An Algebraic Characterisation of First-Order Logic with Neighbour

Amaldev Manuel
Indian Institute of Technology Goa
amal@iitgoa.ac.in

Dhruv Nevatia
Chennai Mathematical Institute
dhruv@cmi.ac.in

Abstract—We give an algebraic characterisation of first-order logic with the neighbour relation, on finite words. For this, we consider languages of finite words over alphabets with an involution on them. The natural algebras for such languages are involution semigroups. To characterise the logic, we define a special kind of semidirect product of involution semigroups, called the locally hermitian product. The characterisation theorem for FO with neighbour states that a language is definable in the logic if and only if it is recognised by a locally hermitian product of an aperiodic commutative involution semigroup, and a locally trivial involution semigroup. We then define the notion of involution varieties of languages, namely classes of languages closed under Boolean operations, quotients, involution, and inverse images of involutory morphisms. An Eilenberg-type correspondence is established between involution varieties of languages and pseudovarieties of involution semigroups.
.

I. INTRODUCTION

We give an algebraic characterisation of a logic over finite words, namely *the first-order logic with the neighbour relation*. Let A be a finite alphabet. Formulas of the logic FO with neighbour, $\mathrm{FO}(A, min, max, N)$, are interpreted over finite words over the alphabet A. The constants min and max denote the first and last positions of a given word respectively. The atomic formulas of the logic are the following: The predicate $P_a(x)$, for $a \in A$, denotes that the position x is labelled by the letter a. The binary predicate $N(x, y)$ denotes that x and y are neighbours, i.e., either $x + 1 = y$ or $y + 1 = x$. Finally, we have the equality predicate $x = y$. The set of formulas of the logic are closed under Boolean operations and first order quantifications, i.e., $\varphi \lor \psi$, $\varphi \land \psi$, $\neg\varphi$, $\forall x\, \varphi$, and $\exists x\, \varphi$ are also formulas of the logic, if φ and ψ are formulas of the logic. For example, the formula

$$P_a(min) \land P_b(max) \land \forall x \forall y\, (N(x, y) \to (P_a(x) \leftrightarrow P_b(y)))$$

defines the language $a(ba)^\star b$ over the alphabet $\{a, b\}$. The language *defined* by a formula φ, denoted as $L(\varphi)$, is the set of all words satisfying the formula φ.

Before we go into the characterisation problem of $\mathrm{FO}(A, min, max, N)$, it is interesting to note the expressibility of related logics. The neighbour and *between* predicates — the ternary predicate $bet(x, y, z)$ is true if position y is strictly between positions x and z — were first studied in [8], [9]. Although one could express that a position is one of the

endpoints using bet as well as N, the constants min and max themselves are not expressible using bet and N. Since both bet and N are their own left-to-right dual, if the constants min and max are not used then languages definable using these predicates are closed under the reverse operation. For the sake of good algebraic properties, we include the constants min and max in our vocabulary. Next we state the results of [8], [9] in this setting.

The monadic second-order logic $\mathrm{MSO}(A, min, max, N)$, is the extension that also allows monadic second-order quantification — $\exists X\varphi$, $\forall X\varphi$ are also formulas whenever φ is a formula of the logic; The formula $\exists X\varphi$ is true if there is a set of positions X that satisfies the formula φ. For example, the formula

$$\exists X \forall x \forall y\, (N(x, y) \to (X(x) \leftrightarrow \neg X(y))) \land X(min) \land \neg X(max)$$

defines all the words of even length. It turns out that $\mathrm{MSO}(A, min, max, N)$ defines precisely all regular languages, i.e., it has the same expressive power as MSO with the *successor* relation, $\mathrm{MSO}(A, +1)$, by Büchi-Elgot-Trakhtenbrot's theorem. The relationship between the predicates N and bet are analogous to their oriented counter parts, namely, the the successor relation $(x + 1 = y)$ and the order relation $(x < y)$. Both N and bet are definable in terms of the other by monadic second-order formulas; thus $\mathrm{MSO}(A, min, max, bet)$ also defines all regular languages. Likewise, $\mathrm{FO}(A, min, max, bet)$ defines precisely all aperiodic regular languages — languages that are definable in the logic $\mathrm{FO}(A, <)$. In fact, this parallelism between $\mathrm{FO}(A, min, max, bet)$ and $\mathrm{FO}(A, <)$ extends to their quantifier alternation hierarchies.

However, the parallel described so far breaks down in the case of $\mathrm{FO}(A, min, max, N)$. There are languages expressible in $\mathrm{FO}(A, +1)$ that are not expressible in the former logic. For instance, the language

$$L = c^* abc^*$$

over the alphabet $\{a, b, c\}$ is expressible in the logic $\mathrm{FO}(A, +1)$. But, using an Ehrenfeucht-Fraïssé argument [7], [8], it can be shown that L is *not* definable in $\mathrm{FO}(A, min, max, N)$. Thus, we have the question of characterising the languages definable in this logic.

Using Hanf's theorem [7] from finite model theory, it is possible to give language-theoretic characterisations of both

$\mathrm{FO}(A, +1)$ and $\mathrm{FO}(A, min, max, N)$. For $t > 0$, we define the *equality with threshold* t on the set $\mathbb{N}$ of natural numbers by

$$i =^t j \;:=\; \begin{cases} i = j & \text{if } i < t, \\ i \geq t \text{ and } j \geq t & \text{otherwise.} \end{cases} \quad (1)$$

The word $y \in A^+$ is a *factor* of the word $u \in A^+$ if $u = xyz$ for some x, z in A^*. We use $\sharp(u, y)$ to denote the number of times the factor y appears in u, i.e. the number of pairs (x, z), where $x, z \in A^*$, such that $u = xyz$.

▶ **Definition 1.** *Let $\approx_k^t$, for $k, t > 0$, be the equivalence on A^*, whereby two words u and v are equivalent if either they both have length at most $k - 1$ and $u = v$, or otherwise they have*

1) *the same prefix of length $k - 1$,*
2) *the same suffix of length $k - 1$,*
3) *and the same number of occurrences, up to threshold t, for all factors of length $\leq k$, i.e. $\sharp(u, y) =^t \sharp(v, y)$ for each word $y \in A^+$ of length at most k.*

A language is *locally threshold testable* (or LTT for short) if it is a union of $\approx_k^t$ classes, for some $k, t > 0$. Locally threshold testable languages are precisely the class of languages definable in $\mathrm{FO}(A, +1)$ [1], [23].

Since the neighbour predicate N is definable using the successor relation in first-order logic, $\mathrm{FO}(A, min, max, N)$ definable languages are a subset of LTT. But this inclusion is strict, as we have seen.

We define a coarser equivalence $\overset{r}{\approx}_k^t$ by the counting factors only up to reverse. Let $\sharp^r(w, v)$ denote the number of occurrences of v or v^r in w, i.e. the number of pairs (x, y), where $x, y \in A^*$, such that $w = xvy$ or $w = xv^r y$.

▶ **Definition 2.** *Let $k, t > 0$. Two words $w, w' \in A^*$ are $\overset{r}{\approx}_k^t$-equivalent if $|w| < k$ and $w = w'$, or w, w' both are of length at least k, and they have*

1) *the same prefix of length $k - 1$,*
2) *the same suffix of length $k - 1$, and*
3) *$\sharp^r(w, v) =^t \sharp^r(w', v)$ for each word $v \in A^+$ of length at most k.*

It is shown in [8] that a language is definable in $\mathrm{FO}(A, N)$ if it is closed under the reverse operation and is a union of equivalence classes of $\overset{r}{\approx}_k^t$ for some $k, t > 0$. Such languages are called *locally-reversible threshold testable* languages. The class of languages accepted by $\mathrm{FO}(A, min, max, N)$ are characterised in a similar manner. A language L is *weakly locally-reversible threshold testable*, wLRTT for short, if it is a union of equivalence classes of $\overset{r}{\approx}_k^t$ for some $k, t > 0$. By a straight-forward adaptation of the proofs in [8], [9] we can show that

▶ **Theorem 1.** *A language is definable in the logic $\mathrm{FO}(A, min, max, N)$ if and only if it is weakly locally-reversible threshold testable.*

Obtaining an algebraic characterisation for the class wLRTT is an interesting problem. The analogous character-isation of $\mathrm{FO}(A, +1)$ is a deep result [1], [5], [15], [26] in the theory of finite semigroups. The first observation is that LTT is a *variety of languages* — a class of languages that is closed under Boolean operations, quotients with respect to words, and inverse images under homomorphisms. Eilenberg's variety theorem states that varieties of languages correspond to pseudovarieties of semigroups — a pseudovariety of finite semigroups is a set of finite semigroups that is closed under finite direct products, subsemigroups and quotients. The characterisation of LTT proceeds by first writing the class as a semidirect product of pseudovarieties of aperiodic commutative monoids (**Acom**) and righty trivial semigroups (**D** is the pseudovariety of semigroups that satisfy the equation $se = e$ for all elements s and idempotents e). This step is the algebraic analogue of synthesising an automaton for an LTT language as a cascade of a scanner and an acceptor [1]. The next step in the proof uses the framework of categories for obtaining the identities capturing the semidirect product [21], [22], [25].

In the case of wLRTT, this approach does not work; because wLRTT does not form a variety of languages — for instance, the language $(xy)^* ab(xy)^*$ over the alphabet $\{a, b, x, y\}$ is in wLRTT. Its inverse image under the morphism $a \mapsto a, b \mapsto b, c \mapsto xy$ is the language $c^* abc^*$, over the alphabet $\{a, b, c\}$. As we have seen, this language is not in wLRTT. Therefore, wLRTT is not closed under inverse image of morphisms, and hence not a variety of languages. This necessitates a rubric, of the like of varieties and operations on varieties, to study the class wLRTT.

Our results. It was already observed in [8] that one needs to extend semigroups with an involution (also called $\star$-semigroups) — an involution on a semigroup S is an operation $\star$ such that $(a^\star)^\star = a$ and $(ab)^\star = b^\star a^\star$, for each $a, b \in S$ — to characterise wLRTT. The involution operation is a gen-eralisation of the reversal operation on words. An involutory alphabet $\mathscr{A}$ is a finite alphabet A with a bijection † on it. The map † extends to words over A as $(a_1 \ldots a_n)^\dagger = a_n^\dagger \ldots a_1^\dagger$, where each $a_i \in A$. This map is an involution on the semi-group A^+. We lift the notions of recognisability and syntactic semigroups to the case of languages over involutory alphabets. Syntactic algebras for such languages are semigroups with an involution. A finite alphabet A can be seen as an involutory alphabet with the identity function id as †; such alphabets are called *hermitian*. This just means that A^+ is equipped with the reverse operation as involution. Syntactic algebras for languages over hermitian alphabets are semigroups with involutions that are generated by their hermitian elements (a is hermitian if $a^\star = a$).

Assume $(A, \dagger)$ and $(B, \star)$ are involutory alphabets and $h \colon A^+ \to B^+$ is a morphism. The morphism h is involutory if $h(w^\dagger) = h(w)^\star$. An *involution variety of languages* maps each involutory alphabet $\mathscr{A} = (A, \dagger)$ to a class of languages over A that is closed under involutions, Boolean operations, quotients with respect to words and inverse images under involutory morphisms. We establish an Eilenberg-type correspondence

between involution varieties of languages and pseudovarieties of finite involution semigroups — classes of finite involution semigroups that are closed under finite direct products, sub-$\star$-semigroups and quotients.

