Algebraic theory of automata: historical perspective and new advances (Part I)

Jean-Éric Pin[1]

[1]LIAFA, CNRS and Université Paris Diderot

Warsaw, 2011

Summary

First part

(1) Languages, automata and operations
(2) Some famous results and problems
(3) The variety approach

Second part

(1) Profinite topology on A^*
(2) Equational theories for lattices of languages
(3) Some examples
(4) Profinite topologies

Part I

Languages, automata, operations

Operations on languages

Boolean operations :

(1) union and intersection (positive Boolean operations)

(2) complement ($L \to L^c$)

Quotient: $u^{-1}L = \{v \mid uv \in L\}$ and $Lu^{-1} = \{v \mid vu \in L\}$

Star: L^* is the submonoid of A^* generated by L.

Concatenation product

- Product:

$$L_1L_2 = \{u_1u_2 \mid u_1 \in L_1, u_2 \in L_2\}$$

- Marked product: given letters $a_1, \ldots, a_k$ and languages $L_0, L_1, \ldots, L_k$ of A^*,

$$L_0a_1L_1 \cdots a_kL_k = \{u_0a_1u_1 \cdots a_ku_k \mid u_0 \in L_0, \ldots, u_k \in L_k\}$$

Let $\mathcal{L}$ be a class of languages. An $\mathcal{L}$-polynomial of languages is a finite union of languages of the form $L_0a_1L_1 \cdots a_kL_k$ where $a_1, \ldots, a_k$ are letters and $L_0, \ldots, L_k$ belong to $\mathcal{L}$.

Shuffle product

The shuffle of two words u and v of A^* is the set $u \sqcup\!\sqcup\, v$ of words of A^* of the form $u_1 v_1 \cdots u_n v_n$, with $n \geqslant 0$, $u_1, \ldots, u_n, v_1, \ldots, v_n \in A^*$, $u_1 \cdots u_n = u$, $v_1 \cdots v_n = v$.

$$ab \sqcup\!\sqcup\, ba = \{\ abba\ baab,\ abab,\ baba\ \}$$

The shuffle of two languages K and L of A^* is the language

$$K \sqcup\!\sqcup\, L = \bigcup_{u \in K,\ v \in L} u \sqcup\!\sqcup\, v.$$

Part II

Famous results and problems

(1) The star-height problem
(2) Concatenation hierarchies
(3) Logic expressiveness

The general setting

Given

(1) a class of languages $\mathcal{B}$ (the base),

(2) a set of operators

(3) some rules to use the operators,

Describe the languages expressible from $\mathcal{B}$ by using the operators according to the rules.

Rational languages

Basic class: languages $\{a\}$, with $a \in A$.
Operators: union, concatenation and star.
Defines the rational (or regular) languages.

Theorem (Kleene 1956)

Let L be a subset of A^*. The following conditions are equivalent

(1) *L is rational,*

(2) *the syntactic monoid of L is finite,*

(3) *L is recognized by a finite automaton.*

Star-free languages

Basic class: languages $\{a\}$, with $a \in A$.
Operators: Boolean operations and product.
Defines the star-free languages.

Theorem (Schützenberger 1965)

A language is star-free iff its syntactic monoid is finite and aperiodic.

Star-height of an extended regular expression

Maximum number of nested star operators occurring in the expression.

Star-free expression for the language $(ab)^*$:

$$(b\emptyset^c \cup \emptyset^c a \cup \emptyset^c aa\emptyset^c \cup \emptyset^c bb\emptyset^c)^c$$

An expression of star-height one:

$$(\{a, ba, abb\}^* bba \cap (aa\{a, ab\}^*))^c bbA^*$$

An expression of star-height two:

$$\Big(a(ba)^* abb\Big)^* bba \cap (aa\{a, ab\}^*))^c bbA^*$$

Star-height problem

The star-height of a language is the minimal star-height of an expression representing the language.

Star-height 0 $=$ star-free languages

Star-height 1 $=$?? Open problem!

It is not known whether there are languages of star-height 2!

Concatenation hierarchies

- Level 0 (basic level): $\emptyset$ and A^*
- Level $n + 1/2$: polynomial of languages of level n.
- Level $n + 1$: Boolean algebra generated by the languages of level $n + 1/2$.

Theorem (Brzozowski and Knast 1978)

The concatenation hierarchy is infinite.

One can built other concatenation hierarchies, starting from different basic levels: finite-cofinite languages (Brzozowski), group languages (Pin).

Decidability problems

Problem: Given a nonnegative integer n and a rational language L, decide whether L has level n (resp. $n + 1/2$).

This problem is equivalent to a very natural problem in finite model theory (decidability of the Σ_n hierarchy of linearly ordered coloured structures).

Trivial case. A language has level 0 iff its syntactic monoid is trivial.

Büchi's sequential calculus

The sentence $\exists i\ \mathbf{a}i$ defines the language A^*aA^*.

The sentence $\exists i\ \exists j\ \big((i < j) \wedge \mathbf{a}i \wedge \mathbf{b}j\big)$ defines the language $A^*aA^*bA^*$

$j = i + 1$ is a macro for
$(i < j) \wedge \forall k\ \Big((i < k) \rightarrow ((j = k) \vee (j < k))\Big)$.

