THE PSEUDOFINITE MONADIC SECOND ORDER THEORY OF WORDS

DEACON LINKHORN

ABSTRACT. We analyse the pseudofinite monadic second order theory of words over a fixed finite alphabet. In particular we present an axiomatisation of this theory, working in a one-sorted first order framework. The analysis hinges on the fact that concatenation of words interacts nicely with monadic second order logic. More precisely, give a signature under which for each natural number $k$, equivalence of (monadic second order versions of) words with respect to formulas of quantifier depth at most $k$ is a congruence for concatenation. We use our analysis to present an alternative proof of a theorem connecting recognisable languages and finitely generated free profinite monoids via extended Stone duality, due to Gehrke, Grigorieff, and Pin.

CONTENTS

1. Introduction 1
2. Preliminaries 2
2.1. Boolean algebras and extended Stone duality 2
2.2. Monoids 3
2.3. Uppercase/lowercase notation and comprehension schema 5
3. Axiomatising the pseudofinite monadic second order theory of $\Sigma$-words 6
4. The free profinite monoid generated by $\Sigma$ 12
References 16

1. INTRODUCTION

In this paper we analyse the pseudofinite monadic second order theory of $\Sigma$-words for a fixed finite alphabet $\Sigma$, by which we mean the shared monadic second order theory of finite $\Sigma$-words. A (finite) $\Sigma$-word is a function $w : \alpha_w \rightarrow \Sigma$ where $\alpha_w$ is a (finite) linear order. In particular we give an axiomatisation $T_{M(\Sigma^*)}$ (definition 3.5) of the pseudofinite monadic second order theory of $\Sigma$-words, working in a one-sorted first order framework. To capture the monadic second order properties of a finite word $w : \alpha_w \rightarrow \Sigma$, we work with a first order structure $M(w)$ in a one-sorted signature $L^\Sigma_*$ (definition 3.2). The structure $M(w)$, which we refer to as the monadic second order version of $w$, is in particular an expansion of the Boolean
algebra of subsets of $\alpha_\omega$. The analysis hinges on the fact that concatenation of finite words interacts nicely with monadic second order logic, in particular when working in the signature $\mathcal{L}_\omega^{\Sigma}$ (definition 3.2). More precisely, we make use of the fact that for each $k \in \mathbb{N}$ equivalence of (monadic second order versions of) words with respect to $\mathcal{L}_\omega^{\Sigma}$-formulas of quantifier depth at most $k$ is a congruence for concatenation (proposition 3.15). The central result of the paper is theorem 3.23 in which we prove that the theory $T_{\mathcal{M}(\Sigma^*)}$ which we introduced is indeed an axiomatisation for the pseudofinite monadic second order theory of $\Sigma$-words.

As an application, we present an alternative proof of a theorem connecting recognisable languages and finitely generated free profinite monoids via extended Stone duality, due to Gehrke, Grigorieff, and Pin (theorem 4.3). Our proof is relatively self-contained and any duality-theoretic notions required are introduced in a minimal fashion as and when they are needed.

Related literature: A similar approach to studying the logical properties of words, making use of equivalence up to bounded quantifier depth being a congruence for concatenation, is used in [GS19] to explore the connection between star-free languages (a special collection of recognisable languages) and free aperiodic monoids (a special collection of profinite monoids). The main difference between the two papers is that in [GS19] it is first order logic on finite words, not monadic second order logic, which aligns with star-free languages and free aperiodic monoids.

In the author’s PhD thesis [Lin21] an axiomatisation is given for the shared monadic second order theory of finite words over a singleton alphabet, i.e. the shared monadic second order theory of finite linear orders. The connection between recognisable languages over this alphabet and the free profinite monoid on one generator, which arises as a special case of the theorem of Gehrke, Grigorieff, and Pin (theorem 4.3 hereafter referred to as the GGP theorem), was also considered in the thesis. There is a significant difference in perspective between the thesis work and the work presented here. In the thesis the GGP theorem was taken as a foundation, while the logical analysis of the shared monadic second order theory of finite linear orders was seen as offering a new perspective on this theorem. In this paper this viewpoint is inverted, we show that a logical analysis of the pseudofinite monadic second order theory of words in fact facilitates an alternative proof of the GGP theorem.

Our logical notation, wherever it is not explicitly defined or referenced, is borrowed from [Hod93].

2. Preliminaries

2.1. Boolean algebras and extended Stone duality.
Throughout the paper we assume that the reader is comfortable switching between viewing Boolean algebras as partial orders and as structures for the signature $\{\lor, \land, \bot, \top, c\}$.

The following material on extended Stone duality is adapted from [Han83] which works in the more general setting of spectral spaces and distributive lattices. We also confine our attention to the case of binary operations and ternary relations,
whereas the general theory allows for consideration of $n$-ary operations and $n+1$-ary relations.

**Definition 2.1.** Let $L$ be a Boolean algebra. A binary operation $\cdot : L^2 \to L$ is called,

1. **normal** if for each $a \in L$, $a \cdot \bot = \bot \cdot a = \bot$,
2. **additive** if for each $a, b, c \in L$, $(a \lor b) \cdot c = (a \cdot c) \lor (b \cdot c)$ and $a \cdot (b \lor c) = (a \cdot b) \lor (a \cdot c)$.

We view Boolean algebras with normal additive operations as structures for the signature $\mathcal{L} = \{ \lor, \land, \bot, \top, \cdot \}$. We define $\text{BAlgO}$ to be the category whose objects are Boolean algebras equipped with a normal additive binary operation, and whose morphisms are the usual homomorphisms for the signature $\mathcal{L}$.

**Definition 2.2.** Let $X$ be a Boolean space. A relation $R \subseteq X^3$ is called **compatible** if for each pair $(A, B) \in \text{Clop}(X)^2$ the set,

$$R(A, B, -) := \{ z \in X : \exists x \in A \exists y \in B \ R(x, y, z) \},$$

is clopen in $X$. In other words, $R \subseteq X^3$ is compatible if the the function $\mathcal{P}(X)^2 \to \mathcal{P}(X)$ defined by $(A, B) \mapsto R(A, B, -)$ restricts to $\text{Clop}(X)$. Given a ternary relation $R \subseteq X^3$ and $z \in X$ we write $R^{-1}(z)$ as shorthand for the set $\{ (x, y) \in X^2 : R(x, y, z) \}$. We define $\text{BoolComp}$ to be the category whose objects are Boolean spaces equipped with a compatible ternary relation, and whose morphisms $f : (X, R) \to (Y, S)$ are continuous functions $f : X \to Y$ such that for each $z \in X$,

$$S^{-1}(f(z)) = f(R^{-1}(z)).$$

Theorem 1.14 presents a more general version of the following theorem.

**Theorem 2.3.** The categories $\text{BAlgO}$ and $\text{BoolComp}$ are dual to one another.

Let $F$ be the functor $\text{BAlgO} \to \text{BoolComp}$ which for objects takes $(L, \cdot)$ to $(\text{UltraF}(L), R)$ where,

1. $\text{UltraF}(L)$ is the Boolean space with underlying set comprising ultrafilters of $L$ and with the topology having an open basis comprising sets $D(a) := \{ p \in \text{UltraF}(L) : a \notin p \}$ where $a \in L$,
2. $R := \{ (p, q, r) \in \text{UltraF}(L)^3 : \forall a \in p, \forall b \in q \ (a \cdot b \in r) \}$,

and for morphisms takes $f : L_1 \to L_2$ in $\text{BAlgO}$ to its preimage map $f^{-1} : \text{UltraF}(L_2) \to \text{UltraF}(L_1)$.

Let $G$ be the functor $\text{BoolComp} \to \text{BAlgO}$ which for objects takes $(X, R)$ to $(\text{Clop}(X), \cdot)$ where,

1. $\text{Clop}(X)$ is the Boolean algebra of clopen subsets of $X$,
2. $\cdot : \text{Clop}(X)^2 \to \text{Clop}(X)$ is the operation given by $A \cdot B := \{ z \in X : \exists x \in A, \exists y \in B \ R(x, y, z) \}$,

and for morphisms takes $f : X_1 \to X_2$ in $\text{BoolComp}$ to its preimage map $f^{-1} : \text{Clop}(X_2) \to \text{Clop}(X_1)$.

Then $FG$ is naturally isomorphic to the identity functor on $\text{BoolComp}$ and $GF$ is naturally isomorphic to the identity functor on $\text{BAlgO}$.

2.2. **Monoids.**

Our general setup for working with topological monoids and category-theoretic notions for the most part closely follows [Geh16] (in particular section 4.2), however our notation slightly differs in places. Most of the setup for this subsection can
be applied (and in Geh16  section 4.2, is applied) in the more general setting of universal algebra, but we will confine our attention to the case of monoids.