If $\mathscr{S} = (S, \star)$ and $\mathscr{T} = (T, \dagger)$ are involution semigroups, a two-sided action of T on S is *compatible* with the involution if $(tst')^\star = t'^\dagger s^\star t^\dagger$, where tst' denotes the left action by $t \in T$ and right action by $t' \in T$ on the element $s \in S$. A compatible two-sided action of T on S is *locally hermitian* if for all idempotents $e \in T$ and elements $s \in S$ it holds that $ese^\dagger = es^\star e^\dagger$. Intuitively this means that within an idempotent context that looks the same on the outside from either direction, one can reverse factors. Bilateral semidirect products with locally hermitian actions are called locally hermitian semidirect products. If $\mathbf{V}$ and $\mathbf{W}$ are pseudovarieties of involution semigroups, the locally hermitian product of $\mathbf{V}$ and $\mathbf{W}$ is the pseudovariety generated by all locally hermitian semidirect products of involution semigroups from $\mathbf{V}$ and $\mathbf{W}$. It is shown that a language is wLRTT if and only if it is recognised by a locally hermitian semidirect product of an aperiodic commutative semigroup and a locally trivial semigroup. This result can also be stated in terms of the corresponding pseudovarieties of involution semigroups $\mathbf{Acom}^\star$ and $\mathbb{L}\mathbf{1}^\star$. Therefore, we have an algebraic characterisation of the class $\mathrm{FO}(A, min, max, N)$.

Related work. The FO and MSO logics using the predicates N and bet were introduced in [8], [9]. Formulas of these logics that do not use the constants min and max, are self-dual, i.e., replacing each predicate by its dual, for instance $N(x, y)$ by $N(y, x)$, results in an equivalent formula. Therefore, $\mathrm{MSO}(A, N)$ and $\mathrm{MSO}(A, bet)$ recognise precisely the class of *reversible* regular languages — a language is reversible if it is invariant under taking the reverse of words in it. Similarly, $\mathrm{FO}(A, bet)$ recognises all aperiodic reversible languages. In [8], the question of deciding membership in these classes was studied. For the logics $\mathrm{MSO}(A, bet), \mathrm{FO}(A, bet)$, the existing decidable characterisations of $\mathrm{MSO}(A, <)$ and $\mathrm{FO}(A, <)$ immediately yield an answer. For the case of $\mathrm{FO}(A, N)$, a partial answer was given; it was shown that if a language L is definable in $\mathrm{FO}(A, N)$, then its syntactic semigroup satisfies the identity $ex^\star e^\star = exe^\star$, in addition to those defining the class LTT.

A different but related *between* predicate (namely $a(x, y)$, for $a \in A$, is true if there is an a-labelled position between positions x and y) was introduced in [11]–[13]. They characterise the expressive power of two-variable first-order logic with the order relation ($\mathrm{FO}^2(<)$) enriched with the between predicates $a(x, y)$ for $a \in A$, and show an algebraic characterisation of the resulting family of languages.

Since the class of involution semigroups contains many important classes of semigroups such as inverse semigroups, there is substantial literature on varieties of involution semigroups from a semigroup theoretic perspective (see [6] for a survey). However we are not aware of a language theoretic

study of pseudovarieties of finite involution semigroups or results similar to ours.

Orrganisation of the paper. In Section II, we describe involution semigroups and extend the notions of recognisability by semigroups to the case of languages with an involution. In Section III, involutory semidirect products are defined, and we introduce the notion of locally hermitian semidirect product. Then, it shown that a language is wLRTT if and only if it is recognised by the locally hermitian semidirect product of an aperiodic commutative semigroup and a locally trivial semigroup. We introduce the notion of involution varieties of languages and pseudovarieties of involution semigroups, and prove an Eilenberg correspondence between them in Section IV. In Section V, we conclude with some avenues for further inquiry.

II. Languages over Involutory Alphabets

A. Recognisable languages

Let A be a finite alphabet. Then, A^+ denotes the set of nonempty finite words over A, and A^* denotes the set of all finite words over A including the empty word ϵ. Extending this notation, $A^{\leq k}$ denotes the subset of A^* consisting of words of length at most k, including ϵ.

A semigroup S is a set together with an associative binary operation. Unless specified, the semigroup operation is written as $x \cdot y$, for semigroup elements x, y; but we omit notating it whenever possible. If the operation has an identity, then S is called a monoid. All semigroups and monoids we consider are finite, except when stated otherwise. For a finite alphabet A, the set A^+ forms a semigroup under concatenation, while the set A^* is a monoid with the empty word ϵ as the identity. The set $S' \subseteq S$ forms a *subsemigroup* of S if S' is closed under the semigroup operation.

Let $(T, +)$ be a semigroup. A *semigroup morphism* from S to T is a map $h \colon S \to T$ that satisfies the equation $h(xy) = h(x) + h(y)$, for all $x, y \in S$. If S and T are monoids, then h is a *monoid morphism* if additionally h maps the identity of S to the identity of T. A semigroup T *divides* the semigroup S if T is the image of a subsemigroup of S under some morphism.

Let A be a finite alphabet. A language $L \subseteq A^+$ is *recognized* by a semigroup S if there is a morphism $h \colon A^+ \to S$ and a subset $P \subseteq S$ such that $L = h^{-1}(P)$. The set P is called the accepting set of L. The class of languages recognised by finite semigroups coincide with the class of regular languages [10], p. 160.

The above definitions are easily adapted to the case of languages over A^*. The definitions are assumed in the subsequent sections.

B. Languages over Involutory Alphabets

A finite *involutory alphabet* $\mathscr{A} = (A, \dagger)$ is a finite alphabet A with a bijection $\dagger \colon A \to A$ such that $(a^\dagger)^\dagger = a$ for each letter $a \in A$. The function $\dagger$ is extended to words over A as

$$w^\dagger = a_n^\dagger \ldots a_1^\dagger \, , \tag{2}$$

if $w = a_1 \cdots a_n \in A^+$. Clearly, $w^{\dagger\dagger} = w$. Moreover, $(uv)^\dagger = v^\dagger u^\dagger$, for all words $u, v \in A^+$. Hence, the operation $\dagger \colon A^+ \to A^+$ is an *involution*: a function that is its own inverse. If $L \subseteq A^+$ is a language, then $L^\dagger$ denotes the set $\{w^\dagger \mid w \in L\}$.

If $\dagger$ is the identity function on A, then $\mathscr{A}$ is called a *hermitian* alphabet. In that case, $\dagger$ is the reverse operation $u \mapsto u^r$, for $u \in A^+$, and $L^\dagger$ is the set L^r, for $L \subseteq A^+$. Unless specified, a finite alphabet is taken to be a hermitian alphabet, i.e. with the reverse operation as the involution.

▶ **Definition 3.** *An involution semigroup (also called a $\star$-semigroup) $\mathscr{S} = (S, \star)$ is a semigroup S extended with an operation $\star \colon S \to S$ such that for all elements a, b of S,*

$$(a^\star)^\star = a \ , \ \ and \ \ (a \cdot b)^\star = b^\star \cdot a^\star \ .$$

We say that S is the semigroup reduct of $\mathscr{S}$. The semigroup A^+ with the operation $\dagger \colon A^+ \to A^+$, defined in (2), forms an involution semigroup, denoted as $\mathscr{A}^+$. Similarly, A^* with $\dagger$ forms a monoid with an involution, denoted as $\mathscr{A}^*$. The involution semigroup $\mathscr{A}^+$ is *not* a *free* involution semigroup in general; it is only so when the relation $\dagger$ is irreflexive on A ([14], p. 172).

▶ **Example 1.** Let $T = \{a, b, ab, ba\}$ be the semigroup with the multiplication

$$aa = a \ , \ \ \ bb = b \ , \ \ \ aba = bab = ba.$$

Let $\star$ be the map $a^\star = b$. The map extends to an involution on T,

$$b^\star = (a^\star)^\star = a \ , \ \ \ (ab)^\star = b^\star a^\star = ab \ , \ \ \ (ba)^\star = a^\star b^\star = ba \ .$$

However, the map $a^\dagger = a, b^\dagger = b$ is not a well-defined involution on T;

$$(ba)^\dagger = a^\dagger b^\dagger = ab \neq ba = aba = a^\dagger b^\dagger a^\dagger = (aba)^\dagger = (ba)^\dagger \ .$$

It is not possible to associate an involution with every semigroup; for instance, the semigroup of right zeros, where the multiplication follows the identity $xy = y$ for any elements x, y, does not admit an involution if it has more than one element. However, it is possible to obtain a $\star$-semigroup, from a given semigroup, that shares similar properties, as we shall see later.

Let $\mathscr{S} = (S, \star)$ and $\mathscr{T} = (T, \dagger)$ be involution semigroups. A morphism of involution semigroups, also called an involutory morphism, from $\mathscr{S}$ to $\mathscr{T}$, denoted as $h \colon \mathscr{S} \to \mathscr{T}$, is a semigroup morphism $h \colon S \to T$ such that $h(a^\star) = h(a)^\dagger$, for all $a \in S$.

▶ **Definition 4.** *Let $\mathscr{A} = (A, \dagger)$ be an involutory alphabet. The involution semigroup $\mathscr{S} = (S, \star)$ recognises the language $L \subseteq A^+$ if there is a morphism of involution semigroups $h \colon \mathscr{A}^+ \to \mathscr{S}$ and a subset $P \subseteq S$ such that $L = h^{-1}(P)$.*

▶ **Example 2.** Let $\mathscr{A}$ be the involutory alphabet $\{a, b\}$ with the map $a^\star = b$. Let $T = \{a, b, ab, ba\}$ be the semigroup from Example 1. Then, T accepts the language $L = a^+ b^+$ with the morphism $h(a) = a, h(b) = b$ and the accepting set $\{ab\}$.

Clearly, if a language is recognised by an involution semigroup, then it is recognised by a semigroup as well. The other direction also holds, as we show next.

Let $\mathscr{S} = (S, \dagger)$ be an involution semigroup and let T be a semigroup. By T^{op}, we denote the opposite of T — the semigroup with the same set of elements but reversed multiplication, i.e., $a \cdot^{op} b = b \cdot a$, for all $a, b \in T$. Assume there is a semigroup morphism from S to T. Let $h^\dagger$ be the map from S to T^{op}, given by, for each $s \in S$,

$$h^\dagger(s) = h(s^\dagger) \ .$$

▶ **Lemma 1.** *If $h \colon S \to T$ is a morphism, then $h^\dagger \colon S \to T^{op}$ is also a morphism.*

Proof. For $x, y \in S$,

$$\begin{aligned} h^\dagger(xy) &= h((xy)^\dagger) \\ &= h(y^\dagger x^\dagger) \\ &= h(y^\dagger)h(x^\dagger) && (\text{product in } T) \\ &= h^\dagger(x)h^\dagger(y) && (\text{product in } T^{op}) \ . \end{aligned}$$

Hence $h^\dagger \colon S \to T^{op}$ is a semigroup morphism. ◀

Let $\mathscr{A} = (A, \dagger)$ be an involutory alphabet. Assume that $L \subseteq A^+$ is recognised by the semigroup S with the morphism h and the accepting set P. Lemma 1 gives that $h^\dagger \colon A^+ \to S^{op}$ is also a morphism; by definition, $h^\dagger(w^\dagger) = h(w)$, for all $w \in A^+$. Hence, $h^\dagger(L^\dagger) = h(L) = P$; hence, the semigroup S^{op} recognises $L^\dagger$ with the morphism $h^\dagger$.

