Automata, monoids, and logic

Sam van Gool

November 18, 2022

These are lecture notes for the second part of MPRI course 2.16, as given in 2022-2023.

Note. The lectures contained more complete discussions of the proofs of the various statements given below. As part of the course, it is expected that the student understands and can reproduce the arguments in those proofs, even if they are not reproduced in these notes.

1 Automata and monoids

In this part, we make a first link between automata and monoids. It is also the occasion to fix some definitions and notations.

Definition 1. A *semigroup* is a pair $(S, \cdot)$ where S is a set and $\cdot$ is a binary associative operation on S, that is, $x \cdot (y \cdot z) = (x \cdot y) \cdot z$ for all $x, y, z \in S$. We often omit notation for $\cdot$.

A *monoid* is a tuple $(M, \cdot, 1)$ where $(M, \cdot)$ is a semigroup and 1 is a neutral element, that is, $1x = x = x1$ for all $x \in M$.

A *group* is a monoid $(G, \cdot, 1)$ in which every element is invertible, that is, for all $g \in G$, there exists $h \in G$ such that $gh = 1 = hg$.

A *homomorphism* between semigroups is a function that preserves $\cdot$. A *homomorphism* between monoids is moreover required to preserve 1. A monoid homomorphism between groups is a *group homomorphism*. (Preservation of inverses is automatic.)

A subset T of a semigroup S is called a *subsemigroup* if $xy \in T$ for all $x, y \in T$, and a *submonoid* if it moreover contains 1. The correct notion of *subgroup* of a semigroup is a bit subtle and will be discussed in detail later.

Example 2. Let Σ be a set. The set Σ^* of all finite *words* (i.e. sequences) over Σ is a monoid under concatenation, with neutral element the empty word, denoted ϵ. The *length* of a word w is denoted $|w|$ and is often identified with the finite set $\{1, \ldots, |w|\}$; for $i \in |w|$ we then denote by w_i the i^{th} letter of w. We identify every $a \in \Sigma$ with the word $a \in \Sigma^*$ of length 1.

Proposition 3. *For any set Σ, the monoid Σ^* is the free monoid over Σ, that is, for any function $f \colon \Sigma \to M$, where M is a monoid, there is a unique homomorphism $\bar{f} \colon \Sigma^* \to M$ extending f, that is, such that $\bar{f}(a) = f(a)$ for all $a \in \Sigma$.*

Example 4. For any set Q, the set $\mathrm{Rel}(Q)$ of binary relations from Q to Q is a monoid under relational composition, defined for $R, S \in \mathrm{Rel}(Q)$ by

$$R \cdot S := \{(q, q') \in Q^2 \ : \quad \text{there exists } q'' \in Q \text{ such that } (q, q'') \in R \text{ and } (q'', q') \in S\}.$$

The neutral element is the diagonal relation $\Delta := \{(q, q) : q \in Q\}$.

An important submonoid of $\mathrm{Rel}(Q)$ is $\mathrm{End}(Q)$, the set of functions from Q to Q. We thus $f \circ g$ for the function that first does f and then does g, but we may also write gf for this function.

Definition 5. An *automaton* is a tuple $\mathcal{A} = (Q, \Sigma, \delta, I, F)$, where Q and Σ are sets, $\delta \colon \Sigma \to \mathrm{Rel}(Q)$ is a function, and $I, F \subseteq Q$. The automaton is *finite* if both Q and Σ are finite and (complete) *deterministic* if $\delta(a)$ is a (total) function for all $a \in \Sigma$. An automaton $\mathcal{A}$ *accepts* a word $w \in \Sigma^*$ iff the relation $\overline{\delta}(w)$ intersects $I \times F$ non-trivially, and *rejects* the word otherwise. The set of words accepted by $\mathcal{A}$ is called the *language recognized* by $\mathcal{A}$.

We briefly explain how this succinct definition corresponds to the usual one. We may view an automaton as defined here a Σ-labeled multigraph with set of nodes Q, and for each pair $(q, q') \in Q$ we create an edge $q \xrightarrow{a} q'$ iff $(q, q') \in \delta(a)$. In addition, it has two distinguished sets of states I of *initial* and F of *final* states. The relation $\overline{\delta}(w)$ contains exactly the pairs (q, q') such that there is a path from q to q' in this graph that is labeled by w. Thus, the notion of acceptance given here is exactly the usual one (for non-deterministic automata).

With these definitions in place, the following theorem becomes almost obvious.

Theorem 6. *Let $L \subseteq \Sigma^*$. The following are equivalent:*

1. *The language L is recognized by some finite automaton;*
2. *There exist a finite monoid M and a homomorphism $h \colon \Sigma^* \to M$ such that $L = h^{-1}(P)$ for some $P \subseteq M$;*
3. *The language L is recognized by some finite complete deterministic automaton.*

In light of this theorem, we also say that a homomorphism $h \colon \Sigma^* \to M$ *recognizes* a language L if there is $P \subseteq M$ such that $L = h^{-1}(P)$.

Definition 7. A *regular language* is a set of finite words that satisfies the equivalent conditions of Theorem 6.

In particular, the set of regular languages in an alphabet Σ is now easily seen to form a Boolean algebra, using the characterization (2) in Theorem 6 and the Cartesian product of monoids for the case of intersection.

2 Automata and logic

We define monadic second order (MSO) logic, the extension of first order logic which adds quantification over subsets of the structure. For simplicity, we only consider logic here over linear orders with unary predicates. We also fix three pairwise disjoint sets Σ, X_1 and X_2.

Definition 8 (Syntax of MSO)**.** We inductively define the set of *formulas over alphabet Σ with first-order variables in X_1 and second-order variables in X_2*, and functions FV_1 and FV_2 which assign to any MSO-formula its *set of free first-order and second-order variables*, respectively, as follows.

For any $x, y \in X_1$, $X \in X_2$, and $a \in \Sigma$,

1. the expression $x < y$ is a formula, $FV_1(x < y) = \{x, y\}$, $FV_2(x < y) = \emptyset$;
2. the expression $a(x)$ is a formula, and $FV_1(a(x)) = \{x\}$, $FV_2(a(x)) = \emptyset$;
3. the expression $X(x)$ is a formula, and $FV_1(X(x)) = \{x\}$, $FV_2(X(x)) = \{X\}$;
4. $\bot$ is a formula with $FV_1(\bot) = FV_2(\bot) = \emptyset$;
5. when ϕ, ψ are formulas, $\phi \vee \psi$ is also a formula, and $FV_r(\phi \vee \psi) = FV_r(\phi) \cup FV_r(\psi)$ for $r = 1, 2$;
6. when ϕ is a formula, $\neg\phi$ is also a formula, and $FV_r(\neg\phi) = FV_r(\phi)$ for $r = 1, 2$;
7. when $x \in FV(\phi)$, $\exists x\phi$ is a formula and $FV_1(\exists x\phi) = FV_1(\phi) \setminus \{x\}$, $FV_2(\exists x\phi) = FV_2(\phi)$;
8. when $X \in FV(\phi)$, $\exists X\phi$ is a formula and $FV_1(\exists X\phi) = FV_1(\phi)$, $FV_2(\exists X\phi) = FV_2(\phi) \setminus \{X\}$.

