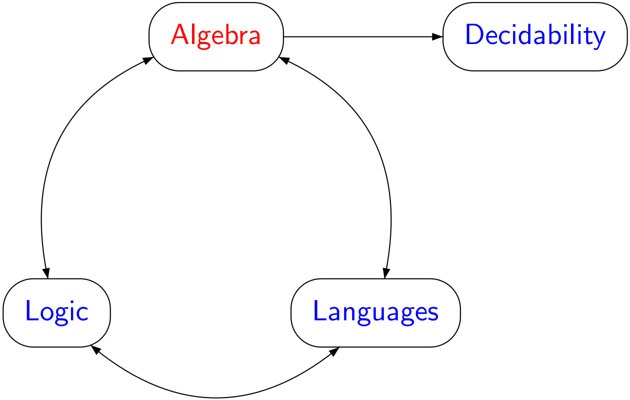
$\mathcal{C}$-varieties and first-order logic with modular predicates

Laura Chaubard
Joint work with J.E. Pin and H. Straubing

LIAFA, Paris 7 and CNRS

Szeged, October 2006

The main idea



Part I

Varieties and $\mathcal{C}$-Varieties

Classify rational languages

Eilenberg's varieties theory aims at classifying rational languages according to the properties of their (ordered) **syntactic monoid**.

$\mathcal{C}$-varieties $\longrightarrow$ syntactic morphism

Classify rational languages

Eilenberg's varieties theory aims at classifying rational languages according to the properties of their (ordered) **syntactic monoid**.

$\mathcal{C}$-**varieties** $\longrightarrow$ **syntactic morphism**

Stutter-invariant languages

- Boolean combinations of languages of the form

$$a_0^+ a_1^+ \cdots a_n^+, \text{ avec } a_0, \ldots a_n \text{ in } A$$

- A language L is stutter-invariant iff, for all **letter** a,

$$a \sim_L a^2$$

Problem: Given the syntactic monoid of a language, one cannot distinguish elements that are congruence classes of **letters**.

Stutter-invariant languages

- Boolean combinations of languages of the form

$$a_0^+ a_1^+ \cdots a_n^+, \text{ avec } a_0, \ldots a_n \text{ in } A$$

- A language L is stutter-invariant iff, for all **letter** a,

$$a \sim_L a^2$$

Problem: Given the syntactic monoid of a language, one cannot distinguish elements that are congruence classes of **letters**.

Stutter-invariant languages

- Boolean combinations of languages of the form

$$a_0^+ a_1^+ \cdots a_n^+, \text{ avec } a_0, \ldots a_n \text{ in } A$$

- A language L is stutter-invariant iff, for all **letter** a,

$$a \sim_L a^2$$

Problem: Given the syntactic monoid of a language, one cannot distinguish elements that are congruence classes of **letters**.

$\mathcal{C}$-varieties of languages

Let $\mathcal{C}$ be a class of morphisms of the form $f : A^* \longrightarrow B^*$, closed under composition.
A **$\mathcal{C}$-variety of languages** is a class of rational languages

1. closed under Boolean operations,
2. closed under residuals,
3. for any morphism $f : A^* \longrightarrow B^*$ that belongs to the class $\mathcal{C}$, and $L \subseteq B^*$,

$$L \in \mathcal{V} \Rightarrow f^{-1}(L) \in \mathcal{V}$$

The class $\mathcal{C}$

We consider certain classes of morphisms of the form

$$f : A^* \rightarrow B^*$$

closed under composition, and containing all length-preserving morphisms.

- *lp* is the class of all **length-preserving** morphisms, i.e morphisms $\varphi : A^* \rightarrow B^*$ such that $\varphi(A) \subseteq B$
- *ne* is the class of all **non-erasing** morphisms, i.e $\varphi(A) \subseteq B^+$,
- *lm* is the class of all **length-multiplying** morphisms, i.e there exists an integer $k > 0$, such that $\varphi(A) \subseteq B^k$,
- *all* is the class of all morphisms,

Positive $\mathcal{C}$-varieties of languages

Let $\mathcal{C}$ be one of the classes of morphisms defined before. A **positive $\mathcal{C}$-variety of languages** is a class of rational languages

1. closed under finite union and intersection,
2. closed under residuals,
3. closed under inverse of morphisms from $\mathcal{C}$.