Next, we show that $(T \times T^{op}, flip \colon (x, y) \mapsto (y, x))$ is an involution semigroup. It suffices to show

▶ **Lemma 2.** *flip is an involution on $T \times T^{op}$.*

Proof. Clearly, for $(a, b) \in T \times T^{op}$

$$\left((a, b)^{flip}\right)^{flip} = (a, b) \ .$$

Let $(a, b), (c, d) \in T \times T^{op}$. Then,

$$\begin{aligned} ((a, b)(c, d))^{flip} &= (ac, db)^{flip} \\ &= (db, ac) \\ &= (d, c)(b, a) \\ &= (c, d)^{flip}(a, b)^{flip} \ . \end{aligned}$$

Hence $flip$ is an involution on $T \times T^{op}$. ◀

▶ **Proposition 1.** *If a language is recognised by a semigroup, then it is recognised by a $\star$-semigroup as well.*

Proof. Let $\mathscr{A} = (A, \dagger)$ be an involutory alphabet. Assume that L is recognised by S using the morphism $h \colon A^+ \to S$. By Lemma 1, $h^\dagger \colon A^+ \to S^{op}$ is a morphism. By Lemma 2, $(S \times S^{op}, flip)$ is an involution semigroup. Then, the product map $g \colon w \mapsto (h(w), h^\dagger(w))$, for $w \in A^+$, is a morphism from A^+ to $S \times S^{op}$. Since

$$g(w^\dagger) = (h(w^\dagger), h^\dagger(w^\dagger)) = (h^\dagger(w), h(w)) = (g(w))^{flip} \ ,$$

g is a morphism of involution semigroups from $\mathscr{A}^+$ to $S \times S^{op}$. Finally, the proof is completed by observing that

L is recognised by the involution semigroup $S \times S^{op}$ with the accepting set $\{(h(w), h^\dagger(w)) \mid w \in L\}$. ◄

Therefore, languages recognised by involution semigroups are precisely the class of recognisable languages.

Let $\mathscr{S} = (S, \star)$ be an involution semigroup. If $S' \subseteq S$, then $S'^\star$ denotes the set $\{s^\star \mid s \in S'\}$. A subset S' of S forms a *sub-$\star$-semigroup* of $\mathscr{S}$ if S' is closed under the semigroup operation and the involution, i.e., if $S'^2 \subseteq S'$ and $S'^\star = S'$. If $h \colon \mathscr{S} \to \mathscr{T}$ is an involution semigroup morphism, then the image of h is a sub-$\star$-semigroup of $\mathscr{T}$.

An element $s \in S$ is *hermitian* if it is its own involution, i.e., if $s^\star = s$. A $\star$-semigroup is hermitian-generated if it is generated by its hermitian elements using the semigroup operation and the involution. For convenience, we shorten hermitian-generated involution semigroup to *hermitian semigroup*. For every hermitian alphabet $\mathscr{A}$, the involution semigroup $\mathscr{A}^+$ is hermitian. Images of hermitian semigroups under involution semigroup morphisms are hermitian as well.

► **Example 3.** In the involution semigroup $\mathscr{T} = (T, \star)$ in Example 1, the set of hermitian elements is $\{ab, ba\}$. Since they don't generate all the elements of the semigroup, $\mathscr{T}$ is not a hermitian semigroup. But $(\{ab, ba\}, \star)$ is a sub-$\star$-semigroup of $\mathscr{T}$ that is hermitian.

For recognising languages that are subsets of A^*, the notion of recognisability by $\star$-semigroups is insufficient; one needs recognisability by $\star$-monoids. The definitions are straightforward adaptations of the ones presented above.

III. Involutory Semidirect Products and The Characterisation of FO

In this section, we first define the involutory semidirect product of two involution semigroups. Then, a particular case, called locally hermitian semidirect product, is defined. It is then shown that wLRTT languages are precisely the ones recognised by locally hermitian products of aperiodic commutative $\star$-semigroups and locally trivial $\star$-semigroups.

A. Bilateral Semidirect Products of Involution Semigroups

Let S and T be finite semigroups. We denote the semigroup operation of S additively (by $+$) and of T multiplicatively. However, we don't assume that S or T is commutative. A *left action l of T on S* is a map $l \colon T \times S \to S$ satisfying the following conditions. Denote the image of (t, s) under l as ts, for $t \in T, s \in S$. Then, for all $s \in S$ and $t, t' \in T$,

$$t(s + s') = ts + ts' \text{ , and,} \tag{3}$$
$$(tt')s = t(t's) \text{ .} \tag{4}$$

Right actions of T on S are defined analogously; a *right action r of T on S* is a map $r \colon S \times T \to S$ such that for all $s \in S$ and $t, t' \in T$,

$$(s + s')t = st + s't \text{ , and,} \tag{5}$$
$$s(tt') = (st)t' \text{ ,} \tag{6}$$

where st denotes the image of (s, t) under the map r.

The actions l and r are *compatible* if for all $t, t' \in T$, and $s \in S$, it is the case that

$$(ts)t' = t(st') \text{ .} \tag{7}$$

A pair of compatible actions (l, r) is called a *bilateral action of T on S*. Given a bilateral action (l, r) of T on S, the *bilateral semidirect product* (also called the *two-sided semidirect product*) $S * {*} T$ is the semigroup with the elements $\{(s, t) \mid s \in S, t \in T\}$ and with the operation

$$(s_1, t_1)(s_2, t_2) = (s_1 t_2 + t_1 s_2, t_1 t_2) \text{ .} \tag{8}$$

Let $\mathscr{S} = (S, \star)$ and $\mathscr{T} = (T, \diamond)$ be involution semigroups. Let (l, r) be a bilateral action of T on S. Then, the pair (l, r) is an *involutory action* of $\mathscr{T}$ on $\mathscr{S}$ if for all $t, t' \in T, s \in S$,

$$(st)^\star = t^\diamond s^\star \text{ .} \tag{9}$$

► **Lemma 3.** *Assume $\mathscr{T} = (T, \diamond)$ has an involutory action on $\mathscr{S} = (S, \star)$. Then, for all $t, t' \in T, s \in S$,*

$$(t_1 s t_2)^\star = t_2^\diamond s^\star t_1^\diamond \text{ .}$$

Proof. By taking $\star$ on both sides in (9), we get

$$st = (t^\diamond s^\star)^\star \text{ .} \tag{10}$$

Exchanging t and $t^\diamond$, and s and $s^\star$ in (10), we obtain

$$s^\star t^\diamond = (ts)^\star \text{ .} \tag{11}$$

Finally,

$$
\begin{aligned}
(t_1 s t_2)^\star &= ((t_1 s) t_2)^\star && \text{by (7)} \\
&= t_2^\diamond (t_1 s)^\star && \text{by (9)} \\
&= t_2^\diamond s^\star t_1^\diamond \text{ .} && \text{by (11)}
\end{aligned}
$$

◄

► **Definition 5.** *Assume $\mathscr{T} = (T, \diamond)$ has an involutory action on $\mathscr{S} = (S, \star)$. Then, their* involutory semidirect product *$\mathscr{S} * {*} \mathscr{T}$ is the semigroup $S * {*} T$ with the involution*

$$\dagger \colon (s, t) \mapsto (s^\star, t^\diamond) \text{ .} \tag{12}$$

► **Lemma 4.** *$\mathscr{S} * {*} \mathscr{T}$ is an involution semigroup.*

Proof. It suffices to show that $\dagger$ is an involution. By definition,

$$(s, t)^{\dagger\dagger} = (s^{\star\star}, t^{\diamond\diamond}) = (s, t) \text{ .}$$

Next, for all $s_1, s_2 \in S, t_1, t_2 \in T$,

$$
\begin{aligned}
((s_1, t_1)(s_2, t_2))^\dagger &= (s_1 t_2 + t_1 s_2, t_1 t_2)^\dagger && \text{by (8)} \\
&= ((s_1 t_2 + t_1 s_2)^\star, (t_1 t_2)^\diamond) && \text{by (12)} \\
&= ((t_1 s_2)^\star + (s_1 t_2)^\star, t_2^\diamond t_1^\diamond) \\
&= (s_2^\star t_1^\diamond + t_2^\diamond s_1^\star, t_2^\diamond t_1^\diamond) \\
&= (s_2^\star, t_2^\diamond)(s_1^\star, t_1^\diamond) \\
&= (s_2, t_2)^\dagger (s_1, t_1)^\dagger \text{ .}
\end{aligned}
$$

Hence $\dagger$ is an involution. ◄

Next, we introduce a particular case of involutory semidirect product.

An element e in a semigroup is an *idempotent* if $e \cdot e = e$. In the case of involution semigroups, if e is an idempotent, then involution of e is also an idempotent.

▶ **Definition 6.** *Let $\mathscr{S} = (S, \star)$ and $\mathscr{T} = (T, \diamond)$ be involution semigroups. Let (l, r) be an involutory action of $\mathscr{T}$ on $\mathscr{S}$. The action is* locally hermitian *if for each idempotent $e \in T$, and each element $s \in S$,*

$$ese^{\diamond} = es^{\star}e^{\diamond} \ . \tag{13}$$

In other words, in the case of a locally hermitian action, the elements of the form $ese^{\diamond}$ are hermitian (observe that, by Lemma 3, $(ese^{\diamond})^{\star} = e^{\diamond\diamond}s^{\star}e^{\diamond} = es^{\star}e^{\diamond}$).

Let $\mathscr{S}$ and $\mathscr{T}$ be involution semigroups. If the action of $\mathscr{T}$ is locally hermitian action on $\mathscr{S}$, then the semidirect product $\mathscr{S} \ast\ast\, \mathscr{T}$ is called *locally hermitian*.

B. Characterisation of FO with Neighbour

A semigroup S is *aperiodic* if there is an $n \in \mathbb{N}$ such that $a^{n+1} = a^{n}$, for all elements a in the semigroup. A semigroup is *commutative* if the semigroup operation is commutative, i.e., $x \cdot y = y \cdot x$, for all elements x and y. A semigroup S is locally trivial if $ese = e$, for each element s and each idempotent e in S. The class of languages LTT has several characterisations in terms of various products of semigroups (see [1] for a detailed discussion); the one relevant to us is,

▶ **Proposition 2.** *A language is in* LTT *if and only if it is recognised by a bilateral semidirect product of an aperiodic commutative semigroup and a locally trivial semigroup.*

Using the construction outlined in the proof of Proposition 1, it can be shown that a language is in LTT if and only if it is recognised by an involutory semidirect product of an aperiodic commutative involution semigroup and a locally trivial involution semigroup.

Next, we characterise the class wLRTT. First, we give an example.

▶ **Example 4.** Consider the language $L = (abc)^{+}$ over the alphabet $A = \{a, b, c\}$. We show that L is defined by a locally hermitian product of two $\star$-semigroups $\mathscr{S}$ and $\mathscr{T}$.

Let T be the semigroup with the set $A^{\leq 2} \setminus \{\epsilon\}$ and the operation $\odot$ given by, for $x, y \in T$

$$x \odot y = \begin{cases} x \cdot y & \text{if } |xy| \leq 2 \\ first(xy) \cdot last(xy) & \text{otherwise,} \end{cases}$$

where $first(w)$ and $last(w)$ denote the first and last letter of a word w. Reverse operation is an involution on the semigroup T. Let $\mathscr{T} = (T, r)$. Idempotents of T are precisely those words of length 2.

