-quotient $\Phi = (h, g, \varphi) : \mathsf{R} \twoheadrightarrow \mathsf{S}$ induces the following commutative diagram of continuous functions:

$$\begin{array}{ccc} Y^* & \xrightarrow{p \circ \mu} & X \\ {\scriptstyle h}\downarrow & & \downarrow \varphi \\ Z^* & \xrightarrow[q \circ \nu]{} & W \end{array}$$

Thus, taking preimages yields the following commutative diagram of homomorphisms of Boolean algebras:

In turn, on the side of Boolean spaces, we have a projective limit system formed by the maps $\tau_{\Delta'}$.

Corollary 5.9. *For every Boolean subalgebra $\Delta \subseteq \mathcal{Q}_{A,x}[\mathcal{N}]$, the set of maps*

$$\{\tau_{\Delta'} : A^* \to C^*_{\Delta'} \mid \Delta' \subseteq \Delta \text{ is a finite Boolean subalgebra}\} \tag{17}$$

forms a projective limit system, where the connecting morphisms are the homomorphisms of monoids $C^*_{\Delta_2} \twoheadrightarrow C^*_{\Delta_1}$ induced by the dual maps of inclusions $\Delta_1 \rightarrowtail \Delta_2$ of finite Boolean subalgebras of Δ. Moreover, the limit of this system is the map $\tau_\Delta : A^* \to X^*_\Delta$ sending a word $w \in A^*$ to the word $\gamma_1 \cdots \gamma_{|w|}$ with $\gamma_i = \{\phi \in \Delta \mid (w, i) \text{ satisfies } \phi\}$. In other words, γ_i is the projection to X_Δ of the principal (ultra)filter of $\mathcal{P}(A^* \otimes \mathbb{N})$ generated by (w, i).

Proof: The fact that (17) forms a projective limit system follows from Lemma 5.5.

Now, we remark that the maps $\tau_{\Delta'}$ are all length preserving, thus the above system factors into projective limit systems $\{\tau_{\Delta'} : A^n \to (C_{\Delta'})^n \mid \Delta' \subseteq \Delta \text{ is a finite Boolean subalgebra}\}$ for each $n \geq 0$, and each of these systems has itself a projective limit, which is induced by the projections $A^* \otimes \mathbb{N} \to X_\Delta$. Thus, the projective limit of (17) is the map $\tau_\Delta : A^* \to X^*_\Delta$ defined in the statement, where the space X^*_Δ is seen as the (disjoint) union over $n \geq 0$ of the spaces X^n_Δ, each one of those being a Boolean space when equipped with the product topology. ∎

The reader should now recall the extension of an lp-strain of languages $\mathcal{V}$ to profinite alphabets presented in Section 3, and in particular that such extension yields an embedding $\mathcal{V}(Y) \rightarrowtail \mathsf{Clopen}(Y^*)$ for every profinite alphabet Y.

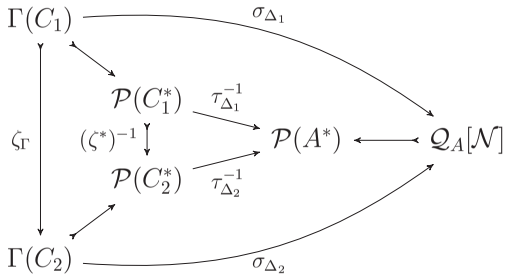
Corollary 5.10. *Let Γ be a class of sentences and $\Delta \subseteq \mathcal{Q}_{A,x}[\mathcal{N}]$ be a Boolean subalgebra. Then, the Boolean algebra $\Gamma \odot \Delta$ is a quotient of $\mathcal{V}_\Gamma(X_\Delta)$, or equivalently, $X_{\Gamma \odot \Delta}$ is isomorphic to a closed subspace of $X_{\mathcal{V}_\Gamma(X_\Delta)}$. More precisely, there exist commutative diagrams*

$$\begin{array}{ccc} \mathsf{Clopen}(X^*_\Delta) & \xrightarrow{\tau_\Delta^{-1}} & \mathcal{P}(A^*) \\ \uparrow & & \uparrow \\ \mathcal{V}_\Gamma(X_\Delta) & \twoheadrightarrow & \Gamma \odot \Delta \end{array} \qquad \begin{array}{ccc} \beta(A^*) & \xrightarrow{\beta\tau_\Delta} & \beta(X^*_\Delta) \\ \downarrow & & \downarrow \\ X_{\Gamma \odot \Delta} & \rightarrowtail & X_{\mathcal{V}_\Gamma(X_\Delta)} \end{array}$$

which are dual to each other.