A formula of type (1) – (4) is called *atomic*; a *literal* is an atomic or negated atomic formula. A formula is *first order* if it does not use (3) and (8) in its construction. A formula ϕ is a *sentence* if $FV_1(\phi) = FV_2(\phi) = \emptyset$. The notation $\phi(\overline{x}, \overline{X})$ means: ϕ is a formula with $FV_1(\phi) \subseteq \overline{x}$ and $FV_2(\phi) \subseteq \overline{X}$.

We also use some standard abbreviations:

- $\phi \wedge \psi = \neg(\neg\phi \vee \neg\psi)$;
- $\forall x\phi := \neg\exists x\neg\phi$ and $\forall X\phi := \neg\exists X\neg\phi$;
- $x > y := y < x$, $x \leq y := \neg(y < x)$, $x = y := x \leq y \wedge y \leq x$;
- $\mathsf{succ}(x, y) := x < y \wedge \forall z, (z < x) \vee (y < z)$
- $\mathsf{first}(x) := \forall y\ x \leq y$, $\mathsf{last}(x) := \forall y\ y \leq x$, $\mathsf{empty} := \forall x\bot$.

The idea is that an MSO-sentence will define a language of Σ-words. To make this definition precise, we will need a slightly more general induction that associates a language of *marked* Σ-words to any MSO-formula, where the marking refers to free variables that might occur in it.

Definition 9 (Semantics of MSO)**.** Let $F_1 = \{x_1, \ldots, x_m\} \subseteq X_1$, $F_2 = \{X_1, \ldots, X_n\} \subseteq X_2$. A *marked* word over Σ, F_1, F_2 is a tuple $(w, \overline{p}, \overline{P})$, where $w \in \Sigma^*$, $\overline{p} \in |w|^{F_1}$ and $\overline{P} \in \mathcal{P}(|w|)^{F_2}$. We inductively define a relation $\models$ between marked words $(w, \overline{p}, \overline{P})$ and formulas ϕ such that $FV_r(\phi) \subseteq F_r$ for $r = 1, 2$:

1. $(w, \overline{p}, \overline{P}) \models x_i < x_j$ iff $p_i < p_j$;
2. $(w, \overline{p}, \overline{P}) \models a(x_i)$ iff $w_i = a$;
3. $(w, \overline{p}, \overline{P}) \models X_i(x_j)$ iff $p_j \in P_i$;
4. $(w, \overline{p}, \overline{P}) \models \bot$ never holds;
5. $(w, \overline{p}, \overline{P}) \models \phi \vee \psi$ iff $(w, \overline{p}, \overline{P}) \models \phi$ or $(w, \overline{p}, \overline{P}) \models \psi$;
6. $(w, \overline{p}, \overline{P}) \models \neg\phi$ iff it is not the case that $(w, \overline{p}, \overline{P}) \models \phi$;

7. $(w, \overline{p}, \overline{P}) \models \exists x_{m+1}\phi$ iff there exists $p_{m+1} \in |w|$ such that $(w, \overline{p}', \overline{P}) \models \phi$, where $\overline{p}'$ is the tuple $\overline{p}$ with p_{m+1} appended;

8. $(w, \overline{p}, \overline{P}) \models \exists X_{m+1}\phi$ iff there exists $P_{m+1} \in |w|$ such that $(w, \overline{p}, \overline{P}') \models \phi$, where $\overline{P}'$ is the tuple $\overline{P}$ with P_{m+1} appended;

The language *defined* by ϕ is the set of marked words $(w, \overline{p}, \overline{P})$ such that $(w, \overline{p}, \overline{P}) \models \phi$.

Note that with these definitions, the abbreviations given after Definition 8 also have the expected meaning. A *formula in negation normal form* is built from literals by applying disjunction, conjunction and quantifiers. Using rewriting rules like $\neg(\phi \vee \psi) \to \neg\phi \wedge \neg\psi$, $\neg\exists x\phi \to \forall x\neg\phi$, etc., it can be easily shown that any formula can be put into an equivalent formula in negation normal form.

Theorem 10. *A language $L \subseteq \Sigma^*$ is regular if, and only if, it is definable by a monadic second order sentence.*

Note that the above result (and the fact that its proof is constructive) in particular gives a decision procedure for satisfiability of a formula of monadic second order logic on finite words: given a formula ϕ, construct its automaton $\mathcal{A}_\phi$ and check whether or not it accepts any word. The arguments of the above proof can also be used to prove that every MSO formula ϕ is equivalent to an MSO formula ϕ' with only existential second order quantifiers, since the formula expressing the behavior of an automaton is of this form.

3 First order logic and aperiodicity

We now investigate first order logic. The goal of this section is to give a decidable criterion on a recognizing monoid for a regular language L that characterizes first order definability.

Example 11. The language

$$L = \{w \in \{a, b\}^* \ : \ w \text{ contains an odd number of } b\text{'s}\}$$

is regular, as can be easily seen using the monoid homomorphism induced by $\{a, b\} \to \mathbb{Z}_2$ sending a to $[0]$ and b to $[1]$, and taking $P = \{[1]\}$. Thus, it is MSO definable by Theorem 10, and an explicit formula is also not too hard to write down; it roughly has the form:

$$\exists X(\text{“}X \text{ contains exactly every other } b\text{-position”})$$

However, the use of the monadic quantifier $\exists X$ is essential: this language can *not* be defined in first order logic. How do we prove that?

3.1 Rank equivalence

In this section, we concentrate on first order logic, so we will use marked words of the form $(w, \overline{p})$, where each p_i is a position in $|w|$, but no subsets are marked. We mostly fix a finite alphabet Σ and we also fix the set of first order variables $X_1 = \{x_1, x_2, \dots\}$, and by an *$n$-marked word* we mean a marked word $(w, p_1, \dots, p_n)$ with $w \in \Sigma^*$.

An important induction parameter in first order logic (and to some extent also in monadic second order logic, but we do not consider that here) is the *quantifier rank* of a formula, defined as the maximum nesting depth of quantifiers. When ϕ is a first order formula, we write $qr(\phi)$ for its quantifier rank; it may be defined formally by a simple induction.

For every $k, n \geq 0$, we introduce an equivalence relation $\equiv_{n,k}$ on n-marked words, defined as follows. Let $(w, p_1, \ldots, p_n)$ and $(v, q_1, \ldots, q_n)$ be n-marked words. Then we say

$$(w, p_1, \ldots, p_n) \equiv_{n,k} (v, q_1, \ldots, q_n) \iff \text{for every FO-formula } \phi(x_1, \ldots, x_n) \text{ with } qr(\phi) \leq k,$$
$$(w, p_1, \ldots, p_n) \models \phi \text{ iff } (v, q_1, \ldots, q_n) \models \phi.$$

We will write $FO_{n,k}$ for the set of FO-formulas with free variables among $\{x_1, \ldots, x_n\}$ and quantifier rank $\leq k$.

