Fragments of first-order logic over infinite words[1]

Volker Diekert (joint work with Manfred Kufleitner)
Universität Stuttgart

Workshop on Automata, Concurrency and Timed Systems
ACTS
Chennai, January 31st, 2009

[1] To appear at STACS 2009, Freiburg im Breisgau, Germany

First-order logic over words

- Atomic predicates: $\lambda(x) = a$, $x < y$
  $L(\varphi) = \{w \in \Gamma^\infty \mid w \models \varphi\}$
- $\varphi = \exists x \exists z \forall y \colon \big(\lambda(x) = a \wedge \lambda(z) = b \wedge x \le y \le z\big)$
- Language $L(\varphi) = a\,\Gamma^* b$.

Fragments:

- *Monomial* $= A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ with $a_i \in \Gamma$, $A_i \subseteq \Gamma$.
  $\exists x_1 \cdots \exists x_k \, \forall y : \quad x_i < y < x_{i+1} \Rightarrow \lambda(y) \in A_i$
- *Polynomial* = union of monomials.
- Unambiguous Polynomials = UPOL = FO^2 for finite words.
- Challenge: Can one decide wether a regular language $L \subseteq \Gamma^*$ is a finite Boolean combination of monomials?
  Problem is open for more than 30 years.

Finite words

Logic	Languages	Algebra
$\mathrm{FO}^1[<]$	$\mathbb{B}\{A^* \mid A \subseteq \Gamma\}$	commutative and idempotent
$\Sigma_1[<]$	simple polynomials	$u \leq 1$
$\mathbb{B}\Sigma_1[<]$	piecewise testable	$\mathcal{J}$-trivial
$\mathrm{FO}^2[<]$ $= \Delta_2[<]$ $= \mathrm{TL}[\mathsf{X}_a, \mathsf{Y}_a]$ $= \mathrm{TL}[\mathsf{XF}, \mathsf{YP}]$	UPol $= \mathrm{Pol} \cap \text{co-}\mathrm{Pol}$ $=$ ranker languages	**DA**
$\Sigma_2[<]$	Pol	$eM_e e \leq e$
FO $= \mathrm{FO}^3[<]$ $= \mathrm{TL}[\mathsf{X}, \mathsf{U}]$	star-free	aperiodic

Examples

i.) $\{a,b\}^*\,ab\,\{a,b\}^*$ is an unambiguous monomial:

$$\{a,b\}^*\,ab\,\{a,b\}^* = b^*\,a\,a^*\,b\,\{a,b\}^*$$

ii.) $\Gamma = \{a,b,c\}$.
Question: $\Gamma^* ab \Gamma^* \in \mathrm{UPol}$? or $\notin \mathrm{UPol}$?

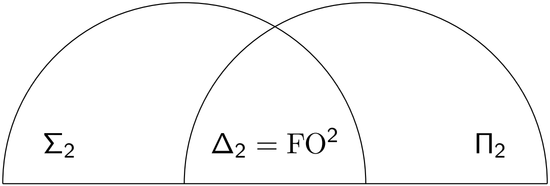
Answer: $\Gamma^* ab \Gamma^* \notin \mathrm{UPol}$, if $|\Gamma| > 2$

Now infinite words:

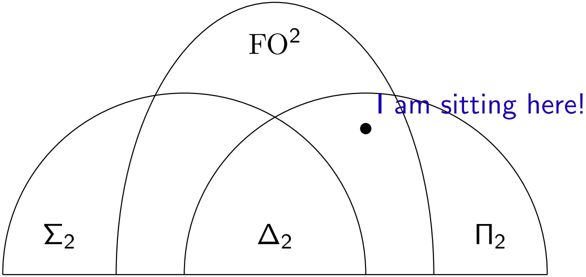
iii.) $\Gamma = \{a,b,c\}$.
Possible question: Where sits $\Gamma^*(b\Gamma^*)^\omega$?