**Notation 2.4.** Let $X$ be a set and $\cdot : X^2 \to X$ a binary operation. We write $\sim$ for the binary operation on $\mathcal{P}(X)$ given by,

$$A \cdot B := \{a \cdot b : a, b \in A, b \in B\},$$

for each $A, B \in \mathcal{P}(X)$. We call $\sim$ the complexification of $\cdot$ (see Gol89, where algebras of complexes are discussed, for an explanation of the terminology).

**Definition 2.5.** A congruence on a monoid $(M, \cdot)$ is an equivalence relation $\rho$ on the underlying set $M$ such that for any $a, b, a', b' \in M$, $(a, a'), (b, b') \in \rho$ implies $(a \cdot b, a' \cdot b') \in \rho$. In other words, a congruence is an equivalence relation $\rho$ for which we get a well-defined binary operation on equivalence classes ‘pointwise’ (i.e. taking $[a]_\rho \cdot [b]_\rho := [a \cdot b]_\rho$). Note that for each congruence $\rho$ on $M$ the set of equivalence classes $M/\rho$ forms a monoid under this ‘pointwise’ operation.

We write $\text{Con}(M)$ for the collection of congruences on a monoid $M$, and $\text{Con}_\omega(M)$ for the subcollection of finite index congruences, i.e. those with finitely many equivalence classes. For $\rho \in \text{Con}(M)$, the homomorphism $\pi_\rho : M \to M/\rho$ given by $a \mapsto [a]_\rho$ is the called the canonical quotient map of $\rho$.

**Definition 2.6.** Let $X$ be a set and take two equivalence relations $\rho$ and $\theta$ on $X$. We say that $\rho$ refines $\theta$, or equivalently that $\theta$ is a coarsening of $\rho$, precisely when $\rho \subseteq \theta$, in other words when $x \sim_\rho x'$ implies $x \sim_\theta x'$. If $\rho$ and $\theta$ are equivalence relations on a set $X$, and $\rho$ refines $\theta$, then there is a unique function $p_{\rho, \theta} : X/\rho \to X/\theta$ such that $p_{\rho, \theta} \circ \pi_\rho = \pi_\theta$. The function $p_{\rho, \theta}$ is given by $p_{\rho, \theta}([x]_\rho) := [x]_\theta$, and is called the projection map.

**Definition 2.7.** A subset $S$ of a poset $(P, \leq)$ is called

- directed if for each $a, b \in S$ there exists $c \in S$ such that $a \leq c$ and $b \leq c$ (note in some parts of the literature this would instead be called upwards directed),

- cofinal if for each $a \in P$ there is $b \in S$ such that $a \leq b$.

**Remark 2.8.** For each monoid $M$, viewing $\text{Con}(M)$ as a partial order under $\supseteq$, the subcollection $\text{Con}_\omega(M)$ is a directed subset. In particular for any $\rho, \theta \in \text{Con}_\omega(M)$ the set-theoretic intersection $\rho \cap \theta$ is an element of $\text{Con}_\omega(M)$ with $\rho \supseteq \rho \cap \theta$ and $\theta \supseteq \rho \cap \theta$.

**Definition 2.9.** The category $\text{TopMon}$ of topological monoids has as objects triples $(M, \cdot, \tau)$ comprising a monoid $(M, \cdot)$ and a topology $\tau$ on $M$ such that $\cdot : M^2 \to M$ is continuous ($M^2$ is given the product topology). Arrows in $\text{TopMon}$ are functions $f : M \to M'$ which are monoid homomorphisms that are moreover continuous with respect to the topologies.

**Definition 2.10.** A directed system of topological monoids is (for us) a contravariant functor $F : I \to \text{TopMon}$ where $I$ is a directed partial order viewed as a category.

Recall that a cone for a (contravariant) functor $F : \mathcal{C} \to \mathcal{D}$ comprises an object $d \in \mathcal{D}$ and for each $c \in \mathcal{C}$ an arrow $f_c : d \to F(c)$, such that for each arrow $f : c' \to c$ in $\mathcal{C}$ the equality $F(f) \circ f_c = f_d$ holds.

The inverse limit (also known as projective limit) of a directed system of topological monoids $F : I \to \text{TopMon}$ is a cone $(M, (\tilde{\pi}_i)_{i \in I})$ for $F$ satisfying the universal property that for any other cone $(N, (g_i)_{i \in I})$ for $F$ there is a unique morphism
h : N → M such that for each i ∈ I we have g_i = \bar{\pi}_i \circ h. Note that unlike direct limits, which are in fact colimits, inverse limits really are limits according to the category theory lexicon. The inverse limit of the directed system F, which always exists in the category TopMon, is denoted by \( \lim F \).

**Lemma 2.11.** Let \( F : I \to \text{TopMon} \) be a directed system of topological monoids. If \( C \subseteq I \) is a cofinal subset then the restriction \( F' \) of \( F \) to \( C \) is a directed system of topological monoids and the two topological monoids \( \lim F \) and \( \lim F' \) are isomorphic.

**Proof.** See for example [Mac98], in particular Theorem 1 on page 217 (which works in a more general setting and under slightly different notation and setup). \( \square \)

**Definition 2.12.** Let \( M \) be a monoid. Let \( F_M : (\text{Con}_\omega(M), \supseteq) \to \text{TopMon} \) be the directed system given by taking \( F_M(\rho) \) to be the finite monoid \( M/\rho \) endowed with the discrete topology, and taking the projection map \( p_{\rho,\theta} : M/\rho \to M/\theta \) (see definition 2.6) whenever \( \theta \supseteq \rho \). (Note this is a contravariant functor, as \( F_M \) sends an arrow \( \rho \to \theta \) to an arrow \( M/\rho \to M/\theta \).) We write \( \hat{M} \) for the inverse limit \( \lim F_M \). The topological monoid \( \hat{M} \) is called the **profinite completion** of \( M \). In the case where \( M \) is the free monoid \( \Sigma^* \) on the set \( \Sigma \), the topological monoid \( \hat{\Sigma}^* \) is called the **free profinite monoid on** \( \Sigma \). In general a topological monoid \( (M, \cdot, \tau) \) is called **profinite** if there exists a directed system \( F : I \to \text{TopMon} \) such that for each \( i \in I \) the topological monoid \( F(i) \) is finite and carries the discrete topology, such that \( (M, \cdot, \tau) = \lim F \).

**Remark 2.13.** The choice of ‘profinite completion’ as terminology for the inverse limit \( \hat{M} \) may seem strange at first. This choice can be justified in multiple ways. In a very general setting, as is noted in [Geh16] pp 28, the forgetful functor from the category of profinite topological algebras of a given type into the category of algebras of that type admits a left adjoint, and this left adjoint sends an algebra \( M \) to its profinite completion. Alternatively, in the case of \( \Sigma^* \) there exist ‘profinite metrics’ (the distance between two words is inversely proportional to the minimal index of a finite congruence on \( \Sigma^* \) for which the two words reside in distinct equivalence classes) for which \( \hat{\Sigma}^* \) is the metric completion. For the latter approach, see for example [Pin09].

**2.3. Uppercase/lowercase notation and comprehension schema.**

We distinguish between upper and lower case variables as a notational shorthand for a certain relativisation of formulas (see [Hod93] Section 5.1 pp 202 for a general overview of the relativisation technique) in the setting of atomic Boolean algebras.

**Definition 2.14.** An **atom** of a partial order \((P, \leq)\) with a smallest element \( \perp \) is a minimal element of \( P \setminus \{ \perp \} \). A Boolean algebra, viewed as a partial order, is called **atomic** if every element lies above an atom. Note that for any partial order \((P, \leq)\) the atoms form a 0-definable subset of \( P \), and moreover this can be done uniformly. By uniformly we mean that there is a single \( \mathcal{L} = \{ \leq \} \) formula \( \text{At}(x) \) which defines the set of atoms in each partial order \((P, \leq)\). Later, for the purpose of eliminating quantifiers, we will simply include \( \text{At} \) as a unary predicate symbol in the signatures we work with.
Remark 2.15. For a Boolean algebra \( B \), being atomic is equivalent to the following holding,
\[
\forall x \forall y ( (\forall z (\text{At}(z) \rightarrow (z \leq x \leftrightarrow z \leq y))) \rightarrow x = y).
\]
This says, if two elements sit above the same atoms then they are equal.

Notation 2.16. To make formulae more easily readable when working over theories of (expansions of) atomic Boolean algebras we make use of some notational conventions. For atoms we reserve lower case letters \( x, y, z, \ldots \) for variables, and lower case letters \( a, b, c, \ldots \) for constants and parameters. For general elements we use upper case letters \( X, Y, Z, \ldots \) for variables and upper case letters \( A, B, C, \ldots \) for constants and parameters. Formally, the use of a lower case variable in a formula is shorthand for relativisation of that variable to the 0-definable set of atoms in that formula. Moreover, we write \( X(x) \) as shorthand for \( x \subseteq X \).

