# New mathematical trends in automata theory

Jean-Éric Pin

LIAFA, CNRS and University Paris 7

22 December 2005, EPFL

# Outline

# A generic decidability problem

## Problem

*Given a class $\mathcal{L}$ of regular languages and a regular language $L$, decide whether or not $L$ belongs to $\mathcal{L}$.*

Various instances of this problem

(1) $\mathcal{L}$ is defined constructively (star-free languages),

(2) $\mathcal{L}$ is the class of languages captured by some fragment of first order logic,

(3) $\mathcal{L}$ is the class of languages captured by some fragment of linear temporal logic.

# Two approaches

Model theoretic approach: use Fraïssé-Ehrenfeucht games (Thomas, Etessami, Wilke, Straubing, etc.)

Algebraic approach : characterize $\mathcal{L}$ by a set of identities (Schützenberger, Simon, Eilenberg, Straubing, Thérien, Pippinger, Pin, etc.)

How do these techniques compare? What are their respective scope?

# A first example: a fragment of temporal logic

Let $A$ be a finite alphabet. A marked word is a pair $(u, i)$ where $u$ is a word and $i$ is a position in $u$ (that is, an element of $\{0, \ldots, |u| - 1\}$).

For each letter $a \in A$, let $p_a$ be a predicate.

## Formulas

(1) The predicates $p_a$ are formula.

(2) If $\varphi$ is a formula, $XF\varphi$ is a formula.

(3) If $\varphi$ and $\psi$ are formula, $\varphi \vee \psi$, $\varphi \wedge \psi$ and $\neg\varphi$ are formula.

# Semantics

(1) If $u = a_0 a_1 \cdots a_{n-1}$, the marked word $(u, i)$ satisfies $p_a$ iff $a_i = a$.

(2) A marked word $(u, i)$ satisfies $XF\varphi$ iff there exists $j > i$ such that $(u, j)$ satisfies $\varphi$. Intuitively, $XF\varphi$ is satisfied at position $i$ iff $\varphi$ is satisfied at some strict future $j$ of $i$.

(3) Connectives have their usual interpretation.

For instance, if $u = ababcb$, $(u, 0)$ and $(u, 2)$ satisfy $p_a$, $(u, 0)$ and $(u, 1)$ satisfy $Fp_a$, $(u, 0)$ and $(u, 2)$ satisfy $p_a \wedge F(p_b \wedge \neg F p_a)$.

# Languages captured by strict future formula

The language defined by a temporal formula $\varphi$ is

$$L(\varphi) = \{u \mid (u, 0) \text{ satisfies } \varphi\}$$

**Examples**. Let $A = \{a, b, c\}$.

(1) $L(p_a) = aA^*$,

(2) $L(XF\varphi) = A^+L(\varphi)$,

(3) $L(p_a \wedge XF(p_b \wedge \neg XFp_a)) = aA^*b\{b, c\}^*$.

The class $\mathcal{L}$ is the smallest Boolean algebra containing the languages $aA^*$ (for each letter $a$) which is closed under the operation $L \to A^+L$.

# A game (Etessami-Wilke)

Let $u$ and $v$ be two words. The game $G_k(u, v)$ is a $k$-turn game between two players, Spoiler ($I$) and Duplicator ($II$). Initially, one pebble is set on the initial position of each word.

In each turn, $I$ chooses one of the pebbles and moves it to the right. Then $II$ moves the other pebble to the right. The letters under the pebbles should match.

A player who cannot play has lost. In particular, if the first letter of the two words don't match, $II$ loses immediately. Player $II$ wins if she was able to play k moves or if $I$ has lost before.