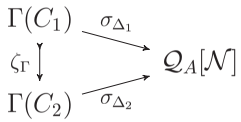
As a straightforward consequence of diagram (15), of Lemma 5.5, and of Proposition 5.2, we have the following:

Theorem 5.6. *Let $\Delta_1 \rightarrowtail \Delta_2$ be an embedding of finite Boolean subalgebras of $\mathcal{Q}_{A,x}[\mathcal{N}]$, and let $\zeta : C_2 \twoheadrightarrow C_1$ be its dual map. Then, the following diagram commutes*



Corollary 5.7. *If $\Delta_1 \subseteq \Delta_2$ are finite Boolean subalgebras of $\mathcal{Q}_{A,x}[\mathcal{N}]$, then the inclusion $\Gamma \odot \Delta_1 \subseteq \Gamma \odot \Delta_2$ holds. In particular, for every Boolean subalgebra $\Delta \subseteq \mathcal{Q}_{A,x}[\mathcal{N}]$, the family $\{\Gamma \odot \Delta' \mid \Delta' \subseteq \Delta$ is a finite Boolean subalgebra$\}$ forms a direct limit system.*

Proof: Let $\Delta_1 \rightarrowtail \Delta_2$ be an embedding of finite Boolean subalgebras of $\mathcal{Q}_{A,x}[\mathcal{N}]$, and $\zeta : C_2 \twoheadrightarrow C_1$ its dual map. Commutativity of the outer triangle of Theorem 5.6 yields



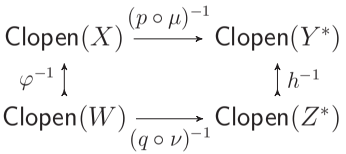
Thus, we have a direct system of Boolean subalgebras of $\mathcal{Q}_A[\mathcal{N}]$ and $\Gamma \odot \Delta_1 \subseteq \Gamma \odot \Delta_2$. ∎

This leads to the following definition.

Definition 5.8. For an arbitrary Boolean subalgebra $\Delta \subseteq \mathcal{Q}_{A,x}[\mathcal{N}]$, we set

$$\Gamma \odot \Delta = \varinjlim \{\Gamma \odot \Delta' \mid \Delta' \subseteq \Delta \text{ is a finite Boolean subalgebra}\}. \tag{16}$$

Note that $\Gamma \odot \Delta$ is simply given by the union of all Boolean subalgebras $\Gamma \odot \Delta' \subseteq \mathcal{Q}_A[\mathcal{N}]$. In particular, it is also a Boolean subalgebra of $\mathcal{Q}_A[\mathcal{N}]$.



Therefore, S being an lp-quotient of R means that every language recognized by S is also recognized by R, under the identification $\mathsf{Clopen}(Z^*) \subseteq \mathsf{Clopen}(Y^*)$.

Finally, we show that BiM-stamps over profinite alphabets can be described in terms of certain projective limits of BiM-stamps over finite alphabets. Although a bit technical, the proof mostly uses standard arguments used in the computation of projective limits in set-based categories.

Proposition 4.3. *Projective limits exist in the category of BiM-stamps with lp-morphisms. Moreover, each BiM-stamp is the projective limit of its BiM-stamp lp-quotients over finite alphabets.*

Proof: We give the general idea of the proof. All the missing details are routine computations. Let $\mathcal{F} = \{\mathsf{R}_i = (Y_i, \mu_i, M_i, p_i, X_i)\}_{i \in I}$ be a projective system in the category of BiM-stamps with lp-morphisms. For $i \geq j$, we denote by $\Phi_{i,j} = (h_{i,j}^*, g_{i,j}, \varphi_{i,j}) : \mathsf{R}_i \to \mathsf{R}_j$ the corresponding connecting morphism, with $h_{i,j} : Y_i \to Y_j$ a continuous function. Then, each of the families $\{Y_i\}_{i \in I}$, $\{M_i\}_{i \in I}$, and $\{X_i\}_{i \in I}$ forms itself a projective system, with the maps $h_{i,j}$, $g_{i,j}$ and $\varphi_{i,j}$ as connecting morphisms, respectively. We set

$$Y = \varprojlim \{Y_i \mid i \in I\}, \quad M_0 = \varprojlim \{M_i \mid i \in I\}, \quad \text{and} \quad X_0 = \varprojlim \{X_i \mid i \in I\}, \tag{8}$$

and for each $i \in I$, we denote by $h_i : Y \twoheadrightarrow Y_i$, $\zeta_i : M_0 \twoheadrightarrow M_i$, and $\pi_i : X_0 \twoheadrightarrow X_i$ the corresponding projections.

Using the explicit description of projective limits displayed in (1), we may check that there are well-defined maps

$$\mu_0 : Y^* \to M_0,\ z \mapsto (\mu_i \circ h_i^*(z))_{i \in I} \qquad \text{and} \qquad p_0 : M_0 \to X_0,\ m \mapsto (p_i \circ \zeta_i(m))_{i \in I}. \tag{9}$$

Graphically, we have the following commutative diagram:

$$\begin{array}{ccccc} Y^* & \xrightarrow{\mu_0} & M_0 & \xrightarrow{p_0} & X_0 \\ {\scriptstyle h_i^*}\downarrow & & \downarrow{\scriptstyle \zeta_i} & & \downarrow{\scriptstyle \pi_i} \\ Y_i^* & \xrightarrow[\mu_i]{} & M_i & \xrightarrow[p_i]{} & X_i \end{array}$$

Moreover, by continuity of $p_i \circ \mu_i \circ h_i^*$, for every $i \in I$, the map $p_0 \circ \mu_0$ is also continuous. We set

$$M = \mu_0[Y^*] \qquad \text{and} \qquad X = \overline{p_0[M]}.$$

Co-restricting μ_0 to M, we have an onto homomorphism $\mu : Y^* \twoheadrightarrow M$, and the restriction and co-restriction of p_0 to M and to X, respectively, yields a map $p : M \to X$ with dense image. We claim that $\mathsf{R} = (Y, \mu, M, p, X)$ is the projective limit of $\mathcal{F}$. The fact that R satisfies the universal property of projective limits is inherited from the universal properties satisfied by Y, M_0 and X_0.

