Logical aspects of regular languages

Boris Dashkovsky

Department of Computer Science

McGill University, Montreal

November 1999

A thesis submitted to the

Faculty of Graduate Studies and Research

in partial fulfilment of the requirements of the degree of

Master of Science.

Contents

Abstract

A thorough review of selected results on the logical aspects of regular languages includes the theorem of Büchi on monadic second order logic over strings, a characterization of $FO[<]$ and the theorem of I. Simon. With the help of the Ehrenfeucht-Fraïssé Game we show that $\exists^{(k+1)}$-sentences of $FO[<]$ cannot be expressed as a boolean combination of $\exists^{(k)}$-sentences. Block product of finite monoids is used to analyze languages defined by the boolean closure of the Σ_2-sentences. Positive varieties and the Mal'cev product are introduced and $\Sigma_{n+1} \cap \Pi_{n+1}$ is shown to be equal to the unambiguous polynomial closure of the nth level of the Straubing-Thérien hierarchy. In particular, $\Sigma_2 \cap \Pi_2 = \mathcal{DA}$, where $\mathcal{DA}$ is the smallest variety of languages closed under the unambiguous product.

Résumé

Nous proposons un aperçu complet de résultats choisis concernant les aspects logiques des langages réguliers incluant le théorème de Büchi sur la logique monadique de second ordre sur les chaînes de caractères, la caractérisation de $FO[<]$ et le théorème de I. Simon. Grâce au jeu de Ehrenfeucht-Fraïssé, nous démontrons que, dans $FO[<]$, les énoncés logiques $\exists^{(k+1)}$ ne peuvent être exprimés comme une combinaison booléene d'énoncés $\exists^{(k)}$. Nous utilisons le produit bloc de monoïdes finis pour analyser les langages définis par la fermeture booléene des énoncés Σ_2. Nous présentons également les variétés positives et le produit de Mal'cev et montrons que $\Sigma_{n+1} \cap \Pi_{n+1}$ est égal à la fermeture polynomiale non-ambigue du $n^{\text{ième}}$ niveau de la hiérarchie de Straubing-Thérien. En particulier, $\Sigma_2 \cap \Pi_2 = \mathcal{DA}$, où $\mathcal{DA}$ est la plus petite variété de langages fermée sous le produit non-ambigu.

Acknowledgements

I wish to express my thanks to my supervisor Denis Thérien for introducing me to this field and for his patience and support during the entire time of my residency. My gratitude also goes to my friends – Emil Ciasca and Flavia Majlis for their encouragement.

Introduction

The topic of this Thesis lies at the juncture of formal language theory, algebraic theory of finite automata and model theory in logic.

In 1956 S. C. Kleene showed that the class of languages recognized by finite automata (*regular* languages) coincides with that given by the rational expressions (*rational* languages). This theorem is usually considered to be the foundation of the theory of finite automata. The definition of the *syntactic monoid* was first given in a paper of M. O. Rabin and D. Scott in 1959, where the notion was credited to Myhill. It was shown in particular that a language is recognizable if and only if its syntactic monoid is finite. M. P. Schützenberger made a non-trivial use of the syntactic monoid to characterize an important subclass of the rational languages, the *star-free* languages: a language is star-free if and only if its syntactic monoid is finite and aperiodic.

In the early 1970's I. Simon proved that a language is piecewise testable if and only if its syntactic monoid is $\mathcal{J}$-trivial. Other important syntactic characterization followed, settling the power of the semigroup approach. But it was S. Eilenberg who formulated the appropriate framework for this type of results. A variety of finite monoids is a class of monoids closed under taking submonoids, quotients and finite direct products. Eilenberg's Theorem states that varieties of finite monoids are in one-to-one correspondence with certain classes of regular languages, the varieties of languages.

For these reasons the part of formal language theory concerned with rational languages is now intimately related to both the theory of finite automata and the

theory of finite monoids.

The connection between automata and formal logic dates back to 1936 when A. Turing proved the undecidability of first-order logic by showing how to describe the behaviour of an abstract computing machine with a formula of this logic. More contributions into the research on the logical aspects of the automata theory ensued, with the works of J. R. Büchi on monadic second-order logic and R. McNaughton and S. Papert on automata admitting first-order behavioral description – among the more famous ones.

In the mid-1990's J. E. Pin developed a theory of so-called positive varieties of languages, which – unlike varieties introduced by S. Eilenberg – do not have to be closed under complement. Their algebraic counterpart had to be modified too – they are varieties of finite ordered monoids. The polynomial closure of a variety of languages is always a positive variety; this property led to establishing some new connections between regular languages and logic.

The main objective of this study is concentrated on proving necessary (and sometimes also sufficient) conditions for a property of words to be expressible in a particular logical formalism. We present two general techniques for accomplishing such results: analysis of logical formulæ with methods of the theory of finite monoids and the model-theoretic method of Ehrenfeucht-Fraïssé Games, described in Chapter 7.

Some developments in the field of logical aspects of regular languages – both classical and relatively new – are echoed in this text.

In Chapter 1 we review the main concepts of formal logic and finite automata. The mathematical machinery needed to maintain a degree of self sufficiency of the manuscript includes elements of the theory of finite monoids presented in Chapter 2.

Identities of finite monoids, the notion of variety and its connection with logic are introduced in Chapter 3.

Our digression into semigroup theory continues in Chapter 4 where we define transformation semigroups, wreath product and block product. Acquired tools will

be used in the subsequent chapters to establish some important algebraic characterization of subclasses of regular languages.

Chapter 5 expounds two topics: the theorem of Büchi on monadic second order logic over strings and the algebraic characterization of first-order logic in signature with $<$.

The subject of Chapter 6 is the theorem of I. Simon and piecewise testable languages; we give both combinatorial and algebraic description of these.

In Chapter 7 we present an algebraic characterization of the first two levels of the Straubing-Thérien hierarchy [1] and their connection to the logical hierarchy. We also give a treatment of some special quantification structures and examine the corresponding varieties of languages. The quest for more ties between the two hierarchies reveals some interesting results as we introduce the notions of ordered finite monoids, positive varieties and the Mal'cev product in Chapter 8.

[1] It should be noted, however, that the "characterization" of level 2 is not effective.

Chapter 1

The Basis

1-I Introduction

This chapter focuses on some fundamental concepts in the study of formal languages. We continue by introducing the notion of finite automaton, followed by a digression into formal logic.

1-II Formal Languages

Let $A = \{a_1, a_2, \ldots, a_i\}$ be a finite set of symbols, called an *alphabet* and its elements - *letters* . A *word* (or a *string*) $w = a_1 a_2 \cdots a_m$ over an alphabet A is a finite sequence of letters. By $|w|$ we denote the length m of the word w. For some $a \in A$, $|w|_a$ denotes the number of occurrences of a in w. We then have:

$$\sum_{a \in A} |w|_a = |w|.$$

The *empty string* , denoted 1, has length 0. By juxtaposition uv, or multiplication $u \cdot v$ we mean concatenation of two words u and v producing a sequence with $|uv| = |u| + |v|$ and clearly $|uv|_a = |u|_a + |v|_a$. For the empty word we have $1 \cdot w = w \cdot 1 = w$.

Notation. For a positive integer k and a word w, the form w^k is a shorthand notation for $\underbrace{ww \cdots w}_{k \text{ times}}$. By convention, $w^0 = 1$.

Given two words u and v:

1. u is a *prefix* of v if $\exists x \in A^* : v = ux$;

2. u is a *suffix* of v if $\exists x \in A^* : v = xu$;

3. u is a *factor* of v if $\exists x, y \in A^* : v = xuy$.

A word $u = a_1 a_2 \dots a_n$ is a *subword* of v if there exist words $v_0, v_1, \dots, v_n \in A^*$ such that $v = v_0 a_1 v_1 a_2 \dots a_n v_n$.

The set of all words over the alphabet A is denoted by A^*, the set of all nonempty words - A^+. A subset of A^* is called a *language* . Various operations can be defined over languages. Besides the classical boolean operations (such as finite union, finite intersection, complement) we shall make use of the ones below.

The *product* (or *concatenation product*) of two languages L and K is the language

$$LK = \{uv \in A^* | u \in L, v \in K\}.$$

The *star* of a language $L \subseteq A^*$, denoted by L^* is the language

$$L^* = \{1\} \cup L \cup LL \cup LLL \cup \cdots$$

If K and L are two languages of A^*, the *left* (*right*) *quotient* of L by K is the language $K^{-1}L$ (respectively LK^{-1}). These are defined by:

$$K^{-1}L = \{v \in A^* | Kv \cap L \neq \emptyset\} = \{v \in A^* | \exists u \in K \text{ such that } uv \in L\}$$

and

$$LK^{-1} = \{v \in A^* | vK \cap L \neq \emptyset\} = \{v \in A^* | \exists u \in K \text{ such that } vu \in L\}.$$

1-III Finite Automata

A *deterministic finite automaton* (or DFA) over a finite alphabet A is a quadruple

$$\mathcal{T} = (Q, i, F, \lambda)$$

where Q is a finite set of *states* of the automaton; $i \in Q$ is the *initial state*; $F \subseteq Q$ is the set of *final states* and λ is the *transition function* $\lambda : Q \times A \mapsto Q$ defined for all $q \in Q$ and for all $a \in A$. We shall adopt the shorthand notation qa or $q \cdot a$ for $\lambda(q, a)$.

The domain of the transition function λ can be extended to the set $Q \times A^*$ by induction on the length of the input word:

$$q \cdot 1 = q \qquad \text{and} \qquad q \cdot (wa) = (qw) \cdot a.$$

The string w is *accepted* by DFA if $i \cdot w \in F$. The *language* L *recognized* by the DFA is the set of all such words w:

$$L = \{w \in A^* \mid i \cdot w \in F\}.$$

A language is said to be *regular* if there exists a DFA recognizing it.

1-III.1 The minimal automaton

Let $\mathcal{T} = (Q, i, F, \lambda)$ be a DFA and $L \subseteq A^*$ the language it recognizes. Define the set $Q' \subseteq Q$ of states of the DFA reachable from the initial state i:

$$Q' = \{i \cdot w \mid w \in A^*\}$$

and the following equivalence relation $\sim$ on Q':

$$q_1 \sim q_2 \iff \{w \in A^* \mid q_1 \cdot w \in F\} = \{w \in A^* \mid q_2 \cdot w \in F\}.$$

$q_1 \sim q_2$ implies $q_1 a \sim q_2 a$ for all $a \in A$ and therefore the transition function $\widetilde{\lambda} : Q'/_\sim \times A \mapsto Q'/_\sim$ is well defined for the equivalence classes $[q]$ of $q \in Q$:

$$\widetilde{\lambda}([q],\ a) = [qa].$$

The DFA

$$\widetilde{\mathcal{T}_L} = (Q'/_\sim,\ [i],\ \{[q] \mid q \in F\},\ \widetilde{\lambda})$$

also recognizes L, but its structure depends **only** on L. $\widetilde{\mathcal{T}_L}$ is called the *minimal* automaton of L. Any automaton $\mathcal{A}$ recognizing L has at least as many states as $\widetilde{\mathcal{T}_L}$ does and if $\mathcal{A}$ and $\widetilde{\mathcal{T}_L}$ have the same number of states, they are isomorphic.

Example 1-III.1. Let $A = \{a, b, c\}$ and $L = A^*abA^*$. A DFA recognizing L is pictured in fig. 1.1. One can easily verify that this is the minimal automaton of L.

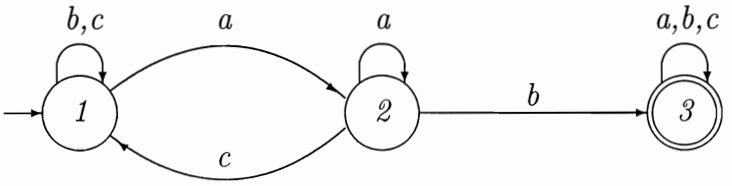

Figure 1.1: The minimal automaton of $L = A^*abA^*$ over $A = \{a, b, c\}$.

1-IV Formal Logic

1-IV.1 Propositional Logic

Define a countable set $X = \{x_1, x_2, \dots\}$ of *boolean* variables (i.e. variables taking on values **True** or **False**).

A *boolean expression* consists of:

(a) a boolean variable x_i; or

(b) an expression of the form: $\neg\phi$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$, where ϕ, ψ are themselves boolean expressions.

The set of boolean variables of an expression ϕ, $X(\phi) \subset X$, is defined inductively as follows:

(a) if ϕ is a boolean variable x_i, then $X(\phi) = \{x_i\}$,

(b) if $\phi = \neg\psi$, then $X(\phi) = X(\psi)$,

(c) if $\phi = (\chi \wedge \psi)$ or $(\chi \vee \psi)$, then $X(\phi) = X(\chi) \cup X(\psi)$.

A *truth assignment* T is a mapping from the set of boolean variables $X(\phi)$ to the set of *truth values* { **True, False** }. We now define what it means for T to *satisfy* ϕ (written $T \models \phi$):

(a) if ϕ is a boolean variable $x_i \in X(\phi)$, then $T \models \phi$ if $T(x_i) =$ **True**,

(b) if $\phi = \neg\psi$, then $T \models \phi$ if it is not the case that $T \models \psi$,

(c) if $\phi = (\chi \vee \psi)$ then $T \models \phi$ if either $T \models \chi$ or $T \models \psi$ holds,

(d) if $\phi = (\chi \wedge \psi)$ then $T \models \phi$ if both $T \models \chi$ and $T \models \psi$ hold.

Notation. An expression of the form x_i or $\neg x_i$ is termed a *literal* . We use $(\phi \Rightarrow \psi)$ to mean $(\neg\phi \vee \psi)$; and $(\phi \iff \psi)$ stands for $((\phi \Rightarrow \psi) \wedge (\psi \Rightarrow \phi))$.

It is well known that the relations $\vee$ and $\wedge$ are commutative, associative, distributive and idempotent (see for instance [Pap94]). Furthermore, it follows that every boolean expression ϕ can be rewritten into an equivalent one in *conjunctive*: $\phi = \wedge_{i=1}^{n} C_i$ or *disjunctive*: $\phi = \vee_{i=1}^{n} D_i$ *normal form*, where C_i (called a *clause*) is the disjunction of one or more literals and D_i (called an *implicant*) is the conjunction of one or more literals.

1-IV.2 First-Order Logic

The language of first-order logic is capable of expressing a wide range of mathematical ideas and facts in much more detail than boolean logic.

1-IV.2-a The Syntax

Let us define three disjoint countable sets: V – a set of variables (ranging over values from the domain of a particular expression); Φ – a set of function symbols; Π – a set of relation symbols and the *arity* function: $r : \Phi \cup \Pi \mapsto \mathbb{Z}_+$. A function $f \in \Phi$ with $r(f) = k$, $k \geq 0$ is called a *k-ary* function (similarly for a relation $R \in \Pi$ with $r(R) = k$, $k > 0$, k-ary relation). The set Π is always assumed to contain the binary *equality* relation $=$. A triplet $\Sigma = (\Phi, \Pi, r)$ is called a *vocabulary.* The set of used function and relation symbols $(\Phi \cup \Pi)$ is called the *signature* of the first-order language.

A *term* over the vocabulary Σ is (a) a variable $x \in V$; or (b) an expression $\underline{f(t_1, t_2, \ldots, t_k)}$, where $f \in \Phi$ and $t_1, t_2, \ldots, t_k$ are themselves terms. (This definition allows for a constant when $k = 0$.)

An *atomic expression* over the vocabulary Σ is an expression of the form $\underline{R(t_1, t_2, \ldots, t_k)}$, where $R \in \Pi$ and $t_1, t_2, \ldots, t_k$ are terms.

A *first-order expression* (or *first-order formula*) is

(a) an atomic expression; or

(b) an expression of the form $\underline{\neg\phi}$, $\underline{(\phi \vee \psi)}$ or $\underline{(\phi \wedge \psi)}$, with ϕ, ψ themselves being first-order expressions; or

(c) an expression of the form $\underline{(\forall x \phi)}$, where $x \in V$ and ϕ is a first-order expression.

Notation. The form $\underline{(\exists x \phi)}$ is used as a shorthand for $\underline{\neg(\forall x \neg\phi)}$. When there is no ambiguity we may write $\forall x, y \cdots$ and $\exists x, y \cdots$ to mean respectively $\forall x \forall y \cdots$ and $\exists x \exists y \cdots$.

The symbols $\forall$ and $\exists$ are the *universal* and *existential quantifier* respectively. An appearance of a variable x in the text of an expression ϕ that does not immediately follow a quantifier is called an *occurrence* of x in ϕ. An occurrence of a variable is said to be *bound* if it is referred to by a quantifier; that is, if $\underline{\forall x \phi}$ is an expression, any

occurrence of x in ϕ is bound [1] (variable x is said to be in the *scope* of a quantifier). If the occurrence is not bound, it is *free*. A variable x that has a free occurrence in ϕ is a *free variable* of ϕ. An expression without free variables is called a *sentence*.

Expressions where a prefix of quantifiers precedes a quantifier-free structure are in *prenex normal form*. Any first-order formula can be transformed into one in prenex normal form. If successive quantifiers of the same type are grouped into n alternating blocks beginning with existential quantifiers, i.e. a formula ϕ is of the form

$$\phi = \exists \widehat{x_1} \forall \widehat{x_2} \cdots \exists \widehat{x_{n-1}} \forall \widehat{x_n} \psi(\widehat{x_1}, \widehat{x_2}, \ldots \widehat{x_n}),$$

where $\widehat{x_i}$ are tuples of variables and ψ is quantifier-free, then ϕ is said to be a Σ_n-formula. In the dual case, when n alternating blocks of quantifiers start with a block of universal quantifiers, the expression is called a Π_n-formula. The negation of a Σ_n formula can be written as a Π_n formula.

Remark 1-IV.1. The first block of quantifiers in a Σ_n (or Π_n) formula may be empty.

1-IV.2-b The Semantics

In first-order logic variables, functions and relations may take on much more complex values than just **True** or **False**. To define the semantics of first-order formulæ we construct an analog of a truth assignment for first-order logic, called a *model*.

A *model appropriate* to a given vocabulary $\Sigma = (\Phi, \Pi, r)$ is a pair $M = (U, \mathcal{I})$, where U is a non-empty set (called the *universe* of M) and $\mathcal{I} : V \cup \Phi \cup \Pi \mapsto U$ is an *interpretation function* associating each symbol α in V, Φ, Π with an actual mathematical object α^M in the universe U. That is, for all $x \in V$, $\mathcal{I}$ assigns an actual element $x^M \in U$; to every function symbol $f \in \Phi$, $\mathcal{I}$ assigns an actual function $f^M : U^k \mapsto U^k$, where k is the arity; and to each relation symbol $R \in \Pi$, $\mathcal{I}$ assigns an actual relation $R^M \subseteq U$.

To define what it means for a model $M = (U, \mathcal{I})$ to *satisfy* a first-order expression ϕ (written $M \models \phi$) we follow the structure of a first-order formula:

[1] An occurrence of x is also bound in any expression containing $\underline{\forall x \phi}$ as a subexpression.

(a) if ϕ is an atomic expression, $\phi = R(t_1, t_2, \ldots, t_k)$, then
$(M \models \phi) \iff (t_1^M, t_2^M, \ldots t_k^M) \in R^M$;

(b) if ϕ is an expression of the form $\neg\alpha$, $(\alpha \vee \beta)$ or $(\alpha \wedge \beta)$, where α, β are first-order expressions, satisfaction is defined by induction on the structure of ϕ;

(c) if ϕ is an expression of the form $(\forall x \psi)$, then
$(M \models \phi) \iff (\forall u \in U : M_{x=u} \models \psi)$, where $M_{x=u}$ is a new model obtained from M by fixing $x^{M_{x=u}} = u$.

Theorem 1-IV.1 (cf. [EFT94]). *Let ϕ be an expression and M, M' – two models appropriate to the vocabulary of ϕ. If M, M' agree on everything except for the values they assign to the variables that are not free in ϕ, then*

$$M \models \phi \iff M' \models \phi.$$

Consequently, for sentences (i.e. expressions with no free variables) satisfaction by a model does not depend on the values assigned to the variables that are bound (or do not appear) in the expression. More generally, if ϕ is a formula with free variables, whether a model satisfies or fails to satisfy ϕ depends both on the interpretation $\mathcal{I}$ and the set of free variables in ϕ. Therefore a "model appropriate to an expression" shall henceforth refer to the part of the model that deals with the functions, relations and free variables (if any).

1-IV.3 Words as a Model

We shall now assemble the following vocabulary $\Sigma = (\Phi, \Pi, r)$: $\Phi = \{\emptyset\}$, i.e. there will be no functions; the set of relation symbols $\Pi = \{=, <, S, Q_a\}$ includes the equality relation $=$, the precedence order $<$, the successor relation S and unary "label" predicates Q_a defined below.

1-IV.3-a Büchi sequential calculus

Let A be a finite alphabet and let $w = a_1 a_2 \cdots a_n$ be a word over A. Variables $x \in V$ range over the set of letter positions of w, or the *domain* of w: $dom(w) = \{1, \ldots, n\}$.

Let us now define a *word model* $\mathcal{W}$ for w appropriate to the vocabulary Σ:

(a) $<^{\mathcal{W}}$ is the natural order on $dom(w)$;

(b) $S^{\mathcal{W}}(i, i+1)$ is the successor relation for $1 \leq i \leq n-1$; and

(c) $Q_a^{\mathcal{W}}$ are unary predicates collecting for each letter $a \in A$ the word positions i in which the letter a appears: $Q_a^{\mathcal{W}} = \{i \in dom(w) | a_i = a\}$.

Remark 1-IV.2. Observe that the successor relation $S(x, y)$ can be expressed in terms of relation $<$ by the formula $(x < y) \wedge \neg\exists z((x < z) \wedge (z < y))$.

If $p_1, \ldots, p_n$ are positions from $dom(w)$ then

$$(\mathcal{W}, p_1, \ldots, p_n) \models \phi(x_1, \ldots, x_n)$$

means that ϕ is satisfied in $\mathcal{W}$ when the signature symbols (i.e. $=, <, S, Q_\alpha$) are interpreted by the relations of equality, $<^{\mathcal{W}}, S^{\mathcal{W}}, Q_a^{\mathcal{W}}$, respectively and positions $p_1, \ldots, p_n$ are interpretation of variables $x_1, \ldots, x_n$ respectively. The word model $\mathcal{W}$ is called *Büchi sequential calculus* (cf. [Büc60], [Büc62]).

1-IV.3-b The $\mathcal{V}$-structure model

As noted above, in view of theorem 1-IV.1, let us concentrate on the part of the model concerned with the free variables of an expression. The following idea of treating the structures in which we interpret formulæ as being words over an extended finite alphabet emanates from Perrin and Pin (cf. [PP86]).

Let ϕ be a first-order formula such that no variable x in ϕ (and all its subexpressions) has bound occurrences in the scope of two different quantifiers.[2] We construct a finite set $\mathcal{V} \subseteq V$ of first-order variables of ϕ:

$$x \in \mathcal{V} \iff x \text{ has only free occurrences in } \phi.$$

[2] Any first-order formula can be written to satisfy this condition by introducing new names for the bound variables, if needed.

A $\mathcal{V}$-*structure* over A is a word w over the extended alphabet $A \times 2^{\mathcal{V}}$:

$$w = (a_1, P_1) \cdots (a_r, P_r),$$

where $r = |\mathcal{V}|$, $a_i \in A$ and P_i satisfy the following:

$$P_i \cap P_j = \emptyset, \text{ if } i \neq j, \qquad \text{and} \qquad \bigcup_{i=1}^{r} P_i = \mathcal{V}.$$

We now define the meaning of $w \models_{\mathcal{I}} \phi$ by induction on the construction of ϕ:

(a) $w \models_{\mathcal{I}} Q_a(x)$ if and only if w contains a letter of the form (a, P) and $x \in P$;

(b) $w \models_{\mathcal{I}} R(x_1, \ldots, x_k) \iff (p_1, \ldots, p_k) \in R^{\mathcal{I}}$, where $R^{\mathcal{I}}$ is the k-ary relation on $\{1, \ldots, |w|\}$ associated to R by $\mathcal{I}$ and $p_1, \ldots, p_k$ are the positions in w where the variables $x_1, \ldots, x_n$, respectively, occur;

(c) $w \models_{\mathcal{I}} \neg\phi$ if and only if w is not a model of ϕ with respect to the interpretation $\mathcal{I}$;

(d) $w \models_{\mathcal{I}} (\phi \wedge \psi) \iff (w \models_{\mathcal{I}} \phi) \wedge (w \models_{\mathcal{I}} \psi)$;

(e) $w \models_{\mathcal{I}} \exists x \phi$ if and only if there exists i, $1 \leq i \leq r$, such that

$$w' = (a_1, P_1) \cdots (a_i, P_i \cup \{x\}) \cdots (a_r, P_r) \models_{\mathcal{I}} \phi.$$

The atomic expressions of this first-order language are of the form:

(a) $\underline{x = y}$ means x and y refer to the same position in w;

(b) $\underline{S(x, y)}$ says that position x is immediately succeeded by position y;

(c) $\underline{x < y}$ tells us that position x is to the left of position y in w;

(d) $\underline{Q_a(x)}$ reveals that in w position x is occupied by the letter a.

Notation. The set of first order formulæ utilizing the set of relational symbols $\Pi = \{=, <, Q_a\}$ ($\Pi = \{=, S, Q_a\}$) is denoted $FO[<]$ (respectively $FO[S]$).

1-IV.4 Languages defined by first-order expressions

If ϕ is a sentence (i.e. ϕ does not have any free variables), then ϕ can be interpreted in a word $w \in A^*$, in which case the *language defined by* ϕ is

$$L(\phi) = \{w \in A^* \mid w \models_{\mathcal{I}} \phi\}.$$

If ϕ is a formula with free variables in $\mathcal{V}$, then by $L(\phi)$ we denote the set of $\mathcal{V}$-structures that satisfy ϕ. This notion depends both on the interpretation function $\mathcal{I}$ and on the set of free variables $\mathcal{V}$.

Below are some examples of languages defined by first-order sentences.

Example 1-IV.1. An $FO[S]$ sentence

$$\phi = \exists x \exists y \exists z \big(S(x,y) \wedge S(y,z) \Rightarrow \neg \exists p S(p,x) \wedge \neg \exists q S(z,q)\big)$$

defines a set of words with exactly three distinct positions in them:

$$L(\phi) = \{w \in A^* : |w| = 3\}.$$

Example 1-IV.2. Consider an $FO[<]$ sentence $\psi = \exists x(\forall z(z \geq x) \wedge Q_a x)$. It describes a language of all words over A^* beginning with the letter a, i.e. $L(\psi) = aA^*$.

Two expressions ϕ and ψ are said to be *equivalent* if their languages coincide, i.e. $L(\phi) = L(\psi)$.

Remark 1-IV.3. The empty word 1 is allowed as member of formal languages and the empty model $\underline{1}$ is admitted as interpretation of sentences. By convention, $\underline{1}$ satisfies universal sentences $\forall x \phi(x)$, but not existential ones $\exists x \phi(x)$.

1-IV.5 MSO Logic

In a first-order formula only individual variables can be quantified. Allowing quantification over sets of variables as well as individual variables, extends the logical formalism by *second-order monadic variables* or *predicates* (usually written as capitalized

X as opposed to x). With the introduction of corresponding atomic expressions: e.g. $X(x)$ (meaning x belongs to the set X), the resulting system becomes *monadic second-order logic* or MSO-logic (sets are monadic objects).

A second-order formula can also be presented in prenex normal form. A Σ^1_n-formula is an expression with a prefix of n second-order quantifier blocks (beginning with a block of existential quantifiers) trailing by a formula where at most first-order quantifiers occur. Σ^1_n-formulæ of MSO-logic are called *existential monadic second-order formulæ* or EMSO-formulæ.

Example 1-IV.3. Consider a language L over the alphabet $A = \{a, b\}$ where any two occurrences of a are separated by an odd number of b's. L can be expressed by the following MSO sentence:

$$\begin{aligned}\phi \;=\; & \forall x \forall y \Big(Q_a(x) \wedge Q_a(y) \wedge (x < y) \wedge \forall z \big((x < z) \wedge (z < y) \Rightarrow \neg Q_a(z) \big) \\ & \Rightarrow \exists X \big(X(x) \wedge X(y) \wedge \forall p \forall q (S(p,q) \Rightarrow (X(p) \Leftrightarrow \neg X(q))) \big) \Big)\end{aligned}$$

Here the first part of the formula says that x and y are two positions carrying the letter a such that no other a appears between them. Then the second part identifies the set X as containing the position of the first a, then every second position and finally the position of the next letter a.

1-IV.5-a Interpretation of MSO formulæ

The following somewhat over-specialized model is justified by our interest in only interpreting expressions in words; and the fact that we do not deal with second-order variables of arity more than one renders it sufficient.

Let $\mathcal{V}_1$ be a finite set of first-order variables, and $\mathcal{V}_2$ – a finite set of monadic second order variables. A $(\mathcal{V}_1, \mathcal{V}_2)$-*structure* over A is a word

$$w = (a_1, S_1, T_1) \cdots (a_n, S_n, T_n) \in (A \times 2^{\mathcal{V}_1} \times 2^{\mathcal{V}_2})^*$$

such that

$$(a_1, S_1) \cdots (a_n, S_n)$$

is a $\mathcal{V}_1$-structure. No constraints are imposed on the occurrences of the second-order variables in the structure. The definition of $w \models_{\mathcal{I}} \phi$ is the same as for the first-order expressions, with the addition of two new clauses:

1. if x is a first-order variable and X is a second-order variable then $w \models_{\mathcal{I}} X(x)$ if and only if w contains a letter (a_i, S_i, T_i) such that $x \in S_i$ and $X \in T_i$;

2. if X is a second-order variable, then $w \models_{\mathcal{I}} \exists X\phi$ if and only if there exists a (possibly empty) set J of positions in w with the following property: the $(\mathcal{V}_1, \mathcal{V}_2)$-structure w' formed by replacing each letter (a_i, S_i, T_i), with $i \in J$, by $(a_i, S_i, T_i \cup \{X\})$ satisfies ϕ.

The language $L(\phi)$ defined by an MSO expression ϕ is the set of $(\mathcal{V}_1, \mathcal{V}_2)$-structures that satisfy ϕ.

Chapter 2

Finite Monoids

2-I Introduction

In this chapter we present a more algebraic approach to languages as recognizable sets, with monoids replacing finite automata. S. Eilenberg (cf. [Eil76]) showed that monoids provide a powerful and systematic tool for language classification.

2-II The structure of finite monoids

The pair $(S, \times)$ where S is a set and $\times$ is a (binary) associative operation is a *semigroup*. It is customary to write "semigroup S" rather than "semigroup $(S, \times)$".

Notation.

1. Juxtaposition ab is a shorthand for $a \times b$.

2. If $P_1, P_2, \ldots, P_n$ are nonempty subsets of a semigroup S then $P_1P_2 \cdots P_n = \{p_1p_2 \cdots p_n | p_i \in P_i, 1 \leq i \leq n\}$. If $P = P_1 = P_2 = \cdots = P_n$ we write P^n instead of $P_1P_2 \cdots P_n$.

A *monoid* $(M, \cdot, 1)$ is a set M with a binary operation, denoted by $\cdot$, and a distinguished element 1, such that $(M, \cdot)$ is a semigroup with an *identity* 1, i.e. for all $x \in M$, $1 \cdot x = x \cdot 1 = x$. We usually write "monoid M" instead of "monoid $(M, \cdot, 1)$".

An element z of a monoid M is a *zero* of M if for all $s \in M$, $z = zs = sz$. We usually denote such an element z by 0.

Let z_1, z_2 be two zeros of a monoid M. By definition: $z_1 z_2 = z_1$ and $z_1 z_2 = z_2$. Whence, $z_1 = z_2$, i.e. a monoid can have at most one zero. A similar argument shows that a monoid contains a single identity element.

We now turn to subsets of a finite monoid (semigroup) exhibiting special properties.

A *subsemigroup* T of a semigroup S is a subset of S such that $x_1 \in T$ and $x_2 \in T$ imply $x_1 x_2 \in T$. This is equivalent to $T^2 \subseteq T$.

A subset T of a monoid M is a *submonoid* of M if it is closed under the operation of M and contains the identity element, i.e.

(a) $1 \in T$ and

(b) $T^2 \subseteq T$.

Clearly, a submonoid of a monoid is a monoid in its own right.

A monoid M *is generated by* its subset G if every element of M can be written as a product of some elements of G.

A nonempty subset T of a monoid M is a *left ideal* of M if $MT \subseteq T$; a *right ideal* of M if $TM \subseteq T$; a *two-sided ideal* (or simply an *ideal*) if it is both a left and a right ideal, i.e. $MT \cup TM \subseteq T$.

The intersection of all ideals of a monoid M is the *kernel* of M.

A monoid M is *simple* (*left-simple, right-simple*) if no proper subset of M is an ideal (respectively, left ideal, right ideal) of M.

Lemma 2-II.1 (cf. [CP67]). *The set of all ideals of a finite monoid M is closed under intersection and arbitrary union. The intersection of a finite number of ideals is an ideal.*

The lemma above holds for the set of all left (right) ideals of M as well.

An element e of a monoid M is *idempotent* if $e^2 = e$. Let s be an element of a finite monoid M and let S be the submonoid generated by s. The sequence $s^0 = 1, s, s^2, s^3, \ldots$ contains only finitely many distinct elements of S, for S is finite and closed under product. Let p be the smallest positive integer such that there exists an integer $m > 0$ satisfying

$$s^p = s^{p+m}.$$

Let us fix the smallest such m and name it q. Choosing $r \geq 0$ such that $p + r \equiv 0 (\mathrm{mod} q)$ yields for some $i \geq 1$:

$$(s^{p+r})^2 = s^{2(p+r)} = s^{(p+r)+iq} = s^{(p+q)+r} = s^{p+r}$$

That is, s^{p+r} is an idempotent element of S.

Furthermore, the elements $1, s, s^2, \ldots, s^{p+q-1}$ are all distinct. For any integer $n \geq q$ we have $n = iq + j$ (with $i \geq 1$, $0 \leq j < q$) and

$$s^{p+n} = s^{p+iq+j} = s^{p+j},$$

whence

$$S = \{1, s, s^2, s^3, \ldots, s^{p+q-1}\}.$$

Observe also that the set $G = \{s^p, s^{p+1}, \ldots, s^{p+q-1}\}$ is a maximal subgroup of M since the mapping $\phi : G \mapsto \mathbb{Z}_q$ defined by $\phi(s^{p+k}) = p + (k \bmod q)$ is an isomorphism. Since every $s \in S \setminus \{1\}$ has a power in G, s^{p+r} is the only other idempotent of S beside 1. The structure of the submonoid S therefore resembles a frying pan with the dish representing the group G as shown in figure 2.1.

We thus have the following results:

Proposition 2-II.2. *If s is an element of a finite monoid M, then the submonoid S generated by s contains a unique maximal subgroup.*

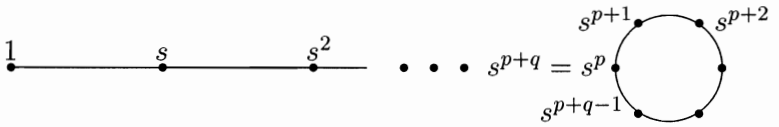

Figure 2.1: The structure of the submonoid S.

Corollary 2-II.3. *Every non-empty finite semigroup contains an idempotent.*

A monoid M is *aperiodic* if for all $x \in M$ there exists an integer n such that $x^n = x^{n+1}$.

An element a of a monoid M is *regular* if $a = asa$ for some $s \in M$. If every element of M is regular, M is *regular*. An element x of M is an *inverse* of a if $a = axa$ and $x = xax$. In a monoid every regular element has an inverse.

A monoid in which every element has a unique inverse is called a *group*. A group is *cyclic* if it is the set of powers of a single element. A cyclic group is commutative.

A *subgroup* H of a group G is a subset of G which is itself a group under the operation of G. Every group has two trivial subgroups: the group itself and the group consisting of the identity. Any non-cyclic group G has necessarily a non-trivial subgroup.

For any group G, and any element $g \in G$, one has

$$Gg = \{g_i g |\ g_i \in G\} = G.$$

Indeed, every g_1 is obtainable as a product $g_1 g^{-1} \cdot g = g_1$ and $g_1 g^{-1}$ is equal to some $g_i \in G$.

If G is a group and H is a subgroup of G, then Ha, where $a \in G$, is called a *right coset* of H in G. (We have a similar definition for a *left coset* .) Assume $c \in Ha \cap Hb$. Then there exists an element $h \in H$ such that $c = ha$, i.e.

$$a = h^{-1}c \qquad \text{and} \qquad Ha = Hh^{-1}c = Hc.$$

In the same way $Hb = Hc$, i.e. if two right cosets of H in G have a common element, they coincide, otherwise they are disjoint.

A subgroup H of G is called a *normal subgroup* if its right cosets coincide with the left ones, i.e. $Ha = aH$. In this case one has $a^{-1}Ha = H$ and hence

$$Ha \cdot Hb = Ha(a^{-1}Ha)b = HHab = Hab,$$

i.e. the product of two right cosets is a right coset. A group which has only trivial normal subgroups is called a *simple group.*

Given a group G and a normal subgroup H, one can use the partition of G into (right) cosets of H to build the factor group G/H, whose elements are the blocks of the partition, i.e. the cosets of H in G.

The next result presents decomposition of finite left-simple semigroups.

Lemma 2-II.4 (cf. [CP67]). *Every finite left-simple semigroup S is isomorphic to a direct product $T \times G$, where G is a group and T is a left-zero semigroup.*

Proof: If s is an element of S, then either $Ss \subset S$ (in which case Ss is a proper left ideal of S), or $Ss = S$, in which case

$$\pi_s : \ t \mapsto ts$$

is a permutation of elements of S. We consider the right action of s on S and

$$\pi_{s'}(\pi_s(t)) = \pi_{ss'}(t) = tss'.$$

Then

$$G = \{\pi_s | \ s \in S\}$$

is a group of permutations of S acting on S on the right. Let T be the set of orbits of this action; $\mathcal{O}_s$ denotes the orbit containing s. We then define a multiplication on T by setting

$$\forall \mathcal{O}_s, \forall \mathcal{O}_t \in T : \ \mathcal{O}_s \cdot \mathcal{O}_t = \mathcal{O}_s,$$

to ensure that T is a left-zero semigroup.