$j \leqslant i$ is a macro for $j < i \vee j = i$.

The sentence $\exists j\ \forall i\ j \leqslant i \wedge \mathbf{a}j$ defines aA^*.

The sentence $\exists i\ \exists j\ j = i + 1 \wedge \mathbf{a}i \wedge \mathbf{a}j$ defines A^*aaA^*.

Syntax and semantics

For each letter $a \in A$, $\mathbf{a}$ denotes a unary predicate.

A word u is represented as a structure $\big(\mathsf{Dom}(u), (\mathbf{a})_{a \in A}, <\big)$ where $\mathsf{Dom}(u) = \{1, \ldots, |u|\}$ and $\mathbf{a} = \{i \in \mathsf{Dom}(u) \mid u(i) = a\}$.

$<$ is a binary relation symbol, interpreted as the usual order.

Thus, if $u = abbaab$, $\mathsf{Dom}(u) = \{1, \ldots, 6\}$, $\mathbf{a} = \{1, 4, 5\}$ and $\mathbf{b} = \{2, 4, 6\}$.

The main questions

What is the expressive power of a given logic fragment? Is this given fragment decidable?

For instance, which languages can be expressed by a first order (resp. second order, Σ_1-, etc.) formula?

Is it decidable whether a given language is expressible by a first order (resp. second order, Σ_1-, etc.) formula?

Monadic second order

Theorem (Büchi 1960, Elgot 1961)

Monadic second order captures the rational languages.

There is an algorithm to pass from a formula to a rational expression and vice-versa. However

Theorem (Meyer 1975)

There is no elementary time bounded decision procedure for deciding whether a given monadic second order sentence is true or not.

First order

Theorem (McNaughton-Papert 1971)

First order captures exactly the star-free languages.

Star-free languages form the smallest class of languages containing the finite languages which is closed under Boolean operations (union, intersection, complement) and concatenation product (no star !)

Star-free languages

Star-free languages are represented by star-free extended regular expressions (intersection, union, complement and product).

(1) $A^* = \emptyset^c$ is star-free.

(2) $b^* = (A^*aA^*)^c$ is star-free.

(3) $(ab)^* = \big(b\emptyset^c \cup \emptyset^c a \cup \emptyset^c aa\emptyset^c \cup \emptyset^c bb\emptyset^c\big)^c$ is star-free.

(4) $(aa)^*$ is not star-free.

Home work. Which languages are star-free ?

$(aba, b)^*$, $(ab, ba)^*$, $(a(ab)^*b)^*$, $(a(a(ab)^*b)^*b)^*$

Schützenberger theorem

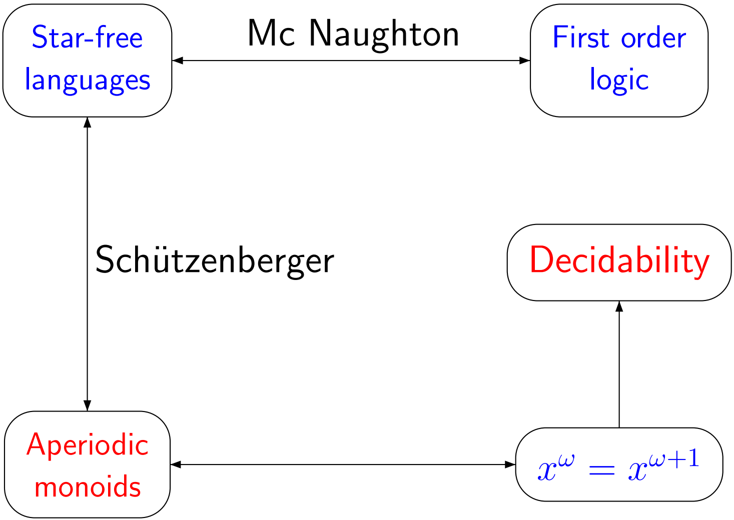
Theorem (Schützenberger 1965)

A language is star-free iff its syntactic monoid is finite and aperiodic (i.e. satisfies an identity $x^n = x^{n+1}$ *for some* $n > 0$*).*

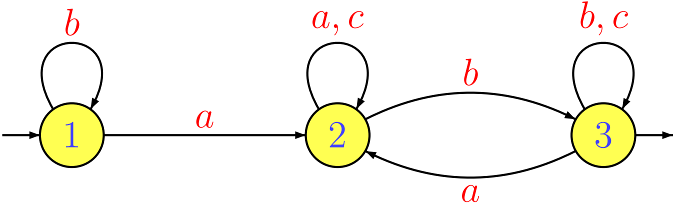
Corollary

It is decidable to know whether a given regular language is first-order expressible.