The following characterization of this equivalence relation is a fundamental tool. It holds more generally for any finite logical syntax, but we only use it here for words.

Theorem 12 (Fraïssé-Hintikka)**.** *For every n, k, the equivalence relation $\equiv_{n,k}$ has finitely many classes, each of which is first order definable.*

More precisely, there exists a family $(\Theta_{n,k})_{n,k\geq 0}$ of finite sets of $FO_{n,k}$-formulas such that, for every $n, k \geq 0$, every marked word $(w, \overline{p})$ satisfies exactly one of the formulas θ in $\Theta_{n,k}$, and its $\equiv_{n,k}$-equivalence class consists of exactly those marked words $(v, \overline{q})$ that also satisfy θ.

Moreover, for any n-marked words $(w, \overline{p}), (v, \overline{q})$, the following are equivalent for any $k \geq 0$:

1. *$(w, \overline{p}) \equiv_{n,k+1} (v, \overline{q})$;*
2. *for every $p_{n+1} \in |w|$, there exists $q_{n+1} \in |v|$ such that $(w, \overline{p}p_{n+1}) \equiv_{n+1,k} (v, \overline{q}q_{n+1})$, and for every $q_{n+1} \in |v|$, there exists $p_{n+1} \in |w|$ such that $(w, \overline{p}p_{n+1}) \equiv_{n+1,k} (v, \overline{q}q_{n+1})$.*

Also, $(w, \overline{p}) \equiv_{n,0} (v, \overline{q})$ if, and only if, $w(p_i) = v(q_i)$ for every $1 \leq i \leq n$ and $p_i < p_j$ iff $q_i < q_j$ for every $1 \leq i, j \leq n$.

While the proof of this theorem was some work, it now allows us to deduce some things about equivalence and definability rather easily.

When w is a word and $p, q \in \{1, \ldots, |w|\}$, we write $w(p, q)$ for the factor of w on the open interval (p, q); more formally, it has length $q - 1 - p$ when $p < q - 1$ and 0 otherwise, and $w(p, q)_i := w(p + i)$ for $1 \leq i \leq q - 1 - p$. Similarly, we write $w(< q)$ for the prefix of w ending just before q, and $w(> p)$ for the suffix of w starting just after p. If $1 \leq p_1 < \cdots < p_n \leq |w|$ is a strictly increasing tuple of positions in $|w|$, we say that it induces the decomposition

$$w = w(< p_1) \cdot w(p_1) \cdot w(p_1, p_2) \cdot \cdots \cdot w(p_{n-1}, p_n) \cdot w(p_n) \cdot w(> p_n).$$

We need one lemma about relativizing first order formulas to intervals and rays.

Lemma 13. *Let $\overline{x} = (x_1, \ldots, x_n)$ and let y and z be variables not occurring in $\overline{x}$. For any first order formula $\phi(\overline{x})$, there exists a formula $\phi^{(y,z)}(\overline{x}, y, z)$ of the same quantifier rank as ϕ such that,*

for any word w, $q, r \in |w|$, and $\overline{p} \in |w(i,j)|^n$,

$$(w, \overline{p}\ q\ r) \models \phi^{(y,z)} \iff w(q,r), \overline{p} \models \phi.$$

Analogous formulas $\phi^{<y}$ and $\phi^{>z}$ exist for the prefixes and suffixes.

Proposition 14. *Let $(w, \overline{p})$ and $(v, \overline{q})$ be n-marked words. Then $(w, \overline{p}) \equiv_{n,k} (v, \overline{q})$ if, and only if, $(w, \overline{p}) \equiv_{n,0} (v, \overline{q})$, and, for every $1 \leq i, j \leq n$ such that (p_i, p_j) does not contain any element of $\overline{p}$, $w(p_i, p_j) \equiv_{0,k} v(q_i, q_j)$, and also $w(< \min \overline{p}) \equiv_{0,k} v(< \min \overline{q})$ and $w(> \max \overline{p}) \equiv_{0,k} v(> \max \overline{q})$.*

The following now follows easily.

Lemma 15 (Ehrenfeucht-Fraïssé lemma)**.** *Let $w, v \in \Sigma^*$. For any $k \geq 0$, $w \equiv_{0,k+1} v$ if, and only if, for every $p \in |w|$, there exists $q \in |v|$ such that $w(p) = v(q)$, $w(< p) \equiv_{0,k} v(< q)$, and $w(> p) \equiv_{0,k} v(> q)$, and for every $q \in |v|$ there exists $p \in |w|$ such that the same holds.*

We deduce two basic but crucial facts from the above: first, $\equiv_{0,k}$ respects monoid multiplication (i.e., it is a *congruence* in the terminology that we will introduce below), and second, first order logic of rank k can not 'count' beyond 2^k.

Proposition 16. *For any $w, w' \in \Sigma^*$, if $w \equiv_{0,k} w'$, then $uw \equiv_{0,k} uw'$ and $wu \equiv_{0,k} w'u$ for any $u \in \Sigma^*$.*

Proposition 17. *For any $k \geq 0$ and $w \in \Sigma^*$, if $m, m' \geq 2^k$, then the words w^m and $w^{m'}$ are $\equiv_{0,k}$-equivalent.*

We can now easily deduce the non-FO-definability of a language.

Example 18. The language $L = \{w \in \{a,b\}^* : w \text{ contains an odd number of } b\text{'s}\}$ is not first order definable. Indeed, let ϕ be any FO-sentence, and $k := qr(\phi)$. Then, by Proposition 17, $b^{2^k} \models \phi$ iff $b^{2^k+1} \models \phi$, but the first is in L while the second is not (conversely when $k = 0$). Thus, ϕ does not define L.

More generally, given a regular language L, we deduce that a necessary condition for L to be first order definable is that, for every word w, there exists M such that either $w^m \in L$ for every $m \in M$, or $w^m \notin L$ for every $m \in M$. Indeed, if ϕ is of quantifier rank k and defines L, take $M := 2^k$. If $w^M \in L$, then $w^M \models \phi$, so $w^m \in L$ for every $m \in M$, since $w^m \equiv_{0,k} w^M$ by Proposition 17. If $w^M \notin L$, then $w^m \notin L$ for every $m \geq M$.

3.2 Syntactic congruence

There are two problems: it is not clear a priori that the necessary condition given above is also sufficient, and also not that it is decidable. The second problem is most easily dealt with using the concept of syntactic congruence for a regular language.

We recall some basic algebra.

