

Profinite approach to language theory

Jean-Éric Pin

LIAFA, CNRS and University Paris Diderot

May 2011, EPIT Spring School



Summary

- (1) Motivations
- (2) Lattices of languages
- (3) Birkhoff's theorem
- (4) The profinite world
- (5) Profinite Equations



Part I

Motivations



High-level vs. algebraic properties [M. Zeitoun]

Language L $Synt(L)$ Logical
definability

Star-free	A	FO [<], LTL
-----------	----------	---------------------------

[Schützenberger 65, McNaughton-P. 71, Kamp 68]

Piecewise testable	J	Bool (Σ_1)
--------------------	----------	----------------------------

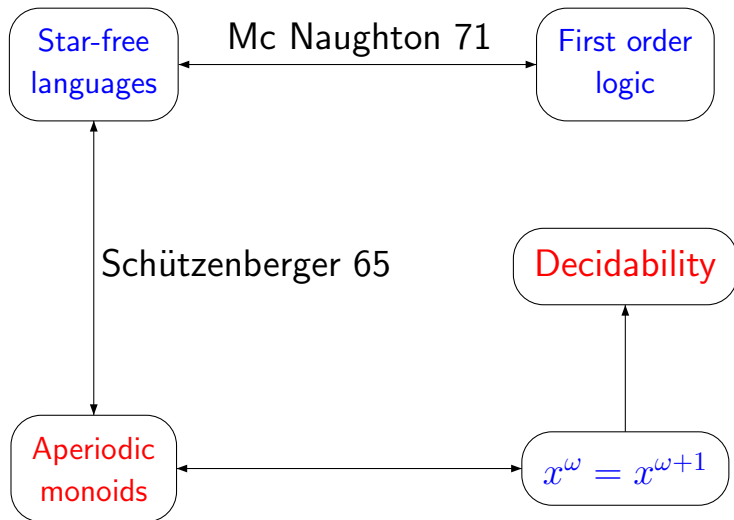
[Simon 75, Thomas 87]

Unambiguous star-free	DA	FO ² [<], UTL , $\Sigma_2 \cap \Pi_2$
-----------------------	-----------	---

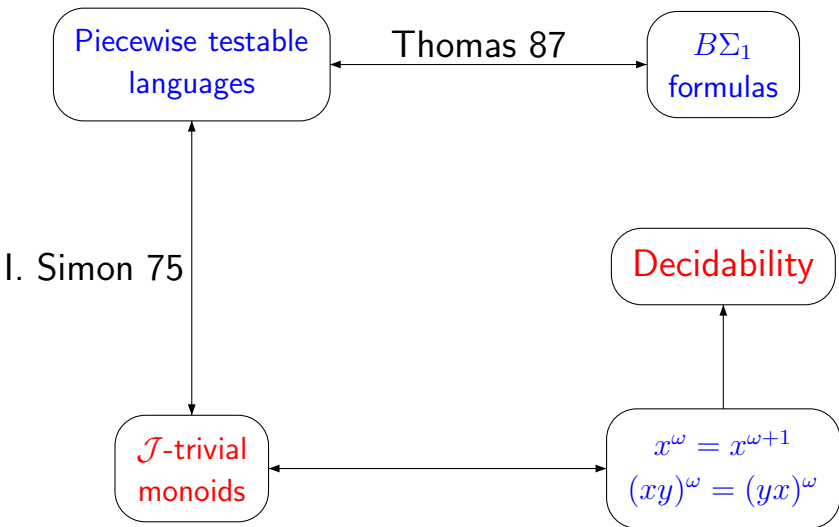
[Sch76, SchThV01, ThW98, EVW97, PW97]



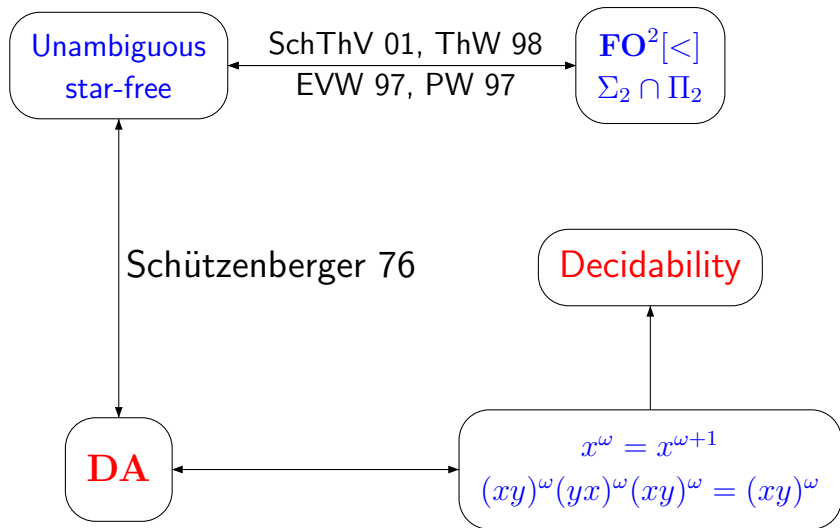
Virtuous circle 1



Virtuous circle 2



Virtuous circle 3



Part II

Lattices of languages

Lattices of languages

Let A be a finite alphabet. A **lattice of languages** is a set of **regular** languages of A^* containing \emptyset and A^* and closed under **finite intersection** and **finite union**.