# An example of game

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $u$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ |
| $v$ | $a$ | $c$ | $a$ | $c$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | |

# An example of game

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $u$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ |
| $v$ | $a$ | $c$ | $a$ | $c$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | |

# An example of game

| $u$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $v$ | $a$ | $c$ | $a$ | $c$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | |

# An example of game

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $u$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ |
| $v$ | $a$ | $c$ | $a$ | $c$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | |

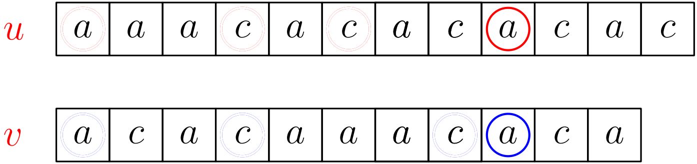
# An example of game

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $u$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ |
| $v$ | $a$ | $c$ | $a$ | $c$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | |

# An example of game

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $u$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ | $a$ | $c$ |
| $v$ | $a$ | $c$ | $a$ | $c$ | $a$ | $a$ | $a$ | $c$ | $a$ | $c$ | $a$ | |

# An example of game



# Languages and Games

Define the (XF-)rank $r(\varphi)$ of a formula $\varphi$ by setting

(1) $r(p_a) = 0$ (for $a \in A$),

(2) $r(XF\varphi) = r(\varphi) + 1$,

(3) $r(\neg\varphi) = r(\varphi)$,
$r(\varphi \vee \psi) = r(\varphi \wedge \psi) = \max(r(\varphi), r(\psi))$.

**Theorem (F-E 80%, Etessami-Wilke 20%)**

*Let $L$ be a language. Then $L = L(\varphi)$ for some formula of rank $\leq n$ iff, for each $u \in L$ and $v \notin L$, Spoiler wins $G_n(u, v)$.*

# An example of game

## Proposition (Wilke)

*For every* $u, v \in A^*$ *and* $a \in A$, *and for* $k \leq n$, $II$ *wins the game* $G_k(a(vu)^n, au(vu)^n)$.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $e$ | $e$ | $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ |

# An example of game

## Proposition (Wilke)

*For every* $u, v \in A^*$ *and* $a \in A$, *and for* $k \leq n$, $II$ *wins the game* $G_k(a(vu)^n, au(vu)^n)$.

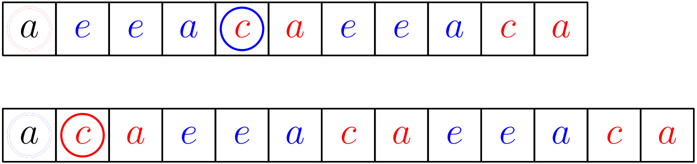
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $e$ | $e$ | $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ |

# An example of game

## Proposition (Wilke)

*For every* $u, v \in A^*$ *and* $a \in A$, *and for* $k \leq n$, $II$ *wins the game* $G_k(a(vu)^n, au(vu)^n)$.



# An example of game

## Proposition (Wilke)

*For every $u, v \in A^*$ and $a \in A$, and for $k \leq n$, $II$ wins the game $G_k(a(vu)^n, au(vu)^n)$.*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $e$ | $e$ | $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ |

# An example of game

## Proposition (Wilke)

*For every* $u, v \in A^*$ *and* $a \in A$, *and for* $k \leq n$, $II$ *wins the game* $G_k(a(vu)^n, au(vu)^n)$.

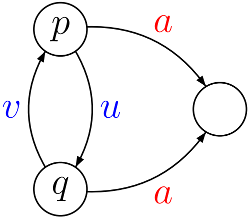
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $e$ | $e$ | $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ | $e$ | $e$ | $a$ | $c$ | $a$ |

# Interpretation on automata

## Theorem (Wilke)

*A language $L$ is in $\mathcal{L}$ iff $L$ is regular and the minimal automaton $\mathcal{A}$ of $L^r$ satisfies the following property: if two states $p$ and $q$ are in the same strongly connected component of $\mathcal{A}$ and if $a$ is any letter, then $p{\cdot}a = q{\cdot}a$.*



# A second example: the star-height problem

Star-height = maximum number of nested star operators occurring in the expression. Operations allowed: Boolean operations, product and star. The complement of $L$ is denoted by $L^c$.

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 over all the expressions representing this language. A language of star-height 0 is also called star-free.

## Problem

*Given a regular language $L$ and an integer $n$, decide whether $L$ has star-height $n$.*

# Examples of star-free languages

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

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

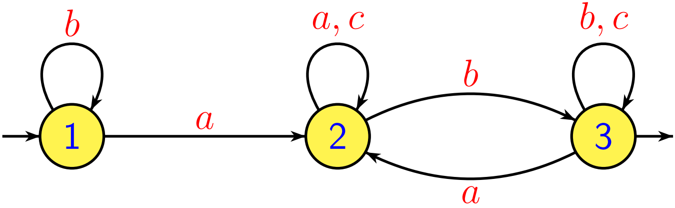
(3) $(ab)^* = \big(b\emptyset^c \cup \emptyset^c a \cup \emptyset^c aa\emptyset^c \cup \emptyset^c bb\emptyset^c\big)^c$ is star-free.