Pictures of the world

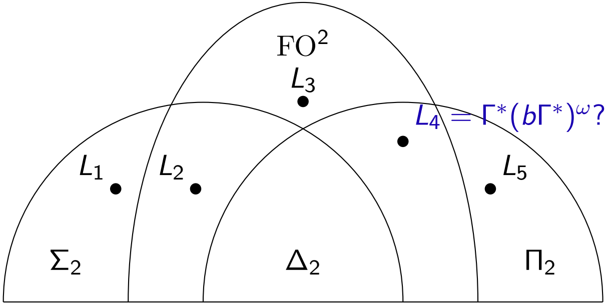
- Finite words:



- Infinite words:



Introduction



Here $\Gamma = \{a, b, c\}$ and

L_1 = "there exists a factor ab"
$= \Gamma^* ab \Gamma^\infty$

L_2 = "finitely many a's"

L_3 = "finitely many a's and infinitely many b's" $= L_2 \cap L_4$

L_4 = "infinitely many b's"

L_5 = "there is no factor ab" $= \Gamma^\infty \setminus L_1$

Prominent fragments FO^2 and Σ_2

- First-order logic:
  - Atomic predicates: $\lambda(x) = a, \ \ x < y$
- Σ_2: $\mathrm{FO}[<]$ sentences starting with a block of existential quantifiers, followed by a block of universal quantifiers and a Boolean combination of atomic formulae.

  $\exists x \exists y \forall z \colon \lambda(x) = a \ \wedge \ y \leq z \ \ \Rightarrow \ \lambda(z) = b$
  defines $\Gamma^* a \, \Gamma^* b^\infty$.
- FO^2: At most two names for variables.

  $\exists x \colon \lambda(x) = a \ \wedge \ \exists y \forall x \colon y \leq x \ \ \Rightarrow \ \lambda(x) = b$
  defines again $\Gamma^* a \, \Gamma^* b^\infty$.

  $\exists x \colon \lambda(x) = a$
  $\wedge \ \exists y \forall x \colon y \leq x \ \ \Rightarrow \ \lambda(x) = b \ \wedge \ \forall y \exists x \colon y < x \ \ \wedge \ \lambda(x) = b$
  defines $\Gamma^* a \, \Gamma^* b^\omega$.

Our starting point

- How can we decide definability in some FO fragment X?
- What is the relation between some FO fragments X and Y?
- How can we unify answers to these questions?

  Finite vs. infinite words

  Traces.

L'objet obscure du désire: **DA**.

Topological ideas

FO^2 can specify exactly which letters occur infinitely often.
Σ_2 can specify which letters may not occur infinitely often.
These properties define (two different) *alphabetic topologies*.
FO^2 yields clopen sets.
Σ_2 yields open sets.

Initial results

1.) For infinite words we have $\mathrm{FO}^2 = \mathbf{DA} \cap$ closed and it was *not easy* to come up with examples that FO^2 and **DA** are different.

2.) Surprise (for us): $L \in \mathbf{DA}$ implies L is closed.

It easy to come up with examples showing $\mathrm{FO}^2 \neq \mathbf{DA}$ for *weak recognizability*:
Let $P = a\Gamma^* b$ and $\Gamma = \{a, b, c\}^*$. Then $L \in \mathrm{FO}^2$.
Consider $L = P^\omega$. It is weakly recognizable, since $L = [ab]^\omega$; and $[ab]$ is idempotent in the syntactic monoid.
L is the language of words with infinitely many occurrences of ba.
This language is not closed.
The syntactic monoid is not in **DA**.

Main results

Algebraic and topological and formal language theoretical characterizations of:

1.) Σ_2,
2.) FO^2

Σ_2 is somewhat easier, so it goes first.

Let $L \subseteq \Gamma^\infty$ be regular. The following are equivalent:

- L is Σ_2-definable.
- L is a polynomial.
- L is open in the alphabetic topology and all idempotents of $\mathrm{Synt}(L)$ are locally top.
- The following three conditions hold for some homomorphism $h : \Gamma^* \to M$ which weakly recognizes L:
  - L is open in the alphabetic topology.
  - All idempotents of M are locally top.
  - L is downward closed for h.

Let $L \subseteq \Gamma^\infty$ be regular. The following are equivalent:

- **L is $\boldsymbol{\Sigma_2}$-definable.**
- L is a polynomial.
- L is open in the alphabetic topology and all idempotents of $\mathrm{Synt}(L)$ are locally top.
- The following three conditions hold for some homomorphism $h : \Gamma^* \to M$ which weakly recognizes L:
  - L is open in the alphabetic topology.
  - All idempotents of M are locally top.
  - L is downward closed for h.