To illustrate with an example,
\[
\forall X \forall Y (\forall z (X(z) \leftrightarrow Y(z)) \rightarrow X = Y),
\]
is the rewriting of \((\dagger)\) using these notational shorthands.

Definition 2.17. Let \( L \) be a signature containing a binary relation symbol \( \subseteq \).
For each \( L \)-formula \( \phi(x, \bar{Y}) \), we define \( \text{Comp}_\phi \) to be the \( L \)-sentence,
\[
\forall \bar{Y} \exists Z \forall x (Z(x) \leftrightarrow \phi(x, \bar{Y})).
\]
(Recall that \( Z(x) \) is shorthand for \( x \subseteq Z \).)
If \( \mathcal{M} \) is an \( L \)-structure in which \( \subseteq \) is the ordering of an atomic Boolean algebra, then \( \text{Comp}_\phi \) says that for each tuple of parameters \( \bar{Y} \) there is an element \( Z \) of the Boolean algebra which lies above precisely the atoms \( x \) in the set defined by the formula \( \phi \) with parameters \( \bar{Y} \).

The \textbf{comprehension schema} for the signature \( L \) comprises the sentence \( \text{Comp}_\phi \) for each \( L \)-formula \( \phi \).

3. Axiomatizing the Pseudofinite Monadic Second Order Theory of \( \Sigma \)-words

We will make use of a one-sorted first-order setup to capture monadic second order structures. Space does not permit a detailed discussion of the motivations behind choosing this approach or the alternatives available. For these the reader is referred to the author’s PhD thesis [Lin21], in particular chapters 1 and 3.

Definition 3.1. Let \( \Sigma \) be a finite set. A \( \Sigma \)-\textbf{word} is a function \( w: \alpha_w \rightarrow \Sigma \) where \( \alpha_w \) is a linear order. We write \( \Sigma^* \) for the collection of finite \( \Sigma \)-words, including the unique empty \( \Sigma \)-word \( \varepsilon: \emptyset \rightarrow \Sigma \). For \( \sigma \in \Sigma \) we sometimes identify \( \sigma \) with the \( \Sigma \)-word \( \{\dagger\} \rightarrow \Sigma, \dagger \mapsto \sigma \).

Definition 3.2. We define \( L^\Sigma_{\ominus} \) to be the signature \( \{\subseteq, \preceq, \text{At}, \bot, (P_\sigma)_{\sigma \in \Sigma}\} \) in which,
- \( \subseteq \) and \( \preceq \) are binary relation symbols,
- \( \text{At} \) is a unary relation symbol,
- \( \bot \) is a constant symbol,
- \( P_\sigma \) is a unary relation symbol for each \( \sigma \in \Sigma \).

For each \( w: \alpha_w \rightarrow \Sigma \) from \( \Sigma^* \) we define \( M(w) \) to be the \( L^\Sigma_{\ominus} \)-structure in which,
1. the universe is \( \mathcal{P}(\alpha_w) \),
(2) \( \subseteq \) is the usual set-theoretic inclusion,
(3) \( \lessdot \) is \( \{ (A, B) \in \mathcal{P}(\alpha)^2 : a < b \text{ for some } a \in A, b \in B \} \), where < is the ordering of \( \alpha_w \),
(4) \( \exists < \) is \( \{ (A, B) \in \mathcal{P}(\alpha) : a < b \text{ for some } a \in A, b \in B \} \), i.e. the singleton subsets of \( \alpha_w \),
(5) \( \bot \) is the empty set,
(6) for each \( \sigma \in \Sigma \) the unary relation symbol \( P_\sigma \) is given the interpretation,
\[ \{ a : a \in w^{-1}(\sigma) \}, \]
i.e. \( P_\sigma \) is a collection of atoms of the Boolean algebra, namely those \( \{ a \} \) for which \( w(a) = \sigma \).

We call the \( \mathcal{L}_{\Sigma^+} \)-structure \( M(w) \) the monadic second order version of the word \( w \). The pseudofinite monadic second order theory of \( \Sigma \)-words is defined to be the \( \mathcal{L}_{\Sigma^+} \)-theory \( \bigcap_{w \in \Sigma^*} \text{Th}(M(w)) \). We write \( \text{Th}(M(\Sigma^*)) \) as a shorthand for this theory.

**Example 3.3.** Let \( \Sigma = \{ \sigma, \gamma, \delta \} \) and consider the \( \Sigma \)-word \( \sigma \gamma \sigma \). In the setup we use this word to be viewed as a function \( w : \{ 0, 1, 2 \} \to \{ \sigma, \gamma \} \). Then the monadic second order version \( M(w) \) comprises an expansion of the 8 element Boolean algebra \( \mathcal{P}(\{ 0, 1, 2 \}) \). In this expansion we have \( P_\sigma = \{ 0, 2 \} \), \( P_\gamma = \{ 1 \} \) and \( P_\delta = \emptyset \).

Rather than give our axiomatisation of \( \text{Th}(M(\Sigma^*)) \) immediately, we first introduce a weaker \( \mathcal{L}_{\Sigma^+} \)-theory \( T_{\text{base}(\Sigma)} \) which will serve as the setting for several technical lemmas.

**Definition 3.4.** We define an \( \mathcal{L}_{\Sigma^+} \)-theory \( T_{\text{base}(\Sigma)} \) to be the set of sentences expressing,
(1) atomic Boolean algebra (possibly trivial in the sense that \( \bot = \top \)) under \( \subseteq \), with the atoms picked out by the unary relation \( \text{At} \),
(2) \( \bot \) is the bottom element of the Boolean algebra,
(3) atoms linearly ordered (possibly trivially in the sense that the collection of atoms is empty) under \( \lessdot \),
(4) \( \forall X, Y (X \lessdot Y \iff \exists x \exists y (X(x) \land Y(y) \land x \lessdot y)) \),
(5) the relations \( P_\sigma \) only hold for atoms and,
\[ \forall x ( \bigvee_{\sigma \in \Sigma} P_\sigma(x) \land \bigwedge_{\sigma, \sigma' \in \Sigma, \sigma \neq \sigma'} \neg (P_\sigma(x) \land P_{\sigma'}(x))) \].

**Definition 3.5.** We define an \( \mathcal{L}_{\Sigma^+} \)-theory \( T_{M(\Sigma^*)} \) to be the extension of \( T_{\text{base}(\Sigma)} \) by sentences expressing,
(6) the linear ordering \( \lessdot \) on the atoms, if not trivial, is discrete with endpoints (we denote the smallest and largest atoms by \( 0 \) and \( 0^* \) respectively in this case),
(7) each element not equal to \( \bot \) lies above a smallest atom with respect to \( \lessdot \),
(8) the comprehension schema for \( \mathcal{L}_{\Sigma^+} \) (see definition 2.17).

**Remark 3.6.** There is some redundancy in our choice of axioms. For example, the discreteness of the linear order and the presence of endpoints can both be seen to follow from the other axioms.
For the remainder of this section, we prove that $T_{M(\Sigma^*)}$ axiomatises the pseudofinite monadic second order theory of linear order. Along the way some of the machinery required in section 4 will be introduced.

First we establish some lemmas working with our base theory $T_{\text{base}(\Sigma)} \subseteq T_{M(\Sigma^*)}$.

**Definition 3.7.** Let $(\alpha, <_\alpha)$ and $(\beta, <_\beta)$ be linear orders. We define $\alpha + \beta$ to be the linear ordering on the disjoint union $\alpha \sqcup \beta$ given by $<_\alpha \cup <_\beta \cup (\alpha \times \beta)$. Given $\Sigma$-words $w : \alpha \to \Sigma$ and $v : \beta \to \Sigma$, we define their concatenation to be the $\Sigma$-word $w \cdot v : \alpha + \beta \to \Sigma$ given by,

$$(w \cdot v)(i) := \begin{cases} w(i) & \text{if } i \in \alpha, \\ v(i) & \text{if } i \in \beta. \end{cases}$$

**Remark 3.8.** It is well known that the theory of Boolean algebras in the signatures $\{\subseteq\}$ and $\{\cup, \cap, \vee, \wedge, \top\}$ are bi-interpretable.

Given Boolean algebras $B_1$ and $B_2$, the direct product $B_1 \times B_2$ given by taking the pointwise ordering is again a Boolean algebra.

