An introduction to the algebraic theory of rational languages

(and a little bit of topology...)

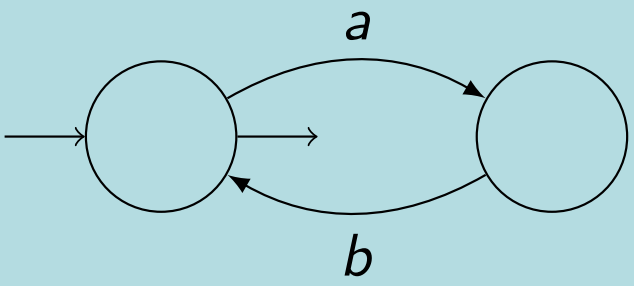
Laure Daviaud

Warsaw University

MFPS 2017

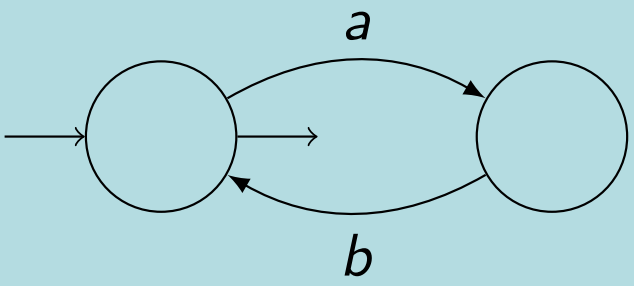
Rational languages

Automata



Rational languages

Automata



Rational languages

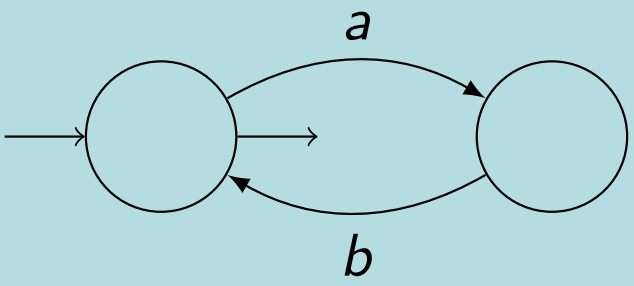
$(ab)^*$

Rational Expressions

Automata

Büchi

Logic



$MSO[<]$ on finite words

- starts with an a
- ends with a b
- the successor of an a is a b
- the successor of a b is an a

Rational languages

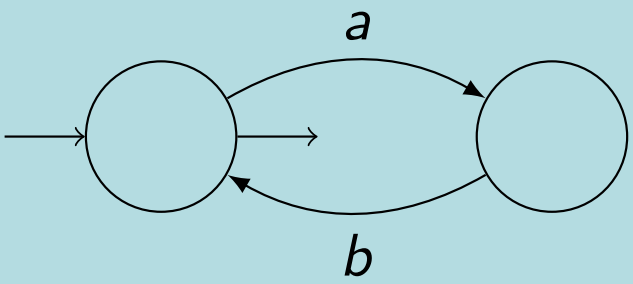
$(ab)^*$

Rational Expressions

Automata

Büchi

Logic



$MSO[<]$ on finite words

- starts with an a
- ends with a b
- the successor of an a is a b
- the successor of a b is an a

Rational languages

$(ab)^*$

Finite monoids

Kleene

Rational Expressions

Algebra

Star-free languages

The set of the star-free languages is the smallest set:

- containing the finite languages (including the empty language),
- closed under finite union, concatenation and complement.

Star-free languages

The set of the star-free languages is the smallest set:

- containing the finite languages (including the empty language),
- closed under finite union, concatenation and complement.

Are the following languages star-free?

- A^*

Star-free languages

The set of the star-free languages is the smallest set:

- containing the finite languages (including the empty language),
- closed under finite union, concatenation and complement.

Are the following languages star-free?

- A^* is star-free $\quad [= \emptyset^c]$

Star-free languages

The set of the star-free languages is the smallest set:

- containing the finite languages (including the empty language),
- closed under finite union, concatenation and complement.

Are the following languages star-free?

- A^* is star-free $\quad [= \emptyset^c]$
- $(ab)^*$

Star-free languages

The set of the star-free languages is the smallest set:

- containing the finite languages (including the empty language),
- closed under finite union, concatenation and complement.

Are the following languages star-free?

- A^* is star-free $\quad [= \emptyset^c]$
- $(ab)^*$ is star-free $\quad [= (bA^* \cup A^*a \cup A^*aaA^* \cup A^*bbA^*)^c]$

Star-free languages

The set of the star-free languages is the smallest set:

- containing the finite languages (including the empty language),
- closed under finite union, concatenation and complement.

Are the following languages star-free?

- A^* is star-free $\quad [= \emptyset^c]$
- $(ab)^*$ is star-free $\quad [= (bA^* \cup A^*a \cup A^*aaA^* \cup A^*bbA^*)^c]$
- $(((ab)^*a((b^ca^*)^c)^*a(a(ba)^*)^c)^*aaab(bab)^c(ab)^*)^*$

Star-free languages

The set of the star-free languages is the smallest set:

- containing the finite languages (including the empty language),
- closed under finite union, concatenation and complement.

Are the following languages star-free?