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

**Home work**. Which of these languages are star-free ?

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

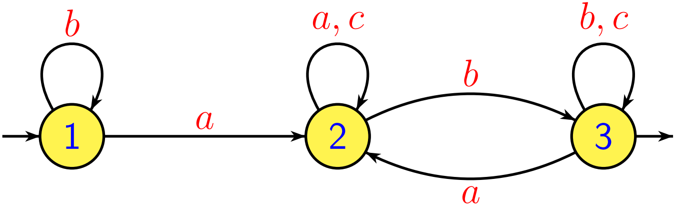
# Transition monoid of an automaton




| 1 | 1 | 2 | 3 |
|---|---|---|---|
| $a$ | 2 | 2 | 2 |
| $b$ | 1 | 3 | 3 |
| $c$ | - | 2 | 3 |

Relations:

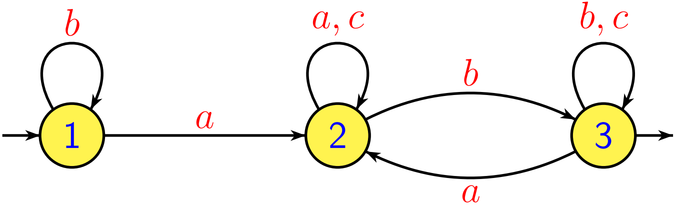
# Transition monoid of an automaton




| 1 | 1 | 2 | 3 |
|---|---|---|---|
| $a$ | 2 | 2 | 2 |
| $b$ | 1 | 3 | 3 |
| $c$ | - | 2 | 3 |