A virtuous circle



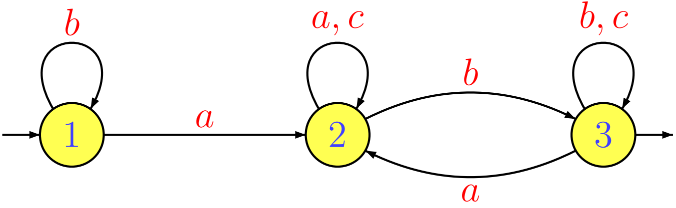
Syntactic monoid of $\{a,b\}^*aA^*b\{b,c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3

Rules:

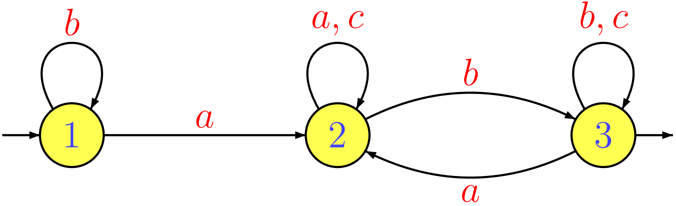
Syntactic monoid of $\{a, b\}^* a A^* b \{b, c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3

Rules:
$aa \rightarrow a$

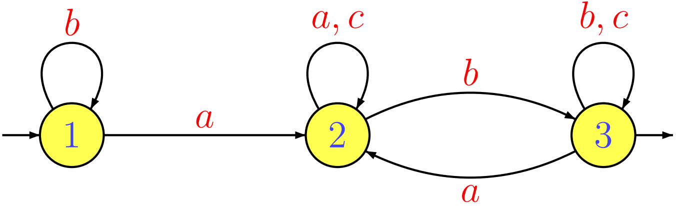
Syntactic monoid of $\{a, b\}^*aA^*b\{b, c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3

Rules:
$aa \rightarrow a$

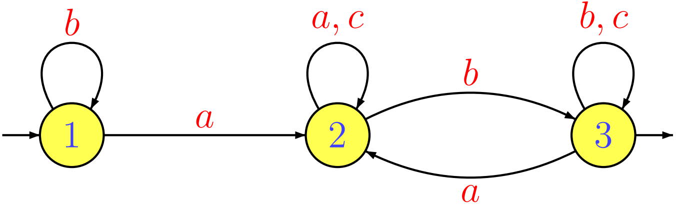
Syntactic monoid of $\{a, b\}^* a A^* b \{b, c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3

Rules:
$aa \to a$
$ac \to a$

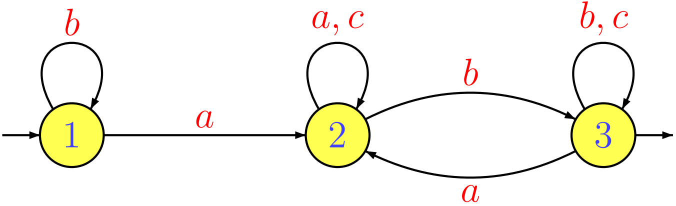
Syntactic monoid of $\{a,b\}^*aA^*b\{b,c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3

Rules:
$aa \to a$
$ac \to a$
$ba \to a$

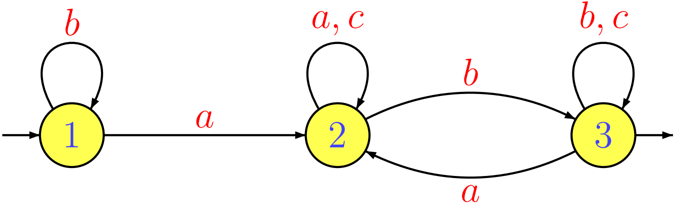
Syntactic monoid of $\{a, b\}^* a A^* b \{b, c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3

Rules:
$aa \to a$
$ac \to a$
$ba \to a$
$bb \to b$

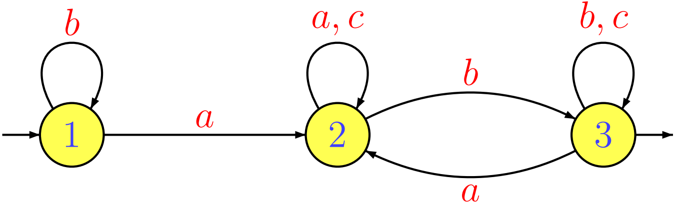
Syntactic monoid of $\{a,b\}^*aA^*b\{b,c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3
bc	-	3	3

Rules:
$aa \to a$
$ac \to a$
$ba \to a$
$bb \to b$

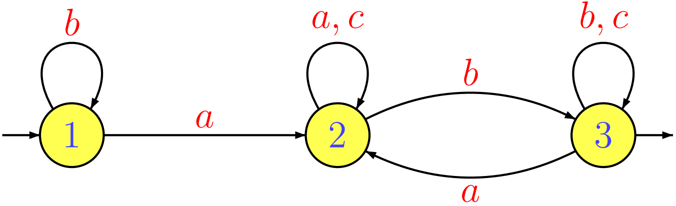
Syntactic monoid of $\{a,b\}^*aA^*b\{b,c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3
bc	-	3	3
ca	-	2	2

Rules:
$aa \longrightarrow a$
$ac \longrightarrow a$
$ba \longrightarrow a$
$bb \longrightarrow b$

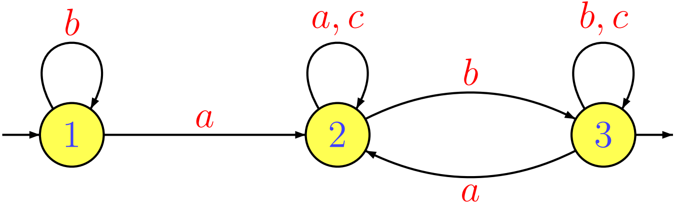
Syntactic monoid of $\{a,b\}^* a A^* b \{b,c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3
bc	-	3	3
ca	-	2	2