Definition 19. Let S be a semigroup. A *congruence* on S is an equivalence relation $\equiv$ such that for any $x, y, \alpha \in S$, if $x \equiv y$, then $\alpha x \equiv \alpha y$ and $x\alpha \equiv y\alpha$. In this case, the quotient $S/\equiv$ is again a

semigroup. We say $\equiv$ has *finite index* if $S/\equiv$ is finite. When $h\colon S \to T$ is a homomorphism, the equivalence relation

$$\ker(h) := \{(x, y) \in S^2 \ :\ h(x) = h(y)\}$$

is always a congruence.

The following is a special case of the general "first isomorphism theorem" of universal algebra.

Theorem 20. *For any homomorphism $h\colon S \to T$, the quotient semigroup $S/\ker(h)$ is isomorphic to the image of h, via the isomorphism sending a class $[x]_{\ker(h)}$ to $h(x)$.*

Definition 21. Let $L \subseteq \Sigma^+$ be a language. The *syntactic congruence* of L is the equivalence relation on Σ^+ defined by

$$u \equiv_L v \iff \text{for every } \alpha, \beta \in \Sigma^*, \alpha u \beta \in L \text{ iff } \alpha v \beta \in L.$$

The *syntactic semigroup* is $S_L := \Sigma^+/\equiv_L$ and the *syntactic morphism* is the quotient map $\pi_L\colon \Sigma^+ \to S_L$. In the same way, when $L \subseteq \Sigma^*$, we define the *syntactic monoid* $M_L := \Sigma^*/\equiv_L$.

Note that $\equiv_L$ is indeed always a congruence and that $L = \bigcup_{u \in L} [u]_{\equiv_L}$.

Proposition 22. *If $h\colon \Sigma^+ \to S$ is any semigroup homomorphism and $L = h^{-1}(P)$ for some $P \subseteq S$, then there exist a subsemigroup S' of S and a surjective homomorphism of S' onto the semigroup S_L. In particular, the syntactic congruence $\equiv_L$ has finite index if, and only if, L is regular. The same holds true for monoids.*

With the terminology that a semigroup T *divides* a monoid S if T is the homomorphic image of a subsemigroup of S, we can phrase the first part of Proposition 22 as: the syntactic semigroup of L divides any semigroup recognizing L.

We are now ready to state and prove one direction of the announced decidable characterization of first order definable languages.

Definition 23. A finite semigroup S is *aperiodic* if there exists $m \geq 1$ such that $x^m = x^{m+1}$ for every $x \in S$.

Lemma 24. *If S is aperiodic and T divides S, then T is aperiodic.*

Theorem 25. *For any first order definable $L \subseteq \Sigma^*$, the syntactic monoid M_L is aperiodic.*

3.3 Schützenberger's Theorem

To finish this section, we will establish the converse of Theorem 25, in the following form.

Theorem 26. *For any finite aperiodic semigroup S and any homomorphism $h\colon \Sigma^+ \to S$, the language $h^{-1}(P)$ is first order definable for every $P \subseteq S$.*

This requires a somewhat finer analysis of the structure of finite aperiodic semigroups. As a first step, let us establish a useful equivalent characterization of aperiodicity.

Definition 27. A *subgroup* of a semigroup, or of a monoid, is a subsemigroup which is a group.

An element e of a semigroup is called *idempotent* if $e \cdot e = e$.

Note that a subgroup of a monoid $(M, \cdot, 1)$ may not have the same neutral element as M.

Lemma 28. *If S is a finite semigroup, then for any $x \in S$, there exist $m, r \geq 1$ such that $\{x^{m+i} : 0 \leq i < r\}$ is a subgroup of S. In particular, x^p is idempotent for some $p \geq 1$, and if x^q is also idempotent, then $x^q = x^p$.*

Proposition 29. *Let S be a finite semigroup. The following are equivalent:*

1. *there exists $m \geq 1$ such that for every $x \in S$, $x^m = x^{m+1}$;*
2. *for any $x, y, \alpha, \beta \in S$, if $\alpha y = x$ and $x\beta = y$, then $x = y$;*
3. *any subgroup of S is trivial.*

We have been a bit careless about the distinction between semigroups and monoids so far, but it becomes important in the following. We in particular will use the following simple construction which adds an identity to a semigroup, even if there already is one.

Definition 30. Let S be a semigroup. We define $S^I := S \uplus \{I\}$, where I is a new element not in S, and a multiplication on S^I defined as on S, with in addition, $x \cdot I = I \cdot x = x$ for all $x \in S^I$. We call I an *external identity* added to S.

Note that $(\Sigma^+)^I \cong \Sigma^*$. When $h \colon \Sigma^+ \to S$ is a homomorphism, by a slight abuse of notation we also write h for the (unique) homomorphism $\Sigma^* \to S^I$ that extends h and sends ϵ to I. More generally, we have a functor $(-)^I$ from semigroups to monoids. It is the adjoint to the forgetful functor from monoids to semigroups.

We show that a few useful concepts are first order definable, that will be used in the proof below.

Lemma 31. *Let Σ be a finite alphabet and $\Sigma = \Sigma_1 \uplus \Sigma_2$ for two disjoint non-empty subsets Σ_1, Σ_2 of Σ. There exist first order formulas $\mathsf{pre}_i(z)$ and $\mathsf{suf}_i(z)$ such that, for any $w \in \Sigma^*$ and $p \in |w|$,*

$$w, p \models \mathsf{pre}_i(z) \iff w(<p) \text{ is the longest prefix of } w \text{ that is in } \Sigma_i^*,$$

$$w, p \models \mathsf{suf}_i(z) \iff w(<p') \text{ is the longest suffix of } w \text{ that is in } \Sigma_i^*.$$

There also exist first order formulas $\mathsf{first}_i(z)$ and $\mathsf{last}_i(z)$ such that, for any $w \in \Sigma^*$ and $p \in |w|$,

$$w, p \models \mathsf{first}_i(z) \iff w(p) \in \Sigma_i \text{ and if } p \text{ has a predecessor } q, \text{ then } w(q) \notin \Sigma_i,$$

$$w, p \models \mathsf{last}_i(z) \iff w(p) \in \Sigma_i \text{ and if } p \text{ has a successor } q, \text{ then } w(q) \notin \Sigma_i.$$

Note that $\neg\exists z, \mathsf{pre}_i(z)$ holds in a word w exactly if $w \in \Sigma_i^*$, in which case w itself is the longest prefix that is in Σ_i^*.

We now prove Theorem 26.[1] The proof is by induction on the pair $(|S|, |h(\Sigma)|)$ in the lexicographic ordering. That is, we will prove, for every $n \geq 1$, the following statement $H(n)$:

> For any aperiodic semigroup S with $|S| \leq n$, for any alphabet Σ and any homomorphism $h\colon \Sigma^+ \to S$ and $s \in S$, $h^{-1}(s)$ is first order definable.

Note that this suffices, because if $P \subseteq S$ and for each $s \in P$, ϕ_s is a first order formula that defines $h^{-1}(s)$, then $\bigvee_{s \in P} \phi_s$ defines $h^{-1}(P)$.