| Relations: |
|---|
| $aa = a$ |

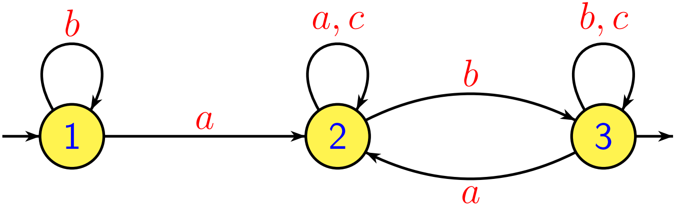
# Transition monoid of an automaton




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

| Relations: |
|---|
| $aa = a$ |

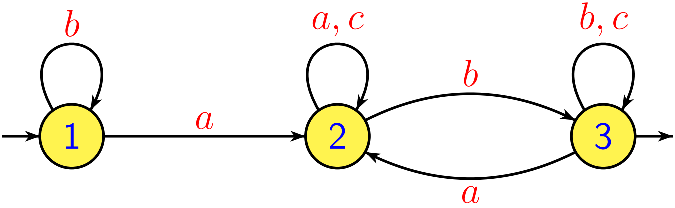
# Transition monoid of an automaton




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

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |

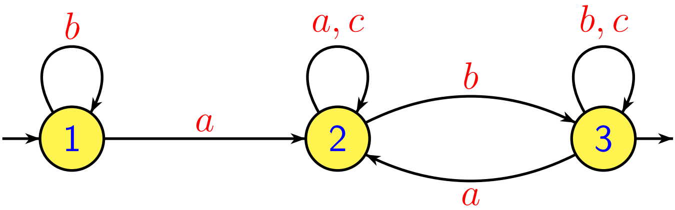
# Transition monoid of an automaton




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

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |
| $ba = a$ |

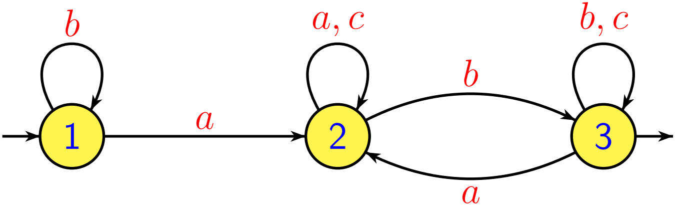
# Transition monoid of an automaton




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

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |
| $ba = a$ |
| $bb = b$ |

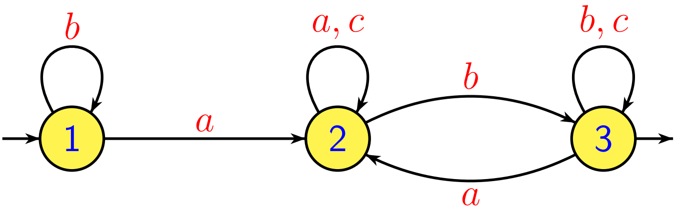
# Transition monoid of an automaton




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

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |
| $ba = a$ |
| $bb = b$ |

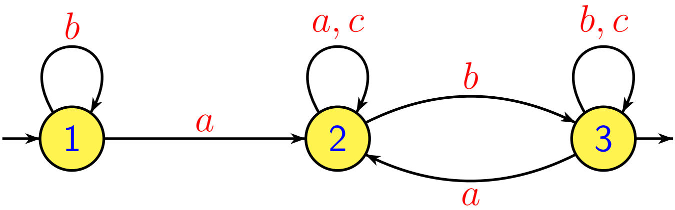
# Transition monoid of an automaton




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

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |
| $ba = a$ |
| $bb = b$ |

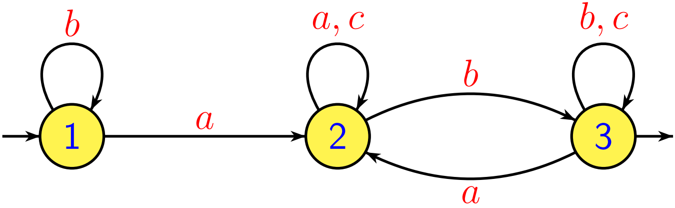
# Transition monoid of an automaton




| 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 |

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |
| $ba = a$ |
| $bb = b$ |
| $cb = bc$ |

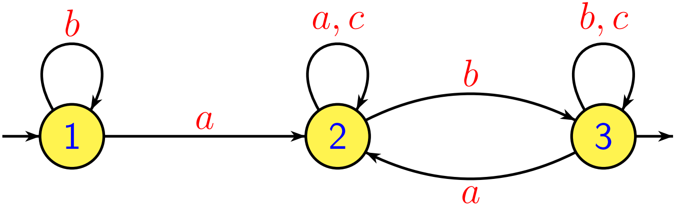
# Transition monoid of an automaton




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

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |
| $ba = a$ |
| $bb = b$ |
| $cb = bc$ |
| $cc = c$ |

# Transition monoid of an automaton

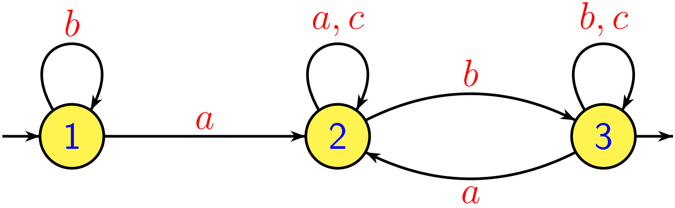



| $1$ | 1 | 2 | 3 |
|---|---|---|---|
| $a$ | 2 | 2 | 2 |
| $b$ | 1 | 3 | 3 |
| $c$ | - | 2 | 3 |
| $ab$ | 3 | 3 | 3 |
| $bc$ | - | 3 | 2 |
| $ca$ | - | 2 | 2 |

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |
| $ba = a$ |
| $bb = b$ |
| $cb = bc$ |
| $cc = c$ |

$abc = ab$

# Transition monoid of an automaton

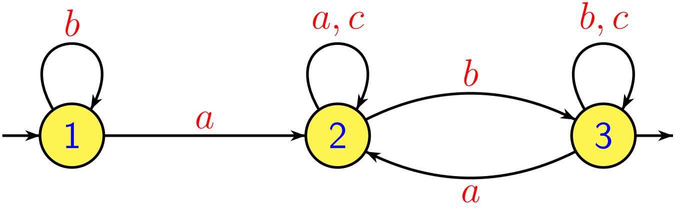



| $1$ | 1 | 2 | 3 |
|---|---|---|---|
| $a$ | 2 | 2 | 2 |
| $b$ | 1 | 3 | 3 |
| $c$ | - | 2 | 3 |
| $ab$ | 3 | 3 | 3 |
| $bc$ | - | 3 | 2 |
| $ca$ | - | 2 | 2 |