**Definition 3.9.** Given $L_\Sigma^\circ$-structures $M$ and $N$, we define an $L_\Sigma^\circ$-structure $M \otimes N$ as follows,

1. the universe of the structure is $M \times N$,
2. $\subseteq$ is defined as it is in the direct product, i.e. for all $(A, B), (C, D) \in M \otimes N$ we declare $M \otimes N \models (A, B) \subseteq (C, D)$ iff $M \models A \subseteq C$ and $N \models B \subseteq D$,
3. $\bot$ is interpreted in $M \otimes N$ as $(\bot, \bot)$,
4. $\top$ is interpreted in $M \otimes N$ as the collection of elements of the form $(A, \bot)$ where $A \in \text{At}^M$,
5. $\leq$ is given by $M \otimes N \models (A, B) \leq (C, D)$ if and only if,
   
   $M \models A = \bot$ and $N \models D = \bot$, or $M \models A \leq C$, or $N \models B \leq D$,
6. for each $\sigma \in \Sigma$, $P_\sigma$ is given by $M \otimes N \models P_\sigma((A, B))$ if and only if,
   
   $M \models P_\sigma(A)$ and $N \models B = \bot$, or, $M \models A = \bot$ and $N \models P_\sigma(B)$.

**Remark 3.10.** For any $w, v \in \Sigma^*$ we have $M(w) \otimes M(v) \cong M(w \cdot v)$. We can therefore view $w \mapsto M(w)$ as a homomorphism from $(\Sigma^*, \cdot)$ to the class of $L_\Sigma^\circ$-structures with the operation $\otimes$.

More generally, if $M, N \models T_{\text{base}(\Sigma^*)}$ then $M \otimes N$ is the unique $L_\Sigma^\circ$-structure expanding the direct product of the Boolean algebras underlying $M$ and $N$, in which $\leq$ and the predicates $P_\sigma$ are interpreted according to the concatenation of the words underlying $M$ and $N$.

Working over a finite signature, the following definition organises unnested first-order formulas into a union of finite sets according to the amount of nesting of quantifiers involved. As each formula is equivalent over the empty theory to an unnested formula ([Hod93] 2.6.2), this allows for inductive proofs targeting all formulas while only dealing with finitely many at a time.

**Definition 3.11** ([Hod93] pp. 103). Recall that an $L$-formula is called unnested if all of its atomic subformulae are of the form $x = y$, $c = y$, $F(\bar{x}) = y$, or $R(x)$ where $x, y, \bar{x}, \bar{y}$ comprise only variables, $c$ is a constant symbol of $L$, $F$ is a function symbol of $L$, and $R$ is a relation symbol of $L$.

Given an $L$-formula $\phi(\bar{x})$, the **quantifier depth** of $\phi$, denoted by $\text{qd}(\phi)$, is defined by induction on the complexity of $\phi$ as follows,
- $\text{qd}(\phi) = 0$ if $\phi$ is quantifier-free,
- $\text{qd}(\phi) = \max\{\text{qd}(\phi_i) : 1 \leq i \leq n\}$ if $\phi$ is a Boolean combination of $\mathcal{L}$-formulae $\phi_1, \ldots, \phi_n$,
- $\text{qd}(\exists \psi \phi) = \text{qd}(\exists \psi \phi) = \text{qd}(\phi) + 1$.

For each $k \in \mathbb{N}$ we can define an equivalence relation $\approx_k$ on the class of $\mathcal{L}$-structures by taking $\mathcal{M} \approx_k \mathcal{N}$ if and only if for each unnested $\mathcal{L}$-sentence $\phi$ of quantifier depth at most $k$, $\mathcal{M} \models \phi$ if and only if $\mathcal{N} \models \phi$.

**Remark 3.12.** If $\mathcal{L}$ is finite, then for each $k$ the equivalence relation $\approx_k$ on the class of $\mathcal{L}$-structures admits finitely many equivalence classes. This is because there are finitely many unnested sentences of quantifier depth at most $k$.

**Theorem 3.13** (Fraïssé-Hintikka Theorem [Hod93 Theorem 3.3.2]). Let $\mathcal{L}$ be a finite first order signature. Let $k \in \mathbb{N}$, and write $C_1, \ldots, C_n$ for the finitely many equivalence classes of $\approx_k$. Then there are unnested $\mathcal{L}$-sentences $\phi_1, \ldots, \phi_n$ of quantifier depth at most $k$, such that for each $\mathcal{L}$-structure $\mathcal{M}$ and each $i$ with $1 \leq i \leq n$,

$$\mathcal{M} \in C_i \iff \mathcal{M} \models \phi_i.$$ 

We refer to $\phi_1, \ldots, \phi_n$ as Hintikka-sentences.

**Remark 3.14.** Hintikka sentences axiomatising equivalence classes can be given effectively by an induction on $k$ but this is not required for our applications, see [Hod93] 3.3.2 for more details.

As each $\mathcal{L}$-formula is equivalent to an unnested $\mathcal{L}$-formula, it is clear that for two $\mathcal{L}$-structures $\mathcal{M}$ and $\mathcal{N}$ the following are equivalent,

1. $\mathcal{M} \equiv \mathcal{N}$,
2. for each $k \in \mathbb{N}$, $\mathcal{M} \approx_k \mathcal{N}$.

For each $k$, the equivalence relation $\approx_k$ can also be defined in terms of the second player having a winning strategy in the so-called unnested Ehrenfeucht-Fraïssé game of length $k$, which we denote by $G_k$, between two $\mathcal{L}$-structures. We make use of this equivalent formulation in the proof of the upcoming Proposition 3.15 but space does not permit a full definition of these games. A full account of the games can be found in [Hod93] pp. 102.

**Proposition 3.15.** Let $\mathcal{M}_1, \mathcal{M}_2, \mathcal{N}_1, \mathcal{N}_2 \models T_{\text{base}(\Sigma)}$. Let $k \in \mathbb{N}$. If $\mathcal{M}_1 \approx_k \mathcal{N}_1$ and $\mathcal{M}_2 \approx_k \mathcal{N}_2$ then $\mathcal{M}_1 \otimes \mathcal{M}_2 \approx_k \mathcal{N}_1 \otimes \mathcal{N}_2$.

**Proof.** We outline a winning strategy of $\exists$ (the player who moves second) for the unnested Ehrenfeucht-Fraïssé game of length $k$ played on the structures $\mathcal{M}_1 \otimes \mathcal{M}_2$ and $\mathcal{N}_1 \otimes \mathcal{N}_2$, i.e. for $G_k(\mathcal{M}_1 \otimes \mathcal{M}_2, \mathcal{N}_1 \otimes \mathcal{N}_2)$. The basic idea is to fuse moves taken from winning strategies for the games $G_k(\mathcal{M}_1, \mathcal{N}_1)$ and $G_k(\mathcal{M}_2, \mathcal{N}_2)$.

More explicitly, the strategy of $\exists$ is as follows. We give the case where $\forall$ makes a move in $\mathcal{M}_1 \otimes \mathcal{M}_2$, the case where $\forall$ moves in $\mathcal{N}_1 \otimes \mathcal{N}_2$ is dealt with symmetrically.

1. Choose winning strategies in the games $G_k(\mathcal{M}_1, \mathcal{N}_1)$ and $G_k(\mathcal{M}_2, \mathcal{N}_2)$, possible by the assumption that $\mathcal{M}_1 \approx_k \mathcal{N}_1$ and $\mathcal{M}_2 \approx_k \mathcal{N}_2$,
2. upon receiving the $i$-th move $A_i = (A_{i1}, A_{i2}) \in \mathcal{M}_1 \otimes \mathcal{M}_2$ from $\forall$, take both $A_{i1}$ and $A_{i2}$ and feed them into simulations of the games $G_k(\mathcal{M}_1, \mathcal{N}_1)$ and $G_k(\mathcal{M}_2, \mathcal{N}_2)$ respectively,
3. take responses $B_{i1} \in \mathcal{N}_1$ and $B_{i2} \in \mathcal{N}_2$ dictated by the winning strategies for $G_k(\mathcal{M}_1, \mathcal{N}_1)$ and $G_k(\mathcal{M}_2, \mathcal{N}_2)$ respectively which were chosen at the outset,
form the move \( B_i = (B_{i1}, B_{i2}) \in \mathcal{M}_1 \otimes \mathcal{M}_2 \) as the response.

Suppose that the strategy just described is deployed, and the resulting play of the game is \((A_1, B_1), \ldots, (A_k, B_k)) \in ((\mathcal{M}_1 \otimes \mathcal{M}_2) \times (\mathcal{M}_1 \otimes \mathcal{M}_2))^k\), with \( A_i = (A_{i1}, A_{i2}) \in \mathcal{M}_1 \times \mathcal{M}_2 \) and \( B_i = (B_{i1}, B_{i2}) \in \mathcal{M}_1 \times \mathcal{M}_2 \) for each \( i \) such that \( 1 \leq i \leq k \).

We must check that for each unnested atomic \( \mathcal{L}_\mathcal{N}^\exists \)-formula \( \phi(X) \) we have,
\[
\mathcal{M}_1 \otimes \mathcal{M}_2 \models \phi(\bar{A}) \iff \mathcal{M}_1 \otimes \mathcal{M}_2 \models \phi(\bar{B}).
\]

This can be done on a case by case basis. Unnested atomic formulae in \( \bar{X} \) are of the form \( X = X', X \subseteq X', \text{At}(X), X \ll X', \text{or } P_r(X) \), where \( X, X' \in \bar{X} \cup \{ \bot \} \).