Rules:
$aa \to a$
$ac \to a$
$ba \to a$
$bb \to b$
$cb \to bc$

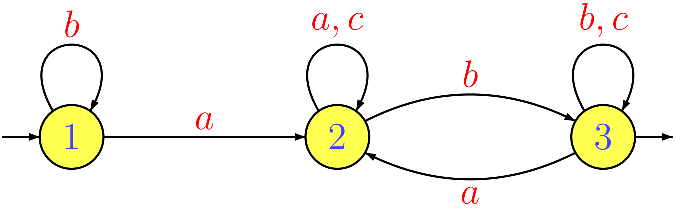
Syntactic monoid of $\{a,b\}^*aA^*b\{b,c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3
bc	-	3	3
ca	-	2	2

Rules:
$aa \longrightarrow a$
$ac \longrightarrow a$
$ba \longrightarrow a$
$bb \longrightarrow b$
$cb \longrightarrow bc$
$cc \longrightarrow c$

Syntactic monoid of $\{a,b\}^*aA^*b\{b,c\}^*$

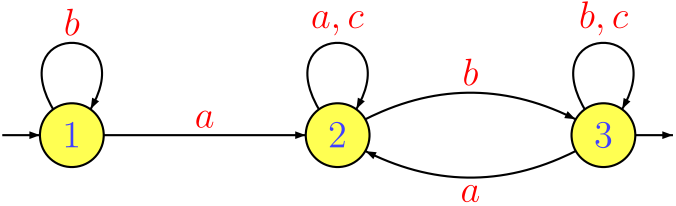


1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3
bc	-	3	3
ca	-	2	2

Rules:
$aa \to a$
$ac \to a$
$ba \to a$
$bb \to b$
$cb \to bc$
$cc \to c$

$abc \to ab$

Syntactic monoid of $\{a, b\}^* a A^* b \{b, c\}^*$

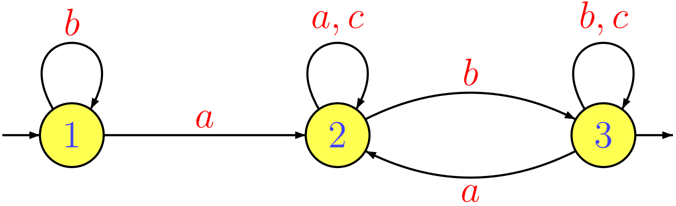


1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3
bc	-	3	3
ca	-	2	2

Rules:
$aa \to a$
$ac \to a$
$ba \to a$
$bb \to b$
$cb \to bc$
$cc \to c$

$abc \to ab$
$bca \to ca$

Syntactic monoid of $\{a,b\}^*aA^*b\{b,c\}^*$

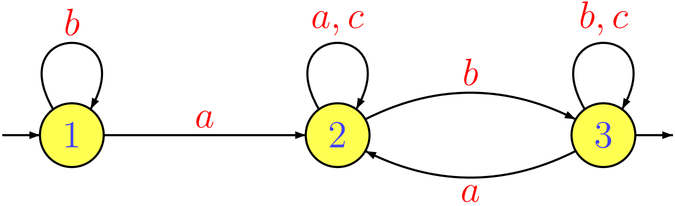


1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3
bc	-	3	3
ca	-	2	2

Rules:
$aa \to a$
$ac \to a$
$ba \to a$
$bb \to b$
$cb \to bc$
$cc \to c$

$abc \to ab$
$bca \to ca$
$cab \to bc$

Syntactic monoid of $\{a, b\}^* a A^* b \{b, c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	-	2	3
ab	3	3	3
bc	-	3	3
ca	-	2	2

Rules:
$aa \rightarrow a$
$ac \rightarrow a$
$ba \rightarrow a$
$bb \rightarrow b$
$cb \rightarrow bc$
$cc \rightarrow c$

$abc \rightarrow ab$
$bca \rightarrow ca$
$cab \rightarrow bc$
The end!

Restricting the number of variables

Theorem (Kamp 68, Immerman-Kozen 89)

Three variables suffice to express first order definable languages.

Proposition (Folklore)

First order with one variable captures the rational languages whose syntactic monoid is idempotent and commutative. These languages are Boolean combinations of the languages A^*aA^**, with* $a \in A$.

Two variables

A product $L_0 a_1 L_1 a_2 \cdots a_k L_k$ is unambiguous if each word in the product has a unique decomposition.

A language is unambiguous star-free iff it is a finite union of unambiguous products of the form $A_0^* a_1 A_1^* \cdots a_k A_k^*$ where $A_0, A_1, \ldots, A_k \subseteq A$ and $a_1, \ldots, a_k \in A$.

Theorem (Thérien-Wilke 1998)

First order with two variables captures the unambiguous star-free languages.

Unambiguous star-free languages

Theorem (Schützenberger 1975)

A language is unambiguous star-free iff its syntactic monoid belongs to $\mathbf{DA}$*.*

The variety of finite monoids $\mathbf{DA}$

Let M be a finite monoid and let $x, y \in M$. We say that x divides y if $y = uxv$ for some $u, v \in M$.