- A^* is star-free $\quad [= \emptyset^c]$
- $(ab)^*$ is star-free $\quad [= (bA^* \cup A^*a \cup A^*aaA^* \cup A^*bbA^*)^c]$
- $(((ab)^*a((b^ca^*)^c)^*a(a(ba)^*)^c)^*aaab(bab)^c(ab)^*)^*$
- $(aa)^*$

Finite monoids in 3 points

Finite monoids in 3 points

1 . $(M, \cdot, 1)$ with M a finite set equipped with a binary associative operation $\cdot$ and a neutral element 1

Finite monoids in 3 points

1. $(M, \cdot, 1)$ with M a finite set equipped with a binary associative operation $\cdot$ and a neutral element 1

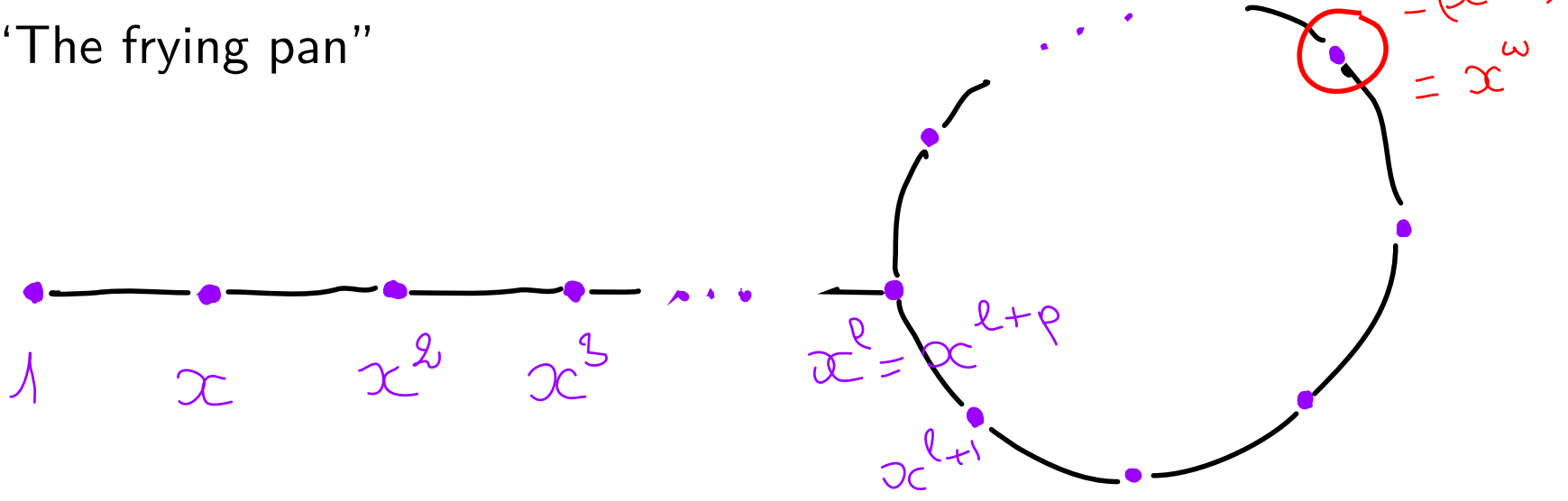
2. Idempotent elements: $e \in M$ such that $e^2 = e$

Finite monoids in 3 points

1. $(M, \cdot, 1)$ with M a finite set equipped with a binary associative operation $\cdot$ and a neutral element 1

2. Idempotent elements: $e \in M$ such that $e^2 = e$

3. "The frying pan"



Every element $x \in M$ has a unique idempotent power $x^\omega = x^{|M|!}$

Recognisability by monoids

Theorem [Kleene]

A language is rational if and only if it is recognised by a finite monoid

Recognisability by monoids

Theorem [Kleene]

A language is rational if and only if it is recognised by a finite monoid

A language L of A^* is recognised by M if there is a morphism $\varphi : A^* \to M$ and $P \subseteq M$ such that $L = \varphi^{-1}(P)$.

$$\begin{array}{ccc} & \overset{\varphi}{} & \\ A^* & \longrightarrow & M \\ \cup\mid & & \cup\mid \\ \varphi^{-1}(P) = L & \dashrightarrow & P \end{array}$$

A canonical object: the syntactic monoid

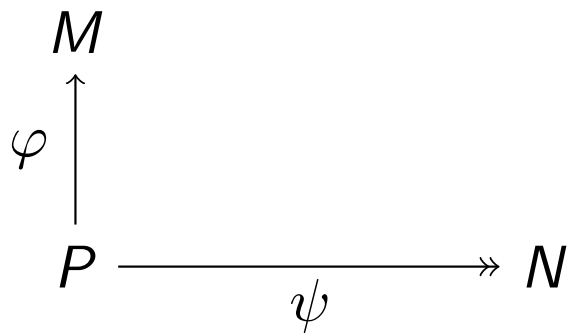
. Three equivalent definitions .

A canonical object: the syntactic monoid

. Three equivalent definitions .

1 . The "smallest one" which recognises L

N divides M if there is an injective morphism φ and a surjective morphism ψ such that:

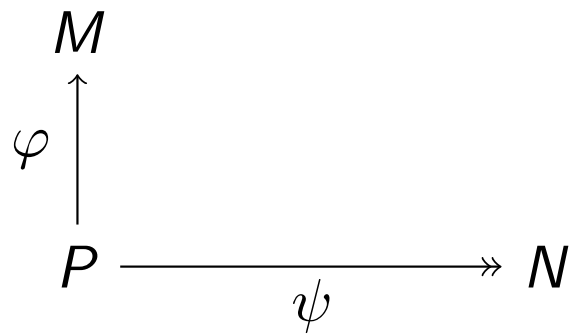


A canonical object: the syntactic monoid

. Three equivalent definitions .