Thus, it remains to check that M continuously bi-acts on X.

We denote by $\lambda_{i,m_i} : X_i \to X_i$ the (continuous) component at $m_i \in M_i$ of the left action of M_i on X_i. Then, for $m = (m_i)_{i\in I} \in M$, setting

$$\lambda_m : X \to X, \qquad x = (x_i)_{i\in I} \mapsto (\lambda_{i,m_i}(x_i))_{i\in I}$$

defines a continuous map which induces a left action of M on X. Indeed, for every $m, m' \in M$, we may compute $\lambda_m \circ p(m') = p(mm')$, using the analogous property for each λ_{i,m_i}. This proves not only that λ_m is well-defined (because $\lambda_m[X] = \lambda_m\left[\overline{p[M]}\right] \subseteq \overline{p[M]} = X$), but also that p is a morphism of sets with a left M-action. Similarly, we can define the right action of M on X. The compatibility between the left and right actions is inherited from the compatibility between the left and right actions for each R_i. Thus, R is a BiM-stamp.

Finally, that each BiM-stamp is the projective limit of all its lp-quotients over finite alphabets is an easy consequence of the explicit computation of projective limits just made. ∎

We remark that, even though each of the maps $\mu_i \circ h_i^*$ is onto, since Y^* is not compact, the map μ_0 defined in (9) is not onto in general (cf. [1, Lemma 1.2]). To see this, consider for instance, for each integer n, the BiM $(\{0\}^* \twoheadrightarrow \mathbb{Z}_n \xrightarrow{id} \mathbb{Z}_n)$. Then, we have a projective limit system $\mathcal{F} = \{(\{0\}^* \twoheadrightarrow \mathbb{Z}_n \xrightarrow{id} \mathbb{Z}_n)\}_{n\in\mathbb{N}}$ and M_0 (which equals X_0) defined in (8) is the additive profinite group $\widehat{\mathbb{Z}}$

(cf. diagram (2) and (LC.2)):

$$\begin{array}{ccc} \Gamma(C_1) & \rightarrowtail & \mathcal{P}(C_1^*) \\ {\scriptstyle \zeta_\Gamma}\downarrow & & \downarrow{\scriptstyle (\zeta^*)^{-1}} \\ \Gamma(C_2) & \rightarrowtail & \mathcal{P}(C_2^*) \end{array} \tag{15}$$

We also have:

Lemma 5.5. *The following diagram is commutative:*

$$\begin{array}{ccc} C_2^* & \xrightarrow{\tau_{\Delta_2}} & \\ {\scriptstyle \zeta^*}\downarrow & & A^* \\ C_1^* & \xrightarrow{\tau_{\Delta_1}} & \end{array}$$

Proof: We first recall from Proposition 5.2 that, for $i = 1, 2$, and $w \in A^*$, we have

$$\tau_{\Delta_i}(w) = \xi_i(w, 1) \dots \xi_i(w, |w|),$$

where $\xi_i : A^* \otimes \mathbb{N} \twoheadrightarrow C_i$ is the quotient dual to the embedding $\Delta_i \rightarrowtail \mathcal{P}(A^* \otimes \mathbb{N})$. On the other hand, by hypothesis, we have a commutative diagram

$$\begin{array}{ccc} \Delta_1 & \searrow & \\ \downarrow & & \mathcal{P}(A^* \otimes \mathbb{N}) \\ \Delta_2 & \nearrow & \end{array}$$

which dually gives

$$\begin{array}{ccc} C_1 & \xleftarrow{\xi_1} & \\ {\scriptstyle \zeta}\uparrow & & A^* \otimes \mathbb{N} \\ C_2 & \xleftarrow{\xi_2} & \end{array}$$

It then follows that

$$\begin{aligned} \tau_{\Delta_1}(w) &= \xi_1(w, 1) \dots \xi_1(w, |w|) \\ &= \zeta^*(\xi_2(w, 1) \dots \xi_2(w, |w|)) = \zeta^*(\tau_{\Delta_2}(w)), \end{aligned}$$

that is, the statement of the lemma is true. ∎

The existence of the map τ_Δ defined in Proposition 5.2 yields the next result.

Corollary 5.3 (Substitution Principle - local version)**.** *Let Γ be a class of sentences and $\Delta \subseteq \mathcal{Q}_{A,\mathbf{x}x}[\mathcal{N}]$ be a finite Boolean subalgebra. Then, the languages definable by a formula of $\Gamma \odot \Delta$ are precisely those of the form $\tau_\Delta^{-1}(K)$, where K is a language definable in $\Gamma(C_\Delta)$.*

Proof: This is an immediate consequence of the commutativity of the diagram dual to that of Proposition 5.2. ∎

Remark 5.4. We warn the reader that the operator $(_ \odot _)$ just defined does not coincide with the operator $(_ \circ _)$ considered both in [5] and in the regular version of [20]. The relationship between these operators is expressed by the following equality:

$$\Gamma \circ \Delta = \langle (\Gamma \odot \Delta) \cup \Delta_{\mathbf{x}} \rangle_{\mathsf{BA}},$$

where $\Delta_{\mathbf{x}}$ denotes the set of formulas of Δ in context $\mathbf{x}$ (i.e., those whose free variable belong to $\mathbf{x}$). We choose to first study the operator $(_ \odot _)$ in order to emphasize the role of Stone duality in the *Substitution Principle*. It should be clear for the reader how to state the corresponding results for $(_ \circ _)$.