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |
| $ba = a$ |
| $bb = b$ |
| $cb = bc$ |
| $cc = c$ |

| |
|---|
| $abc = ab$ |
| $bca = ca$ |

# Transition monoid of an automaton

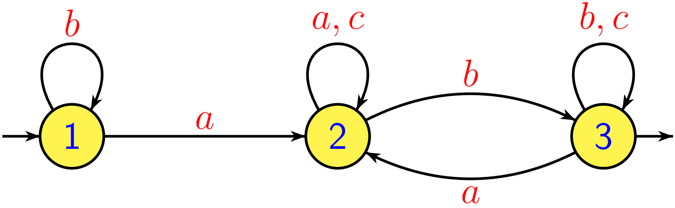



| $1$ | 1 | 2 | 3 |
|---|---|---|---|
| $a$ | 2 | 2 | 2 |
| $b$ | 1 | 3 | 3 |
| $c$ | - | 2 | 3 |
| $ab$ | 3 | 3 | 3 |
| $bc$ | - | 3 | 2 |
| $ca$ | - | 2 | 2 |

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |
| $ba = a$ |
| $bb = b$ |
| $cb = bc$ |
| $cc = c$ |

| |
|---|
| $abc = ab$ |
| $bca = ca$ |
| $cab = bc$ |

# Transition monoid of an automaton




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

| Relations: |
|---|
| $aa = a$ |
| $ac = a$ |
| $ba = a$ |
| $bb = b$ |
| $cb = bc$ |
| $cc = c$ |

| |
|---|
| $abc = ab$ |
| $bca = ca$ |
| $cab = bc$ |
| The end! |

# Schützenberger Theorem

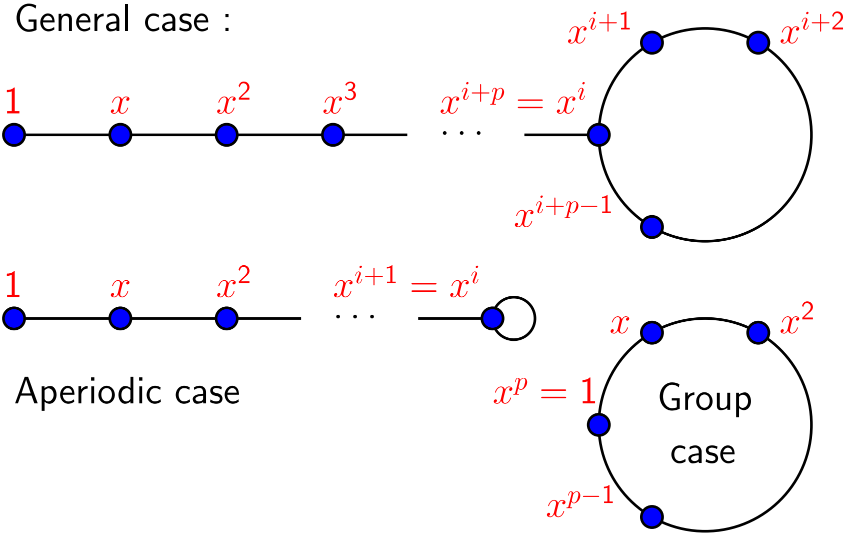
The syntactic monoid of a language is the transition monoid of its minimal automaton (it can also be defined directly).

## Theorem (Schützenberger 1965)

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

A finite monoid $M$ is aperiodic if, for each $x \in M$, there exists $n > 0$ such that $x^n = x^{n+1}$.

# Submonoid generated by $x$



# The algebraic approach

What is the proper setting for Schützenberger's Theorem ?

## Problem

*Is it possible to characterize other classes of regular languages through an algebraic property of their syntactic monoid ?*

Two approaches:

(1) Search for robust classes of finite monoids.