1 . The "smallest one" which recognises L

N divides M if there is an injective morphism φ and a surjective morphism ψ such that:



2 . $A^*/\sim_L$

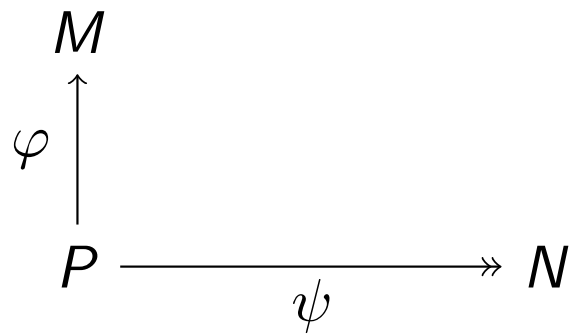
where $\sim_L$ is the equivalence relation of finite index:
$u \sim_L v$ if for all $w, w' \in A^*$, $wuw' \in L$ iff $wvw' \in L$.

A canonical object: the syntactic monoid

. Three equivalent definitions .

1 . The "smallest one" which recognises L

N divides M if there is an injective morphism φ and a surjective morphism ψ such that:

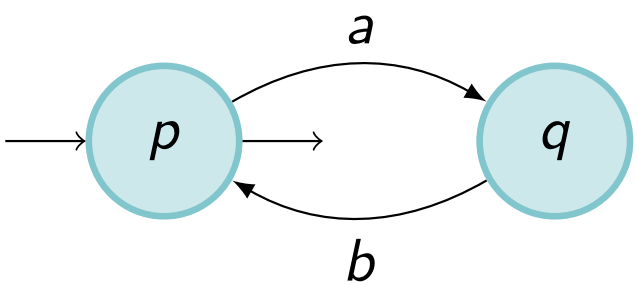


2 . $A^*/\sim_L$

where $\sim_L$ is the equivalence relation of finite index:
$u \sim_L v$ if for all $w, w' \in A^*$, $wuw' \in L$ iff $wvw' \in L$.

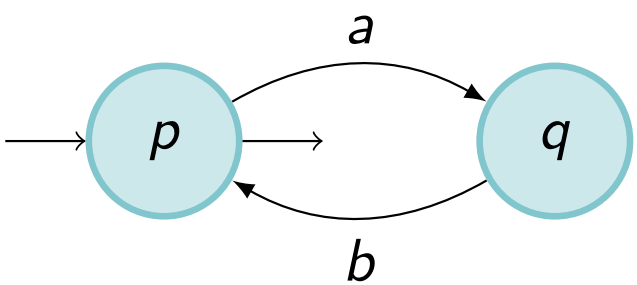
3 . The monoid of the transitions of the minimal automaton