Claim:

$$\phi : \ s \mapsto (\mathcal{O}_s, \pi_s)$$

is a bijection between S and $T \times G$.

We first show that ϕ is surjective. Consider $(\mathcal{O}, \pi_s) \in T \times G$ and $t \in \mathcal{O}$. Then for all $x \in S$

$$\pi_s(x) = xs = \pi_s(x\pi_t^{-1}(t)).$$

Since G is a group, there exists $u \in S$ such that $\pi_u = \pi_t^{-1}$ and hence $\pi_s = \pi_{tus}$ with $tus \in \mathcal{O}$. Thus $(\mathcal{O}, \pi_s) = \phi(tus)$ and ϕ is surjective.

To see that ϕ is injective, assume $(\mathcal{O}_s, \pi_s) = (\mathcal{O}_{s'}, \pi_{s'})$. Then $su = s'$ for some $u \in S$ and thus $\pi_s = \pi_{s'} = \pi_u(\pi_s)$, i.e. π_u is the identity permutation. Hence $s' = su = s$.

And finally ϕ is a function preserving multiplication since

$$ss' \in \mathcal{O}_s$$

and

$$\phi(ss') = (\mathcal{O}_{ss'}, \pi_{ss'}) = (\mathcal{O}_s, \pi_{ss'}) = (\mathcal{O}_s, \pi_s)(\mathcal{O}_{s'}, \pi_{s'}) = \phi(s)\phi(s').$$

Q.E.D.

2-III Homomorphisms and the syntactic congruence

A *homomorphism*[1] φ from a semigroup $(S, \cdot)$ to a semigroup $(S', \star)$ is a mapping φ from the set S into the set S' such that

$$\varphi(x \cdot y) = \varphi(x) \star \varphi(y)$$

[1]The word *morphism* is also used by some authors.

for every $x, y \in S$. To denote such a mapping we write $\varphi : S \mapsto S'$. If φ is also a surjective mapping, then φ is called a homomorphism from S *onto* S', and S' is called the *homomorphic image* of S. In case the mapping φ above is injective, it is called a *one-to-one homomorphism*. An *isomorphism* from S to S' is a homomorphism which is both surjective and injective.

A homomorphism φ from a monoid $(M, \cdot, 1)$ to a monoid $(M', \star, 1')$ is a semigroup homomorphism $\varphi : M \mapsto M'$ such that

$$\varphi(1) = 1'.$$

The terminology for surjective and injective homomorphisms of monoids is the same as above. It will be clear from the context whether the intended meaning is "monoid homomorphism" or "semigroup homomorphism".

We shall say that a monoid N is a *quotient* of a monoid M if there exists a surjective homomorphism $\phi : M \mapsto N$.

A monoid M is said to *divide* a monoid N (written $M \prec N$) if M is a quotient of a submonoid of N.

The notions of quotient and division are defined similarly for semigroups.

Let A be a finite alphabet and let $L \subseteq A^*$. Consider the following equivalence relation $\equiv_L$ on A^*:

$$x \equiv_L y \iff \{(u, v) \in A^* \times A^* : uxv \in L\} = \{(u, v) \in A^* \times A^* : uyv \in L\}.$$

It is easy to show that if $x \equiv_L y$ and $a \in A$, then

$$xa \equiv_L ya \qquad \text{and} \qquad ax \equiv_L ay.$$

It follows that the equivalence relation $\equiv_L$ is a congruence on A^*. It is called the *syntactic congruence* of L. The quotient of A^* by $\equiv_L$, denoted $M(L)$, is the *syntactic monoid* (or *syntactic semigroup* for A^+) of L and the projection $\eta_L : A^* \mapsto M(L)$ is termed the *syntactic morphism* of L.

2-IV Equivalence of automaton and monoid

A monoid M is said to recognize $L \subseteq A^*$ if there exists a subset X of M and a homomorphism $\phi : A^* \mapsto M$ such that $L = \phi^{-1}(X)$. (We also say that the homomorphism ϕ recognizes a language L.)

We next show that the two notions of recognizable sets – by finite automata and by finite monoids – are equivalent.

Theorem 2-IV.1 (cf. [MP71]). *A subset L of A^* is regular if and only if it is recognized by a finite monoid.*

Proof: Let $L \subseteq A^*$ be a regular language and $\mathcal{A} = (Q, i, F, \lambda)$ be a deterministic finite automaton recognizing L. We define an equivalence relation $\sim$ on A^* by

$$x \sim y \iff \forall q \in Q : q \cdot x = q \cdot y.$$

The number of equivalence classes of the equivalence relation $\sim$ does not exceed $|Q|^{|Q|}$. Suppose now $x \sim y$ and $uxv \in L$ for some $u, v \in A^*$. We then derive:

$$i \cdot (uyv) = ((i \cdot u) \cdot y) \cdot v = ((i \cdot u) \cdot x) \cdot v = i \cdot (uxv) \in F.$$

Thus $uyv \in L$. A similar derivation will show that $uyv \in L$ implies $uxv \in L$. Therefore,

$$x \sim y \Rightarrow x \equiv_L y,$$

which shows that the equivalence relation $\sim$ refines $\equiv_L$, and hence $|M(L)| \leq |Q|^{|Q|}$.

Conversely, let us assume $M(L)$ is finite. First observe that if $x \in L$ and $x \equiv_L y$, then $y \in L$, because $x = 1 \cdot x \cdot 1$. We construct a deterministic finite automaton $\mathcal{T} = (Q, i, F, \lambda)$ recognizing L by setting: the set of states Q is the set of elements of $M(L)$, the initial state i is 1, the set of final states F is the set of classes of words in L and the transition function λ is given for all $a \in A$ by

$$\lambda([w], a) = [wa],$$

where $[v]$ denotes the $\equiv_L$-class of a word v. Thus a word w is accepted by $\mathcal{T}$ if and only if $1 \cdot w = [w]$ is the class of a word in L. By the observation above, this is true if

and only if $w \in L$. Therefore, $\mathcal{T}$ recognizes L; and since $M(L)$ is finite, L is regular. *Q.E.D.*

Let $\mathcal{A} = (Q, i, F, \lambda)$ be a deterministic finite automaton operating over a finite alphabet A. For each word $w \in A^*$ we define a corresponding *state-transition function* $\mu_w : Q \mapsto Q$, denoted by a two-row matrix

$$\mu_w = \begin{pmatrix} m_{1j} \\ m_{2j} \end{pmatrix},$$

where the first row m_{1j} is an (ordered) permutation of $q_j \in Q$ $(1 \leq j \leq |Q|)$ and elements of the second row are $m_{2j} = \lambda(q_j, w)$. The set of these maps under the operation of functional composition

$$\mu_v \circ \mu_u = \mu_{uv}$$

forms a monoid, termed the *transition monoid* of $\mathcal{A}$, denoted by $M(\mathcal{A})$.

Theorem 2-IV.2 (cf. [MP71]). *Let $\mathcal{A}$ be the minimal automaton of L. Then $M(\mathcal{A})$ and $M(L)$, the syntactic monoid of L, are isomorphic.*

Theorem 2-IV.3. *Let $L \subseteq A^*$ be a language and $\eta_L : A^* \mapsto M(L)$ – its syntactic morphism. Let $\phi : A^* \mapsto M$ be a homomorphism. Then:*

1. *ϕ recognizes L if and only if there exists a homomorphism $\psi : \phi(A^*) \mapsto M$ such that $\psi \circ \phi = \eta_L$ (i.e. η_L factors through ϕ).*

2. *A monoid M recognizes L if and only if $M(L) \prec M$.*

Proof: If ϕ recognizes L then there exists $X \subseteq M$ such that $L = \phi^{-1}(X)$. Suppose $\phi(w_1) = \phi(w_2)$. Then $xw_1y \in L$ implies $\phi(xw_2y) \in X$ since $\phi(xw_1y) \in X$ and $\phi(xw_1y) = \phi(xw_2y)$. Thus $xw_2y \in L$. Similarly, $xw_2y \in L \Rightarrow xw_1y \in L$. Therefore, $\phi(w_1) = \phi(w_2) \Rightarrow w_1 \equiv_L w_2$. Hence η_L factors through ϕ, and $M(L)$ is a homomorphic image of $\phi(A^*)$, proving $M(L) \prec M$.

Conversely, suppose there exists a homomorphism $\psi : \phi(A^*) \mapsto M$ such that $\psi \circ \phi = \eta_L$. If $\phi(w) \in \phi(L)$ then $\eta_L(w) \in \eta_L(L)$, whence $\phi(w) \in \phi(L) \iff w \in L$. That is, ϕ recognizes L. Let M be a monoid and $M(L) \prec M$, then there exists a

submonoid M' of M and a surjective homomorphism $\psi : M' \mapsto M$. For every $a \in A$ fix $\phi(a) \in M'$ such that $\psi(\phi(a)) = \eta_L(a)$. We then extend the domain of ϕ to A^*, i.e. ϕ is a homomorphism $\phi : A^* \mapsto M$ such that η_L factors through ϕ. Then M recognizes L since ϕ recognizes L. *Q.E.D.*

The next results apply to operations on languages.

Proposition 2-IV.4 (cf. [Arb68]). *Let L, K be two languages of A^* recognized respectively by monoids M_L and M_K and let M be a monoid. Then*

1. *if M recognizes L, M recognizes $A^* \setminus L$;*
2. *$L \cap L$ and $L \cup K$ are recognized by $M_L \times M_K$;*
3. *if M recognizes L, M recognizes $K^{-1}L$ and LK^{-1}.*

2-V Green's relations

The equivalence relations we are about to introduce were first formulated by J. A. Green in 1951 ([Gre51]) and have become fundamental in the theory of semigroups.

Definition 2-V.1. Let M be a monoid. Green's relations are defined by the following equivalences:

$$\begin{aligned} a\mathcal{R}b &\iff aM = bM & \mathcal{D} &= \mathcal{R} \vee \mathcal{L} \\ a\mathcal{L}b &\iff Ma = Mb & \mathcal{H} &= \mathcal{R} \cap \mathcal{L} \\ a\mathcal{J}b &\iff MaM = MbM \end{aligned}$$

(cf. figure 2.2)

We also introduce reflexive and transitive relations based on the above:

$$\begin{aligned} a \leq_{\mathcal{R}} b &\iff aM \subseteq bM \\ a \leq_{\mathcal{L}} b &\iff Ma \subseteq Mb \\ a \leq_{\mathcal{J}} b &\iff MaM \subseteq MbM \\ a \leq_{\mathcal{H}} b &\iff a \leq_{\mathcal{R}} b \text{ and } a \leq_{\mathcal{L}} b \end{aligned}$$

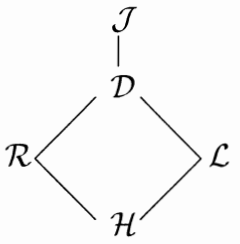

Figure 2.2: The inclusion of various Green's equivalences.

Notation. If a is an element of a monoid M, then by R_a, L_a, H_a, J_a and D_a we mean respectively the $\mathcal{R}$-class, $\mathcal{L}$-class, $\mathcal{H}$-class, $\mathcal{J}$-class and $\mathcal{D}$-class containing a.

Lemma 2-V.1 ([Gre51]). *In a finite monoid, the relations $\mathcal{R}$ and $\mathcal{L}$ commute. Consequently the relation $\mathcal{D} = \mathcal{RL} = \mathcal{LR}$ is the smallest one containing $\mathcal{R}$ and $\mathcal{L}$.*

Proposition 2-V.2 ([Gre51]). *In a finite monoid, $\mathcal{D} = \mathcal{J}$.*

Proposition 2-V.3. *Let R be an $\mathcal{R}$-class and L be an $\mathcal{L}$-class of a finite monoid M. Then $R \cap L \neq \emptyset$ if and only if R and L are within the same $\mathcal{J}$-class.*

Proof: If $a \in R \cap L$ the result is immediate: $R = R_a$ and $L = L_a$ and therefore J_a must contain both of them.

Conversely, suppose R and L are in the same $\mathcal{J}$ class of M. Then for every $x \in R$ and $y \in L$ there exists $a \in M$ such that $x\mathcal{R}a$ and $a\mathcal{L}y$ (since $x\mathcal{J}y$ and $\mathcal{J} = \mathcal{RL}$). Hence, $a \in R \cap L$. *Q.E.D.*

A $\mathcal{D}$-class (or a $\mathcal{J}$-class) of a finite monoid can thus be viewed as a table where rows represent $\mathcal{R}$-classes and columns – $\mathcal{L}$-classes. $\mathcal{H}$-classes lie at the intersections (fig 2.3). The presence of an idempotent in an $\mathcal{H}$-class is indicated by a star (*).

Lemma 2-V.4 (cf. [CP67]). *Let m be an element of a finite monoid M. If $L_m = J_m$ and L_m contains an idempotent, then L_m is a subsemigroup of M.*

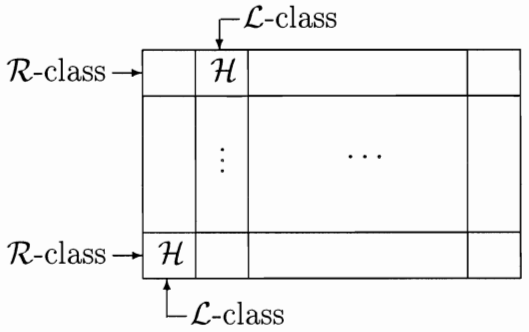

Figure 2.3: The $\mathcal{D}$-class structure.

Proof: Let $e \in L_m$ be idempotent, so $L_e = L_m = J_m = J_e$. Consider two elements t_1, t_2 of L_e: $t_1 = ue, t_2 = ve$ for some $u, v \in M$. Thus $t_1 t_2 = ueve \in Me$. On the other hand, $e = xt_1 = yt_2$ for some $x, y \in M$. Thus $e = e^2 = xt_1 yt_2$. Since $xt_1 y = xuey \in MeM$ and $e = xt_1 yt_2 \in Mxt_1 yM$, we conclude that $xt_1 y$ and e generate the same two-sided ideal of M: $J_{xt_1 y} = J_e = L_e = L_{t_1}$. Hence there exists $w \in M$ such that $xt_1 y = wt_1$. Thus $e = wt_1 t_2$ and e is in the left ideal generated by $t_1 t_2$ and $t_1 t_2$ is in the left ideal generated by e. This implies $t_1 t_2 \in L_e$ and therefore $L_e = L_m$ is closed under product. *Q.E.D.*

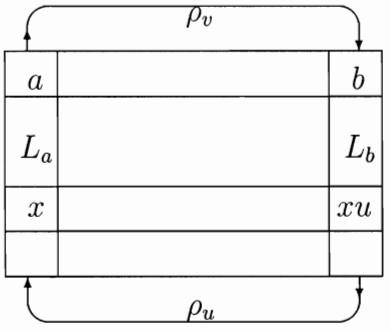

Figure 2.4: Green's Lemma.

Theorem 2-V.5 (Green's Lemma, [Gre51]). *Let $a, b \in M$ be such that $a\mathcal{R}b$. Then there exist $u, v \in M$ satisfying $au = b$ and $bv = a$. If ρ_u, ρ_v are the right translations defined respectively by $\rho_u(x) = xu$ and $\rho_v(x) = xv$, then $\rho_u : L_a \mapsto L_b$ and $\rho_v : L_b \mapsto L_a$ are inverse bijections preserving the $\mathcal{H}$-classes, i.e.*

$$\forall x, y \in L_a : x\mathcal{H}y \iff \rho_u(x)\mathcal{H}\rho_u(y)$$

and

$$\forall x, y \in L_b : x\mathcal{H}y \iff \rho_v(x)\mathcal{H}\rho_v(y)$$

Proof: (figure 2.4) Let $x\mathcal{L}a$. By definition, $Mx = Ma$; therefore $Mxu = Mau$, or $xu\mathcal{L}au = b$. Hence ρ_u is a function from L_a to L_b. Since there exist $t \in M$ such that $ta = x$ we have

$$\rho_v(\rho_u(x)) = \rho_v(xu) = \rho_v(tau) = \rho_v(tb) = tbv = ta = x,$$

i.e. the composition $\rho_u \circ \rho_v$ is the identity function on L_a. A similar argument shows ρ_v to be a function from L_b to L_a and $\rho_v \circ \rho_u$ to be the identity on L_b.

Since every $x \in L_a$ is $\mathcal{R}$-equivalent to xu and every $z \in L_b$ is $\mathcal{R}$-equivalent to zv, we conclude:

$$(x\mathcal{H}y) \Rightarrow (xu\mathcal{H}yu) \qquad \text{and} \qquad (xu\mathcal{H}yu) \Rightarrow (x = xuv\mathcal{H}yuv = y).$$

Q.E.D.

The case of two $\mathcal{L}$-equivalent elements is symmetric.

Proposition 2-V.6 ([CM56]). *If a, b are two elements of a $\mathcal{J}$-class of a monoid M, then:*

$$(ab \in R_a \cap L_b) \iff \exists e \in R_b \cap L_a : e^2 = e$$

The situation is summarized in the figure 2.5.

a	R_a	ab
L_a		L_b
$*e$	R_b	b

Figure 2.5: Proposition 2-V.6.

Proof: Suppose $ab \in R_a \cap L_b$. By Green's Lemma $\rho_b : L_a \mapsto L_b$ is a bijection. Chose an element $e \in R_b \cap L_a$ such that $\rho_b(e) = eb = b$. Since e and b are $\mathcal{R}$-equivalent there exists $u \in M$ such that $e = bu$. Then $e^2 = ebu = bu = e$, i.e. e is idempotent.

Conversely, suppose e is an idempotent element, $e \in R_b \cap L_a$. Then $e\mathcal{R}b \Rightarrow \exists u : b = eu$. Hence, $eb = eeu = eu = b$. Similarly, $e\mathcal{L}a \Rightarrow \exists v : a = ve$ whence $ae = vee = ve = a$. Also, $e\mathcal{R}b \Rightarrow a = ae\mathcal{R}ab$ and $e\mathcal{L}a \Rightarrow b = eb\mathcal{L}ab$. That is, $ab \in R_a \cap L_b$. *Q.E.D.*

Lemma 2-V.7 (cf. [CP67]). *Let x and m be elements of a finite monoid M. Then*

$$xm\mathcal{J}m \Rightarrow xm\mathcal{L}m.$$

Proof: $\mathcal{J}$-equivalence of xm and m implies the existence of $p, q \in M$ such that $m = p \cdot xm \cdot q$. Then there exists a positive integer k such that both $e = (px)^k$ and $f = q^k$ are idempotent and we have $m = (px)^k mq^k = emf$. Thus

$$m = em = (px)^{k-1}p \cdot xm,$$

so m belongs to the left ideal generated by xm. Hence, $L_{xm} = L_m$. *Q.E.D.*

Lemma 2-V.8 (cf. [Lal79]). *Let H be an $\mathcal{H}$-class of a monoid M. The following conditions are equivalent:*

1. $\exists e \in H : e^2 = e$
2. $\exists a, b \in H : ab \in H$
3. *H is a maximal group in M*

Proof: 3 $\Rightarrow$ 1. If H is a group, it contains an idempotent.

1 $\Rightarrow$ 2. $H = R_a \cap L_b = R_b \cap L_a$ and by proposition 2-V.6, $ab \in H$.

2 $\Rightarrow$ 3. By proposition 2-V.6, H must contain an idempotent e. For two arbitrary elements of H, x and y: $e \in R_x \cap L_y = R_y \cap L_x$ implies (by the same proposition) $xy \in H$. Thus H is a semigroup. Furthermore, $e\mathcal{R}x$ means there exists $u \in H$ such that $x = eu$; then $ex = eeu = eu = x$. Similarly, from $e\mathcal{L}x$ we derive $xe = x$. That is, $ex = x = xe$ and H is a monoid. Let $\rho_x : H \mapsto H$ be a bijection defined by Green's Lemma. Then there exist x' such that

$$\rho_x(x') = x'x = e,$$

which shows that H is a group. Since every element of a group containing e is $\mathcal{H}$-equivalent to e, H is a maximal group. *Q.E.D.*

Proposition 2-V.9 (cf. [Lal79]). *Two maximal subgroups of a finite monoid M contained in the same $\mathcal{J}$-class are isomorphic.*

Proof: By Lemma 2-V.8 two maximal subgroups of a finite monoid M are $\mathcal{H}$-classes H_e and H_f containing respectively idempotents e, f. Since both H_e and H_f are within the same $\mathcal{J}$-class there exists $a \in H_a$, where $H_a = R_e \cap L_f$ (Lemma 2-V.3). Then:

$$a\mathcal{R}e \Rightarrow ea = a \qquad \text{and} \qquad a\mathcal{L}f \Rightarrow (\exists a' \in M : a'a = f) \text{ and } (af = a).$$

By Green's Lemma $\rho_a(x) = xa$ is a bijection from H_e onto H_a. Similarly, by the dual version of Green's Lemma we have that $\lambda_{a'} = a'x$ is a bijection from H_a onto H_f. Therefore the composition $\rho_a \circ \lambda_{a'}$ is a bijection mapping every x in H_e onto $a'xa$ in H_f. Clearly,

$$\rho_a \circ \lambda_{a'}(e) = a'ea = a'a = f.$$

To see that $\rho_a \circ \lambda_{a'}$ is an isomorphism, we first observe that aa' is an idempotent of R_a:

$$(aa')^2 = aa'aa' = afa' = aa'.$$

Hence, for every element $x \in R_a$ we have $aa'x = x$. For arbitrary $x, y \in H_e$, the product $xy \in H_e$. Their images under $\rho_a \circ \lambda_{a'}$ exhibit the same property:

$$(a'xa)(a'ya) = a'x(aa'y)a = a'xya.$$

Q.E.D.

A $\mathcal{J}$-class is called *regular* if all its elements are regular. (We have similar definitions for regular $\mathcal{R}$, $\mathcal{L}$ and $\mathcal{H}$-classes). The next proposition further explores the structure of a regular $\mathcal{J}$-class.

Proposition 2-V.10. *Let J be a $\mathcal{J}$-class of a finite monoid M. The following are equivalent:*

1. *J is regular*
2. *J contains a regular element*
3. *every $\mathcal{L}$-class contained in J has an idempotent*
4. *every $\mathcal{R}$-class contained in J has an idempotent*
5. *J contains an idempotent*
6. *$\exists x, y \in J : xy \in J$*

Proof: $1 \Rightarrow 2$. By definition.

$2 \Rightarrow 3, 4$. Suppose a is a regular element of J. Then $a = asa \Rightarrow a\mathcal{L}sa$. Note also that sa is idempotent:

$$(sa)^2 = sasa = s(asa) = sa.$$

Similarly, $a = asa \Rightarrow a\mathcal{R}as$ and

$$(as)^2 = asas = (asa)s = as.$$

$3, 4 \Rightarrow 2$. Let e be an idempotent element of M in J. Then $a\mathcal{R}e \Rightarrow \exists u \in M$: $au = e$ and $ea = a$, whence

$$a = ea = eea = auea = asa.$$

By the same reasoning $a\mathcal{L}f$ (where f is idempotent) implies $\exists v \in M : va = f$ and $af = a$. Therefore,

$$a = af = aff = afva = ata.$$

$2 \Rightarrow 1$. Let a be a regular element of M in J and b - an element in J. Then $a\mathcal{J}b \iff \exists c \in J : a\mathcal{R}c \wedge c\mathcal{L}b$. Since a is regular, $R_a = R_c$ contains an idempotent and therefore c is regular. Also, b must be regular because $L_c = L_b$ has an idempotent.

$3, 4 \Rightarrow 5$. Obvious.

$5 \Rightarrow 2$. Same reasoning as $3, 4 \Rightarrow 2$ applies.

$5 \iff 6$. By proposition 2-V.6. *Q.E.D.*

Proposition 2-V.11. *Let M and N be two monoids and $\phi : M \mapsto N$ be a surjective homomorphism. If J_N is a regular $\mathcal{J}$-class of N, then there exist a regular $\mathcal{J}$-class J_M of M such that $\phi(J_M) = J_N$.*

Proof: Consider a $\mathcal{J}$-minimal element s in $\phi^{-1}(J_N)$ (that is an element s such that $\forall q \in \phi^{-1}(J_N) : s \leq_{\mathcal{J}} q$). Then

$$\phi(MsM) = \phi(M)\phi(s)\phi(M) = N\phi(s)N$$

and

$$N(N\phi(s)N)N \subseteq N\phi(s)N,$$

i.e. $\phi(MsM)$ is an ideal of N. Since this ideal intersects J_N, it must contain J_N entirely. If there exists an element t such that $t <_{\mathcal{J}} s$, then $t \notin \phi^{-1}(J_N)$, for s is $\mathcal{J}$-minimal. Thus if J_M is the $\mathcal{J}$-class of s then $J_N \subseteq \phi(J_M)$. On the other hand, $s\mathcal{J}r$ for some $r \in J_M$ implies $\phi(s)\mathcal{J}\phi(r)$, whence $\phi(J_M) \subseteq J_N$. Combining the latter two we obtain $\phi(J_M) = J_N$. *Q.E.D.*

A monoid M is *$\mathcal{R}$-trivial* if for two elements $a, b \in M$ we have

$$a\mathcal{R}b \Rightarrow a = b.$$

Definitions for *$\mathcal{L}$-trivial*, *$\mathcal{H}$-trivial* and *$\mathcal{J}$-trivial* monoids are similar.

Example 2-V.1 (Computing the syntactic monoid of a language). Let $A = \{a, b, c\}$ and $L = A^*abA^*$. The minimal automaton of L was presented in example 1-III.1. Figure 2.6 shows the transitions and relations defining the syntactic monoid $M(L)$ and its $\mathcal{J}$-class structure. It's easy to see that $M(L)$ is not $\mathcal{J}$-trivial, since, for instance, b and c are $\mathcal{J}$-equivalent.

w	1	2	3
ϵ	1	2	3
a	2	2	3
b	1	3	3
c	1	1	3
ab	3	3	3
ba	2	3	3

	a	b	c	ab	ba
a	a	0	c	0	0
b	ba	b	b	0	0
c	a	c	c	0	0
ab	0	0	0	0	0
ba	0	0	0	0	0

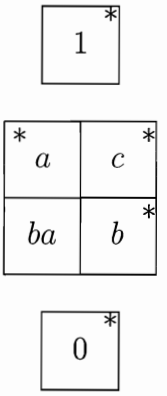

Figure 2.6: Transition relations, the syntactic monoid and $\mathcal{J}$-class structure of the language $L = A^*abA^*$ over $A = \{a, b, c\}$.

While we normally treat monoids in this thesis, certain statements require manipulating semigroups specifically. If S is a semigroup, we can adjoin a new identity element I to S and thereby obtain a monoid, denoted S^I, by setting

$$I \cdot I = I,$$

and for all $s \in S$

$$s \cdot I = I \cdot s = s.$$

We then set

$$S^1 = \begin{cases} S^I, & \text{if } S \text{ is not a monoid,} \\ S, & \text{if } S \text{ is a monoid.} \end{cases}$$

Lemma 2-V.12.

Let S be a finite semigroup. Then either of the following holds:

1. *S is cyclic, i.e. $S = \{s^k | \ k \geq 0\}$ for some $s \in S$,*
2. *S is left-simple, i.e. no proper subset of S is its left ideal,*
3. *$S = P \cup T$, where P is a proper left ideal of S and T is a proper subsemigroup of S.*

Proof: Let M denote the smallest monoid containing S as a subsemigroup, i.e. $M = S^1$. We choose an element $s \in S$ to maximize the set MsM, i.e.

$$\forall t \in S\, [MsM \subseteq MtM \;\Rightarrow\; MsM = MtM].$$

Consider the following cyclic subsemigroup of S:

$$C = \{s^k |\; k \geq 0\}.$$

If $C = S$, then S is cyclic. Otherwise, consider the set $I = S \setminus C$. If $SI \subseteq I$, then I is a proper left ideal of S and $S = I \cup C$.

We henceforth suppose this is not the case, i.e. I is not a left ideal and therefore there exists $t \neq s$ such that $s = rt$ for some $r \in S$. Since s is chosen to ensure the maximality of MsM, we have $t = m_1 s m_2$ for some $m_1, m_2 \in M$. Thus

$$s = rt = rm_1sm_2 = rm_1(\underbrace{rm_1sm_2}_{s})m_2 = \cdots = (rm_1)^k s (m_2)^k,$$

for all $k > 0$. So k can be chosen in order that $(rm_1)^k = e$ is an idempotent of S. By the maximality condition, $e \in MsM$ and hence e and s are $\mathcal{J}$-equivalent. Thus J_s contains an idempotent e and there are several cases to consider.

Case 1. If $L_s = J_s$ we have the condition of Lemma 2-V.4 and L_s is a subsemigroup of S.

> Case 1a. If $L_s = S$ then no proper subset of S is its left ideal, i.e. S is left-simple.
>
> Case 1b. Otherwise, we claim that $P = S \setminus L_s$ is a left ideal. Suppose the contrary. Then $s = xq$ for some $x \in S$, $q \notin L_s$ and we arrive at a contradiction since by Lemma 2-V.7, $q \in J_s = L_s$. Therefore in this case we take $P = S \setminus L_s$ and $T = L_s$.

Case 2. $L_s \subset J_s$. Consider the set

$$W = S \setminus J_s = \{w |\; s \notin MwM\}.$$

We claim that $W \cup L_s$ is a left ideal of S. Here is why. First, for all $x \in S$ and $w \in W$ we have $xw \in W$, i.e. W is a left ideal of S. By Lemma 2-II.1, ideals are closed under union.

This claim can also be shown without the aforementioned Lemma. If $t \in L_s$ then $t = ys$ for some $y \in S$ and $xt = xys$. If $xys \in J_s$, then by Lemma 2-V.7, $xt = xys \in L_s$. If $xys \notin J_s$, then $xys \in W$, by the definition of W. Hence $xt \in W \cup L_s$ for all $t \in W \cup L_s$ and $W \cup L_s$ is a left ideal of S.

In a similar way we can prove that $(J_s \setminus L_s) \cup W$ is a left ideal of S. Assuming the contrary means there exists $t \in J_s \setminus L_s$ such that $xt \in L_s$. Again, by Lemma 2-V.7, xt and t are $\mathcal{L}$-equivalent, so $t \in L_s$, which contradicts our assumption.

We thus take $P = (J_s \setminus L_s) \cup W$ and $T = L_s \cup W$. *Q.E.D.*

Chapter 3

Variety

3-I Introduction

To each regular language corresponds a finite monoid – its syntactic monoid. Naturally one may attempt classification of regular languages according to the algebraic properties of their syntactic monoids. In this chapter we introduce the proper framework to formalize this idea.

3-II Identities of finite monoids

A *variety of monoids* [1] is a class of monoids closed under the operations of taking submonoids, quotients and finite products. A class $\mathbf{V}$ of monoids is a variety if $\mathbf{V}$ satisfies the following:

1. if $M \in \mathbf{V}$ and N is a submonoid of M, then $N \in \mathbf{V}$
2. if $M \in \mathbf{V}$ and N is a quotient of M, then $N \in \mathbf{V}$
3. if $(M_i)_{i \in I}$ is a finite family of elements of $\mathbf{V}$, then $\prod_{i \in I} M_i \in \mathbf{V}$

The definition of variety of semigroups is similar, with the word "subsemigroup" replacing "submonoid". Varieties of semigroups and monoids will be denoted by

[1] Varieties of finite monoids are also referred to as *pseudovarieties*.

boldface capital letters, like $\mathbf{V}$.

Let $u, v \in A^*$. A finite monoid M *separates* u and v if there exists a monoid homomorphism $\phi : A^* \mapsto M$ such that $\phi(u) \neq \phi(v)$. We define a distance on A^* as follows. If u, v are two words of A^*, let

$$r(u, v) = \min\{\ |M|\ |M \text{ separates } u \text{ and } v\}$$

and

$$d(u, v) = 2^{-r(u,v)}.$$

By convention, $\min(|\emptyset|) = -\infty$ and $2^{-\infty} = 0$.

Below are some properties of $d(u, v)$ for all $u, v, w, \in A^*$.

1. $d(u, v) = 0 \iff u = v$,
2. $d(u, v) = d(v, u)$,
3. $d(u, v) \leq \max(d(u, w),\ d(v, w))$,
4. $d(uw, vw) \leq d(u, v)$ and $d(wu, wv) \leq d(u, v)$.

That is, $d(u, v)$ is an ultrametric distance function. For this metric, multiplication in A^* is uniformly continuous. The completion of the metric space (A^*, d), denoted $\widehat{A^*}$, is called the *free profinite monoid* on A.

We consider each finite monoid M as being equipped with a discrete distance, defined for all $x, y \in M$ by

$$d(x, y) = \begin{cases} 0, & \text{if} \quad x = y \\ 1, & \text{if} \quad x \neq y. \end{cases}$$

Let M be a finite monoid. Then a map $\phi : \widehat{A^*} \mapsto M$ is *continuous* if and only if, for every converging sequence $(u_n)_{n\geq 0}$ of $\widehat{A^*}$, the sequence $\phi(u_n)_{n\geq 0}$ is ultimately constant, i.e. if there exists an integer n_0 such that, for all $n, m \geq n_0$, $\phi(u_n) = \phi(u_m)$.

Let $x, y \in \widehat{A^*}$. We say that a finite monoid M *satisfies the identity* $x = y$ if, for every continuous homomorphism $\phi : \widehat{A^*} \mapsto M$,

$$\phi(x) = \phi(y).$$

Note that such a continuous homomorphism is entirely determined by the values of $\phi(a)$, for $a \in A$. Any map $\psi : A \mapsto M$ can be extended in a unique way into a monoid homomorphism $\phi : A^* \mapsto M$. Since M is finite, such a homomorphism is uniformly continuous: if two elements of A^* cannot be separated by M, their images under ϕ must be the same. Then every monoid homomorphism $\phi : A^* \mapsto M$ can be extended in a unique way into a continuous homomorphism from $\widehat{A^*}$ onto M. Since $\widehat{A^*}$ is a completion of A^*, its elements are words and limits of sequences of words. The ω-power, whose definition relies on the following lemma, is an example of such a limit.

Lemma 3-II.1 (cf. [Rei82]). *Let A be a finite alphabet and $x \in \widehat{A^*}$. The sequence $(x^{n!})_{n\geq 0}$ converges in $\widehat{A^*}$ to an idempotent, denoted x^ω.*

Given a set E of identities, we denote by $[\![E]\!]$ the class of all finite monoids which satisfy all the identities of E. The fundamental theorem below, an extension of an earlier result due to Birkhoff [Bir35], states that

Theorem 3-II.2 ([Rei82]).
A class of finite monoids is a variety if and only if it can be defined by a set of identities of $\widehat{A^*}$.

Example 3-II.1.

$[\![x = y]\!]$	defines the variety of finite monoids containing only the trivial monoid.
$[\![x = x]\!]$	defines the variety of all finite monoids.
$[\![xy = yx]\!]$	defines the variety of finite commutative monoids.
$[\![x^2 = x]\!]$	defines the variety of finite idempotent monoids.

Example 3-II.2. We say that a semigroup S is *locally trivial* if for every idempotent e of S and for every element $s \in S$, we have $ese = e$. It is not difficult to see that locally trivial semigroups are closed under taking subsemigroups, quotients and finite products and therefore form a variety of semigroups, denoted **LI**. It follows from the definition that the variety **LI** is defined by the identity $[\![x^\omega y x^\omega = x^\omega]\!]$.

3-III The variety theorem

If $\mathbf{V}$ is a variety of finite monoids and A is an alphabet, we denote by $A^*\mathcal{V}$ the set of (regular) languages of A^* whose syntactic monoid belongs to $\mathbf{V}$ or, equivalently, the set of languages of A^* recognized by the monoids of $\mathbf{V}$. The correspondence $\mathbf{V} \mapsto \mathcal{V}$ associates with each variety of finite monoids a class of regular languages. We have the following theorem due to Eilenberg.

Theorem 3-III.1 ([Eil76]). *The correspondence $\mathbf{V} \mapsto \mathcal{V}$ defines a one-to-one correspondence between the varieties of finite monoids and the varieties of languages.*

When a variety $\mathbf{V}$ is generated by a single monoid, we have a direct description of the corresponding languages.

Theorem 3-III.2 (cf. [Eil76]).
Let $\mathbf{V} = (M)$ be the variety of monoids generated by a monoid M and $\mathcal{V}$ – be the corresponding variety of languages. Then for every alphabet A, $A^*\mathcal{V}$ is the boolean algebra generated by the languages of the form $\phi^{-1}(m)$ where $\phi : A^* \mapsto M$ is an arbitrary homomorphism and $m \in M$.

3-IV Varieties defined by Green's relations

Recall that a monoid M is aperiodic if for every $s \in M$, there exists an integer n such that $s^n = s^{n+1}$. We begin with various characterizations of aperiodic monoids.

Proposition 3-IV.1. *Let S be a finite monoid. The following conditions are equivalent:*

1. $\forall x \in S(\exists n \in \mathbb{N} : x^n = x^{n+1})$, *i.e. S is aperiodic.*
2. $\exists m \in \mathbb{N}(\forall x \in S : x^m = x^{m+1})$.
3. *If G is a group in S, G is trivial.*

4. S is $\mathcal{H}$-trivial.

Proof: $1 \Rightarrow 2$. For every element $x \in S$ denote by n_x the smallest integral exponent in $x^{n_x} = x^{n_x+1}$ and let

$$m = \max_{x \in S} n_x.$$

Then for all $x \in S$, $x^m = x^{m+1}$.