Let u and v be words of A^* . A language L of A^* **satisfies the equation** $u \rightarrow v$ if

$$u \in L \Rightarrow v \in L$$

Proposition

Let E be a set of equations of the form $u \rightarrow v$.
Then the languages of A^* satisfying the equations
of E form a *lattice of languages*.

Proposition

Let L, L_1, \dots, L_n be languages. If L satisfies all the
equations satisfied by L_1, \dots, L_n , then L belongs to
the *lattice of languages* generated by L_1, \dots, L_n .

Proof

Let L, L_1, \dots, L_n be languages. If L satisfies all the equations satisfied by L_1, \dots, L_n , then L belongs to the *lattice of languages* generated by L_1, \dots, L_n .

Indeed

$$L = \bigcup_{I \in \mathcal{F}} \bigcap_{i \in I} L_i$$

where \mathcal{F} is the set of all subsets I of $\{1, \dots, n\}$ for which there exists a word $v \in L$ such that $v \in L_i$ if and only if $i \in I$.

Characterization of finite lattices of languages

Proposition

A finite set of languages of A^* is a *lattice of languages* iff it can be defined by a set of equations of the form $u \rightarrow v$ with $u, v \in A^*$.

The lattice generated by the languages A^*aA^* (with $a \in A$) is defined by the equations $ab \leftrightarrow ba$, $aa \leftrightarrow a$ and $xy \rightarrow xay$, for all $a, b \in A$ and all $x, y \in A^*$.

FO¹ is defined by the equations $ab \leftrightarrow ba$ and $aa \leftrightarrow a$, for all $a, b \in A$.

Lattices and equations

There is an equational theory for **finite lattices** of languages.

What about **infinite lattices**?



Lattices and equations

There is an equational theory for **finite lattices** of languages.

What about **infinite lattices**?

One needs the **profinite world**...



Birkhoff's variety theorem

A monoid is **commutative** iff it satisfies the **identity** $xy = yx$. A monoid is **idempotent** iff it satisfies the **identity** $x^2 = x$.

Birkhoff's variety theorem. A class of monoids is defined by a set of **identities** iff it is a **variety**, i.e. is closed under taking **submonoids**, **quotient** monoids and (possibly infinite!) direct **products**.

What about **finite monoids**?



Birkhoff's variety theorem

A monoid is **commutative** iff it satisfies the **identity** $xy = yx$. A monoid is **idempotent** iff it satisfies the **identity** $x^2 = x$.

Birkhoff's variety theorem. A class of monoids is defined by a set of **identities** iff it is a **variety**, i.e. is closed under taking **submonoids**, **quotient** monoids and (possibly infinite!) direct **products**.

What about **finite monoids**?

One needs the **profinite world**...



Part III

The profinite world

Citation (M. Stone)

*A cardinal principle of modern mathematical research may be stated as a maxim: **One must always topologize.***

Overview

The set of **profinite words** is

- The **completion** of the discrete space A^* , equipped with the **profinite metric**.
- The **projective limit** of the directed system of finite monoids.
- The set of **implicit operations** on A^*
- The **Stone dual** of the Boolean algebra of regular languages.



Separating words

A monoid M separates two words u and v of A^* if there exists a monoid morphism $\varphi : A^* \rightarrow M$ such that $\varphi(u) \neq \varphi(v)$.

- The morphism $u \rightarrow |u| \pmod 2$ (from A^* into $\mathbb{Z}/2\mathbb{Z}$) separates $abaabaaba$ and $abaabaabab$.
- Similarly, for each letter a , one can count the number of a modulo n .
- Let $M = \{1, x, y\}$ with $xx = yx = x$ and $xy = yy = y$. Let $\varphi : \{a, b\}^* \rightarrow M$ defined by $\varphi(a) = x$ and $\varphi(b) = y$. For each u , $\varphi(ua) = x$ and $\varphi(ub) = y$. Thus φ separates ua from ub .

Proposition

One can always *separate* two *distinct* words by a finite monoid.

Proof. Let $u \neq v$ and let $\eta : A^* \rightarrow M$ be the syntactic morphism of the language $\{u\}$.

Suppose that $\eta(u) = \eta(v)$. Then $v \in \eta^{-1}(\eta\{u\})$ and thus $v = u$ since $\eta^{-1}(\eta\{u\}) = \{u\}$.

The profinite metric

Let u and v be two words. Put

$$r(u, v) = \min\{|M| \mid M \text{ is a finite monoid} \\ \text{that separates } u \text{ and } v\}$$

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

Then d is an **ultrametric**, that is, for all $x, y, z \in A^*$,

- (1) $d(x, y) = 0 \iff x = y$,
- (2) $d(x, y) = d(y, x)$,
- (3) $d(x, z) \leq \max\{d(x, y), d(y, z)\}$



The free profinite monoid

The completion of the metric space (A^*, d) is the free **profinite monoid** on A and is denoted by $\widehat{A^*}$. It is a **compact** space, whose elements are called **profinite words**.

The concatenation product is **uniformly continuous** on A^* and can be extended by continuity to $\widehat{A^*}$.