Polynomials

- A *monomial* is a language of the form $P = A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ with $a_i \in \Gamma$ and $A_i \subseteq \Gamma$.
- P is *unambiguous* if for every $\alpha \in P$ there exists a unique factorization $\alpha = u_1 a_1 \cdots u_k a_k \beta$ such that $u_i \in A_i^*$, $\beta \in A_{k+1}^\infty$.
- A *polynomial* is a finite union of monomials.
- A polynomial is *unambiguous* if it is a finite union of unambiguous monomials.
- Example:
  - $A_1^* a_1 \cdots A_k^* a_k$ is a monomial with $A_{k+1} = \emptyset$
  - A^* is a polynomial since $A^* = \emptyset^\infty \cup \bigcup_{a \in A} A^* a$

Let $L \subseteq \Gamma^\infty$ be regular. The following are equivalent:

- **L is Σ_2-definable.**
- **L is a polynomial.**
- L is open in the alphabetic topology and all idempotents of $\mathrm{Synt}(L)$ are locally top.
- The following three conditions hold for some homomorphism $h : \Gamma^* \to M$ which weakly recognizes L:
  - L is open in the alphabetic topology.
  - All idempotents of M are locally top.
  - L is downward closed for h.

Topology

- Basis B of a topology:
  - sets in B are open
  - every open set is a union of sets in B
- Basis of *alphabetic topology* on Γ^∞:
  sets of the form uA^∞, $u \in \Gamma^*$, $A \subseteq \Gamma$.
- L is *open* iff $L = \bigcup W_A A^\infty$.
- A set is *closed*, iff its complement is open.
- A set is *clopen*, iff it is both open and closed.

By definition of the alphabetic topology, polynomials are open. Actually, it is the coarsest topology with this property. The crucial observation is that we have a syntactic description of closure of a polynomial as a finite union of other polynomials.

Lemma

Let $P = A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ be a monomial and $L = P \cap B^{\mathrm{im}}$ for some $B \subseteq A_{k+1}$. Then the closure of a L is given by

$$\bigcup_{\{a_i,\ldots,a_k\} \cup B \subseteq A \subseteq A_i} A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^\infty \cap A^{\mathrm{im}}.$$

"$\subseteq$": Assume $\{a_i, \ldots, a_k\} \cup B \subseteq A \subseteq A_i$.
Let $\alpha \in A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^\infty \cap A^{\mathrm{im}}$.
We have to show that α is in the closure of L.
Let $\alpha = u\beta$ with $u \in A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^*$ and $\beta \in A^\infty \cap A^{\mathrm{im}}$.
We show that $uA^\infty \cap L \neq \emptyset$.
Choose some $\gamma \in B^\infty \cap B^{\mathrm{im}}$, as $B \subseteq A_{k+1}$ by hypothesis, we see

$$u a_i \cdots a_k \gamma \in P$$

Hence:

$$u a_i \cdots a_k \gamma \in uA^\infty \cap L.$$

The other direction: For $\alpha \in \overline{L}$ with $\mathrm{im}(\alpha) = A$.
Write $\alpha \in uv_1 \cdots v_{k+1} A^\infty \cap A^{\mathrm{im}}$ with $\mathrm{alph}(v_j) = A$.
As α is in the closure, there exists $\gamma \in A^\infty$ such that $uv_1 \cdots v_{k+1}\gamma \in P \cap B^{\mathrm{im}}$. This implies $B \subseteq A$.
Since $uv_1 \cdots v_{k+1}\gamma \in A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ some v_j is inside some A_i^*.

Thus, there are some $1 \leq i, j \leq k+1$ such that $uv_1 \cdots v_{j-1} \in A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^*$, $v_j \in A_i^*$, and $v_{j+1} \cdots v_{k+1}\gamma \in A_i^* a_i \cdots A_k^* a_k A_{k+1}^\infty \cap A^\infty$.
Therefore $\{a_i, \ldots, a_k\} \subseteq A \subseteq A_i$, too. It follows $\alpha \in A_1^* a_1 \cdots A_{i-1}^* a_{i-1} A_i^\infty \cap A^{\mathrm{im}}$. □

Let $L \subseteq \Gamma^\infty$ be regular. The following are equivalent:

- **L is Σ_2-definable.**
- **L is a polynomial.**
- **L is open in the alphabetic topology** and all idempotents of $\mathrm{Synt}(L)$ are locally top.
- The following three conditions hold for some homomorphism $h : \Gamma^* \to M$ which weakly recognizes L:
  - **L is open in the alphabetic topology.**
  - All idempotents of M are locally top.
  - L is downward closed for h.

