

On recognizable trace languages

Pascal Weil

(joint work with M. Kufleitner)

LaBRI, Université de Bordeaux and CNRS

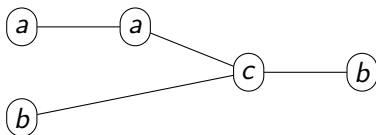
Mondello, 18 June 2007

Outline

- ▶ Trace languages, a model for concurrent behaviors. A renewal of interest in the face of applications to verification; especially the characterization of logically definable trace languages
- ▶ The algebraic approach to recognizability is an essential tool for word languages; is it also the case for trace languages?
- ▶ the basic definitions: do trace languages have a syntactic monoid? how are varieties defined?
- ▶ why have these definitions not given rise to a full-fledge theory as for word languages?
- ▶ we propose a new framework to discuss trace languages (built-up on the old framework of the trace monoid), we see how it allows us to give a robust framework of the classification of recognizable trace languages, and we give a first set of applications
- ▶ this is on-going work (still rough around the edges. . .)

Traces: a classical model

- ▶ Independence alphabet (A, I) : A is a finite alphabet, I an irreflexive and symmetric relation on A
- ▶ the trace monoid $\mathbb{M}(A, I)$ is the quotient of A^* by the congruence generated by $ab = ba$ whenever $(a, b) \in I$
- ▶ traces are one of the most important models used to represent concurrent behavior
- ▶ Each trace is naturally represented as a poset. If $A = \{a, b, c\}$ and $I = \{a, b\}$, then $abacb$ is represented by



Recognizable trace languages: a classical notion

Let $\mu: A^* \longrightarrow \mathbb{M}(A, I)$

- ▶ A trace language $L \subseteq \mathbb{M}(A, I)$ is recognizable if its set of linearizations is recognizable, that is, if $\mu^{-1}(L)$ is recognizable.
- ▶ Equivalent to the monoid-theoretic notion: L is recognizable if there exists $\varphi: \mathbb{M}(A, I) \rightarrow M$ into a finite monoid such that $L = \varphi^{-1}\varphi(L)$. Trace languages have a syntactic monoid.
- ▶ Recognizability is equivalent with MSO-definability (Thomas, 1989)
- ▶ Automata: see the notion of diamond property in automata; equivalence with a beautiful model of automata which captures the notion of independence: Zielonka's automata
- ▶ Rational expressions: there is a problem. If a, b are independent letters, then $(ab)^*$ is not recognizable, see Ochmański's concurrent rational expressions.

A classification of recognizable trace languages?

- ▶ So we seem to be in a similar situation to word languages: the algebraic properties of the syntactic monoid should be useful to characterize and to decide significant classes of recognizable trace languages, in a framework similar to Eilenberg's variety theory.
- ▶ and indeed Guaiana, Restivo, Salemi showed that star-free trace languages are characterized by the aperiodicity of their syntactic monoid. Ebinger, Muscholl showed that this class coincides with FO-definable trace languages.
- ▶ But that is essentially the only example of such a correspondence (until Kufleitner's recent result). There has been no satisfactory Eilenberg-like statement, . . . Why?

So what goes on in words?

- ▶ Schützenberger's result on star-free vs. aperiodic, is an instance of a general correspondence
- ▶ Variety of languages: a class of recognizable languages closed under Boolean operations, left and right residuals and inverse morphisms
- ▶ Pseudovariety of monoids: a class of finite monoids closed under submonoids, homomorphic images and finite direct products.
- ▶ Eilenberg's theorem: (a) the languages whose syntactic monoid lies in a given pseudovariety of monoids V form a variety of languages \mathcal{V} ; (b) the correspondence $V \mapsto \mathcal{V}$ is one-to-one and onto between pseudovarieties and varieties
- ▶ a conceptual framework for many famous results: Simon on piecewise testable languages; Simon and McNaughton on locally testable languages; many others. . .