Computing the syntactic monoid



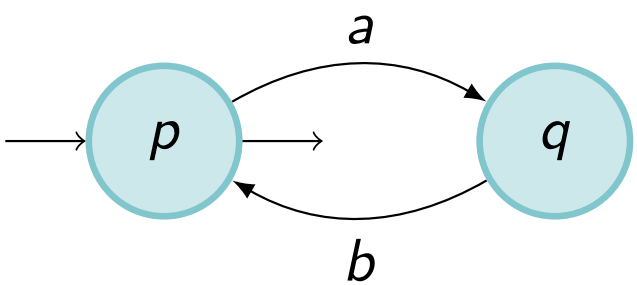
	p	q

Computing the syntactic monoid



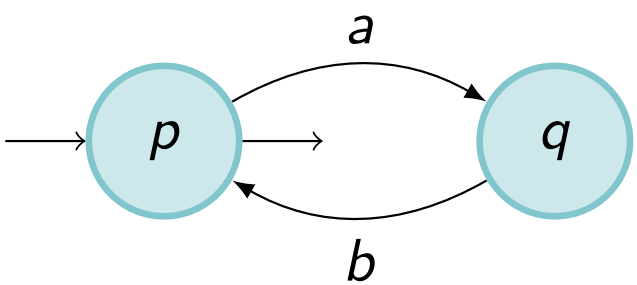
	p	q
ε	p	q

Computing the syntactic monoid



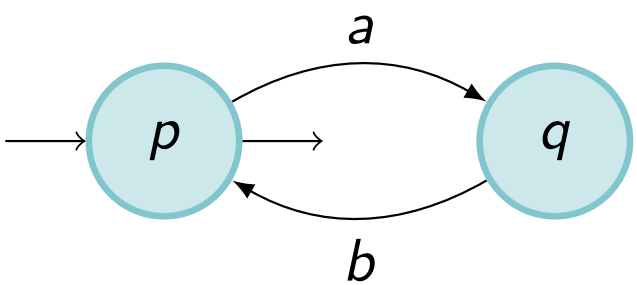
	p	q
ε	p	q
a	q	$\bot$

Computing the syntactic monoid



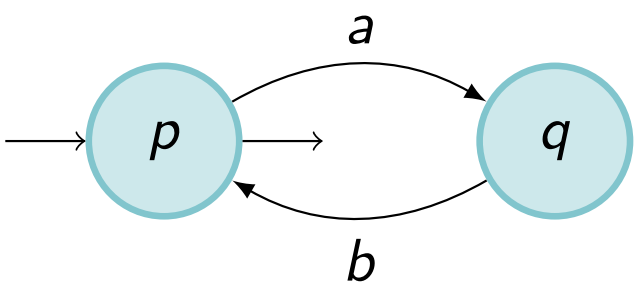
	p	q
ε	p	q
a	q	$\perp$
b	$\perp$	p

Computing the syntactic monoid



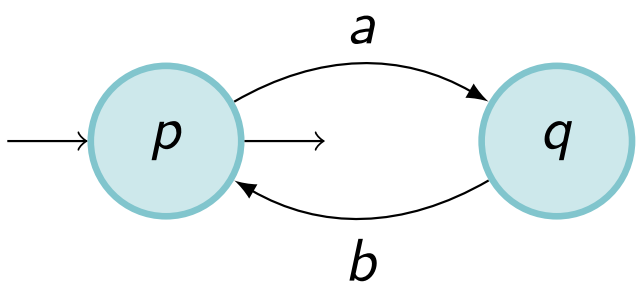
	p	q
ε	p	q
a	q	$\bot$
b	$\bot$	p
aa	$\bot$	$\bot$

Computing the syntactic monoid



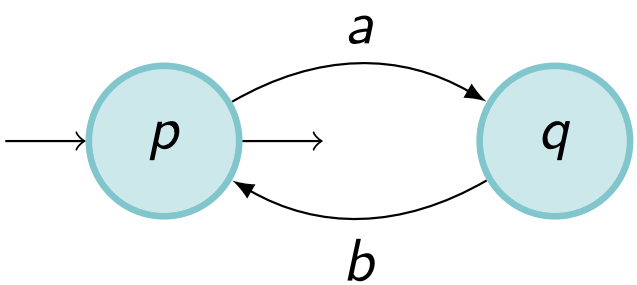
	p	q
ε	p	q
a	q	$\perp$
b	$\perp$	p
aa	$\perp$	$\perp$
ab	p	$\perp$

Computing the syntactic monoid



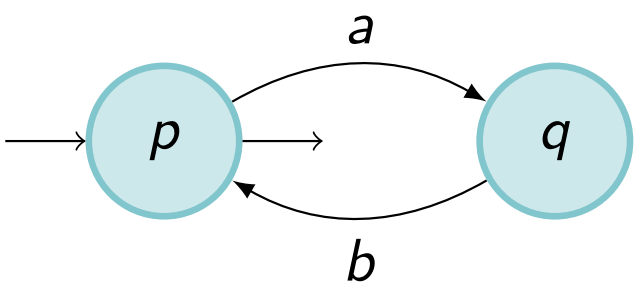
	p	q
ε	p	q
a	q	$\perp$
b	$\perp$	p
aa	$\perp$	$\perp$
ab	p	$\perp$
ba	$\perp$	q

Computing the syntactic monoid



	p	q	
ε	p	q	
a	q	$\perp$	$= aba$
b	$\perp$	p	$= bab$
aa	$\perp$	$\perp$	$= bb = 0$
ab	p	$\perp$	
ba	$\perp$	q	

Computing the syntactic monoid

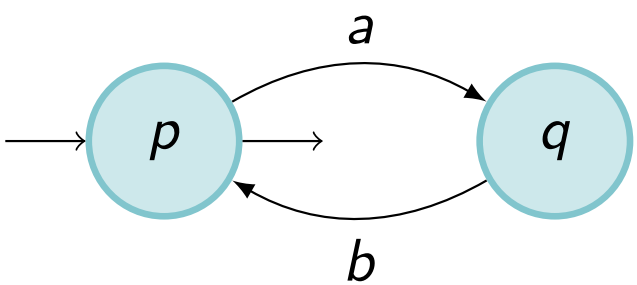


	p	q
ε	p	q
a	q	$\perp$
b	$\perp$	p
aa	$\perp$	$\perp$
ab	p	$\perp$
ba	$\perp$	q

$$M = \{1^*, a, b, ab^*, ba^*, 0^*\}$$

with $aa = bb = 0$, $aba = a$, $bab = b$.

Computing the syntactic monoid



	p	q
ε	p	q
a	q	$\perp$
b	$\perp$	p
aa	$\perp$	$\perp$
ab	p	$\perp$
ba	$\perp$	q

$$M = \{1^*, a, b, ab^*, ba^*, 0^*\}$$

with $aa = bb = 0$, $aba = a$, $bab = b$.

If $\varphi : A^* \to M$ then $(ab)^* = \varphi^{-1}(ab)$.

What about our original question?...

Theorem [Schützenberger]

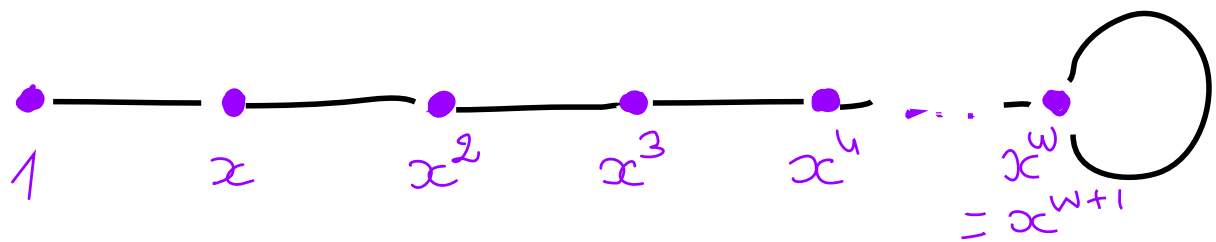
A language is star-free if and only if its syntactic monoid is aperiodic.