The statement $H(1)$ is trivial, because $h^{-1}(P)$ is either empty or Σ^+.

Assume $H(k)$ has been proved for all $k < n$. We establish $H(n)$ by a second induction on the parameter $|h(\Sigma)|$. If $|h(\Sigma)| = 1$, let x be the unique element in $h(\Sigma)$ and let $m \geq 1$ be minimal such that $x^m = x^{m+1}$. Then, for any word $w \in \Sigma^+$,

$$h(w) = \begin{cases} x^{|w|} & \text{if } |w| < m, \\ x^m & \text{if } |w| \geq m. \end{cases}$$

Thus, for $s \in S$, we have

$$h^{-1}(s) = \begin{cases} \{w \in \Sigma^+ \ : \ |w| = k\} & \text{if } s = x^k \text{ for some } k < m, \\ \{w \in \Sigma^+ \ : \ |w| \geq m\} & \text{if } s = x^m, \\ \emptyset & \text{otherwise.} \end{cases}$$

Each of these languages is first order definable (exercise).

Now assume $H(n)$ has been proved for all homomorphisms $h\colon \Sigma^+ \to S$ such that $|h(\Sigma)| < r$, for some $r \geq 2$, and let $h\colon \Sigma^+ \to S$ a homomorphism with $|h(\Sigma)| = r$. By restricting the codomain to the subsemigroup $\mathrm{im}(h)$ if necessary, we will assume that the homomorphism h is surjective.

For any $x \in S$, we write $\lambda_x\colon S \to S$ and $\rho_x\colon S \to S$ for the left and right multiplication by x.

We distinguish two cases.

Case 1. For every $a \in \Sigma$, $\lambda_{h(a)}$ and $\rho_{h(a)}$ are surjective.

In this case, note that in fact $\lambda_{h(w)}$ and $\rho_{h(w)}$ are surjective for every $w \in \Sigma^+$, since they are compositions of surjective functions. Since h is surjective, this means that λ_x and ρ_x are surjective for every $x \in S$. We show that the aperiodicity of S now implies that S must be a singleton. Indeed, let $x, y \in S$ be arbitrary. Since λ_x and ρ_y are surjective, pick $\alpha, \beta \in S$ such that $\lambda_x(\alpha) = y$ and $\rho_y(\beta) = x$. This means that $x\alpha = y$ and $\beta y = x$, so $x = y$ by Proposition 29. Thus, we conclude by $H(1)$.

Case 2. There exists $a \in \Sigma$ such that at least one of $\lambda_{h(a)}$ and $\rho_{h(a)}$ is not surjective.

By symmetry, we may assume that $\lambda_{h(a)}$ is not surjective. We now define two alphabets,

$$\Sigma_1 := h^{-1}(h(a)) \text{ and } \Sigma_2 := \Sigma \setminus \Sigma_1.$$

Note that Σ_1 and Σ_2 are proper non-empty subsets of Σ, since $|h(\Sigma)| = r \geq 2$. Since Σ_1 and Σ_2 are a partition of Σ, for any word $w \in \Sigma^+$, there are unique $v_1 \in \Sigma_1^*, v_2 \in \Sigma_2^*$, and $u \in (\Sigma_1^+\Sigma_2^+)^*$

[1] Our proof adapts an argument due to Wilke, also see v.G. & Steinberg, Canad. Math. Bull. vol 62(1), pp. 199-208 (2019).

such that $w = v_2uv_1$. Note that this decomposition is moreover first order definable, in the sense that $v_2 = w(<p)$, $u = w[p, p']$ and $v_1 = w(>p')$ where p is the unique position in w that satisfies $\mathsf{pre}_2(p)$ and p' is the unique position in w that satisfies $\mathsf{suf}_1(p')$, or one of v_1 or v_2 is the empty word, which are cases that can be dealt with separately.

For $i = 1, 2$, let us write S_i for the subsemigroup of S generated by $h(\Sigma_i)$ and h_i for the restriction of h to a homomorphism $h_i \colon \Sigma_i^+ \to S_i$. Since $|h(\Sigma_i)| < |h(\Sigma)|$, we know by induction that $h_i^{-1}(q)$ is first order definable for every $q \in S$; say by a formula $\phi_{q,i}$. Write S_0 for the image of $\lambda_{h(a)}$, which is strictly contained in S by this case's assumption.

The idea of the proof is now the following. Let $s \in S$. Define

$$T_s := \{(s_2, t, s_1) \in S_2^I \times S_0^I \times S_1^I \ : \ s_2ts_1 = s\}.$$

Assume for a moment that there is, for each $t \in S_0^I$, a formula ψ_t defining exactly the language $h^{-1}(t) \cap (\Sigma_1^+\Sigma_2^+)^*$. Then we will have, for any $w \in \Sigma^+$,

$$h(w) = s \iff w \models \bigvee_{(s_2,t,s_1)\in T_s} \exists z \exists z' \mathsf{pre}_2(z) \wedge \mathsf{suf}_1(z') \wedge \phi_{s_2,2}^{(<z)} \wedge \psi_t^{[z,z']} \wedge \phi_{s_1,1}^{(>z')}.$$

(Note that some care needs to be taken in the disjuncts where one or more of the coordinates of the triple (s_2, t, s_1) equal I, but we leave the details of this as an exercise.)

The rest of the proof shows that such formulas ψ_t indeed exist. We decompose $h|_{(\Sigma_1^+\Sigma_2^+)^*}$ into two functions, f and μ. For any $w_1w_2 \in \Sigma_1^+\Sigma_2^+$, define $f(w_1w_2) := h(w_1w_2) \in S_0$, and extend this uniquely to a homomorphism $f \colon (\Sigma_1^+\Sigma_2^+)^* \to S_0^*$. For example, if $h(a) = s$ and $h(b) = t$ with $s \neq t$, then $f(a^5b^2ab^3ab)$ is defined as the finite word (s^5t^2, st^3, st) over the alphabet S_0: each element of the sequence is indeed in S_0 because it starts with $s = h(a)$. Write $\mu \colon S_0^* \to S_0^I$ for the unique homomorphism extending the identity function $S_0 \to S_0$. By the assumption that $\lambda_{h(a)}$ is not surjective, we have $|S_0| < |S|$, so we know by the induction hypothesis $H(|S_0|)$ applied to μ that $\mu^{-1}(t)$ is first order definable for every $t \in S_0$, and so is of course $\mu^{-1}(I) = \{\epsilon\}$. We need to show that $h^{-1}(t) \cap (\Sigma_1^+\Sigma_2^+)^* = f^{-1}(\mu^{-1}(t))$ is also first order definable for every $t \in S_0$.