(2) Search for robust classes of regular languages.

# Identities

A monoid $M$ is commutative iff for every $x, y \in M$, $xy = yx$. In other words, iff it satisfies the identity $xy = yx$.

**Example**. A monoid $M$ satisfies the identity $xyzyz = yxzxy$ iff for every $x, y, z \in M$, $xyzyz = yxzxy$.

**Formal definition**. Let $u$ and $v$ be two words on the alphabet $A$. A monoid $M$ satisfies the identity $u = v$, if for every monoid morphism $\varphi : A^* \to M$, $\varphi(u) = \varphi(v)$.

# Varieties of monoids

A variety of monoids is a class of monoids defined by a set of identities (given a set $E$ of identities, the variety defined by $E$ consists of the monoids satisfying all the identities of $E$).

**Theorem (Birkhoff 1935)**

*A class of monoids is a variety iff it is closed under taking submonoids, quotients and (direct) products.*

Nice result, but what remains for finite monoids ?

# Varieties of finite monoids

Let's try the other way around...

## Definition

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

**Examples.**

(1) Finite aperiodic monoids form a variety of finite monoids.

(2) Finite groups form a variety of finite monoids.

# Identities for varieties of finite monoids

Can one extend Birkhoff's theorem ?

A finite monoid $M$ is aperiodic iff there exists a positive integer $m$ such that, for all $n \geq m$, $M$ satisfies all the identities $x^n = x^{n+1}$. One says that the sequence of identities $x^n = x^{n+1}$ ultimately defines the variety of finite aperiodic monoids.

**Theorem (Eilenberg-Schützenberger 1975)**

*Any variety of finite monoids can be ultimately defined by a sequence of identities.*

# More on identities

Intuitively, finite aperiodic monoids satisfy the identity $x^n = x^{n+1}$ "in the limit". Is it possible to turn this intuition into a precise statement?

## Quotation (M. Stone)

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

# Profinite topology

A finite monoid $M$ separates two words $u, v \in A^*$ if $\varphi(u) \neq \varphi(v)$ for some morphism $\varphi : A^* \to M$. Now set, for $u, v \in A^*$,

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

and $d(u, v) = 2^{-r(u,v)}$, with the usual conventions $\min \emptyset = +\infty$ and $2^{-\infty} = 0$.

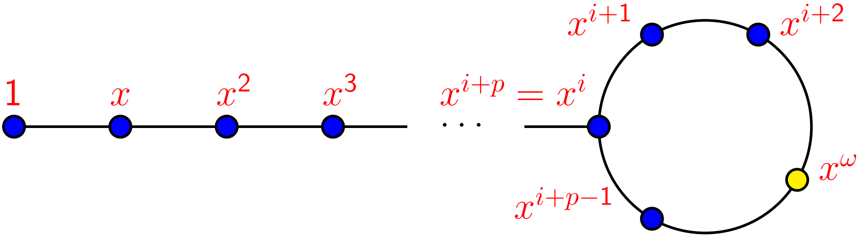
**Intuition**. Two words are close if a large monoid (or a large automaton) is needed to separate them.

# Properties of the profinite metric

Then $d$ is a metric and $(A^*, d)$ is a metric space.
Let $\hat{A}^*$ be its completion.

**Proposition (M. Hall 1950, Reutenauer 1979)**

*For each $x \in A^*$, the sequence $x^{n!}$ converges in $\hat{A}^*$ to a limit, denoted by $x^\omega$.*



# Reiterman's theorem

An identity is still a formal equality of the form $u = v$, but $u$ and $v$ are now elements of the completion $\hat{A}^*$ of the free monoid $A^*$.

## Theorem (Reiterman 1982)

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

Example. The variety of finite groups is defined by the identity $x^\omega = 1$. The variety of finite aperiodic monoids is defined by the identity $x^\omega = x^{\omega+1}$.

# Varieties of languages

Let $\mathbf{V}$ be a variety of finite monoids. Consider the class $\mathcal{V}$ of regular languages the syntactic monoid of which belongs to $\mathbf{V}$. Then $\mathcal{V}$ is closed under the following operations :

(1) Boolean operations,
(2) Residuals ($L \to u^{-1}L$ and $L \to Lu^{-1}$)
(3) Inverse of morphisms.