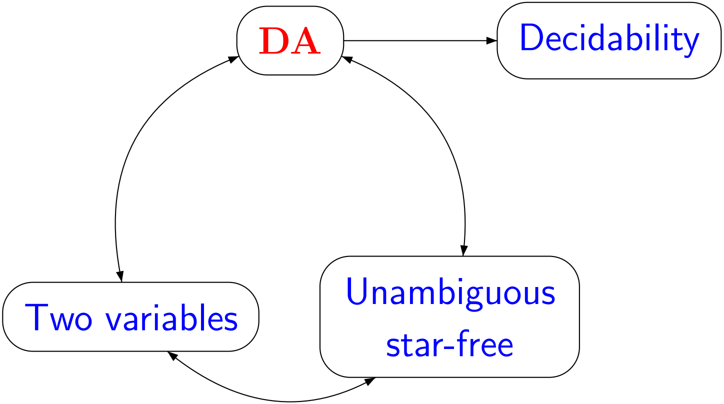
A finite monoid belongs to $\mathbf{DA}$ if it satisfies one of the following equivalent conditions:

- if two elements divide each other, and one is idempotent, then so is the other,
- its regular $\mathcal{J}$-classes are idempotent semigroups

*	*	*	*
*	*	*	*
*	*	*	*

- it satisfies an identity of the form $(xyz)^n y (xyz)^n = (xyz)^n$ for some $n > 0$.

Summary for two variables



The Σ_n-hierarchy of first order formulas

Formulas are written in prenex normal form.

Σ_0: Quantifier-free formulas

Σ_1: Formulas $\exists x_1 \cdots \exists x_n \; \varphi(x_1, \ldots, x_n)$ where φ is quantifier-free.

Π_1: Formulas $\forall x_1 \cdots \forall x_n \; \varphi(x_1, \ldots, x_n)$

Σ_2: Formulas $\exists x_1 \cdots \exists x_p \forall y_1 \cdots \forall y_q \; \varphi(x, y)$

Π_2: Formulas $\forall x_1 \cdots \forall x_n \exists y_1 \cdots \exists y_p \; \varphi(x, y)$

The hierarchies Σ_n, Π_n and Δ_n

Σ_n: Formulas $\exists^* \ \forall^* \ \exists^* \ \cdots \varphi$ with n alternating blocks of quantifiers.

Π_n: Formulas $\forall^* \ \exists^* \ \forall^* \cdots \varphi$ with n alternating blocks of quantifiers.

Δ_n: Formulas which are equivalent to a Σ_n-formula and to a Π_n-formula.

$\mathcal{B}\Sigma_n$: Boolean combinations of Σ_n-formulas.

Problem. What are the classes of languages captured by these families of formulas?

Polynomial closure

Let $\mathcal{L}$ be a class of languages.

The polynomial closure of $\mathcal{L}$ is the class of all unions of products of the form $L_0a_1L_1a_2 \cdots a_kL_k$ where $L_0, \ldots, L_k$ are in $\mathcal{L}$ and $a_1, \ldots, a_k$ are letters.

The unambiguous polynomial closure of $\mathcal{L}$ is the class of all unions of unambiguous products of the form $L_0a_1L_1a_2 \cdots a_kL_k$ where $L_0, \ldots, L_k$ are in $\mathcal{L}$ and $a_1, \ldots, a_k$ are letters.

Straubing-Thérien' hierarchy

Level $0 = \emptyset$ and A^*.

Level $1/2 =$ union of languages of the form $A^*a_1A^* \cdots A^*a_kA^*$.

Level $1 =$ Boolean combinations of languages of the form $A^*a_1A^* \cdots A^*a_kA^*$.

Level $n + 1/2 =$ union of products $L_0a_1L_1 \cdots a_kL_k$, where each L_i has level n and the a_i's are letters.

Level $n + 1 =$ Boolean combinations of languages of level $n + 1/2$.

This hierarchy is infinite (Brzozowski-Knast 1978).

Correspondence between the two hierarchies

Theorem (Thomas 1982, Perrin-Pin 1986)

(1) *The class* $\mathcal{B}\Sigma_n$ *captures the level* n.
(2) *The class* Σ_n *captures the level* $n - 1/2$.

Theorem (Pin-Weil 1997)

The class Δ_n *captures the unambiguous polynomial closure of level* n.

What about decidability ?

Level one

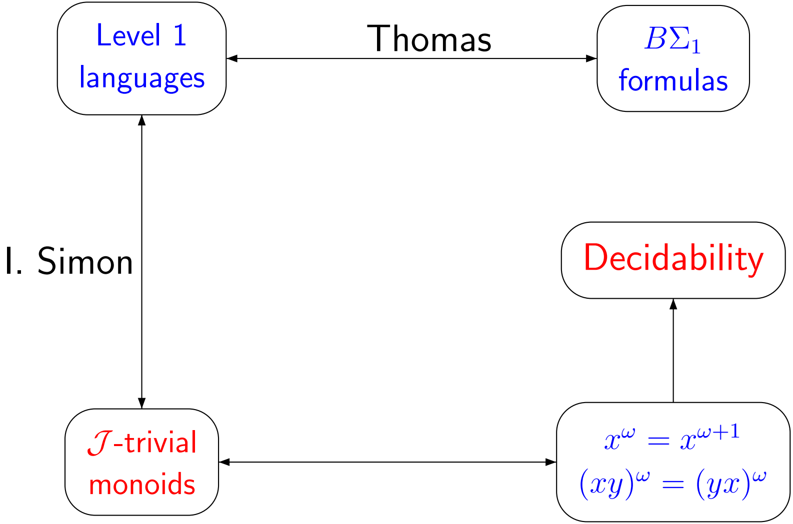
Theorem (Simon, 1972)

A rational language has level one iff its syntactic monoid is $\mathcal{J}$-trivial.