Extensions and applications (for word languages)

- ▶ Eilenberg's theory extends to operations on varieties of languages
- ▶ *Variants*: semigroups instead of monoids; ordered monoids and positive varieties; \mathcal{C} -varieties
- ▶ Let $V \mapsto \mathcal{V}$, let $\text{Pol}\mathcal{V}$ be the class of unions of products of the form $L_0 a_1 L_1 \cdots a_k L_k$, where $k \geq 0$, the a_i are letters and the L_i are in \mathcal{V} . And let $\text{UPol}V$ be the class of unions of unambiguous products of the same form
- ▶ Then $\text{UPol}V$ is a variety of languages, and the corresponding pseudovariety of monoids is $\text{LI} \textcircled{m} V$ (computable, decidable if V is, etc). And $\text{Pol}\mathcal{V}$ is a positive variety and the corresponding pseudovariety of ordered monoids is $\llbracket x^\omega yx^\omega \leq x^\omega \rrbracket \textcircled{m} V$.

And the results on traces?

- ▶ Except for the result on star-free trace languages (1992), no result until Kufleitner's dissertation (2006).
- ▶ Kufleitner shows an analogue of the results on $\mathcal{V} \mapsto V$ vs $\text{Pol}\mathcal{V} \mapsto \llbracket x^\omega y x^\omega \leq x^\omega \rrbracket \textcircled{m} V$ but only when V consists only of commutative monoids
- ▶ and an analogue of the results on $\mathcal{V} \mapsto V$ vs $\text{UPol}V \mapsto \text{LI} \textcircled{m} V$ but only when $V = J_1$ (idempotent and commutative monoids)
- ▶ In fact, there is no notion of variety of trace languages with an Eilenberg-like theorem, to provide a clean framework

Why doesn't the theory extend (smoothly) to trace languages?

- ▶ Proofs on recognizable trace languages that mimic the proofs on word languages, usually stumble on elementary technical lemmas, that are obvious for morphisms defined on A^* and fail on $\mathbb{M}(A, I)$.
- ▶ Our idea is that the monoid-theoretic framework is not sufficient to deal with trace languages,
- ▶ that the trace monoids have more than a monoid structure: they also have an independence structure.

Independent traces: not a new idea

- ▶ Given (A, I) an independence alphabet (I irreflexive and symmetric), extend I to $\mathbb{M}(A, I)$ by saying that traces u and v are independent if $\text{alph}(u) \times \text{alph}(v) \subseteq I$. Then
- ▶ $(u, v) \in I \implies uv = vu$
- ▶ $(u, vw) \in I \iff (u, v) \in I$ and $(u, w) \in I$
- ▶ $(u, u) \in I$ iff $u = 1$. And $(u, 1) \in I$ for each u .

We propose to abstract out this notion

- ▶ On $\mathbb{M}(A, I)$, all the information on the independence relation is contained in the alphabetic information.
- ▶ It has been considered several times in the literature (Diekert, Gastin, Muscholl, Petit, ...), to transfer this alphabetic information onto the finite monoids recognizing trace languages: if $\varphi: \mathbb{M}(A, I) \rightarrow M$ recognizes L , then so does $\varphi': \mathbb{M}(A, I) \rightarrow M \times 2^A$, where $\varphi'(u) = (\varphi(u), \text{alph}(u))$
- ▶ in certain technical situation, this idea helps, but it increases the size of the recognizing monoid, and the algebraic structure that it yields is intrinsically alphabet-dependent.
- ▶ To have a proper algebraic framework, abstract that out!

Independence monoids: definition

- ▶ An *independence monoid*, or *I-monoid* is a pair (M, I) such that M is a monoid and I is a symmetric relation satisfying
- ▶ $(u, v) \in I \implies uv = vu$
- ▶ $(u, vw) \in I \iff (u, v) \in I$ and $(u, w) \in I$
- ▶ $(u, u) \in I$ iff $u = 1$. And $(u, 1) \in I$ for each u .
- ▶ Morphisms of *I-monoids*: $\varphi: (M, I) \rightarrow (N, J)$ is an *I-morphism* if $\varphi: M \rightarrow N$ is a monoid morphism, and if $(u, v) \in I \implies (\varphi(u), \varphi(v)) \in J$
Already considered for morphisms between trace monoids (Bruyère, de Felice, Guaiana)
- ▶ Also consider *strong I-morphisms*, where $(u, v) \in I \iff (\varphi(u), \varphi(v)) \in J$.