$2 \Rightarrow 3$. Let x be an element of a group G in S. Then there exist an integer $k > 0$ such that $x^k = 1$. It follows that

$$x = x\left(\left(x^k\right)^m\right) = x\left(x^{km}\right) = x\left(x^{km+1}\right) = x^2\left(x^{km}\right) = x^2.$$

That is, every element of G is idempotent and G is trivial.

$3 \Rightarrow 4$. Let H be an $\mathcal{H}$-class of S and $x, y \in H$. Then

$$(x\mathcal{H}y) \iff \left\{ \begin{array}{rcl} x\mathcal{L}y & \iff & \exists a, b \in S(ax = y \wedge by = x) \\ x\mathcal{R}y & \iff & \exists c, d \in S(xc = y \wedge yd = x) \end{array} \right\} \iff (x = axd)$$

Therefore, $\forall n > 0 : x = a^n x d^n$. By Green's Lemma and Corollary 2-V.6, a regular $\mathcal{H}$-class is in one-to-one correspondence with a group in S. By assumption the groups in S are trivial and hence there exist $m > 0$ such that $a^m = a^{m+1}$. We thus can deduce

$$y = ax = aa^m x d^m = a^{m+1} x d^m = a^m x d^m = x.$$

That is, $x\mathcal{H}y \Rightarrow x = y$ and S is $\mathcal{H}$-trivial.

$4 \Rightarrow 1$. Suppose S is $\mathcal{H}$-trivial and $s \in S$. Let T be the subsemigroup generated by s. We showed in section 2 that $T = \{s, s^2, \ldots, s^{p+q-1}\}$ with $s^p = s^{p+q}$ and $G = \{s^p, s^{p+1}, \ldots, s^{p+q-1}\}$ is a maximal subgroup in T. All elements of G are $\mathcal{H}$-equivalent in T and therefore in S. By assumption, S is $\mathcal{H}$-trivial and hence G is trivial and $q = 1$, whence $s^p = s^{p+1}$. *Q.E.D.*

Notation. **R**, **L** and **J** denote respectively the varieties of $\mathcal{R}$-trivial, $\mathcal{L}$-trivial and $\mathcal{J}$-trivial monoids.

Proposition 3-IV.2.

1. $\mathbf{R} = [\![(xy)^\omega x = (xy)^\omega]\!]$.

2. $\mathbf{L} = [\![y(xy)^\omega = (xy)^\omega]\!]$.

3. $\mathbf{J} = [\![(xy)^\omega x = (xy)^\omega = y(xy)^\omega]\!]$ *and also*
   $\mathbf{J} = [\![(xy)^\omega = (yx)^\omega,\ x^\omega = x^{\omega+1}]\!]$.

Proof: 1. If a finite monoid M is $\mathcal{R}$-trivial, it is aperiodic. (To see this, let $c \in M$ and let m be an integer such that c^m is idempotent. We then have $(c^m)\mathcal{R}(c^m)c$, whence $c^m = c^{m+1}$.) Let x, y be two arbitrary elements of M. Then $(xy)^n = (xy)^{n+1} = ((xy)^n x)y$, that is $(xy)^n x \mathcal{R} (xy)^n$. Since M is $\mathcal{R}$-trivial, $(xy)^n x = (xy)^n$.

Conversely, suppose M is a finite monoid satisfying $(xy)^n x = (xy)^n$ and let e be an idempotent element of M and $x \in M$ such that $e\mathcal{R}x$. Then:

$$\begin{aligned} x &= ex && \text{from } e\mathcal{R}x \\ &= xyx && \text{from } e\mathcal{R}x \Rightarrow \exists y \in M \text{ such that } xy = e \\ &= (xy)^n x && e \text{ is idempotent of } M \\ &= (xy)^n && \text{by assumption} \\ &= e \end{aligned}$$

Thus $e\mathcal{R}x$ implies $x = e$ and M is $\mathcal{R}$-trivial.

2. The proof is similar.

3. Since $\mathbf{J} = \mathbf{R} \cap \mathbf{L}$ we have

$$(xy)^n x = (xy)^n = y(xy)^n.$$

Assuming $y = 1$ leads to $x^n = x^{n+1}$ and to $(xy)^n = y(xy)^n = (yx)^n y = (yx)^n$.

Conversely, if a variety is ultimately defined by the equations $x^n = x^{n+1}$ and $(xy)^n = (yx)^n$ then

$$(xy)^n = (xy)^{n+1} = (yx)^{n+1} = y(yx)^n x = y(xy)^n x \tag{3.1}$$

Continuing in the same manner we obtain: $(xy)^n = y^n(xy)^n x^n$, and then $y^n(xy)^n x^n = y^{n+1}(xy)^n x^n = yy^n(xy)^n x^n$. Now applying 3.1 from right to left n times, we arrive at $(xy)^n = y(xy)^n$. A similar derivation yields $(xy)^n = (xy)^n x$. *Q.E.D.*

We denote by **DA** the class of aperiodic monoids whose every regular $\mathcal{J}$-class is an idempotent monoid.

Lemma 3-IV.3. *Let M be a submonoid of a monoid N and J – a regular $\mathcal{J}$-class of M. The restrictions to J of Green's relations in M and N coincide.*

Proof: Let $\mathcal{R}_M$ and $\mathcal{R}_N$ denote Green's relations $\mathcal{R}$ in M and N respectively. Suppose x and y are elements of J such that $x\mathcal{R}_N y$. Since J is a regular $\mathcal{J}$-class of M, there exist idempotents e, f such that $e\mathcal{R}_M x$ and $f\mathcal{R}_M y$. Hence, $e\mathcal{R}_N x\mathcal{R}_N y\mathcal{R}_N f$ and therefore $ef = f$ and $fe = e$. Consequently, $e\mathcal{R}_M f$ and $x\mathcal{R}_M y$. The proof for other Green's relations is similar. *Q.E.D.*

Lemma 3-IV.4. *If a regular $\mathcal{J}$-class is a monoid, it is a simple monoid.*

Proof: Let M be a monoid and J – a regular $\mathcal{J}$-class which is a submonoid of M. Since the restrictions to J of Green's relations in M and J coincide (by Lemma 3-IV.3), we have in particular, $JaJ = JbJ = J$ for every $a, b \in J$. Thus the only ideals of J are the empty set and J itself. *Q.E.D.*

Proposition 3-IV.5. DA *is a variety of monoids.*

Proof: We need to show that **DA** is closed under the operations of taking submonoid, quotient and finite product.

Let $S \in \mathbf{DA}$ be a monoid, let T be a submonoid of S and J – a regular $\mathcal{J}$-class of T. If a is an element of J, then $a\mathcal{L}_T e$ for some idempotent $e \in T$ and therefore $a\mathcal{L}_S e$. Since $S \in \mathbf{DA}$, the $\mathcal{J}$-class of e contains only idempotents and hence a is an idempotent of T.

Let S be a monoid, $S \in \mathbf{DA}$, and let T be a quotient of S. Then there exists a surjective homomorphism $\phi : S \mapsto T$. If J_T is a regular $\mathcal{J}$-class of T then by proposition 2-V.11 there exists a regular $\mathcal{J}$-class J_S of S such that $\phi(J_S) = J_T$. Since by assumption J_S only contains idempotents, so does J_T and therefore $T \in \mathbf{DA}$.

Suppose S and T are two monoids, $S, T \in \mathbf{DA}$. The $\mathcal{J}$-classes of the product $S \times T$ are of the form $J_S \times J_T$. Therefore if J_S, J_T are idempotent monoids of S

and T respectively, then a $\mathcal{J}$-class of $S \times T$ is an idempotent monoid and hence, $S \times T \in \mathbf{DA}$. *Q.E.D.*

Proposition 3-IV.6.

$$\mathbf{DA} = [\![(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega,\ x^\omega = x^{\omega+1}]\!].$$

Proof: Let $M \in \mathbf{DA}$ be a monoid. The two idempotents $e = (xy)^\omega$, $f = (yx)^\omega$ are in the same regular $\mathcal{J}$-class J of M (Proposition 2-V.6). Since J is a simple monoid (by Lemma 3-IV.4), the product efe appears in the same $\mathcal{R}$-class and in the same $\mathcal{L}$-class as e, that is in a group where e is an idempotent. This group is an $\mathcal{H}$-class of J and it consists of only one element - e itself (J is an idempotent monoid). Therefore $e = efe$, which establishes the first identity. The second equation follows directly from the fact that J is idempotent.

Conversely, suppose a monoid M satisfies the identities of the proposition. Let x be an element of a $\mathcal{J}$-class J of M. By hypothesis J contains idempotents. Then there exists $y \in J$ such that $xy = e$ and $yx = f$, where e and f are idempotent elements of J. It follows from the first equation that the product $xy \cdot yx \cdot xy = xy$ is in J, i.e. J is closed under multiplication. By the second equation, $x^2 = x$, which means J is a union of trivial groups and an idempotent monoid. *Q.E.D.*

3-V Variety and formal logic

In this section we build a connection between boolean operations and varieties of regular languages defined by formulæ of formal logic. For an expression ϕ with free variables in a set $\mathcal{V}$, $M(\phi)$ and η_ϕ denote, respectively, the syntactic monoid and the syntactic morphism of $L(\phi) \subseteq (A \times 2^{\mathcal{V}})^*$. By θ_ϕ we mean the restriction of η_ϕ to A^* and by $N(\phi)$ – the image of this restriction.

Proposition 3-V.1. *Let $L(\phi)$ and $L(\psi)$ be regular languages over the alphabet $A \times 2^{\mathcal{V}}$ defined by the formulæ ϕ and ψ. Let $\mathbf{V}$ be a variety of finite monoids. Then*

(1.a) $M(\phi), M(\psi) \in \mathbf{V} \Rightarrow M(\phi \wedge \psi) \in \mathbf{V}$.

(1.b) $N(\phi), N(\psi) \in \mathbf{V} \Rightarrow N(\phi \wedge \psi) \in \mathbf{V}$.

(2.a) $N(\phi) \in \mathbf{V} \Rightarrow N(\neg\phi) \in \mathbf{V}$.

(2.b) If $\mathbf{V}$ *contains all the commutative aperiodic monoids, then*

$$M(\phi) \in \mathbf{V} \Rightarrow M(\neg\phi) \in \mathbf{V}.$$

Proof: (1.a) We define a homomorphism

$$(\eta_\phi,\ \eta_\psi) : (A \times 2^{\mathcal{V}})^* \mapsto M(\phi) \times M(\psi)$$

given by

$$\forall w \in (A \times 2^{\mathcal{V}})^* :\ (\eta_\phi,\ \eta_\psi)(w) = (\eta_\phi(w),\ \eta_\psi(w)).$$

Let $u, v \in (A \times 2^{\mathcal{V}})^*$ be such that $(\eta_\phi,\ \eta_\psi)(u) = (\eta_\phi,\ \eta_\psi)(v)$. Then, for some $z_1, z_2 \in (A \times 2^{\mathcal{V}})^*$,

$$\begin{aligned} z_1 u z_2 \models (\phi \wedge \psi) &\iff z_1 u z_2 \models (\phi) \text{ and } z_1 u z_2 \models (\psi) \\ &\iff z_1 v z_2 \models (\phi) \text{ and } z_1 v z_2 \models (\psi) \\ &\iff z_1 v z_2 \models (\phi \wedge \psi). \end{aligned}$$

Therefore,

$$\eta_{\phi\wedge\psi}(u) = \eta_{\phi\wedge\psi}(v)$$

and hence,

$$M(\phi \wedge \psi) \prec M(\phi) \times M(\psi).$$

Thus, if $M(\phi), M(\psi) \in \mathbf{V}$ then $M(\phi \wedge \psi) \in \mathbf{V}$.

(1.b) The proof is similar to the one above with N substituted for M, θ_ϕ for η_ϕ and θ_ψ for η_ψ.

(2.a) Let $u, v \in A^*$ be such that $\theta_\phi(u) = \theta_\phi(v)$. We embed A into $A \times 2^{\mathcal{V}}$ by identifying $a \in A$ with $(a, \emptyset)$. Thus, if $z_1 u z_2 \models \neg\phi$ then $z_1 v z_2$ is a $\mathcal{V}$-structure and

the case $z_1vz_2 \models \phi$ is excluded as it implies $z_1uz_2 \models \phi$. Therefore, $z_1vz_2 \models \neg\phi$. A similar argument leads to

$$z_1vz_2 \models \neg\phi \Rightarrow z_1uz_2 \models \neg\phi.$$

Hence,

$$\theta_{\neg\phi}(u) = \theta_{\neg\phi}(v).$$

Since $(\neg\neg\phi) \equiv \phi$, we conclude $\theta_\phi = \theta_{\neg\phi}$ and if $N(\phi) \in \mathbf{V}$ then $N(\neg\phi) \in \mathbf{V}$.

(2.*b*) Let $\mathcal{L}$ denote the set of all $\mathcal{V}$-structures. Suppose $w, w' \in (A \times 2^{\mathcal{V}})^*$ are such that $\eta_\phi(w) = \eta_\phi(w')$ and $\eta_{\mathcal{L}}(w) = \eta_{\mathcal{L}}(w')$. If

$$z_1wz_2 \models \neg\phi$$

then $z_1w'z_2$ cannot satisfy ϕ and since $z_1w'z_2$ is a $\mathcal{V}$-structure,

$$z_1w'z_2 \models \neg\phi.$$

Similarly, we conclude

$$z_1w'z_2 \models \neg\phi \Rightarrow z_1wz_2 \models \neg\phi.$$

Hence, $\eta_{\neg\phi}$ factors through $(\eta_\phi,\ \eta_{\mathcal{L}})$. By assumption, $\mathbf{V}$ contains all commutative aperiodic monoids. Thus we have: $M(\phi) \in \mathbf{V}$, $M(\mathcal{L}) \in \mathbf{V}$ and $\eta_{\neg\phi}$ factors through $(\eta_\phi,\ \eta_{\mathcal{L}})$. Therefore, $M(\neg\phi) \in \mathbf{V}$. *Q.E.D.*

Chapter 4

The Krohn-Rhodes Decomposition

4-I Introduction

In this chapter we define transformation semigroups and two operations on finite monoids: wreath product and block product. We then present a fundamental theorem due to Krohn and Rhodes which states that any finite monoid can be decomposed into "smaller" components.

4-II Transformation Semigroups

Given a finite set Q, by *transformation* of Q we mean a map $s : Q \mapsto Q$. We write qs or $q \cdot s$ for the image of $q \in Q$ under a transformation s. If s and t are transformations, then for all $q \in Q$

$$(qs)t = q(st),$$

i.e. transformations are composed from left to right.

A *transformation semigroup* (abbreviated *ts*)

$$X = (Q, S)$$

consists of a finite set Q and a semigroup S which is a (sub)set of transformations of Q closed under the operation of composition of transformations. The set Q is called the set of *states* of the ts X and S is called the *underlying semigroup* of the ts X.

By X^1 we denote the ts obtained by adjoining to S the identity transformation 1_Q on Q. $\overline{X}$ denotes the ts obtained by adjoining to S all the *constant* transformations on Q. That is, for each $q \in Q$ we adjoin the transformation c_q defined by

$$p \cdot c_q = q,$$

for all $p \in Q$. We observe that for all $s \in S$, $sc_q = c_q$ and $c_q s = c_{qs}$, so the adjunction of these transformations indeed yields a new ts. Note also that these operations on transformation semigroups are idempotent:

$$(X^1)^1 = X \qquad \text{and} \qquad \overline{\overline{X}} = \overline{X},$$

for all tss X.

Let $X = (P, S)$ and $Y = (Q, T)$ be tss. We say X *divides* Y, written $X \prec Y$, if there exists a subset Q' of Q and a surjective map $\psi : Q' \mapsto P$, such that for each $s \in S$ there is $\widehat{s} \in T$ satisfying for all $q \in Q'$:

$$q\widehat{s} \in Q' \qquad \text{and} \qquad \psi(q\widehat{s}) = \psi(q)s.$$

The two notions of division – one for semigroups, the other for transformation semigroups – are related, as the following lemma attests.

Lemma 4-II.1.

If $(P, S) \prec (Q, T)$, *then* $S \prec T$.

Proof: We need to show that there exists a surjective homomorphism ϕ from S onto a subsemigroup of T. Let $Q' \subseteq Q$ and $\psi : Q' \mapsto P$ be as in the definition of division of tss. Let T' be a subsemigroup of T generated by the set $\{\widehat{s} |\ s \in S\}$. Consider a map $\phi : T' \mapsto S$ given by

$$\phi(\widehat{s_1} \cdots \widehat{s_r}) = s_i \cdots s_r.$$

We first show that ϕ is well-defined. Suppose

$$\widehat{s_1} \cdots \widehat{s_r} = \widehat{t_1} \cdots \widehat{t_m}.$$

If p is a transformation in P then there exists a transformation q in Q' such that $\psi(q) = p$. Since $(P, S) \prec (Q, T)$ we have

$$\begin{aligned} ps_i \cdots s_r &= \psi(q\widehat{s_1} \cdots \widehat{s_r}) && \text{from the definition of division of tss} \\ &= \psi(\widehat{qt_1} \cdots \widehat{t_m}) && \text{by assumption} \\ &= pt_1 \cdots t_m && \text{from the definition of division of tss.} \end{aligned}$$

Thus

$$s_1 \cdots s_r = t_1 \cdots t_m,$$

which establishes that ϕ is well defined. It follows immediately that ϕ is a surjective homomorphism and hence $S \prec T$. *Q.E.D.*

4-III Wreath Product

Let S and T be semigroups. We shall write the operation on S additively; in particular, if S is a monoid, its identity will be denoted by 0, and if S is a group then the inverse of the element s will be written $-s$. This is done to provide a more readable notation, but it is **not** meant to suggest that S is commutative. A *left action* of T on S is a map $(t, s) \mapsto ts$ from $T^1 \times S$ into S such that for all $s, s' \in S$ and for all $t, t' \in T$:

$$(tt')s = t(t's) \tag{4.1}$$

$$t(s + s') = ts + ts' \tag{4.2}$$

$$1s = s \tag{4.3}$$

The action is called *monoidal* if S and T are monoids and for all $t \in T$ we have:

$$t0 = 0 \tag{4.4}$$

The *right action* of T on S is defined dually. The left action and the right action of T on S are *compatible* if for all $t, t' \in T$ and for all $s \in S$

$$(ts)t' = t(st') \tag{4.5}$$

The *semidirect product* of S and T with respect to such a left action is the semigroup $S * T$ defined on the set $S \times T$ by

$$(s,t)(s',t') = (s + ts', tt') \tag{4.6}$$

Given a pair of compatible (left and right) actions of T on S, the *bilateral semidirect product* $S**T$ is defined to be the set $S \times T$ with the multiplication given by

$$(s,t)(s',t') = (st' + ts', tt') \tag{4.7}$$

Lemma 4-III.1. *The bilateral semidirect product $S**T$ is a semigroup. If S and T are monoids, and the underlying left and right actions are monoidal, then $S**T$ is a monoid.*

Proof: The product is associative:

$$\begin{aligned}((s_1,t_1)(s_2,t_2))(s_3,t_3) &= (s_1t_2 + t_1s_2, t_1t_2)(s_3,t_3)\\ &= (s_1t_2t_3 + t_1s_2t_3 + t_1t_2s_3, t_1t_2t_3)\\ &= (s_1,t_1)(s_2t_3 + t_2s_3, t_2t_3)\\ &= (s_1,t_1)((s_2,t_2)(s_3,t_3)).\end{aligned}$$

For the monoidal actions we have:

$$\begin{aligned}(0,1)(s,t) &= (0t + 1s, 1t) && \text{by 4.7}\\ &= (0 + s, t) && \text{by 4.3}\\ &= (s,t) && \text{since } 0 + s = s\\ &= (s1 + t0, t1) && \text{by 4.3 and 4.4}\\ &= (s,t)(0,1) && \text{by 4.7}\end{aligned}$$

That is, $S**T$ is a monoid with $(0,1)$ as the identity element. *Q.E.D.*

Lemma 4-III.2. *Let S and T be finite groups and the underlying action of T on S – monoidal. Then $S**T$ is a group.*

Proof: By Lemma 4-III.1, $S**T$ is a monoid with the identity $(0,1)$. We thus have:

$$\begin{aligned}(s,t)(s',t') = (0,1) &\iff (st' + ts', tt') = (0,1)\\ &\iff \begin{cases} tt' &= 1\\ st' + ts' &= 0 \end{cases}\\ &\iff \begin{cases} t' &= t^{-1}\\ st' &= -ts' \end{cases}\end{aligned}$$

Observe that if $ts_1 = ts_2$ then

$$s_1 = t^{-1}(ts_1) = t^{-1}(ts_2) = (t^{-1}t)s_2 = 1s_2 = s_2,$$

whence for every $t \in T$ the map $s \mapsto ts$ is one-to-one and in particular, surjective. Thus, every element of the bilateral semidirect product $S**T$ has a unique right inverse, which implies that $S**T$ is a group. *Q.E.D.*

Let G be a finite group and H a subgroup of G. Recall that G/H denotes the set of all right cosets Hg, with $g \in G$.

Lemma 4-III.3. *Let G be a group contained in a bilateral semidirect product $S**T$ of finite semigroups. Then there is a normal subgroup H of G such that H is isomorphic to a group contained in S and G/H is isomorphic to a group contained in T.*

Proof: Consider the surjective homomorphism $\pi : S**T \mapsto T$. Its restriction to G maps G onto a group contained in T. It remains to show that the kernel H of this restriction is isomorphic to a group in S. Let (f,e) be the identity of G. It follows from 4.7 that e is idempotent. We have

$$H = \{(s,e) |\ (s,e) \in G\}.$$

Let $\psi : H \mapsto S$ be a function given by $\psi(s,e) = ese$. Then

$$\begin{aligned}
\psi((s,e)(s',e)) &= \psi(se + es', ee) && \text{by 4.7} \\
&= \psi(se + es', e) && \text{since } e \text{ is idempotent} \\
&= e(se + es')e && \text{by the definition of } \psi \\
&= (ese + es')e && \text{by 4.2} \\
&= ese + es'e && \text{by 4.2 for right action} \\
&= \psi(s,e) + \psi(s',e) && \text{by the definition of } \psi.
\end{aligned}$$

Thus ψ is a homomorphism. We need to show that ψ is one-to-one. If (s,e) is in the kernel K of ψ, then $ese = efe$ and

$$\begin{aligned}
(s,e) &= (f,e)(s,e)(f,e) && \text{since } (s,e) \in K \\
&= (fe + ese + ef,\ e) && \text{by 4.7} \\
&= (fe + efe + ef,\ e) && \text{since } (s,e) \in K \\
&= (f,e)(f,e)(f,e) && \text{by 4.7} \\
&= (f,e) && \text{since } (f,e) \text{ is the identity of } G,
\end{aligned}$$

which means $s = f$. Hence K is trivial and ϕ is an isomorphism. Therefore H is isomorphic to a group contained in S. *Q.E.D.*

The *wreath product* $S \circ T$ is the semidirect product $S^{T^1} * T$ defined by the action of T on S^{T^1} given by

$$tf(t') = f(tt')$$

for $f : T^1 \mapsto S$ and $t, t' \in T$. The multiplication in $S \circ T$ is given by

$$(f_1, t_1)(f_2, t_2) = (f_1(t) + f_2(t_1 t), t_1 t_2) \tag{4.8}$$

for all $t \in T^1$.

Wreath product can also be defined in terms of transformation semigroups. Let $X = (P, S)$ and $Y = (Q, T)$ be transformation semigroups. Then the wreath product of Y and X, $Y \circ X$, is a new ts:

its set of states is $Q \times P$ and

the underlying semigroup is the set of transformations of the form (F, s), where $s \in S$ and $F : P \mapsto T$, whose action is defined for all $(q, p) \in Q \times P$ by

$$(q, p)(F, s) = (q \cdot F(p),\ ps).$$

To see that the set of the above transformations of $Q \times P$ is closed under composition, consider the application

$$\begin{aligned}((q, p)(F_1, s_1))(F_2, s_2) &= (q \cdot F_1(p),\ ps_1)(F_2, s_2) \\ &= (q \cdot F_1(p) \cdot F_2(ps_1),\ ps_1 s_2) \\ &= (q, p)(\underbrace{F_1(p)F_2(ps_1)}_{G},\ s_1 s_2) \\ &= (q, p)(G,\ s_1 s_2),\end{aligned}$$

where $G(r)$ is such that for all $r \in P$, $G(r) = F_1(r)F_2(rs_1)$.

We also observe that wreath product is associative when it comes to transformation semigroups, since $Z \circ (Y \circ X)$ contains exactly the same transformations on the set of states $Q_Z \times Q_Y \times Q_X$ as $(Z \circ Y) \circ X$.

Lemma 4-III.4. *Let $X = (P, S)$ and $Y = (Q, T)$ be transformation semigroups. Then*

$$\begin{aligned}\overline{Y \circ X} &\prec \overline{Y} \circ \overline{X}, \\ (Y \circ X)^1 &\prec Y^1 \circ X^1.\end{aligned}$$

Proof: Since the set of states in each of these four tss: $\overline{Y \circ X}$, $\overline{Y} \circ \overline{X}$, $(Y \circ X)^1$ and $Y^1 \circ X^1$ is $Q \times P$, we need only to show that the underlying semigroups of the left-hand sides of both formulæ are contained in the underlying semigroups of the corresponding right-hand sides.

If $(q, p) \in Q \times P$ then for all $p \in P$, $F(p) = c_q$ and

$$c_{(q,p)} = (c_q,\ c_p) = (F,\ c_p),$$

which proves containment for the first expression. As for the second one, we have

$$1_{Q\times P} = (1_Q,\ 1_P) = (G,\ 1_P),$$

where $G(p) = 1_Q$ for all $p \in P$. *Q.E.D.*

4-IV Krohn-Rhodes Theorem

The notion of wreath product permits decomposition of semigroups into "smaller pieces". In this section we further develop tools necessary for such decomposition. The proof of the Krohn-Rhodes Theorem is from [Eil76] and [Arb68].

Lemma 4-IV.1. *Let G be a finite group and N a normal subgroup of G. Then*

$$(G,\ G) \prec (N,\ N) \circ (G/N,\ G/N).$$

Proof: Recall that in a normal subgroup N of G right and left cosets coincide and hence the product of two right cosets is a right coset (see Section 2-II):

$$Ng_i \cdot Ng_j = NNg_ig_j = Ng_k,$$

for some $g_i,\ g_j,\ g_ig_j = g_k \in G$. If two cosets have a common element, they coincide, otherwise they are disjoint. Let

$$R = \{g_1, g_2, \ldots, g_r\}$$

be a set of representatives of the cosets of N in G. We define a product $\otimes$ on R by setting $g_i \otimes g_j$ to be the representative of g_ig_j. Thus $(R, \otimes)$ is a group identical to G/N. Now consider a map

$$\psi : N \times (G/N) \mapsto G$$

given by

$$\psi(n, g_i) = ng_i.$$

This map is surjective, because every element of G belongs to one of the right cosets Ng_i. Let g be an element of G and g_i – the representative of the right coset Ng in R. We set

$$\widehat{g} = (F, g_i),$$

where

$$F(g_i) = g_j g g_k^{-1} \qquad \text{with} \qquad g_k = g_j \otimes g_i.$$

Observe that $g_j g g_k^{-1} \in N$ and for some $n \in N$ we have

$$\begin{aligned} \psi(n, g_i)g &= n g_i g \\ &= n g_i g g_k^{-1} g_k \\ &= n \cdot F(g_j) \cdot g_k \\ &= \psi(n \cdot F(g_i),\ g_k) \\ &= \psi(n \cdot F(g_i),\ g_j \otimes g_i) \\ &= \psi((n,\ g_i) \cdot \widehat{g}), \end{aligned}$$

which shows that ψ is a surjective map satisfying the definition of division of tss. *Q.E.D.*

Lemma 4-IV.2. *Let G be a finite group. Then*

$$\overline{(G, G)} \prec \overline{(G, \emptyset)}^1 \circ (G, G).$$

Proof: As in the previous Lemma, we have to exhibit a surjective map ψ, meeting the criteria set forth in the definition of division for tss. Consider a map $\psi : G \times G \mapsto G$ defined for all $g_1, g_2 \in G$ by

$$\psi(g_1, g_2) = g_1 g_2.$$

This map is obviously surjective. For an element g of G we set

$$\widehat{g} = (F,\ g),$$

with $F(h) = 1_G$ for all $h \in G$. We verify that ψ satisfies the division condition:

$$\psi(g_1, g_2)g = g_1 g_2 g = \psi(g_1,\ g_2 g) = \psi((g_1, g_2) \cdot \widehat{g}).$$

Setting for $g \in G$

$$\widehat{c_g} = (T, 1_G), \qquad \text{with} \qquad T(h) = c_{gh^{-1}},$$

for all $h \in G$, yields

$$\psi(g_1, g_2)c_g = g_1 g_2 \cdot (g_1 g_2)^{-1} \cdot g = g = \psi(g g_2^{-1},\ g_2) = \psi((g_1, g_2) \cdot \widehat{g}).$$

Q.E.D.

Lemma 4-IV.3. *Let* $X_i = (P_i, S_i)$, $Y_i = (Q_i, T_i)$ *for* $i = 1, 2$. *Then*

$$X_1 \prec Y_1 \wedge X_2 \prec Y_2 \ \Rightarrow\ X_2 \circ X_1 \prec Y_2 \circ Y_1.$$

Proof: By the antecedent of the lemma we have well-defined subsets $Q_i' \subseteq Q_i$, surjective maps $\psi_i : Q_i' \mapsto P_i$ and maps $s \mapsto \widehat{s} = \alpha(s)$ from S_i to T_i satisfying the condition of division for tss. Let us define for all $(q_2, q_1) \in Q_2' \times Q_1'$ a new map:

$$\psi(q_2, q_1) = (\psi_2(q_2),\ \psi_1(q_1)).$$

Clearly, ψ is surjective. For a transformation $(F,\ s)$ in the underlying semigroup of $X_2 \circ X_1$ we set

$$\overline{(F,\ s)} = (F',\ \alpha(s)), \qquad \text{where} \qquad F' = \alpha(F(\psi_1(q)))$$

for all $q \in Q_1'$. Then

$$\begin{aligned}
\psi(q_2, q_1) \cdot (F,\ s) &= (\psi(q_2),\ \psi(q_1)) \cdot (F,\ s) \\
&= (\psi_2(q_2) \cdot F(\psi_1(q_1)),\ \psi_1(q_1)s) \\
&= (\psi_2(q_2 \cdot \alpha(F(\psi_1(q_1)))),\ \psi_1(q_1\alpha(s))) \\
&= \psi(q_2 \cdot F'(q_1),\ q_1\alpha(s)) \\
&= \psi((q_2,\ q_1) \cdot \overline{(F,\ s)}),
\end{aligned}$$

for all $(q_2, q_1) \in Q_2' \times Q_1'$. *Q.E.D.*

Lemma 4-IV.4. *Let S be a finite semigroup. Then*

$$(S^I,\, S) \prec \overline{(\{a,\, b\},\, \emptyset)}^1 \circ (S^1,\, S).$$

Proof: If S is not a monoid, then $S^I = S^1$ and the result is trivial, since

$$X \prec Y \circ X$$

for all tss X and Y. Otherwise, $S^1 = S$, i.e. S is a monoid, and we define $\psi : \{a,\, b\} \times S \mapsto S^I$ in the following way:

$$\begin{aligned} \psi(a,\, 1) &= I, \\ \psi(b,\, s) &= s \end{aligned}$$

for all $s \in S$. Also for all $s \in S$ we set

$$\widehat{s} = (F,\, s).$$

We now verify the condition of division for tss for all $(q_1,\, q_2) \in \{a,\, b\} \times S$:

$$\begin{aligned} \psi(a,\, 1)s &= I \cdot s \\ &= s \\ &= \psi(a \cdot F(1),\, s) \\ &= \psi((a,\, 1)(F,\, s)) \\ &= \psi((a,\, 1)\widehat{s}), \end{aligned}$$

and

$$\begin{aligned} \psi(b,\, s')s &= s's \\ &= \psi(b \cdot F(s'),\, s's) \\ &= \psi((b,\, s')(F,\, s)) \\ &= \psi((b,\, s')\widehat{s}). \end{aligned}$$

Q.E.D.

Theorem 4-IV.5 (Krohn-Rhodes Theorem, [KR65]).
Let S be a finite semigroup. Then

$$(S^1, S) \prec X_k \circ \cdots \circ X_1,$$

where each transformation semigroup X_i is

either

$$X_i = (G, G), \quad \text{where } G \text{ is a simple group; } G \prec S,$$

or

$$X_i = \overline{(R, \emptyset)}^1, \quad \text{where } R \text{ is a finite set.}$$

Proof: The proof proceeds by first expounding the cases where *(1)* S is a group, and *(2)* S is a left-simple semigroup and then by presenting the reduction of a general case to these two special cases.

Case 1: S is a group. We repeatedly apply Lemma 4-IV.1 until the group S/N has only trivial normal subgroups, i.e. until it is simple.

Let us first analyze the case where S is a cyclic group, i.e.

$$S = \{s, s^2, \ldots, s^k = s^{k+1}\}.$$

If $k = 1$ then (S^1, S) divides any transformation semigroup with a non-empty semigroup of transformations.

Claim: If $k > 1$ then the following decomposition holds

$$(S^1, S) \prec \overline{(\{a, b\}, \emptyset)}^1 \circ (T^1, T),$$

where

$$T = \{t, t^2, \ldots, t^{k-1} = t^k\}.$$

The result for S will follow by induction on k once the claim is established. Towards that end we define a map $\psi : \{a, b\} \times T \mapsto S$ given by

$$\begin{aligned} \psi(a, t^j) &= s^j, \quad \text{for } 0 \leq j \leq k-1, \\ \psi(b, t^{k-1}) &= s^k \end{aligned}$$

and set $\widehat{s} = (F, t)$, where

$$\begin{aligned} F(t^j) &= 1_{\{a,b\}}, \quad \text{for } 0 \leq j < k-1, \\ F(t^{k-1}) &= c_b. \end{aligned}$$

We now verify that $\psi(q\widehat{s}) = \psi(q)s$ for all q in the domain of ψ:

$$\begin{aligned} \psi((a, t^j)\widehat{s}) &= \psi((a, t^j)(F, t)) \\ &= \psi(a \cdot F(t^j),\ t^{j+1}) \\ &= \psi(a \cdot 1_{\{a,b\}},\ t^{j+1}) \\ &= s^{j+1} \\ &= \psi(a, t^j)s, \end{aligned}$$

and

$$\begin{aligned} \psi((b, t^{k-1})\widehat{s}) &= \psi((b, t^{k-1})(F, t)) \\ &= \psi(b \cdot F(t^{k-1}),\ t^k) \\ &= \psi(b \cdot c_b,\ t^k) \\ &= s^{k+1} \\ &= s^k \\ &= \psi(b, t^{k-1})s. \end{aligned}$$

Then by setting $\widehat{s^j} = (\widehat{s})^j$ for $2 \leq j \leq k$ we obtain the division, which completes the proof of the claim.

Let us now turn to the general case, where $S = \{s, s^2, \ldots, s^k\}$ is a group with $s^{k+1} = s^m$ for some $1 < m < k$. Let T be the cyclic group

$$T = \{t, t^2, \ldots, t^{k-m+1} = 1\},$$

and let U be the cyclic aperiodic semigroup

$$U = \{u, u^2, \ldots, u^m = u^{m+1}\}.$$

Claim:

$$(S^1, S) \prec (T^1, T) \circ (U^1, U).$$

Once this claim is proved, the theorem (for case 1) then follows from Lemma 4-IV.3 as well as the associativity of the wreath product and the special cases treated above.

To establish the claim, we define a map

$$\begin{aligned} \psi(1, u^i) &= s^i, \quad \text{for } 0 \leq i \leq m-1, \\ \psi(t^j, u^m) &= s^{m+j} \quad \text{for } j \geq 0 \end{aligned}$$

and set $\widehat{s} = (F, u)$, where

$$\begin{aligned} F(u^i) &= 1, \quad \text{for } 0 \leq i \leq m-1, \\ F(u^m) &= t. \end{aligned}$$

Again, we verify that $\psi(q\widehat{s}) = \psi(q)s$ for all q in the domain of ψ:

$$\begin{aligned} \psi((1, u^i)\widehat{s}) &= \psi((1, u^i)(F, u)) \\ &= \psi(1 \cdot F(u^i),\ u^{i+1}) \\ &= \psi(1,\ u^{i+1}) \\ &= s^{i+1} \\ &= \psi(1, u^i)s, \end{aligned}$$

and

$$\begin{aligned} \psi((t^j, u^m)\widehat{s}) &= \psi((t^j, u^m)(F, u)) \\ &= \psi(t^j \cdot F(u^m),\ u^{m+1}) \\ &= \psi(t^{j+1},\ u^m) \\ &= s^{m+j+1} \\ &= \psi(t^j, u^m)s. \end{aligned}$$

The desired division is then obtained by setting $\widehat{s^j} = (\widehat{s})^j$ for $2 \leq j \leq k$.

Case 2: S is a left-simple semigroup. By Lemma 2-II.4, S is isomorphic to a direct product $T \times G$, where G is a group and T is a left-zero semigroup. By setting

$$\begin{aligned} \psi(t, g) &= (t, g) \quad \text{for all } (t, g) \in T^1 \times G, \\ \widehat{(t, g)} &= (F, g), \quad \text{where } F(h) = t \text{ for all } h \in G. \end{aligned}$$

we have

$$(S^1, S) \prec (T^1, T) \circ (G, G).$$

It follows from repeated application of Lemmas 4-IV.1 and 4-IV.3 that

$$(S^1, S) \prec (T^1, T) \circ (G_k, G_k) \circ \cdots \circ (G_1, G_1),$$

where the groups G_i are simple groups that divide S. It will thus suffice to show that the theorem holds for the left-zero semigroup T.