In a finite monoid, division is a quasi-order. If it is a partial order, the monoid is called $\mathcal{J}$-trivial because the Green relation $\mathcal{J}$ is the identity relation.

$\mathcal{J}$-trivial monoids are characterized by the identities $(xy)^n = (yx)^n$ and $x^{n+1} = x^n$, for some $n > 0$.

Another virtuous circle



Level 2

Theorem (Pin-Straubing 1981)

A language is of level 2 iff, for some $n > 0$, it is recognized by the monoid of $n \times n$ upper triangular Boolean matrices.

It is still an open problem to know whether this condition is decidable.

The class Δ_2

Theorem (Pin-Weil 97 + Schützenberger 75)

A language is Δ_2-definable iff its syntactic monoid belongs to the variety $\mathbf{DA}$.

It follows that Δ_2 is decidable.

Theorem (Thérien-Wilke 1998)

First order with two variables captures the unambiguous star-free languages.

Summary of known decidability results

(1) Σ_0 and Π_0 are decidable (easy)

(2) $\mathcal{B}\Sigma_1$ is decidable (Simon 1972)

(3) Σ_2, Π_2 and Δ_2 are decidable (Pin-Weil 1995)

(4) $\mathcal{B}\Sigma_2$ is decidable for a two-letter alphabet (Straubing 1988)

(5) Σ_3 is decidable for a two-letter alphabet (Glasser-Schmidt 2001)

Decidability of $\mathcal{B}\Sigma_2$ and beyond is open for more than two letters...

Part III

The variety approach

Ordered minimal automaton

Let A be a finite alphabet. Denote by A^* the free monoid on A. A subset of A^* is a language.

Let $\mathcal{A} = (Q, A, \cdot, i, F)$ be a minimal deterministic automaton. Define a relation $\leqslant$ on Q by $p \leqslant q$ iff for each $u \in A^*$, $(q{\cdot}u \in F \Rightarrow p{\cdot}u \in F)$.

The order is $2 \leqslant 4$, $1 \leqslant 3$ and $1, 2, 3, 4 \leqslant 0$.

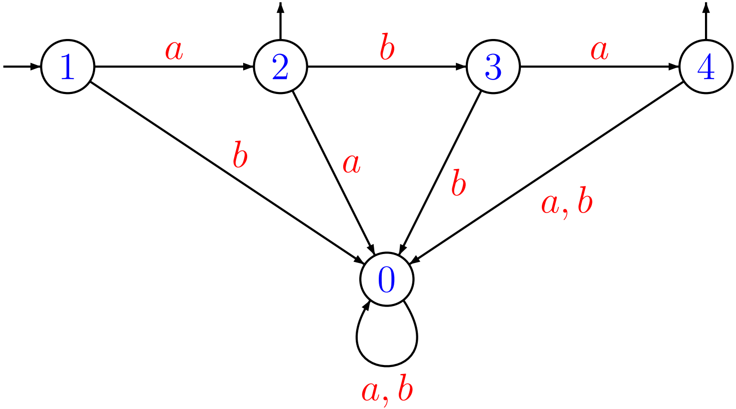


Figure: Minimal automaton of $\{a, aba\}$.

Varieties of finite monoids

A monoid is a set M equipped with an associative binary operation and an identity for this product: for all $x, y, z \in M$, $(xy)z = x(yz)$ and $1x = x = x1$.

A variety of finite monoids is a class of finite monoids closed under taking submonoids, quotients and finite products.

An ordered monoid is a monoid equipped with a stable order $\leqslant$: $x \leqslant y \Rightarrow zx \leqslant zy$ and $xz \leqslant yz$

A variety of finite ordered monoids is a class of finite ordered monoids closed under taking ordered submonoids, quotients and finite products.