Let us first show that there exists, for any $s \in S_0$, a first order formula $\mathsf{block}_s(x)$ such that:

$$w, p \models \mathsf{block}_s(x) \iff p \text{ is a first } \Sigma_1\text{-position and } h(w[p, p']) = s, \text{ where } p' \text{ is the subsequent last } \Sigma_2\text{-position.}$$

Indeed, first note that for a word of the form $w = uv$ with $u \in \Sigma_1^+$ and $v \in \Sigma_2^+$, we have $h(w) = s$ iff $w \models \theta_s$, where

$$\theta_s := \exists z, \mathsf{pre}_1(z) \wedge \bigvee \{\phi_{s_1,1}^{<z} \wedge \phi_{s_2,2}^{\geq z} \ : \ (s_1, s_2) \in S_1^I \times S_2^I \text{ such that } s_1s_2 = s\}.$$

Now $\mathsf{block}_s(x)$ can be defined by relativizing this formula to the definitions of 'first Σ_1 position' and 'last Σ_2 position', that is,

$$\mathsf{block}_s(x) := \mathsf{first}_1(x) \wedge \exists x', \mathsf{last}_2(x') \wedge \forall y(x < y < x' \to \neg \mathsf{last}_2(y)) \wedge \theta_s^{[x,x']}.$$

Let $t \in S_0$ and let χ_t be a first order sentence defining the language $\mu^{-1}(t)$ in the alphabet S_0.

We replace in χ_t any occurrence of an atomic predicate $s(x)$, for $s \in S_0$, by the first order formula $\mathsf{block}_s(x)$, and we call the resulting formula ψ_t. It then follows that a word $w \in (\Sigma_1^+\Sigma_2^+)^*$ satisfies ψ_t iff $f(w) \in \mu^{-1}(t)$, iff $\mu(f(w)) = t$. It is easy to define $(\Sigma_1^+\Sigma_2^+)^* \subseteq \Sigma^*$ in FO. □

4 Green's relations and fragments of first order logic

We will now show how low-level fragments of first order logic also correspond to semigroup properties. This will in particular require a somewhat finer study of the structure of finite semigroups. The main result is Simon's theorem, characterizing the languages definable in the fragment $B\Sigma_1$ as those recognized by $\mathcal{J}$-trivial semigroups.

4.1 Suffix-unambiguity and $\mathcal{L}$-triviality

As a warm-up, we will consider the simpler case of $\mathcal{L}$-trivial semigroups.

Definition 32. Let S be a semigroup. A subset I of S is a *left ideal* if for any $s \in I$, $\alpha \in S$, $\alpha s \in S$. For any $s \in S$, the set

$$S^I s := \{\alpha s : \ \alpha \in S^I\}$$

is the smallest left ideal containing S.

For $s, t \in S$, we write $t \leq_{\mathcal{L}} s$ if and only if $t \in S^I s$. We write $s\mathcal{L}t$ iff $s \leq_{\mathcal{L}} t$ and $t \leq_{\mathcal{L}} s$.

A semigroup S is called $\mathcal{L}$-trivial iff the $\mathcal{L}$-classes are singletons, i.e., for any $s, t \in S$, $s\mathcal{L}t$ implies $s = t$.

We have the obvious analogous notions of right ideal, $\leq_{\mathcal{R}}$, $\mathcal{R}$, and $\mathcal{R}$-trivial.

Note that $\leq_{\mathcal{L}}$ is a preorder that is *right compatible*, i.e., if $t \leq_{\mathcal{L}} s$ then $t\beta \leq_{\mathcal{L}} s\beta$ for any $\beta \in S$. Also note that $t \leq_{\mathcal{L}} s$ iff $S^I t \subseteq S^I s$, and thus $s\mathcal{L}t$ iff $S^I s = S^I t$.

Example 33. In the free semigroup Σ^+, we have that $u \leq_{\mathcal{L}} v$ iff v is a suffix of u. Consequently, Σ^+ is $\mathcal{L}$-trivial and $\mathcal{R}$-trivial.

In a group G, we have $u\mathcal{L}v$ (and $u\mathcal{R}v$) for all u, v.

We also give a non-symmetric example. For any set X, define a semigroup $LZ(X)$ on X by $x \cdot y := x$ for all $x, y \in X$. Then, for any $x, y \in X$, we have $x\mathcal{L}y$, because $xy = x$ and $yx = y$, but the semigroup $LZ(X)$ is $\mathcal{R}$-trivial, for if $x\beta = y$ for some β, then $y = x$. This is called the *left zero semigroup* on X.

We note also that if 1 is a neutral element in a semigroup S, then $s \leq_{\mathcal{L}} 1$ (and $s \leq_{\mathcal{R}} 1$) for all $s \in S$.

We now characterize the languages recognized by $\mathcal{L}$-trivial semigroups. We do not actually give a logic fragment in this case, but we just directly describe the languages recognized by $\mathcal{L}$-trivial semigroups; see Bojanczyk Section 2.3 for a connection with linear temporal logic.

Definition 34. Let Σ be a finite alphabet. We will call a sequence $(\Sigma_0, a_1, \Sigma_1, \ldots, a_n, \Sigma_n)$ a *suffix prescription* if $\Sigma_0, \ldots, \Sigma_n$ are subsets of Σ and for each $1 \leq i \leq n$, $a_i \in \Sigma \setminus \Sigma_i$. (In particular, we must have $\Sigma_i \subsetneq \Sigma$ for $1 \leq i \leq n$, but possibly $\Sigma_0 = \Sigma$, and each Σ_i is allowed to be empty.)

For $w \in \Sigma^+$ we say that $(u_0, u_1, \ldots, u_n)$ is a *valid decomposition* (for this suffix prescription) of w if $w = u_0 a_1 u_1 \ldots a_n u_n$ and $u_i \in \Sigma_i^*$ for every $0 \leq i \leq n$. (Here and in what follows, we interpret $\emptyset^*$ as $\{\epsilon\}$.)

We write

$$L(\Sigma_0, \ldots, \Sigma_n, a_1, \ldots, a_n) := \Sigma_0^* a_1 \Sigma_1^* a_2 \cdots a_n \Sigma_n^* = \{w \in \Sigma^* \ : \ w \text{ has a valid decomposition}\},$$

and call this *the left-simple language prescribed by* the suffix prescription $(\Sigma_0, a_1, \Sigma_1, \ldots, a_n, \Sigma_n)$.

We call a language L *suffix unambiguous* if it is a finite union of left-simple languages.

Lemma 35. *A word w has at most one valid decomposition.*

As for aperiodic semigroups, $\mathcal{L}$-triviality may be characterized in an equational way. The following notation is useful for writing equations on finite semigroups. It is well-defined by Lemma 28.

Definition 36. Let S be a finite semigroup. We denote by $\omega(S)$ the smallest positive number ω such that x^ω is idempotent for every $x \in S$.

When S is clear from the context, we write $\omega = \omega(S)$. One may deduce from the proof of Lemma 28 that $\omega(S) \leq |S|!$.

Lemma 37. *A finite semigroup S is $\mathcal{L}$-trivial, if, and only if, for every $x, y \in S$, $(xy)^\omega = y(xy)^\omega$. In particular, finite $\mathcal{L}$-trivial semigroups are aperiodic.*

The nice thing about such an equational characterization is that the following kind of result becomes obvious.