Any morphism $\varphi : A^* \rightarrow M$, where M is a (discrete) finite monoid is uniformly continuous and extends in a unique way to a **uniformly continuous** morphism $\widehat{\varphi} : \widehat{A^*} \rightarrow M$.

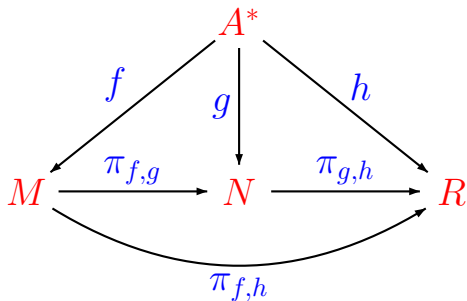
Main properties of d

Intuitively, two words are close for d if one needs a **large** monoid to separate them.

A sequence of profinite words u_n **converges** to a profinite word u iff, for every morphism φ from A^* to a finite monoid, the sequence $\widehat{\varphi}(u_n)$ is ultimately equal to $\widehat{\varphi}(u)$.

The free profinite monoid as a projective limit

For each pair of surjective morphisms $f : A^* \rightarrow M$ and $g : A^* \rightarrow N$ such that N is a quotient of M , choose a surjective morphism $\pi_{f,g} : M \rightarrow N$ such that $g = \pi_{f,g} \circ f$. Make this choice **compatible**, in the sense that $\pi_{f,h} = \pi_{g,h} \circ \pi_{f,g}$.



The free profinite monoid as a projective limit

The monoid $\widehat{A^*}$ can be defined as the **projective limit** of the **directed system** formed by the **surjective morphisms** between finite A -generated monoids.

Consider the **compact monoid**

$$P = \prod_{f:A^* \rightarrow M} M$$

An element $(s_f)_{f:A^* \rightarrow M}$ of P is **compatible** if $\pi_{f,g}(s_f) = s_g$. The set of compatible elements is a closed submonoid of P , equal to $\widehat{A^*}$.

Profinite words

A profinite word u is completely determined by the elements $\widehat{\varphi}(u)$, where $\varphi : A^* \rightarrow M$ is a morphism into a finite monoid.

Alternatively, one can define a profinite word as a equivalence class of Cauchy sequences of words, for the following equivalence: two Cauchy sequences $x = (x_n)_{n \geq 0}$ and $y = (y_n)_{n \geq 0}$ are equivalent if the interleave sequence $x_0, y_0, x_1, y_1, \dots$ is also a Cauchy sequence.

The free profinite monoid on one generator

The free monoid a^* is isomorphic to \mathbb{N} . The free profinite monoid $\widehat{a^*}$ embeds into the product

$$(\mathbb{N} \cup \{\infty\}) \times \prod_{p \text{ prime}} \mathbb{Z}_p$$

where \mathbb{Z}_p is the group of p -adic numbers.

The product $\widehat{\mathbb{Z}} = \prod_{p \text{ prime}} \mathbb{Z}_p$ is the group of factoriadic integers. It is a subgroup of $\widehat{a^*}$ and thus $\widehat{a^*}$ is uncountable.

Factoriadic integers (after Lenstra)

Each natural number has a **unique representation** as

$$n = c_k k! + \cdots + c_2 2! + c_1 1!$$

where $c_k \neq 0$ and $0 \leq c_i \leq i$ for $1 \leq i \leq k$. We write $n = (c_k \cdots c_2 c_1)!$. For instance $5 = (21)!$ and $25 = (1001)!$.

A **factoriadic integer** is an infinite sequence $(\dots, c_4, c_3, c_2, c_1)$, where $0 \leq c_i \leq i$ for all $i > 0$. The set of **factoriadic integers** is denoted by $\widehat{\mathbb{Z}}$.



Addition of factoriadic integers

One can identify the natural number $(c_k \cdots c_2 c_1)!$ with $(\dots, 0, 0, c_k, \dots, c_1)$. Thus \mathbb{N} embeds into $\widehat{\mathbb{Z}}$.

Addition is performed **termwise**, from **right to left**. When the sum of the terms of rank i exceeds i , one subtracts $i + 1$ from it and adds a carry of 1 to the sum of the terms of rank $i + 1$. Observe that

$$(\dots, 5, 4, 3, 2, 1) + (\dots, 0, 0, 1) = (\dots, 0, 0, 0)$$

Therefore $(\dots, 5, 4, 3, 2, 1) = -1$, $\widehat{\mathbb{Z}}$ is a **commutative group** and \mathbb{Z} embeds into $\widehat{\mathbb{Z}}$.



The free profinite monoid \widehat{a}^*

The free profinite monoid \widehat{a}^* is the disjoint union of \mathbb{N} and $\widehat{\mathbb{Z}}$.

To avoid any confusion, we set $\omega = (\dots, 0, 0, 0)$.
Then an element of \widehat{a}^* is either a natural number n or an element of the form $\omega + c$, where c is a factoriadic integer.