Syntactic monoid of a language L of A^*

Two possible definitions

(1) The transition monoid of the minimal automaton of L

(2) The monoid $A^*/{\sim_L}$ where $\sim_L$ is the syntactic congruence of L:

By definition, $u \sim_L v$ if and only if, for every $x, y \in A^*$,

$$xvy \in L \Leftrightarrow xuy \in L$$

Syntactic ordered monoid

Two possible definitions

(1) The transition monoid of the minimal ordered automaton of L, ordered by $u \leqslant v$ iff for each $q \in Q$, $q{\cdot}u \leqslant q{\cdot}v$

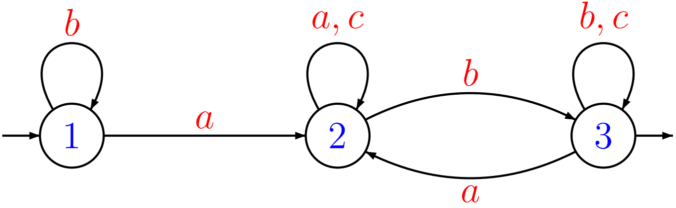
(2) Abstract definition: syntactic preorder of L: $u \leqslant_L v$ iff, for every $x, y \in A^*$,

$$xvy \in L \Rightarrow xuy \in L$$

Syntactic ordered monoid of L : $(A^*/{\sim_L}, \leqslant_L /sim_L)$

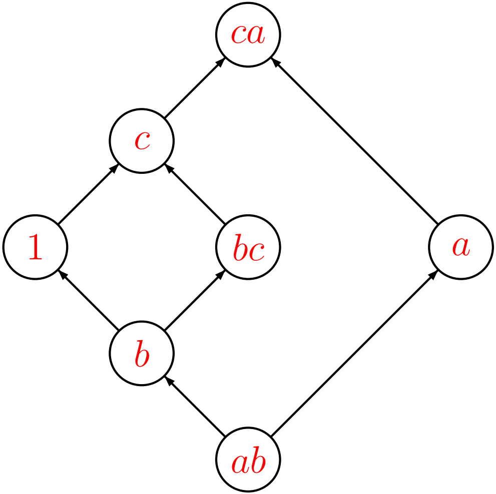
Syntactic monoid of $\{a, b\}^* a A^* b \{b, c\}^*$

The order is $3 < 2 < 1$.



1	a	b	c	ab	bc	ca
1	2	1	–	3	–	–
2	2	3	2	3	3	2
3	2	3	3	3	2	2

Syntactic order of $\{a, b\}^* a A^* b \{b, c\}^*$



1	1	2	3
a	2	2	2
b	1	3	3
c	–	2	3
ab	3	3	3
bc	–	3	2
ca	–	2	2

The syntactic ordered monoid of $(ab)^*$

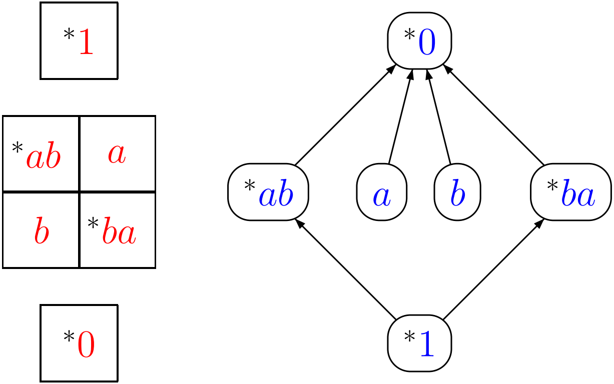


Figure: $\mathcal{J}$-classes and syntactic order.

Positive varieties of languages

A positive variety of languages is a class $\mathcal{V}$ of recognizable languages such that for every alphabet A, $\mathcal{V}(A^*)$ is a lattice closed under quotients and inverse of morphisms.

A variety of languages is also closed under complement.

Eilenberg's theorem

Theorem (Eilenberg 1976)

Let $\mathbf{V}$ *be a variety of finite monoids. For each alphabet* A, *denote by* $\mathcal{V}(A^*)$ *the set of all languages whose syntactic monoid is in* $\mathbf{V}$.
(1) $\mathcal{V}$ *is a variety of languages.*
(2) The correspondence $\mathbf{V} \to \mathcal{V}$ *is a bijection between varieties of finite monoids and varieties of languages.*

Ordered version

Theorem (Pin 1995)

Let $\mathbf{V}$ *be a variety of finite ordered monoids. For each alphabet* A*, denote by* $\mathcal{V}(A^*)$ *the set of all languages whose syntactic ordered monoid is in* $\mathbf{V}$*.*
(1) $\mathcal{V}$ *is a positive variety of languages.*
(2) The correspondence $\mathbf{V} \to \mathcal{V}$ *is a bijection between varieties of finite ordered monoids and positive varieties of languages.*

Some examples

- The star-free languages form a variety of languages.

- The level n of the concatenation hierarchy is a variety of languages.

- The level $n + 1/2$ of the concatenation hierarchy is a positive variety of languages.

Decidability results

Theorem (Pin-Weil 1995))

A language has level $1/2$ iff its ordered syntactic monoid M satisfies the following condition: for every $x \in M$, $x \leqslant 1$.

Theorem (Simon 1972)

A language has level 1 iff its syntactic monoid is finite and $\mathcal{J}$-trivial.

Decidability problems and results

Theorem (Pin-Weil 2001)

A language is of level $3/2$ *if and only if its ordered syntactic monoid belongs to the variety of ordered monoids* $[ese \leqslant e] \textcircled{M} [x^2 = x, xy = yx]$*. This condition is decidable.*

The notation $\textcircled{M}$ stands for the Mal'cev product.

Back to the star-height problem

Theorem (Pin 1978)

If the languages of star-height $\leqslant 1$ *form a variety of languages, then there is no language of star-height* 2.