Free I -monoids and skeleton monoids

- ▶ An I -monoid (M, J) is generated (resp. strongly generated) by (A, I) if there exists a map $\psi: A \rightarrow M$ such that $\psi(A)$ generates M and $\psi(I) \subseteq J$ (resp. and in addition $\psi^{-1}(J) \subseteq I$)
- ▶ There is a largest I -monoid generated by (A, I) , and it is $\mathbb{M}(A, I)$ (an initial object in the category of (A, I) -generated I -monoids)
- ▶ Every I -monoid (M, I) has a well-defined least strong homomorphic image $\mathbb{S}(M, I)$, called its skeleton
- ▶ Skeleton monoids are idempotent and commutative, and they encapsulate some fundamental information on the independence structure of the I -monoid

With morphisms comes recognizability

Let (M, I) be an I -monoid and $L \subseteq M$.

- ▶ L is recognizable if $L = \varphi^{-1}\varphi(L)$ for a monoid morphism φ into a finite monoid, I -recognizable if φ is an I -morphism, strongly I -recognizable if φ is a strong I -morphism.
- ▶ For finitely generated I -monoids, and in particular for the trace monoid $\mathbb{M}(A, I)$, these 3 notions are equivalent (luckily!)
- ▶ But a recognizable trace language $L \subseteq \mathbb{M}(A, I)$ has a syntactic I -monoid that may be different from its usual syntactic monoid.
- ▶ Syntactic congruence of L . Let $u \sim_L v$ iff $xuy \in L \iff xvy \in L$ for all $x, y \in \mathbb{M}(A, I)$
 $(x, u) \in I \iff (x, v) \in I$ for all $x \in \mathbb{M}(A, I)$
- ▶ the syntactic I -morphism $\mathbb{M}(A, I) \rightarrow \mathbb{M}(A, I)/\sim_L$ is always a strong I -morphism

An Eilenberg theorem for trace languages

Let us (re)define varieties in a now natural fashion

- ▶ Let a variety of trace languages be a collection $\mathcal{V} = (\mathcal{V}(A, I))_{(A, I)}$ of recognizable trace languages closed under Boolean operations, left and right residuals and inverse I -morphisms
- ▶ Let a pseudovariety of I -monoids be a class V of finite I -monoids closed under taking sub- I -monoids, finite direct products and images under strong I -morphisms. We also require that V contains all skeleton monoids
- ▶ If V is a pseudovariety, let $\mathcal{V} =$ all trace languages recognized by an I -monoid in V . Then $\mathcal{V} =$ all trace languages whose syntactic I -monoid is in V and
- ▶ $V \mapsto \mathcal{V}$ is a one-to-one and onto correspondence between varieties of trace languages and pseudovarieties of I -monoids

Identities for pseudovarieties of I -monoids

- ▶ Pseudovarieties of I -monoids can be defined by sequences of identities (and probably pseudo-identities in the profinite completion)
- ▶ Example: $\llbracket yxyzy = zzyxy \rrbracket$, which is in fact equal to $\llbracket xz = zx \rrbracket$
- ▶ but also $\llbracket yxyzy = zzyxy \rrbracket_{(x,z) \in I}$, which is a different pseudoidentity