Addition is performed in the natural way:

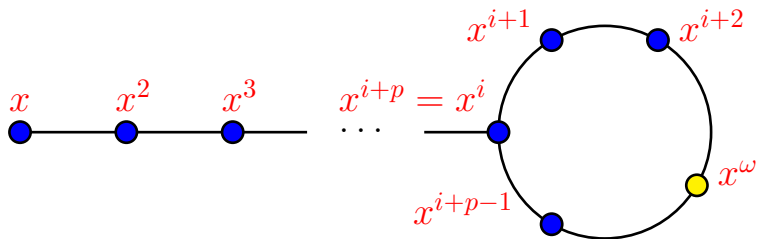
$$n + m,$$

$$(\omega + c) + n = \omega + (c + n),$$

$$(\omega + c) + (\omega + d) = \omega + (c + d)$$

Idempotents in finite semigroups

Let x be an element of a finite semigroup S .



For $k \geq i$, $x^k = x^{k+p}$. Thus, if ω is the smallest multiple of p such that $\omega \geq i$, say $\omega = qp$, one has

$$(x^\omega)^2 = x^{2\omega} = x^{\omega+qp} = x^\omega$$

and thus x^ω is **idempotent**. Note that $\omega \leq |S|$.

A converging sequence in \widehat{A}^*

Proposition

For each $u \in \widehat{A}^*$, the sequence $u^{n!}$ is a *converging sequence*.

Let $\varphi : A^* \rightarrow M$ be a morphism and let $x = \widehat{\varphi}(u)$. Then $\widehat{\varphi}(u^{n!}) = x^{n!}$.

Further, x has an *idempotent power* x^ω , with $\omega \leq |M|$. Thus for $n \geq |M|$, $n!$ is a multiple of ω and $x^{n!} = x^\omega$.

Thus $x^{n!}$ is ultimately equal to x^ω and $u^{n!}$ is a *converging sequence*.



The profinite operator ω

For each $u \in A^*$, the sequence $u^{n!}$ converges in $\widehat{A^*}$ to a limit, denoted by u^ω .

$$u^\omega = \lim_{n \rightarrow \infty} u^{n!}$$

This element is **idempotent** in $\widehat{A^*}$.

In the same way, we set

$$u^{\omega-1} = \lim_{n \rightarrow \infty} u^{n!-1} \quad u^{\omega+1} = \lim_{n \rightarrow \infty} u^{n!+1}$$

Then $u^{\omega-1}u^\omega = u^\omega u^{\omega-1} = u^{\omega-1}$, $u^{\omega-1}u^{\omega+1} = u^\omega$
and $u^{\omega+1}u^\omega = u^\omega u^{\omega+1} = u^{\omega+1}$.

Examples of profinite words

Every **term** in the signature $(x, y) \rightarrow xy, x \rightarrow x^\omega$
and $x \rightarrow x^{\omega-1}$ is a **profinite word**.

Example: $\left(\left((ab)^\omega (ba)^\omega \right)^{\omega+1} (ab)^\omega \right)^{\omega-1}$

Do we exhaust all profinite words this way?

Examples of profinite words

Every **term** in the signature $(x, y) \rightarrow xy, x \rightarrow x^\omega$ and $x \rightarrow x^{\omega-1}$ is a **profinite word**.

Example: $\left(((ab)^\omega (ba)^\omega)^{\omega+1} (ab)^\omega \right)^{\omega-1}$

Do we exhaust all profinite words this way?

No, since \widehat{A}^* is **uncountable**. Worse, if $|A| > 1$, no such term belongs to the minimal ideal K of \widehat{A}^* .

Indeed, a **profinite word** w belongs to K iff it is $\leq_{\mathcal{J}}$ -below every finite word x : that is, $w = pxs$ for some profinite words p and s .

Another profinite word

Let us fix a total order on the alphabet A . Let u_0, u_1, \dots be the ordered sequence of all words of A^* in the induced shortlex order.

$1, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, \dots$

Reilly and Zhang (see also Almeida-Volkov) proved that the sequence $(v_n)_{n \geq 0}$ defined by

$$v_0 = u_0, \quad v_{n+1} = (v_n u_{n+1} v_n)^{(n+1)!}$$

converges to an idempotent ρ_A of the minimal ideal of $\widehat{A^*}$.

Part IV

Equational theory



Reiterman's theorem

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

Let u and v be profinite words of \widehat{A}^* . A finite monoid satisfies the **identity** $u = v$ if for each morphism $\varphi = A^* \rightarrow M$, $\widehat{\varphi}(u) = \widehat{\varphi}(v)$.

Theorem (Reiterman)

*A class of finite monoids is defined by a set of **profinite identities** iff it is a **variety of finite monoids**.*



Examples of identities

A finite monoid is **aperiodic** iff it satisfies the identity $x^{\omega+1} = x^\omega$.

A finite monoid is **\mathcal{J} -trivial** iff it satisfies the identities

$$x^{\omega+1} = x^\omega \text{ and } (xy)^\omega x = (xy)^\omega = y(xy)^\omega$$

iff it satisfies the identities

$$x^{\omega+1} = x^\omega \text{ and } (xy)^\omega = (yx)^\omega .$$

A finite monoid is in **DA** iff it satisfies the identities $x^{\omega+1} = x^\omega$ and $(xy)^\omega (yx)^\omega (xy)^\omega = (xy)^\omega$.

A finite monoid is a group iff it satisfies the identity $x^\omega = 1$.



Reiterman's theorem for ordered monoids

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

Let u and v be profinite words of \widehat{A}^* . A finite monoid satisfies the **identity** $u \leq v$ if for each morphism $\varphi = A^* \rightarrow M$, $\widehat{\varphi}(u) \leq \widehat{\varphi}(v)$.