Let $\hat{A} = \{\hat{a}, \hat{b}, \hat{c}\}$. Let U be the set $A^{\leq 1} \cdot \hat{A} \cdot A^{\leq 1}$. Elements of T act on words in U in the following way: for $w \in T$ and $x\hat{y}z \in U$, where $x, z \in A^{\leq 1}$ and $\hat{y} \in \hat{A}$, we define the right action $\otimes$ as

$$w \otimes x\hat{y}z = \begin{cases} last(w) \cdot \hat{y}z & \text{if } x \text{ is } \epsilon \\ x\hat{y}z & \text{otherwise,} \end{cases}$$

The right action is analogous.

$$x\hat{y}z \otimes w = \begin{cases} x\hat{y} \cdot first(w) & \text{if } z \text{ is } \epsilon \\ x\hat{y}z & \text{otherwise,} \end{cases}$$

Let S' be the semigroup formed by the powerset of U with set union as the semigroup operation. For each element $I \subseteq U$ of S', let $I^{r} = \{w^{r} \mid w \in I\}$. The operation $I \mapsto I^{r}$, for $I \in S'$, is an involution on the semigroup S'. Let $\mathscr{S}' = (S', r)$. The action of T on U, extends to the elements of $\mathscr{S}'$ pointwise. The elements $A \cdot \hat{A} \cdot A$ are called the *zeros* for the action. Let $\sim$ be a congruence on $\mathscr{S}'$, given by, for $I, J \in S'$,

- $x \in I$ if and only if $x \in J$, for every nonzero $x \in U$, and
- x or x^{r} is in I if and only if x or x^{r} is in J, for each zero $x \in U$.

Let $\mathscr{S}$ be the quotient of $\mathscr{S}'$ with respect to $\sim$. It is easy to see that for each idempotent $e \in T$, and each $I \in \mathscr{S}$, $e \otimes I \otimes e^{r} = e \otimes I^{r} \otimes e^{r}$. Hence the actions are locally hermitian. Finally L is accepted by $\mathscr{S} \ast\ast\, \mathscr{T}$ with the morphism $h \colon a \mapsto (\{\hat{a}\}, a)$.

▶ **Theorem 2.** *Each* wLRTT *language is recognised by a locally hermitian semidirect product of an aperiodic commutative $\star$-semigroup and a locally trivial $\star$-semigroup. Conversely, if L is recognised by a locally hermitian semidirect product $\mathscr{S} \ast\ast\, \mathscr{T}$, where $\mathscr{S}$ is aperiodic and commutative and $\mathscr{T}$ is locally trivial, then L is* wLRTT.

Proof. Assume $L \subseteq A^{+}$, for an involutory alphabet $\mathscr{A} = (A, id)$, is a wLRTT language. To prove the first claim, we define $\star$-semigroups $\mathscr{S}$ and $\mathscr{T}$ such that $\mathscr{S}$ is aperiodic and commutative and $\mathscr{T}$ is locally trivial, and show that their locally hermitian semidirect product recognises L. We let $k, m > 0$ be such that L is a union of $\overset{r}{\approx}{}^{m}_{2k+1}$-classes. Since the equivalence $\overset{r}{\approx}{}^{t}_{d+1}$ refines the equivalence $\overset{r}{\approx}{}^{t}_{d}$, for $t, d > 0$, there is no loss of generality.

Let $\odot$ be the operation on the set $A^{\leq 2k} \setminus \{\epsilon\}$ defined as

$$t \odot t' = \begin{cases} t \cdot t' & \text{if } |tt'| < 2k, \\ pref_{k}(t \cdot t') \cdot suff_{k}(t \cdot t') & \text{otherwise,} \end{cases}$$

where $pref_{k}()$ and $suff_{k}()$ denote the k-length prefix and suffix respectively. It is not difficult to see that the operation $\odot$ is associative. Hence, $T = (A^{\leq 2k} \setminus \{\epsilon\}, \odot)$ is a semigroup. It is straight-forward to check that $(u \odot v)^{r} = v^{r} \odot u^{r}$ and $(u^{r})^{r} = u$, for the reverse operation r. Let $\mathscr{T} = (T, r)$ be the involution semigroup with the reverse operation as the involution. The idempotents of T are precisely the words of length $2k$. Also, $e \odot t \odot e = e$, for words $t \in T$ and $e \in A^{2k}$. Hence, $\mathscr{T}$ is locally trivial.

Next, we set up some notation in preparation for defining $\mathscr{S}$. Let $\hat{A}$ denote the set $\{\hat{a} \mid a \in A\}$ of *anchored letters*. A *k-anchored word* $\hat{w}$ is a word from the set $A^{\leq k} \cdot \hat{A} \cdot A^{\leq k}$. Let W_{k} denote the set of all k-anchored words. We denote anchored words by $\hat{w}, \hat{u}$ etc.

Let $\hat{w} = u\hat{a}v$, where $u, v \in A^*$ and $\hat{a} \in \hat{A}$. We write $l(\hat{w})$ and $r(\hat{w})$ to denote how far is the anchored letter from the left-end and right-end respectively; Hence, $l(\hat{w}) = |u|$ and $r(\hat{w}) = |v|$. Clearly, $0 \le l(\hat{w}), r(\hat{w}) \le k$, for all $\hat{w} \in W_k$.

The semigroup T acts on the set W_k, on the left as well as on the right, in the following manner. For each letter $a \in A$ and $\hat{w} \in W_k$, the left action $a \otimes \hat{w} \in W_k$ of a on $\hat{w}$ is given by,

$$a \otimes \hat{w} = \begin{cases} \hat{w} & \text{if } l(\hat{w}) = k, \\ a \cdot \hat{w} & \text{otherwise.} \end{cases} \tag{14}$$

Similarly the right action $\hat{w} \otimes a$ appends a to the right if there are less than k positions to the right of the anchored position, otherwise it leaves $\hat{w}$ intact. We extend these actions to all the elements of T by letting, for $u \in T$, $a \in A$ and $\hat{w} \in W_k$,

$$(u \odot a) \otimes \hat{w} = u \otimes (a \otimes \hat{w}) , \text{ and } \hat{w} \otimes (a \odot u) = (\hat{w} \otimes a) \otimes u . \tag{15}$$

Clearly the left and right actions are compatible with each other, i.e.,

$$(u \otimes \hat{w}) \otimes v = u \otimes (\hat{w} \otimes v) \tag{16}$$

for all $\hat{w} \in W_k$ and $u, v \in T$. Also the actions are involutory, with respect to the involution on T, since

$$(u \otimes \hat{w})^r = \hat{w}^r \otimes u^r . \tag{17}$$

An anchored word is a *zero* for the action $\otimes$ if for all $a \in A$, $a \otimes \hat{w} = \hat{w} \otimes a = \hat{w}$. The zero elements of W_k are precisely the anchored words whose anchored letter is at distance k from each end.

Next, we define the semigroup S. A *multiset* in the universe U, where U is a set, is a collection of elements from U that allows for multiple instances. For instance, the multiset $\{3, 3, 4\}$, in the universe $\mathbb{N}$, contains two occurrences of 3 and one occurrence of 4. Given a multiset I and an element $x \in I$, the *multiplicity* of x in I, denoted as $m_I(x)$, is the number of times x occurs in I. The *support* of I, denoted as $Supp(I)$, is the set of all elements occurring in I. Given two multisets I and J, the *union* of I and J, denoted as $I \cup J$, is the multiset K with the support $Supp(I) \cup Supp(J)$ and multiplicity $m_K(x) = m_I(x) + m_J(x)$, for all $x \in Supp(K)$. Let $M(U)$ denote the set of all *non-empty* multisets in the universe U. The set $M(U)$ is a semigroup with $\cup$ as the semigroup operation.

Consider the semigroup $M(W_k)$. Clearly, $M(W_k)$ is infinite. For each multiset I in $M(W_k)$, we let I^r denote the multiset with support $\{x^r \mid x \in I\}$ and multiplicity $m_{I^r}(x^r) = m_I(x)$, for each $x \in I$. The reverse operation is an involution on $M(W_k)$.

We define the following equivalence relation $\equiv^m$ on the multisets in $M(W_k)$; for I, J in $M(W_k)$, $I \equiv^m J$ if

1) $m_I(x) =^m m_J(x)$, for each non-zero element x of W_k, and,
2) $m_I(x) + m_I(x^r) =^m m_J(x) + m_J(x^r)$, for each zero element x of W_k.

The equivalence relation $\equiv^m$ is a congruence relation on the involution semigroup $M(W_k)$: If $I \equiv^m J$, then $I^r \equiv^m J^r$, and also, if $I \equiv^m J$ and $K \equiv^m L$, then $I \cup K \equiv^m J \cup L$, for all $I, J, K, L \in M(W_k)$.

We let S to be the quotient $M(W_k)/\equiv^m$. Clearly, S is a finite semigroup. We denote the semigroup operation on S additively. The $\star$-semigroup $\mathscr{S} = (S, r)$ is aperiodic and commutative.

We can extend the involutory action $\otimes$ of $\mathscr{T}$ on W_k, to the $\star$-semigroup $\mathscr{S}$ pointwise, i.e., we let the left action $u \otimes I$, for $u \in T$ and $I \in S$, to be the multiset of anchored words with the support $\{u \otimes \hat{w} \mid \hat{w} \in I\}$ and multiplicity

$$m_{u \otimes I}(\hat{v}) = \sum_{\hat{w} \in I, u \otimes \hat{w} = \hat{v}} m_I(\hat{w}) , \tag{18}$$

for $\hat{v} \in W_k$. Right action $I \otimes u$ is defined dually. By Equations 15, 16, 17, the actions of T on S are involutory. It is easily observed that, by definition of the action $\otimes$ and the congruence $\equiv^m$, for each idempotent $e \in T$ and each multiset $I \in S$,

$$e \otimes I \otimes e^r = e \otimes I^r \otimes e^r . \tag{19}$$

Hence the action of $\mathscr{T}$ on $\mathscr{S}$ is locally hermitian. For convenience, below we identify the anchored word $\hat{w}$ and the singleton multiset $\{\hat{w}\}$. Also, we will omit writing the action $\otimes$ explicitly.

Let $\mathscr{S} ** \mathscr{T}$ be the locally hermitian semidirect product of $\mathscr{S}$ and $\mathscr{T}$ with the action $\otimes$. We define the $\star$-morphism h from $\mathscr{A}^+$ to $\mathscr{S} ** \mathscr{T}$ as

$$h \colon a \mapsto (\hat{a}, a) \in S \times T .$$

It is easy to verify that h is a morphism of involution semigroups.

Next, we show that the morphism h recognises the language L; it suffices to show that if $w \overset{r}{\approx}_{2k+1}^m w'$, then $h(w) = h(w')$.

Assume $w = a_1 \cdots a_n \in A^+$ and $w' = a_1' \cdots a_\ell'$ are two words such that $w \overset{r}{\approx}_{2k+1}^m w'$. If $|w| < 2k+1$ or $|w'| < 2k+1$, then clearly $w = w'$ and $h(w) = h(w')$. Therefore, assume that both words have length at least $2k + 1$. The image of w under h is of the form

$$\begin{aligned} h(w) &= (\hat{a_1}, a_1) \cdots (\hat{a_n}, a_n) \\ &= (\hat{a_1}a_2 \cdots a_n + a_1\hat{a_2} \cdots a_n + \cdots + a_1 \cdots a_{n-1}\hat{a_n}, \\ &\quad a_1 \cdots a_n) . \end{aligned} \tag{20}$$

Observe that if $u, v \in T$ are such that $|u|, |v| \ge k$, then $u \otimes \hat{w} \otimes v = suff_k(u) \otimes \hat{w} \otimes pref_k(v)$, and $u \odot v = pref_k(u) \cdot suff_k(v)$, for each $\hat{w} \in S$. Since $n \ge 2k + 1$, we derive,

$$\begin{aligned} h(w) = (&\hat{a_1}a_2 \cdots a_{k+1} + a_1\hat{a_2} \cdots a_{k+2} + \cdots \\ &+ a_{n-k} \cdots a_{n-1}\hat{a_n}, \quad a_1 \cdots a_k a_{n-k+1} \cdots a_n) . \end{aligned}$$

Similarly for $w' = a_1' \cdots a_\ell'$ the image is of the form

$$\begin{aligned} h(w') = (&\hat{a_1'}a_2' \cdots a_{k+1}' + a_1'\hat{a_2'} \cdots a_{k+2}' + \cdots \\ &+ a_{\ell-k}' \cdots a_{\ell-1}'\hat{a_\ell'}, \quad a_1' \cdots a_k' a_{\ell-k+1}' \cdots a_\ell') \end{aligned}$$

Since w and w' have a common prefix and suffix of length $2k$, their images in T are identical. Hence, all it remains to conclude the claim is to show that the summations

$$L = \hat{a}_1 a_2 \cdots a_{k+1} + a_1 \hat{a}_2 \cdots a_{k+2} + \cdots + a_{n-k} \cdots a_{n-1} \hat{a}_n$$
$$R = \hat{a'_1} a'_2 \cdots a'_{k+1} + a'_1 \hat{a'_2} \cdots a'_{k+2} + \cdots + a'_{\ell-k} \cdots a'_{\ell-1} \hat{a'_\ell}$$

have the same value in S. Observe that, since S is aperiodic and commutative, we can freely commute the terms in L and R.