Examples

1. Stutter-invariant languages form an lp-variety of languages.
2. Languages of generalized star-height $\leqslant n$ form an lp-variety of languages.

Examples

1. Stutter-invariant languages form an *lp*-variety of languages.
2. Languages of generalized star-height $\leqslant n$ form an *lp*-variety of languages.

Examples (2)

- All finite unions of languages of the form

$$A^*a_1A^* \cdots a_kA^*$$

with $k \geqslant 0$ and $a_1, \ldots, a_k$ letters from A, form a positive *all*-variety of languages denoted $\mathbf{J}^+$.

- Boolean combination of languages of $\mathbf{J}^+$ form an *all*-variety of languages denoted $\mathbf{J}$.

Examples (2)

- All finite unions of languages of the form

$$A^* a_1 A^* \cdots a_k A^*$$

with $k \geqslant 0$ and $a_1, \ldots, a_k$ letters from A, form a positive *all*-variety of languages denoted $\mathbf{J}^+$.
- Boolean combination of languages of $\mathbf{J}^+$ form an *all*-variety of languages denoted $\mathbf{J}$.

A new syntactic invariant

- A **stamp** is a onto morphism from a finitely-generated free monoid onto a finite monoid $\varphi : A^* \to M$.
- **The syntactic stamp** of a rational language $L \subseteq A^*$ is the natural morphism

$$\varphi : A^* \to M(L)$$

- **The ordered syntactic stamp** of a rational language $L \subseteq A^*$ is the natural morphism

$$\varphi : A^* \to (M(L), \leqslant_L)$$

A new syntactic invariant

- A **stamp** is a onto morphism from a finitely-generated free monoid onto a finite monoid $\varphi : A^* \to M$.
- **The syntactic stamp** of a rational language $L \subseteq A^*$ is the natural morphism

$$\varphi : A^* \to M(L)$$

- **The ordered syntactic stamp** of a rational language $L \subseteq A^*$ is the natural morphism

$$\varphi : A^* \to (M(L), \leqslant_L)$$

Example: the *lm*-variety **MOD**

- Consider the class of languages whose syntactic stamp is of the form

$$\varphi : A^* \to G$$

where G is a cyclic group and $\varphi(a) = \varphi(b)$ for all letters a and b.
This class forms a *lm*-variety of languages denoted **MOD**.

- **MOD** is the Boolean algebra generated by languages of the form

$$(A^n)^* A^i, \text{ for } 0 \leqslant i < n$$

Example: the *lm*-variety **MOD**

- Consider the class of languages whose syntactic stamp is of the form

$$\varphi : A^* \to G$$

  where G is a cyclic group and $\varphi(a) = \varphi(b)$ for all letters a and b.
  This class forms a *lm*-variety of languages denoted **MOD**.
- **MOD** is the Boolean algebra generated by languages of the form

$$(A^n)^* A^i, \text{ for } 0 \leqslant i < n$$

Stable submonoid

- Let $\varphi : A^* \to M$ be a (ordered) stamp. The set $\varphi(A)$ is an element of the monoid $\mathcal{P}(M)$ of subsets of M. Therefore, it has a unique idempotent power s such that

$$\varphi(A)^s = \varphi(A)^{2s}$$

- The set $\varphi(A)^s \cup \{1\}$ is a submonoid of M called the (ordered) stable submonoid of the stamp φ.
- s is the stability index of φ.

Part II

Logic on words

Logic on words: $FO[<]$

We consider the first-order logic on words with

- classical predicates on letters positions $(\mathbf{a})_{a\in A}$,
- the usual order on positions $<$.

A formula in $FO[<]$

$$\exists x \exists y\ (x < y) \wedge \mathbf{a}x \wedge \mathbf{b}y$$

Interpretation on a word u:

There exist two integers $x < y$ such that, u contains an a in position x and a b in position y.

The set of all words satisfying this formula on a finite alphabet A is the language

$$A^*aA^*bA^*$$

Logic with modular predicates: $FO[< + MOD]$

To the logic defined before, we add two new symbols: the modular predicates

- A unary numerical predicate,

$$\mathbf{MOD}_{\mathbf{r}}^{\mathbf{d}}$$

  interpreted as the set of integers that are congruent to r modulo d.