We deal with the case \( X \ll X' \), and leave the remaining cases as an exercise for the reader. Suppose,
\[
\mathcal{M}_1 \otimes \mathcal{M}_2 \models (A_{i1}, A_{i2}) \ll (A_{j1}, A_{j2}).
\]

By definition of \( \otimes \) this is equivalent to at least one of the following conditions holding,
\[
\begin{align*}
(1) \mathcal{M}_1 \not\models A_{i1} &= \bot \text{ and } \mathcal{M}_2 \not\models A_{j2} = \bot, \\
(2) \mathcal{M}_1 \models A_{i1} &\ll A_{j1}, \\
(3) \mathcal{M}_2 \models A_{j2} &\ll A_{j2}.
\end{align*}
\]

Now by our assumption that the pairs \((\bar{A}_1, \bar{B}_1)\) and \((\bar{A}_2, \bar{B}_2)\) are winning plays for \( \exists \) in \( G_k[\mathcal{M}_1, \mathcal{N}_1] \) and \( G_k[\mathcal{M}_2, \mathcal{N}_2] \) respectively, these conditions are equivalent to the following respectively,
\[
\begin{align*}
(1) \mathcal{N}_1 \not\models B_{i1} &= \bot \text{ and } \mathcal{N}_2 \not\models B_{j2} = \bot, \\
(2) \mathcal{N}_1 \models B_{i1} &\ll B_{j1}, \\
(3) \mathcal{N}_2 \models B_{j2} &\ll B_{j2}.
\end{align*}
\]

By definition of \( \otimes \), the disjunction of these conditions is equivalent to,
\[
\mathcal{N}_1 \otimes \mathcal{N}_2 \models (B_{i1}, B_{i2}) \ll (B_{j1}, B_{j2}).
\]

By symmetry we can therefore conclude,
\[
\mathcal{M}_1 \otimes \mathcal{M}_2 \models (A_{i1}, A_{i2}) \ll (A_{j1}, A_{j2}) \iff \mathcal{N}_1 \otimes \mathcal{N}_2 \models (B_{i1}, B_{i2}) \ll (B_{j1}, B_{j2}).
\]

Hence we have outlined a winning strategy for \( \exists \) in \( G_k[\mathcal{M}_1 \otimes \mathcal{M}_2, \mathcal{N}_1 \otimes \mathcal{N}_2] \).

**Lemma 3.16.** Let \( \mathcal{C} \) be a class of \( \mathcal{L} \)-structures. For each \( \mathcal{L} \)-structure \( \mathcal{M} \) the following are equivalent,
\[
\begin{align*}
(1) \ &\mathcal{M} \models \bigcap_{\mathcal{N} \in \mathcal{C}} \text{Th}(\mathcal{N}), \\
(2) \ &\text{Th}(\mathcal{M}) \subseteq \bigcup_{\mathcal{N} \in \mathcal{C}} \text{Th}(\mathcal{N}).
\end{align*}
\]

**Proof.** We write \( \text{Th}(\mathcal{C}) \) for \( \bigcap_{\mathcal{N} \in \mathcal{C}} \text{Th}(\mathcal{N}) \).

(1) implies (2):
Towards a contradiction suppose that \( \mathcal{M} \models \text{Th}(\mathcal{C}) \) holds while \( \text{Th}(\mathcal{M}) \subseteq \bigcup_{\mathcal{N} \in \mathcal{C}} \text{Th}(\mathcal{N}) \) fails. Then there exists a sentence \( \phi \in \text{Th}(\mathcal{C}) \setminus \bigcup_{\mathcal{N} \in \mathcal{C}} \text{Th}(\mathcal{N}) \). But now note that for any sentence \( \phi, \phi \notin \bigcup_{\mathcal{N} \in \mathcal{C}} \text{Th}(\mathcal{N}) \) implies that \( \neg \phi \in \bigcap_{\mathcal{N} \in \mathcal{C}} \text{Th}(\mathcal{N}) = \text{Th}(\mathcal{C}) \). Therefore we get that \( \mathcal{M} \models \phi \land \neg \phi \), giving us a contradiction.

(2) implies (1):
Towards a contradiction suppose that \( \text{Th}(\mathcal{M}) \subseteq \bigcup_{\mathcal{N} \in \mathcal{C}} \text{Th}(\mathcal{N}) \) while \( \mathcal{M} \not\models \text{Th}(\mathcal{C}) \). Then there exists a sentence \( \phi \in \text{Th}(\mathcal{C}) \setminus \text{Th}(\mathcal{M}) \). Therefore \( \neg \phi \in \text{Th}(\mathcal{M}) \subseteq \bigcup_{\mathcal{N} \in \mathcal{C}} \text{Th}(\mathcal{N}) \). This implies that for some \( \mathcal{N} \in \mathcal{C} \) we have \( \mathcal{N} \models \phi \land \neg \phi \), giving us a contradiction. \( \square \)
Remark 3.17. Lemma 3.16 can be rephrased into a more order-theoretic presentation. If $\mathcal{B}$ is a Boolean algebra, \{\mathcal{V}_i : i \in I\} is a collection of ultrafilters of $\mathcal{B}$, and $\mathcal{U}$ is an ultrafilter of $\mathcal{B}$, then the following are equivalent,

(1) $\bigcap_{i \in I} \mathcal{V}_i \subseteq \mathcal{U}$,
(2) $\mathcal{U} \subseteq \bigcup_{i \in I} \mathcal{V}_i$.

It is well known that each Boolean algebra is isomorphic to the Lindenbaum-Tarski algebra of some theory $T$ in some signature $\mathcal{L}$ (see forthcoming definition 4.33). Ultrafilters of the Lindenbaum-Tarski algebra of a theory $T$ are precisely complete extensions of $T$, i.e. theories of the form $\text{Th}(\mathcal{M})$ for some $\mathcal{M} \models T$. Therefore the order-theoretic presentation is equivalent to the logical presentation lemma 3.16.

Definition 3.18. Let $\mathcal{M} \models T_{\Sigma^*}$. For each $a \in \text{At}(\mathcal{M})$ it follows immediately from the comprehension scheme (in particular from Comp$\phi$ for $\phi(x,a) : x \not\leq a \lor x = a$) that the initial interval induced by $a$ is an element of $\mathcal{M}$. We denote this element by $[0,a]$. Note that $[0,a]$ is in the definable closure of the element $a$.

Lemma 3.19. Let $\mathcal{M} \models T_{\Sigma^*}$. For each $a \in \text{At}(\mathcal{M})$, the set of elements which are beneath the element $[0,a]$ with respect to the ordering $\subseteq$ form the underlying set of a substructure of $\mathcal{M}$. We denote this substructure by $\mathcal{M} \restriction [0,a]$.

By relativisation of quantifiers (see [Hod92], Section 5.1 pp 202), for each $\mathcal{L}_*^\Sigma$-formula $\phi(\bar{X})$ there is a formula $\phi(\bar{X}) \restriction [0,x]$ where $x$ is a ‘fresh’ variable (i.e. $x \notin \bar{X}$) such that for each $a \in \text{At}(\mathcal{M})$ and $\bar{A} \in \mathcal{M} \restriction [0,a]$,

$\mathcal{M} \restriction [0,a] \models \phi(\bar{A}) \iff \mathcal{M} \models \phi(\bar{A}) \restriction [0,a]$.

This generalises in the following way. For each $B \in \mathcal{M}$ the set of elements beneath $B$ form the underlying set of a substructure of $\mathcal{M}$. We denote this sub-structure by $\mathcal{M} \restriction B$. Relativising quantifiers, for each $\mathcal{L}_*^\Sigma$-formula $\phi(\bar{X})$ there is a formula $\phi(\bar{X}) \restriction Y$ where $Y \notin \bar{X}$ such that for each $B \in \mathcal{M}$ and each $\bar{A} \in \mathcal{M} \restriction B$,

$\mathcal{M} \restriction B \models \phi(\bar{A}) \iff \mathcal{M} \models \phi(\bar{A}) \restriction B$.

Proof. This is a straightforward application of relativisation of quantifiers.

Definition 3.20. We write $S_k$ for the set of equivalence classes of the equivalence relation $\approx_k$ which contain a structure of the form $M(w)$ where $w \in \Sigma^*$. For each $k \in \mathbb{N}$ the set $S_k$ is closed under $\hat{\otimes}$. Moreover for each $k \in \mathbb{N}$ the structure $(S_k, \hat{\otimes})$ is a monoid.

Lemma 3.21. For each $k \in \mathbb{N}$ the set $S_k$ is closed under $\hat{\otimes}$. Moreover for each $k \in \mathbb{N}$ the structure $(S_k, \hat{\otimes})$ is a monoid.