Varieties of word languages vs varieties of trace languages

- ▶ A word language $L \subseteq A^*$ is a particular case of a trace language (over (A, I) with $I = \emptyset$)
- ▶ Say that a variety \mathcal{W} of trace languages is *independence-blind* if $L \in \mathcal{W}(A, I)$ iff $\mu^{-1}(L) \in \mathcal{W}(A, \emptyset)$, where $\mu: A^* \rightarrow \mathbb{M}(A, I)$
- ▶ Let \mathcal{V} be a pseudovariety of monoids (containing J_1) and \mathcal{V} the corresponding variety of word languages. Let \mathcal{V}^{ind} be class of all I -monoids (M, I) such that $M \in \mathcal{V}$ and let \mathcal{V}^{ind} be the corresponding variety of trace languages
- ▶ Then $L \in \mathcal{V}^{ind}$ iff $\mu^{-1}(L) \in \mathcal{V}$
- ▶ $\mathcal{V} \mapsto \mathcal{V}^{ind}$ maps bijectively the lattice of pseudovarieties of monoids into the lattice of pseudovarieties of I -monoids, and the corresponding map $\mathcal{V} \mapsto \mathcal{V}^{ind}$ is onto the independence-blind varieties of trace languages

What can we hope to do with this theory?

- ▶ Of course, this variety-theoretic framework is appropriate to account for the known correspondence $\text{Star-free} = \text{FO}[<]\text{-definable trace languages} = \text{Ap}^{ind}$
- ▶ More examples can be derived, and we have concentrated on discussing operations on trace language varieties
- ▶ The hope is that all the (algebraic) hard work has already been done in the theory of word languages, and we can re-use the same concepts, techniques and proofs for trace languages
- ▶ On-going work on wreath products (with Bharat Adsul)
- ▶ Some positive results on Malcev products

It works for Malcev products and polynomial closure! (1/2)

- ▶ Let \mathcal{V} be a variety of trace languages, and let $\text{Pol}\mathcal{V}(A, I)$ consist of all unions of products of the form $L_0 a_1 L_1 \cdots a_n L_n$ (as in the word case)
- ▶ in the word/monoid case, $\text{Pol}\mathcal{V}$ is a positive variety, and the corresponding ordered pseudovariety is described by a Malcev product. Recall the following
- ▶ a monoid M is in $W \overset{\text{m}}{\circlearrowleft} V$ if there exists a relational morphism $\tau: M \rightarrow N$ with $N \in V$ and $\tau^{-1}(e) \in W$ for each $e = e^2 \in N \dots$
- ▶ ... and a relational morphism $\tau: M \rightarrow N$ is a relation such that $\text{graph}(\tau)$ is a submonoid of $M \times N$ whose first projection is onto M

It works for Malcev products and polynomial closure! (2/2)

- ▶ If (M, I) and (N, J) are I -monoids, define a relational I -morphism $\tau: (M, I) \rightarrow (N, J)$ to be a relational morphism such that $(u, v) \in I$ implies $\tau(u) \times \tau(v) \subseteq J$
- ▶ and if V, W are pseudovarieties of (ordered) I -monoids, let (M, I) be an element of $W \textcircled{m} V$ if there exists a relational I -morphism $\tau: (M, I) \rightarrow (N, J)$ with $(N, J) \in V$ and $(\tau^{-1}(e), I) \in W$ for each $e = e^2 \in N$
- ▶ Then we get the exact same statement as in the word case, with essentially the same proof ideas
- ▶ If $\mathcal{V} \mapsto V$, then $\text{Pol}\mathcal{V} \mapsto \llbracket x^\omega y x^\omega \leq x^\omega \rrbracket \textcircled{m} V$ generalizing Kufleitner's earlier results (only for V commutative)

As in the word case, logical consequences

- ▶ Let $\Sigma_n[E]$ be the class of first-order formulas in normal prenex form, with n blocks of quantifier, starting with a block of existential quantifiers — where E is the edge relation in the dependence graph of a trace. Let $\Sigma_n[E]$ also denote the class of trace languages definable by such formulas
- ▶ Then $\Sigma_n[E]$ is a positive variety of trace languages, $\text{Bool}(\Sigma_n[E])$ is a variety of trace languages, and $\Sigma_{n+1} = \text{PolBool}(\Sigma_n[E])$
- ▶ Moreover and if V_n, BV_n are the corresponding pseudovarieties of (ordered) I -monoids, then $V_{n+1} = \llbracket x^\omega y x^\omega \leq x^\omega \rrbracket \circledast BV_n$
- ▶ Decidability results for the lower levels (up to V_2) should follow

Thank you for your attention!