Proposition 38. *Any homomorphic image, subsemigroup, or finite product of finite $\mathcal{L}$-trivial semigroups is again $\mathcal{L}$-trivial.*

Proof. Exercise; use Lemma 37. □

Remark 39. In general, a collection of finite semigroups is called a *pseudovariety* if it is closed under homomorphic images, subsemigroups, and finite products. The above proposition shows that $\mathcal{L}$-trivial semigroups form a pseudovariety, thanks to the equational characterization given in Lemma 37. There is a general theory that shows that pseudovarieties of finite semigroups (more generally, of finite algebraic structures) can always be characterized by "equations", where equation has to be interpreted in the correct way. See classical articles by Banaschewski, Eilenberg, Reiterman, and more recent work by Gehrke, Grigorieff and Pin.

The following general example of an $\mathcal{L}$-trivial semigroup will be useful in one direction of the theorem.

Example 40. Let $(X, \leq)$ be a finite partially ordered set (i.e., $\leq$ is a reflexive, transitive, antisymmetric relation on X). A *contraction* is a function $f\colon X \to X$ such that $f(x) \leq x$ for all $x \in X$. Note that the set $\mathcal{C}(X, \leq)$ of contractions on X is a submonoid of the monoid of $\mathrm{End}(X)$. Here, we define the multiplication $f \cdot g$ to mean: first do g, then f.

Then $C := \mathcal{C}(X, \leq)$ is $\mathcal{L}$-trivial. Indeed, if $f, f' \in C$ and $\alpha f = f'$ for some $\alpha \in C$, then for any $x \in X$ we have $f'(x) = \alpha(f(x)) \leq f(x)$. Thus, $f' \leq_{\mathcal{L}} f$ implies $f' \leq f$ pointwise. It follows that $f' \mathcal{L} f$ implies $f' = f$ pointwise.

In fact, one may also prove that any $\mathcal{L}$-trivial semigroup is isomorphic to a subsemigroup of one of the form $\mathcal{C}(X, \leq)$ (Exercise).

It follows, using Proposition 38, that any subsemigroup of $\mathcal{C}(X, \leq)$ is also $\mathcal{L}$-trivial. We can now prove the crucial step in one direction of the characterization.

Proposition 41. *Any left-simple language can be recognized by a finite $\mathcal{L}$-trivial semigroup.*

Theorem 42. *Let $L \subseteq \Sigma^+$ be a language. Then L is recognized by a finite $\mathcal{L}$-trivial semigroup if, and only if, L is suffix unambiguous.*

4.2 Equations for $\mathcal{J}$-triviality

We now come to $\mathcal{J}$-trivial semigroups.

Definition 43. Let S be a semigroup. A subset J of S is a (two-sided) *ideal* if it is both a left and a right ideal, equivalently, $s \in J$ implies $\alpha s \beta \in J$ for any $\alpha, \beta \in S^I$. The smallest ideal containing an element $s \in S$ is

$$S^I s S^I := \{\alpha s \beta : \alpha, \beta \in S^I\}.$$

We write $s \leq_{\mathcal{J}} t$ iff $s \in S^I t S^I$, and $s \mathcal{J} t$ if both $s \leq_{\mathcal{J}} t$ and $t \leq_{\mathcal{J}} s$. A semigroup is *$\mathcal{J}$-trivial* if $s \mathcal{J} t$ implies $s = t$.

Remark 44. The relation $\leq_{\mathcal{H}}$ is defined as the intersection of $\mathcal{L}$ and $\mathcal{R}$. Aperiodicity is the same as $\mathcal{H}$-triviality. One direction follows from Proposition 29, the other is left as an exercise.

Note that in any semigroup S, and for any $n \geq 1$, $x, y \in S$, we have

$$(xy)^{n+1} = x(yx)^n y. \tag{switch}$$

The following can be proved with the argument that we already saw in Lemma 37 above.

Lemma 45. *For any $x, y \in S$,*

$$y(xy)^\omega \; \mathcal{L} \; (xy)^\omega \; \mathcal{R} \; (xy)^\omega x.$$

We say that S is $\mathcal{J}$-, $\mathcal{R}$- or $\mathcal{L}$-trivial if the corresponding pre-order is antisymmetric and we say that S is aperiodic if $x^\omega x = x^\omega$ for all $x \in S$.

Proposition 46. *The following are equivalent for any finite semigroup S:*

1. *S is $\mathcal{J}$-trivial;*
2. *S is both $\mathcal{L}$-trivial and $\mathcal{R}$-trivial;*
3. *S is aperiodic and for all $x, y \in S$, $(xy)^\omega = (yx)^\omega$.*
4. *for all $x, y \in S$, $(xy)^\omega x = (xy)^\omega = y(xy)^\omega$;*

We conclude in particular that finite $\mathcal{J}$-trivial monoids form a pseudovariety, see Remark 39 above.

4.3 Simon's Theorem

We say that a sentence ϕ of first order logic is $B\Sigma_1$ if it is a Boolean combination of formulas of the form $\exists x_1 \dots \exists x_n \psi$ with ψ quantifier free.

Theorem 47 (I. Simon)**.** *A language $L \subseteq \Sigma^*$ is definable by a $B\Sigma_1$ sentence if, and only if, the syntactic monoid of L is $\mathcal{J}$-trivial.*

A useful intermediate step in proving this theorem will be giving another characterization of the $B\Sigma_1$-definable languages: they are the *piecewise testable* ones.

Definition 48. When $u = a_1 \dots a_n$ is a word, we define

$$L(u) := \Sigma^* a_1 \Sigma^* a_2 \dots \Sigma^* a_n \Sigma^*.$$

(In particular, $L(\epsilon) := \Sigma^*$.) When $v \in L(u)$, we say that u is a *subword* of v, and we write $\downarrow v$ for the set of subwords of v, which is clearly finite. For $w, w' \in \Sigma^*$, we define

$$w \preceq_k w' \iff \downarrow w \cap \Sigma^{\leq k} \subseteq \downarrow w' \cap \Sigma^{\leq k},$$

and $w \sim_k w'$ iff $w \preceq_k w'$ and $w' \preceq_k w$. We call a language (finitely) *piecewise testable* if it is a (finite) union of $\sim_k$-classes, for some $k \geq 0$.

We leave the equivalence of $B\Sigma_1$ definability and piecewise testability as a (to be guided) exercise.

One direction of Simon's theorem is given by the following lemma.

Lemma 49. *For any $k \geq 0$, the quotient $\Sigma^*/\sim_k$ is a finite $\mathcal{J}$-trivial monoid.*

Finally, the difficult direction of Simon's theorem.

Proposition 50. *If a language L is recognized by a finite $\mathcal{J}$-trivial monoid, then it is piecewise testable.*

The following proof is based on a proof given by Howard Straubing (private communication), combined with ideas found in the book of Jean-Éric Pin cited below.