Proof. Suppose that $\mathcal{M}$ and $\mathcal{N}$ are $\mathcal{L}_*^\Sigma$-structures such that $\mathcal{M} \approx_k M(w)$ and $\mathcal{N} \approx_k M(v)$ for some $w,v \in \Sigma^*$. Then by proposition 3.13 it follows that $\mathcal{M} \otimes \mathcal{N} \approx_k M(w) \otimes M(v) \cong M(w \cdot v)$, hence $S_k$ is indeed closed under $\hat{\otimes}$. For each $k \in \mathbb{N}$ the operation $\hat{\otimes}$ on $S_k$ is associative and admits the equivalence class containing $M(\varepsilon)$ as an identity element, hence $(S_k, \hat{\otimes})$ is a monoid.

Notation 3.22. For each $k \in \mathbb{N}$, we will write $\otimes_k$ for the restriction of $\hat{\otimes}$ to $S_k$.

Theorem 3.23. The theory $T_{M(\Sigma^*)}$ is an axiomatisation of the pseudofinite monadic second order theory of $\Sigma$-words, $\bigcap_{w \in \Sigma^*} \text{Th}(M(w))$.

Proof. It is clear that for each $w \in \Sigma^*$, $M(w) \models T_{M(\Sigma^*)}$, and hence $T_{M(\Sigma^*)} \subseteq \text{Th}(M(\Sigma^*))$. So we must show that if $\mathcal{M}$ is an $\mathcal{L}_*^\Sigma$-structure and $\mathcal{M} \models T_{M(\Sigma^*)}$ then $\mathcal{M} \models \text{Th}(M(\Sigma^*))$. By lemma 3.16 $\mathcal{M} \models \text{Th}(M(\Sigma^*))$ is equivalent to $\mathcal{M} \models \text{Th}(M(\Sigma^*)) \subseteq \bigcup_{w \in \Sigma^*} \text{Th}(M(w))$. 

Fix $\mathcal{M} \models T_{M(\Sigma^*)}$. It is enough to prove that for each $k \in \mathbb{N}$ there exists $w_k \in \Sigma^*$ (depending on $\mathcal{M}$) such that $\mathcal{M} \models \psi_k(w_k)$. Then $\text{Th}(\mathcal{M}) \subseteq \bigcup_{k \in \mathbb{N}} \text{Th}(M(w_k)) \subseteq \bigcup_{w \in \Sigma^*} \text{Th}(M(w))$.

For each $k \in \mathbb{N}$, there is a sentence $\psi_k$ such that for each $\mathcal{L}_k^0$-structure $\mathcal{M}$ the following are equivalent,

- $\mathcal{M} \models \psi_k$,
- $\mathcal{M} \models \psi_k(w)$ for some $w \in \Sigma^*$.

The sentence $\psi_k$ is given by taking the Hintikka sentences $\phi_1, \ldots, \phi_n$ axiomatising the equivalence classes $C_1, \ldots, C_n$ of $\approx_k$, and forming the disjunction of those $\phi_i$ such that $C_i$ contains $M(w)$ for some $w \in \Sigma^*$.

Consider the formula $\psi_k \models [0, x]$. It is clear that for each $\mathcal{M} \models T_{M(\Sigma^*)}$ we have that $\mathcal{M} \models [0, 0] \models M(\sigma)$ for some $\sigma \in \Sigma^*$. Clearly $M(\sigma) \models \psi_k$ for each $k \in \mathbb{N}$, hence $\mathcal{M} \models \psi_k \models [0, 0]$. On the other hand, $\mathcal{M} \models [0, 0^*] = \mathcal{M}$, so that $\mathcal{M} \models \psi_k$ is equivalent to $\mathcal{M} \models \psi_k \models [0, 0^*]$.

Suppose towards a contradiction that $\mathcal{M} \models \psi_k$. Equivalently, suppose $\mathcal{M} \not\models \psi_k \models [0, 0^*]$. By comprehension there is an element of $\mathcal{M}$ which lies above precisely those $a \in \text{At}(\mathcal{M})$ for which $\mathcal{M} \not\models \psi_k \models [0, a]$. This element cannot be $\bot$, as we supposed (towards a contradiction) that it contains $0^*$. Therefore, by the axioms of $T_{M(\Sigma^*)}$, it must contain a smallest atom $a \in \text{At}(\mathcal{M})$ (with respect to $\preceq$). Now since $\mathcal{M} \models \psi_k \models [0, 0]$ we must have $a \neq 0$. As $a \neq 0$, we can take its immediate predecessor $b \in \text{At}(\mathcal{M})$. We then get a contradiction using proposition 3.13 and the observation that $\mathcal{M} \models [0, a] \models [0, b] \otimes M(\sigma)$ where $\sigma$ is the unique element of $\Sigma^*$ such that $\mathcal{M} \models \psi_k \models [0, b]$. Since $\mathcal{M} \models [0, b] \models \psi_k$ there exists $w \in \Sigma^*$ such that $\mathcal{M} \models [0, b] \models \psi_k \models M(w)$, and therefore $\mathcal{M} \models [0, a] \models \psi_k \models [0, b] \models \psi_k \models M(w)'$ where $w' \in \Sigma^*$ is given by appending $\sigma$ to $w$. This then implies $\mathcal{M} \models [0, a] \models \psi_k$, whence we have arrived at a contradiction.

Therefore $\mathcal{M} \models \psi_k$ as required.

**Remark 3.24.** We write $S(T_{M(\Sigma^*)})$ for the space of complete theories extending $T_{M(\Sigma^*)}$, i.e. the space with underlying set comprising theories of the form $\text{Th}(\mathcal{M})$ for some $\mathcal{M} \models T_{M(\Sigma^*)}$ and with topology given by a basis of clopen subsets of the form $\langle \phi \rangle \setminus \{ T \in S(T_{M(\Sigma^*)}) : T \models \phi \}$. For each $k \in \mathbb{N}$, by theorem 3.23 $\text{Th}(\mathcal{M}) \to \{ \mathcal{M} \models \psi_k \}$ is a well-defined function $S(T_{M(\Sigma^*)}) \to S_k$.

The Fraïssé-Hintikka theorem tells us that this function is continuous when we equip $S_k$ with the discrete topology. This is because it tells us that the preimage of each singleton subset of $S_k$ is axiomatised by a sentence, and hence is basic clopen.

### 4. The Free Profinite Monoid Generated by $\Sigma$

**Definition 4.1.** Let $\Sigma$ be a finite set. A subset $L$ of $\Sigma^*$ is called a **recognisable** $\Sigma$-language (or a recognisable subset of $\Sigma^*$) if there exists a finite semigroup $S$, a subset $F \subseteq S$, and a homomorphism $f : \Sigma^* \to S$ such that $L = f^{-1}(F)$.

**Remark 4.2.** It is well-known that the collection $\text{Rec}(\Sigma)$ of recognisable subsets of $\Sigma^*$ forms a Boolean algebra under the set-theoretic operations of intersection, union, and complement (with the empty set and $\Sigma^*$ being the bottom and top elements respectively).

Recognisable subsets can also be characterised as the languages accepted by certain finite automata, or as those sets of finite words which are finitely axiomatisable in monadic second order logic (see theorem 4.6).
It is also well known that the collection of recognisable languages is closed under the complexification of concatenation $\hat{\cdot}$ : $\mathcal{P}(\Sigma^*)^2 \to \mathcal{P}(\Sigma^*)$ (see notation 2.4). It is also straightforward to check that the complexification of concatenation is both normal and additive (see definition 2.1). It therefore makes sense to talk about the extended Stone dual of $(\operatorname{Rec}(\Sigma), \hat{\cdot})$ in the sense of section 2.4.

In [GGP10] the following theorem is stated, but a proof is not given.

**Theorem 4.3** ([GGP10] Theorem 6.1). *The Boolean algebra $\operatorname{Rec}(\Sigma)$ of recognisable $\Sigma$-languages together with (the complexification of) the binary operation of concatenation is the extended Stone dual of $\hat{\Sigma^*}$, the free profinite monoid on $\Sigma$.*

Here the Boolean space underlying $\hat{\Sigma^*}$ is viewed together with the graph of the monoid operation which it carries, which is a compatible ternary relation. It is said in [GGP10] that a proof would require advanced machinery from duality theory.

This theorem can be viewed as a special case of the following considerably more general theorem (by taking $A$ to be $\Sigma^*$).