Algebra

- Ordered monoid $(M, \leq)$:

$$u \leq v \ \text{ implies } \ sut \leq svt$$

- $M_e = \{s \in M \mid e \in MsM\}^*$ for idempotent $e \in M$
- e is called *locally top*,if $ese \leq e$ for all $s \in M_e$
- $M \in \mathbf{DA}$ iff $ese = e$ for all idempotents $e \in M$ and all $s \in M_e$,
- i.e., all idempotents are locally top and locally bottom.
- Syntactic (ordered) monoid $\mathrm{Synt}(L)$ of ∞-language $L \subseteq \Gamma^\infty$

Let $L \subseteq \Gamma^\infty$. The *syntactic preorder* $u \leq_L v$ is defined by both implications:

$$xvyz^\omega \in L \;\Rightarrow\; xuyz^\omega \in L$$
$$x(vy)^\omega \in L \;\Rightarrow\; x(uy)^\omega \in L$$

$u \equiv_L v$ means $u \leq_L v$ and $v \leq_L u$.
The congruence classes $[u]_L = \{v \in \Gamma^* \mid u \equiv_L v\}$ form the *syntactic monoid* $\mathrm{Synt}(L)$, and
$(\mathrm{Synt}(L), \leq_L)$ becomes an ordered monoid.
A regular language L can be written as a finite union of languages of type $[u]_L\,[v]_L^\omega$ where $u, v \in \Gamma^*$ with $uv \equiv_L u$ and $v^2 \equiv_L v$.

Σ_2

Let $L \subseteq \Gamma^\infty$ be regular. The following are equivalent:

- **L is $\boldsymbol{\Sigma_2}$-definable.**
- **L is a polynomial.**
- **L is open in the alphabetic topology and all idempotents of $\mathrm{Synt}(\mathbf{L})$ are locally top.**
- The following three conditions hold for some homomorphism $h : \Gamma^* \to M$ which weakly recognizes L:
  - **L is open in the alphabetic topology.**
  - **All idempotents of M are locally top.**
  - L is downward closed for h.

Algebra (cont'd)

- $h : \Gamma^* \to (M, \leq)$ surjective homomorphism, $[s] = h^{-1}(s)$
- $(s, e) \in M \times M$ is a *linked pair*, if $se = s$ and $e^2 = e$
- *h weakly recognizes L*, if

$$L = \bigcup \{[s][e]^\omega \mid (s, e) \text{ is a linked pair and } [s][e]^\omega \subseteq L\}$$

- *h strongly recognizes L*, if

$$L = \bigcup \{[s][e]^\omega \mid (s, e) \text{ is a linked pair and } [s][e]^\omega \cap L \neq \emptyset\}$$

- L is *downward closed (on finite prefixes)* for h, if $[s][e]^\omega \subseteq L$ implies $[t][e]^\omega \subseteq L$ for all $s, t, e \in M$ where $t \leq s$.

Let $L \subseteq \Gamma^\infty$ be regular. The following are equivalent:

- **L is Σ_2-definable.**
- **L is a polynomial.**
- **L is open in the alphabetic topology and all idempotents of $\mathrm{Synt}(L)$ are locally top.**
- **The following three conditions hold for some homomorphism $h : \Gamma^* \to M$ which weakly recognizes L:**
  - **L is open in the alphabetic topology.**
  - **All idempotents of M are locally top.**
  - **L is downward closed for h.**

Topology (cont'd)

- $\mathrm{im}(\alpha)$ = letters which have infinitely many occurrences in α.
- $A^{\mathrm{im}} = \{\alpha \in \Gamma^\infty \mid \mathrm{im}\,\alpha = A\}$.
- Basis of *strict alphabetic topology*: sets $uA^\infty \cap A^{\mathrm{im}}$.

Lemma

If $L \subseteq \Gamma^\infty$ *is strongly recognized by some homomorphism* $h : \Gamma^* \to M \in \mathbf{DA}$*, then* L *is clopen in the strict alphabetic topology.*

Strong rec. in **DA** implies clopen

Proof.

- Since h also strongly recognizes $\Gamma^\infty \setminus L$ as well, it is enough to show that L is open.
- Let $\alpha \in L$ with $\alpha \in [s][e]^\omega$ for some linked pair (s, e) and let $A = \mathrm{im}(\alpha)$.
- We show that $[s]A^\infty \cap A^{\mathrm{im}} \subseteq L$:
- Let $\beta \in [s]A^\infty \cap A^{\mathrm{im}}$.
- Then $\beta = uv\gamma$ with $h(u) = s$, $h(v) = r$, $\gamma \in [f]^\omega$ where $v \in A^*$, $\mathrm{alph}(\gamma) = \mathrm{im}(\gamma) = A$, and (r, f) is a linked pair.
- Since $M \in \mathbf{DA}$: $s = se = serfe = srfe$ and $efe = e$ and $fef = f$.
- Now, by strong recognition:

$$\beta \in [sr][f]^\omega = [sr][fef]^\omega = [srfe][efe]^\omega = [s][e]^\omega \subseteq L.$$