- A constant symbol m interpreted as the last position in a word.

An example: The formula

$$\exists x \; \mathbf{MOD}_2^3 x \wedge \mathbf{b}x \wedge \mathbf{MOD}_1^2 m$$

defines the language $(A^3)^* AbA^* \cap (A^2)^* A$.

Logic with modular predicates: $FO[< + MOD]$

To the logic defined before, we add two new symbols: the modular predicates

- A unary numerical predicate,

$$\mathbf{MOD_r^d}$$

  interpreted as the set of integers that are congruent to r modulo d.

- A constant symbol m interpreted as the last position in a word.

An example: The formula

$$\exists x \ \mathbf{MOD}_2^3 x \wedge \mathbf{b}x \wedge \mathbf{MOD}_1^2 m$$

defines the language $(A^3)^* AbA^* \cap (A^2)^* A$.

First order

Theorem (McNaughton-Papert 71, Schützenberger 65)

A language is definable in $FO[<]$ iff its syntactic semigroup is aperiodic.

Theorem (Barrington, Compton, Straubing, Thérien 92)

A language is definable in $FO[< + \text{MOD}]$ iff the stable subsemigroup of its syntactic stamp is aperiodic.

Fragments of first-order logic

- Σ_1 denotes the set of existential formulas:

$$\exists x_1 \cdots \exists x_n \varphi(x_1, \cdots x_n)$$

where φ is quantifier-free.

- $\mathcal{B}\Sigma_1$ denotes the set of Boolean combinations of Σ_1-formulas.

Expressive power of $\Sigma_1[< + MOD]$

Given a class of languages $\mathcal{L}$, the polynomial closure $\mathrm{Pol}(\mathcal{L})$ of $\mathcal{L}$ is the class of finite unions of langages of the form

$$L_0 a_1 L_1 a_2 \cdots a_k L_k$$

with $L_0, \ldots, L_k \in \mathcal{L}$ and $a_1, \ldots, a_k$ letters.

Proposition

A language is definable in $\Sigma_1[< + MOD]$ if and only if it belongs to $Pol(\mathcal{M}od)$.

Decidability of $\Sigma_1[< + MOD]$

Theorem

A language belongs to Pol($\mathcal{M}od$) if and only if the stable ordered monoid of its ordered syntactic stamp satisfies the identity $x \leqslant 1$.

Corollary

The class $\Sigma_1[< + MOD]$ is decidable.

Theorem (Thomas 82, Perrin-Pin 86)

A language is definable in $\Sigma_1[<]$ iff its ordered syntactic monoid satisfies the identity $x \leqslant 1$.

Decidability of $\Sigma_1[< + MOD]$

Theorem

A language belongs to $Pol(\mathcal{M}od)$ if and only if the stable ordered monoid of its ordered syntactic stamp satisfies the identity $x \leqslant 1$.

Corollary

The class $\Sigma_1[< + MOD]$ is decidable.

Theorem (Thomas 82, Perrin-Pin 86)

A language is definable in $\Sigma_1[<]$ iff its ordered syntactic monoid satisfies the identity $x \leqslant 1$.

Part III

Wreath Product

Wreath product on monoids

$$M, N \to M \circ N$$

- Crucial operation on monoids that “codes“ for composition of sequential functions.

- Essential tool to decompose semigroups:

Theorem (Krohne-Rhodes 64)

Any finite semigroup divides an alternating wreath product of finite groups and aperiodic semigroups.

Wreath product on monoids

$$M, N \to M \circ N$$

- Crucial operation on monoids that “codes“ for composition of sequential functions.
- Essential tool to decompose semigroups:

Theorem (Krohne-Rhodes 64)

Any finite semigroup divides an alternating wreath product of finite groups and aperiodic semigroups.

Wreath product on varieties

$$\mathbf{V}, \mathbf{W} \to \mathbf{V} * \mathbf{W}$$

- Tool: Given a description of languages belonging to $\mathbf{V}$ and $\mathbf{W}$, the wreath product principle provides a description of languages in $\mathbf{V} * \mathbf{W}$.
- Problem: If $\mathbf{V}, \mathbf{W}$ are both decidable varieties, is $\mathbf{V} * \mathbf{W}$ decidable?
  Answer: NO!