Theorem

*A class of finite ordered monoids is defined by a set of **profinite identities** of the form $u \leq v$ iff it is a **variety of ordered finite monoids**.*

Equations for regular languages

Let L be a regular language of A^* and let $\eta : A^* \rightarrow M$ be its syntactic morphism.

Let \overline{L} be the topological closure of L for the profinite metric.

Définition

Let u and v be profinite words of $\widehat{A^*}$. Then L satisfies the equation $u \rightarrow v$ if the condition $u \in \overline{L}$ implies $v \in \overline{L}$.

An equivalent definition

Let L be a regular language of A^* and let $\eta : A^* \rightarrow M$ be its syntactic morphism.

Définition

Let u and v be profinite words of $\widehat{A^*}$. Then L satisfies the profinite equation $u \rightarrow v$ iff the condition $\widehat{\eta}(u) \in \eta(L)$ implies $\widehat{\eta}(v) \in \eta(L)$.

Equational theory of lattices

Given a set E of equations of the form $u \rightarrow v$ (where u and v are profinite words), the set of all regular languages of A^* satisfying all the equations of E is called the set of languages **defined by E** .

Theorem (Gehrke, Grigorieff, Pin 2008)

A set of regular languages of A^ is a **lattice of languages** iff it can be defined by a **set of equations** of the form $u \rightarrow v$, where $u, v \in \widehat{A^*}$.*

An example: regular prefix codes

A set P of nonempty words is a **prefix code** if no word of P is a strict prefix of another word of P .

Theorem (Daviaud and Paperman)

*The lattice generated by the **regular prefix codes** is defined by the equations $ux^\omega \rightarrow y$ for $x, u, y \in \widehat{A}^*$, $x \neq 1$.*



Proof of the main theorem

Fancy proof: by Stone duality.

Proof by compactness.

Step 1. If L is regular, then \overline{L} is clopen.

Step 2. The set

$$E_L = \{(u, v) \in \widehat{A}^* \times \widehat{A}^* \mid L \text{ satisfies } u \rightarrow v\}$$

is equal to $(\overline{L}^c \times \widehat{A}^*) \cup (\widehat{A}^* \times \overline{L})$. Therefore it is clopen in $\widehat{A}^* \times \widehat{A}^*$.

Let \mathcal{L} be a lattice.



Proof of the main theorem (2)

Step 3. Let \mathcal{K} be a subset of \mathcal{L} . The property :

\mathcal{K} satisfies $u \rightarrow v \Rightarrow L$ satisfies $u \rightarrow v$
holds iff (u, v) belongs to $E_L \cup \bigcup_{K \in \mathcal{K}} E_K^c$.

Let L be a language satisfying the equations of \mathcal{L} .

Step 4. If $\mathcal{K} = \mathcal{L}$, the relation \Rightarrow holds for every pair (u, v) . Therefore, $E_L \cup \{E_K^c \mid K \in \mathcal{L}\}$ is an open cover of $\widehat{A}^* \times \widehat{A}^*$. By **compactness**, one can extract a finite subcover $E_L \cup \{E_{L_1}^c, \dots, E_{L_n}^c\}$. Then L satisfies the equations of the **lattice generated by** L_1, \dots, L_n . Thus L belongs to \mathcal{L} .

Second day

Second day. . .

Ordered minimal automaton

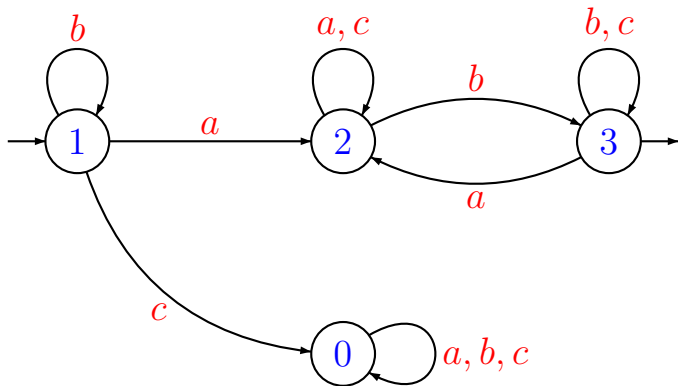
Let A be a finite alphabet. Denote by A^* the free monoid on A . A subset of A^* is a **language**.

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

Minimal automaton of $\{a, b\}^* a A^* b \{b, c\}^*$

$p \leq q$ iff for each $u \in A^*$, $(q \cdot u \in F \Rightarrow p \cdot u \in F)$.

Here, the order is $3 < 2 < 1 < 0$.