**Theorem 4.4** ([Geh16] Theorem 4.5). *Let $A$ be an abstract algebra. The profinite completion $\hat{A}$ (which is defined in a similar way to the profinite completion of a monoid in definition 2.12) is, up to isomorphism as topological algebras, the extended Stone dual of $\operatorname{Rec}(A)$, the Boolean algebra of recognisable subsets of $A$ equipped with residuation operations of recognisable subsets.*

Discussion of residuation is outside the scope of this article, and is unnecessary for the specific cases we are interested in (free profinite monoids on finitely many generators). A full proof of this theorem is presented at length in [Geh16] alongside other results from duality theory. The purpose of this section is to offer a proof of the more specific theorem from [GGP10], making use of our model-theoretic analysis of $\operatorname{Th}(M(\Sigma^*))$ to present a concise proof which does not rely on advanced machinery from duality theory.

**Definition 4.5.** *Let $T$ be an $\mathcal{L}$-theory. The *Lindenbaum-Tarski algebra* of $T$, which we denote by $\operatorname{LT}(T)$, is defined to be the Boolean algebra of equivalence classes of $\mathcal{L}$-sentences under the equivalence relation $\sim_T$ given by $\phi \sim_T \psi$ if and only if,

$$T \models \phi \iff \psi.$$

Note that this tacitly exploits the fact that $\sim_T$ is a congruence (see definition 2.5, the generalisation of which to the setting of universal algebra is straightforward) for the Boolean operations of conjunction, disjunction, and negation. When working with $\operatorname{LT}(T)$ we write $\phi$ in place of $[\phi]_{\sim_T}$ to avoid bloated notation.

The following is a concise (but anachronistic) formulation of a theorem of Büchi on the monadic second order logic of finite words. It conceals the automata theoretic machinery used in the original proofs. It is also worth noting that Büchi did not work in the signature $\mathcal{L}_{\Sigma^*}$ which we make use of here.

**Theorem 4.6** (Büchi). *Let $\Sigma$ be a finite set. The map $\operatorname{SPEC} : \operatorname{LT}(M(\Sigma^*)) \to \operatorname{Rec}(\Sigma)$ given by $\phi \mapsto \{w \in \Sigma^* : M(w) \models \phi\}$ is an isomorphism of Boolean algebras.*

In the following proposition we exploit the shorthand for restrictions and relativisations laid out in lemma 3.19.
**Proposition 4.7.** The map \( \text{SPEC} \) is still an isomorphism if we enrich the two Boolean algebras as follows,

(i) to \( \text{LT}(T_{M(\Sigma^*)}) \) we append the operation \( + \) given by,

\[
\phi + \psi := \exists X (\text{dwl}(X) \land \phi \land X \land \psi \land X^c),
\]

where \( \text{dwl}(X) \) (downwards closed) is the formula,

\[
\forall y, z((X(y) \land z \preceq y) \rightarrow X(z)).
\]

(ii) to \( \text{Rec}(\Sigma) \) we append the operation \( \cdot \) given by,

\[
L_1 \cdot L_2 := \{ w \cdot v : w \in L_1, v \in L_2 \}.
\]

**Proof.** All that is required, in light of theorem 4.6, is to show that for any \( \phi, \psi \in \text{LT} \),

\[
\text{SPEC}(\phi + \psi) = \text{SPEC}(\phi) \cdot \text{SPEC}(\psi).
\]

If \( w \in \text{SPEC}(\phi + \psi) \) this means that \( M(w) \models \phi + \psi \), i.e. there is \( A \in M(w) \) such that,

\[
M(w) \models \text{dwl}(A) \land \phi \land A \land \psi \land A^c.
\]

Therefore we get that \( M(w) \upharpoonright A \models \phi \) and \( M(w) \upharpoonright A^c \models \psi \). Now it is clear, since \( M(w) \models \text{dwl}(A) \), that \( M(w) \upharpoonright A \cong M(v_1) \) and \( M(w) \upharpoonright A^c \cong M(v_2) \) for some \( v_1, v_2 \in \Sigma^* \) such that \( v_1 \cdot v_2 = w \). As \( M(w) \models \phi \land A \) is equivalent to \( M(v_1) \models \phi \) and \( M(w) \models \psi \land A^c \) is equivalent to \( M(v_2) \models \psi \), we therefore have \( w \in \text{SPEC}(\phi) \cdot \text{SPEC}(\psi) \).

For the converse suppose that \( w \in \text{SPEC}(\phi) \cdot \text{SPEC}(\psi) \). Then there exists \( v_1 \in \text{SPEC}(\phi) \) and \( v_2 \in \text{SPEC}(\psi) \) such that \( w = v_1 \cdot v_2 \). Then it is clear that there exists \( A \in M(w) \), namely the initial segment corresponding to \( v_1 \), such that,

\[
M(w) \models \text{dwl}(A) \land \phi \land A \land \psi \land A^c.
\]

Hence \( w \in \text{SPEC}(\phi + \psi) \) as required. \( \square \)

Proposition 4.7 says that \( \text{(Rec}(\Sigma), \cdot) \) and \( \text{(LT}(T_{M(\Sigma^*)}), +) \) differ only superficially. So towards a proof of theorem 4.3 we can consider the extended Stone dual of \( \text{LT}(T_{M(\Sigma^*)}) \).

**Theorem 4.8.** The topological monoid \( (S(T_{M(\Sigma^*)}), \otimes) \) is the extended Stone dual of the Boolean algebra with operator \( \text{(LT}(T_{M(\Sigma^*)}), +) \).

**Proof.** The extended Stone dual of \( \text{(LT}(T_{M(\Sigma^*)}), +) \) is, by definition (see theorem 2.3), the Boolean space with compatible operation \( (S(T_{M(\Sigma^*)}), R_+) \), where \( R_+ \) is the ternary relation on \( S(T_{M(\Sigma^*)}) \) given by,

\[
R_+(x, y, z) \iff x + y \subseteq z.
\]

It is therefore enough to show that for each \( x, y \in S(T_{M(\Sigma^*)}) \), \( x \otimes y \) is the unique ultrafilter on \( \text{LT}(T_{M(\Sigma^*)}) \) such that, \( x + y \subseteq x \otimes y \). In other words, we will show that \( R_+ \) is the graph of \( \otimes \).

We will in fact show that \( \langle x + y \rangle = x \otimes y \), where \( \langle x + y \rangle \) is the filter generated by the set \( x + y \). Uniqueness then follows from the fact that there can be no proper inclusions of ultrafilters of a Boolean algebra.

Firstly we show \( x + y \subseteq x \otimes y \). As \( x \otimes y \) is a filter this implies that \( x + y \subseteq x \otimes y \). Let \( \phi \in x \) and \( \psi \in y \) be given. Recall that \( \phi + \psi \) is defined as the sentence,

\[
\exists X (\text{dwl}(X) \land \phi \land X \land \psi \land X^c).
\]
Let $\mathcal{M} \models x$ and $\mathcal{N} \models y$, so that $\mathcal{M} \otimes \mathcal{N} \models x \otimes y$. Then consider the element $A = (\top, \top) \in \mathcal{M} \otimes \mathcal{N}$. It is immediate that $A^c = (\bot, \top)$, $\mathcal{M} \otimes \mathcal{N} \models A \cong \mathcal{M}$, and $\mathcal{M} \otimes \mathcal{N} \models A^c \cong \mathcal{N}$. Therefore,

$$\mathcal{M} \otimes \mathcal{N} \models \text{dwcl}(A) \land \phi \land \psi \land A^c.$$

So $\mathcal{M} \otimes \mathcal{N} \models \phi + \psi$, and hence $\phi + \psi \in x \otimes y$ as required.

To conclude we must show that $x \otimes y \subseteq (x + y)$. We show that each sentence $\eta \in x \otimes y$ is equivalent over $T_{\mathcal{M}(\Sigma^c)}$ to a finite disjunction of sentences contained in $x + y$. Take $\eta \in x \otimes y$. Let $\mathcal{M} \models x$ and $\mathcal{N} \models y$, so that $\mathcal{M} \otimes \mathcal{N} \models \eta$. We want to show that there exist $m \in \mathbb{N}$, $\phi_1, \ldots, \phi_m \in x$, and $\psi_1, \ldots, \psi_m \in y$ such that for $i = 1, \ldots, m$,

$$T_{\mathcal{M}(\Sigma^c)} \models \eta \leftrightarrow \bigvee_{i=1}^{m} (\phi_i + \psi_i).$$