Wreath product on stamps

It's a very technical operation!

Let $\mathbf{V}, \mathbf{W}$ be two $\mathcal{C}$-varieties of stamps. A $(\mathbf{V}, \mathbf{W})$-product is a stamp $\varphi : A^* \to M$ such that:

(1) M is a submonoid of a wreath product $N \circ K$.

(2) Let $\pi : N \circ K \to K$ be the canonical projection. then the stamp $\pi \circ \varphi : A^* \to \pi(M)$ is in $\mathbf{W}$.

(3) Given a in A, one can write $\varphi(a) = (f_a, \pi \circ \varphi(a))$ where f_a is in N^K. now, define the stamp

$$\Phi : (K \times A)^* \to \mathrm{Im}(\Phi) \subseteq N$$

$$\text{by} \quad \Phi(k, a) = f_a(k).$$

Then Φ is required to be in $\mathbf{V}$.

$\mathbf{V} * \mathbf{W}$ is the class of all stamps that $\mathcal{C}$-divide a $(\mathbf{V}, \mathbf{W})$-product.

Wreath product on $\mathcal{C}$-varieties

- Here, the wreath product will remain a “black box“:

$$\mathbf{V}, \mathbf{W} \to \mathbf{V} * \mathbf{W}$$

- Nevertheless, the wreath product principle extends to $\mathcal{C}$-varieties (Esik-Ito 03, Chaubard-Pin-Straubing 05) .
- This new version of the wreath product principle yields

$$\mathrm{Pol}(\mathcal{M}od) = \mathbf{J}^+ * \mathbf{MOD}$$

Whence

$$\Sigma_1[< + \textit{MOD}] = \mathbf{J}^+ * \mathbf{MOD}$$

Wreath product on $\mathcal{C}$-varieties

- Here, the wreath product will remain a “black box“:

$$\mathbf{V}, \mathbf{W} \to \mathbf{V} * \mathbf{W}$$

- Nevertheless, the wreath product principle extends to $\mathcal{C}$-varieties (Esik-Ito 03, Chaubard-Pin-Straubing 05) .
- This new version of the wreath product principle yields

$$\mathrm{Pol}(\mathcal{M}od) = \mathbf{J}^+ * \mathbf{MOD}$$

Whence

$$\Sigma_1[< + \textit{MOD}] = \mathbf{J}^+ * \mathbf{MOD}$$

Wreath product on $\mathcal{C}$-varieties

- Here, the wreath product will remain a “black box“:

$$\mathbf{V}, \mathbf{W} \to \mathbf{V} * \mathbf{W}$$

- Nevertheless, the wreath product principle extends to $\mathcal{C}$-varieties (Esik-Ito 03, Chaubard-Pin-Straubing 05) .
- This new version of the wreath product principle yields

$$\mathrm{Pol}(\mathcal{M}od) = \mathbf{J}^+ * \mathbf{MOD}$$

Whence

$$\Sigma_1[< + \, \mathit{MOD}] = \mathbf{J}^+ * \mathbf{MOD}$$

Part IV

Deciding $\mathcal{B}\Sigma_1[< + \textit{MOD}]$

Languages of $\mathcal{B}\Sigma_1[< + MOD]$

Theorem (Simon 72, Thomas 82)

A language is definable in $\mathcal{B}\Sigma_1[<]$ iff its syntactic monoid is in $\mathbf{J}$.

Boolean combinations of languages in Pol($\mathcal{M}od$).

The extended wreath product principle provides the following algebraic characterisation:

Theorem

A language is a Boolean combination of languages in Pol($\mathcal{M}od$) iff its syntactic stamp belongs to the lm-variety $\mathbf{J} * \mathbf{MOD}$.

decidability???

Languages of $\mathcal{B}\Sigma_1[< + MOD]$

Theorem (Simon 72, Thomas 82)

A language is definable in $\mathcal{B}\Sigma_1[<]$ iff its syntactic monoid is in $\mathbf{J}$.

Boolean combinations of languages in Pol($\mathcal{M}od$).

The extended wreath product principle provides the following algebraic characterisation:

Theorem

A language is a Boolean combination of languages in Pol($\mathcal{M}od$) iff its syntactic stamp belongs to the lm-variety $\mathbf{J} * \mathbf{MOD}$.

decidability???