Syntactic ordered monoid

Two possible definitions

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

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

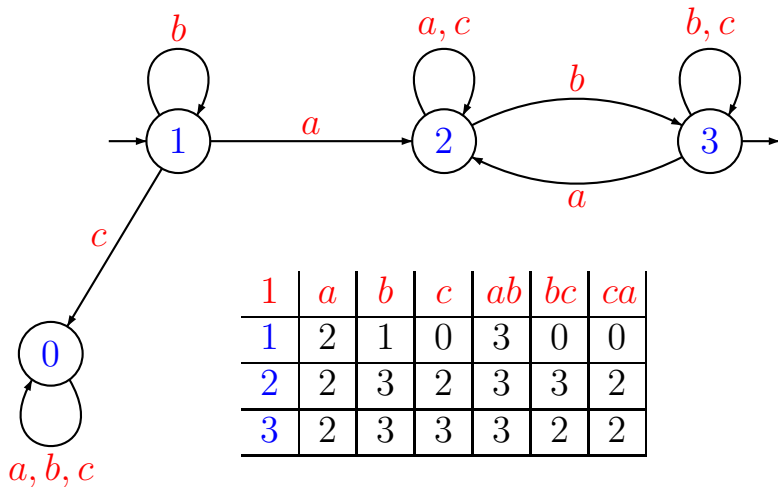
$$xvy \in L \Rightarrow xuy \in L$$

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



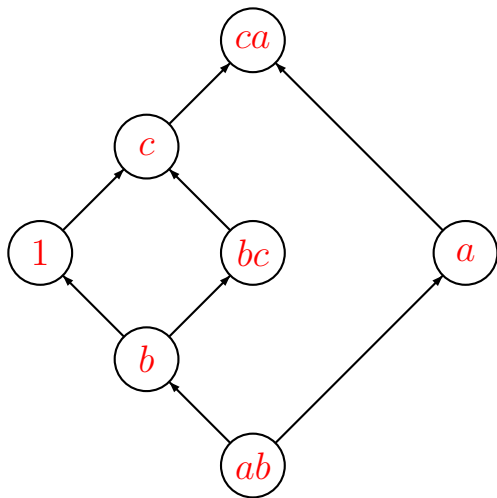
Syntactic monoid of $\{a, b\}^* a A^* b \{b, c\}^*$

The order is $3 < 2 < 1 < 0$.



	1	a	b	c	ab	bc	ca
1	1	2	1	0	3	0	0
2	2	2	3	2	3	3	2
3	3	2	3	3	3	2	2

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



<i>1</i>	1	2	3
<i>a</i>	2	2	2
<i>b</i>	1	3	3
<i>c</i>	0	2	3
<i>ab</i>	3	3	3
<i>bc</i>	0	3	2
<i>ca</i>	0	2	2

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

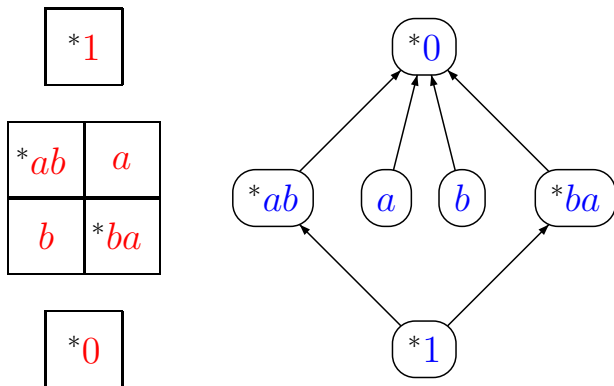


Figure: \mathcal{J} -classes and syntactic order.

Equations of the form $u \leq v$

We say that L satisfies the equation $u \leq v$ if, for all $x, y \in \widehat{A}^*$, it satisfies the equation $xvy \rightarrow xuy$.

Proposition

Let L be a regular language of A^* , let (M, \leq_L) be its ordered syntactic monoid and let $\eta : A^* \rightarrow M$ be its syntactic morphism. Then L satisfies the equation $u \leq v$ iff $\widehat{\eta}(u) \leq_L \widehat{\eta}(v)$.

Lattices of languages closed under quotient

Let $x, y \in A^*$. The quotient of L by x and y is the language $x^{-1}Ly^{-1} = \{u \in A^* \mid xuy \in L\}$.

A lattice of regular languages \mathcal{L} is **closed under quotient** if for all $x, y \in A^*$, $L \in \mathcal{L}$ implies $x^{-1}Ly^{-1} \in \mathcal{L}$.

Theorem

A set of regular languages of A^ is a **lattice closed under quotients** iff it can be defined by a **set of profinite equations** of the form $u \leq v$, where $u, v \in \widehat{A^*}$.*

Boolean algebras closed under quotient

If L satisfies $u \rightarrow v$, then L^c satisfies $v \rightarrow u$.

We write $u = v$ for $u \leq v$ and $v \leq u$.

Theorem

A set of regular languages of A^ is a Boolean algebra closed under quotient iff it can be defined by a set of profinite equations of the form $u = v$.*

Languages with zero

A **language with zero** is a language whose **syntactic monoid** has a zero. The class of regular **languages with zero** is closed under **Boolean operations** and **quotients**. Therefore, it can be defined by a set of equations of the form $u = v$.

Proposition

*A regular language **has a zero** iff it satisfies the equation $x\rho_A = \rho_A = \rho_Ax$ for all $x \in A^*$.*



Nondense languages

A language L of A^* is **dense** if, for each word $u \in A^*$, $L \cap A^*uA^* \neq \emptyset$.

The **nondense or full** languages form a lattice closed under **quotients**.

Theorem

A language of A^* is **nondense or full** iff it satisfies the equations $x\rho_A = \rho_A = \rho_Ax$ and $x \leq \rho_A$ for all $x \in A^*$.