Claim:

$$(T^1, T) \prec \overline{(T^1, \emptyset)}^1 \circ \overline{(\{a, b\}, \emptyset)}^1.$$

To prove this claim we set

$$\begin{aligned} \psi(1, a) &= 1, \\ \psi(t, b) &= t, \end{aligned}$$

for $t \in T$. And also for $t \in T$:

$$\widehat{t} = (F, c_b),$$

where $F(a) = c_t$ and $F(b) = 1_{T^1}$. We then have

$$\psi(1, a)t = t = \psi(1 \cdot c_t,\ a \cdot c_b) = \psi((1, a)\widehat{t}),$$

and

$$\psi(t', b)t = t't = t' = \psi(t' \cdot 1_{T^1},\ b \cdot c_b) = \psi((t', b)\widehat{t}),$$

thus establishing the claim.

General case. The proof is by induction on the order of S.

Base case. If $|S| = 1$, then (S^1, S) divides any ts with nonempty underlying semigroup.

Inductive step. We assume $|S| > 1$ and that the theorem holds for all semigroups S'

with $|S'| < |S|$. By Lemma 2-V.12, S is either a cyclic group, or left-simple semigroup, or

$$S = P \cup T, \tag{4.9}$$

where P is a proper left ideal of S, and T is a proper subsemigroup of S. With the first two situations dealt with in cases *(1)* and *(2)* above, we now claim for the condition 4.9:

$$(S^1, S) \prec (P^I, P)^1 \circ \overline{(T^I, T)}. \tag{4.10}$$

Firstly, we observe that by the inductive hypothesis, the theorem holds for P and T. Secondly, by Lemma 4-IV.4, (P^I, P) and (T^I, T) both divide wreath products of the appropriate form and hence, so do $(P^I, P)^1$ and $\overline{(T^I, T)}$, by Lemmas 4-III.4 and 4-IV.2. Thus the theorem will follow with the establishment of division 4.10. Let us define a surjective map $\psi : P^I \times T^I \mapsto S$ by setting

$$\begin{aligned}
\psi(I, I) &= I; \\
\psi(I, t) &= t, \quad \text{for } t \in T; \\
\psi(p, I) &= p, \quad \text{for } p \in P; \\
\psi(p, t) &= pt, \quad \text{for } p \in P, t \in T.
\end{aligned}$$

We also set

$$\begin{aligned}
\widehat{s} &= (G, s), \quad \text{if } s \in T, \quad \text{where } G(t) = 1_{P^I} \text{ for all } t \in T^I, \\
\widehat{s} &= (F, c_I), \quad \text{if } s \notin T, \quad \text{where } F(t) = ts \in P \text{ for all } t \in T^I.
\end{aligned}$$

It remains to verify that $\psi(q\widehat{s}) = \psi(q)s$ for all $q \in P^I \times T^I$ and $s \in S$. We consider several cases:

$$\psi((I, t)\widehat{s}) = \left\{ \begin{array}{lllll} \psi((I, t)(G, s)) & = \psi(I \cdot G(t), ts) & = \psi(I, ts) & = & ts \\ \psi((I, t)(F, c_I)) & = \psi(I \cdot F(t), tc_I)) & = \psi(ts, I) & = & ts \end{array} \right\} = \psi(I, t)s,$$

and

$$\psi((p, I)\widehat{s}) = \left\{ \begin{array}{lllll} \psi((p, I)(G, s)) & = \psi(p \cdot G(I), s) & = \psi(p, s) & = & ps \\ \psi((p, I)(F, c_I)) & = \psi(p \cdot F(I), c_I)) & = \psi(ps, I) & = & ps \end{array} \right\} = \psi(p, I)s,$$

and

$$\psi((p,\,t)\widehat{s}) = \left\{ \begin{array}{lllllll} \psi((p,\,t)(G,\,s)) & = & \psi(p\cdot G(t),\,ts) & = & \psi(p,\,ts) & = & pts \\ \psi((p,\,t)(F,\,c_I)) & = & \psi(p\cdot F(t),\,tc_I)) & = & \psi(pts,\,I) & = & pts \end{array} \right\} = \psi(p,\,t)s,$$

which completes the proof. *Q.E.D.*

We denote by U_1 the monoid $\{0,1\}$ with the usual multiplication and by U_2 – the monoid $\{1,a,b\}$ defined by $a^2 = ba = a$ and $ab = b^2 = b$. Then the following consequence of the Krohn-Rhodes Theorem holds.

Corollary 4-IV.6. *If a monoid M is aperiodic, M divides a wreath product of copies of U_2.*

The concept of the semidirect product can be extended to varieties of semigroups and monoids. If $\mathbf{V}$ and $\mathbf{W}$ are two varieties of finite semigroups (monoids), let us denote by $\mathbf{V} * \mathbf{W}$ the variety of finite semigroups (monoids) generated by the semidirect products $S * T$, where $S \in \mathbf{V}$ and $T \in \mathbf{W}$. Since every semidirect product $S*T$ is a subsemigroup of $S \circ T$, $\mathbf{V}*\mathbf{W}$ is generated by all wreath products of the form $S \circ T$. Even though the semidirect product of finite semigroups is not associative, it becomes associative at the level of variety:

Theorem 4-IV.7 (cf. [Eil76]).
Let $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$ be varieties of finite semigroups. Then

$$(\mathbf{V}_1 * \mathbf{V}_2) * \mathbf{V}_3 = \mathbf{V}_1 * (\mathbf{V}_2 * \mathbf{V}_3).$$

4-V Block Product

Let M and N be monoids. Consider the set of all mappings $f : N \times N \mapsto M$ with the componentwise product (written additively $f = f_1 + f_2$) defined by

$$f(n_1, n_2) = f_1(n_1, n_2) \cdot f_2(n_1, n_2)$$

for all $n_1, n_2 \in N$. That is, $M^{N\times N}$ is isomorphic to the direct product of $|N|^2$ copies of M. This set together with the operation $+$ form a monoid whose identity is the mapping f such that for all pairs $(n, n') \in N \times N : f(n, n') = 1$. Let us verify that the equations

$$(fn)(n_1, n_2) = f(n_1, nn_2) \tag{4.11}$$

$$(nf)(n_1, n_2) = f(n_1 n, n_2) \tag{4.12}$$

define left and right compatible monoidal actions of N on $M^{N\times N}$:

$$\begin{aligned}(n \cdot f(x, y)) \cdot n' &= f(xn, y) \cdot n' \\ &= (fn')(xn, y) \\ &= f(xn, n'y) \\ &= (nf)(x, n'y) \\ &= n \cdot f(x, n'y) \\ &= n \cdot (f(x, y) \cdot n').\end{aligned}$$

Thus, the resulting bilateral semidirect product is called the *block product* of M and N, denoted by $M\Box N$.

We now state a formulation of the Theorem 4-IV.5 in terms of the iterated block product.

Theorem 4-V.1 (Krohn-Rhodes Theorem, [Str94]). *Let M be a finite monoid. Then there exists a sequence $M_0, \ldots, M_k$ of finite monoids such that M_0 is the trivial monoid, $M \prec M_k$, and for all $i = 0, \ldots, k-1$,*

$$M_{i+1} = N\Box M_i,$$

where N is either a simple group that divides M, or $N = U_1$. Furthermore, if M is a group, then we do not need to use any factor of the form $N = U_1$.

We note in conclusion of this section that unlike wreath product, block product is not associative even at the variety level.

Chapter 5

Automata and Logic

5-I Introduction

The connection between automata and formal logic has been a subject of research in theoretical computer science since even before these three words became a collocation and considerably before there were any electronic computers to model the theory.

In the beginning of the twentieth century David Hilbert – one of the greatest mathematicians of the last century – set out on an ambitious program: to find a way to mechanically verify the consistency of the axiomatic systems in use. In particular, he was looking for a procedure to determine if an arbitrary expression in the first-order predicate calculus, applied to integers, was true. Hilbert's project was a major intellectual effort which has had a tremendous influence on mathematical thought and culture; it yielded several important positive results for first-order logic including not only algorithms for special cases, but also the Completeness Theorem by Hilbert's own student, Kurt Gödel.

Ironically, it was Gödel who put an abrupt end to Hilbert's quest by constructing a formula in the predicate calculus applied to integers whose very definition stated that it could neither be proved nor disproved within this logical system. As the original proof of the Incompleteness Theorem published in 1931 was purely logical, with no recourse to computation, it did not immediately imply the undecidability of validity in first-order logic, which had to wait for the works of Alonzo Church and

Allan Turing.

The subsequent clarification and formalization of our intuitive notion of an effective procedure – one of the great intellectual achievements of the last century – brought about the understanding that there was no algorithm for computing many specific functions. Furthermore, some problems and functions with genuine significance in mathematics, computer science and other disciplines are noncomputable.[1]

The Turing machine has become the accepted formalization of an algorithm; it is equivalent in computing power to the digital computer as we know it today and also to all the most general mathematical notions of computation. While one cannot prove that the Turing machine is equivalent to our intuitive notion of a computer, there are some compelling arguments for this equivalence, which has become known as *Church-Turing Thesis* .[2]

First research on the logical aspects of the theory of finite-state automata, which is the subject of this chapter, dates back to the early 1960's when J. R. Büchi [Büc60] and C. C. Elgot [Elg61] showed that finite automata and monadic second-order logic (interpreted over finite words) have the same expressive power and that the transformation from formulæ to automata and vice versa are effective.[3] The reduction of formulæ to finite automata was the key to proving the decidability of two other theories: monadic second-order (MSO) of one successor function and MSO of two successor functions (denoted respectively, S1S and S2S)[4].

The equivalence between automata and logical formalisms has influenced the research in language theory as well. For example, the classification of formal languages

[1] An example is the following problem proposed by Hilbert at the World Congress of Mathematics in Paris in 1900 [Hil02]: "Is there an algorithm to decide if a multivariate polynomial equation such as $x^2y + 3yz - y^2 - 17 = 0$ has an integer solution?" This problem, which became known as *Hilbert's tenth problem*, is a special case of the problem of telling whether $\mathbf{N} \models \phi$, where $\mathbf{N}$ is a model of number theory and ϕ is a sentence (in particular, ϕ is restricted to have no Boolean connectives, no exponentiation, no quantifiers, and no inequality). But even this special case is undecidable [Mat70].

[2] Although this thesis does not admit of mathematical proof, it is *refutable*, if false.

[3] Later, such an equivalence was also shown between finite automata and monadic second-order logic over infinite words and trees.

[4] Cf. [Büc62] and [Rab69]

was deepened by including logical notions and techniques, and the logical approach helped in generalizing language theoretical results from the domain of words to more general structures like trees and partial orders.

5-II MSO-logic over finite strings

In this section we obtain a characterization of the regular languages in terms of logic. We shall consider monadic second-order sentences in which the only numerical predicates are the equality relation $(x = y)$ and the successor relation $(S(x, y))$.

In the definition of word properties, it is often convenient to allow predicates $first(x)$ and $last(x)$ which apply only to the first, respectively last position of a word model. Thus, $first(x)$ is an abbreviation of $\neg\exists y S(y, x)$ and $last(x)$ stands for $\neg\exists y S(x, y)$.

We shall use $MSO[S]$ informally to refer to this logical apparatus.

Theorem 5-II.1 ([Büc60]). *Let $L \subseteq A^*$. Then $L \in MSO[S]$ if and only if L is regular.*

Proof: Suppose L is regular and let $\mathcal{T} = (Q, q_0, F, \lambda)$ be a deterministic finite automaton. Let us assume without loss of generality that the set of states of the automaton is $Q = \{0, 1, \ldots, k\}$ and the initial state is $q_0 = 0$. To show that L is $MSO[S]$-definable we exhibit a monadic second-order sentence that expresses in any model $\underline{w}$ over A that $\mathcal{T}$ accepts w. Over a word $w = a_1 a_2 \ldots a_n$ (where $a_i \in A$), the sentence will state the existence of a successful run $q_0, \ldots, q_n$ of $\mathcal{T}$, i.e. with $q_0 = 0$, $q_i = \lambda(q_{i-1}, a_i)$, for $1 \leq i \leq n$, and $q_n \in F$. We may code such a state sequence up to q_{n-1} by a tuple $(X_0, \ldots, X_k)$ of pairwise disjoint subsets of the set $\{0, \ldots, n-1\}$ such that X_i contains those positions of w where state i is assumed by $\mathcal{T}$.[5] From the last state q_{n-1} the automaton should be able to reach a final state via the word's last

[5] A more efficient coding would use a correspondence between states and 0-1 vectors of suitable length, which allows to describe a run over 2^m states by an m-tuple of subsets of the word domain.

letter a_n. Thus $\mathcal{T}$ accepts the non-empty word w if and only if

$$\begin{aligned}\underline{w} \models \exists X_0 \ldots \exists X_k \{ \quad & \bigwedge_{i \neq j} \quad \forall x \neg (X_i(x) \wedge X_j(x)) \\ & \wedge \quad \forall x (first(x) \Rightarrow X_0(x)) \\ & \wedge \quad \forall x \forall y (S(x,y) \Rightarrow \bigvee_{j=\lambda(i,a)} (X_i(x) \wedge Q_a(x) \wedge X_j(y))) \\ & \wedge \quad \forall x (last(x) \Rightarrow \bigvee_{\exists j \in F : j=\lambda(i,a)} (X_i(x) \wedge Q_a(x))) \}\end{aligned} \tag{5.1}$$

The empty word satisfies this sentence with $X_i = \emptyset$. Thus, if $\mathcal{T}$ does not accept ϵ, a corresponding clause, such as $\exists x(x = x)$, should be added.

Now suppose L is $MSO[S]$-definable. We shall prove by induction on the construction of MSO formulæ that for any sets $\mathcal{V}_1$ and $\mathcal{V}_2$ of first- and second-order variables, and every expression ϕ with free first-order variables in $\mathcal{V}_1$ and free second-order variables in $\mathcal{V}_2$, $L(\phi)$ is a regular language.[6]

We first consider base cases, i.e. the atomic formulæ. Let $\mathcal{L}$ denote the set of all $(\mathcal{V}_1, \mathcal{V}_2)$-structures. A finite automaton over the input alphabet $A \times 2^{\mathcal{V}_1} \times 2^{\mathcal{V}}_2$ can verify that each first-order variable in $\mathcal{V}_1$ is found exactly once in an input string. Thus, $\mathcal{L}$ itself is a regular language. Whether a particular first-order variable x occurs in a letter whose first component is a, can also be easily checked by a finite automaton. Note that the intersection of the set of all such strings with $\mathcal{L}$ is the set of all $(\mathcal{V}_1, \mathcal{V}_2)$-structures satisfying $Q_a x$. One can determine with a finite automaton whether the first-order variables x and y happen to be in consecutive letters, or in the same letter, and whether any letter has x in the second component and X in the third component. Thus the sets of $(\mathcal{V}_1, \mathcal{V}_2)$-structures that satisfy each of the following: $S(x,y)$, $x = y$ and $X(x)$ are regular languages, and we therefore conclude that the claim is true for the atomic expressions.

For the inductive step, it suffices to consider the connectives $\neg$, $\wedge$ and the existential (set and variable) quantification, since the other connectives and the universal (set and variable) quantifier are definable in terms of them. Treatment of $\neg$ and $\wedge$

[6]The theorem is the case $\mathcal{V}_1 = \mathcal{V}_2 = \emptyset$.

in turn amounts to the proof that the class of regular languages shares well-known closure properties, namely closure under complement and under intersection. Indeed, if the claim is true for the MSO expressions ϕ and ψ, then it is true for $\phi \wedge \psi$ and $\neg\phi$:

$$L(\phi \wedge \psi) = L(\phi) \cap L(\psi) \cap \mathcal{L}$$

and

$$L(\neg\phi) = \mathcal{L} \backslash L(\phi).$$

Let us elaborate on the case when ϕ has the form $\exists x\psi$ and the claim is true for ψ. Then the set of $(\mathcal{V}_1 \cup \{x\},\ \mathcal{V}_2)$-structures that satisfy ψ is a regular language. Let $\mathcal{A} = (Q, q_0, F, \lambda)$ be a deterministic finite automaton recognizing this language. We now define a new automaton

$$\mathcal{T} = (Q \times \{0,1\},\ (q_0, 0),\ F \times \{1\},\ \Lambda),$$

where the new transition function Λ is defined as follows:
if $x \notin S$,

$$\Lambda((q,u),\ (a,\ S,\ T)) = (q', u);$$

if $x \in S$,

$$\Lambda((q,0),\ (a, S\backslash\{x\}, T)) = (q', 1).$$

where

$$u \in \{0,1\} \qquad \text{and} \qquad q' = \lambda(q,\ (a,\ S,\ T)).$$

It is straightforward to verify that w is accepted by $\mathcal{T}$ if and only if there is a way to adjoin x to the middle component of a letter of w so as to obtain a word recognized by $\mathcal{A}$. Thus w is accepted by $\mathcal{T}$ if and only if $w \models \exists x\psi$.

A similar construction may be proposed for the case where ϕ is of the form $\exists X\psi$. We replace the DFA $\mathcal{A} = (Q, q_0, F, \lambda)$ recognizing $L(\psi)$ by a new one, $\mathcal{T} = (Q, q_0, F, \Lambda)$, whose transition function is

$$\Lambda(q,\ (a,\ S,\ T\backslash\{X\})) = \lambda(q,\ (a,\ S,\ T)).$$

Thus, $\mathcal{T}$ recognizes $L(\exists X\psi)$. *Q.E.D.*

The expression 5.1 in the above proof is an EMSO-formula of a special type. Invoking the second part of the proof, we see that it provides a normal form of $MSO[S]$-formulæ, an *automata normal form* in Büchi's terminology.

Corollary 5-II.2.

Any $MSO[S]$-formula can be written as an $EMSO[S]$-formula.

5-III Algebraic Characterization of $FO[<]$

In order to tackle the issue of an algebraic characterization of the family of languages $FO[<]$ we first introduce a connection between existential quantification and the block product of the form $U_1 \Box M$.

Let ϕ be a formula of $FO[<]$ in which free variables belong to a finite set $\mathcal{V}$. Recall (section 1-IV.3-b) that by $L(\phi)$ we mean the set of $\mathcal{V}$-structures that satisfy ϕ.

Notation. $M(\phi)$, $\eta_\phi : (A \times 2^{\mathcal{V}})^* \mapsto M(\phi)$ and $\sim_\phi$ denote, respectively, the syntactic monoid, the homomorphism and the syntactic congruence of $L(\phi)$. We embed A into $A \times 2^{\mathcal{V}}$ by identifying $a \in A$ with $(a, \emptyset)$. By θ_ϕ and $N(\phi)$ we denote, respectively, the restriction of η_ϕ to A^* and the image of this restriction.

Lemma 5-III.1 ([Str94]). *Let $\phi \in FO[<]$ and x be a free variable in ϕ. Let π be the projection homomorphism from $U_1 \Box M(\phi)$ onto $M(\phi)$. Then there exists a homomorphism $\zeta : A^* \mapsto U_1 \Box M(\phi)$ such that $\pi \circ \zeta = \theta_\phi$, and $\theta_{\exists x\phi}$ factors through ζ.*

Proof: Let $\mathcal{V}$ be the set of free variables of $\exists x\phi$, so that

$$\eta_\phi : (A \times 2^{\mathcal{V} \cup \{x\}})^* \mapsto M(\phi)$$

is a homomorphism and there exists $T \subseteq M(\phi)$ such that $L(\phi) = \eta_\phi^{-1}(T)$. We define a function $F : M(\phi) \times M(\phi) \mapsto U_1$ given by

$$F^{(\alpha,\beta)} = \begin{cases} 0\,, \text{ if } \alpha \cdot \eta_\phi(a, S \cup \{x\}) \cdot \beta \in T; \\ 1\,, \text{ otherwise} \end{cases}$$

Consider a homomorphism $\theta_{\exists x\phi} : (A \times 2^{\mathcal{V}})^* \mapsto U_1 \Box M(\phi)$ such that for every letter (a, S) over the alphabet $A \times 2^{\mathcal{V}}$,

$$\theta_{\exists x\phi}(a, S) = (F^{(\alpha,\beta)}, \eta_\phi(a, S)).$$

Suppose $w = (a_1, S_1)(a_2, S_2)(a_3, S_3)$ and let us write η_i in lieu of $\eta_\phi(a_i, S_i)$ to facilitate readability. Then

$$\begin{aligned}
\theta_{\exists x\phi}(w) &= ((F_1^{(\alpha,\,\beta)},\ \eta_1)(F_2^{(\alpha,\,\beta)},\ \eta_2))(F_3^{(\alpha,\,\beta)},\ \eta_3) && \text{by assumption} \\
&= (F_1^{(\alpha,\,\beta)} \cdot \eta_2 + \eta_1 \cdot F_2^{(\alpha,\,\beta)},\ \eta_1\eta_2)(F_3^{(\alpha,\,\beta)},\ \eta_3) && \text{by 4.7} \\
&= (F_1^{(\alpha,\,\eta_2\beta)} + F_2^{(\alpha\eta_1,\,\beta)},\ \eta_1\eta_2)(F_3^{(\alpha,\,\beta)},\ \eta_3) && \text{by 4.11 and 4.12} \\
&= ((F_1^{(\alpha,\,\eta_2\beta)} + F_2^{(\alpha\eta_1,\,\beta)}) \cdot \eta_3 + \eta_1\eta_2 F_3^{(\alpha,\,\beta)},\ \eta_1\eta_2\eta_3) && \text{by 4.7} \\
&= (F_1^{(\alpha,\,\eta_2\beta)}\eta_3 + F_2^{(\alpha\eta_1,\,\beta)}\eta_3 + \eta_1\eta_2 F_3^{(\alpha,\,\beta)},\ \eta_\phi(w)) && \text{by 4.2} \\
&= (F_1^{(\alpha,\,\eta_3\eta_2\beta)} + F_2^{(\alpha\eta_1,\,\eta_3\beta)} + F_3^{(\alpha\eta_1\eta_2,\,\beta)},\ \eta_\phi(w)) && \text{by 4.11 and 4.12} \\
&= (F_1^{(\alpha,\,\eta_2\eta_3\beta)} + F_2^{(\alpha\eta_1,\,\eta_3\beta)} + F_3^{(\alpha\eta_1\eta_2,\,\beta)},\ \eta_\phi(w)) && \text{by commutativity of } \eta
\end{aligned}$$

If $w = (a_1, S_1) \cdots (a_n, S_n) \in A \times 2^{\mathcal{V}}$, $n > 3$, we have:

$$\begin{aligned}
\theta_{\exists x\phi}(w) &= (F_1^{(\alpha,\beta)},\ \eta_\phi(a_1, S_1)) \cdots (F_n^{(\alpha,\beta)},\ \eta_\phi(a_n, S_n)) \\
&= (\underbrace{\prod_{i=1}^{n} F_i^{(\alpha\eta_\phi((a_1,S_1)\cdots(a_{i-1},S_{i-1})),\ \eta_\phi((a_{i+1},S_{i+1})\cdots(a_n,S_n))\beta)}}_{G(\alpha,\,\beta)},\ \eta_\phi(w)) \\
&= (G(\alpha,\ \beta),\ \eta_\phi(w)).
\end{aligned}$$

Then

$$\begin{aligned}
G(1,1) = 0 \iff\ & \exists i, 1 \le i \le n, \text{ such that} \\
& F_i^{(\eta_\phi((a_1,S_1)\cdots(a_{i-1},S_{i-1})),\ \eta_\phi((a_{i+1},S_{i+1})\cdots(a_n,S_n)))} = 0 \\
\iff\ & \exists w', w'' \text{ such that } \eta_\phi(w'(a, S \cup \{x\})w'') \in T \\
\iff\ & w = w'(a, S \cup \{x\})w'' \\
\iff\ & w \models \exists x\phi.
\end{aligned}$$

Hence the language defined by $\exists x\phi$ is

$$L(\exists x\phi) = \theta_{\exists x\phi}^{-1}(K), \text{ where } K = \{(G,\ m) \in U_1 \Box M(\phi) |\ G(1,1) = 0\}.$$

By Theorem 2-IV.3, $\eta_{\exists x\phi}$ factors through $\theta_{\exists x\phi}$, and $\pi \circ \theta_{\exists x\phi}$ is the restriction of η_ϕ to $(A \times 2^{\mathcal{V}})^*$. Then by setting ζ to be the restriction of $\theta_{\exists x\phi}$ to A^*, we have the desired result. *Q.E.D.*

Let $\mathcal{V}$ be a set of first-order variables that does not include the variable x; let w be a $(\mathcal{V} \cup \{x\})$-structure in which all the variables of $\mathcal{V}$ appear to the left of the position containing x. Denote by w' the prefix of w consisting of the letters to the left of the position that contains the variable x.

Lemma 5-III.2 ([Str94]). *Let ϕ be an $FO[<]$ expression whose free variables are in $\mathcal{V}$. Then, there exists a formula $\phi[< x]$ with free variables in $\mathcal{V} \cup \{x\}$ such that, with w, w' as above,*

$$w \models \phi[< x] \iff w' \models \phi.$$

Proof: The proof proceeds by induction on the structure of the formula ϕ. If ϕ is an atomic expression, we take $\phi[< x]$ to be itself. The desired result also holds for the following two cases, which are easily verifiable. Take $(\phi \wedge \psi)[< x]$ to be $\phi[< x] \wedge \psi[< x]$ and $(\neg\phi)[< x]$ to be $\neg(\phi[< x])$.

If ϕ is of the form $\exists y\psi$ then we set $\phi[< x]$ to be

$$\exists y((y < x) \wedge \psi[< x]).$$

Let us assume that w satisfies the stated property, with the variable x occurring in the k^{th} position. If

$$w \models \phi[< x],$$

then by adjoining the variable y to any position to the left of that occupied by x, we obtain a new structure $\overline{w}$ such that

$$\overline{w} \models \phi[< x].$$

By inductive hypothesis, the prefix $\overline{w'}$ of $\overline{w}$ of length $k-1$ satisfies

$$\overline{w'} \models \psi.$$

Now, by removing the variable y we arrive at

$$w' \models \exists y \psi.$$

For the converse, suppose

$$w' \models \exists y \psi,$$

where w' is the prefix of w of length $k-1$. Variable y can be adjoined to some position of w' to obtain a structure $\overline{w'}$ such that

$$\overline{w'} \models \psi.$$

If v is the suffix of w of length $|w| - k + 1$, then by the inductive hypothesis

$$\overline{w'}v \models \psi[< x].$$

Removal of the variable y yields

$$w = w'v \models \exists y((y < x) \wedge \psi[< x]).$$

Q.E.D.

The formula $\phi[< x]$ is called the *relativization* of ϕ. Relativizations of $\phi[> x]$, $\phi[\leq x]$ and $\phi[\geq x]$ can be defined similarly.

Lemma 5-III.3 ([Str94]). *Let M be a finite monoid such that every language $L \subseteq A^*$ recognized by M is defined by a sentence of $FO[<]$. Then every language in A^* recognized by the block product $U_1 \Box M$ is in $FO[<]$.*

Proof: Suppose M satisfies the hypothesis of the lemma. The block product $U_1 \Box M$ is isomorphic to a bilateral semidirect product $V * * M$, where V is idempotent and commutative. If $L \subseteq A^*$ is recognized by a homomorphism

$$\gamma : A^* \mapsto V * * M,$$

then there exists $T \subseteq V**M$ such that $L = \gamma^{-1}(T)$. We need to show that for each $(v, m) \in V**M$, $\gamma^{-1}(v, m)$ is defined by a sentence $\phi_{(v,m)}$ of Σ_{k+1}. Then

$$L = \bigvee_{(v,m)\in T} \phi_{(v,m)}.$$

If $a \in A$, we denote by v_a the left-hand co-ordinate of $\gamma(a)$. Let π be a projection homomorphism from $V**M$ onto M. Then for $w \in A^*$, we have:

$$\gamma(w) = (v,\, m) \iff \begin{cases} \pi \circ \gamma(w) = m \\ \sum_{w=w'aw''} \pi \circ \gamma(w') \cdot v_a \cdot \pi \circ \gamma(w'') = v. \end{cases}$$

By hypothesis, $\pi \circ \gamma(w) = m$ if and only if $w \models \delta_m$, where δ_m is a sentence of Σ_k. Since V is idempotent and commutative, the second equation depends only on the set of summands that appear, and thus w satisfies the second condition if and only if it satisfies a boolean combination of the conditions of the form

$$w = w'aw'' \tag{5.2}$$

where $\pi \circ \gamma(w') = m' \in M$ and $\pi \circ \gamma(w'') = m'' \in M$. A condition 5.2 is expressed by the sentence

$$\exists x(Q_a x \wedge \delta_{m'}[< x] \wedge \delta_{m''}[> x]),$$

where $\delta_{m'}[< x]$ and $\delta_{m''}[> x]$ are the relativized formulæ of Lemma 5-III.2. Thus L is defined by the conjunction

$$\mathcal{BC}\big(\exists x(Q_a x \wedge \delta_{m'}[< x] \wedge \delta_{m''}[> x])\big) \bigwedge \delta_m,$$

where $\mathcal{BC}$ stands for boolean combination. Hence $L \in FO[<]$. *Q.E.D.*

Theorem 5-III.4 ([MP71]). *Let $L \subseteq A^*$ be a language and $M(L)$ a monoid that recognizes L. Then $L \in FO[<]$ if and only if $M(L)$ is finite and aperiodic.*

Proof: Let us first assume that $M(L)$ is finite and aperiodic. By Theorem 4-V.1,

$$M(L) \prec U_1 \Box (U_1 \Box \cdots (U_1 \Box \{1\}) \cdots) \tag{5.3}$$

and by Theorem 2-IV.3, L is recognized by the block product above. Since the only languages recognized by the trivial monoid $\{1\}$ are $\emptyset$ and A^*, a repeated application of Lemma 5-III.3 leads to the desired conclusion: $L \in FO[<]$.

Conversely, let ϕ be an $FO[<]$ formula. We shall prove by induction on the construction of ϕ that $N(\phi)$ is aperiodic. (We use notation established in the beginning of this section. If ϕ is a sentence, then $M(\phi) = N(\phi)$ since $\mathcal{V} = \emptyset$.)

Base case: ϕ is an atomic expression. Then $N(\phi)$ is trivial and therefore aperiodic.

For the inductive step, we claim that aperiodicity is invariant under the boolean operations as well as existential quantification. The former is given by Proposition 3-V.1 and the fact that the aperiodic monoids form a variety. The latter follows from Lemma 5-III.1 and the fact that if M is aperiodic, then $U_1 \Box M$ is also aperiodic (by Lemma 4-III.3). *Q.E.D.*

Corollary 5-III.5. *There is an algorithm to decide whether a given regular language L is in $FO[<]$.*

Proof: The multiplication table of the syntactic monoid of L can be effectively computed and analyzed for the presence of a non-trivial group. The latter would indicate that $L \notin FO[<]$ since an aperiodic monoid contains no non-trivial groups. *Q.E.D.*

5-III.1 A Hierarchy in $FO[<]$

In this section we show that the logical hierarchy Σ_k of $FO[<]$ is infinite.

Let $A = \{a, b\}$ and let $\mathcal{B}_0$ be the family of atomic expressions of $FO[<]$. For $k \geq 0$, $\mathcal{B}_{k+1}$ denotes the family of boolean combinations of the expressions of the form

$$\exists x_1 \cdots \exists x_r \phi,$$

where $r \geq 0$ and $\phi \in \mathcal{B}_k$. We shall also use $\mathcal{B}_k$ to denote the family of languages in A^* defined by the sentences of $\mathcal{B}_k$.

Theorem 5-III.6 ([Tho82]). *For all $k \geq 0$, $\mathcal{B}_k$ is strictly contained in $FO[<]$.*

Proof: We prove the theorem by exhibiting a language contained in $FO[<]$, but not in $\mathcal{B}_k$. The proof is split into Lemmas 5-III.7, 5-III.8 and 5-III.9.

Let $k \geq 1$ and let L_k be the set of all $w \in A^*$ such that for every prefix v of w,

$$0 \leq |v|_a - |v|_b \leq k.$$

L_k is a regular language. It's minimal automaton for the case $k = 3$ is pictured in the fig. 5.1.

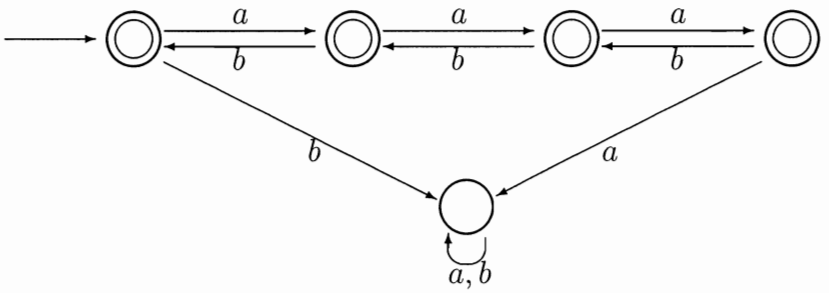

Figure 5.1: The minimal automaton of L_3 over $A = \{a, b\}$.

Lemma 5-III.7. $L_k \in FO[<]$.

Proof: By Theorem 5-III.4, establishing the aperiodicity of $M(L_k)$ will show that $L_k \in FO[<]$. Suppose $v \in A^*$ is a word such that $|v|_a \neq |v|_b$. Then v^{k+1} maps all states of the minimal automaton to the unique nonaccepting state. If, on the other hand, $|v|_a = |v|_b$, then v and v^2 induce the same transition of the set of states of the minimal automaton of L_k. Thus we have that for every word $v \in A^*$, $v^{k+1} \equiv_{L_k} v^{k+2}$ and hence, $M(L_k)$ is aperiodic. *Q.E.D.*

To show that $L_k \notin \mathcal{B}_k$ we define for each $m \geq 1$, a sequence of triples of words in A^*:

$$\{(u_{m,r},\ v_{m,r},\ w_{m,r})\}_{r \geq 0},$$

where

$$\begin{array}{llllll} u_{m,0} = 1, & v_{m,0} = a, & w_{m,0} = b, \\ u_{m,1} = (ab)^{2^m}, & v_{m,1} = u_{m,1}au_{m,1}, & w_{m,1} = u_{m,1}bu_{m,1}, \\ u_{m,r+1} = (v_{m,r}w_{m,r})^{2^m}, & v_{m,r+1} = u_{m,r+1}au_{m,r+1}, & w_{m,r+1} = u_{m,r+1}bu_{m,r+1}. \end{array}$$

Lemma 5-III.8. *For $\phi \in \mathcal{B}_k$ let $\theta_\phi : A^* \mapsto M(\phi)$ be the restriction of the syntactic morphism of ϕ to A^*. Then there exists $n \geq 1$ such that for all $m \geq n$,*

$$\theta_\phi(u_{m,k}) = \theta_\phi(v_{m,k}) = \theta_\phi(w_{m,k}).$$

The **proof** is by induction on k.

Case $k = 0$: ϕ is an atomic expression and every word in A^* is mapped to the identity of $M(\phi)$.