What about our original question?...

Theorem [Schützenberger]

A language is star-free if and only if its syntactic monoid is aperiodic.

A monoid M is aperiodic if for all $x \in M$, $x^{\omega+1} = x^{\omega}$.



$\rightarrow (ab)^*$ is star-free.

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

Identities ?

A monoid M satisfies a word-identity $u = v$ with $u, v \in A^*$, if for all morphisms $\varphi : A^* \to M$, $\varphi(u) = \varphi(v)$.

Identities ?

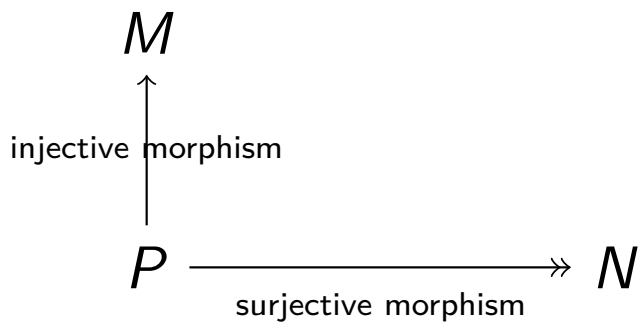
A monoid M satisfies a word-identity $u = v$ with $u, v \in A^*$, if for all morphisms $\varphi : A^* \to M$, $\varphi(u) = \varphi(v)$.

Birkhoff variety of monoids: class of monoids closed under:

- direct product

$$M_1 \times M_2 \times M_3 \times \cdots$$

- division



Identities ?

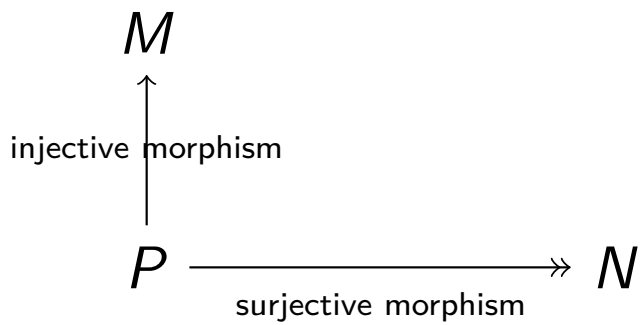
A monoid M satisfies a word-identity $u = v$ with $u, v \in A^*$, if for all morphisms $\varphi : A^* \to M$, $\varphi(u) = \varphi(v)$.

Birkhoff variety of monoids: class of monoids closed under:

- direct product

$$M_1 \times M_2 \times M_3 \times \cdots$$

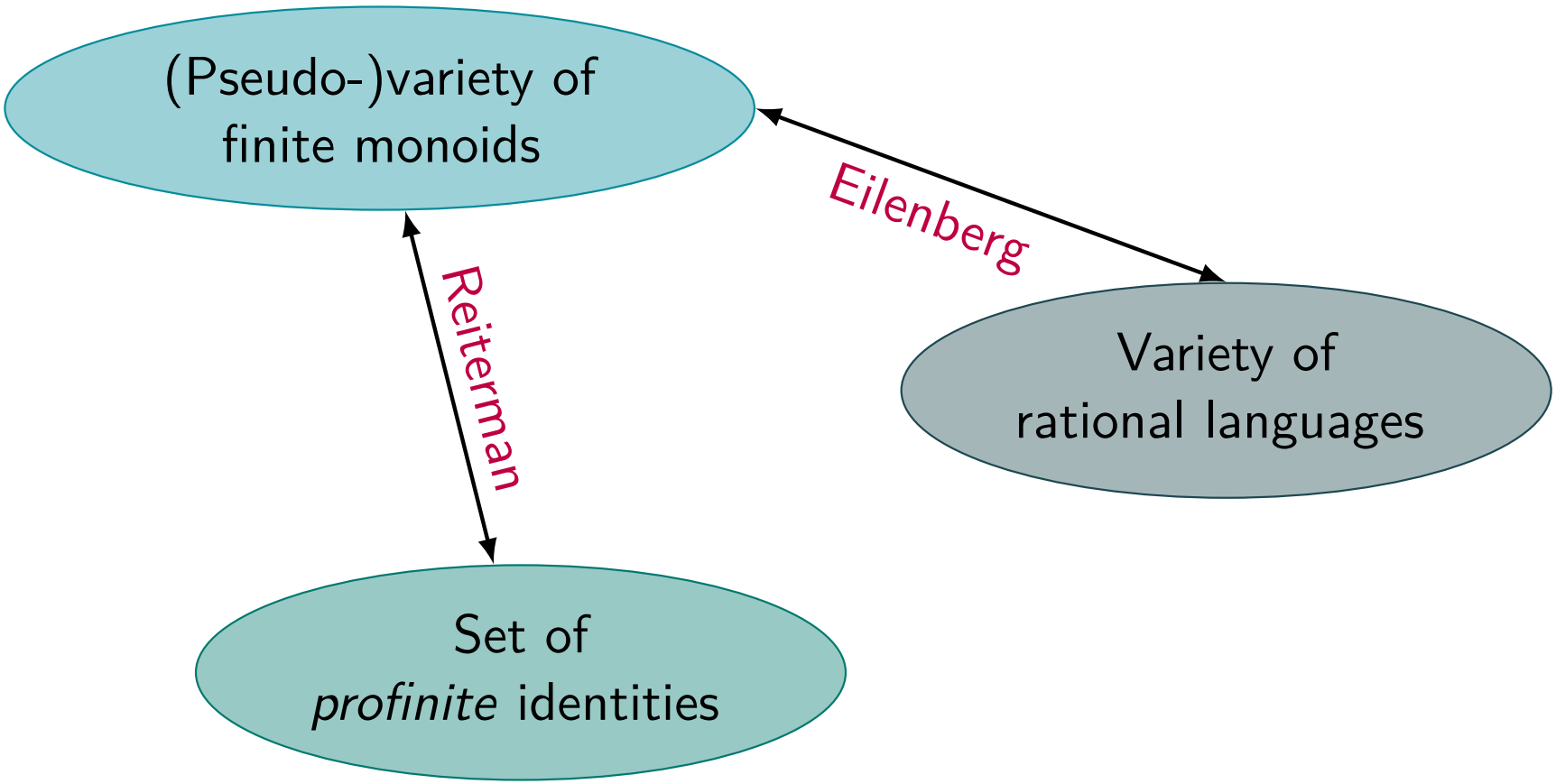
- division



Birkhoff varieties of monoids are defined by a set of identities.

Example: the variety of commutative monoids: $xy = yx$

The variety theorem(s)



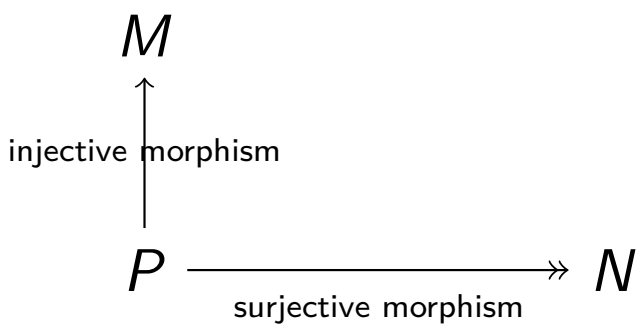
Varieties of finite monoids

A class of finite monoids forms a (pseudo-)variety if:

- it is closed under FINITE product

$$M_1 \times M_2 \times \cdots \times M_\ell$$

- it is closed under division



Varieties of languages

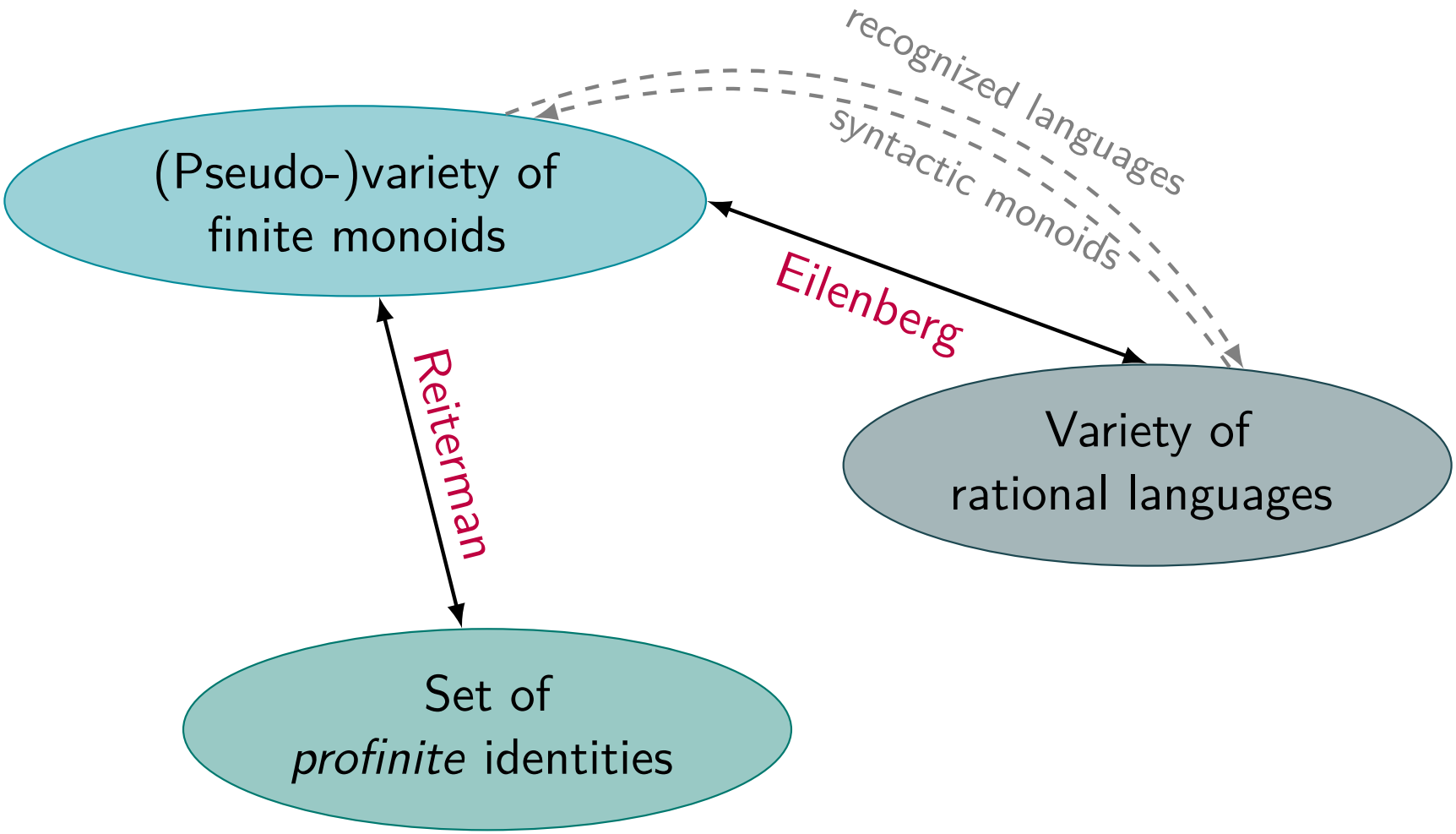
A variety of languages is a class of rational languages