Slender languages

The **density** of a language is the function

$$d_L(n) = |L \cap A^n|$$

A language is **slender** if its density is **bounded by a constant**.

Proposition

*A regular language is **slender** iff it is a finite union of languages of the form xu^*y , where $x, u, y \in A^*$.*



Sparse languages

A language is **sparse** if its density is **bounded by a polynomial**.

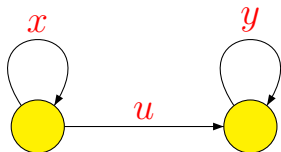
Proposition

*A regular language is **sparse** iff it is a finite union of languages of the form $u_0 v_1^* u_1 \cdots v_n^* u_n$, where $u_0, v_1, \dots, v_n, u_n$ are words.*

Automata for regular slender languages

Theorem

A regular language is *slender* iff its minimal deterministic trim automaton does not contain *any pair of connected simple cycles*.

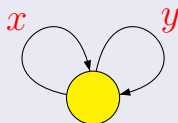


Two connected cycles, where $x, y \in A^+$ and $u \in A^*$.

Automata for regular sparse languages

Theorem

A regular language is *sparse* iff its minimal deterministic trim automaton does not contain *any pattern of the form*



where x and y are nonempty words such that $i(x) \neq i(y)$.

Equations for slender languages

The full language is A^* . Regular slender or full languages form a lattice of languages closed under quotients. Denote by $i(u)$ the first letter (or initial letter) of u .

Theorem

Suppose that $|A| \geq 2$. A regular language of A^ is slender or full iff it satisfies the equations $x \leq 0$ for all $x \in A^*$ and the equation $x^\omega u y^\omega = 0$ for each $x, y \in A^+$, $u \in A^*$ such that $i(uy) \neq i(x)$.*

Theorem

Suppose that $|A| \geq 2$. A regular language of A^* is *sparse or full* iff it satisfies the equations $x \leq 0$ for all $x \in A^*$ and the equations $(x^\omega y^\omega)^\omega = 0$ for each $x, y \in A^+$ such that $i(x) \neq i(y)$.

Theorem

Suppose that $|A| \geq 2$. A regular language of A^* is *slender or coslender* iff its syntactic monoid has a zero and satisfies the equations $x^\omega u y^\omega = 0$ for each $x, y \in A^+$, $u \in A^*$ such that $i(uy) \neq i(x)$.

Theorem

Suppose that $|A| \geq 2$. A regular language of A^* is *sparse or cosparse* iff its syntactic monoid has a zero and satisfies the equations $(x^\omega y^\omega)^\omega = 0$ for each $x, y \in A^+$ such that $i(x) \neq i(y)$.

Closure under inverses of morphisms

A class of languages \mathcal{K} associates to each alphabet A a lattice $\mathcal{K}(A^*)$ of languages of A^* .

\mathcal{K} is closed under inverses of morphisms if for every morphism $\varphi : A^* \rightarrow B^*$, $L \in \mathcal{K}(B^*)$ implies $\varphi^{-1}(L) \in \mathcal{K}(A^*)$.

A class of profinite equations E associates to each alphabet A a set $E(A)$ of profinite equations on $\widehat{A^*}$.

Closure under inverses of morphisms (2)

A class E of profinite equations is **closed under morphism** if, for each morphism $\varphi : A^* \rightarrow B^*$, $u \rightarrow v \in E(A)$ implies $\varphi(u) \rightarrow \varphi(v) \in E(B)$. Such equations are called **identities**.

Theorem

*A class of **regular languages** is defined by a set of **profinite identities** iff it is closed under **inverses of morphisms**.*

Examples

FO[<] (star-free languages) is defined by the identity $x^{\omega+1} = x^{\omega}$.

Σ_1 [<] is defined by the identity $x \leq 1$.

B Σ_1 [<] (piecewise testable languages) is defined by the identities $x^{\omega+1} = x^{\omega}$ and $(xy)^{\omega} = (yx)^{\omega}$.

FO²[<] is defined by the identities $x^{\omega+1} = x^{\omega}$ and $(xy)^{\omega}(yx)^{\omega}(xy)^{\omega} = (xy)^{\omega}$.

\mathcal{C} -morphisms

Let \mathcal{C} be a class of morphisms closed under composition containing the length preserving morphisms.

Examples of such classes \mathcal{C} :

- All morphisms
- Length preserving ($\varphi(A) \subseteq B$)
- Length increasing ($\varphi(A) \subseteq B^+$)
- Length decreasing ($\varphi(A) \subseteq B \cup \{1\}$)
- Uniform ($\varphi(A) \subseteq B^k$)

Closure under inverses of \mathcal{C} -morphisms

A class of languages \mathcal{K} is **closed under inverses of \mathcal{C} -morphisms** if, for each \mathcal{C} -morphism $\varphi : A^* \rightarrow B^*$, the condition $L \in \mathcal{K}(B^*)$ implies $\varphi^{-1}(L) \in \mathcal{K}(A^*)$.

A class E of profinite equations is **closed under \mathcal{C} -morphism** if, for each \mathcal{C} -morphism $\varphi : A^* \rightarrow B^*$, $u \rightarrow v \in E(A)$ implies $\varphi(u) \rightarrow \varphi(v) \in E(B)$. Such equations are called **\mathcal{C} -identities**.