Inductive step: Suppose the proposition is true for some $k \geq 0$. Then let $\phi \in \mathcal{B}_{k+1}$ and

$$\phi = \exists x_1, \cdots \exists x_r \psi,$$

where $\psi \in \mathcal{B}_k$. By the inductive hypothesis, there exists $n \geq 1$ such that for all $m \geq n$

$$\theta_\psi(u_{m,k}) = \theta_\psi(v_{m,k}) = \theta_\psi(w_{m,k}).$$

By Theorem 5-III.4, there exists $s \geq 1$ such that for all $w \in A^*$,

$$\theta_\phi(w^s) = \theta_\phi(w^{s+1}).$$

Let $n' = \max(n, s)$ and $m \geq n'$. Observe that since $u_{m,k} = x^m$ for some word x, $\theta_\phi(u_{m,k})$ is idempotent:

$$\begin{aligned} \theta_\phi(u_{m,k}) &= \theta_\phi(x^m) \\ &= \theta_\phi(x^{m-s}x^s) \\ &= \theta_\phi(x^{m-s}x^{s+1}) \\ &= \theta_\phi(x^{m-s+1}x^s) \\ &\cdots\cdots\cdots\cdots \\ &= \theta_\phi(x^{2m}) \\ &= (\theta_\phi(u_{m,k}))^2. \end{aligned}$$

Now suppose

$$z_1 v_{m,k+1} z_2 \models \phi.$$

Substituting $u_{m,k+1} a u_{m,k+1}$ for $v_{m,k+1}$,

$$z_1 u_{m,k+1} a u_{m,k+1} z_2 \models \phi,$$

and $(v_{m,k} w_{m,k})^{2^m}$ for $u_{m,k+1}$,

$$z_1 (v_{m,k} w_{m,k})^{2^m} a (v_{m,k} w_{m,k})^{2^m} z_2 \models \phi.$$

Yet another substitution and the idempotence of $\theta_\phi(u_{m,k})$ gives

$$z_1 (u_{m,k} a u_{m,k} b u_{m,k})^{2^m} a u_{m,k}^{2r+1} (u_{m,k} a u_{m,k} b u_{m,k})^{2^m} z_2 \models \phi.$$

To obtain a structure that satisfies ψ, we first adjoin the variables $x_1, \ldots, x_r$ to positions of the structure displayed above. Observe, that at least one of the factors of $u_{m,k}$ in the block $u_{m,k}^{2r+1}$ will not be affected. Since by the inductive hypothesis,

$$\theta_\psi(u_{m,k}) = \theta_\psi(w_{m,k}) = \theta_\psi(u_{m,k} b u_{m,k}),$$

the non-affected factor $u_{m,k}$ can be replaced by $u_{m,k} b u_{m,k}$ and thereby another structure satisfying ψ is obtained. Now, removal of variables $x_1, \ldots, x_r$ yields:

$$z_1 (u_{m,k} a u_{m,k} b u_{m,k})^{2^m} a u_{m,k}^{t} b u_{m,k}^{t'} (u_{m,k} a u_{m,k} b u_{m,k})^{2^m} z_2 \models \phi$$

for some $t, t' \geq 1$. We use the idempotence of $\theta_\phi(u_{m,k})$ again to obtain

$$z_1 (v_{m,k} w_{m,k})^{2^{m+1}+1} z_2 \models \phi.$$

And since $\theta_\phi(u_{m,k+1})$ is idempotent as well,

$$z_1 u_{m,k+1} z_2 = z_1 (v_{m,k} w_{m,k})^{2^{m+1}} z_2 \models \phi.$$

Thus we have shown the implication:

$$z_1 v_{m,k+1} z_2 \models \phi \;\Rightarrow\; z_1 u_{m,k+1} z_2 \models \phi.$$

For the converse, let us suppose

$$z_1 u_{m,k+1} z_2 \models \phi.$$

By the idempotence of $\theta_\phi(u_{m,k+1})$ we first arrive at

$$z_1 u_{m,k+1} u_{m,k+1} z_2 \models \phi,$$

and then

$$z_1 u_{m,k+1} u_{m,k+1}^{2r+1} u_{m,k+1} z_2 \models \phi.$$

We adjoin the variables $x_1, \ldots, x_r$ to positions in the structure above and thus obtain a structure that satisfies ψ. Since one of the $2r+1$ consecutive occurrences of $u_{m,k+1}$ is unaffected, by the inductive hypothesis we can replace it by $u_{m,k+1} a u_{m,k+1}$ and obtain another structure satisfying ψ. Once the adjoined variables are removed, we have

$$z_1 u_{m,k+1} u_{m,k+1}^{t} a u_{m,k+1}^{t'} u_{m,k+1} z_2 \models \phi,$$

for some $t, t' \geq 1$. It follows from the idempotence of $\theta_\phi(u_{m,k+1})$ that

$$z_1 v_{m,k+1} z_2 = z_1 u_{m,k+1} a u_{m,k+1} z_2 \models \phi.$$

Therefore, we have $z_1 u_{m,k+1} z_2 \models \phi \;\Rightarrow\; z_1 v_{m,k+1} z_2 \models \phi$ and hence,

$$u_{m,k+1} \equiv_\phi v_{m,k+1}.$$

A similar argument will lead to the conclusion that

$$u_{m,k+1} \equiv_\phi w_{m,k+1}.$$

To complete the proof of the lemma, suffice it to show that the stated property of formulæ in $\mathcal{B}_{k+1}$ is preserved under boolean operations. Let ϕ, ψ be expressions of $FO[<]$ such that the following equations hold, respectively, for $m \geq n_\phi$ and $m \geq n_\psi$.

$$\begin{aligned} \theta_\phi(u_{n_\phi,k+1}) &= \theta_\phi(v_{n_\phi,k+1}) = \theta_\phi(w_{n_\phi,k+1}) \\ \theta_\psi(u_{n_\psi,k+1}) &= \theta_\psi(v_{n_\psi,k+1}) = \theta_\psi(w_{n_\psi,k+1}). \end{aligned}$$

If we choose $n = \max(n_\phi, n_\psi)$, then for $m \geq n$:

$$u_{m,k+1} \equiv_{\phi\wedge\psi} v_{m,k+1} \equiv_{\phi\wedge\psi} w_{m,k+1},$$

which proves that the property is preserved under conjunction. Since $\theta_\phi = \theta_{\neg\phi}$ (see proof of Proposition 3-V.1 on page 47), it is also preserved under negation and we have the desired result. *Q.E.D.*

Lemma 5-III.9. $L_k \notin \mathcal{B}_k$.

Proof: One can show by induction on k that for all $k \geq 1$,

$$|u_{m,k}|_a = |u_{m,k}|_b,$$

and for all prefixes v of $u_{m,k}$,

$$0 \leq |v|_a - |v|_b \leq k.$$

Therefore for all $m, k \geq 1$, $u_{m,k} \in L_k$. By Lemma 5-III.8, $w_{m,k}$ is also in L_k, but from the definition of $w_{m,k}$, it contains a prefix in which there are more occurrences of b than a. So, clearly, $w_{m,k} \notin L_k$ for any m, k. A contradiction. Therefore $L_k \notin \mathcal{B}_k$ and

$$L_k \in FO[<] \setminus \mathcal{B}_k.$$

Q.E.D.

Chapter 6

Piecewise Testable Languages

6-I Introduction

Our presentation of piecewise testable languages hinges upon one of the fundamental results in the theory of formal languages: the theorem of I. Simon [Sim75]. Its proof relies on another remarkable combinatorial result involving words whose importance itself warrants attention.

Simon's theorem enables us to describe the variety of languages corresponding to the variety of monoids **J**. These languages also occupy a special place in the logical hierarchy of languages, as we shall see in the next chapter.

6-II Simon's Theorem

Let A be an alphabet. Recall that a word $a_1 \dots a_k \in A^*$ is a subword of a word v of A^* if there exist words $v_0, v_1, \dots, v_k \in A^*$ such that $v = v_0 a_1 v_1 \dots a_k v_k$. Let $\aleph(v)$ denote the set of subwords of v and $\aleph_{\leq n}(v)$ denote the set of subwords of length less than or equal to n of the word v. For each integer $n \geq 0$, we define an equivalence relation $\sim_n$ on A^* given by

$$u \sim_n v \iff \aleph_{\leq n}(v) = \aleph_{\leq n}(u).$$

One can easily verify that $\sim_n$ is a congruence of finite index with the number of equivalence classes bounded by 2^N, where $N = |\{w \in A^* |\ |w| \leq n\}|$.

A language is *piecewise testable* if it is the union of classes modulo $\sim_n$, for some $n \in \mathbb{N}$. In other words, L is piecewise testable if there exists an integer n such that one can test whether a word w belongs to L by simple inspection of its subwords of length at most n.

Proposition 6-II.1.
A language $L \subseteq A^*$ is piecewise testable if and only if it is in the boolean algebra generated by the languages of the form $A^*a_1A^*a_2 \ldots A^*a_nA^*$, where $n \geq 0$ and $a_i \in A$.

Proof: Let u be a word of A^*. We then observe

$$\{v \in A^* |\ v \sim_n u\} = \Big(\bigcap_{\substack{(a_1,\ldots,a_m),\\ 0\leq m\leq n,\\ a_1\cdots a_m \in \aleph(u)}} A^*a_1A^*a_2 \ldots A^*a_mA^* \Big) \setminus \Big(\bigcup_{\substack{(a_1,\ldots,a_m),\\ 0\leq m\leq n,\\ a_1\cdots a_m \notin \aleph(u)}} A^*a_1A^*a_2 \ldots A^*a_mA^* \Big).$$

From this it follows that if L is a union of classes modulo $\sim_n$, L is in the boolean algebra generated by the languages of the form $A^*a_1A^*a_2 \ldots A^*a_nA^*$.

For the converse, suppose $L = A^*a_1A^*a_2 \ldots A^*a_nA^*$ and $u \in L$. Then $a_1 \ldots a_n \in \aleph(u)$. Therefore if $u \sim_n v$, then $a_1 \ldots a_n \in \aleph(v)$ and hence $v \in L$. This shows that L is saturated by the relation $\sim_n$ and thus L is a finite union of classes modulo $\sim_n$. *Q.E.D.*

We shall now turn to establishing the properties of the congruence $\sim_n$ which provide a basis for the syntactic characterization of piecewise testable languages.

Proposition 6-II.2. *Let $u, v \in A^*$ and $a \in A$. Then*

$$uav \sim_{2n-1} uv \Rightarrow ua \sim_n u \vee av \sim_n v.$$

Proof: Let us prove the stated property by showing the validity of its negation, i.e. we will show

$$ua \not\sim_n u \wedge av \not\sim_n v \Rightarrow uav \not\sim_{2n-1} uv. \tag{6.1}$$

Assume the antecedent in 6.1. Then there exist a word x in A^*, $|x| \leq n$, such that x is a subword of ua, but x is not a subword of u. Note also, the following factorization holds: $x = x'a$.

Likewise, there exist a word y in A^*, such that $y \in \aleph(av) \wedge y \notin \aleph(v)$ and $y = ay'$. Therefore the word $x'ay'$, whose length is $|x'ay'| \leq 2n - 1$, is a subword of uav but not of uv, as was to be shown. *Q.E.D.*

Let u be a word over the alphabet A. Then by $\alpha(u)$ we denote the set of all letters appearing in u:

$$\alpha(u) = \{a \in A |\ |u|_a > 0\}.$$

Theorem 6-II.3. *Let $u, v \in A^*$. Then*

$$u \sim_n vu \iff \exists\, u_1, \ldots, u_n \in A^*[\ (u = u_1 \cdots u_n) \wedge(\alpha(v) \subseteq \alpha(u_1) \subseteq \cdots \subseteq \alpha(u_n))\ ].$$

Proof: The result is trivial if $u = \epsilon$ and therefore we can proceed with the assumption that $u \in A^+$.

We prove the theorem in both direction by induction on n. First, we show that the condition is *necessary.*

Base case: If $n = 1$, we have

$$u \sim_1 vu \Rightarrow \aleph_{\leq 1}(u) = \aleph_{\leq 1}(vu) \Rightarrow \alpha(u) = \alpha(vu) \Rightarrow \alpha(v) \subseteq \alpha(u).$$

Inductive step. We assume the condition is necessary for some $n > 1$, i.e.

$$u_1 \cdots u_n \sim_n vu_1 \cdots u_n \Rightarrow \alpha(v) \subseteq \alpha(u_1) \subseteq \cdots \subseteq \alpha(u_n)$$

and show that it is the case for $n + 1$.

Suppose the congruence $u \sim_{n+1} vu$ holds and let u_{n+1} be the shortest right factor of u such that $\alpha(u_{n+1}) = \alpha(u)$. Since $u \in A^+$, $\alpha(u)$ is non-empty and therefore $u_{n+1} \in A^+$. Thus, we can assume $u_{n+1} = au'$ for some $a \in A$ and $u' \in A^*$. Then

$u = wu_{n+1} = wau'$. By definition, u_{n+1} is the shortest right factor of u containing the same set of letters as u. Therefore the letter a cannot appear in u', since otherwise u' would be the shortest such.

We now show that $w \sim_n vw$. Let x be a word in A^* such that $x \in \aleph_{\leq n}(vw)$. Then

$$xa \in \aleph_{\leq n+1}(vwa) \;\Rightarrow\; xa \in \aleph_{\leq n+1}(vu).$$

Since by induction hypothesis $u \sim_{n+1} vu$, xa is a subword of $u = wau'$ and, since a is not a letter of u', xa is a subword of wa. Thus x is a subword of w.

For the converse, we first note that every subword of w is a subword of vw, and therefore $w \sim_n vw$, as stated. By the inductive hypothesis, there exist $u_1, \ldots, u_n \in A^*$ such that $w = u_1 \cdots u_n$ and $\alpha(v) \subseteq \alpha(u_1) \subseteq \cdots \subseteq \alpha(u_n)$. Since $u = wu_{n+1}$ and $\alpha(u_n) \subseteq \alpha(u) = \alpha(u_{n+1})$, we have the desired result.

Let us now prove that the condition is *sufficient.*

Base case: $n = 1$, then $u_1 = u$ and $\alpha(v) \subseteq \alpha(u)$ implies $\alpha(u) = \alpha(vu)$, i.e. $u \sim_1 vu$.

Inductive step. Assuming the condition is sufficient for some $n > 1$, i.e.

$$\alpha(v) \subseteq \alpha(u_1) \subseteq \cdots \subseteq \alpha(u_n) \;\Rightarrow\; u_1 \cdots u_n \sim_n vu_1 \cdots u_n,$$

we shall prove it holds for $n + 1$.

Suppose $u = u_1 \cdots u_{n+1}$ and $\alpha(v) \subseteq \alpha(u_1) \subseteq \cdots \subseteq \alpha(u_{n+1})$. Then $\alpha(vu) = \alpha(u) = \alpha(u_{n+1})$, i.e. the set of letters appearing in the factor u_{n+1} is identical to that of vu. Let $x \in A^+$ be such that $x \in \aleph_{\leq n+1}(vu)$ and let x' be the longest right factor of x such that x' is a subword of u_{n+1}. Then x admits the factorization $x = x''x'$, where x'' is a subword of $vu_1 \cdots u_n$. Since u_{n+1} contains all the letters of vu, it must contain all the letters of x at least once. Therefore, x' is non-empty. By definition, $|x| \leq n + 1$ and since $|x'| \geq 1$, we have $|x''| \leq n$. Thus

$$x'' \in \aleph_{\leq n}(vu_1 \cdots u_n).$$

Since $u_1 \cdots u_n \sim_n vu_1 \cdots u_n$ by the inductive hypothesis, we conclude that x'' is in fact a subword of $u_1 \cdots u_n$. Hence, $x = x''x'$ is a subword of $u = u_1 \cdots u_{n+1}$, so

$$\forall x[\; (x \in \aleph_{\leq n+1}(vu_1 \cdots u_{n+1}) \wedge x \neq \epsilon) \;\Rightarrow\; x \in \aleph_{\leq n+1}(u_1 \cdots u_{n+1})\;].$$

Conversely, every subword of u is clearly a subword of vu and therefore $u \sim_{n+1} vu$. *Q.E.D.*

Corollary 6-II.4.

$$\forall u, v \in A^*[\ (uv)^n u \sim_n (uv)^n \sim_n v(uv)^n\ ]$$

Proof: The congruence $(uv)^n \sim_n v(uv)^n$ follows immediately from Theorem 6-II.3. The derivation of the expression $(uv)^n \sim_n (uv)^n u$ is similar. *Q.E.D.*

The following is a remarkable combinatorial property of the congruence $\sim_n$.

Theorem 6-II.5 ([Sim75]).

$$x \sim_n y \ \Rightarrow\ \exists\, h[\ x \in \aleph(h) \wedge y \in \aleph(h) \wedge x \sim_n h \sim_n y\ ].$$

Proof: By induction on $k = |x| + |y| - 2|x \oplus y|$, where $|x \oplus y|$ is the largest left factor common to both x and y.

Base case: $k = 0$, then $x = y$ and we can take $h = x = y$. The cases $x \in \aleph(y)$ or $y \in \aleph(x)$ are trivial and therefore excluded from further consideration.

Inductive step. Let

$$x = uav \qquad \text{and} \qquad y = ubw,$$

where $u, v, w \in A^*$ and a and b are two distinct letters of A. We shall show that $ubw \sim_n ubav$ or $uav \sim_n uabw$.

Suppose neither of the above assertions is true. Since $ubw = y \sim_n x$ and $uav = x \in \aleph(ubav)$, there exists a word r such that $r \in \aleph_{\leq n}(ubav)$ and $r \notin \aleph(ubw)$. Similarly, there exists a word s such that $s \in \aleph_{\leq n}(uabw)$ and $s \notin \aleph(uav)$. Let

$$\begin{aligned} r &= r_1 b r_2, \quad \text{where } r_1 \in \aleph(u) \quad \text{and } r_2 \in \aleph(av) \\ s &= s_1 a s_2, \quad \text{where } s_1 \in \aleph(u) \quad \text{and } s_2 \in \aleph(bw). \end{aligned}$$

From this we deduce that $r_1 b \notin \aleph(u)$ (otherwise $r = r_1 b r_2 \in \aleph(uav)$, and since $uav = x \sim_n y$, r would be a subword of y). Similarly, $s_1 a \notin \aleph(u)$.

Since $r_2 \in \aleph(av)$, r_2 admits factorization $r_2 = r_2''r_2'$ with $r_2'' = \epsilon$ or $r_2'' = a$ and $r_2' \in \aleph(v)$. And likewise, since $s_2 \in \aleph(bw)$ we have $s_2 = s_2''s_2'$, where $s_2'' = \epsilon$ or $s_2'' = b$ and $s_2' \in \aleph(w)$. Thus

$$\begin{aligned} |r_1bs_2'| + |s_1ar_2'| &\leq |r_1as_2| + |s_1br_2| \\ &= |r| + |s| \\ &\leq 2n, \end{aligned}$$

whence

$$|r_1bs_2'| \leq n \qquad \text{or} \qquad |s_1ar_2'| \leq n.$$

Suppose for example $|r_1bs_2'| \leq n$, then $r_1bs_2' \in \aleph_{\leq n}(ubw)$ and since $ubw = y \sim_n x = uav$, we have $r_1bs_2' \in \aleph_{\leq n}(uav)$. But $r_1b \notin \aleph(u)$. Therefore $bs_2' \in \aleph(v)$, which in turn forces s_2 to be a subword of v. Thus $s = s_1as_2$ is a subword of $uav = x$, contradicting our assumption. Therefore one of the assertions $ubw \sim_n ubav$ or $uav \sim_n uabw$ must be true. Suppose for instance, $x = uav \sim_n uabw$. Then

$$\begin{aligned} |uav| + |uabw| - 2|uav \oplus uabw| &\leq |x| + |y| + 1 - 2|ua| \\ &\leq |x| + |y| + 1 - (2|x \oplus y| + 2) \\ &\leq \underbrace{|x| + |y| - 2|x \oplus y|}_{k} - 1 \\ &< k. \end{aligned}$$

By the inductive hypothesis, there exist h such that $x = uav$ is a subword of h; $uabw$ is a subword of h and $x \sim_n h \sim_n uabw$. Since y is a subword of $uabw$, we conclude $x \sim_n h \sim_n y$. *Q.E.D.*

Theorem 6-II.6 ([Sim75]).

A language $L \subseteq A^*$ is piecewise testable if and only if its syntactic monoid is $\mathcal{J}$-trivial.

Proof: By definition L is the union of classes modulo $\sim_n$ for some positive integer n. Thus L is recognized by the quotient $A^*/\sim_n$, which by Corollary 6-II.4 satisfies

$$(pq)^n p = (pq)^n = q(pq)^n$$

and therefore is $\mathcal{J}$-trivial (Proposition 3-IV.2). Since $M(L)$ divides $A^*/\sim_n$, $M(L)$ is also $\mathcal{J}$-trivial.

For the converse, let M be a $\mathcal{J}$-trivial monoid and let $L \subseteq A^*$ be a language recognized by homomorphism $\wp : A^* \mapsto M$ that we shall denote by $u \mapsto \overline{u}$. We shall show that L is the union of classes modulo $\sim_{2n-1}$, where n is the maximal length of chains of elements of M for the ordering $\leq_{\mathcal{J}}$. In other words, n is such that if

$$m_0 \leq_{\mathcal{J}} m_1 \leq_{\mathcal{J}} \cdots \leq_{\mathcal{J}} m_n$$

is a chain of $n+1$ elements of M, at least two of them are equal. Suffice it to verify the implication

$$x \sim_{2n-1} y \;\Rightarrow\; \overline{x} = \overline{y}.$$

On the basis of Theorem 6-II.5 we may take $x \in \aleph(y)$. Note further that if $x \in \aleph(h)$ and $h \in \aleph(y)$ we also have $x \sim_{2n-1} h$. This enables us to assume that $x = uv$ and $y = uav$, in which case $ua \sim_n u$ or $av \sim_n v$ (by Proposition 6-II.2). Suppose $av \sim_n v$. Then by Theorem 6-II.3, there exist $v_1, \ldots, v_n \in A^*$ such that $v = v_1 \cdots v_n$ and $\{a\} \subseteq \alpha(v_1) \subseteq \cdots \subseteq \alpha(v_n)$. Consider the chain

$$\overline{v_1 \cdots v_n} \leq_{\mathcal{J}} \ldots \leq_{\mathcal{J}} \overline{v_{n-1} v_n} \leq_{\mathcal{J}} \overline{v_n} \leq_{\mathcal{J}} 1.$$

From the choice of n there exist $i < j$ such that

$$\overline{v_i \cdots v_j \cdots v_n} = \overline{v_j \cdots v_n} = s.$$

Let $b \in \alpha(v_i)$. Then $v_i = v_i' b v_i''$ and we have

$$\overline{v_i \cdots v_n} \leq_{\mathcal{J}} \overline{b v_i'' \cdots v_n} \leq_{\mathcal{J}} \overline{v_i'' \cdots v_n} \leq_{\mathcal{J}} \overline{v_j \cdots v_n},$$

and since M is $\mathcal{J}$-trivial

$$\overline{v_i \cdots v_n} = \overline{b v_j \cdots v_n} = \overline{v_j \cdots v_n}.$$

Therefore

$$\overline{b} s = s$$

for all $b \in \alpha(v_i)$ and consequently

$$\overline{v} = \overline{v_1 \cdots v_n} = \overline{av_1 \cdots v_n} = \overline{av}.$$

From this we conclude that

$$\wp(y) = \overline{y} = \overline{uav} = \overline{uv} = \overline{x} = \wp(x).$$

Q.E.D.

Corollary 6-II.7. *For every alphabet A, $A^*\mathcal{J}$ is the boolean algebra generated by the languages of the form $A^*a_1A^*a_2A^* \cdots A^*a_nA^*$, where a_i are letters of A.*

We will have more to say about piecewise testable languages in the next chapter.

Chapter 7

Quantifier Complexity of the Straubing-Thérien Hierarchy

7-I Introduction

A language $L \subseteq A^*$ is called *star-free* if it can be constructed from finite languages by applications of boolean operations and concatenation only. Star-free expressions over a given alphabet A are built up from constants $\emptyset$, ϵ and $a \in A$ (denoting the empty set, the singleton with the empty word and the set $\{a\}$, respectively) by means of the operations $\cup$, $\cap$, $\bar{\cdot}$ (for complement with respect to A^*) and concatenation dot $\cdot$.[1] For example, $L = (ab)^*$ over $A = \{a, b\}$ is star-free since

$$(ab)^* = ((aA^* \cap A^*b) \setminus (A^*aaA^* \cup A^*bbA^*)) \cup \{\epsilon\}.$$

The aforementioned operations naturally correspond to the logical connectives $\vee$, $\wedge$, $\neg$ and $\exists$, which makes it easy to transform a star-free expression into a first-order formula. For example, over $A = \{a, b, c\}$ the expression $A^*ab\overline{(A^*aA^*)}$ defines the same language as

$$\exists x \exists y (S(x, y) \wedge Q_a x \wedge Q_b y \wedge \neg \exists z (y < z \wedge Q_a z)),$$

[1] We note that (a) the expression A^* is admitted as abbreviation of $\overline{\emptyset}$ and (b) the concatenation dot $\cdot$ is commonly omitted when the context is clear.

where $S(x,y)$ is an abbreviation of $x < y \wedge \neg\exists z(x < z \wedge z < y)$. A famous theorem due to R. McNaughton and S. Papert [MP71] shows also the converse translation:

Theorem 7-I.1. *A language is star-free if and only if it is definable in $FO[<]$.*

Within the class of star-free languages, a number of hierarchies have been considered in the literature. Below we define one of them, called the *Straubing-Thérien hierarchy*, whose levels measure the concatenation depth of defining star-free expressions. It was first implicitly suggested by D. Thérien in [Thé81] and later proposed by H. Straubing in [Str85].

For a given alphabet A we set:

$$\begin{aligned} A^*\mathcal{V}_0 &= \{\emptyset, A^*\}, \\ A^*\mathcal{V}_{k+1} &= \{L \subseteq A^* \mid\ L \text{ is a boolean combination of languages} \\ &\qquad \text{of the form } L_0a_1L_1a_2\cdots a_nL_n, (n \geq 0) \\ &\qquad L_i \in A^*\mathcal{V}_k \text{ and } a_i \in A, n \geq i \geq 0\}. \end{aligned}$$

Let

$$A^*\mathcal{V} = \bigcup_{k\geq 0} A^*\mathcal{V}_k.$$

A language $L \subseteq A^*$ is star-free if and only if there exists a non-negative integer k such that $L \in A^*\mathcal{V}_k$. The *dot-depth* of L is then the smallest such k. The levels of the Straubing-Thérien hierarchy have been characterized in terms of quantifier-prefixes of formulæ in $FO[<]$ (cf. [PP86]):

Theorem 7-I.2. *A star-free language belongs to $A^*\mathcal{V}_k$ if and only if it is definable by a boolean combination of Σ_k sentences in $FO[<]$.*

For $k \geq 1$, let us define subhierarchies of $A^*\mathcal{V}$ as follows: for all $m \geq 1$, let

$$\begin{aligned} A^*\mathcal{V}_{k,m} &= \{L \subseteq A^* \mid\ L \text{ is a boolean combination of languages} \\ &\qquad \text{of the form } L_0a_1L_1a_2\cdots a_nL_n, (m \geq n \geq 0) \\ &\qquad L_i \in A^*\mathcal{V}_{k-1} \text{ and } a_i \in A, n \geq i \geq 0\}. \end{aligned}$$

We have:

$$A^*\mathcal{V}_k = \bigcup_{m\geq 1} A^*\mathcal{V}_{k,m}\ ,$$

$$A^*\mathcal{V}_{k,m} \subseteq A^*\mathcal{V}_{k+1,m} \qquad \text{and} \qquad A^*\mathcal{V}_{k,m} \subseteq A^*\mathcal{V}_{k,m+1}.$$

In this chapter we analyze the first two levels of the Straubing-Thérien hierarchy. Level $A^*\mathcal{V}_1$ is the class of piecewise testable languages ([Sim75]) whose algebraic description was presented in the previous chapter.

7-II A subhierarchy in $A^*\mathcal{V}_1$

This section examines a logical characterization of a natural subhierarchy in $A^*\mathcal{V}_1$ based on the length of a quantifier block. By $\exists^{(k)}$ we denote the set of languages $L \subseteq A^*$ expressible by a sentence

$$\phi = \exists x_1 \exists x_2 \cdots \exists x_k \psi,$$

where ψ is a boolean combination of atomic formulæ of first-order logic in the signature with $<$ (but not S), i.e. $x = y$, $x < y$ and $Q_a x$ (see section 1-IV.3). $\mathcal{BC}(\exists^{(k)})$ denotes the set of languages corresponding to the boolean combinations of $\exists^{(k)}$-formulæ.

We begin with the case of $\mathcal{BC}(\exists)$.

7-II.1 Case of $\mathcal{BC}(\exists)$

Let $\mathbf{J_1}$ denote the variety of idempotent and commutative monoids (or *semilattices*). We denote by U_1 the monoid with two elements $\{0, 1\}$ under the usual multiplication ($1 \cdot 0 = 0 \cdot 1 = 0 \cdot 0 = 0$ and $1 \cdot 1 = 1$). We next prove the following result.

Theorem 7-II.1. *The variety* $\mathbf{J_1}$ *is generated by the monoid* U_1.

Proof: Let $\mathbf{V}$ be the variety generated by U_1: $\mathbf{V} = (U_1)$. By Theorem 3-II.2 $\mathbf{V}$ is defined by a sequence of equations. Since $U_1 \in \mathbf{J_1}$, $\mathbf{V}$ is contained in $\mathbf{J_1}$ and therefore satisfies the identities

$$xy = yx \tag{7.1}$$

$$x = x^2 \tag{7.2}$$

If $\mathbf{V} \neq \mathbf{J_1}$, we can find an equation $u = v$ satisfied by $\mathbf{V}$ which cannot be derived from 7.1 and 7.2. Let us choose such an equation with the minimal total length of u and v. Observe that in this case u and v must contain at most one occurrence of each letter since otherwise we could use identities 7.1 and 7.2 to obtain an equation equivalent to $u = v$ but shorter. Let x be a letter of u. If we take $y = 1$ for every $y \neq x$ in $u = v$, then $x = x^{|v|_x}$ and (as $x = 1$ is not an equation of U_1) we have $|v|_x > 0$. There is therefore an occurrence of x in v and the same argument shows that that every letter of v has an occurrence in u; consequently u and v contain exactly the same letters. It follows from this that $u = v$ can be easily deduced from 7.1, which contradicts the hypothesis. Therefore $\mathbf{V} = \mathbf{J_1}$. *Q.E.D.*

We denote by $\mathcal{J}_1$ the variety of languages corresponding to $\mathbf{J_1}$.

Theorem 7-II.2.

For every alphabet A, $A^*\mathcal{J}_1$ is the boolean algebra generated by the languages of the form A^*aA^* where a is a letter. Equivalently, $A^*\mathcal{J}_1$ is the boolean algebra generated by the languages of the form B^* where B is a subset of A.

Proof: The equality of the two boolean algebras in the statement results from the formulæ

$$B^* = A^* \setminus \bigcup_{a \in A \setminus B} A^*aA^* \qquad \text{and} \qquad A^*aA^* = A^* \setminus (A \setminus a)^*.$$

Since $\mathbf{J_1} = (U_1)$, we apply Theorem 3-III.2 to describe $A^*\mathcal{J}_1$. Let $\phi : A^* \mapsto U_1$ be an arbitrary homomorphism and let $B = \{a \in A |\ \phi(a) = 1\}$. Then clearly, $\phi^{-1}(1) = B^*$ and $\phi^{-1}(0) = A^* \setminus B^*$, which establishes the theorem. *Q.E.D.*

Corollary 7-II.3. $\mathcal{BC}(\exists) = \mathcal{J}_1$.

Proof: Languages in A^* defined by the sentences of $FO[<]$ of the form $\exists x\psi$, where ψ is a boolean combination of the atomic formulæ, are of the form A^*aA^* where a is a letter. *Q.E.D.*

7-II.2 The Ehrenfeucht-Fraïssé Game

We investigate the question of $\exists^{(k)}$-definability employing an interesting model-theoretic technique: the *Ehrenfeucht-Fraïssé game*. It is perhaps the most versatile method in proving non-definability in systems of first-order logic.[2]

Let ϕ be an $FO[<]$-sentence. If ϕ is a boolean combination of the Σ_k-sentences $\phi_1, \ldots, \phi_n$, we define the *quantifier rank* $qr(\phi)$ to be the maximum number of quantifiers occurring in the leading block of one of the formulæ ϕ_i.

Let $\bar{m} = (m_1, \ldots, m_k)$, where $k \geq 0$, be a sequence of positive integers. We define the set of $\bar{m}$-formulæ of $FO[<]$ by induction on k: if $k = 0$, it is the set of the boolean combinations of the atomic formulæ, and for $\bar{m} = (m, m_1, \ldots, m_k)$, an $\bar{m}$-formula is a boolean combination of formulæ

$$\exists x_1 \cdots \exists x_m \phi,$$

where ϕ is an $(m_1, \ldots, m_k)$-formula. We write $u \equiv_{\bar{m}} v$ if u and v satisfy the same $\bar{m}$-sentences of $FO[<]$. For $\bar{m} = (m_1, \ldots, m_k)$, the $\bar{m}$-formulæ of $FO[<]$ are the set of boolean combinations of formulæ ϕ such that $\phi \in \Sigma_k$ and $qr(\phi) \leq m_1$.

Let us now describe how to play the Ehrenfeucht-Fraïssé game (short: EFG). For a sequence $\bar{m} = (m_1, \ldots, m_k)$ of positive integers, where $k \geq 0$, the game $\mathcal{G}_{\bar{m}}(u,\ v)$ is played between two players called Spoiler and Duplicator (as suggested in [FSV95]) on the structures u and v. Spoiler will attempt to prove that the structures differ while Duplicator will try to show them equal. There are k rounds carried out as follows. At the i^{th} round Spoiler chooses, in u or in v, a sequence of m_i distinct positions; Duplicator, in turn, picks m_i positions in the other structure. After k rounds they concatenate positions chosen in u into a sequence $\bar{p} = p_1, \ldots, p_n$, where $n \leq \Sigma_{i=1}^{k} m_i$, and similarly positions picked in v are assembled into a sequence $\bar{q} = q_1, \ldots, q_n$. Duplicator has won the game if the map $p_i \mapsto q_i$ respects the relation $<$ and predicates Q_a, $a \in A$, i.e. if

$$p_i <^u p_j \iff q_i <^v q_j \qquad \text{and} \qquad Q_a^u p_i \iff Q_a^v q_i,$$

[2]The reader is referred to [EFT94] or [EF95] for more background.

where $a \in A$ and $1 \leq i, j \leq n$. In other words, Duplicator wins if the two subwords – one in u given by position sequence $\bar{p}$ and the other in v given by $\bar{q}$ – coincide. Otherwise Spoiler wins.

For a given number of rounds k, an initial configuration $(u,\ v)$ can be represented by the tree of height $2k$ of all possible sequences of play. Since a game cannot end in a tie, we mark the leaves of this tree according to the winner: 'S' or 'D'. The interior nodes are then labelled recursively (beginning at the leaves) in the following manner. A node corresponding to a play by Spoiler is labelled 'S' if and only if it has a child marked 'S'; it is labelled 'D' otherwise. A node corresponding to the Duplicator's move is labelled 'D' if and only if at least one of its children is labelled 'D'; otherwise it is marked 'S'. A label at the root thus determines who has a winning strategy. The fact that Duplicator possesses a winning strategy in $\mathcal{G}_{\bar{m}}(u,\ v)$ is denoted $u \sim_{\bar{m}} v$. Naturally, $u \sim_{\bar{m}} v$ defines a congruence on A^* which we denote also by $u \sim_{\bar{m}} v$.

The above version of the Ehrenfeucht-Fraïssé game was proposed by W. Thomas in [Tho84]. The original EFG is the special case of $\mathcal{G}_{\bar{m}}(u,\ v)$ with $\bar{m} = (1, \ldots, 1)$. Fraïssé showed in the 1950's that for a non-negative integer r the relations $\equiv_r$ and $\sim_r$ coincide on relational structures of finite signature; later Ehrenfeucht introduced the game theoretical formulation of $\sim_r$. J. G. Rosenstein [Ros82] and R. Fraïssé [Fra72] contain a more detailed discussion of model-theoretic games.

For our version of the game we have the following result which will be used as a tool in the next section.

Theorem 7-II.4 (Ehrenfeucht-Fraïssé Theorem, [Ehr61]).
For all $\bar{m} = (m_1, \ldots, m_k)$ with $k \geq 1$ and $m_i \geq 1$, $(1 \leq i \leq k)$, we have:

$$u \equiv_{\bar{m}} v \iff u \sim_{\bar{m}} v.$$

7-II.3 Application of EFG to $\mathcal{BC}(\exists^{(k)})$

The following is a very trivial application of the powerful EFG idea.

Theorem 7-II.5.

$$\exists^{(m+1)} \not\subseteq \mathcal{BC}(\exists^{(m)})$$

Proof: Let $A = \{a, b\}$. Consider the language $L = A^*(aA^*)^{m+1}$. $L \in \exists^{(m+1)}$ since

$$L = \exists x_1 \cdots \exists x_{m+1} \bigwedge_{i=1}^{m} (x_i < x_{i+1}) \bigwedge_{i=1}^{m+1} Q_a x_i.$$

We claim that $L \notin \mathcal{BC}(\exists^{(m)})$. Suppose the claim is false, i.e. there exists a sentence $\phi \in \mathcal{BC}(\exists^{(m)})$ that defines L. Then $u \models \phi$ if and only if $|u|_a \geq m+1$. Since a one-round game $\mathcal{G}_{(m)}(a^{m+1},\ a^m)$ is easily won by Duplicator who just picks m positions in the available structure, we have for all $m > 0$,

$$u = a^{m+1} \sim_{(m)} a^m = v$$

and thus $v \models \phi$. But $|v|_a = m$, a contradiction. *Q.E.D.*

7-II.4 Connection with matrices

We denote by M_n, $n \geq 1$, the set of all $n \times n$ matrices over the boolean semiring $B = \{0, 1\}$ (where 1+1=1) and by K_n – the set of all upper triangular matrices in M_n all of whose diagonal entries equal 1. That is,

$$K_n = \left\{ m \in M_n \;\middle|\; \begin{array}{rcll} m_{ij} & = & 0 & \text{for } 1 \leq j < i \leq n \\ m_{ii} & = & 1 & \text{for } 1 \leq i \leq n \end{array} \right\}.$$

Since K_n is closed under multiplication of matrices, it is a submonoid of the multiplicative monoid M_n. We set $\mathbf{U}$ to be the set of all finite monoids that are divisors of K_n for certain n:

$$\mathbf{U} = \{M |\ \exists\, n \in \mathbb{N} : M \prec K_n\}.$$

It's easy to see that $\mathbf{U}$ is closed under division. It is also closed under direct product. To establish this, consider an injective homomorphism $\phi : M_m \times M_n \mapsto M_{m+n}$ given

by

$$\phi_{ij}(p,\ q) = \begin{cases} 0 & \text{if } i \leq m \quad \text{and } j > m, \\ 0 & \text{if } i > m \quad \text{and } j \leq m, \\ p_{ij} & \text{if } i \leq m \quad \text{and } j \leq m, \\ q_{i-m,j-m} & \text{if } i > m \quad \text{and } j > m. \end{cases}$$

Observe that ϕ embeds $K_m \times K_n$ into K_{m+n}:

$$(p_{i,j}) \times (q_{i,j}) \mapsto \begin{pmatrix} (p_{i,j}) & | & 0 \\ ----- & + & ----- \\ 0 & | & (q_{i,j}) \end{pmatrix}$$

Since

$$N_1 \prec M_1 \wedge N_2 \prec M_2 \ \Rightarrow\ N_1 \times N_2 \prec M_1 \times M_2$$

we conclude that $\mathbf{U}$ contains the direct product of any two of its members. Thus we have:

Lemma 7-II.6. $\mathbf{U}$ *is a variety of finite monoids.*

If $\mathbf{V}$ is a variety of finite monoids and A is a finite alphabet, then we denote by $A^*\mathcal{V}$ the family of all recognizable languages in A^* whose syntactic monoid belongs to $\mathbf{V}$. It is well known that every variety of finite monoids is generated by the syntactic monoids it contains (cf. [Eil76], Ch. VII). Thus if $\mathbf{V}_1$, $\mathbf{V}_2$ are varieties of finite monoids, $\mathbf{V}_1 \subseteq \mathbf{V}_2$ if and only if $A^*\mathcal{V}_1 \subseteq A^*\mathcal{V}_2$ for every finite alphabet A. This enables us to demonstrate that two varieties are equal by showing that the corresponding families of recognizable languages are equal. In case of the variety $\mathbf{U}$ we have the following theorem.

Theorem 7-II.7 (Straubing, [Str80]).
$A^*\mathcal{U}$ is the boolean closure of the family of languages of the form $A^*a_1A^* \cdots a_nA^*$, where $a_i \in A$, $1 \leq i \leq n$.