5.2. The extension to arbitrary Boolean algebras using profinite alphabets. As the reader may have noticed, the context $\mathbf{x}$ fixed along the previous section is not playing an active role. For that reason, and in order to simplify the notation, we now assume that $\mathbf{x}$ is the empty context. Later, in Section 5.3, we will see that, if the logic at hand is expressive enough, this assumption may be done without loss of generality.

Here we show that substitution as defined in Section 5.1 extends to arbitrary Boolean algebras in a meaningful way. Fix a class of sentences Γ. We start by comparing the substitution maps obtained for two finite Boolean algebras, one contained in the other. To this end, suppose $\Delta_1 \rightarrowtail \Delta_2$ is such an inclusion of finite Boolean subalgebras of $\mathcal{Q}_{A,x}[\mathcal{N}]$ and let $\zeta : C_2 \twoheadrightarrow C_1$ be the dual of the inclusion. Recall that the semantics of logic on words provides embeddings of $\Gamma(C_i)$ in $\mathcal{P}(C_i^*)$ and that, since Γ is a class of sentences, the surjection ζ yields an embedding $\zeta_\Gamma \colon \Gamma(C_1) \rightarrowtail \Gamma(C_2)$ making the following diagram commute

(seen as a monoid). The map μ_0 is then the inclusion $\{0\}^* \rightarrowtail \widehat{\mathbb{Z}}$, which is not surjective. Finally, the projective limit of $\mathcal{F}$ is the BiM $(\{0\}^* \twoheadrightarrow \mathbb{N} \rightarrowtail \widehat{\mathbb{N}})$, where $\widehat{\mathbb{N}}$ denotes the closure of $\mathbb{N}$ in $\widehat{\mathbb{Z}}$.

4.2. Languages of marked words and quotienting operations. Recall that the set $A^* \otimes \mathbb{N}$ of marked words may be seen as a regular language in $(A \times 2)^*$. The fact that the syntactic morphism of this language is $\mu : (A \times 2)^* \to \{e, m, z\}$ given by $(a, 0) \mapsto e$ and $(a, 1) \mapsto m$ corresponds to saying that

$$(A \times 2)^* \cong A^* \uplus (A^* \otimes \mathbb{N}) \uplus A_z,$$

as defined in Section 2.9, and that concatenation on $(A \times 2)^*$ decomposes as follows: it yields biactions of A^* on each of these components, thus accounting for all pairs involving an element of A^*, and any other pair is sent to A_z.

Dually, the above disjoint union decomposition of $(A \times 2)^*$ yields the following Cartesian product decomposition

$$\mathcal{P}((A \times 2)^*) \cong \mathcal{P}(A^*) \times \mathcal{P}(A^* \otimes \mathbb{N}) \times \mathcal{P}(A_z),$$

We are interested in Boolean subalgebras $\mathcal{D}$ of $\mathcal{P}((A \times 2)^*)$. Dual to the inclusions of the disjoint components, we get projections of $\mathcal{D}$ onto $\mathcal{D}_0 = \{L \cap A^* \mid L \in \mathcal{D}\}$, $\mathcal{D}_1 = \{L \cap (A^* \otimes \mathbb{N}) \mid L \in \mathcal{D}\}$, and $\mathcal{D}_z = \{L \cap A_z \mid L \in \mathcal{D}\}$, respectively. While $\mathcal{D}$ is always contained in $\mathcal{D}_0 \times \mathcal{D}_1 \times \mathcal{D}_z$ it is not difficult to see that we have equality if and only if $\mathcal{D}$ contains the language $A^* \otimes \mathbb{N}$ and its orbit under the biaction of $(A \times 2)^*$.

We will be particularly interested in such algebras for which $\mathcal{D}_z = 2$. In this case, any component of the dual biaction by quotients on $\mathcal{P}((A \times 2)^*)$ with domain $\mathcal{P}(A_z)$ just sends the bounds to bounds of the appropriate component. The remaining components are as follows. First, we have the biactions of A^* on $\mathcal{P}(A^*)$ and on $\mathcal{P}(A^* \otimes \mathbb{N})$ given, for $u \in A^*$, respectively, by

$$\begin{array}{rcl} \ell_u^0 : \mathcal{P}(A^*) \to \mathcal{P}(A^*) \\ L \mapsto u^{-1}L \end{array} \qquad \begin{array}{rcl} r_u^0 : \mathcal{P}(A^*) \to \mathcal{P}(A^*), \\ L \mapsto Lu^{-1} \end{array} \tag{10}$$

and by

$$\begin{array}{rcl} \ell_u^1 : \mathcal{P}(A^* \otimes \mathbb{N}) \to \mathcal{P}(A^* \otimes \mathbb{N}) \\ L \mapsto u^{-1}L \end{array} \qquad \begin{array}{rcl} r_u^1 : \mathcal{P}(A^* \otimes \mathbb{N}) \to \mathcal{P}(A^* \otimes \mathbb{N}). \\ L \mapsto Lu^{-1} \end{array} \tag{11}$$