□

Let $L \subseteq \Gamma^\infty$. The following assertions are equivalent:

- L is FO^2-definable.
- L is regular and $\mathrm{Synt}(L) \in \mathbf{DA}$.
- L is strongly recognized by some homomorphism $h : \Gamma^* \to M \in \mathbf{DA}$.
- L is clopen in the strict alphabetic topology and L is weakly recognized by some homomorphism $h : \Gamma^* \to M \in \mathbf{DA}$.
- L is closed in the strict alphabetic topology and L is weakly recognized by some homomorphism $h : \Gamma^* \to M \in \mathbf{DA}$.
- L is a finite union of sets of the form $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty \cap A_{k+1}^{\mathrm{im}}$, where each language $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ is an unambiguous monomial.

Weak is not strong

For infinite words we have $\mathrm{FO}^2 \neq$ **DA** w.r.t. weak recognizability; and it *easy* to come up with examples that weak recognizability goes beyond **DA**.

Open:

- Characterize weakly recognizable languages for **DA**.

Let $L \subseteq \Gamma^\infty$. The following assertions are equivalent:

- L is both FO^2-definable and Σ_2-definable.
- L is FO^2-definable and open in the alphabetic topology.
- L is a finite union of unambiguous monomials of the form $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$.
- L is the interior of some FO^2-definable language.

There is a dual statement for Π_2.

Let $L \subseteq \Gamma^\infty$. The following assertions are equivalent:

- L is Δ_2-definable.
- L is FO^2-definable and L is clopen in the alphabetic topology.
- L is a finite union of unambiguous closed monomials $A_1^* a_1 \cdots A_k^* a_k A^\infty$, i.e., there is no $1 \leq i \leq k$ such that $\{a_i, \ldots, a_k\} \subseteq A_i$.
- L is regular, $\mathrm{Synt}(L) \in \mathbf{DA}$, and for all linked pairs (s, e), (t, f) with $s \mathrel{\mathcal{R}} t$ (i.e., there exist $x, y \in \mathrm{Synt}(L)$ such that $s = tx$ and $t = sy$) we have

$$[s][e]^\omega \subseteq L \;\Leftrightarrow\; [t][f]^\omega \subseteq L.$$

Let $A \subseteq \Gamma$. The following assertions are equivalent:

- $L \cap A^{\mathrm{im}}$ is FO^2-definable.
- There are languages $L_\sigma \in \Sigma_2$ and $L_\pi \in \Pi_2$ such that

$$L \cap A^{\mathrm{im}} = L_\sigma \cap A^{\mathrm{im}} = L_\pi \cap A^{\mathrm{im}}.$$

- For $A = \emptyset$ we have $\emptyset^{\mathrm{im}} = \Gamma^*$ and we recover

$$\mathrm{FO}^2 = \mathbf{DA} = \Delta_2 = \Sigma_2 \cap \Pi_2.$$

Work in progress

- $\mathrm{TL}[\mathsf{X}_a, \mathsf{Y}_a]$, rankers, quantifier alternation within FO^2.
- Generalizing the topologies.
- ω-semigroups.
- Extension to Mazurkiewicz traces.

Open problems

- Topologies for dot-depth hierarchy over infinite words.
- Fragments with successor.
- Languages weakly recognizable by monoids in **DA**.

Appendix: one more proof

Lemma

It is PSPACE*-hard to decide whether a regular language* $L \subseteq \Gamma^\omega$ *is closed.*

Proof.

- reduction of "$L(\mathcal{A}) = \Gamma^*$?" for some NFA $\mathcal{A}$
- new letter $c \notin \Gamma$, we can assume $\varepsilon \in L(\mathcal{A})$
- Büchi automaton $\mathcal{B}$ such that

$$L(\mathcal{B}) = \{w_1 c w_2 c \cdots \in (\Gamma \cup \{c\})^\omega \mid \exists i \colon w_i \in L(\mathcal{A})\}$$

- Now

$$\overline{L(\mathcal{B})} = (\Gamma^* c)^\omega$$

- Hence

$$L(\mathcal{A}) = \Gamma^* \;\Leftrightarrow\; L(\mathcal{B}) = \overline{L(\mathcal{B})} \;\Leftrightarrow\; L(\mathcal{B}) \text{ is closed}$$

□