Theorem 6-II.6 (The Theorem of Simon) and Corollary 6-II.7 assert that this is precisely the family of languages whose syntactic monoids belong to the variety $\mathbf{J}$,

yielding the equality:

$$\mathbf{U} = \mathbf{J}.$$

This class of languages is also defined by

$$\bigcup_{n\geq 0} \mathcal{BC}(\exists^{(n)})$$

The next theorem summarizes our analysis of the first level of the Straubing-Thérien hierarchy.

Theorem 7-II.8. *The following conditions are equivalent:*

1. $L \in A^*\mathcal{V}_1$
2. $L \in \mathcal{BC}(\exists^{(k)})$ *for some* $k \geq 0$
3. $M(L)$ *is* $\mathcal{J}$*-trivial*
4. $M(L) \prec K_n$ *for some* $n \geq 1$

7-III Characterization of $A^*\mathcal{V}_2$

The *polynomial closure* of a class of languages $\mathcal{L}$ of A^* is the set of languages that are finite unions of languages of the form

$$L_0 a_1 L_1 \cdots a_n L_n,$$

where the a_i's are the letters and the L_i's are elements of $\mathcal{L}$. By extension, if $\mathcal{V}$ is a $*$-variety, we denote by Pol$\mathcal{V}$ the class of languages such that, for every alphabet A, A^*Pol$\mathcal{V}$ is the polynomial closure of $A^*\mathcal{V}$. We also denote by Co-Pol$\mathcal{V}$ the class of languages such that, for every alphabet A, A^*Co-Pol$\mathcal{V}$ is the set of languages L whose complement is in A^*Pol$\mathcal{V}$. Finally, we denote by BPol$\mathcal{V}$ the class of languages such that, for every alphabet A, A^*BPol$\mathcal{V}$ is the closure of A^*Pol$\mathcal{V}$ under finite boolean operations (finite union and complement).

Example 7-III.1. Let $\mathcal{B}$ be a class of languages defined, for every alphabet A, by $A^*\mathcal{B} = \{B^* | B \subseteq A\}$. Then $\mathrm{Pol}(A^*\mathcal{B})$ is a finite union of languages of the form

$$A_0^* a_1 A_1^* \cdots a_n A_n^*, \tag{7.3}$$

where n is a non-negative integer, A_i's are subsets of A for $i \in \{0, \ldots, n\}$ and a_i's are letters for $i \in \{1, \ldots, n\}$.

The marked product $L = L_0 a_1 L_1 \ldots a_n L_n$ of n languages L_0, $L_1, \ldots, L_n$ of A^* is *unambiguous* if every word u of L admits a unique factorization of the form $u_0 a_1 u_1 \cdots a_n u_n$ with $u_0 \in L_0$, $u_1 \in L_1, \ldots, u_n \in L_n$.

The *unambiguous polynomial closure* of a class of languages $\mathcal{L}$ of A^* is the set of languages that are finite disjoint unions of unambiguous products of the form $L_0 a_1 L_1 \ldots a_n L_n$, where a_i's are letters and L_i's are elements of $\mathcal{L}$. Again, by extension, if $\mathcal{V}$ is a variety of languages, we denote by $\mathrm{UPol}\mathcal{V}$ the class of languages such that, for every alphabet A, $A^*\mathrm{UPol}\mathcal{V}$ is the unambiguous polynomial closure of $A^*\mathcal{V}$.

Recall (section 3-IV) that **DA** is the variety of finite aperiodic monoids whose regular $\mathcal{J}$-classes are idempotent semigroups (or rectangular bands). We denote by $\mathcal{DA}$ the corresponding variety of languages. By Proposition 3-IV.6, **DA** is defined by the identities

$$(xy)^\omega (yx)^\omega (xy)^\omega = (xy)^\omega \qquad \text{and} \qquad x^\omega = x^{\omega+1} \tag{7.4}$$

We next prove the following result.

Lemma 7-III.1. $\mathbf{J_1} \square \mathbf{J_1} \subseteq \mathbf{DA}$

Proof: Let $N, M \in \mathbf{J_1}$ be two arbitrary monoids. The block product $N \square M$ is isomorphic to a bilateral semidirect product $V * * M$, where V is idempotent and commutative. Let $x = (v_1, m_1)$ and $y = (v_2, m_2)$ be elements of $V * * M$. Then using

additive notation for V we obtain:

$$\begin{aligned}
xy &= (v_1m_2 + m_1v_2,\ m_1m_2)\\
xyx &= (v_1m_1m_2 + m_1v_2m_1 + m_1m_2v_1,\ m_1m_2)\\
(xy)^2 &= (v_1m_1m_2 + m_1v_2m_1m_2 + m_1m_2v_1m_2 + m_1m_2v_2,\ m_1m_2)\\
(xy)^2x &= (v_1m_1m_2 + m_1v_2m_1m_2 + m_1m_2v_1m_1m_2\\
&\quad + m_1m_2v_2m_1 + m_1m_2v_1,\ m_1m_2)\\
(xy)^3 &= (v_1m_1m_2 + m_1v_2m_1m_2 + m_1m_2v_1m_1m_2\\
&\quad + m_1m_2v_2m_1m_2 + m_1m_2v_1m_2 + m_1m_2v_2,\ m_1m_2)\\
(xy)^\omega &= (v_1m_1m_2 + m_1v_2m_1m_2 + m_1m_2v_1m_1m_2\\
&\quad + m_1m_2v_2m_1m_2 + m_1m_2v_1m_2 + m_1m_2v_2,\ m_1m_2).
\end{aligned}$$

Similarly,

$$\begin{aligned}
yx &= (v_2m_1 + m_2v_1,\ m_1m_2)\\
yxy &= (v_2m_1m_2 + m_2v_1m_2 + m_1m_2v_2,\ m_1m_2)\\
(yx)^2 &= (v_2m_1m_2 + m_2v_1m_1m_2 + m_1m_2v_2m_1 + m_1m_2v_1,\ m_1m_2)\\
(yx)^\omega &= (v_2m_1m_2 + m_2v_1m_1m_2 + m_1m_2v_2m_1m_2\\
&\quad + m_1m_2v_1m_1m_2 + m_1m_2v_2m_1 + m_1m_2v_1,\ m_1m_2).
\end{aligned}$$

And finally,

$$\begin{aligned}
(xy)^\omega(yx)^\omega &= (v_1m_1m_2 + m_1v_2m_1m_2 + m_1m_2v_1m_1m_2\\
&\quad + m_1m_2v_2m_1m_2 + m_1m_2v_2m_1 + m_1m_2v_1,\ m_1m_2)\\
(xy)^\omega(yx)^\omega(xy)^\omega &= (v_1m_1m_2 + m_1v_2m_1m_2 + m_1m_2v_1m_1m_2\\
&\quad + m_1m_2v_2m_1m_2 + m_1m_2v_1m_2 + m_1m_2v_2,\ m_1m_2)\\
&= (xy)^\omega.
\end{aligned}$$

Clearly, $(xy)^\omega = (xy)^{\omega+1}$. Thus

$$\mathbf{J_1}\Box\mathbf{J_1} \subseteq [\![x^\omega = x^{\omega+1} \text{ and } (xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega]\!]$$

and therefore by Proposition 3-IV.6, $\mathbf{J_1}\Box\mathbf{J_1} \subseteq \mathbf{DA}$. *Q.E.D.*

Theorem 7-III.2.

$$\mathcal{BC}(\exists\forall) \subseteq \mathcal{DA}.$$

Proof: Let L be a language in $\mathcal{BC}(\exists\forall)$. Since $L \in FO[<]$, $M(L)$ is aperiodic (Theorem 5-III.4) and by Theorem 4-V.1 $M(L) \prec U_1 \Box M'$, where M' again contains no non-trivial subgroups. $U_1 \Box M'$ is isomorphic to $V * * M'$, where V is idempotent and commutative. Let

$$\gamma : A^* \mapsto V * * M'$$

be a homomorphism such that $L = \gamma^{-1}(T)$ for some $T \subseteq V * * M$. Then

$$L = \bigcup_{(v,m)\in T} \gamma^{-1}(v, m).$$

Let $w \in A^*$. By hypothesis L is defined by a boolean combination of the sentences

$$\phi = \exists x \forall y \psi,$$

where ψ is a boolean combination of the atomic formulæ. Suppose ψ is in conjunctive normal form. If ψ contains a clause $Q_a x$, where a is a letter, then $w \models \phi$ if and only if w admits a factorization $w'aw''$ with

$$\begin{aligned} w' &\models \forall y((y < x) \wedge \psi), \\ w'' &\models \forall y((y > x) \wedge \psi). \end{aligned} \tag{7.5}$$

If ψ does not contain a clause $Q_a x$, then a position x may be occupied by any letter s of the alphabet and hence, $w \models \phi$ if and only if $w = w's w''$ and the condition above holds.

Therefore $w \in L$ if and only if it satisfies a boolean combination of the conditions of the form

$$w = w' b w'',$$

where $b \in A$ and w', w'' satisfy 7.5.

Let π be a projection homomorphism from $V**M'$ onto M'. Then $\pi\circ\gamma(w') = m' \in M'$ and $\pi \circ \gamma(w'') = m'' \in M'$. Since w' and w'' are in $\mathcal{BC}(\exists)$, languages recognized by M' belong to $A^*\mathcal{J}_1$ (by Corollary 7-II.3) and hence $M' \in \mathbf{J_1}$. Applying Lemma 7-III.1 we have the desired result. *Q.E.D.*

Let M_n and K_n be the sets of matrices defined in section 7-II.4. Consider the family T_n of all *upper triangular* $n \times n$ matrices over the semiring $B = \{0, 1\}$. T_n is a submonoid of M_n which contains K_n. We set

$$\mathbf{W} = \{M | \exists n \in I\!N : M \prec T_n\}.$$

Evidently $\mathbf{W}$ is closed under division. Observe also that the homomorphism ϕ : $M_m \times M_n \mapsto M_{m+n}$ defined in section 7-II.4 maps $T_m \times T_n$ into T_{m+n} and therefore $\mathbf{W}$ is closed under product. Thus $\mathbf{W}$ is a variety of finite monoids.

The next theorem describes the family of recognizable languages corresponding to $\mathbf{W}$.

Theorem 7-III.3 ([PS81]).
$A^*\mathcal{W}$ is the boolean closure of the family of languages of the form $A_0^*a_1A_1^*\cdots a_kA_k^*$, where $k \geq 0$ and for $1 \leq i \leq k$: $a_i \in A$ and A_i are (possibly empty) subsets of A. If $A_i = \emptyset$, then $A_i^* = \{1\}$.

Proof: Let $\mathcal{F}$ denote the boolean closure of the family of languages of the form specified in the theorem. We first show that $\mathcal{F} \subseteq A^*\mathcal{W}$.

Since $A^*\mathcal{V}$ is closed under boolean operations for any variety $\mathbf{V}$ it suffices to show that the syntactic monoid of any language of the form

$$L = A_0^*a_1A_1^*\cdots a_kA_k^*$$

is in $\mathbf{W}$. We shall show that L is recognized by the monoid T_{k+1} by exhibiting a homomorphism $\psi : A^* \mapsto T_{k+1}$ and a set $X \subseteq T_{k+1}$ such that $\psi^{-1}(X) = L$:

$$\psi_{ij}(a) = \begin{cases} 1 & \text{if} \quad i = j \quad \text{and} \quad a \in A_{i-1}, \\ 1 & \text{if} \quad j = i+1 \quad \text{and} \quad a = a_i, \\ 0 & \text{otherwise} \end{cases}$$

for all $a \in A$, $i, j \in \{1, \ldots, k+1\}$. It is easy to verify that if $w \in A^*$ then $\psi_{ij}(w) = 1$ if and only if there is a path labelled w from state i to state j in the nondeterministic automaton pictured in the figure 7.1.

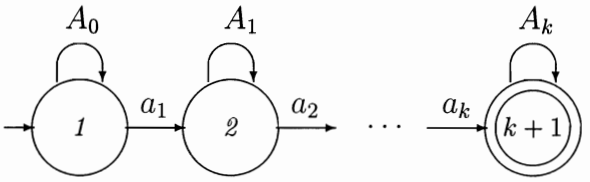

Figure 7.1: An automaton recognizing $L = A_0^* a_1 A_1^* \cdots a_k A_k^*$.

In particular,

$$\psi_{1,k+1}(w) = 1 \iff w \in A_0^* a_1 A_1^* \cdots a_k A_k^* = L.$$

Thus $L = \psi^{-1}(X)$, where $X = \{m \in T_{k+1} |\ m_{1,k+1} = 1\}$.

To prove the inclusion $A^*\mathcal{W} \subseteq \mathcal{F}$ let us suppose that $L \in A^*\mathcal{W}$. Then $M(L) \in \mathbf{W}$ and there exists $n \geq 1$ such that L is recognized by T_n. This in turn means there exist a homomorphism $\eta : A^* \mapsto T_n$ and a subset X of T_n with the property $L = \eta^{-1}(X)$. We need to show that $\eta^{-1}(X) \in \mathcal{F}$. Since

$$\eta^{-1}(X) = \bigcup_{x \in X} \eta^{-1}(x)$$

and since $\mathcal{F}$ is closed under boolean operations, it suffices to prove that $\eta^{-1}(x) \in \mathcal{F}$ for each $x \in T_n$.

$$\begin{aligned} \eta^{-1}(x) &= \bigcap_{1 \leq i,j \leq n} \{w |\ \eta_{ij}(w) = x_{ij}\} \\ &= \bigcap_{\{(i,j) |\ x_{ij}=1\}} \{w |\ \eta_{ij}(w) = 1\} \setminus \bigcup_{\{(i,j) |\ x_{ij}=0\}} \{w |\ \eta_{ij}(w) = 1\}. \end{aligned}$$

Thus it suffices to show that each set of the form $\{w |\ \eta_{ij}(w) = 1\}$ belongs to $\mathcal{F}$. Let $A_{k,l} = \{a \in A |\ \eta_{kl}(a) = 1\}$ and let Q_{ij} be the set of all strictly increasing sequences $(i_0, \ldots, i_t)$ such that $i_0 = i$ and $i_t = j$.[3] Then

$$\{w |\ \eta_{ij}(w) = 1\} = \bigcup_{(i_0, \ldots, i_t) \in Q_{ij}} A_{i_0 i_0}^* A_{i_0 i_1} A_{i_1 i_1}^* \cdots A_{i_{k-1} i_k} A_{i_k i_k}^*.$$

[3] If $i > j$ then Q_{ij} is empty. If $i = j$ then Q_{ij} consists of the single sequence (i).

Since each language $A^*_{i_0 i_0} A_{i_0 i_1} A^*_{i_1 i_1} \cdots A_{i_{k-1} i_k} A^*_{i_k i_k}$ is a finite union of the form $A^*_{i_0 i_0} a_1 A^*_{i_1 i_1} \cdots a_k A^*_{i_k i_k}$, it follows that $\{w |\ \eta_{ij}(w) = 1\} \in \mathcal{F}$. *Q.E.D.*

Let $\mathcal{L}$ and $\mathcal{R}$ be the varieties of languages corresponding to the varieties of $\mathcal{L}$-trivial and $\mathcal{R}$-trivial monoids respectively; and let $\mathcal{B}$ be the family of languages defined in Example 7-III.1.

Theorem 7-III.4 ([Arf91]).

$$Pol(\mathcal{J}_1) = Pol(\mathcal{J}) = Pol(\mathcal{R}) = Pol(\mathcal{L}) = Pol(\mathcal{DA}) = Pol(\mathcal{B}).$$

Proof: From the definitions of the corresponding varieties of monoids we have two series of inclusions:

$$\mathbf{J_1} \subset \mathbf{J} \subset \mathbf{R} \subset \mathbf{DA} \qquad \text{and} \qquad \mathbf{J_1} \subset \mathbf{J} \subset \mathbf{L} \subset \mathbf{DA}. \tag{7.6}$$

Since $A^*\mathcal{J}_1$ is the boolean algebra generated by $A^*\mathcal{B}$ (Theorem 7-II.2), we have the inclusion

$$\mathrm{Pol}(A^*\mathcal{B}) \subseteq \mathrm{Pol}(A^*\mathcal{J}_1). \tag{7.7}$$

Schützenberger showed (cf. [Sch76]) that every language of $A^*\mathcal{DA}$ is a finite disjoint union of languages of the form

$$A^*_0 a_1 A^*_1 \ldots a_n A^*_n, \tag{7.8}$$

where $n \geq 0$, $A_0, \ldots, A_n \subseteq A$, $a_1, \ldots, a_n \in A$ and where the product $A^*_0 a_1 A^*_1 \ldots a_n A^*_n$ is unambiguous. The form 7.8 is the same as 7.3 with an added restriction. From this result we conclude that

$$A^*\mathcal{DA} \subseteq \mathrm{Pol}(A^*\mathcal{B}) \tag{7.9}$$

and therefore

$$\mathrm{Pol}(A^*\mathcal{DA}) \subseteq \mathrm{Pol}(A^*\mathcal{B}). \tag{7.10}$$

The theorem then follows from the inclusions 7.6, 7.7 and 7.10. *Q.E.D.*

Theorem 7-III.5 ([Tho82], [PP86]).

1. *A language is in $\mathcal{BC}(\Sigma_k)$ if and only if it is in $\mathcal{V}_k$.*
2. *A language is in Σ_{k+1} if and only if it is in $Pol\mathcal{V}_k$.*
3. *A language is in Π_{k+1} if and only if it is in Co-$Pol\mathcal{V}_k$.*

Theorem 7-III.6. *The following conditions are equivalent:*

1. $L \in A^*\mathcal{V}_2$
2. $L \in \mathcal{BC}(\Sigma_2)$
3. $M(L) \prec T_n$ *for some* $n \geq 1$

Proof: (1 $\iff$ 3). By definition $A^*\mathcal{V}_2$ is the boolean-polynomial closure of $A^*\mathcal{V}_1$, that is of the family of piecewise testable languages. Applying Theorem 6-II.6 (The Theorem of Simon) and Corollary 6-II.7 we obtain $A^*\mathcal{V}_2 = \mathrm{BPol}(\mathcal{J})$. It follows from Theorem 7-III.4 that a language L belongs to $A^*\mathcal{V}_2$ if and only if L is a boolean combination of the languages of the form

$$A_0^* a_1 A_1^* \cdots a_n A_n^*,$$

where n is a non-negative integer, A_i's are subsets of A for $i \in \{0, \ldots, n\}$ and a_i's are letters for $i \in \{1, \ldots, n\}$. By Theorem 7-III.3, this family of languages corresponds to the variety of monoids $\mathbf{W}$. Therefore a language $L \in A^*\mathcal{V}_2$ if and only if there exists $n \in I\!N$ such that $M(L) \prec T_n$.

(1 $\iff$ 2) is a particular case of the first condition of Theorem 7-III.5. *Q.E.D.*

Corollary 7-III.7. *For any non-negative integer k,*

$$\mathcal{BC}(\exists^{(k)}\forall) = \mathcal{BC}(\exists^{(k)}\forall^*).$$

Proof: The set of $\exists^{(k)}\forall^*$-sentences consists of $\exists^{(k)}\forall^+$-sentences (which are in Σ_2) and $\exists^{(k)}$-sentences (which are in Σ_1). Let $\phi = \exists^{(k)}\psi$ be a sentence of $FO[<]$ and y – a variable that does not appear in ϕ. We then construct a new sentence $\phi' = \exists^{(k)}\forall y(\psi \wedge (y = y))$. By Theorem 1-IV.1, we have for all $w \in A^*$

$$w \models \phi \iff w \models \phi'.$$

Thus every Σ_1-sentence can be transformed into an equivalent Σ_2-sentence. So we need to show

$$\mathcal{BC}(\exists^{(k)}\forall) = \mathcal{BC}(\exists^{(k)}\forall^+). \tag{7.11}$$

It follows from Theorems 7-III.6 and 7-III.3 that languages in $\mathcal{BC}(\exists^{(k)}\forall^+)$ are recognized by the monoid T_{k+1}. By Theorem 3-III.1, there is a bijection between the variety of monoids $\mathbf{W}$, generated by T_n, and the corresponding variety of languages $\mathcal{W}$. Therefore for a fixed integer k and any positive integer m, all sentences of the form $\mathcal{BC}(\exists^{(k)}\forall^{(m)})$ define the same subset of A^*, which proves the equality 7.11. *Q.E.D.*

Theorem 7-III.8.

$$\mathcal{DA} \subseteq \Sigma_2 \cap \Pi_2.$$

Proof: By Theorem 7-III.5 $\Sigma_2 = \mathrm{Pol}\mathcal{V}_1$, and by Theorem 7-II.8 $\mathcal{V}_1 = \mathcal{J}$, finally by Theorem 7-III.4, $\mathcal{DA} \subseteq \mathrm{Pol}(\mathcal{J})$. Therefore,

$$\mathcal{DA} \subseteq \Sigma_2.$$

By Theorem 7-III.5, $\Pi_2 = \text{Co-Pol}\mathcal{V}_1$. Since $\mathcal{DA}$ is a variety of languages, it is closed under complementation and hence

$$\mathcal{DA} \subseteq \Pi_2.$$

Thus $\mathcal{DA}$ is contained in both Σ_2 and Π_2. *Q.E.D.*

We prove the opposite inclusion in the next chapter, where we extend the "standard" notions of finite monoids, homomorphisms and variety and introduce a new operation on varieties – the Mal'cev product.

Chapter 8

Ordered monoids and positive varieties

8-I Introduction

The most important tool for classifying the regular languages is Eilenberg's variety theorem (Theorem 3-III.1), which gives a one-to-one correspondence between varieties of finite monoids and varieties of regular languages.

Certain families of regular languages, which are not varieties of languages, also admit a syntactic characterization. J. E. Pin ([Pin95]) showed that such results are not isolated, but are as general as Eilenberg's theorem.

This chapter introduces *positive* varieties of languages, which have the same properties as varieties of languages, but need not be closed under complement. Positive varieties are in bijection with varieties of finite ordered monoids. We shall show that the polynomial closure of a variety of languages is a positive variety. This property will be exploited to find new connections between classes of languages and the logical hierarchy Σ_k and in particular, to prove the opposite direction of Theorem 7-III.8.

8-II Ordered monoids

An *ordered monoid* $(M, \leq)$ is a monoid M equipped with an order relation $\leq$ such that, for every $u, v, x \in M$,

$$u \leq v \Rightarrow ux \leq vx \wedge xu \leq xv.$$

The ordered monoid $(M, \geq)$ is called the *dual* of $(M, \leq)$.

An *order ideal* of $(M, \leq)$ is a subset I of M such that, if $x \leq y$ and $y \in I$, then $x \in I$.

A *homomorphism of ordered monoids* $\phi : (M, \leq) \mapsto (N, \leq)$ is a monoid homomorphism from M to N such that, for every $x, y \in M$,

$$x \leq y \Rightarrow \phi(x) \leq \phi(y) \qquad \text{and} \qquad \phi(1) = 1.$$

A monoid M can be regarded as an ordered monoid with $=$ as order relation.

Order ideals are closed under union, intersection, inverse homomorphisms and residual (cf. [Pin95]).

An ordered monoid $(M, \leq)$ is an *ordered submonoid of* $(N, \leq)$ if M is a submonoid of N and the order on M is the restriction to M of the order on N.

An ordered monoid $(N, \leq)$ is an *ordered quotient* of $(M, \leq)$ if there exists a surjective homomorphism of ordered monoids $\phi : (M, \leq) \mapsto (N, \leq)$. For example, any ordered monoid $(M, \leq)$ is a quotient of $(M, =)$. An ordered monoid $(M, \leq)$ *divides* an ordered monoid $(N, \leq)$ if $(M, \leq)$ is an ordered quotient of an ordered submonoid of $(N, \leq)$.

Given a family $(M_i, \leq)_{i \in I}$ of ordered monoids, the product $\prod_{i \in I}(M_i, \leq)$ is the ordered monoid defined on the set $\prod_{i \in I} M_i$ by the law

$$(s_i)_{i \in I}(s'_i)_{i \in I} = (s_i s'_i)_{i \in I}$$

and the order given by

$$(s_i)_{i\in I} \leq (s'_i)_{i\in I} \iff \forall i \in I : s_i \leq s'_i.$$

Let A be a finite set and let A^* be the free monoid on A. Then $(A^*, =)$ is an ordered monoid. As the following lemma shows, it is in fact the free ordered monoid on A.

Lemma 8-II.1 ([Pin95]). *Let $\phi : A \mapsto M$ be a function from A into an ordered monoid $(M, \leq)$. Then there exists a unique homomorphism of ordered monoids $\bar{\phi} : (A^*, =) \mapsto (M, \leq)$ such that for all $a \in A$, $\phi(a) = \bar{\phi}(a)$.*

Proof: Since A^* is the free monoid on A, there exists a unique homomorphism $\bar{\phi} : A^* \mapsto M$, such that $\phi(a) = \bar{\phi}(a)$ for every $a \in A$. Hence, if $u = v$, then $\phi(u) = \phi(v)$ and thus $\phi(u) \leq \phi(v)$. Therefore ϕ is a homomorphism of ordered monoids. *Q.E.D.*

A *variety of finite ordered monoids* is a class of finite ordered monoids closed under the operations of taking ordered submonoids, ordered quotients and finite products.

If $\mathbf{V}$ is a variety of finite monoids, the class of all ordered monoids of the form $(M, \leq)$, where $M \in \mathbf{V}$, is a variety of ordered monoids, called the variety of ordered monoids *generated* by $\mathbf{V}$ and also denoted $\mathbf{V}$. It will be clear from the context whether $\mathbf{V}$ is a variety of finite monoids or a variety of finite ordered monoids.

Given a variety of finite ordered monoids, the class of all duals of members of $\mathbf{V}$ form a variety of finite ordered monoids, called the *dual* of $\mathbf{V}$, denoted $\breve{\mathbf{V}}$.

Recall (section 3-II) that a finite monoid M satisfies the identity $x = y$, where $x, y \in \widehat{A^*}$, if, for every continuous homomorphism $\phi : \widehat{A^*} \mapsto M$, $\phi(x) = \phi(y)$. Similarly, a finite ordered monoid $(M, \leq)$ *satisfies the identity* $x \leq y$ if, for every continuous homomorphism $\phi : \widehat{A^*} \mapsto M$, $\phi(x) \leq \phi(y)$. Again, it should be clear from the context which sense of "identity" is intended.

8-III Relational homomorphisms and Mal'cev products

A *relational homomorphism* between monoids M and N is a relation $\tau : M \mapsto N$ such that

1. $\tau(s)\tau(t) \subseteq \tau(st)$ for all $s, t \in M$,
2. $\tau(s)$ is non-empty for all $s \in M$,
3. $1 \in \tau(1)$.

Let $\mathbf{V}$ be a variety of monoids and $\mathbf{W}$ be a variety of semigroups. The *Mal'cev product* $\mathbf{W}\textcircled{M}\,\mathbf{V}$ is the class of all monoids M such that there exists a relational homomorphism $\tau : M \mapsto V$ with $V \in \mathbf{V}$ and $\tau^{-1}(e) \in \mathbf{W}$ for each idempotent e of V. It's easy to see that $\mathbf{W}\textcircled{M}\,\mathbf{V}$ is a variety of monoids.

In a more general sense, if $\mathbf{V}$ is a variety of monoids and $\mathbf{W}$ is a variety of ordered semigroups, the *Mal'cev product* $\mathbf{W}\textcircled{M}\,\mathbf{V}$ is the class of all ordered monoids $(M, \leq)$ such that there exists a relational homomorphism $\tau : M \mapsto V$ with $V \in \mathbf{V}$ and $\tau^{-1}(e) \in \mathbf{W}$ for each idempotent e of V. Then we have:

Theorem 8-III.1. $\mathbf{W}\textcircled{M}\,\mathbf{V}$ *is a variety of ordered monoids.*

Proof: We have to show that $\mathbf{W}\textcircled{M}\,\mathbf{V}$ is closed under taking ordered submonoids, ordered quotients and finite products. If $(N, \leq)$ is an ordered submonoid of $(M, \leq) \in \mathbf{W}\textcircled{M}\,\mathbf{V}$, then N is a submonoid of M and hence the restriction of τ to N satisfies the required condition.

If $(N, \leq)$ is an ordered quotient of $(M, \leq) \in \mathbf{W}\textcircled{M}\,\mathbf{V}$, then there exists a surjective homomorphism of ordered monoids $\phi : (M, \leq) \mapsto (N, \leq)$. Then $M = \phi^{-1}(N)$ and there exists a relational homomorphism $\tau' = \tau \circ \phi^{-1}$ with the stated property.

Finally, let $(M_i, \leq)_{i \in I}$ be a finite family of ordered monoids such that for all $i \in I$, $(M_i, \leq) \in \mathbf{W}\textcircled{M}\,\mathbf{V}$. By definition of the relational homomorphism,

$$\prod_{i \in I} \tau(M_i) \subseteq \tau(\prod_{i \in I} M_i)$$

and therefore, $\tau : \prod_{i \in I} M_i \mapsto V$ is a relational morphism with $V \in \mathbf{V}$ and $\tau^{-1}(e) \in \mathbf{W}$ for each idempotent e of V. *Q.E.D.*

We now describe the defining set of identities for a Mal'cev product.

Theorem 8-III.2 ([PW95]). *Let* $\mathbf{V}$ *be a variety of monoids and* $\mathbf{W}$ *– a variety of ordered semigroups. Let* E *be the set of identities such that* $\mathbf{W} = [\![E]\!]$. *Then* $\mathbf{W} \textcircled{M} \mathbf{V}$ *is defined by the identities of the form* $\sigma(x) \leq \sigma(y)$, *where* $x \leq y$ *is an identity of* E *with* $x, y \in \widehat{B^*}$ *for some finite alphabet* B *and* $\sigma : \widehat{B^*} \mapsto \widehat{A^*}$ *is a continuous homomorphism such that, for all* $b, b' \in B$, $\mathbf{V}$ *satisfies* $\sigma(b) = \sigma(b') = \sigma(b^2)$.

Recall (Example 3-II.2 on page 42) that $\mathbf{LI} = [\![x^\omega y x^\omega = x^\omega]\!]$ is the variety of locally trivial semigroups.

Corollary 8-III.3. *Let* $\mathbf{V}$ *be a variety of monoids. Then* $\mathbf{LI} \textcircled{M} \mathbf{V}$ *is defined by the identities of the form* $x^\omega y x^\omega = x^\omega$, *where* $x, y \in \widehat{A^*}$ *for some finite set* A *and* $\mathbf{V}$ *satisfies* $x = y = x^2$.

Proof: follows from Theorem 8-III.2.

Corollary 8-III.4. *Let* $\mathbf{V}$ *be a variety of monoids. Then* $[\![x^\omega y x^\omega \leq x^\omega]\!] \textcircled{M} \mathbf{V}$ *is defined by the identities of the form* $x^\omega y x^\omega \leq x^\omega$, *where* $x, y \in \widehat{A^*}$ *for some finite set* A *and* $\mathbf{V}$ *satisfies* $x = y = x^2$.

Proof: follows immediately from Theorem 8-III.2.

8-IV Syntactic ordered monoids

Let $(M, \leq)$ be an ordered monoid and let $\eta : (M, \leq) \mapsto (N, \leq)$ be a surjective homomorphism of ordered monoids. An order ideal Q of M is said to be *recognized* by η if there exists an order ideal P of N such that $Q = \eta^{-1}(P)$. Observe that this condition implies $\eta(Q) = \eta(\eta^{-1}(P)) = P$. By extension, the order ideal Q of M is said to be *recognized* by $(N, \leq)$ if there exists a surjective homomorphism of ordered

monoids from $(M, \leq)$ onto $(N, \leq)$ that recognizes Q. This definition can be applied in particular to languages. A language L of A^* is recognized by an ordered monoid $(M, \leq)$ if there exists a surjective homomorphism of ordered monoids $\eta : (A^*, =) \mapsto (N, \leq)$ and an order ideal P of N such that $L = \eta^{-1}(P)$. A language is *regular* if there exists a finite ordered monoid that recognizes it. This definition is equivalent to the one given in section 2-IV, page 27. To see this, consider η as a homomorphism of ordered monoids from $(A^*, =)$ onto $(N, =)$. The condition on order is trivially satisfied in this case (since $x = y$ implies $\eta(x) = \eta(y)$) and any subset of $(M, =)$ is an order ideal.

Let $(N, \leq)$ be an ordered monoid and let P be an order ideal of N. The *syntactic quasiordering* of P is the relation $\preceq_P$ defined by setting

$$u \preceq_P v \iff \forall x, y \in N : xvy \in P \Rightarrow xuy \in P.$$

The associated equivalence relation $\sim_P$, defined by

$$u \sim_P v \iff u \preceq_P v \wedge v \preceq_P u$$

is a congruence[1] termed the *syntactic congruence* of P. The quotient monoid $M(P) = N/\sim_P$ is called the *syntactic monoid* of P. The ordered monoid $(M(P), \leq_P)$, where $\leq_P$ is the order induced by $\preceq_P$, is called the *syntactic ordered monoid* of P. The natural homomorphism $\eta_P : (N, =) \mapsto (M(P), \leq_P)$ is called the *syntactic homomorphism* of P. Note that here again we have a situation where definitions given previously are subsumed within these new ones.

The next proposition shows that to obtain the syntactic ordered monoid of the complement of an order ideal, one simply reverses the order.

Proposition 8-IV.1 ([PW97]). *Let P be an order ideal of $(N, \leq)$. Then $N \setminus P$ is an order ideal of $(N, \geq)$ and the syntactic ordered monoid of $N \setminus P$ is the dual of the syntactic ordered monoid of P.*

[1] It can be shown that the quasiorder $\preceq_P$ is reflexive, transitive and stable.

Proof: By definition, $u \preceq_{M\setminus P} v$ if and only if for all $x, y \in N$,

$$xvy \in M \setminus P \;\Rightarrow\; xuy \in M \setminus P,$$

which is equivalent to the statement

$$xuv \in P \;\Rightarrow\; xvy \in P.$$

Thus,

$$u \preceq_{M\setminus P} v \iff v \preceq_P u.$$

Q.E.D.

Corollary 8-IV.2. *Let $L \in A^*$ and let $(M(L), \leq_L)$ be its syntactic ordered monoid. Then the syntactic ordered monoid of $A^* \setminus L$ is $(M(L), \geq_L)$.*

Eilenberg's original variety theorem (Theorem 3-III.1) deals with varieties of finite monoids. To obtain a similar statement for varieties of ordered monoids, we define the notion of positive variety, introduced by J. E. Pin in [Pin95].

A *positive variety* is a class of recognizable languages $\mathcal{V}$ such that

1. for every alphabet A, $A^*\mathcal{V}$ is closed under finite union and finite intersection,
2. if $\phi : A^* \mapsto B^*$ is a monoid homomorphism, then
$$L \in B^*\mathcal{V} \;\Rightarrow\; \phi^{-1}(L) \in A^*\mathcal{V},$$
3. if $L \in A^*\mathcal{V}$ and $a \in A$, then $a^{-1}L$ and La^{-1} are in $A^*\mathcal{V}$.

Thus, unlike variety, a positive variety is not required to be closed under complement. To each variety of ordered monoids $\mathbf{V}$, we associate the class $\mathcal{V}$ such that, for each alphabet A, $A^*\mathcal{V}$ is the set of regular languages of A^* whose ordered syntactic monoid belongs to $\mathbf{V}$. The class $\mathcal{V}$ is a positive variety:

Theorem 8-IV.3 ([Pin95]). *The mapping* $\mathbf{V} \mapsto \mathcal{V}$ *defines a bijective correspondence between the varieties of finite ordered monoids and the positive varieties.*

Theorem 8-IV.4. *For each alphabet* A, $A^*\check{\mathcal{V}}$ *is the class of all complements in* A^* *of the languages of* $A^*\mathcal{V}$.

Proof: Follows from Corollary 8-IV.2.

Let, for $0 \leq i \leq n$, L_i be recognizable languages of A^*, let $\eta_i : A^* \mapsto M(L_i)$ be their syntactic homomorphism and let

$$\eta : A^* \mapsto M(L_0) \times M(L_1) \times \cdots \times M(L_n)$$

be the homomorphism defined by

$$\eta(u) = (\eta_0(u), \eta_1(u), \ldots, \eta_n(u)).$$

Let $a_1, a_2, \ldots, a_n$ be letters of A and let $L = L_0 a_1 L_2 \cdots a_n L_n$. Let $\mu : A^* \mapsto M(L)$ be the syntactic homomorphism of L. We now consider the relational homomorphism

$$\tau = \eta(\mu^{-1}) : M(L) \mapsto M(L_0) \times M(L_1) \times \cdots \times M(L_n).$$

Proposition 8-IV.5 ([PW97]). *For every idempotent* e *of* $M(L_0) \times M(L_1) \times \cdots \times M(L_n)$, $\tau^{-1}(e)$ *is an ordered semigroup that satisfies the inequality* $x^\omega y x^\omega \leq x^\omega$.

Proof: Let e be an idempotent of $M(L_0) \times M(L_1) \times \cdots \times M(L_n)$, and let x and y be words in A^* such that $\eta(x) = \eta(y) = e$. Let $k > n$ be an integer such that $\mu(x^k)$ is idempotent. Suffice it to show that for all $u, v \in A^*$

$$ux^k v \in L \;\Rightarrow\; ux^k y x^k v \in L.$$

Since $ux^k v \in L$, there exists a factorization of the form

$$ux^k v = w_0 a_1 w_1 \cdots a_n w_n,$$

where $w_i \in L_i$ for $0 \leq i \leq n$. By the choice of k, there exist integers t and j such that $0 \leq t \leq n$, $0 \leq j \leq k-1$ and

$$\begin{aligned} w_t &= w_t' x w_t'', \\ ux^j &= w_0 a_1 \cdots w_{t-1} a_t w_t', \\ x^{k-j-1} v &= w_t'' a_{t+1} \cdots a_n w_n, \end{aligned}$$

for some $w_t', w_t'' \in A^*$. Now since

$$\eta_t(x) = \eta_t(y) = \eta_t(x^2),$$

the condition $w_t' x w_t'' \in L_t$ implies $w_t' x^{k-j} y x^{j+1} w_t'' \in L_t$. Therefore, $ux^k y x^k v \in L$. *Q.E.D.*

Corollary 8-IV.6. *Let* $\mathbf{V}$ *be a variety of finite monoids and let* $\mathcal{V}$ *be the corresponding variety. If* $L \in A^*Pol\mathcal{V}$, *then* $M(L)$ *belongs to the variety of finite ordered monoids* $[\![x^\omega y x^\omega \leq x^\omega]\!]\textcircled{M}\,\mathbf{V}$.