A class of regular languages closed under (1-3) is called a variety of languages.

# Eilenberg's variety theorem

## Theorem (Eilenberg 1976)

*The correspondence* $\mathbf{V} \to \mathcal{V}$ *is a bijection between varieties of finite monoids and varieties of languages.*

## Corollary (How to use Eilenberg's theorem)

*Any class of regular languages closed under Boolean operations, residuals and inverse morphisms can be defined by identities.*

# First improvement: ordered varieties

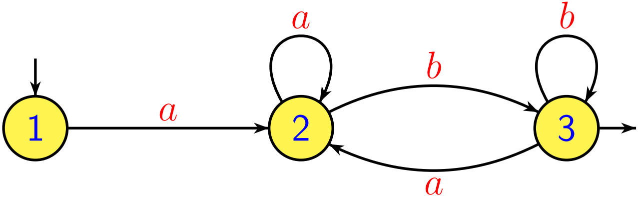
An ordered monoid is a monoid equipped with a stable order $\leq$: $x \leq y \Rightarrow zx \leq zy$ and $xz \leq yz$

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

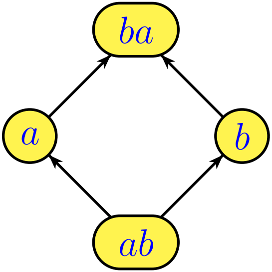
**Theorem (Pin & Weil, 1996)**

*A class of finite ordered monoids is a variety iff it is defined by a set of identities of the form $u \leq v$, where $u, v \in \hat{A}^*$.*

# Syntactic ordered monoid




The syntactic monoid, ordered by $u \leq v$ iff for each $q \in Q$, $q{\cdot}u \leq q{\cdot}v$.



# Positive varieties

Let $\mathbf{V}$ be a variety of finite ordered monoids. The class of regular languages whose ordered syntactic monoid belong to $\mathbf{V}$ is closed under finite union, finite intersection, residuals and inverse of morphisms. Such a class is called a positive variety of languages.

## Theorem (Pin 1995)

*The correspondence $\mathbf{V} \longrightarrow \mathcal{V}$ is a bijection between varieties of ordered finite monoids and positive varieties of languages.*

# Example: reversible automata

A reversible automaton is a finite (possibly incomplete) automaton in which each letter induces a partial one-to-one map from the set of states into itself. Several initial (resp. final) states are possible.

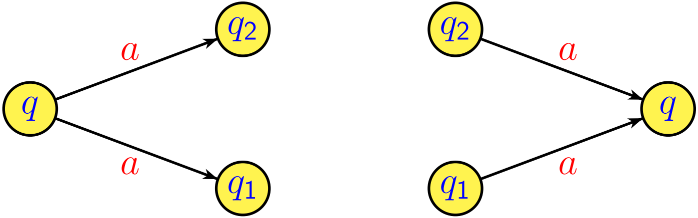



Figure: Forbidden configurations in a reversible automaton.

# Algebraic characterization

## Theorem (Pin 1987)

*A regular language is accepted by some reversible automaton iff its ordered syntactic monoid satisfies the identities.* $x^\omega y^\omega = y^\omega x^\omega$ *and* $1 \leq x^\omega$*. (idempotents commute and* $1$ *is the smallest).*

Consequence. One can decide whether a given regular language is accepted by some reversible automaton.

# First order logic on words

The formula $\exists x \; \mathbf{a}x$ defines the language $A^*aA^*$.

The formula $\exists x \; \exists y \; (x < y) \wedge \mathbf{a}x \wedge \mathbf{b}y$ defines the language $A^*aA^*bA^*$.

The formula $\exists x \; \forall y \; (x < y) \vee (x = y) \wedge \mathbf{a}x$ defines the language $aA^*$.

**Theorem (McNaughton-Papert 1971)**

*First order captures star-free languages.*

# Fragments of first order logic

## Proposition (Easy)

*Existential formulas capture a class defined by the identity* $x \leq 1$.

## Theorem (Simon 1972 + Thomas 1986)

*Boolean combinations of existential formulas capture a class defined by the identities* $(xy)^\omega = (yx)^\omega$ *and* $x^\omega = x^{\omega+1}$.