The zero elements in the summation L are precisely the terms of the form

$$a_{i-k} \cdots a_{i-1} \hat{a}_i a_{i+1} \cdots a_{i+k} \ ,$$

where $k < i \le n - k$. We observe that there is a one-to-one correspondence between the factors of w of length $2k+1$ and zero elements in the summation L; a factor of w of length $2k+1$ centred at a position $k < i \le n-k$, corresponds to the term $a_{i-k} \cdots a_{i-1} \hat{a}_i a_{i+1} \cdots a_{i+k}$. Similarly, there is a one-to-one correspondence between the factors of w' of length $2k+1$ and zero elements in the summation R. The non-zero elements in L and R are identical, owing to the fact that w and w' have the same prefix and suffix of length $2k$. Finally, since sets of $2k + 1$-factors of w and w' are equal up to reverse and up to threshold m, it mean that sets of zeros of L and R are also identical up to reverse and threshold m. Since in the semigroup $\mathscr{S}$, by definition of $\equiv^m$, zero elements are identified with their reverse and they are counted only up to threshold m, we deduce that $L = R$. Hence the claim is proved.

Next we prove the converse direction.

Let $\mathscr{A} = (A, id)$ be an involutory alphabet. Assume $L \subseteq A^+$ is recognised by the locally hermitian product $\mathscr{S} ** \mathscr{T}$, for some aperiodic commutative $\star$-semigroup $\mathscr{S} = (S, \star)$ and locally trivial $\star$-semigroup $\mathscr{T} = (T, \diamond)$, with the morphism $h \colon \mathscr{A}^+ \to \mathscr{S} ** \mathscr{T}$ and the accepting set P.

Note that since $\mathscr{A}$ is hermitian, for each $(s, t) \in S \times T$ such that $(s, t) = h(a)$ for some $a \in A$,

$$(s, t) = (s^\star, t^\diamond) \ . \tag{21}$$

We use the following property of locally trivial semigroups: If T is locally trivial, then there is a $k \ge 1$ such that

$$x_1 \cdots x_k \cdot z \cdot y_1 \cdots y_k = x_1 \cdots x_k \cdot y_1 \cdots y_k \text{ for all } x_i, z, y_i \in T. \tag{22}$$

Let k be such a number for the locally trivial semigroup T.

Since the semigroup S is aperiodic there exists a natural number $t > 0$, such that $s^t = s^{t+1}$ for all elements $s \in S$.

Let $w, w' \in \mathscr{A}^+$ be two words such that $w \overset{r}{\approx}{}^t_{4k+1} w'$. We claim that $w \in L$ if and only if $w' \in L$; proving the claim implies the lemma. We assume that w and w' are of length at least $4k + 1$, otherwise $w = w'$, and the claim is immediate.

Let $w = a_1 \cdots a_n \in \mathscr{A}^+$ be a word. Let $h(a_i) = (s_i, t_i)$. The image of w under h is

$$h(w) = (s_1, t_1) \cdots (s_n, t_n)$$
$$= (s_1 t_2 \cdots t_n + t_1 s_2 t_3 \cdots t_n + \cdots + t_1 \cdots t_{n-1} s_n,$$
$$t_1 \cdots t_n) \ . \tag{23}$$

Similarly, let $w' = a'_1 \cdots a'_m \in \mathscr{A}^+$. Assuming that $h(a'_i) = (s'_i, t'_i)$, the image of w' under h is

$$h(w') = (s'_1 t'_2 \cdots t'_m + t'_1 s'_2 t'_3 \cdots t'_m + \cdots + t'_1 \cdots t'_{m-1} s'_m,$$
$$t'_1 \cdots t'_m) \ . \tag{24}$$

Since $m, n \ge 4k + 1$ and $w \overset{r}{\approx}{}^t_{4k+1} w'$, by definition, the words w and w' have the same prefix and suffix of length $4k$. Using Equation 22, we conclude that $t_1 \cdots t_n = t'_1 \cdots t'_m$. Therefore, it only remains to show that the summations

$$L = s_1 t_2 \cdots t_n + t_1 s_2 t_3 \cdots t_n + \cdots + t_1 \cdots t_{n-1} s_n$$
$$R = s'_1 t'_2 \cdots t'_m + t'_1 s'_2 t'_3 \cdots t'_m + \cdots + t'_1 \cdots t'_{m-1} s'_m$$

have the same value in S. Observe that, since S is aperiodic and commutative, we can freely commute the terms in L and R. Let $p_i = t_1 \cdots t_{i-1}$ and $q_i = t_{i+1} \cdots t_n$, for $1 \le i \le n$. Then,

$$L = \sum_{i=1}^n p_i s_i q_i$$

Similarly, letting $p'_i = t'_1 \cdots t'_{i-1}$ and $q'_i = t'_{i+1} \cdots t'_m$, for $1 \le i \le m$. Then,

$$R = \sum_{i=1}^m p'_i s'_i q'_i$$

We are going to consider p_i, q_i, p'_i, q'_i etc. as words over the alphabet T and manipulate them. Let $f \colon T^+ \to T^+$ be the function defined as, for each $w \in T^+$,

$$f(w) = \begin{cases} w & \text{if } |w| < 2k, \\ pref_k(w) \cdot suff_k(w) & \text{otherwise.} \end{cases}$$

Using Equation 22, the expressions L and R can be rewritten as

$$L = \sum_{i=1}^n f(p_i) \, s_i \, f(q_i) \ , \tag{25}$$
$$R = \sum_{i=1}^m f(p'_i) \, s'_i \, f(q'_i) \ . \tag{26}$$

Let L_i denote the term $f(p_i) \, s_i \, f(q_i)$. Similarly, let R_i denote the term $f(p'_i) \, s'_i \, f(q'_i)$.

Since the words w and w' have the same prefix of length $4k$, the partial sums

$$\sum_{i=1}^{2k-1} L_i \text{ and } \sum_{i=1}^{2k-1} R_i$$

are identical. Similarly, since the words have the same suffix of length $4k$, the partial sums

$$\sum_{i=n-2k+1}^{n} L_i \text{ and } \sum_{i=m-2k+1}^{m} R_i$$

are also identical. Thus, all it remains is to show is the equality of the following partial sums:

$$L_f = \sum_{i=2k}^{n-2k} L_i \text{ and } R_f = \sum_{i=2k}^{m-2k} R_i \ .$$

The terms occurring in L_f and R_f are called *special*.

Let $x = t_1 \cdots t_k = t'_1 \cdots t'_k$ and $y = t_{n-k+1} \cdots t_n = t'_{m-k+1} \cdots t'_m$. Observe that, for each special term L_i,

$$L_i = x \cdot (t_{i-k} \cdots t_{i-1}) \, s_i (t_{i+1} \cdots t_{i+1+k}) \cdot y \ .$$

Similarly, for each special term R_i,

$$R_i = x \cdot (t'_{i-k} \cdots t'_{i-1}) \, s'_i (t'_{i+1} \cdots t'_{i+1+k}) \cdot y \ .$$

Before proceeding, we make an observation about special terms. Let e and $e^\diamond$ be two idempotents in T. Let $p, q, u, v \in T^+$ be such that $|p|, |q|, |u|, |v| \geq k$, and let $s \in S$, then,

$$
\begin{align}
pqsuv &= pe(qsu)e^\diamond v \quad \text{(using Equation 22)} \tag{27} \\
&= pe(qsu)^\star e^\diamond v \quad (\because \text{the actions are locally hermitian)} \tag{28} \\
&= peu^\diamond s^\star q^\diamond e^\diamond v \quad (\because \text{the actions are involutory)} \tag{29} \\
&= pu^\diamond s^\star q^\diamond v \ . \quad \text{(using Equation 22)} \tag{30}
\end{align}
$$

Applying the above observation and Equation 21 to the term L_i implies that

$$
\begin{align}
&x \cdot (t_{i-k} \cdots t_{i-1}) \, s_i (t_{i+1} \cdots t_{i+1+k}) \cdot y \notag \\
&= x \cdot (t^\diamond_{i+1+k} \cdots t^\diamond_{i+1}) \, s_i^\star (t^\diamond_{i-1} \cdots t^\diamond_{i-k}) \cdot y \tag{31} \\
&= x \cdot (t_{i+1+k} \cdots t_{i+1}) \, s_i (t_{i-1} \cdots t_{i-k}) \cdot y \ .
\end{align}
$$

Similarly for R_i,

$$
\begin{align}
&x \cdot (t'_{i-k} \cdots t'_{i-1}) \, s'_i (t'_{i+1} \cdots t'_{i+1+k}) \cdot y \notag \\
&= x \cdot (t'^\diamond_{i+1+k} \cdots t'^\diamond_{i+1}) \, {s'_i}^\star (t'^\diamond_{i-1} \cdots t'^\diamond_{i-k}) \cdot y \tag{32} \\
&= x \cdot (t'_{i+1+k} \cdots t'_{i+1}) \, s'_i (t'_{i-1} \cdots t'_{i-k}) \cdot y
\end{align}
$$

We are ready to conclude the claim $L_f = R_f$. Factors of w, of length $2k+1$, and centred at positions i, for $2k \leq i < n - 2k$, are called *special*. Similarly, factors of w', of length $2k + 1$, and centred at positions i, for $2k \leq i < m - 2k$, are also called *special*. We observe that the value of each special term L_i is completely determined by the special factor at i. A similar claim holds for the special term R_i and the special factor of w' at i. Since the semigroup S is aperiodic, it is enough to compare the special factors in each word up to the threshold t. Moreover, by Equations 31 and 32, it is enough to consider the special factors up to reverse. Finally, since $w \stackrel{r}{\approx}^t_{4k+1} w'$, the claim follows. ◄

IV. Varieties of Involutory Languages

Next, we introduce the notion of varieties of involutory languages, and relate it to pseudovarieties of finite involution semigroups by proving an Eilenberg-type correspondence. For this purpose, we define the notion of a syntactic involution semigroup, and show that it possesses properties similar to that of syntactic semigroups. Finally, Theorem 2 is stated in terms of pseudovarieties of involution semigroups.

It is possible to state the results in this section in a much more general setting using categories, following [3], [4], where the notions of recognisability, languages and language varieties etc. are developed for monads and a variety theorem for them is proved. Variety theorems of many settings, for instance finite/infinite words, finite trees etc. can be seen as instances of this unified result. In this framework, the involution operation is a monad that satisfies a certain distributive law. One advantage of this approach is that the variety theorem for involutary languages can be deduced from the general case. But it also requires the significantly more complex language of categories. For the sake of simplicity we do the variety theorem in the usual way.