For every marked word (w,i), the biaction of A^* on $A^*\otimes\mathbb{N}$ also induces two functions $A^*\to A^*\otimes\mathbb{N}$, which are given by left and by right multiplication by (w,i). Dually, these define

$$\begin{aligned} \ell^1_{(w,i)} : \mathcal{P}(A^*\otimes\mathbb{N}) &\to \mathcal{P}(A^*) & r^1_{(w,i)} : \mathcal{P}(A^*\otimes\mathbb{N}) &\to \mathcal{P}(A^*). \\ L &\mapsto (w,i)^{-1}L & L &\mapsto L(w,i)^{-1} \end{aligned} \tag{12}$$

Lemma 4.4. *Let $\mathcal{D}\subseteq\mathcal{P}((A\times 2)^*)$ be a Boolean subalgebra such that $\mathcal{D}_z = 2$. Then, the following are equivalent:*

(a) $\mathcal{D}$ is closed under quotients and $(A^*\otimes\mathbb{N})\in\mathcal{D}$ or $A_z\in\mathcal{D}$;
(b) $\mathcal{D}\cong\mathcal{D}_0\times\mathcal{D}_1\times 2$, $\mathcal{D}_0$ and $\mathcal{D}_1$ are closed under the respective biactions of A^*, and for all $L\in\mathcal{D}_1$ and all $(w,i)\in A^*\otimes\mathbb{N}$ we have $(w,i)^{-1}L,\ L(w,i)^{-1}\in\mathcal{D}_0$.

Proof: Suppose we have (a). If $A_z\in\mathcal{D}$ then, as $\mathcal{D}$ is closed under quotients,

$$(a,1)^{-1}A_z = (A^*\otimes\mathbb{N})\cup A_z\in\mathcal{D},$$

and thus $A^*\otimes\mathbb{N}\in\mathcal{D}$. Since $\mathcal{D}$ contains the language $A^*\otimes\mathbb{N}$ and is closed under quotients, it also contains its orbit under the biaction of $(A\times 2)^*$. Thus, we have that $\mathcal{D}$ is isomorphic to $\mathcal{D}_0\times\mathcal{D}_1\times\mathcal{D}_z$ which, by hypothesis, is $\mathcal{D}_0\times\mathcal{D}_1\times 2$. The rest of the assertion in (b) follows from $\mathcal{D}$ being closed under quotients and the splitting of the biaction of $(A\times 2)^*$ on $\mathcal{P}((A\times 2)^*)$ as given by (10), (11), and (12).

Conversely, let us assume that (b) holds. Notice that having $\mathcal{D}\cong\mathcal{D}_0\times\mathcal{D}_1\times\mathcal{D}_z$ amounts to having that, for every $L_1,L_2,L_3\in\mathcal{D}$, there exists some $L\in\mathcal{D}$ such that

$$L_1\cap A^* = L\cap A^*,\qquad L_2\cap(A^*\otimes\mathbb{N}) = L\cap(A^*\otimes\mathbb{N}),\qquad\text{and}\qquad L_3\cap A_z = L\cap A_z.$$

So, in particular, by taking $L_1 = L_3 = \emptyset$ and $L_2 = (A\times 2)^*$, we may conclude that $A^*\otimes\mathbb{N}\in\mathcal{D}$. To show that $\mathcal{D}$ is closed under quotients, it suffices to consider quotients by letters. Also, since quotient operations are homomorphisms, it suffices to consider languages belonging to each component. By hypothesis $\mathcal{D}_0$ and $\mathcal{D}_1$ are closed under quotients by letters $a\in A$ and clearly so is 2 within $\mathcal{P}(A_z)$. Also by hypothesis, $(a,1)^{-1}L,\ L(a,1)^{-1}\in\mathcal{D}_0$ for $L\in\mathcal{D}_1$. Finally, for $L\in\mathcal{D}_0$ we have

$$(a,1)^{-1}L = L(a,1)^{-1} = \emptyset\in\mathcal{D}$$

and

$$(a,1)^{-1}A_z = A_z(a,1)^{-1} = (A^*\otimes\mathbb{N})\cup A_z\in\mathcal{D}. \quad \blacksquare$$

Proposition 5.2. *Let Γ be a class of sentences, Δ a finite Boolean subalgebra of $\mathcal{Q}_{A,\mathbf{x}x}[\mathcal{N}]$, and $\sigma_\Delta : \Gamma(C_\Delta)\to\Gamma\odot\Delta$ the associated substitution as defined above. Then, the function $\tau_\Delta : A^*\otimes\mathbb{N}^{|\mathbf{x}|}\to C_\Delta^*$ defined by $\tau_\Delta(w) = \xi_\Delta(w,1)\cdots\xi_\Delta(w,|w|)$ makes the following diagram commute:*

$$\begin{array}{ccc} \Gamma(C_\Delta) & \xrightarrow{\ \sigma_\Delta\ } & \Gamma\odot\Delta \\ \downarrow & & \downarrow \\ \mathcal{P}(C_\Delta^*) & \xrightarrow[\tau_\Delta^{-1}]{} & \mathcal{P}(A^*\otimes\mathbb{N}^{|\mathbf{x}|}) \end{array}$$