Without loss of generality we can assume that $\eta$ is an unnested sentence of quantifier rank $k \in \mathbb{N}$. The equivalence relation $\approx_k$ on the class of $L^\omega_{\infty}$-structures admits finitely many equivalence classes, each of which is axiomatised by an $L^\omega_{\infty}$-sentence. Taking the finite disjunction of all sentences of the form $\phi + \psi$, where $\phi \in x$ and $\psi \in y$ each axiomatises an equivalence class of $\approx_k$, which $\otimes_k$ (see definition \ref{def:otimes_k}) sends to an equivalence class of $\approx_k$ containing a model of $\eta$. This disjunction is equivalent to $\eta$ over $T_{\mathcal{M}(\Sigma^c)}$ so we are done. \hfill \Box

\begin{remark}
By taking $w \approx_k v$ if and only if $M(w) \approx_k M(v)$, for each $w, v \in \Sigma^*$, we may think of $\approx_k$ as an element of $\text{Con}_w(\Sigma^*)$. In theorem \ref{thm:4.10} and the remainder of the paper we will switch freely between these two views of $\approx_k$.
\end{remark}

\begin{theorem}
The congruences $\approx_k$ for $k \in \mathbb{N}$ form a cofinal subset of the poset $(\text{Con}_w(\Sigma^*), \supseteq)$.
\end{theorem}

\begin{proof}
Let $\rho \in \text{Con}_w(\Sigma^*)$. We must show that there is $k \in \mathbb{N}$ such that $w \approx_k w'$ implies $w \sim_\rho w'$ for each $w, w' \in \Sigma^*$. Consider the quotient $\Sigma^*/\rho$. For each $[w]_\rho \in \Sigma^*/\rho$, we have that the preimage $\pi_\rho^{-1}([w]_\rho) \subseteq \Sigma^*$ is a recognisable language. We can enumerate these equivalence classes as $L_1, \ldots, L_n \subseteq \Sigma^*$. Note that these equivalence classes form a partition of $\Sigma^*$. By theorem \ref{thm:4.10}, there are $L^\omega_{\infty}$-formulas $\phi_1, \ldots, \phi_n$ which axiomatise each of these regular languages, meaning for each $i$ such that $1 \leq i \leq n$ we have $L_i = \{ w \in \Sigma^* : M(w) \models \phi_i \}$. Without loss of generality we may assume that each of the sentences $\phi_1, \ldots, \phi_n$ are unnested. Now let $k = \max_{1 \leq i \leq n} qd(\phi_i)$. We claim that $w \approx_k w'$ implies $w \sim_\rho w'$. But this is immediate, for if $w \approx_k w'$ then we have $w \models \phi_i$ if and only if $w' \models \phi_i$ for each $i$ such that $1 \leq i \leq n$, and hence there is $i$ such that $1 \leq i \leq n$ for which $w, w' \in L_i$. \hfill \Box
\end{proof}

\begin{theorem}
The topological monoid $(S(T_{\mathcal{M}(\Sigma^c)}), \otimes)$ is the inverse limit of the directed system of topological monoids given by restricting $F_{\Sigma^c} : \text{Con}_w(\Sigma^*) \to \text{TopMon}$ (see definition \ref{def:omega_infinity}) to the cofinal subset $C$ of $(\text{Con}_w(\Sigma^*), \supseteq)$ comprising the congruences $\approx_k$ for $k \in \mathbb{N}$.
\end{theorem}

\begin{proof}
For each $k$ we have a morphism (a continuous map which respects the binary operation) $p_k : S(T_{\mathcal{M}(\Sigma^c)}) \to S_k$ given by taking a completion of $T_{\mathcal{M}(\Sigma^c)}$ to the $\approx_k$ class of any of its models. These morphisms commute with the projection maps which make up the directed system (see definition \ref{def:omega_infinity}).

Now suppose $(X, \times)$ is some other topological monoid, and that for each $k \in \mathbb{N}$ we have a morphism $q_k : X \to S_k$. Moreover suppose these morphisms $q_k$ commute
with the projection maps. We must show that there exists a unique morphism $g : X \to S(T_{M(\Sigma^*)})$ such that $p_k \circ g = q_k$ for each $k \in \mathbb{N}$. Taking $x \in X$, we can form $\tilde{q}(x) = (q_k(x))_{k \in \mathbb{N}}$. Similarly for $x \in S(T_{M(\Sigma^*)})$ we can form $\tilde{p}(x) = (p_k(x))_{k \in \mathbb{N}}$.

By a straightforward application of the compactness theorem, for each $x$ in $X$ there exists $g(x) \in S(T_{M(\Sigma^*)})$ such that $p_k \circ g(x) = q_k(x)$ for each $k \in \mathbb{N}$. That $g(x)$ is unique is clear.

So far we have that $g$ is a well defined function, but we must check that it is also continuous and respects the binary operations. Let $x, x' \in X$ and consider $g(x \times x')$, we want to show that $g(x \times x') = g(x) \otimes g(x')$. For each $k \in \mathbb{N}$ we have,

$$p_k(g(x \times x')) = q_k(x \times x')$$

$$= q_k(x) \otimes_k q_k(x')$$

$$= p_k(g(x)) \otimes_k p_k(g(x'))$$

$$= p_k(g(x) \otimes g(x')).$$

But if two elements of $S(T_{M(\Sigma^*)})$ agree with respect to all of the morphisms $p_k$, then they must in fact be equal, hence $g(x \times x') = g(x) \otimes g(x')$. The topology on $S(T_{M(\Sigma^*)})$ has a basis of clopens comprising sets of the form $\langle \phi \rangle := \{ x : x \models \phi \}$ where $\phi$ is a sentence. Consider $g^{-1}(\langle \phi \rangle)$. Taking $k = \text{qd}(\phi)$, there exists a subset $F \subseteq S_k$ such that $\langle \phi \rangle = p_k^{-1}(F)$. Then it follows at once from the fact that $p_k \circ g = q_k$ that $g^{-1}(\langle \phi \rangle) = q_k^{-1}(F)$, which is clopen as it is the preimage of a continuous map into a finite discrete space. Hence $g$ is continuous.

Having verified directly that $(S(T_{M(\Sigma^*)}), \otimes)$ satisfies the universal property for inverse limits the proof is complete. \qed

**Theorem 4.12.** The topological monoids $(S(T_{M(\Sigma^*)}), \otimes)$ and $\widehat{\Sigma}^*$ are isomorphic.

**Proof.** By definition 2.12 $\widehat{\Sigma}^*$ is $\varprojlim F_{\Sigma^*}$, where $F_{\Sigma^*} : \text{Con}_{\Sigma}(\Sigma^*) \to \text{TopMon}$ takes a finite index congruence $\rho$ to the quotient $\Sigma^*/\rho$. In theorem 4.10 it was shown that the congruences of the form $\approx_k$, for $k \in \mathbb{N}$, form a cofinal subset $C$ of the poset $\text{Con}_{\Sigma}(\Sigma^*, \supseteq)$. Hence $\widehat{\Sigma}^* = \varprojlim F'$, Lemma 2.11 then tells us that $\varprojlim F_{\Sigma^*} = \varprojlim F'$ where $F'$ is the restriction of $F_{\Sigma^*}$ to $C$. Finally, theorem 4.11 says that $\varprojlim F'$ is $(S(T_{M(\Sigma^*)}), \otimes)$. Uniqueness of inverse limits up to isomorphism then allows us to conclude that $(S(T_{M(\Sigma^*)}), \otimes)$ and $\widehat{\Sigma}^*$ are isomorphic as required. \qed

**Corollary 4.13.** $\widehat{\Sigma}^*$ is the extended Stone dual of $(\text{Rec}(\Sigma), \subseteq)$.

**References**

[Geh16] M. Gehrke. “Stone duality, topological algebra, and recognition”. In: *J. Pure Appl. Algebra* 220.7 (2016), pp. 2711–2747.

[GGP10] M. Gehrke, S. Grigorieff, and J.-É. Pin. “A Topological Approach to Recognition”. In: *Automata, Languages and Programming*. Ed. by S. Abramsky et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 151–162.

[Gol89] R. Goldblatt. “Varieties of complex algebras”. In: *Ann. Pure Appl. Logic* 44.3 (1989), pp. 173–242.

[GS19] S. J. van Gool and B. Steinberg. “Pro-aperiodic monoids via saturated models”. In: *Israel J. Math.* 234.1 (2019), pp. 451–498.
REFERENCES

[Han83] G. Hansoul. “A duality for Boolean algebras with operators”. In: Algebra Universalis 17.1 (1983), pp. 34–49.

[Hod93] W. Hodges. Model theory. Vol. 42. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1993, pp. xiv+772.

[Lin21] D. Linkhorn. “Monadic Second Order Logic and Linear Orders”. (PhD thesis, available from the author’s website or by email request.) University of Manchester, 2021.

[Mac98] S. Mac Lane. Categories for the working mathematician. Second edition. Vol. 5. Graduate Texts in Mathematics. Springer-Verlag, New York, 1998, pp. xii+314.

[Pin09] J.-É. Pin. Profinite methods in automata theory. https://www.irif.fr/~jep/PDF/Exposes/STACS2009Transp.pdf Accessed: 3rd of March 2022. 2009.

THE UNIVERSITY OF MANCHESTER, SCHOOL OF MATHEMATICS, OXFORD ROAD, MANCHESTER, M13 9PL, UK
Email address: deacon.linkhorn@manchester.ac.uk
URL: https://deaconlinkhorn.com