$$\nu(A_1) \cup \nu(A_2) \cup \nu(A_3) \ldots$$

such that:

- for each alphabet A_i, $\nu(A_i)$ is a boolean algebra over A_i
  (closed under finite union, intersection, complement)
- for each alphabet A_i, $\nu(A_i)$ is closed under quotient:
  if $L \in \nu(A_i)$ and $u \in A_i$ then Lu^{-1} and $u^{-1}L \in \nu(A_i)$
- it is closed under inverse image: for each monoid morphism
  $\varphi : A_i^* \to A_j^*$, $L \in \nu(A_j)$ implies $\varphi^{-1}(L) \in \nu(A_i)$

The variety theorem(s)



Separation of words by monoids

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

Separation of words by monoids

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

Example 1: separate $u \neq v$?

Separation of words by monoids

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

Example 1: separate $u \neq v$?

Syntatic monoid of $\{u\}$ (or $\{v\}$...)

Separation of words by monoids

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

Example 1: separate $u \neq v$?

Syntatic monoid of $\{u\}$ (or $\{v\}$...)

Example 2: $a \in A$ - separate a^{99} and a^{100}?

Separation of words by monoids

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

Example 1: separate $u \neq v$?

Syntatic monoid of $\{u\}$ (or $\{v\}$...)

Example 2: $a \in A$ - separate a^{99} and a^{100}?

$\mathbb{Z}/2\mathbb{Z}$

Separation of words by monoids

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

Example 1: separate $u \neq v$?

Syntatic monoid of $\{u\}$ (or $\{v\}$...)

Example 2: $a \in A$ - separate a^{99} and a^{100}?

$\mathbb{Z}/2\mathbb{Z}$

Example 3: $u \in A^*$, $n \in \mathbb{N}$ - separate $u^{n!}$ and $u^{(n+1)!}$?

Separation of words by monoids

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

Example 1: separate $u \neq v$?

Syntatic monoid of $\{u\}$ (or $\{v\}$...)

Example 2: $a \in A$ - separate a^{99} and a^{100}?

$\mathbb{Z}/2\mathbb{Z}$

Example 3: $u \in A^*$, $n \in \mathbb{N}$ - separate $u^{n!}$ and $u^{(n+1)!}$?

$x \in M$ then $x^{|M|!} = x^{(|M|+1)!} =$ the idempotent power of x in M

$\implies \varphi(u)^{|M|!} = \varphi(u)^{(|M|+1)!}$

$u^{n!}$ and $u^{(n+1)!}$ cannot be separated by a monoid of size less than n

Distance over words

A monoid M separates u and v if:

there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

Distance over words

A monoid M separates u and v if:
there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

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

where n is the minimal size of a monoid that separates u and v.

Distance over words

A monoid M separates u and v if:
there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

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

where n is the minimal size of a monoid that separates u and v.

d is an ultrametric distance:

- $d(u, v) = 0$ iff $u = v$
- $d(u, v) = d(v, u)$
- $d(u, v) \leqslant \max(d(u, w), d(w, v))$

Distance over words

A monoid M separates u and v if:
there is a morphism $\varphi : A^* \to M$ such that $\varphi(u) \neq \varphi(v)$.

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

where n is the minimal size of a monoid that separates u and v.

d is an ultrametric distance:

- $d(u, v) = 0$ iff $u = v$
- $d(u, v) = d(v, u)$
- $d(u, v) \leqslant \max(d(u, w), d(w, v))$

The words $u^{n!}$ and $u^{(n+1)!}$ are closer and closer...

Profinite monoids

Definition

Profinite monoid $\widehat{A^*}$:
completion of A^* with respect to the distance d.

- Monoid if u and v sequences of words, $(u.v)_n = u_n v_n$
- Metric space
- A^* dense subset
- Compact

Profinite monoids

Definition

Profinite monoid $\widehat{A^*}$:
completion of A^* with respect to the distance d.