Proof: Let $\mathbf{W} = [\![x^\omega y x^\omega \leq x^\omega]\!]\textcircled{M}\,\mathbf{V}$ and let $\mathcal{W}$ be the positive variety corresponding to $\mathbf{W}$. By Theorem 8-IV.3, it suffices to show that $L \in A^*\mathcal{W}$. Being a positive variety, $A^*\mathcal{W}$ is closed under finite union, so we only need to show that the theorem holds when L is of the form $L_0 a_1 L_2 \cdots a_n L_n$, where $n \geq 0$ and, for $0 \leq t \leq n$, $a_t \in A$ and $L_t \in A^*\mathcal{V}$. But in this case Proposition 8-IV.5 shows that $M(L) \in \mathbf{W}$. *Q.E.D.*

The following result was established in [Pin80] and [PST88] as a generalization of an earlier theorem due to M. P. Schützenberger, [Sch76].

Theorem 8-IV.7. *Let* $\mathbf{V}$ *be a variety of monoids and let* $\mathcal{V}$ *be the corresponding variety. Then* $UPol\mathcal{V}$ *is a variety of languages, and the associated variety of monoids is* $\mathbf{LI}\textcircled{M}\,\mathbf{V}$.

Theorem 8-IV.8 ([PW97]). *Let $\mathcal{V}$ be a variety of languages. Then*

$$Pol\mathcal{V} \cap Co\text{-}Pol\mathcal{V} = UPol\mathcal{V}.$$

Proof: By definition, $A^*\mathrm{UPol}\mathcal{V}$ is contained in $A^*\mathrm{Pol}\mathcal{V}$. Since $A^*\mathrm{UPol}\mathcal{V}$ is a variety of languages (Theorem 8-IV.7), it is closed under complement. Therefore, $A^*\mathrm{UPol}\mathcal{V}$ is also contained in $A^*\mathrm{Co\text{-}Pol}\mathcal{V}$, which proves the inclusion

$$A^*\mathrm{UPol}\mathcal{V} \subseteq A^*\mathrm{Pol}\mathcal{V} \cap A^*\mathrm{Co\text{-}Pol}\mathcal{V}.$$

For the converse, suppose $L \in A^*\mathrm{Pol}\mathcal{V} \cap A^*\mathrm{Co\text{-}Pol}\mathcal{V}$. The ordered syntactic monoid $M(L)$ of L belongs to the variety of finite ordered monoids $[\![x^\omega y x^\omega \leq x^\omega]\!]\textcircled{M}\,\mathbf{V}$ (Corollary 8-IV.6). By Corollary 8-III.4, the identities defining this variety are of the form $x^\omega y x^\omega \leq x^\omega$, where $x, y \in \widehat{A^*}$ for some finite set A and $\mathbf{V}$ satisfies $x = y = x^2$. Let B be a finite alphabet and let $x, y \in \widehat{B^*}$ be such that $\mathbf{V}$ satisfies $x = y = x^\omega$. Then $M(L)$ satisfies $[\![x^\omega y x^\omega \leq x^\omega]\!]$. Since $L \in A^*\mathrm{Co\text{-}Pol}\mathcal{V}$, the complement of L belongs to $A^*\mathrm{Pol}\mathcal{V}$ and thus by Corollary 8-IV.2 and Theorem 8-IV.4, $M(L)$ satisfies $[\![x^\omega \leq x^\omega y x^\omega]\!]$. Then necessarily $M(L)$ satisfies $[\![x^\omega y x^\omega = x^\omega]\!]$. Thus, by Corollary 8-III.3, $M(L) \in \mathbf{LI}\textcircled{M}\,\mathbf{V}$ and by Theorem 8-IV.7, $L \in A^*\mathrm{UPol}\mathcal{V}$. *Q.E.D.*

8-V Application to the logical hierarchy Σ_k

Theorem 8-V.1.

$$\mathcal{DA} = \Sigma_2 \cap \Pi_2.$$

Proof: $A^*\mathcal{DA}$ is the smallest class of languages of A^* containing languages of the form B^*, with $B \subseteq A$, and closed under disjoint union and unambiguous product (cf. [Sch76]). Thus

$$A^*\mathcal{DA} = A^*\mathrm{UPol}(\mathcal{B}).$$

By Theorems 7-III.5 and 7-II.8 $\Sigma_2 = \mathrm{Pol}(\mathcal{J})$, and by Theorem 7-III.4, $\mathrm{Pol}(\mathcal{J}) = \mathrm{Pol}(\mathcal{B})$, so $\Sigma_2 = \mathrm{Pol}(\mathcal{B})$. By the same theorems, $\Pi_2 = \mathrm{Co\text{-}Pol}(\mathcal{B})$. Now applying Theorem 8-IV.8, we have the stated property. *Q.E.D.*

We can also derive a more general statement. Let $\Delta_n = \Sigma_n \cap \Pi_n$ for all non-negative integers n.

Theorem 8-V.2 ([PW97]). *Let L be a language of A^*. Then*

$$L \in \Delta_{n+1} \iff L \in UPol\mathcal{V}_n.$$

Proof: By Theorems 7-III.5 and 8-IV.8 we obtain:

$$\Delta_{n+1} = \Sigma_{n+1} \cap \Pi_{n+1} = \mathrm{Pol}\mathcal{V}_n \cap \text{Co-Pol}\mathcal{V}_n = \mathrm{UPol}\mathcal{V}_n.$$

Q.E.D.

Conclusion

Theorem 7-III.5 shows that the Straubing-Thérien hierarchy $\mathcal{V}_k$ is in one-to-one correspondence with a well known hierarchy of first-order logic, the Σ_n hierarchy. Theorems 7-II.8 and 7-III.6 assert that both the boolean closure of Σ_1 and the boolean closure of Σ_2 define varieties of languages; they correspond, respectively, to levels 1 and 2 of the Straubing-Thérien hierarchy. Level $\mathcal{V}_1$ is precisely the class of piecewise testable languages, i.e. languages recognized by $\mathcal{J}$-trivial monoids. Level $\mathcal{V}_2$ is recognized by the monoids of upper-triangular matrices over the semiring $\{0, 1\}$.

We defined the level $\mathcal{V}_{i+1}$ of the Straubing-Thérien hierarchy as the boolean-polynomial closure of the level $\mathcal{V}_i$. An alternative definition may be stated in the following way: the level $n + 1/2$ is the polynomial closure of the level n and the level $n + 1$ is the boolean closure of the level $n + 1/2$.

The main problems associated with any hierarchy are the finiteness and the decidability of each level.

The Straubing-Thérien hierarchy is infinite. This result follows from the fact that the logical hierarchy Σ_k is infinite (Theorem 5-III.6).

Levels 0, 1/2 and 1 are known to be decidable in polynomial time. The level 3/2 has also been shown decidable, in time polynomial in $2^{|A|}n$, where A is the alphabet and n is the number of states of the deterministic automaton (cf. [PW97]). Decidability of level 2 is still an open question, as is the problem of identities for the variety of monoids corresponding to languages of level 2.

Bibliography

[Arb68] M. A. Arbib, *Algebraic Theory of Machines, Languages and Semigroups*, Academic Press, New York, 1968

[Arf91] M. Arfi, Opération polynomiales et hiérarchies de concaténation, *Theoret. Comp. Sci.* **91** (1991) 71-84

[Bir35] G. Birkhoff, On the structure of abstract algebras, *Proc. Cambridge Phil. Soc.* **31**, 1935, 433-454.

[Büc60] J. R. Büchi, Weak second-order arithmetic and finite automata, *Zeit. Math. Logik. Grund. Math.*, **6** (1960), 66-92

[Büc62] J. R. Büchi, On a decision method in restricted second-order arithmetic, in *Proc. 1960 Int. Cong. for Logic, Methodology and Philosophy of Science*, Stanford Univ. Press, Stanford, (1962), 1-11.

[CM56] A. H. Clifford and D. D. Miller, Regular $\mathcal{D}$-classes in semigroups, *Trans. Am. Math. Soc.* **82** (1956), 1-15

[CP67] A. H. Clifford and G. B. Preston, *The Algebraic Theory of Semigroups*, Vol. 1 and 2, American Mathematical Society, Mathematical Surveys 7, 1967

[EF95] H. D. Ebbinghaus, J. Flum, *Finite Model Theory*, Springer-Verlag, New York 1995

[EFT94] H. D. Ebbinghaus, J. Flum, W. Thomas, *Mathematical Logic (2nd Ed.)*, Springer-Verlag, New York 1994

[Ehr61] A. Ehrenfeucht, An application of games to the completeness problem for formalized theories, *Fund. Math.*, **49** (1961), 129-141

[Eil76] S. Eilenberg, *Automata, languages and machines*, Vol. B, Academic Press, New York, 1976

[Elg61] C. Elgot, Decision problems of finite automata design and related arithmetics, *Trans. Amer. Math. Soc.* **98** (1961), 21-52

[FSV95] R. Fagin, L. J. Stockmeyer, M. Y. Vardi, On monadic NP vs monadic co-NP, *Information and computation* **120** (1995), 78-92

[Fra72] R. Fraïssé, *Cours de Logique Mathématique*, Tome 2, Gauthiers-Villars, Paris 1972

[Gre51] J. A. Green, On the structure of semigroups, *Ann. Math.* **54** (1951), 163-172

[Hil02] D. Hilbert, Mathematical Problems, *Bull. Amer. Math. Soc.* **8**, 437-479, 1902

[HU79] J. E. Hopcroft and J. D. Ullman. *Introduction to automata theory, Languages and Computation.* Addison-Wesley, Reading, MA, 1979.

[KR65] K. Krohn and J. Rhodes, The Algebraic Theory of Machines I, *Trans. Amer. Math. Soc.* **116** (1965), 450-464

[Lal79] G. Lallement, *Semigroups and combinatorial applications*, John Wiley & Sons, New York, 1979

[Lot83] M. Lothaire, *Combinatorics on Words, Encyclopedia of Mathematics 17*, Addison-Wesley, Reading, MA, 1983

[Mat70] Y. Matiyasevich, Enumerable sets are Diophantine, *Dokl. Akad. Nauk SSSR* **191** (1970), 279-282

[MP71] R. McNaughton and S. Papert, *Counter-free Automata*, MIT Press, Cambridge, Mass. 1971

[Pap94] C. H. Papadimitriou, *Computational complexity*, Addison-Wesley, 1994.

[PP86] D. Perrin and J. E. Pin, First-order logic and star-free sets, *J. Comp. and Syst. Sci.* **32** (1986), 393-406

[Pin80] J. E. Pin, Propriétés syntactique du produit non ambigu. *7th ICALP, Lecture Notes in Computer Science* **85** (1980), 483-499

[Pin84] J. E. Pin, *Variété de langages formels*, Masson, Paris, 1984.

[Pin95] J. E. Pin, A variety theorem without complementation, *Izvestiya VUZ Matematika* **39** (1995), 80-90; English version, *Russian Mathem. (Iz. VUZ)* **39** (1995), 74-83

[PS81] J. E. Pin and H. Straubing, Monoids of upper triangular matrices, *Colloquia Mathematica Societatis Janos Bolyai* **39**, *Semigroups*, (1981) 259-272

[PST88] J. E. Pin, H. Straubing, D. Thérien, Locally trivial categories and unambiguous concatenation, *Journal of Pure and Applied Algebra* **52** (1988), 297-311

[PW95] J. E. Pin and P. Weil, Profinite semigroups, Mal'cev products and identities, to appear in *J. of Algebra*

[PW97] J. E. Pin and P. Weil, Polynomial closure and unambiguous product, *Theory Comput. Systems* **30** (1997), 383-422

[Rab69] M. O. Rabin, Decidability of second-order theories and automata on infinite trees, *Trans. Amer. Math. Soc.* **141** (1969), 1-35

[Rei82] J. Reiterman, The Birkoff theorem for finite algebras, *Algebra Universalis* **14**, 1982, 1-10

[Ros82] J. R. Rosenstein, *Linear Orderings*, Academic Press, New York 1982

[Sch65] M. P. Schützenberger, On finite monoids having only trivial subgroups, *Information and Control* **8** (1965), 190-194

[Sch76] M. P. Schützenberger, Sur le produit de concaténation non ambigu, *Semigroup Forum* **13** (1976), 47-75

[Sim75] I. Simon, Piecewise testable events, *Proc. 2nd GI Conf.*, Lecture Notes in Computer Science, **33**, Springer-Verlag, Berlin 1975, pp. 214-222

[Str80] H. Straubing, On finite $\mathcal{J}$-trivial monoids, *Semigroup foruum* **19** (1980), 107-110

[Str85] H. Straubing, Finite semigroups varieties of the form $\mathbf{V} * \mathbf{D}$, *J. Pure Appl. Algebra* **36**, (1985) 53-94

[Str94] H. Straubing, *Finite automata, formal logic and circuit complexity*, Birkhäuser, Boston, 1994

[Thé81] D. Thérien, Classification of finite monoids: the language approach, *Theoretical Comp. Sci.* **14** (1981), 195-208

[Tho82] W. Thomas, Classifying regular events in symbolic logic, *J. Comput. Sysytem Sci.* **25** (1982), 360-376

[Tho84] W. Thomas, An application of Ehrenfeucht-Fraïssé game in formal language theory, *Bull. Soc. Math. France* **16** (1984), 11-21

[Tur36] A. M. Turing, On computable numbers with an application to the *entscheidungsproblem*, *Proc. London Math. Soc.* **2**:32, 230-265

List of Figures

Index

Logical aspects of regular languages

Boris Dashkovsky

Department of Computer Science

McGill University, Montreal

November 1999

A thesis submitted to the

Faculty of Graduate Studies and Research

in partial fulfilment of the requirements of the degree of

Master of Science.

Contents

Abstract

A thorough review of selected results on the logical aspects of regular languages includes the theorem of Büchi on monadic second order logic over strings, a characterization of $FO[<]$ and the theorem of I. Simon. With the help of the Ehrenfeucht-Fraïssé Game we show that $\exists^{(k+1)}$-sentences of $FO[<]$ cannot be expressed as a boolean combination of $\exists^{(k)}$-sentences. Block product of finite monoids is used to analyze languages defined by the boolean closure of the Σ_2-sentences. Positive varieties and the Mal'cev product are introduced and $\Sigma_{n+1} \cap \Pi_{n+1}$ is shown to be equal to the unambiguous polynomial closure of the nth level of the Straubing-Thérien hierarchy. In particular, $\Sigma_2 \cap \Pi_2 = \mathcal{DA}$, where $\mathcal{DA}$ is the smallest variety of languages closed under the unambiguous product.

Résumé

Nous proposons un aperçu complet de résultats choisis concernant les aspects logiques des langages réguliers incluant le théorème de Büchi sur la logique monadique de second ordre sur les chaînes de caractères, la caractérisation de $FO[<]$ et le théorème de I. Simon. Grâce au jeu de Ehrenfeucht-Fraïssé, nous démontrons que, dans $FO[<]$, les énoncés logiques $\exists^{(k+1)}$ ne peuvent être exprimés comme une combinaison booléene d'énoncés $\exists^{(k)}$. Nous utilisons le produit bloc de monoïdes finis pour analyser les langages définis par la fermeture booléene des énoncés Σ_2. Nous présentons également les variétés positives et le produit de Mal'cev et montrons que $\Sigma_{n+1} \cap \Pi_{n+1}$ est égal à la fermeture polynomiale non-ambigue du $n^{\text{ième}}$ niveau de la hiérarchie de Straubing-Thérien. En particulier, $\Sigma_2 \cap \Pi_2 = \mathcal{DA}$, où $\mathcal{DA}$ est la plus petite variété de langages fermée sous le produit non-ambigu.

Acknowledgements

I wish to express my thanks to my supervisor Denis Thérien for introducing me to this field and for his patience and support during the entire time of my residency. My gratitude also goes to my friends – Emil Ciasca and Flavia Majlis for their encouragement.

Introduction

The topic of this Thesis lies at the juncture of formal language theory, algebraic theory of finite automata and model theory in logic.

In 1956 S. C. Kleene showed that the class of languages recognized by finite automata (*regular* languages) coincides with that given by the rational expressions (*rational* languages). This theorem is usually considered to be the foundation of the theory of finite automata. The definition of the *syntactic monoid* was first given in a paper of M. O. Rabin and D. Scott in 1959, where the notion was credited to Myhill. It was shown in particular that a language is recognizable if and only if its syntactic monoid is finite. M. P. Schützenberger made a non-trivial use of the syntactic monoid to characterize an important subclass of the rational languages, the *star-free* languages: a language is star-free if and only if its syntactic monoid is finite and aperiodic.

In the early 1970's I. Simon proved that a language is piecewise testable if and only if its syntactic monoid is $\mathcal{J}$-trivial. Other important syntactic characterization followed, settling the power of the semigroup approach. But it was S. Eilenberg who formulated the appropriate framework for this type of results. A variety of finite monoids is a class of monoids closed under taking submonoids, quotients and finite direct products. Eilenberg's Theorem states that varieties of finite monoids are in one-to-one correspondence with certain classes of regular languages, the varieties of languages.

For these reasons the part of formal language theory concerned with rational languages is now intimately related to both the theory of finite automata and the

theory of finite monoids.

The connection between automata and formal logic dates back to 1936 when A. Turing proved the undecidability of first-order logic by showing how to describe the behaviour of an abstract computing machine with a formula of this logic. More contributions into the research on the logical aspects of the automata theory ensued, with the works of J. R. Büchi on monadic second-order logic and R. McNaughton and S. Papert on automata admitting first-order behavioral description – among the more famous ones.

In the mid-1990's J. E. Pin developed a theory of so-called positive varieties of languages, which – unlike varieties introduced by S. Eilenberg – do not have to be closed under complement. Their algebraic counterpart had to be modified too – they are varieties of finite ordered monoids. The polynomial closure of a variety of languages is always a positive variety; this property led to establishing some new connections between regular languages and logic.

The main objective of this study is concentrated on proving necessary (and sometimes also sufficient) conditions for a property of words to be expressible in a particular logical formalism. We present two general techniques for accomplishing such results: analysis of logical formulæ with methods of the theory of finite monoids and the model-theoretic method of Ehrenfeucht-Fraïssé Games, described in Chapter 7.

Some developments in the field of logical aspects of regular languages – both classical and relatively new – are echoed in this text.

In Chapter 1 we review the main concepts of formal logic and finite automata. The mathematical machinery needed to maintain a degree of self sufficiency of the manuscript includes elements of the theory of finite monoids presented in Chapter 2.

Identities of finite monoids, the notion of variety and its connection with logic are introduced in Chapter 3.

Our digression into semigroup theory continues in Chapter 4 where we define transformation semigroups, wreath product and block product. Acquired tools will

be used in the subsequent chapters to establish some important algebraic characterization of subclasses of regular languages.

Chapter 5 expounds two topics: the theorem of Büchi on monadic second order logic over strings and the algebraic characterization of first-order logic in signature with $<$.

The subject of Chapter 6 is the theorem of I. Simon and piecewise testable languages; we give both combinatorial and algebraic description of these.

In Chapter 7 we present an algebraic characterization of the first two levels of the Straubing-Thérien hierarchy [1] and their connection to the logical hierarchy. We also give a treatment of some special quantification structures and examine the corresponding varieties of languages. The quest for more ties between the two hierarchies reveals some interesting results as we introduce the notions of ordered finite monoids, positive varieties and the Mal'cev product in Chapter 8.

[1] It should be noted, however, that the "characterization" of level 2 is not effective.

Chapter 1

The Basis

1-I Introduction

This chapter focuses on some fundamental concepts in the study of formal languages. We continue by introducing the notion of finite automaton, followed by a digression into formal logic.

1-II Formal Languages

Let $A = \{a_1, a_2, \ldots, a_i\}$ be a finite set of symbols, called an *alphabet* and its elements - *letters* . A *word* (or a *string*) $w = a_1 a_2 \cdots a_m$ over an alphabet A is a finite sequence of letters. By $|w|$ we denote the length m of the word w. For some $a \in A$, $|w|_a$ denotes the number of occurrences of a in w. We then have:

$$\sum_{a \in A} |w|_a = |w|.$$

The *empty string* , denoted 1, has length 0. By juxtaposition uv, or multiplication $u \cdot v$ we mean concatenation of two words u and v producing a sequence with $|uv| = |u| + |v|$ and clearly $|uv|_a = |u|_a + |v|_a$. For the empty word we have $1 \cdot w = w \cdot 1 = w$.

Notation. For a positive integer k and a word w, the form w^k is a shorthand notation for $\underbrace{ww \cdots w}_{k \text{ times}}$. By convention, $w^0 = 1$.

Given two words u and v:

1. u is a *prefix* of v if $\exists x \in A^* : v = ux$;

2. u is a *suffix* of v if $\exists x \in A^* : v = xu$;

3. u is a *factor* of v if $\exists x, y \in A^* : v = xuy$.

A word $u = a_1 a_2 \dots a_n$ is a *subword* of v if there exist words $v_0, v_1, \dots, v_n \in A^*$ such that $v = v_0 a_1 v_1 a_2 \dots a_n v_n$.

The set of all words over the alphabet A is denoted by A^*, the set of all nonempty words - A^+. A subset of A^* is called a *language* . Various operations can be defined over languages. Besides the classical boolean operations (such as finite union, finite intersection, complement) we shall make use of the ones below.

The *product* (or *concatenation product*) of two languages L and K is the language

$$LK = \{uv \in A^* | u \in L, v \in K\}.$$

The *star* of a language $L \subseteq A^*$, denoted by L^* is the language

$$L^* = \{1\} \cup L \cup LL \cup LLL \cup \cdots$$

If K and L are two languages of A^*, the *left* (*right*) *quotient* of L by K is the language $K^{-1}L$ (respectively LK^{-1}). These are defined by:

$$K^{-1}L = \{v \in A^* | Kv \cap L \neq \emptyset\} = \{v \in A^* | \exists u \in K \text{ such that } uv \in L\}$$

and

$$LK^{-1} = \{v \in A^* | vK \cap L \neq \emptyset\} = \{v \in A^* | \exists u \in K \text{ such that } vu \in L\}.$$

1-III Finite Automata

A *deterministic finite automaton* (or DFA) over a finite alphabet A is a quadruple

$$\mathcal{T} = (Q, i, F, \lambda)$$

where Q is a finite set of *states* of the automaton; $i \in Q$ is the *initial state*; $F \subseteq Q$ is the set of *final states* and λ is the *transition function* $\lambda : Q \times A \mapsto Q$ defined for all $q \in Q$ and for all $a \in A$. We shall adopt the shorthand notation qa or $q \cdot a$ for $\lambda(q, a)$.

The domain of the transition function λ can be extended to the set $Q \times A^*$ by induction on the length of the input word:

$$q \cdot 1 = q \qquad \text{and} \qquad q \cdot (wa) = (qw) \cdot a.$$

The string w is *accepted* by DFA if $i \cdot w \in F$. The *language* L *recognized* by the DFA is the set of all such words w:

$$L = \{w \in A^* \mid i \cdot w \in F\}.$$

A language is said to be *regular* if there exists a DFA recognizing it.

1-III.1 The minimal automaton

Let $\mathcal{T} = (Q, i, F, \lambda)$ be a DFA and $L \subseteq A^*$ the language it recognizes. Define the set $Q' \subseteq Q$ of states of the DFA reachable from the initial state i:

$$Q' = \{i \cdot w \mid w \in A^*\}$$

and the following equivalence relation $\sim$ on Q':

$$q_1 \sim q_2 \iff \{w \in A^* \mid q_1 \cdot w \in F\} = \{w \in A^* \mid q_2 \cdot w \in F\}.$$

$q_1 \sim q_2$ implies $q_1 a \sim q_2 a$ for all $a \in A$ and therefore the transition function $\widetilde{\lambda} : Q'/_\sim \times A \mapsto Q'/_\sim$ is well defined for the equivalence classes $[q]$ of $q \in Q$:

$$\widetilde{\lambda}([q],\ a) = [qa].$$

The DFA

$$\widetilde{\mathcal{T}_L} = (Q'/_\sim,\ [i],\ \{[q] \mid q \in F\},\ \widetilde{\lambda})$$

also recognizes L, but its structure depends **only** on L. $\widetilde{\mathcal{T}_L}$ is called the *minimal* automaton of L. Any automaton $\mathcal{A}$ recognizing L has at least as many states as $\widetilde{\mathcal{T}_L}$ does and if $\mathcal{A}$ and $\widetilde{\mathcal{T}_L}$ have the same number of states, they are isomorphic.

Example 1-III.1. Let $A = \{a, b, c\}$ and $L = A^*abA^*$. A DFA recognizing L is pictured in fig. 1.1. One can easily verify that this is the minimal automaton of L.

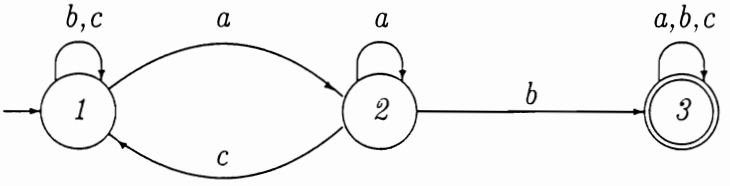

Figure 1.1: The minimal automaton of $L = A^*abA^*$ over $A = \{a, b, c\}$.

1-IV Formal Logic

1-IV.1 Propositional Logic

Define a countable set $X = \{x_1, x_2, \dots\}$ of *boolean* variables (i.e. variables taking on values **True** or **False**).

A *boolean expression* consists of:

(a) a boolean variable x_i; or

(b) an expression of the form: $\neg\phi$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$, where ϕ, ψ are themselves boolean expressions.

The set of boolean variables of an expression ϕ, $X(\phi) \subset X$, is defined inductively as follows:

(a) if ϕ is a boolean variable x_i, then $X(\phi) = \{x_i\}$,

(b) if $\phi = \neg\psi$, then $X(\phi) = X(\psi)$,

(c) if $\phi = (\chi \wedge \psi)$ or $(\chi \vee \psi)$, then $X(\phi) = X(\chi) \cup X(\psi)$.

A *truth assignment* T is a mapping from the set of boolean variables $X(\phi)$ to the set of *truth values* { **True, False** }. We now define what it means for T to *satisfy* ϕ (written $T \models \phi$):

(a) if ϕ is a boolean variable $x_i \in X(\phi)$, then $T \models \phi$ if $T(x_i) = \mathbf{True}$,

(b) if $\phi = \neg\psi$, then $T \models \phi$ if it is not the case that $T \models \psi$,

(c) if $\phi = (\chi \vee \psi)$ then $T \models \phi$ if either $T \models \chi$ or $T \models \psi$ holds,

(d) if $\phi = (\chi \wedge \psi)$ then $T \models \phi$ if both $T \models \chi$ and $T \models \psi$ hold.

Notation. An expression of the form x_i or $\neg x_i$ is termed a *literal* . We use $(\phi \Rightarrow \psi)$ to mean $(\neg\phi \vee \psi)$; and $(\phi \iff \psi)$ stands for $((\phi \Rightarrow \psi) \wedge (\psi \Rightarrow \phi))$.

It is well known that the relations $\vee$ and $\wedge$ are commutative, associative, distributive and idempotent (see for instance [Pap94]). Furthermore, it follows that every boolean expression ϕ can be rewritten into an equivalent one in *conjunctive*: $\phi = \wedge_{i=1}^{n} C_i$ or *disjunctive*: $\phi = \vee_{i=1}^{n} D_i$ *normal form*, where C_i (called a *clause*) is the disjunction of one or more literals and D_i (called an *implicant*) is the conjunction of one or more literals.

1-IV.2 First-Order Logic

The language of first-order logic is capable of expressing a wide range of mathematical ideas and facts in much more detail than boolean logic.

1-IV.2-a The Syntax

Let us define three disjoint countable sets: V – a set of variables (ranging over values from the domain of a particular expression); Φ – a set of function symbols; Π – a set of relation symbols and the *arity* function: $r : \Phi \cup \Pi \mapsto \mathbb{Z}_+$. A function $f \in \Phi$ with $r(f) = k$, $k \geq 0$ is called a *k-ary* function (similarly for a relation $R \in \Pi$ with $r(R) = k$, $k > 0$, k-ary relation). The set Π is always assumed to contain the binary *equality* relation $=$. A triplet $\Sigma = (\Phi, \Pi, r)$ is called a *vocabulary*. The set of used function and relation symbols $(\Phi \cup \Pi)$ is called the *signature* of the first-order language.

A *term* over the vocabulary Σ is (a) a variable $x \in V$; or (b) an expression $\underline{f(t_1, t_2, \ldots, t_k)}$. where $f \in \Phi$ and $t_1, t_2, \ldots, t_k$ are themselves terms. (This definition allows for a constant when $k = 0$.)

An *atomic expression* over the vocabulary Σ is an expression of the form $\underline{R(t_1, t_2, \ldots, t_k)}$, where $R \in \Pi$ and $t_1, t_2, \ldots, t_k$ are terms.

A *first-order expression* (or *first-order formula*) is

(a) an atomic expression; or

(b) an expression of the form $\underline{\neg\phi}$, $\underline{(\phi \vee \psi)}$ or $\underline{(\phi \wedge \psi)}$, with ϕ, ψ themselves being first-order expressions; or

(c) an expression of the form $\underline{(\forall x \phi)}$, where $x \in V$ and ϕ is a first-order expression.

Notation. The form $\underline{(\exists x \phi)}$ is used as a shorthand for $\underline{\neg(\forall x \neg\phi)}$. When there is no ambiguity we may write $\forall x, y \cdots$ and $\exists x, y \cdots$ to mean respectively $\forall x \forall y \cdots$ and $\exists x \exists y \cdots$.

The symbols $\forall$ and $\exists$ are the *universal* and *existential quantifier* respectively. An appearance of a variable x in the text of an expression ϕ that does not immediately follow a quantifier is called an *occurrence* of x in ϕ. An occurrence of a variable is said to be *bound* if it is referred to by a quantifier; that is, if $\underline{\forall x \phi}$ is an expression, any

occurrence of x in ϕ is bound [1] (variable x is said to be in the *scope* of a quantifier). If the occurrence is not bound, it is *free*. A variable x that has a free occurrence in ϕ is a *free variable* of ϕ. An expression without free variables is called a *sentence*.

Expressions where a prefix of quantifiers precedes a quantifier-free structure are in *prenex normal form*. Any first-order formula can be transformed into one in prenex normal form. If successive quantifiers of the same type are grouped into n alternating blocks beginning with existential quantifiers, i.e. a formula ϕ is of the form

$$\phi = \exists \widehat{x_1} \forall \widehat{x_2} \cdots \exists \widehat{x_{n-1}} \forall \widehat{x_n} \psi(\widehat{x_1}, \widehat{x_2}, \ldots \widehat{x_n}),$$

where $\widehat{x_i}$ are tuples of variables and ψ is quantifier-free, then ϕ is said to be a Σ_n-formula. In the dual case, when n alternating blocks of quantifiers start with a block of universal quantifiers, the expression is called a Π_n-formula. The negation of a Σ_n formula can be written as a Π_n formula.

Remark 1-IV.1. The first block of quantifiers in a Σ_n (or Π_n) formula may be empty.

1-IV.2-b The Semantics

In first-order logic variables, functions and relations may take on much more complex values than just **True** or **False**. To define the semantics of first-order formulæ we construct an analog of a truth assignment for first-order logic, called a *model*.

A *model appropriate* to a given vocabulary $\Sigma = (\Phi, \Pi, r)$ is a pair $M = (U, \mathcal{I})$, where U is a non-empty set (called the *universe* of M) and $\mathcal{I} : V \cup \Phi \cup \Pi \mapsto U$ is an *interpretation function* associating each symbol α in V, Φ, Π with an actual mathematical object α^M in the universe U. That is, for all $x \in V$, $\mathcal{I}$ assigns an actual element $x^M \in U$; to every function symbol $f \in \Phi$, $\mathcal{I}$ assigns an actual function $f^M : U^k \mapsto U^k$, where k is the arity; and to each relation symbol $R \in \Pi$, $\mathcal{I}$ assigns an actual relation $R^M \subseteq U$.

To define what it means for a model $M = (U, \mathcal{I})$ to *satisfy* a first-order expression ϕ (written $M \models \phi$) we follow the structure of a first-order formula:

[1] An occurrence of x is also bound in any expression containing $\forall x \phi$ as a subexpression.

(a) if ϕ is an atomic expression, $\underline{\phi = R(t_1, t_2, \ldots, t_k)}$, then
$(M \models \phi) \iff (t_1^M, t_2^M, \ldots t_k^M) \in R^M$;

(b) if ϕ is an expression of the form $\underline{\neg\alpha}$, $\underline{(\alpha \vee \beta)}$ or $\underline{(\alpha \wedge \beta)}$, where α, β are first-order expressions, satisfaction is defined by induction on the structure of ϕ;

(c) if ϕ is an expression of the form $\underline{(\forall x \psi)}$, then
$(M \models \phi) \iff (\forall u \in U : M_{x=u} \models \psi)$, where $M_{x=u}$ is a new model obtained from M by fixing $x^{M_{x=u}} = u$.

Theorem 1-IV.1 (cf. [EFT94]). *Let ϕ be an expression and M, M' – two models appropriate to the vocabulary of ϕ. If M, M' agree on everything except for the values they assign to the variables that are not free in ϕ, then*

$$M \models \phi \iff M' \models \phi.$$

Consequently, for sentences (i.e. expressions with no free variables) satisfaction by a model does not depend on the values assigned to the variables that are bound (or do not appear) in the expression. More generally, if ϕ is a formula with free variables, whether a model satisfies or fails to satisfy ϕ depends both on the interpretation $\mathcal{I}$ and the set of free variables in ϕ. Therefore a "model appropriate to an expression" shall henceforth refer to the part of the model that deals with the functions, relations and free variables (if any).

1-IV.3 Words as a Model

We shall now assemble the following vocabulary $\Sigma = (\Phi, \Pi, r)$: $\Phi = \{\emptyset\}$, i.e. there will be no functions; the set of relation symbols $\Pi = \{=, <, S, Q_a\}$ includes the equality relation $=$, the precedence order $<$, the successor relation S and unary "label" predicates Q_a defined below.

1-IV.3-a Büchi sequential calculus

Let A be a finite alphabet and let $w = a_1 a_2 \cdots a_n$ be a word over A. Variables $x \in V$ range over the set of letter positions of w, or the *domain* of w: $dom(w) = \{1, \ldots, n\}$.

Let us now define a *word model* $\mathcal{W}$ for w appropriate to the vocabulary Σ:

(a) $<^{\mathcal{W}}$ is the natural order on $dom(w)$;

(b) $S^{\mathcal{W}}(i, i+1)$ is the successor relation for $1 \leq i \leq n-1$; and

(c) $Q_a^{\mathcal{W}}$ are unary predicates collecting for each letter $a \in A$ the word positions i in which the letter a appears: $Q_a^{\mathcal{W}} = \{i \in dom(w) | a_i = a\}$.

Remark 1-IV.2. Observe that the successor relation $S(x, y)$ can be expressed in terms of relation $<$ by the formula $(x < y) \wedge \neg\exists z((x < z) \wedge (z < y))$.

If $p_1, \ldots, p_n$ are positions from $dom(w)$ then

$$(\mathcal{W}, p_1, \ldots, p_n) \models \phi(x_1, \ldots, x_n)$$

means that ϕ is satisfied in $\mathcal{W}$ when the signature symbols (i.e. $=, <, S, Q_\alpha$) are interpreted by the relations of equality, $<^{\mathcal{W}}, S^{\mathcal{W}}, Q_a^{\mathcal{W}}$, respectively and positions $p_1, \ldots, p_n$ are interpretation of variables $x_1, \ldots, x_n$ respectively. The word model $\mathcal{W}$ is called *Büchi sequential calculus* (cf. [Büc60], [Büc62]).

1-IV.3-b The $\mathcal{V}$-structure model

As noted above, in view of theorem 1-IV.1, let us concentrate on the part of the model concerned with the free variables of an expression. The following idea of treating the structures in which we interpret formulæ as being words over an extended finite alphabet emanates from Perrin and Pin (cf. [PP86]).

Let ϕ be a first-order formula such that no variable x in ϕ (and all its subexpressions) has bound occurrences in the scope of two different quantifiers.[2] We construct a finite set $\mathcal{V} \subseteq V$ of first-order variables of ϕ:

$$x \in \mathcal{V} \iff x \text{ has only free occurrences in } \phi.$$

[2]Any first-order formula can be written to satisfy this condition by introducing new names for the bound variables, if needed.