Theorem (Pin, Straubing, Thérien 1989)

The languages of star-height $\leqslant 1$ *are closed under Boolean operations, quotients and inverse of length-preserving morphisms.*

$\mathcal{C}$-varieties (Straubing 2001).

Let $\mathcal{C}$ be a class of morphisms between free monoids, closed under composition and containing all length-preserving morphisms.

Examples

(1) length-preserving morphisms
(2) length-multiplying morphisms
(3) non-erasing morphisms
(4) length-decreasing morphisms
(5) all morphisms

$\mathcal{C}$-varieties of languages (Straubing 2001).

A positive $\mathcal{C}$-variety of languages is a class $\mathcal{V}$ of recognizable languages such that for every alphabet A, $\mathcal{V}(A^*)$ is a positive Boolean algebra closed under quotients and inverse of $\mathcal{C}$-morphisms.

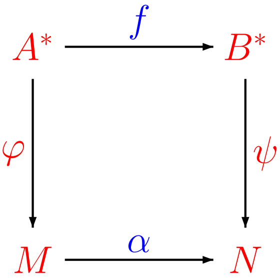
A $\mathcal{C}$-variety of languages is also closed under complement.

Is there an algebraic counterpart ?

Stamps

A stamp is a surjective morphism from A^* onto a finite monoid.

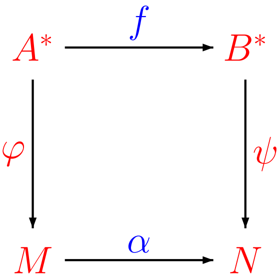
A $\mathcal{C}$-morphism of stamps is a pair (f, α) such that $f \in \mathcal{C}$ and the diagram below commutes.



Projections and inclusions

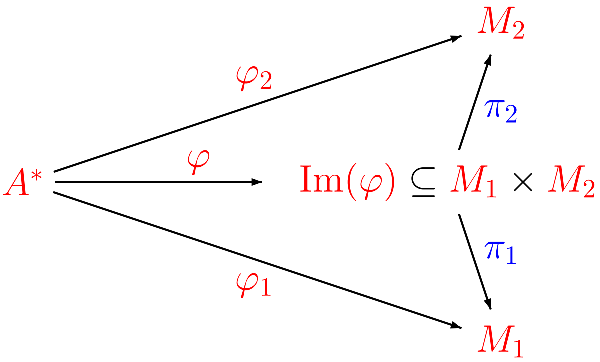
A $\mathcal{C}$-morphism (f, α) is a $\mathcal{C}$-projection if the map $f : A^* \to B^*$ satisfies $f(A) = B$.

A $\mathcal{C}$-morphism (f, α) is a $\mathcal{C}$-inclusion if the morphism $\alpha : M \to N$ is injective.



Restricted products

The restricted direct product of two stamps φ_1 and φ_2 is the stamp φ defined by $\varphi(a) = (\varphi_1(a), \varphi_2(a))$.



Varieties of stamps

A $\mathcal{C}$-variety of stamps is a class of stamps closed under $\mathcal{C}$-inclusion, $\mathcal{C}$-projection and finite restricted direct products.

When $\mathcal{C}$ is the class of all (resp. length-preserving, length-multiplying, non-erasing, length-decreasing) morphisms, we use the term all-variety (resp. lp-variety, lm-variety, ne-variety, de-variety).

Straubing's $\mathcal{C}$-variety theorem

Theorem

Let $\mathbf{V}$ *be a* $\mathcal{C}$*-variety of stamps. For each alphabet* A*, denote by* $\mathcal{V}(A^*)$ *the set of all languages whose syntactic morphism is in* $\mathbf{V}$*.*
(1) $\mathcal{V}$ *is a* $\mathcal{C}$*-variety of languages.*
(2) The correspondence $\mathbf{V} \to \mathcal{V}$ *is a bijection between* $\mathcal{C}$*-varieties of stamps and* $\mathcal{C}$*-varieties of languages.*

Reiterman theorem for $\mathcal{C}$-varieties (Kunc)

Every $\mathcal{C}$-variety can be defined by a set of identities in which symbols are interpreted as follows:

$\mathcal{C}$	Interpretation
all morphisms	arbitrary words
nonerasing	nonempty words
length-multiplying	words of a fixed length
length-preserving	letters
length-decreasing	letters or empty word

Quasi-aperiodic stamps

Let $\varphi : A^* \to M$ be a stamp. The set $\varphi(A)$, as an element of $\mathcal{P}(M)$, has a unique idempotent power. If this subsemigroup of M is aperiodic, the stamp is called quasi-aperiodic. Such stamps form the lm-variety $\mathbf{QA}$.

Theorem (Kunc 2003)

$\mathbf{QA}$ *is defined by the lm-identity* $(a^{\omega-1}b)^\omega = (a^{\omega-1}b)^{\omega+1}$.

Quasi-aperiodic stamps (2)

Theorem (Straubing 1994)

A rational language belongs to AC^0 *iff its syntactic stamp belongs to* $\mathbf{QA}$*.*

Back to the star-height problem

Corollary (Pin, Straubing, Thérien 1989)

The languages of star-height $\leqslant 1$ from a lp-variety of languages.

What is the corresponding lp-variety of stamps?

Is there any nontrivial identity satisfied by all languages of star-height $\leqslant 1$?