A. Syntactic Involution Semigroups

The *syntactic congruence* of a language $L \subseteq A^+$, denoted as $\sim_L$, is the congruence relation on A^+, defined as

$$x \sim_L y \quad := \quad uxv \in L \text{ if and only if } uyv \in L, \tag{33}$$

for all words $u, v \in A^*$.

The quotient semigroup $A^+/\sim_L$, denoted as $S(L)$, is called the *syntactic semigroup* of L. Let $[u]_L$ denote the equivalence class of $u \in A^+$ under the congruence $\sim_L$. Then, the surjective morphism $\varphi_L \colon u \mapsto [u]_L$, for $u \in A^+$, from A^+ to $S(L)$, recognises L with the accepting set $\varphi_L(L)$. The morphism φ_L is called the *syntactic morphism* of L.

It is possible to lift the notion of syntactic semigroup to the case of recognition by involution semigroups.

▶ **Definition 7.** *Let $\mathscr{A} = (A, \dagger)$ be an involutory alphabet. The* syntactic $\star$-congruence *of $L \subseteq A^+$ is the relation $\approx_L$ on A^+, given by*

$$x \approx_L y \quad := \quad x \sim_L y \text{ and } x \sim_{L^\dagger} y \ .$$

Since $\approx_L$ is an intersection of two congruence relations on A^+, it is a congruence on A^+. Hence, $A^+/\approx_L$ is a semigroup.

Let $[\![w]\!]_L$ denote the equivalence class of $w \in A^+$ under the congruence $\approx_L$. Firstly, we observe

▶ **Lemma 5.** *Let $\mathscr{A} = (A, \dagger)$ be an involutory alphabet. Let $L \subseteq A^+$. Then, $x \sim_{L^\dagger} y$ if and only if $x^\dagger \sim_L y^\dagger$, for all $x, y \in A^+$.*

Proof. Assume $x \sim_{L^\dagger} y$, for $x, y \in A^+$. Let $u, v \in A^*$. Then, $uxv \in L^\dagger$ if and only if $uyv \in L^\dagger$, by definition of $\sim_{L^\dagger}$. It follows that $v^\dagger x^\dagger u^\dagger \in L$ if and only if $v^\dagger y^\dagger u^\dagger \in L$. Since $u^\dagger, v^\dagger$ are arbitrary, it follows that $x^\dagger \sim_L y^\dagger$. The other directon is similar. ◄

▶ **Lemma 6.** *The map $[\![w]\!]_L \mapsto [\![w^\dagger]\!]_L$, for $w \in A^+$, is an involution on the equivalence classes of $\approx_L$.*

Proof. It suffices to show that $x \approx_L y$ if and only if $x^\dagger \approx_L y^\dagger$, and $(xy)^\dagger \approx_L y^\dagger x^\dagger$, for all words $x, y \in A^+$.

Firstly,

$$
\begin{align}
x \approx_L y &\iff x \sim_L y \text{ and } x \sim_{L^\dagger} y && \text{(by Definition 7)} \\
&\iff x^\dagger \sim_{L^\dagger} y^\dagger \text{ and } x^\dagger \sim_L y^\dagger && \text{(by Lemma 5)} \\
&\iff x^\dagger \approx_L y^\dagger \ . && \text{(by Definition 7)}
\end{align}
$$

Lastly, since $(xy)^\dagger = y^\dagger x^\dagger$, $(xy)^\dagger \approx_L y^\dagger x^\dagger$. ◄

▶ **Definition 8.** *The syntactic $\star$-semigroup of L is the quotient semigroup $A^+/\approx_L$ along with the involution $\dagger\colon [\![w]\!]_L \mapsto [\![w^\dagger]\!]_L$, for $w \in A^+$.*

We can obtain the syntactic $\star$-semigroup of L from the syntactic semigroup of L as well. Let $\varphi_L^\dagger$ denote the morphism from A^+ to $S(L)^{op}$ given by Lemma 1. Let χ_L be the product morphism from A^+ to $S(L) \times S(L)^{op}$ given by, for $w \in A^+$,

$$\chi_L\colon w \mapsto \left(\varphi_L(w), \varphi_L^\dagger(w)\right) . \tag{34}$$

By definition of the morphism $\varphi_L^\dagger$,

$$\chi_L(w) = \left([w]_L, [w^\dagger]_L\right) . \tag{35}$$

It is easy to see that χ_L is a morphism from $\mathscr{A}^+$ to $(S(L) \times S(L)^{op}, flip)$. Let $\mathscr{S}(L)$ denote the sub-$\star$-semigroup that is the image of the morphism χ_L.

▶ **Proposition 3.** *$\mathscr{S}(L)$ is isomorphic to the syntactic $\star$-semigroup of L.*

Proof. We claim that the map $f\colon [\![w]\!]_L \mapsto ([w]_L, [w^\dagger]_L)$, for $w \in A^+$, is an isomorphism between $A^+/\approx_L$ and $\mathscr{S}(L)$. If $w \approx_L w'$, then $w \sim_L w'$ and $w \sim_{L^\dagger} w'$ by definition; by Lemma 5, $w \sim_{L^\dagger} w'$ if and only if $w^\dagger \sim_L w'^\dagger$. Therefore, the map f is well-defined. Clearly, it is surjective. By the definition of $\approx_L$, it is injective as well. Since

$$\begin{aligned}
f([\![uv]\!]_L) &= ([uv]_L, \ [(uv)^\dagger]_L) \\
&= ([uv]_L, \ [v^\dagger u^\dagger]_L) \\
&= ([u]_L, \ [u^\dagger]_L)([v]_L, [v^\dagger]_L) \\
&= f([\![u]\!]_L)f([\![v]\!]_L) ,
\end{aligned}$$

f is a morphism. Finally, since

$$\begin{aligned}
f([\![w^\dagger]\!]_L) &= ([w^\dagger]_L, [(w^\dagger)^\dagger]_L) \\
&= ([w^\dagger]_L, [w]_L) \\
&= f([\![w]\!]_L)^{flip} ,
\end{aligned}$$

f is a morphism of involution semigroups. Therefore f is an isomorphism from $A^+/\approx_L$ to $\mathscr{S}(L)$. ◀

As one would expect, the syntactic $\star$-semigroups $\mathscr{S}(L)$ and $\mathscr{S}(L^\dagger)$ are closely related. They are anti-isomorphic to each other, i.e, $\mathscr{S}(L)$ is isomorphic to the opposite of $\mathscr{S}(L^\dagger)$. For this, first we show

▶ **Lemma 7.** *The map $\alpha\colon [w]_L \mapsto [w^\dagger]_{L^\dagger}$, for $w \in A^+$, is an isomorphism from $S(L)^{op}$ to $S(L^\dagger)$.*

Proof. By Lemma 5, the map is well-defined. Clearly, α is surjective. By the same lemma, it is injective. We verify that

$$\begin{aligned}
\alpha([u]_L[v]_L) &= \alpha([vu]_L) && \text{(Product in } S(L)^{op}) \\
&= [(vu)^\dagger]_{L^\dagger} \\
&= [u^\dagger v^\dagger]_{L^\dagger} \\
&= [u^\dagger]_{L^\dagger}[v^\dagger]_{L^\dagger}
\end{aligned}$$

Hence, α is an isomorphism. ◀

Therefore, the semigroup $S(L) \times S(L)^{op}$ is isomorphic to $S(L) \times S(L^\dagger)$. Likewise $S(L^\dagger) \times S(L^\dagger)^{op}$ is isomorphic to $S(L)^{op} \times S(L)$. Finally, the operation $flip$ is an isomorphism from $S(L) \times S(L)^{op}$ to $S(L)^{op} \times S(L)$. Therefore, $\mathscr{S}(L)$ and $\mathscr{S}(L^\dagger)$ are anti-isomorphic to each other.

We look at some special cases. When $L^\dagger = L$, the semigroup reduct of $\mathscr{S}(L)$ is isomorphic to $S(L)$. Therefore, in the case of a reversible language, $\mathscr{S}(L)$ degenerates into the syntactic semigroup.

When $\mathscr{A}$ is hermitian, for every letter $a \in A$, the element $\chi_L(a) \in \mathscr{S}(L)$ is hermitian. Since such elements generate $\mathscr{S}(L)$, it is hermitian.

▶ **Proposition 4.** *$\mathscr{S}(L)$ divides any $\star$-semigroup recognising L.*
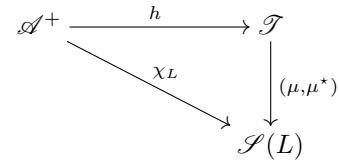

Fig. 1. Universality of syntactic $\star$- semigroups

Proof. The proof is analogous to the semigroup case [17], Proposition 2.7. Assume the language L is recognised by a $\star$-semigroup $\mathscr{T} = (T, \star)$ with a $\star$-morphism $h\colon \mathscr{A}^+ \to \mathscr{T}$. Without loss of generality we can assume that the map h is surjective; otherwise we take $\mathscr{T}$ to be the image of A^+ under h, that is a sub-$\star$-semigroup of $\mathscr{T}$.

Since T recognises the language L, there is a surjective morphism μ from T to $S(L)$ such that $\varphi_L = \mu \circ h$. Similarly, by Lemma 1, there is a surjective morphism $\mu^\star$ from T to $S(L)^{op}$ such that $\varphi_L^\dagger = \mu^\star \circ h$.

Let $(\mu, \mu^\star)$ be the product map from T to $\mathscr{S}(L)$, defined as, for $t \in T$,

$$(\mu, \mu^*)\colon t \mapsto (\mu(t), \mu^\star(t)) . \tag{36}$$

Since h is surjective this implies that there is a surjective $\star$-morphism namely (μ, μ^*) from $\mathscr{T}$ to $\mathscr{S}(L)$. ◀

B. Involution Varieties and Pseudovarieties

A *pseudovariety of finite semigroups* is a set of finite semigroups that is closed under finite direct products, sub-semigroups and quotients.

If $\mathscr{S} = (S, \star)$ and $\mathscr{T} = (T, \dagger)$ are two $\star$-semigroups then the *direct product* of $\mathscr{S}$ and $\mathscr{T}$, denoted $\mathscr{S} \times \mathscr{T}$, is the semigroup $S \times T$ with the involution $(x, y) \to (x^\star, y^\dagger)$.

▶ **Definition 9.** *A class of finite involution semigroups $\mathbf{V}$ is an* involution pseudovariety *if it is closed under taking finite direct products, sub-$\star$-semigroups and quotients.*

If $\mathbf{S}$ is a pseudovariety of finite semigroups, then $\mathbf{S}^\star$ is the pseudovariety of finite involution semigroups whose semigroup reducts are in $\mathbf{S}$.

► **Example 5.** 1) $\mathbb{L}\mathbf{1}$ is the pseudovariety of all finite semigroups S such that S is locally trivial, i.e., for each idempotent $e \in S$ and element $x \in S$, $exe = e$. Hence, $\mathbb{L}\mathbf{1}^\star$ is the set of all involution semigroups whose semigroup reducts are in $\mathbb{L}\mathbf{1}$.

2) $\mathbf{Acom}^\star$ is the pseudovariety of all finite involution semigroups $(S, \star)$ such that S is commutative and aperiodic. Since S is commutative, the identity function id is always an involution on S.

We let Reg_A denote the family of all regular languages over the alphabet A. A *variety of recognisable languages* is a function $\mathcal{V}$ that maps each finite alphabet A to a subset of Reg_A such that the following conditions hold:

1) For each alphabet A, the family of languages $\mathcal{V}(A)$ is closed under union, intersection and complement, i.e., $\mathcal{V}(A)$ is a Boolean algebra.
2) For each $L \in \mathcal{V}(A)$ and each letter $a \in A$, the quotients $a^{-1}L$ and La^{-1} are in $\mathcal{V}(A)$.
3) For each morphism $h\colon A^+ \to B^+$, if $L \in \mathcal{V}(B)$, then $h^{-1}(L) \in \mathcal{V}(A)$.