We will need two lemmas.

Lemma 51. *For any $k \geq 0$, there exists $m \geq k$ such that for any $w \in \Sigma^*$, there is a subword $w' \subseteq w$ such that $w \sim_k w'$ and $|w'| \leq m$.*

Proof. Define $m := |\Sigma^*/\sim_k|$. We prove the statement by induction on the length of $w \in \Sigma^*$. If $|w| \leq m$, we may take $w' = w$. Suppose now $|w| > m$. For $1 \leq i \leq |w|$, denote by w_i the prefix of w of length i. By the pigeon-hole principle, there are $1 \leq i < j \leq |w|$ such that $w_i \sim_k w_j$. Write $w = w_j \beta$ for some $\beta \in \Sigma^*$. Then

$$w = w_j \beta \sim_k w_i \beta.$$

Since $w_i \beta$ is strictly shorter than w, pick $w' \subseteq w_i$ of length at most n such that $w' \sim_k w_i \beta$. Then w' is still a subword of w and it is k-equivalent to w. □

When M is a $\mathcal{J}$-trivial monoid, we define its *$\mathcal{J}$-height* $h(M)$ to be the length of the longest strict $\mathcal{J}$-chain in M.

Lemma 52. *Let M be a $\mathcal{J}$-trivial monoid and $f\colon \Sigma^* \to M$. For any $u, v \in \Sigma^*$ and $a \in \Sigma$, if $uv \sim_{2h(M)-1} uav$, then $f(uv) = f(uav)$.*

Proof. Write $n := h(M)$ and $k := 2n - 1$ and suppose $uv \sim_k uav$. Note that either $u \sim_n ua$ or $av \sim_n v$: if to the contrary neither of these hold, pick $x \subseteq ua$ and $y \subseteq av$ of length $\leq h(M)$ with $x \not\subseteq u$ and $y \not\subseteq v$. Then $x = u'a$ with $u' \subseteq u$ and $y = av'$ with $v' \subseteq v$. Thus, $u'av'$ is a subword of uav, but it can not be a subword of uv.

Without loss of generality, suppose $u \sim_n ua$. Note that in particular $u \neq \epsilon$ since $n \geq 1$. We will show that $f(u) = f(ua)$.

Write $ua = a_1 a_2 \cdots a_k a_{k+1}$, so that in particular $a_{k+1} = a$. For $1 \leq i \leq k+1$, we say that i is a *falling point* if $f(a_1 \cdots a_{i-1}) >_{\mathcal{J}} f(a_1 \cdots a_i)$. In particular, we say that $i = 1$ is a falling point if $f(a_1) <_{\mathcal{J}} f(1) = 1$. Note that for any $i < j$, we have $f(a_1 \dots a_i) = f(a_1 \dots a_j)$ iff there is no falling point in $[i+1, j]$. Our goal is to prove that $k+1$ is not a falling point.

Let $\{i(1) < \cdots < i(p)\}$ be the totally ordered set of falling points in ua. If $p = 0$ then $f(ua) = 1$ so $f(u) = 1$ by $\mathcal{J}$-triviality. We thus assume $p > 0$ and we use as a notational convention $i(0) := 0$. Since $1 <_{\mathcal{J}} f(a_1 \dots a_{i(1)}) <_{\mathcal{J}} \cdots <_{\mathcal{J}} f(a_1 \dots a_{i(p)})$ we have $p + 1 \leq h(M) = n$.

We show that for each falling point $i(r)$, where $1 \leq r \leq p$, there is no occurrence of $a_{i(r)}$ in the interval $(i(r-1), i(r))$. Suppose towards a contradiction that $i(r-1) < j < i(r)$ and $a_j = a_{i(r)}$. Since there is no falling point in $[j, i(r) - 1]$, we must have $f(a_1 \cdots a_{i(r)-1}) = f(a_1 \cdots a_{j-1})$. Thus

$$\begin{aligned} f(a_1 \cdots a_{j-1} a_j) &= f(a_1 \cdots a_{j-1}) f(a_j) \\ &= f(a_1 \cdots a_{i(r)-1}) f(a_{i(r)}) \\ &<_{\mathcal{J}} f(a_1 \cdots a_{i(r)-1}) \\ &= f(a_1 \cdots a_{j-1}) \end{aligned}$$

which shows that j should have been a falling point in the first place.

Finally, we show that $i(p) \neq k+1$. Since $p \leq n-1$ and $ua \sim_n u$, the word $u' := \prod_{r=1}^{p} a_{i(r)}$, which is by definition a subword of ua, is also a subword of u. Let $j(1) < \cdots < j(p)$ be indices in u such that the subword on these indices is u'. Then $j(1) \geq i(1)$ because we showed above that the letter $a_{i(1)}$ does not appear in u before $i(1)$. Inductively, for any $1 < r \leq p$, given that $j(r-1) \geq i(r-1)$, we also have $j(r) \geq i(r)$: since $j(r) > j(r-1) \geq i(r-1)$, and the letter $a_{j(r)} = a_{i(r)}$ does not appear strictly between $i(r-1)$ and $i(r)$, we must have $j(r) \geq i(r)$. Hence $j(p) \geq i(p)$, but $j(p) \leq k$, so $i(p) \neq k+1$. □

Proof of Proposition 50. Let M be a finite $\mathcal{J}$-trivial monoid and let $f\colon \Sigma^* \to M$ be a homomorphism. Choose m as in Lemma 51 for $k := 2h(M) - 1$. We will show that $\sim_m \subseteq \ker(f)$, which clearly suffices. Suppose $w_1 \sim_m w_2$. Pick a subword w' of w such that $w' \sim_k w_1$ and $|w'| \leq n$. Then w' is also a subword of w_2 since $w_1 \sim_m w_2$. Moreover, $w' \sim_k w_1 \sim_m w_2$ implies $w' \sim_k w_2$, since $m \geq k$. Thus, for $i = 1, 2$, since w' can be obtained from w_i by repeatedly removing letters, while staying $\sim_k$-equivalent, we see by repeated applications of Lemma 52 that $f(w') = f(w_i)$. Hence, $f(w_1) = f(w_2)$. □

References and acknowledgments

The main references for this material (and much more):

1. Jean-Éric Pin's book Mathematical Foundations of Automata Theory,
2. Mikolaj Bojanczyk's book Languages recognised by finite semigroups and their generalisations to objects such as trees and graphs, with an emphasis on definability in monadic second-order logic,
3. Howard Straubing's book Finite Automata, Formal Logic, and Circuit Complexity,
4. John Rhodes and Benjamin Steinberg's book The q-theory of finite semigroups,
5. Mai Gehrke's lecture notes on the topic, which will soon be available as chapter 8 of our joint book Topological Duality for Distributive Lattices: Theory and Applications.

I have learned much of the material in these notes from the authors mentioned above, and also from discussions with Thomas Colcombet. Jérémie Marques also helped in further clarifying the proof of Simon's Theorem.