A $\mathcal{V}$-*structure* over A is a word w over the extended alphabet $A \times 2^{\mathcal{V}}$:

$$w = (a_1, P_1) \cdots (a_r, P_r),$$

where $r = |\mathcal{V}|$, $a_i \in A$ and P_i satisfy the following:

$$P_i \cap P_j = \emptyset, \text{ if } i \neq j, \qquad \text{and} \qquad \bigcup_{i=1}^{r} P_i = \mathcal{V}.$$

We now define the meaning of $w \models_{\mathcal{I}} \phi$ by induction on the construction of ϕ:

(a) $w \models_{\mathcal{I}} Q_a(x)$ if and only if w contains a letter of the form (a, P) and $x \in P$;

(b) $w \models_{\mathcal{I}} R(x_1, \ldots, x_k) \iff (p_1, \ldots, p_k) \in R^{\mathcal{I}}$, where $R^{\mathcal{I}}$ is the k-ary relation on $\{1, \ldots, |w|\}$ associated to R by $\mathcal{I}$ and $p_1, \ldots, p_k$ are the positions in w where the variables $x_1, \ldots, x_n$, respectively, occur;

(c) $w \models_{\mathcal{I}} \neg\phi$ if and only if w is not a model of ϕ with respect to the interpretation $\mathcal{I}$;

(d) $w \models_{\mathcal{I}} (\phi \wedge \psi) \iff (w \models_{\mathcal{I}} \phi) \wedge (w \models_{\mathcal{I}} \psi)$;

(e) $w \models_{\mathcal{I}} \exists x\phi$ if and only if there exists i, $1 \leq i \leq r$, such that

$$w' = (a_1, P_1) \cdots (a_i, P_i \cup \{x\}) \cdots (a_r, P_r) \models_{\mathcal{I}} \phi.$$

The atomic expressions of this first-order language are of the form:

(a) $\underline{x = y}$ means x and y refer to the same position in w;

(b) $\underline{S(x, y)}$ says that position x is immediately succeeded by position y;

(c) $\underline{x < y}$ tells us that position x is to the left of position y in w;

(d) $\underline{Q_a(x)}$ reveals that in w position x is occupied by the letter a.

Notation. The set of first order formulæ utilizing the set of relational symbols $\Pi = \{=, <, Q_a\}$ ($\Pi = \{=, S, Q_a\}$) is denoted $FO[<]$ (respectively $FO[S]$).

1-IV.4 Languages defined by first-order expressions

If ϕ is a sentence (i.e. ϕ does not have any free variables), then ϕ can be interpreted in a word $w \in A^*$, in which case the *language defined by* ϕ is

$$L(\phi) = \{w \in A^* \mid w \models_{\mathcal{I}} \phi\}.$$

If ϕ is a formula with free variables in $\mathcal{V}$, then by $L(\phi)$ we denote the set of $\mathcal{V}$-structures that satisfy ϕ. This notion depends both on the interpretation function $\mathcal{I}$ and on the set of free variables $\mathcal{V}$.

Below are some examples of languages defined by first-order sentences.

Example 1-IV.1. An $FO[S]$ sentence

$$\phi = \exists x \exists y \exists z \big(S(x,y) \wedge S(y,z) \Rightarrow \neg\exists p S(p,x) \wedge \neg\exists q S(z,q)\big)$$

defines a set of words with exactly three distinct positions in them:

$$L(\phi) = \{w \in A^* : |w| = 3\}.$$

Example 1-IV.2. Consider an $FO[<]$ sentence $\psi = \exists x(\forall z(z \geq x) \wedge Q_a x)$. It describes a language of all words over A^* beginning with the letter a, i.e. $L(\psi) = aA^*$.

Two expressions ϕ and ψ are said to be *equivalent* if their languages coincide, i.e. $L(\phi) = L(\psi)$.

Remark 1-IV.3. The empty word 1 is allowed as member of formal languages and the empty model $\underline{1}$ is admitted as interpretation of sentences. By convention, $\underline{1}$ satisfies universal sentences $\forall x \phi(x)$, but not existential ones $\exists x \phi(x)$.

1-IV.5 MSO Logic

In a first-order formula only individual variables can be quantified. Allowing quantification over sets of variables as well as individual variables, extends the logical formalism by *second-order monadic variables* or *predicates* (usually written as capitalized

X as opposed to x). With the introduction of corresponding atomic expressions: e.g. $X(x)$ (meaning x belongs to the set X), the resulting system becomes *monadic second-order logic* or MSO-logic (sets are monadic objects).

A second-order formula can also be presented in prenex normal form. A Σ^1_n-formula is an expression with a prefix of n second-order quantifier blocks (beginning with a block of existential quantifiers) trailing by a formula where at most first-order quantifiers occur. Σ^1_n-formulæ of MSO-logic are called *existential monadic second-order formulæ* or EMSO-formulæ.

Example 1-IV.3. Consider a language L over the alphabet $A = \{a, b\}$ where any two occurrences of a are separated by an odd number of b's. L can be expressed by the following MSO sentence:

$$\begin{aligned}\phi \;=\; & \forall x \forall y \Big(Q_a(x) \wedge Q_a(y) \wedge (x < y) \wedge \forall z \big((x < z) \wedge (z < y) \Rightarrow \neg Q_a(z) \big) \\ & \Rightarrow \exists X \big(X(x) \wedge X(y) \wedge \forall p \forall q (S(p,q) \Rightarrow (X(p) \Leftrightarrow \neg X(q))) \big) \Big)\end{aligned}$$

Here the first part of the formula says that x and y are two positions carrying the letter a such that no other a appears between them. Then the second part identifies the set X as containing the position of the first a, then every second position and finally the position of the next letter a.

1-IV.5-a Interpretation of MSO formulæ

The following somewhat over-specialized model is justified by our interest in only interpreting expressions in words; and the fact that we do not deal with second-order variables of arity more than one renders it sufficient.

Let $\mathcal{V}_1$ be a finite set of first-order variables, and $\mathcal{V}_2$ – a finite set of monadic second order variables. A $(\mathcal{V}_1, \mathcal{V}_2)$-*structure* over A is a word

$$w = (a_1, S_1, T_1) \cdots (a_n, S_n, T_n) \in (A \times 2^{\mathcal{V}_1} \times 2^{\mathcal{V}_2})^*$$

such that

$$(a_1, S_1) \cdots (a_n, S_n)$$

is a $\mathcal{V}_1$-structure. No constraints are imposed on the occurrences of the second-order variables in the structure. The definition of $w \models_{\mathcal{I}} \phi$ is the same as for the first-order expressions, with the addition of two new clauses:

1. if x is a first-order variable and X is a second-order variable then $w \models_{\mathcal{I}} X(x)$ if and only if w contains a letter (a_i, S_i, T_i) such that $x \in S_i$ and $X \in T_i$;

2. if X is a second-order variable, then $w \models_{\mathcal{I}} \exists X\phi$ if and only if there exists a (possibly empty) set J of positions in w with the following property: the $(\mathcal{V}_1, \mathcal{V}_2)$-structure w' formed by replacing each letter (a_i, S_i, T_i), with $i \in J$, by $(a_i, S_i, T_i \cup \{X\})$ satisfies ϕ.

The language $L(\phi)$ defined by an MSO expression ϕ is the set of $(\mathcal{V}_1, \mathcal{V}_2)$-structures that satisfy ϕ.

Chapter 2

Finite Monoids

2-I Introduction

In this chapter we present a more algebraic approach to languages as recognizable sets, with monoids replacing finite automata. S. Eilenberg (cf. [Eil76]) showed that monoids provide a powerful and systematic tool for language classification.

2-II The structure of finite monoids

The pair $(S, \times)$ where S is a set and $\times$ is a (binary) associative operation is a *semigroup*. It is customary to write "semigroup S" rather than "semigroup $(S, \times)$".

Notation.

1. Juxtaposition ab is a shorthand for $a \times b$.

2. If $P_1, P_2, \ldots, P_n$ are nonempty subsets of a semigroup S then $P_1P_2 \cdots P_n = \{p_1p_2 \cdots p_n | p_i \in P_i, 1 \leq i \leq n\}$. If $P = P_1 = P_2 = \cdots = P_n$ we write P^n instead of $P_1P_2 \cdots P_n$.

A *monoid* $(M, \cdot, 1)$ is a set M with a binary operation, denoted by $\cdot$, and a distinguished element 1, such that $(M, \cdot)$ is a semigroup with an *identity* 1, i.e. for all $x \in M$, $1 \cdot x = x \cdot 1 = x$. We usually write "monoid M" instead of "monoid $(M, \cdot, 1)$".

An element z of a monoid M is a *zero* of M if for all $s \in M$, $z = zs = sz$. We usually denote such an element z by 0.

Let z_1, z_2 be two zeros of a monoid M. By definition: $z_1 z_2 = z_1$ and $z_1 z_2 = z_2$. Whence, $z_1 = z_2$, i.e. a monoid can have at most one zero. A similar argument shows that a monoid contains a single identity element.

We now turn to subsets of a finite monoid (semigroup) exhibiting special properties.

A *subsemigroup* T of a semigroup S is a subset of S such that $x_1 \in T$ and $x_2 \in T$ imply $x_1 x_2 \in T$. This is equivalent to $T^2 \subseteq T$.

A subset T of a monoid M is a *submonoid* of M if it is closed under the operation of M and contains the identity element, i.e.

(a) $1 \in T$ and

(b) $T^2 \subseteq T$.

Clearly, a submonoid of a monoid is a monoid in its own right.

A monoid M *is generated by* its subset G if every element of M can be written as a product of some elements of G.

A nonempty subset T of a monoid M is a *left ideal* of M if $MT \subseteq T$; a *right ideal* of M if $TM \subseteq T$; a *two-sided ideal* (or simply an *ideal*) if it is both a left and a right ideal, i.e. $MT \cup TM \subseteq T$.

The intersection of all ideals of a monoid M is the *kernel* of M.

A monoid M is *simple* (*left-simple, right-simple*) if no proper subset of M is an ideal (respectively, left ideal, right ideal) of M.

Lemma 2-II.1 (cf. [CP67]). *The set of all ideals of a finite monoid M is closed under intersection and arbitrary union. The intersection of a finite number of ideals is an ideal.*

The lemma above holds for the set of all left (right) ideals of M as well.

An element e of a monoid M is *idempotent* if $e^2 = e$. Let s be an element of a finite monoid M and let S be the submonoid generated by s. The sequence $s^0 = 1, s, s^2, s^3, \ldots$ contains only finitely many distinct elements of S, for S is finite and closed under product. Let p be the smallest positive integer such that there exists an integer $m > 0$ satisfying

$$s^p = s^{p+m}.$$

Let us fix the smallest such m and name it q. Choosing $r \geq 0$ such that $p + r \equiv 0 (\mathrm{mod} q)$ yields for some $i \geq 1$:

$$(s^{p+r})^2 = s^{2(p+r)} = s^{(p+r)+iq} = s^{(p+q)+r} = s^{p+r}$$

That is, s^{p+r} is an idempotent element of S.

Furthermore, the elements $1, s, s^2, \ldots, s^{p+q-1}$ are all distinct. For any integer $n \geq q$ we have $n = iq + j$ (with $i \geq 1$, $0 \leq j < q$) and

$$s^{p+n} = s^{p+iq+j} = s^{p+j},$$

whence

$$S = \{1, s, s^2, s^3, \ldots, s^{p+q-1}\}.$$

Observe also that the set $G = \{s^p, s^{p+1}, \ldots, s^{p+q-1}\}$ is a maximal subgroup of M since the mapping $\phi : G \mapsto \mathbb{Z}_q$ defined by $\phi(s^{p+k}) = p + (k \bmod q)$ is an isomorphism. Since every $s \in S \setminus \{1\}$ has a power in G, s^{p+r} is the only other idempotent of S beside 1. The structure of the submonoid S therefore resembles a frying pan with the dish representing the group G as shown in figure 2.1.

We thus have the following results:

Proposition 2-II.2. *If s is an element of a finite monoid M, then the submonoid S generated by s contains a unique maximal subgroup.*

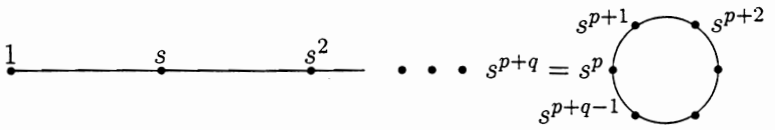

Figure 2.1: The structure of the submonoid S.

Corollary 2-II.3. *Every non-empty finite semigroup contains an idempotent.*

A monoid M is *aperiodic* if for all $x \in M$ there exists an integer n such that $x^n = x^{n+1}$.

An element a of a monoid M is *regular* if $a = asa$ for some $s \in M$. If every element of M is regular, M is *regular*. An element x of M is an *inverse* of a if $a = axa$ and $x = xax$. In a monoid every regular element has an inverse.

A monoid in which every element has a unique inverse is called a *group*. A group is *cyclic* if it is the set of powers of a single element. A cyclic group is commutative.

A *subgroup* H of a group G is a subset of G which is itself a group under the operation of G. Every group has two trivial subgroups: the group itself and the group consisting of the identity. Any non-cyclic group G has necessarily a non-trivial subgroup.

For any group G, and any element $g \in G$, one has

$$Gg = \{g_i g |\ g_i \in G\} = G.$$

Indeed, every g_1 is obtainable as a product $g_1 g^{-1} \cdot g = g_1$ and $g_1 g^{-1}$ is equal to some $g_i \in G$.

If G is a group and H is a subgroup of G, then Ha, where $a \in G$, is called a *right coset* of H in G. (We have a similar definition for a *left coset* .) Assume $c \in Ha \cap Hb$. Then there exists an element $h \in H$ such that $c = ha$, i.e.

$$a = h^{-1}c \qquad \text{and} \qquad Ha = Hh^{-1}c = Hc.$$

In the same way $Hb = Hc$, i.e. if two right cosets of H in G have a common element, they coincide, otherwise they are disjoint.

A subgroup H of G is called a *normal subgroup* if its right cosets coincide with the left ones, i.e. $Ha = aH$. In this case one has $a^{-1}Ha = H$ and hence

$$Ha \cdot Hb = Ha(a^{-1}Ha)b = HHab = Hab,$$

i.e. the product of two right cosets is a right coset. A group which has only trivial normal subgroups is called a *simple group*.

Given a group G and a normal subgroup H, one can use the partition of G into (right) cosets of H to build the factor group G/H, whose elements are the blocks of the partition, i.e. the cosets of H in G.

The next result presents decomposition of finite left-simple semigroups.

Lemma 2-II.4 (cf. [CP67]). *Every finite left-simple semigroup S is isomorphic to a direct product $T \times G$, where G is a group and T is a left-zero semigroup.*

Proof: If s is an element of S, then either $Ss \subset S$ (in which case Ss is a proper left ideal of S), or $Ss = S$, in which case

$$\pi_s : \ t \mapsto ts$$

is a permutation of elements of S. We consider the right action of s on S and

$$\pi_{s'}(\pi_s(t)) = \pi_{ss'}(t) = tss'.$$

Then

$$G = \{\pi_s |\ s \in S\}$$

is a group of permutations of S acting on S on the right. Let T be the set of orbits of this action; $\mathcal{O}_s$ denotes the orbit containing s. We then define a multiplication on T by setting

$$\forall \mathcal{O}_s, \forall \mathcal{O}_t \in T : \ \mathcal{O}_s \cdot \mathcal{O}_t = \mathcal{O}_s,$$

to ensure that T is a left-zero semigroup.

Claim:

$$\phi : \ s \mapsto (\mathcal{O}_s, \pi_s)$$

is a bijection between S and $T \times G$.

We first show that ϕ is surjective. Consider $(\mathcal{O}, \pi_s) \in T \times G$ and $t \in \mathcal{O}$. Then for all $x \in S$

$$\pi_s(x) = xs = \pi_s(x\pi_t^{-1}(t)).$$

Since G is a group, there exists $u \in S$ such that $\pi_u = \pi_t^{-1}$ and hence $\pi_s = \pi_{tus}$ with $tus \in \mathcal{O}$. Thus $(\mathcal{O}, \pi_s) = \phi(tus)$ and ϕ is surjective.

To see that o is injective, assume $(\mathcal{O}_s, \pi_s) = (\mathcal{O}_{s'}, \pi_{s'})$. Then $su = s'$ for some $u \in S$ and thus $\pi_s = \pi_{s'} = \pi_u(\pi_s)$, i.e. π_u is the identity permutation. Hence $s' = su = s$.

And finally o is a function preserving multiplication since

$$ss' \in \mathcal{O}_s$$

and

$$\phi(ss') = (\mathcal{O}_{ss'}, \pi_{ss'}) = (\mathcal{O}_s, \pi_{ss'}) = (\mathcal{O}_s, \pi_s)(\mathcal{O}_{s'}, \pi_{s'}) = \phi(s)\phi(s').$$

Q.E.D.

2-III Homomorphisms and the syntactic congruence

A *homomorphism*[1] φ from a semigroup $(S, \cdot)$ to a semigroup $(S', \star)$ is a mapping φ from the set S into the set S' such that

$$\varphi(x \cdot y) = \varphi(x) \star \varphi(y)$$

[1]The word *morphism* is also used by some authors.

for every $x, y \in S$. To denote such a mapping we write $\varphi : S \mapsto S'$. If φ is also a surjective mapping, then φ is called a homomorphism from S *onto* S', and S' is called the *homomorphic image* of S. In case the mapping φ above is injective, it is called a *one-to-one homomorphism.* An *isomorphism* from S to S' is a homomorphism which is both surjective and injective.

A homomorphism φ from a monoid $(M, \cdot, 1)$ to a monoid $(M', \star, 1')$ is a semigroup homomorphism $\varphi : M \mapsto M'$ such that

$$\varphi(1) = 1'.$$

The terminology for surjective and injective homomorphisms of monoids is the same as above. It will be clear from the context whether the intended meaning is "monoid homomorphism" or "semigroup homomorphism".

We shall say that a monoid N is a *quotient* of a monoid M if there exists a surjective homomorphism $\phi : M \mapsto N$.

A monoid M is said to *divide* a monoid N (written $M \prec N$) if M is a quotient of a submonoid of N.

The notions of quotient and division are defined similarly for semigroups.

Let A be a finite alphabet and let $L \subseteq A^*$. Consider the following equivalence relation $\equiv_L$ on A^*:

$$x \equiv_L y \iff \{(u, v) \in A^* \times A^* : uxv \in L\} = \{(u, v) \in A^* \times A^* : uyv \in L\}.$$

It is easy to show that if $x \equiv_L y$ and $a \in A$, then

$$xa \equiv_L ya \qquad \text{and} \qquad ax \equiv_L ay.$$

It follows that the equivalence relation $\equiv_L$ is a congruence on A^*. It is called the *syntactic congruence* of L. The quotient of A^* by $\equiv_L$, denoted $M(L)$, is the *syntactic monoid* (or *syntactic semigroup* for A^+) of L and the projection $\eta_L : A^* \mapsto M(L)$ is termed the *syntactic morphism* of L.

2-IV Equivalence of automaton and monoid

A monoid M is said to recognize $L \subseteq A^*$ if there exists a subset X of M and a homomorphism $\phi : A^* \mapsto M$ such that $L = \phi^{-1}(X)$. (We also say that the homomorphism ϕ recognizes a language L.)

We next show that the two notions of recognizable sets – by finite automata and by finite monoids – are equivalent.

Theorem 2-IV.1 (cf. [MP71]). *A subset L of A^* is regular if and only if it is recognized by a finite monoid.*

Proof: Let $L \subseteq A^*$ be a regular language and $\mathcal{A} = (Q, i, F, \lambda)$ be a deterministic finite automaton recognizing L. We define an equivalence relation $\sim$ on A^* by

$$x \sim y \iff \forall q \in Q : q \cdot x = q \cdot y.$$

The number of equivalence classes of the equivalence relation $\sim$ does not exceed $|Q|^{|Q|}$. Suppose now $x \sim y$ and $uxv \in L$ for some $u, v \in A^*$. We then derive:

$$i \cdot (uyv) = ((i \cdot u) \cdot y) \cdot v = ((i \cdot u) \cdot x) \cdot v = i \cdot (uxv) \in F.$$

Thus $uyv \in L$. A similar derivation will show that $uyv \in L$ implies $uxv \in L$. Therefore,

$$x \sim y \Rightarrow x \equiv_L y,$$

which shows that the equivalence relation $\sim$ refines $\equiv_L$, and hence $|M(L)| \leq |Q|^{|Q|}$.

Conversely, let us assume $M(L)$ is finite. First observe that if $x \in L$ and $x \equiv_L y$, then $y \in L$, because $x = 1 \cdot x \cdot 1$. We construct a deterministic finite automaton $\mathcal{T} = (Q, i, F, \lambda)$ recognizing L by setting: the set of states Q is the set of elements of $M(L)$, the initial state i is 1, the set of final states F is the set of classes of words in L and the transition function λ is given for all $a \in A$ by

$$\lambda([w], a) = [wa],$$

where $[v]$ denotes the $\equiv_L$-class of a word v. Thus a word w is accepted by $\mathcal{T}$ if and only if $1 \cdot w = [w]$ is the class of a word in L. By the observation above, this is true if

and only if $w \in L$. Therefore, $\mathcal{T}$ recognizes L; and since $M(L)$ is finite, L is regular. *Q.E.D.*

Let $\mathcal{A} = (Q, i, F, \lambda)$ be a deterministic finite automaton operating over a finite alphabet A. For each word $w \in A^*$ we define a corresponding *state-transition function* $\mu_w : Q \mapsto Q$, denoted by a two-row matrix

$$\mu_w = \begin{pmatrix} m_{1j} \\ m_{2j} \end{pmatrix},$$

where the first row m_{1j} is an (ordered) permutation of $q_j \in Q$ $(1 \leq j \leq |Q|)$ and elements of the second row are $m_{2j} = \lambda(q_j, w)$. The set of these maps under the operation of functional composition

$$\mu_v \circ \mu_u = \mu_{uv}$$

forms a monoid, termed the *transition monoid* of $\mathcal{A}$, denoted by $M(\mathcal{A})$.

Theorem 2-IV.2 (cf. [MP71]). *Let $\mathcal{A}$ be the minimal automaton of L. Then $M(\mathcal{A})$ and $M(L)$, the syntactic monoid of L, are isomorphic.*

Theorem 2-IV.3. *Let $L \subseteq A^*$ be a language and $\eta_L : A^* \mapsto M(L)$ - its syntactic morphism. Let $\phi : A^* \mapsto M$ be a homomorphism. Then:*

1. *ϕ recognizes L if and only if there exists a homomorphism $\psi : \phi(A^*) \mapsto M$ such that $\psi \circ \phi = \eta_L$ (i.e. η_L factors through ϕ).*

2. *A monoid M recognizes L if and only if $M(L) \prec M$.*

Proof: If ϕ recognizes L then there exists $X \subseteq M$ such that $L = \phi^{-1}(X)$. Suppose $\phi(w_1) = \phi(w_2)$. Then $xw_1y \in L$ implies $\phi(xw_2y) \in X$ since $\phi(xw_1y) \in X$ and $\phi(xw_1y) = \phi(xw_2y)$. Thus $xw_2y \in L$. Similarly, $xw_2y \in L \Rightarrow xw_1y \in L$. Therefore, $\phi(w_1) = \phi(w_2) \Rightarrow w_1 \equiv_L w_2$. Hence η_L factors through ϕ, and $M(L)$ is a homomorphic image of $\phi(A^*)$, proving $M(L) \prec M$.

Conversely, suppose there exists a homomorphism $\psi : \phi(A^*) \mapsto M$ such that $\psi \circ \phi = \eta_L$. If $\phi(w) \in \phi(L)$ then $\eta_L(w) \in \eta_L(L)$, whence $\phi(w) \in \phi(L) \iff w \in L$. That is, ϕ recognizes L. Let M be a monoid and $M(L) \prec M$, then there exists a

submonoid M' of M and a surjective homomorphism $\psi : M' \mapsto M$. For every $a \in A$ fix $\phi(a) \in M'$ such that $\psi(\phi(a)) = \eta_L(a)$. We then extend the domain of ϕ to A^*, i.e. ϕ is a homomorphism $\phi : A^* \mapsto M$ such that η_L factors through ϕ. Then M recognizes L since ϕ recognizes L. *Q.E.D.*

The next results apply to operations on languages.

Proposition 2-IV.4 (cf. [Arb68]). *Let L, K be two languages of A^* recognized respectively by monoids M_L and M_K and let M be a monoid. Then*

1. *if M recognizes L, M recognizes $A^* \setminus L$;*
2. *$L \cap L$ and $L \cup K$ are recognized by $M_L \times M_K$;*
3. *if M recognizes L, M recognizes $K^{-1}L$ and LK^{-1}.*

2-V Green's relations

The equivalence relations we are about to introduce were first formulated by J. A. Green in 1951 ([Gre51]) and have become fundamental in the theory of semigroups.

Definition 2-V.1. Let M be a monoid. Green's relations are defined by the following equivalences:

$$\begin{aligned} a\mathcal{R}b &\iff aM = bM & \mathcal{D} &= \mathcal{R} \vee \mathcal{L} \\ a\mathcal{L}b &\iff Ma = Mb & \mathcal{H} &= \mathcal{R} \cap \mathcal{L} \\ a\mathcal{J}b &\iff MaM = MbM \end{aligned}$$

(cf. figure 2.2)

We also introduce reflexive and transitive relations based on the above:

$$\begin{aligned} a \leq_{\mathcal{R}} b &\iff aM \subseteq bM \\ a \leq_{\mathcal{L}} b &\iff Ma \subseteq Mb \\ a \leq_{\mathcal{J}} b &\iff MaM \subseteq MbM \\ a \leq_{\mathcal{H}} b &\iff a \leq_{\mathcal{R}} b \text{ and } a \leq_{\mathcal{L}} b \end{aligned}$$

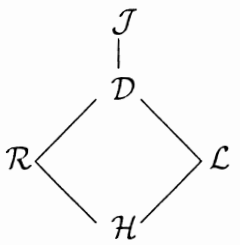

Figure 2.2: The inclusion of various Green's equivalences.

Notation. If a is an element of a monoid M, then by R_a, L_a, H_a, J_a and D_a we mean respectively the $\mathcal{R}$-class, $\mathcal{L}$-class, $\mathcal{H}$-class, $\mathcal{J}$-class and $\mathcal{D}$-class containing a.

Lemma 2-V.1 ([Gre51]). *In a finite monoid, the relations $\mathcal{R}$ and $\mathcal{L}$ commute. Consequently the relation $\mathcal{D} = \mathcal{R}\mathcal{L} = \mathcal{L}\mathcal{R}$ is the smallest one containing $\mathcal{R}$ and $\mathcal{L}$.*

Proposition 2-V.2 ([Gre51]). *In a finite monoid, $\mathcal{D} = \mathcal{J}$.*

Proposition 2-V.3. *Let R be an $\mathcal{R}$-class and L be an $\mathcal{L}$-class of a finite monoid M. Then $R \cap L \neq \emptyset$ if and only if R and L are within the same $\mathcal{J}$-class.*

Proof: If $a \in R \cap L$ the result is immediate: $R = R_a$ and $L = L_a$ and therefore J_a must contain both of them.

Conversely, suppose R and L are in the same $\mathcal{J}$ class of M. Then for every $x \in R$ and $y \in L$ there exists $a \in M$ such that $x\mathcal{R}a$ and $a\mathcal{L}y$ (since $x\mathcal{J}y$ and $\mathcal{J} = \mathcal{R}\mathcal{L}$). Hence, $a \in R \cap L$. *Q.E.D.*

A $\mathcal{D}$-class (or a $\mathcal{J}$-class) of a finite monoid can thus be viewed as a table where rows represent $\mathcal{R}$-classes and columns – $\mathcal{L}$-classes. $\mathcal{H}$-classes lie at the intersections (fig 2.3). The presence of an idempotent in an $\mathcal{H}$-class is indicated by a star (*).

Lemma 2-V.4 (cf. [CP67]). *Let m be an element of a finite monoid M. If $L_m = J_m$ and L_m contains an idempotent, then L_m is a subsemigroup of M.*

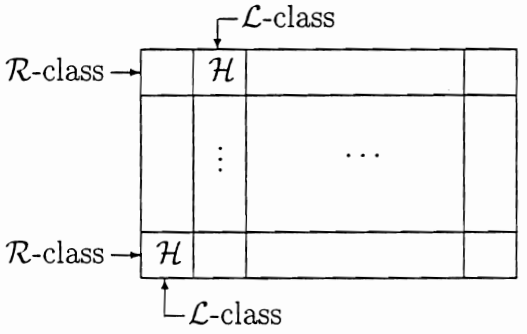

Figure 2.3: The $\mathcal{D}$-class structure.

Proof: Let $e \in L_m$ be idempotent, so $L_e = L_m = J_m = J_e$. Consider two elements t_1, t_2 of L_e: $t_1 = ue, t_2 = ve$ for some $u, v \in M$. Thus $t_1 t_2 = ueve \in Me$. On the other hand, $e = xt_1 = yt_2$ for some $x, y \in M$. Thus $e = e^2 = xt_1 yt_2$. Since $xt_1 y = xuey \in MeM$ and $e = xt_1 yt_2 \in Mxt_1 yM$, we conclude that $xt_1 y$ and e generate the same two-sided ideal of M: $J_{xt_1 y} = J_e = L_e = L_{t_1}$. Hence there exists $w \in M$ such that $xt_1 y = wt_1$. Thus $e = wt_1 t_2$ and e is in the left ideal generated by $t_1 t_2$ and $t_1 t_2$ is in the left ideal generated by e. This implies $t_1 t_2 \in L_e$ and therefore $L_e = L_m$ is closed under product. *Q.E.D.*

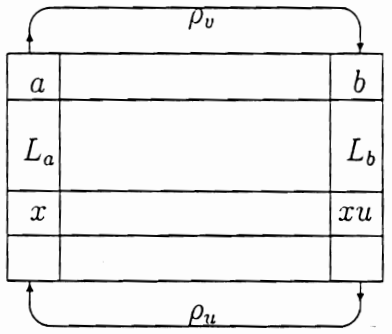

Figure 2.4: Green's Lemma.

Theorem 2-V.5 (Green's Lemma, [Gre51]). *Let $a, b \in M$ be such that $a\mathcal{R}b$. Then there exist $u, v \in M$ satisfying $au = b$ and $bv = a$. If ρ_u, ρ_v are the right translations defined respectively by $\rho_u(x) = xu$ and $\rho_v(x) = xv$, then $\rho_u : L_a \mapsto L_b$ and $\rho_v : L_b \mapsto L_a$ are inverse bijections preserving the $\mathcal{H}$-classes, i.e.*

$$\forall x, y \in L_a : x\mathcal{H}y \iff \rho_u(x)\mathcal{H}\rho_u(y)$$

and

$$\forall x, y \in L_b : x\mathcal{H}y \iff \rho_v(x)\mathcal{H}\rho_v(y)$$

Proof: (figure 2.4) Let $x\mathcal{L}a$. By definition, $Mx = Ma$; therefore $Mxu = Mau$, or $xu\mathcal{L}au = b$. Hence ρ_u is a function from L_a to L_b. Since there exist $t \in M$ such that $ta = x$ we have

$$\rho_v(\rho_u(x)) = \rho_v(xu) = \rho_v(tau) = \rho_v(tb) = tbv = ta = x,$$

i.e. the composition $\rho_u \circ \rho_v$ is the identity function on L_a. A similar argument shows ρ_v to be a function from L_b to L_a and $\rho_v \circ \rho_u$ to be the identity on L_b.

Since every $x \in L_a$ is $\mathcal{R}$-equivalent to xu and every $z \in L_b$ is $\mathcal{R}$-equivalent to zv, we conclude:

$$(x\mathcal{H}y) \Rightarrow (xu\mathcal{H}yu) \qquad \text{and} \qquad (xu\mathcal{H}yu) \Rightarrow (x = xuv\mathcal{H}yuv = y).$$

Q.E.D.

The case of two $\mathcal{L}$-equivalent elements is symmetric.

Proposition 2-V.6 ([CM56]). *If a, b are two elements of a $\mathcal{J}$-class of a monoid M, then:*

$$(ab \in R_a \cap L_b) \iff \exists e \in R_b \cap L_a : e^2 = e$$

The situation is summarized in the figure 2.5.

a	R_a	ab
L_a		L_b
$*e$	R_b	b

Figure 2.5: Proposition 2-V.6.

Proof: Suppose $ab \in R_a \cap L_b$. By Green's Lemma $\rho_b : L_a \mapsto L_b$ is a bijection. Chose an element $e \in R_b \cap L_a$ such that $\rho_b(e) = eb = b$. Since e and b are $\mathcal{R}$-equivalent there exists $u \in M$ such that $e = bu$. Then $e^2 = ebu = bu = e$, i.e. e is idempotent.

Conversely, suppose e is an idempotent element, $e \in R_b \cap L_a$. Then $e\mathcal{R}b \Rightarrow \exists u : b = eu$. Hence, $eb = eeu = eu = b$. Similarly, $e\mathcal{L}a \Rightarrow \exists v : a = ve$ whence $ae = vee = ve = a$. Also, $e\mathcal{R}b \Rightarrow a = ae\mathcal{R}ab$ and $e\mathcal{L}a \Rightarrow b = eb\mathcal{L}ab$. That is, $ab \in R_a \cap L_b$. *Q.E.D.*

Lemma 2-V.7 (cf. [CP67]). *Let x and m be elements of a finite monoid M. Then*

$$xm\mathcal{J}m \Rightarrow xm\mathcal{L}m.$$

Proof: $\mathcal{J}$-equivalence of xm and m implies the existence of $p, q \in M$ such that $m = p \cdot xm \cdot q$. Then there exists a positive integer k such that both $e = (px)^k$ and $f = q^k$ are idempotent and we have $m = (px)^k mq^k = emf$. Thus

$$m = em = (px)^{k-1}p \cdot xm,$$

so m belongs to the left ideal generated by xm. Hence, $L_{xm} = L_m$. *Q.E.D.*

Lemma 2-V.8 (cf. [Lal79]). *Let H be an $\mathcal{H}$-class of a monoid M. The following conditions are equivalent:*

1. $\exists e \in H : e^2 = e$
2. $\exists a, b \in H : ab \in H$
3. *H is a maximal group in M*

Proof: $3 \Rightarrow 1$. If H is a group, it contains an idempotent.

$1 \Rightarrow 2$. $H = R_a \cap L_b = R_b \cap L_a$ and by proposition 2-V.6, $ab \in H$.

$2 \Rightarrow 3$. By proposition 2-V.6, H must contain an idempotent e. For two arbitrary elements of H, x and y: $e \in R_x \cap L_y = R_y \cap L_x$ implies (by the same proposition) $xy \in H$. Thus H is a semigroup. Furthermore, $e\mathcal{R}x$ means there exists $u \in H$ such that $x = eu$; then $ex = eeu = eu = x$. Similarly, from $e\mathcal{L}x$ we derive $xe = x$. That is, $ex = x = xe$ and H is a monoid. Let $\rho_x : H \mapsto H$ be a bijection defined by Green's Lemma. Then there exist x' such that

$$\rho_x(x') = x'x = e,$$

which shows that H is a group. Since every element of a group containing e is $\mathcal{H}$-equivalent to e, H is a maximal group. *Q.E.D.*

Proposition 2-V.9 (cf. [Lal79]). *Two maximal subgroups of a finite monoid M contained in the same $\mathcal{J}$-class are isomorphic.*

Proof: By Lemma 2-V.8 two maximal subgroups of a finite monoid M are $\mathcal{H}$-classes H_e and H_f containing respectively idempotents e, f. Since both H_e and H_f are within the same $\mathcal{J}$-class there exists $a \in H_a$, where $H_a = R_e \cap L_f$ (Lemma 2-V.3). Then:

$$a\mathcal{R}e \Rightarrow ea = a \qquad \text{and} \qquad a\mathcal{L}f \Rightarrow (\exists a' \in M : a'a = f) \text{ and } (af = a).$$

By Green's Lemma $\rho_a(x) = xa$ is a bijection from H_e onto H_a. Similarly, by the dual version of Green's Lemma we have that $\lambda_{a'} = a'x$ is a bijection from H_a onto H_f. Therefore the composition $\rho_a \circ \lambda_{a'}$ is a bijection mapping every x in H_e onto $a'xa$ in H_f. Clearly,

$$\rho_a \circ \lambda_{a'}(e) = a'ea = a'a = f.$$

To see that $\rho_a \circ \lambda_{a'}$ is an isomorphism, we first observe that aa' is an idempotent of R_a:

$$(aa')^2 = aa'aa' = afa' = aa'.$$

Hence, for every element $x \in R_a$ we have $aa'x = x$. For arbitrary $x, y \in H_e$, the product $xy \in H_e$. Their images under $\rho_a \circ \lambda_{a'}$ exhibit the same property:

$$(a'xa)(a'ya) = a'x(aa'y)a = a'xya.$$

Q.E.D.

A $\mathcal{J}$-class is called *regular* if all its elements are regular. (We have similar definitions for regular $\mathcal{R}$, $\mathcal{L}$ and $\mathcal{H}$-classes). The next proposition further explores the structure of a regular $\mathcal{J}$-class.

Proposition 2-V.10. *Let J be a $\mathcal{J}$-class of a finite monoid M. The following are equivalent:*

1. *J is regular*
2. *J contains a regular element*
3. *every $\mathcal{L}$-class contained in J has an idempotent*
4. *every $\mathcal{R}$-class contained in J has an idempotent*
5. *J contains an idempotent*
6. *$\exists x, y \in J : xy \in J$*

Proof: $1 \Rightarrow 2$. By definition.

$2 \Rightarrow 3, 4$. Suppose a is a regular element of J. Then $a = asa \Rightarrow a\mathcal{L}sa$. Note also that sa is idempotent:

$$(sa)^2 = sasa = s(asa) = sa.$$

Similarly, $a = asa \Rightarrow a\mathcal{R}as$ and

$$(as)^2 = asas = (asa)s = as.$$

$3, 4 \Rightarrow 2$. Let e be an idempotent element of M in J. Then $a\mathcal{R}e \Rightarrow \exists u \in M$: $au = e$ and $ea = a$, whence

$$a = ea = eea = auea = asa.$$

By the same reasoning $a\mathcal{L}f$ (where f is idempotent) implies $\exists v \in M : va = f$ and $af = a$. Therefore,

$$a = af = aff = afva = ata.$$

$2 \Rightarrow 1$. Let a be a regular element of M in J and b - an element in J. Then $a\mathcal{J}b \iff \exists c \in J : a\mathcal{R}c \wedge c\mathcal{L}b$. Since a is regular, $R_a = R_c$ contains an idempotent and therefore c is regular. Also, b must be regular because $L_c = L_b$ has an idempotent.

$3, 4 \Rightarrow 5$. Obvious.