Proof: We show that the dual diagram

$$\begin{array}{ccc} X_{\Gamma\odot\Delta} & \xrightarrow{\ \Sigma_\Delta\ } & X_{\Gamma(C_\Delta)} \\ {\scriptstyle p_\Delta}\uparrow & & \uparrow{\scriptstyle q_\Delta} \\ A^*\otimes\mathbb{N}^{|\mathbf{x}|} & \xrightarrow[\tau_\Delta]{} & C_\Delta^* \end{array}$$

commutes. To this end, let $w\in A^*\otimes\mathbb{N}^{|\mathbf{x}|}$, $i\le|w|$, and $c\in C_\Delta$ with ϕ_c the corresponding atom of Δ. First, we argue that

$$(\tau_\Delta(w),i)\models\mathsf{P}_c(x)\iff(w,i)\models\phi_c.$$

This is so because, by the definition of letter predicates, the marked word over C_Δ^*, $(\tau_\Delta(w),i)$, is a model of $\mathsf{P}_c(x)$ if and only if its i-th letter is a c. By definition of τ_Δ, this is equivalent to having $\xi_\Delta(w,i)=c$, which, by (13), means that $(w,i)\models\phi_c$, as required.

Now, since the validity in a marked word of a quantified formula $Qx\ \psi$ is fully determined once we know the truth value of the given formula ψ at each point of the marked word, and since $\psi\in\Gamma(C_\Delta)$ and $\sigma_\Delta(\psi)$ are built up identically once the substitutions of $\mathsf{P}_c(x)$ by ϕ_c have been made, it follows that, for all $\psi\in\Gamma(C_\Delta)$, we have

$$\tau_\Delta(w)\in L_\psi\iff w\in L_{\sigma_\Delta(\psi)}. \tag{14}$$

However

$$w\in L_{\sigma_\Delta(\psi)}\iff L_{\sigma_\Delta(\psi)}\in p_\Delta(w)\iff L_\psi\in\Sigma_\Delta(p_\Delta(w))$$

so that

$$\tau_\Delta(w)\in L_\psi\iff L_\psi\in\Sigma_\Delta(p_\Delta(w))$$

and thus $\Sigma_\Delta(p_\Delta(w)) = q_\Delta(\tau_\Delta(w))$ as required. $\blacksquare$

induction that σ_Δ is a homomorphism of Boolean algebras. We denote the image of this homomorphism by $\Gamma \odot \Delta$. In the sequel we will consider σ_Δ as denoting the co-restriction of the substitution to its image, that is,

$$\sigma_\Delta : \Gamma(C_\Delta) \twoheadrightarrow \Gamma \odot \Delta.$$

Next we describe the languages of $\Gamma \odot \Delta$ via those of Γ and those of Δ. Since $\Gamma(C_\Delta)$ and $\Gamma \odot \Delta$ define, respectively, a Boolean algebra of languages over C_Δ and a Boolean algebra of marked words over A, we have embeddings

$$\begin{array}{ccc} \Gamma(C_\Delta) & \xrightarrow{\sigma_\Delta} & \Gamma \odot \Delta \\ \downarrow & & \downarrow \\ \mathcal{P}(C_\Delta^*) & & \mathcal{P}(A^* \otimes \mathbb{N}^{|\mathbf{x}|}) \end{array}$$

Our first goal is to prove that the substitution map σ_Δ extends to a complete homomorphism of Boolean algebras $\mathcal{P}(C_\Delta^*) \to \mathcal{P}(A^* \otimes \mathbb{N}^{|\mathbf{x}|})$. Formulated dually, this means we are looking for a map on the level of (marked) words $\tau_\Delta : A^* \otimes \mathbb{N}^{|\mathbf{x}|} \to C_\Delta^*$ making the following diagram commute:

$$\begin{array}{ccc} X_{\Gamma(C_\Delta)} & \xleftarrow{\Sigma_\Delta} & X_{\Gamma \odot \Delta} \\ {\scriptstyle p_\Delta}\uparrow & & \uparrow{\scriptstyle q_\Delta} \\ C_\Delta^* & \xleftarrow{\tau_\Delta} & A^* \otimes \mathbb{N}^{|\mathbf{x}|} \end{array}$$

Here the maps $p_\Delta \colon C_\Delta^* \to X_{\Gamma(C_\Delta)}$ and $q_\Delta \colon A^* \otimes \mathbb{N}^{|\mathbf{x}|} \to X_{\Gamma \odot \Delta}$ are, respectively, the restrictions of the dual maps of the embeddings $\Gamma(C_\Delta) \rightarrowtail \mathcal{P}(C_\Delta^*)$ and $\Gamma \odot \Delta \rightarrowtail \mathcal{P}(A^* \otimes \mathbb{N}^{|\mathbf{x}|})$.