# Fragments of first order logic (2)

**Theorem (Pin-Weil 1996)**

*Identities for the fragment $\exists^*\forall^*$ are known.*

Major open problem: find the identities for the Boolean combinations of $\exists^*\forall^*$ formulas.

# $\mathcal{C}$-varieties (Straubing 2003).

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

Examples

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

# $\mathcal{C}$-varieties of languages.

A $\mathcal{C}$-variety of languages is a class of regular languages closed under Boolean operations, residuals and inverse of $\mathcal{C}$-morphisms. Positive $\mathcal{C}$-varieties are defined analogously.

For instance, languages captured by $XF$ form a length-preserving variety but do not form a variety.

Is there an algebraic counterpart?

# Stamps

A stamp is a surjective monoid morphism $\varphi : B^* \to M$ from a free monoid onto a finite monoid.

Let $u, v$ in $A^*$ (or, more generally, in $\hat{A}^*$). Then $\varphi$ satisfies the $\mathcal{C}$-identity $u = v$ if, for any $\mathcal{C}$-morphism $f : A^* \to B^*$, $\varphi(f(u)) = \varphi(f(v))$.

# Example of $lp$-identity

Let $u = acbda$, $v = adc$ in $A^*$. A stamp $\varphi : B^* \to M$ satisfies the $lp$-identity

$$a(vu)^\omega = au(vu)^\omega$$

if, for any map $\sigma : A \to B$,
$\varphi(\sigma(a(vu)^\omega)) = \varphi(\sigma(au(vu)^\omega))$.

For instance, if $\sigma(a) = a$, $\sigma(b) = b$, $\sigma(c) = a$ and $\sigma(d) = c$, then $\sigma(u) = aabca$ and $\sigma(v) = aca$ and thus the elements $a(acaaabca)^\omega$ and $aaabca(acaaabca)^\omega$ should be equal in $M$.

# New variety theorems

**Theorem (Esik-Ito 2003, Straubing 2002)**

*$\mathcal{C}$-varieties of languages are in one-to-one correspondence with varieties of stamps.*

**Theorem (Kunc 2003, Pin-Straubing 2005)**

*Varieties of stamps can be described by $\mathcal{C}$-identities.*

# Back to the future

## Theorem (Wilke, revisited)

*The class of regular languages captured by* $XF$ *is defined by the set of* $lp$*-identities of the form* $a(vu)^\omega = au(vu)^\omega$*, where* $a$ *is a letter and* $u$ *and* $v$ *are words.*

## Theorem (Cohen, Perrin, Pin, 1993)

*The class of regular languages captured by* $X, F$ *is defined by the identities* $x^\omega v(x^\omega u x^\omega v)^\omega = (x^\omega u x^\omega v)^\omega$.

# The star-height problem

## Theorem (Pin 1978)

*If the languages of star-height* $\leq 1$ *form a variety, then every language has star-height* $0$ *or* $1$.

## Theorem (Pin, Straubing, Thérien 1989)

*For each* $n \geq 0$*, the languages of star-height* $\leq n$ *are closed under Boolean operations, residuals and inverse of length-decreasing morphisms.*

# The star-height 2 problem

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

If you think there are languages of star-height 2, you should look for some non-trivial identities for these languages.

Otherwise, you should try to prove that every language has star-height one.

# Conclusion (1)

(1) The algebraic approach guarantees the given class can be characterized by identities.

(2) However, it doesn't explicitely gives these identities. Explicitely finding identities can be very hard!

(3) The algebraic approach allows one to use some powerful algebraic/topological tools.

(4) The scope is large: applies to $\omega$-regular languages and more recently to tree-languages [Benedikt, Bojanczyk, Segoufin, Walukiewicz]

# Conclusion (2)

(1) Any class of regular languages closed under finite intersection, finite union, residuals and inverse of $lp$-morphisms can be described by identities.

(2) Most of the time, identities lead to a decidability algorithm (but not always!)

(3) However, algorithms should be converted to automata for better performance.

# Conclusion (3)

(1) F-E games are flexible tools to guess identities. Their scope is very large.

(2) F-E games are very efficient to separate classes, not so much to characterize them.

(3) F-E are not that far from identities. . .