► **Example 6.** 1) $\mathcal{L}\mathbf{1}$ is the variety of languages where $\mathcal{L}\mathbf{1}(A)$ is the set of all languages $L \subseteq A^+$ that are finite boolean combinations of languages of the form uA^* or A^*u, for some $u \in A^+$. The language variety $\mathcal{L}\mathbf{1}$ corresponds to the pseudovariety $\mathbb{L}\mathbf{1}$ [16].

2) Let $L(a,k) \subseteq A^+$ be the set of all words in which letter $a \in A$ appears exactly k times. $\mathcal{A}com$ is the variety where $\mathcal{A}com(A)$ is the set of all languages $L \subseteq A^+$ that are finite boolean combinations of languages of the form $L(a,k)$ [16], [22].

► **Definition 10.** *A involution variety of languages, also called a $\star$-variety of languages, is a function $\mathcal{V}$ that maps each involutory alphabet $\mathscr{A} = (A, \dagger)$ to a family of recognisable languages over the alphabet A such that the following holds.*

1) *For each $\mathscr{A}$, the set $\mathcal{V}(\mathscr{A})$ is a Boolean algebra.*
2) *For each $L \in \mathcal{V}(\mathscr{A})$ and each letter $a \in A$, the quotients $a^{-1}L$ and La^{-1} are in $\mathcal{V}(\mathscr{A})$.*
3) *Let $h\colon (A^+, \dagger) \to (B^+, \star)$ be a morphism. Then, if $L \in \mathcal{V}(\mathscr{B})$, then $h^{-1}(L) \in \mathcal{V}(\mathscr{A})$.*
4) *If $L \in \mathcal{V}(\mathscr{A})$, then $L^\dagger$ is in $\mathcal{V}(\mathscr{A})$.*

C. Eilenberg-type correspondence

We define a correspondence between $\star$-varieties of languages and pseudovarieties of finite involution semigroups. We write $\mathbf{V}, \mathbf{W}, \ldots$ for pseudovarieties of finite involution semigroups and $\mathcal{V}, \mathcal{W}, \ldots$ for involution varieties of languages. With each involution variety of languages $\mathcal{V}$, we associate a pseudovariety of involution semigroups $\mathbf{V}$, denoted as $\mathcal{V} \mapsto \mathbf{V}$, where $\mathbf{V}$ is the pseudovariety generated by the syntactic $\star$-semigroups of languages in $\mathcal{V}$. Likewise, we associate an involution variety of languages $\mathcal{V}$ with every pseudovariety of involution semigroups $\mathbf{V}$, denoted as $\mathbf{V} \mapsto \mathcal{V}$, where $\mathcal{V}$ maps each involutory finite alphabet $(A, \dagger)$ to all the languages over A recognised by $\star$-semigroups in $\mathbf{V}$.

► **Theorem 3.** *The correspondences $\mathbf{V} \mapsto \mathcal{V}$ and $\mathcal{V} \mapsto \mathbf{V}$ define mutually inverse bijective correspondences between pseudovarieties of involution semigroups and $\star$-varieties of languages.*

Rest of this subsection is devoted to the proof of the above theorem. The following lemma is easy to verify.

► **Lemma 8.** *The correspondences are well defined. i.e.*

1) *If $\mathcal{V}$ is a $\star$-variety of languages and $\mathcal{V} \mapsto \mathbf{V}$, then $\mathbf{V}$ is a pseudovariety of involution semigroups.*
2) *If $\mathbf{V}$ is a pseudovariety of involution semigroups and $\mathbf{V} \mapsto \mathcal{V}$, then $\mathcal{V}$ is a $\star$-variety of languages.*

Next we show that the maps are mutually inverse bijections. The proof follows a standard recipe; the one outlined here closely follows the proof of the Variety theorem for finite semigroups given in [16].

► **Proposition 5.** *Every involution semigroup $\mathscr{S}$ divides a finite direct product of syntactic $\star$-semigroups of languages recognised by $\mathscr{S}$.*

Proof. Let $\mathscr{S} = (S, \star)$ be an involution semigroup. Let $A \subseteq S$ be such that A generates S and $\dagger$ be the involution $\star$ restricted to A. Let $\mathscr{A} = (A, \dagger)$. Let $h\colon \mathscr{A}^+ \to \mathscr{S}$ be the surjective $\star$-morphism that maps each sequence of elements to their product in S. For $s \in S$, let L_s denote the language recognised by the set $\{s\}$, in other words $L_s = h^{-1}(s)$. Let $\mathscr{S}_s = (S_s, \star_s)$ be the syntactic $\star$-semigroup of L_s and let χ_s be the syntactic $\star$-morphism.

Since $\mathscr{S}$ recognises the languages L_s, there exists a surjective morphism ψ_s from $\mathscr{S}$ to $\mathscr{S}_s$, by Proposition 4. Let $\Pi_{s \in S}\mathscr{S}_s$ denote the direct product of the syntactic $\star$-semigroups $\mathscr{S}_s$ corresponding to each element of $\mathscr{S}$. Let $\prod_{s \in S} \chi_s$ denote the product map $u \mapsto (\chi_s(u))_{s \in S}$, for $u \in A^+$. Similarly, $\psi = \prod_{s \in S} \psi_s$ denotes the product map from $\mathscr{S}$ to $\prod_{s \in S}\mathscr{S}_s$. As product of involutory morphisms these maps are involutory morphisms.

The various maps and semigroups are detailed in Figure IV-C.

We claim that ψ is injective. Assume $\psi(s) = \psi(s')$ for some $s, s' \in S$. Then in particular $\psi_s(s) = \psi_s(s')$. This implies that for all $x, y \in S^1$, $xsy = s \Leftrightarrow xs'y = s$. In particular when $x = y = 1$, this implies $s = s'$. Hence, ψ is injective. Therefore $\mathscr{S}$ is isomorphic to a sub-$\star$-semigroup of $\prod_{s \in S}\mathscr{S}_s$. ◄
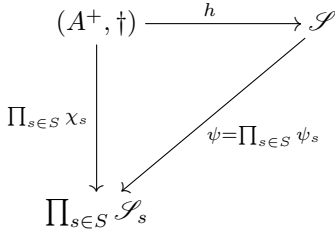
► **Lemma 9.** *The map $\mathbf{V} \mapsto \mathcal{V}$ is injective.*

Proof. Assume that $\mathbf{V} \mapsto \mathcal{V}$ and $\mathbf{W} \mapsto \mathcal{W}$. It suffices to show that if $\mathcal{V} \subseteq \mathcal{W}$, then $\mathbf{V} \subseteq \mathbf{W}$. If so, then
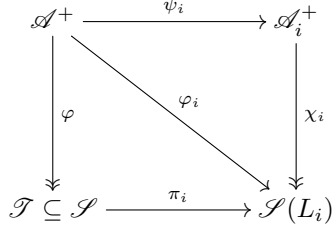
$$\mathcal{V} = \mathcal{W} \implies \mathcal{V} \subseteq \mathcal{W} \text{ and } \mathcal{W} \subseteq \mathcal{V}$$
$$\implies \mathbf{V} \subseteq \mathbf{W} \text{ and } \mathbf{W} \subseteq \mathbf{V}$$
$$\implies \mathbf{V} = \mathbf{W},$$

and the map is injective. Next, we prove this.

Assume that $\mathcal{V} \subseteq \mathcal{W}$. Let $\mathscr{S} = (S, \star)$ be a $\star$-semigroup in $\mathbf{V}$. We show that $\mathscr{S}$ is in $\mathbf{W}$ as well. Let K be the

(a) Various maps in the proof of Proposition 5.

(b) Diagram in Lemma 10

Fig. 2.

set of all languages recognised by $\mathscr{S}$. The languages in K are in $\mathcal{V}$ and hence in $\mathcal{W}$ also. This means there exist $\star$-semigroups recognising languages in K in the pseudovariety $\mathbf{W}$, and hence the syntactic $\star$-semigroups of languages in K are in $\mathbf{W}$, by Proposition 4. Therefore, their product is also in $\mathbf{W}$. Since $\mathscr{S}$ divides this product by Proposition 5, $\mathscr{S}$ is in $\mathbf{W}$. Since $\mathscr{S}$ was arbitrary, this shows that $\mathbf{V} \subseteq \mathbf{W}$. ◄

► **Lemma 10.** *If $\mathcal{V} \mapsto \mathbf{V} \mapsto \mathcal{W}$ then $\mathcal{V} = \mathcal{W}$.*

Proof. Assume that $\mathcal{V} \mapsto \mathbf{V} \mapsto \mathcal{W}$. Firstly, if $L \in \mathcal{V}(\mathscr{A})$, for some $\mathscr{A} = (A, \dagger)$, then one has $\mathscr{S}(L) \in \mathbf{V}$ by definition of the map $\mathcal{V} \mapsto \mathbf{V}$. Hence, $L \in \mathcal{W}(\mathscr{A})$, by definition of the map $\mathbf{V} \mapsto \mathcal{W}$. Therefore, $\mathcal{V}(\mathscr{A}) \subseteq \mathcal{W}(\mathscr{A})$, for every involutory alphabet $\mathscr{A}$.

Next we prove the inclusion $\mathcal{W}(\mathscr{A}) \subseteq \mathcal{V}(\mathscr{A})$. Let $L \in \mathcal{W}(\mathscr{A})$. Then, $\mathscr{S}(L) \in \mathbf{V}$, and since $\mathbf{V}$ is the pseudovariety generated by syntactic $\star$-semigroups of involutory languages in $\mathcal{V}$, there exists finitely many of languages $L_1 \in \mathcal{V}(\mathscr{A}_1), \ldots, L_n \in \mathcal{V}(\mathscr{A}_n)$, where $n > 0$ and $\mathscr{A}_i = (A_i, \dagger_i)$, such that $\mathscr{S}(L)$ divides $\mathscr{S}(L_1) \times \cdots \times \mathscr{S}(L_n)$. Let $\mathscr{S} = \mathscr{S}(L_1) \times \cdots \times \mathscr{S}(L_n)$. Since $\mathscr{S}(L)$ divides $\mathscr{S}$, it is a quotient of a sub-$\star$-semigroup $\mathscr{T} = (T, \star)$ of $\mathscr{S}$.

Observe that $\mathscr{T}$ recognises L. Therefore there exists a surjective $\star$-morphism $\varphi \colon \mathscr{A}^+ \to \mathscr{T}$ and a subset P of T such that $L = \varphi^{-1}(P)$. Let $\pi_i \colon \mathscr{S} \to \mathscr{S}(L_i)$ be the ith projection defined by $\pi_i(s_1, \ldots, s_n) = s_i$. Put $\varphi_i = \pi_i \circ \varphi$ and let $\chi_i \colon \mathscr{A}_i^+ \to \mathscr{S}(L_i)$ be the syntactic morphism of L_i. Since χ_i is surjective, there exists a morphism of semigroups $\psi_i \colon \mathscr{A}^+ \to \mathscr{A}_i^+$ such that $\varphi_i = \chi_i \circ \psi_i$. We can summarise the situation in a diagram in Figure 2.