In order to be able to define the map τ_Δ, we need to understand the Boolean algebra Δ via duality. Recall that $\mathcal{Q}_{A,\mathbf{x}x}[\mathcal{N}]$ embeds in $\mathcal{P}(A^* \otimes \mathbb{N}^{|\mathbf{x}x|})$ as a Boolean subalgebra, and therefore, so does Δ:

$$\Delta \rightarrowtail \mathcal{Q}_{A,\mathbf{x}x}[\mathcal{N}] \rightarrowtail \mathcal{P}(A^* \otimes \mathbb{N}^{|\mathbf{x}x|}).$$

Applying discrete duality to this composition, we obtain a map

$$\xi_\Delta : A^* \otimes \mathbb{N}^{|\mathbf{x}x|} \twoheadrightarrow C_\Delta = \mathrm{At}(\Delta)$$

defined, for $w \in A^* \otimes \mathbb{N}^{|\mathbf{x}|}$, $i \leq |w|$, $c \in C_\Delta$, and ϕ_c the atom of Δ corresponding to c, by

$$\xi_\Delta(w,i) = c \iff (w,i) \in L_{\phi_c} \iff (w,i) \vDash \phi_c. \tag{13}$$

Notice that in the case where $\mathcal{D}$ satisfies the equivalent conditions of Lemma 4.4, the Boolean subalgebra with atoms A^*, $A^* \otimes \mathbb{N}$ and A_z is a subalgebra of $\mathcal{D}$. Also note that the Boolean subalgebra of $\mathcal{P}((A \times 2)^*)$ generated by A^* is closed under quotients and it is such that $\mathcal{D}_z = 2$, but it does not satisfy the equivalent conditions of the the lemma.

Corollary 4.5. *Let $\mathcal{C} \subseteq \mathcal{P}(A^* \cup (A^* \otimes \mathbb{N}))$ be a Boolean subalgebra. Then, the Boolean subalgebra $\mathcal{D}$ of $\mathcal{P}((A \times 2)^*)$ closed under quotients generated by $\mathcal{C}$ is generated, as a lattice, by $\mathcal{S}_0 \cup \mathcal{S}_1 \cup \{A_z\}$, where*

$$\mathcal{S}_0 := \{u^{-1}Lv^{-1} \mid u, v \in A^*,\ L \in \mathcal{C}_0\}$$
$$\cup\ \{(w,i)^{-1}Lu^{-1},\ u^{-1}L(w,i)^{-1} \mid u \in A^*, (w,i) \in A^* \otimes \mathbb{N},\ L \in \mathcal{C}_1\} \subseteq \mathcal{P}(A^*)$$

and

$$\mathcal{S}_1 := \{u^{-1}Lv^{-1} \mid u, v \in A^*,\ L \in \mathcal{C}_1\} \subseteq \mathcal{P}(A^* \otimes \mathbb{N}).$$

Proof: Since the quotient operations are Boolean homomorphisms, $\mathcal{D}$ is generated as a Boolean algebra by the quotients of the languages in $\mathcal{C}$. Also, since, for $a \in A$ and $L \subseteq A^* \cup (A^* \otimes \mathbb{N})$, we have

$$a^{-1}L \subseteq A^* \cup (A^* \otimes \mathbb{N}) \ \text{ and } \ (a,1)^{-1}L \subseteq A^*,$$

it follows that $\mathcal{D}_z = 2$. Also, since $\mathcal{C}$ is a Boolean subalgebra of $\mathcal{P}(A^* \cup (A^* \otimes \mathbb{N}))$, it contains $A^* \cup (A^* \otimes \mathbb{N})$ and thus $\mathcal{D}$ contains A_z. It follows that Lemma 4.4 applies to $\mathcal{D}$ and the conclusion of the corollary easily follows. ∎

Another immediate consequence of the equivalence between (a) and (b) in Lemma 4.4 is the following:

Corollary 4.6. *Let $\mathcal{D} \subseteq \mathcal{P}((A \times 2)^*)$ be a Boolean subalgebra closed under quotients that contains $A^* \otimes \mathbb{N}$ and such that $\mathcal{D}_z = 2$. We let $\pi : (A \times 2)^* \to M_\mathcal{D}$ denote the syntactic morphism of $\mathcal{D}$ and we set*

$$M := \pi[A^*] \qquad \text{and} \qquad T := \pi[A^* \otimes \mathbb{N}].$$

Then,

- (a) *the biaction of $M_\mathcal{D}$ on $X_\mathcal{D}$ restricts and co-restricts to a biaction of M on $X_{\mathcal{D}_0}$ and to a biaction of M on $X_{\mathcal{D}_1}$;*
- (b) *for every $t \in T$, the components of the biaction of $M_\mathcal{D}$ on itself at t restrict and co-restrict as follows*

$$\lambda_t : X_{\mathcal{D}_0} \to X_{\mathcal{D}_1} \qquad \text{and} \qquad \rho_t : X_{\mathcal{D}_0} \to X_{\mathcal{D}_1}.$$

We remark that, if $\mathcal{D} \subseteq \mathcal{P}((A \times 2)^*)$ is a Boolean subalgebra closed under quotients then $\mathcal{D}_0$, seen as a Boolean subalgebra of $\mathcal{P}(A^*)$, is also closed under quotients: it is so because, for every $u, v \in A^*$ and $L \in \mathcal{D}$, the following equality holds:

$$u^{-1}(L \cap A^*)v^{-1} = (u^{-1}Lv^{-1}) \cap A^*.$$

Moreover, since the quotient $\mathcal{P}((A \times 2)^*) \twoheadrightarrow \mathcal{P}(A^*)$ restricts and co-restricts to a quotient $\mathcal{D} \twoheadrightarrow \mathcal{D}_0$, the syntactic morphism of $\mathcal{D}_0 \subseteq \mathcal{P}(A^*)$ is a restriction and co-restriction of the syntactic morphism of $\mathcal{D} \subseteq \mathcal{P}((A \times 2)^*)$. In particular, the monoid M defined in Corollary 4.6 is the syntactic monoid $M_{\mathcal{D}_0}$ of $\mathcal{D}_0$ and the biaction of M on $X_{\mathcal{D}_0}$ is simply the natural biaction of $M_{\mathcal{D}_0}$ on $X_{\mathcal{D}_0}$.