- Monoid if u and v sequences of words, $(u.v)_n = u_n v_n$
- Metric space
- A^* dense subset
- Compact

Idempotent power of $u \in A^*$

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

Profinite identities

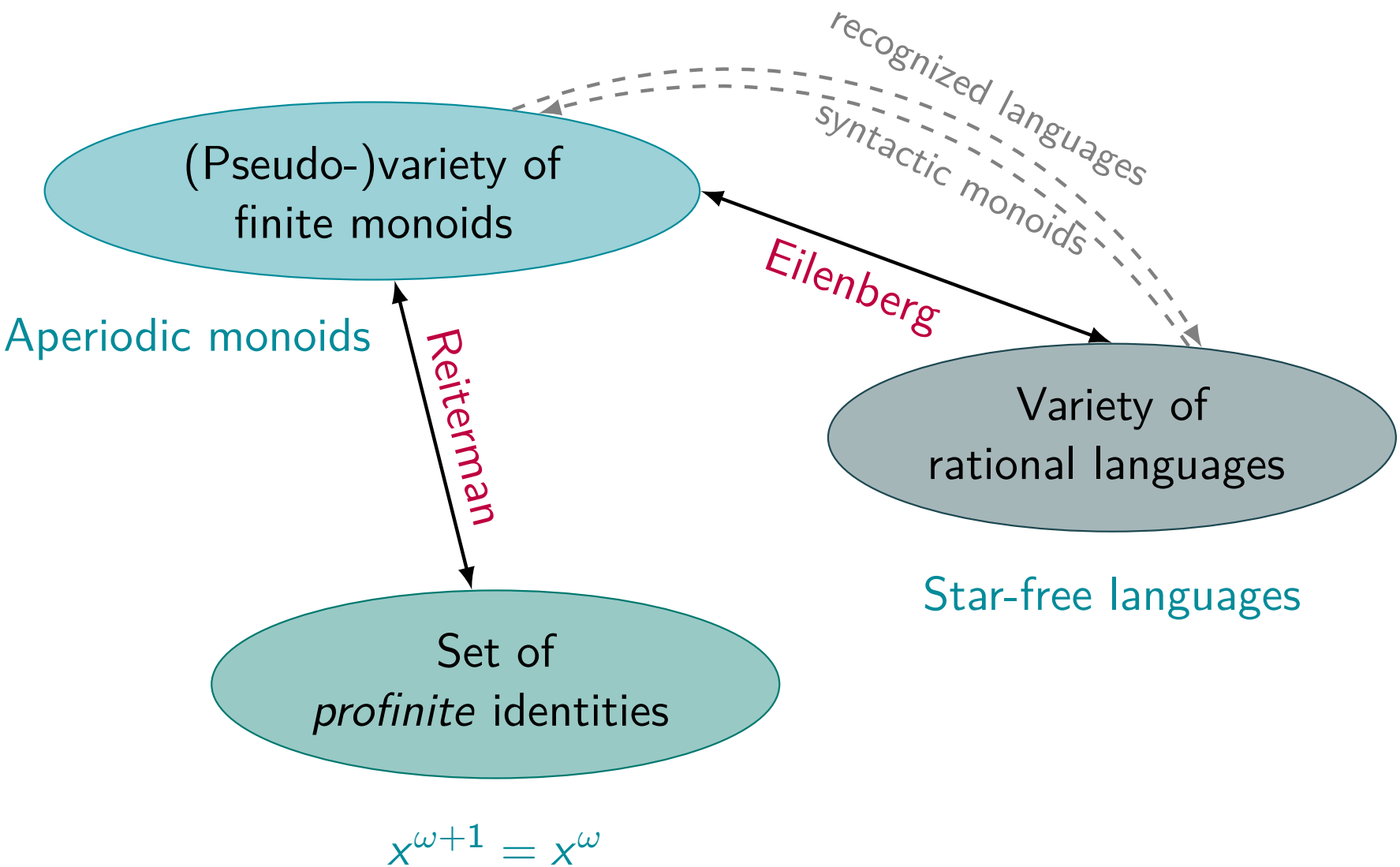
Profinite identity: $u = v$ with $u, v \in \widehat{A^*}$.

A monoid M satisfies a profinite identity $u = v$ with $u, v \in \widehat{A^*}$, if for all morphisms $\varphi : A^* \to M$, $\widehat{\varphi}(u) = \widehat{\varphi}(v)$.

unique way to extend φ by continuity

About $x^\omega = x^{\omega+1}$

The variety theorem(s)



The generalised star-height problem

- One can decide if a given rational language is star-free.
- $(aa)^*$ is not star-free.

The generalised star-height problem

- One can decide if a given rational language is star-free.

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

- *Generalised star-height*: minimal number of nested stars in a generalised expression $(\cup, \cdot, ^c, ^*)$ representing a rational language.

Examples of rational languages of a given generalised star-height?

The generalised star-height problem

- One can decide if a given rational language is star-free.

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

- *Generalised star-height*: minimal number of nested stars in a generalised expression $(\cup, \cdot, ^c, ^*)$ representing a rational language.

Examples of rational languages of a given generalised star-height?

$\longrightarrow$ OPEN : we do not even know if there exist a rational language with star-height at least 2.

The generalised star-height problem

- One can decide if a given rational language is star-free.

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

- *Generalised star-height*: minimal number of nested stars in a generalised expression ($\cup$, $\cdot$, ^{c}, *) representing a rational language.

Examples of rational languages of a given generalised star-height?

$\longrightarrow$ OPEN : we do not even know if there exist a rational language with star-height at least 2.

Mathematical Fondations of Automata Theory - `https://www.irif.fr/~jep/PDF/MPRI/MPRI.pdf`