Languages of $\mathcal{B}\Sigma_1[< + \mathit{MOD}]$

Theorem (Simon 72, Thomas 82)

A language is definable in $\mathcal{B}\Sigma_1[<]$ iff its syntactic monoid is in $\mathbf{J}$.

Boolean combinations of languages in Pol($\mathcal{M}od$).

The extended wreath product principle provides the following algebraic characterisation:

Theorem

A language is a Boolean combination of languages in Pol($\mathcal{M}od$) iff its syntactic stamp belongs to the lm-variety $\mathbf{J} * \mathbf{MOD}$.

decidability???

Languages of $\mathcal{B}\Sigma_1[< + \mathit{MOD}]$

Theorem (Simon 72, Thomas 82)

A language is definable in $\mathcal{B}\Sigma_1[<]$ iff its syntactic monoid is in $\mathbf{J}$.

Boolean combinations of languages in Pol($\mathcal{M}od$).

The extended wreath product principle provides the following algebraic characterisation:

Theorem

A language is a Boolean combination of languages in Pol($\mathcal{M}od$) iff its syntactic stamp belongs to the lm-variety $\mathbf{J} * \mathbf{MOD}$.

decidability???

Derived category

- Given an integer n, let $\pi_n : A^* \to \mathbb{Z}/n\mathbb{Z}$ be the stamp defined by

$$\pi_n(u) = |u| \bmod n$$

- let $\varphi : A^* \to M$ be a stamp. We consider the relational morphism

$$\varphi_n = \pi_n \circ \varphi^{-1}$$

$$\begin{tikzcd}
 & A^* \arrow[dl, "\varphi"'] \arrow[dr, "\pi_n"] & \\
M \arrow[rr, "\varphi_n"] & & \mathbb{Z}/n\mathbb{Z}
\end{tikzcd}$$

Derived category of φ_n

$$\begin{tikzcd} & A^* \arrow[dl, "\varphi"'] \arrow[dr, "\pi_n"] & \\ M \arrow[rr, "\varphi_n"] & & \mathbb{Z}/n\mathbb{Z} \end{tikzcd}$$

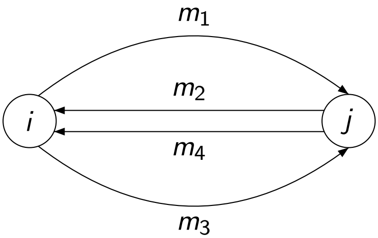
Consider the graph $C_n(\varphi)$ whose vertices are elements of $\mathbb{Z}/n\mathbb{Z}$ and whose edges are the triplets

$$(i, m, j) \quad \text{such that} \quad j - i \in \varphi_n(m)$$

"m has an inverse image by φ whose length is congruent to $j - i$ modulo n."

Knast's equation

The graph $C_n(\varphi)$ satisfies Knast's equation if for all pattern in $C_n(\varphi)$ of the form



we have

$$(m_1 m_2)^\omega (m_3 m_4)^\omega = (m_1 m_2)^\omega m_1 m_4 (m_3 m_4)^\omega$$

Derived category theorem on stamps

Theorem (Chaubard-Pin-Straubing 06)

A stamp φ belongs to $\mathbf{J} * \mathbf{MOD}$ if and only if there exists a positive integer n such that $C_n(\varphi)$ satisfies Knast's equation.

This result is adapted (in two different ways!) from the derived category theorem on monoids.

Decidability of $\mathcal{B}\Sigma_1[< +MOD]$

Theorem

let φ be a stamp with stability index s. Then φ belongs to $\mathbf{J} * \mathbf{MOD}$ *if and only if $C_s(\varphi)$ satisfies Knast's equation.*

Corollary

The class $\mathcal{B}\Sigma_1[< + MOD]$ is decidable.

Future work

- We have proved decidability of both classes but we have no idea of their complexity!
- For Σ_1, we have *lm*-identities, but they do not translate easily into an algorithm.
- For $\mathcal{B}\Sigma_1$, we don't even have identities!
- Open and relevant question: If $\mathbf{V}$ is decidable, is $\mathbf{V} * \mathbf{MOD}$ decidable?