Closure under inverses of \mathcal{C} -morphisms (2)

Theorem

A class of *regular* languages is defined by a set of *profinite \mathcal{C} -identities* iff it is closed under *inverses of \mathcal{C} -morphisms*.

$\mathbf{FO}[\langle + \mathbf{MOD} \rangle] = \mathbf{AC}^0 \cap \mathbf{Reg}$ is defined by the uniform identity $(x^{\omega-1}y)^\omega = (x^{\omega-1}y)^{\omega+1}$ [BCST 92 + Kunc 03]

$\Sigma_1[\langle + \mathbf{MOD} \rangle]$ is defined by the uniform identities $x^{\omega-1}y \leq 1$ and $yx^{\omega-1} \leq 1$. [CPS 06]

Interpretation of equations (Notation: $P = \eta(L)$)

Closed under		Interpretation
\cup, \cap	$u \rightarrow v$	$\hat{\eta}(u) \in P \Rightarrow \hat{\eta}(v) \in P$
quotient	$u \leq v$	$xvy \rightarrow xuy$
complement	$u \leftrightarrow v$	$u \rightarrow v$ and $v \rightarrow u$
quotient and L^c	$u = v$	$xuy \leftrightarrow xvy$
φ^{-1}	Identity	Variables = words
φ^{-1} len. pres.	lp-Identity	Variables = letters
φ^{-1} uniform	uni-Identity	Variables = words of the same length

Polynomial closure

The **polynomial closure** $\text{Pol}(\mathcal{L})$ of a lattice of languages \mathcal{L} of A^* is the set of languages that are **finite unions of marked products** of the form $L_0 a_1 L_1 \cdots a_n L_n$, where the a_i are letters and the L_i are in \mathcal{L} .

Theorem (Branco-Pin 09)

If \mathcal{L} is a *closed under quotients*, then $\text{Pol}(\mathcal{L})$ is defined by the equations of the form $x^\omega y x^\omega \leq x^\omega$, where x, y are profinite words such that the equations $x = x^2$ and $y \leq x$ are satisfied by \mathcal{L} .



Polynomial closure

The **content** of a word is the set of all letters occurring in this word.

Theorem (Arfi 87, Pin-Weil 97)

$\Sigma_2[<]$ is defined by the identities $x^\omega y x^\omega \leq x^\omega$, where x, y are words with the **same content**.

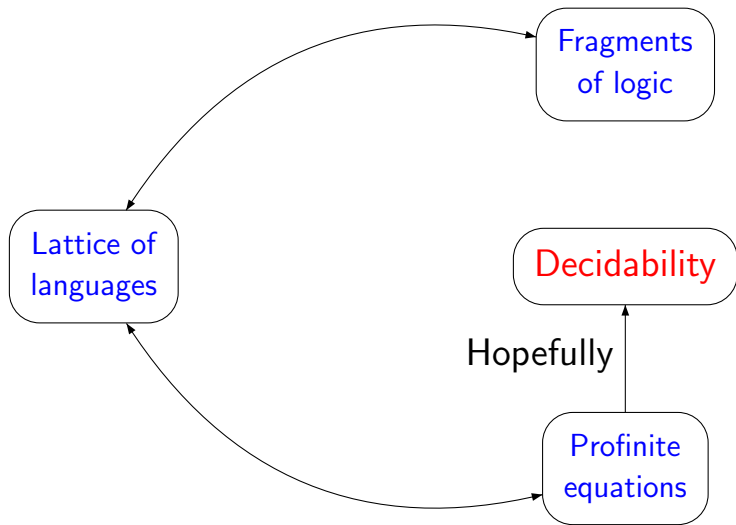
Every **lattice of regular languages** admits an **equational description** (by **profinite equations**).

In particular, every class of regular languages defined by a **fragment of logic** which is closed under **conjunctions** and **disjunctions** (first order, monadic second order, temporal logic, etc.) admits an **equational description**.

Equations for $\Sigma_1[+1]$

$$\begin{aligned}x^\omega y x^\omega &= x^\omega y x^\omega y x^\omega \\x^\omega y x^\omega z x^\omega &= x^\omega z x^\omega y x^\omega \\x^\omega y x^\omega &\leq x^\omega \\x^\omega u y^\omega v x^\omega &\leftrightarrow y^\omega v x^\omega u y^\omega \\y(xy)^\omega &\leftrightarrow (xy)^\omega \leftrightarrow (xy)^\omega x\end{aligned}$$

Back to the virtuous circles



Difficulties

We proved that any **lattice of languages** admits an **equational** description. However...

- This result does not provide any **algorithm** to find a set of equations defining a given class.
- **Profinite words** can be difficult to handle, or even to define.
- Even if you are given a recursively enumerable set of profinite equations, it remains to find an **algorithm for testing** whether a given regular language satisfies these equations.



Extensions

This result extends to **infinite words** (by using Wilke algebras) and very likely to words over **ordinals** or **linear orders** by using the work of Bedon, Bruyère, Carton, Colcombet, Rispal, etc.

It should extend also to **tree** languages and to other structures.

There is also an equational theory for any **lattice of (non regular!) languages**. The price to pay is to use **procompact equations**.