5. The Substitution Principle

The concept of substitution for the study of logic on words, as laid out by Tésson and Thérien in [20], is quite different from substitution in predicate logic. Substitution in predicate logic works on terms, whereas the notion of substitution in [20] works at the level of predicates. As such it provides a method for decomposing complex formulas into simpler ones. The core idea is to enrich the alphabet over which the logic is defined in order to be able to substitute large subformulas through letter predicates.

In this section we start by defining substitution maps with respect to finite Boolean algebras, and we prove a local version of the Substitution Principle (cf. Corollary 5.3), whose main ingredient is the duality between finite sets and finite Boolean algebras. Then, by proving that the substitution maps form a direct limit system, we are able to extend the construction to arbitrary Boolean algebras using full fledged Stone duality. In this process, we are naturally lead to consider profinite alphabets and we are able to state a global version of the Substitution Principle (cf. Corollary 5.11). Finally, in Section 5.4, we show how in practical terms these techniques may be useful in the study of fragments of logic.

5.1. Substitution with respect to finite Boolean algebras. As an example, consider the sentence $\psi = \exists x\ \phi(x)$. Then, ψ may be obtained from the sentence $\exists x\ \mathsf{P}_b(x)$ by replacing $\mathsf{P}_b(x)$ by $\phi(x)$, and thus, understanding ψ amounts to understanding both the sentence $\exists x\ \mathsf{P}_b(x)$ and the formula $\phi(x)$. If we want to substitute away several subformulas in this way, we must account for their logical relations. For instance, suppose that $\phi(x)$ is the conjunction $\phi_1(x) \wedge \phi_2(x)$, and that $\phi_1(x)$ and $\phi_2(x)$ are the simpler subformulas we wish to consider. Should ψ be obtained from a simpler sentence as above, such a sentence would be $\exists x\ (\mathsf{P}_{b_1}(x) \wedge \mathsf{P}_{b_2}(x))$ for two letters b_1, b_2. But, since ϕ_1 and ϕ_2 may be related we would then need to impose relations on letters. Then, no complexity is removed. Thus instead, we will consider a *finite Boolean algebra of formulas* to be substituted away. These can all be accounted for by having *a letter predicate for each atom* of this finite Boolean algebra. The fact that letter predicates model the atoms of a finite Boolean algebra of formulas in a free variable is built into their interpretation. This explains why, when substituting a formula of a given finite set $\mathcal{F}$ for each occurrence of a letter predicate from a corresponding alphabet in a sentence, we should only consider sets $\mathcal{F}$ of formulas that have the same logical behavior as letter predicates, meaning that $\mathcal{F}$ satisfies

(A.1) $\bigvee_{\phi \in \mathcal{F}} \phi$ is the *always-true* proposition;
(A.2) for every $\phi_1, \phi_2 \in \mathcal{F}$ distinct, $\phi_1 \wedge \phi_2$ is the *always-false* proposition.

In other words, we require that $\mathcal{F}$ is the set of atoms of the finite Boolean algebra it generates.

We now formalize this concept of substitution. Throughout this section we fix a context $\mathbf{x}$ and a variable x which does not belong to $\mathbf{x}$. Moreover, Γ will be a fixed class of sentences and $\Delta \subseteq \mathcal{Q}_{A,\mathbf{x}x}[\mathcal{N}]$ a finite Boolean subalgebra. We regard the set of atoms of Δ as a finite alphabet, and in order to emphasize both the fact that it is an alphabet and the fact that it is determined by Δ, we will denote it by C_Δ. On the other hand, when we wish to view an element c of C_Δ as a formula of Δ, we will write ϕ_c instead of c.

Definition 5.1. The Γ-*substitution given by* Δ *(with respect to the variable* x*)* is the map

$$\sigma_{\Gamma,\Delta} : \Gamma(C_\Delta) \to \mathcal{Q}_{A,\mathbf{x}}[\mathcal{N}]$$

sending a sentence to the formula in context $\mathbf{x}$ obtained by substituting for any occurrence of a letter predicate $\mathsf{P}_c(z)$, the formula $\phi_c[x/z]$ (that is, the formula obtained by substituting z for x in the formula $\phi_c \in \mathrm{At}(\Delta)$). Note that here we assume (without loss of generality) that things have been arranged so that the variables occurring in $\psi \in \Gamma(C_\Delta)$, such as z, do not occur in the formulas of Δ. When Γ is clear from the context, as will very often be the case, we will simply write σ_Δ instead of $\sigma_{\Gamma,\Delta}$.

Since the only constraints on the interpretation of letters in a word are given by the properties (A.1) and (A.2), it follows by a simple structural