Next we show that $L \in \mathcal{V}(\mathscr{A})$ that proves the claim $\mathcal{W}(\mathscr{A}) \subseteq \mathcal{V}(\mathscr{A})$. Firstly,

$$L = \varphi^{-1}(P) = \bigcup_{s \in P} \left(\varphi^{-1}(s) \right)$$

Since $\mathcal{V}(\mathscr{A})$ is closed under union, it suffices to establish that for every $s \in P$, one has that $\varphi^{-1}(s)$ belongs to $\mathcal{V}(\mathscr{A})$. Let $s = (s_1, \ldots, s_n)$, then $s = \bigcap_{1 \le i \le n} \pi_i^{-1}(s_i)$. Hence

$$\varphi^{-1}(s) = \bigcap_{1 \le i \le n} \left(\varphi^{-1}\left(\pi_i^{-1}(s_i) \right) \right) = \bigcap_{1 \le i \le n} \varphi_i^{-1}(s_i) \ .$$

As $\mathcal{V}(\mathscr{A})$ is closed under union and intersection, it is enough to show that $\varphi_i^{-1}(s_i)$ is in $\mathcal{V}(\mathscr{A})$ for each $1 \le i \le n$. Since $\varphi_i = \chi_i \circ \psi_i$, one has $\varphi_i^{-1}(s_i) = \psi_i^{-1}(\chi_i^{-1}(s_i))$. Now since $\mathcal{V}$ is a involution variety of languages, it suffices to prove that $\chi_i^{-1}(s_i)$ is in $\mathcal{V}(\mathscr{A}_i)$, that results from the Lemma 11. ◄

► **Lemma 11.** *Let $\mathcal{V}$ be an involution variety of languages and let $\chi \colon \mathscr{A}^+ \to \mathscr{S}(L)$ be the syntactic $\star$-morphism of a language L of $\mathcal{V}(\mathscr{A})$. Then for every $x \in \mathscr{S}(L)$, $\chi^{-1}(x)$ belongs to $\mathcal{V}(\mathscr{A})$.*

Proof. Assume $\mathscr{S}(L) = (S, \star)$. Fix an element $x \in S$. Let $P = \chi(L)$. Then $L = \chi^{-1}(P)$. Setting $E = \{(s, t) \in S^2 \mid sxt \in P\}$, we claim that

$$\{x\} = \bigcap_{(s,t) \in E} s^{-1} P t^{-1} \setminus \bigcup_{(s,t) \in E^c} s^{-1} P t^{-1}$$

Let R be the right-hand side of the above equation. It is clear that x belongs to R. Conversely, let u be an element of R. That means if $sxt \in P$, then $sut \in P$ and if $sxt \notin P$, then $sut \notin P$. It follows that $u \sim_L x$ and thus $u = x$, which proves the claim.

Since χ^{-1} commutes with Boolean operations and quotients, $\chi^{-1}(x)$ is

$$\chi^{-1}(x) = \bigcap_{(s,t) \in E} \chi^{-1}\left(s^{-1} P t^{-1} \right) \setminus \bigcup_{(s,t) \in E^c} \chi^{-1}\left(s^{-1} P t^{-1} \right)$$

which is a Boolean combination of quotients of L and hence belongs to $\mathcal{V}(\mathscr{A}^+)$. ◄

From Lemma 9 and Lemma 10 we conclude Theorem 3.

D. Bilateral Semidirect Products of Involution Pseudovarieties

If $\mathbf{V}$ and $\mathbf{W}$ are pseudovarieties of finite semigroups, then the *bilateral semidirect product of the pseudovarieties* is the pseudovariety generated by all the bilateral semidirect products $S \ast\ast T$ with $S \in \mathbf{V}$, $T \in \mathbf{W}$. Likewise, if $\mathbf{V}$ and $\mathbf{W}$ are pseudovarieties of involution semigroups, then their *bilateral semidirect product* is the pseudovariety generated by all involutory semidirect products of the form $\mathscr{S} \ast\ast \mathscr{T}$, where $\mathscr{S} \in \mathbf{V}$ and $\mathscr{T} \in \mathbf{W}$.

► **Definition 11.** *The* locally-hermitian semidirect product *of the pseudovarieties of involution semigroups $\mathbf{V}$ and $\mathbf{W}$ is the pseudovariety generated by all locally hermitian products $\mathscr{S} \ast\ast \mathscr{T}$, where $\mathscr{S} \in \mathbf{V}$ and $\mathscr{T} \in \mathbf{W}$.*

From Proposition 2, we deduce

► **Proposition 6.** *The involution variety* LTT *corresponds to the bilateral semidirect product of the pseudovarieties* **Acom*** *and* $\mathbb{L}\mathbf{1}^\star$.

Definition 2 and Theorem 2 can be generalized to the case of arbitrary involutory alphabets. We call the resulting involution variety of languages $w\mathcal{LRTT}$. Theorem 2 yields

► **Theorem 4.** *The involution variety* $w\mathcal{LRTT}$ *corresponds to the locally hermitian semidirect product of the pseudovarieties* **Acom*** *and* $\mathbb{L}\mathbf{1}^\star$.

V. Conclusion

The membership problem for a class of languages $\mathcal{C}$ asks: is the given regular language a member of $\mathcal{C}$. The membership question is decidable for many families, the archetypical example is the class of aperiodic languages [20]. The membership problem for locally threshold testable (LTT) languages is equivalent to the problem of finding out if a given semigroup divides a semigroup in **Acom** $* \mathbb{L}\mathbf{1}$. Since $\mathbf{V} * \mathbb{L}\mathbf{1} = \mathbf{V} * \mathbf{D}$ [21] (where $\mathbf{D}$ is the pseudovariety of *definite* semigroups that satisfies $se = e$ for all $e^2 = e$) the problem is slightly simplified. There are two main approaches to solving this problem, the first one is using the Derived Category theorem [18], [22], [24], [25] and the other is using Presburger arithmetic [2], [19]. Both these approaches crucially depend on the fact that a semigroup S divides a semigroup in $\mathbf{V} * \mathbf{D}$ if and only if it divides a semigroup in $\mathbf{V} * \mathbf{D}_n$ (pseudovariety $\mathbf{D}_n$ satisfies the identity $yx_1 \cdots x_n = x_1 \cdots x_n$), where $n = |S|$ [21]. This is known as the delay theorem.

The key idea behind the delay theorem is that given a finite semigroup S, one can insert an idempotent in any product $x_1 \cdots x_n$, $x_i \in S$, of length $n = |S|$ without affecting the value of the product. This construction is not a left-to-right symmetric one. Hence, it is not clear if a delay theorem holds in the case of involutory semidirect products. This is the major hurdle in obtaining the decidability of membership of wLRTT.

The algebraic characterisation of the class wLRTT presented in this work, is not decidable, i.e., the characterisation does not yield an algorithm to check if a given recognisable language is wLRTT or not. In the case of LTT languages, the decidability is obtained by an algebraic characterisation in terms of categories [25]. The next logical step in our work is to extend the theory to categories.

References

[1] Danièle Beauquier and Jean-Éric Pin. Languages and scanners. *Theoretical Computer Science*, 84(1):3–21, 1991.

[2] Mikołaj Bojańczyk. A new algorithm for testing if a regular language is locally threshold testable. *Information Processing Letters*, 104(3):91 – 94, 2007. URL: http://www.sciencedirect.com/science/article/pii/S002001900700141X, doi:https://doi.org/10.1016/j.ipl.2007.05.015.

[3] Mikołaj Bojańczyk. Recognisable languages over monads. In Igor Potapov, editor, *Developments in Language Theory*, pages 1–13, Cham, 2015. Springer International Publishing.

[4] Mikołaj Bojańczyk. Recognisable languages over monads. *CoRR*, abs/1502.04898, 2015. URL: http://arxiv.org/abs/1502.04898, arXiv:1502.04898.

[5] Janusz A. Brzozowski and Imre Simon. Characterizations of locally testable events. In *Proceedings of the 12th Annual Symposium on Switching and Automata Theory (Swat 1971)*, SWAT '71, pages 166–176, 1971.

[6] Siniša Crvenković and Igor Dolinka. Varieties of involution semigroups and involution semirings: a survey. In *Proceedings of the International Conference "Contemporary Developments in Mathematics" (Banja Luka, 2000)*, Bulletin of Society of Mathematicians of Banja Luka, pages 7–47, 2000.

[7] Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite model theory*. Springer Science & Business Media, 2005.

[8] Paul Gastin, Amaldev Manuel, and R. Govind. Logics for reversible regular languages and semigroups with involution. In *Developments in Language Theory*, pages 182–191, 2019.

[9] Paul Gastin, Amaldev Manuel, and R. Govind. Logics for reversible regular languages and semigroups with involution. *CoRR*, abs/1907.01214, 2019. URL: http://arxiv.org/abs/1907.01214, arXiv:1907.01214.

[10] Mike Holcombe and William Michael Lloyd Holcombe. *Algebraic automata theory*, volume 1. Cambridge University Press, 2004.

[11] Andreas Krebs, Kamal Lodaya, Paritosh Pandya, and Howard Straubing. Two-variable logic with a between relation. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '16, pages 106–115, 2016.

[12] Andreas Krebs, Kamal Lodaya, Paritosh K. Pandya, and Howard Straubing. An Algebraic Decision Procedure for Two-Variable Logic with a Between Relation. In *27th EACSL Annual Conference on Computer Science Logic (CSL 2018)*, volume 119 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 28:1–28:17, 2018.

[13] Andreas Krebs, Kamal Lodaya, Paritosh K Pandya, and Howard Straubing. Two-variable logics with some betweenness relations: Expressiveness, satisfiability and membership. *arXiv preprint arXiv:1902.05905*, 2019.

[14] Mark V Lawson. *Inverse semigroups: the theory of partial symmetries*. World Scientific, 1998.

[15] Robert McNaughton and Seymour A. Papert. *Counter-Free Automata (M.I.T. Research Monograph No. 65)*. The MIT Press, 1971.

[16] Jean-Éric Pin. *Mathematical Foundations of Automata Theory*. URL: https://www.irif.fr/~jep/PDF/MPRI/MPRI.pdf.

[17] Jean-Éric Pin. *Varieties of Formal Languages*. Foundations of computer science. North Oxford Academic, 1986.

[18] Jean-Éric Pin. Expressive power of existential first-order sentences of Büchi's sequential calculus. *Discrete Math.*, 291(1–3):155–174, March 2005. doi:10.1016/j.disc.2004.04.027.

[19] Thomas Place, Lorijn van Rooijen, and Marc Zeitoun. On separation by locally testable and locally threshold testable languages. *Logical Methods in Computer Science*, 10(3), Sep 2014. URL: http://dx.doi.org/10.2168/LMCS-10(3:24)2014, doi:10.2168/lmcs-10(3:24)2014.

[20] Marcel-Paul Schützenberger. On finite monoids having only trivial subgroups. *Information and Control*, 8(2):190–194, 1965.

[21] Howard Straubing. Finite semigroup varieties of the form $\mathbf{V} \star \mathbf{D}$. *Journal of Pure and Applied Algebra*, 36:53 – 94, 1985.

[22] Howard Straubing. *Finite Automata, Formal Logic, and Circuit Complexity*. Birkhäuser Verlag, Basel, Switzerland, 1994.

[23] Wolfgang Thomas. Classifying regular events in symbolic logic. *Journal of Computer and System Sciences*, 25(3):360–376, 1982.

[24] Denis Thérien and Alex Weiss. Graph congruences and wreath products. *Journal of Pure and Applied Algebra*, 36:205 – 215, 1985. URL: http://www.sciencedirect.com/science/article/pii/0022404985900714, doi:https://doi.org/10.1016/0022-4049(85)90071-4.

[25] Bret Tilson. Categories as algebra: An essential ingredient in the theory of monoids. *Journal of Pure and Applied Algebra*, 48(1-2):83–198, 1987.

[26] Yechezkel Zalcstein. Locally testable semigroups. *Semigroup Forum*, 5(1):216–227, Dec 1972. doi:10.1007/BF02572893.