Algebra Meets Logic: The Case of Regular Languages (With Applications to Circuit Complexity)

Denis Thérien

McGill University
The following are equivalent:

- **Description by regular languages**
  - Closure of finite subsets of $A^*$ under $\cup, \cdot, *$
  - Also closed under complement

- **Recognition by finite automata**
  - $A = (Q, A, \delta, q_0, F)$
  - $L(A) = \{w \in A^* : \delta(q_0, w) \in F\}$
Monoid: Set $M$ with binary associative operation and identity element.

$L \subseteq A^*$ is recognized by the monoid $M$ iff

- there is a morphism $\varphi : A^* \rightarrow M$, and
- there is set $F \subseteq M$ such that $L = \varphi^{-1}(F)$.

NB: For each language $L$ there is a canonical monoid $M(L)$ that recognizes it.

**Theorem:** $L$ is regular iff $M(L)$ is finite.
The model:

- variables stand for positions in words,
- $Q_a x$ is true in word $w$ iff the $x$th position of $w$ contains letter $a$,
- numerical predicates have their usual meaning, and
- if $\Theta$ is a sentence, $L(\Theta) = \{ w \in A^* : w \models \theta \}$.

**Theorem:** (Büchi) $L$ is regular iff $L = L(\Theta)$ for some $\Theta$ in $ESOM[+1]$.
Turing machines are the traditional model for space/time complexity.

**Theorem:** (Shepherdson) $L$ is regular iff it can be recognized by a TM operating in constant space.

Regular languages are also doable in linear time.

No more to say. BUT there are other models, e.g.:

- Boolean circuits
- Communication complexity
Restricting to First-Order

**Theorem:** *(McNaughton)* $L$ is star-free iff it is in $FO[<]$.

- $FO[+1] \subset FO[<]$.
- The theorem of McNaughton does not give a decision procedure to test if $L$ is in $FO[<]$.

**Theorem:** *(Schützenberger)* $L$ is star-free iff $M(L)$ is group-free.
Base Case: One Variable Only

Can do exactly boolean combinations of $\exists x \ [Q_ax]$. 

Let $S = \{0, 1\}$ be the 2-element semilattice:

\[
\begin{array}{c|cc}
S & 0 & 1 \\
\hline
0 & 0 & 1 \\
1 & 1 & 1 \\
\end{array}
\]

Let $\varphi : A^* \rightarrow S$, with $\varphi(a) = 1$ and $\varphi(b) = 0$. 
Then, $w \models \exists x \ [Q_ax]$ iff $\varphi(w) = 1$.

**Theorem:** (folklore) $L \in FO_1[<]$ iff $M(L)$ is a semilattice.
The General Case: Quantifier Depth \( k \)

- \( L \) is boolean combination of \( \psi = \exists x [\Theta(x)] \).

- This normalizes to \( \psi = \exists x [Q_a x \land \Theta_{left}(x) \land \Theta_{right}(x)] \).

- Now, \( w \models \psi \) iff there exists a factorization \( w = uav \) where \( u \models \Theta_{left} \) and \( v \models \Theta_{right} \).
\(L(\psi)\) is recognized by the monoid \(S \Box T\), where:

- \(S\) is a semilattice (to deal with the external quantifier), and
- \(T\) is the monoid constructed inductively (to deal with formulas of depth \(k - 1\)).

**Theorem:** (Krohn-Rhodes, Schützenberger) \(M\) is group-free iff \(M\) divides a block product of semilattices.
Allowing Modular Quantifiers

\[ w \models \exists c \mod q x \left[ \Theta(x) \right] \text{ iff the number of positions } x \text{ in } w \text{ that satisfy } \Theta(x) \text{ is congruent to } c \mod q. \]

Base case: one variable only

**Theorem:** *(folklore)* \( L \in \text{MOD}_1[<] \text{ iff } M(L) \text{ is an abelian group.} \)
**General Case: Quantifier Depth** $k$

$L$ is a boolean combination of $\psi = \exists^c \text{mod } q \ x [\Theta(x)]$.

$L(\psi)$ is recognized by the monoid $G \Box T$, where:

- $G$ is an abelian group (to deal with the external quantifier), and
- $T$ is the monoid constructed inductively (to deal with formulas of depth $k - 1$).

**Theorem:** (Jordan-Holder, STT) $L \in \text{MOD}[<] \iff M(L)$ is a solvable group.

$L \in \text{FO + MOD}[<] \iff M(L) \in \mathcal{M}_{\text{sol}}$ (i.e. every group in $M(L)$ is solvable).
**Boolean Circuits**

\[
\begin{align*}
AC^0: & \quad \text{constant depth, poly size, unbounded AND and OR gates.} \\
CC^0: & \quad \text{constant depth, poly size, unbounded MOD}_q \text{ gates.} \\
ACC^0: & \quad \text{constant depth, poly size, unbounded AND, OR, and MOD}_q \text{ gates.}
\end{align*}
\]

**Known:** \( AC^0 \subset ACC^0, \)  
\( \text{Parity} \notin AC^0. \)

**Conjectured:** \( CC^0 \subset ACC^0, \)  
\( \text{AND} \notin CC^0. \)
Semi-Obvious Results

- $FO[\text{arbitrary}] = AC^0$,
- $MOD[\text{arbitrary}] = CC^0$,
- $FO + MOD[\text{arbitrary}] = ACC^0$.

Other results:

- $FO[+, \times] = \text{logtime uniform } AC^0$.

What about: $FO[+]$, $MOD[+]$, and $FO + MOD[+]$?
Saving on Variables (by Reuse)

Trivial: \( FO_1[<] \subset FO[<] \), and
\( MOD_1[<] \subset MOD[<] \).

Theorem: (Immerman - Kozen) \( FO_3[<] = FO[<] \).

Theorem: (ST) \( MOD_2[<] = MOD[<] \), and
\( FO + MOD_3[<] = FO + MOD[<] \).

What about \( FO_2[<] \) and \( FO + MOD_2[<] \)?
$M$ is in the variety $\text{DA}$ iff $\forall e = e^2, s \in M$, $MeM = MsM \Rightarrow s = s^2$.

**Theorem:** (Schützenberger) $L$ is a finite union of languages of the form $L_0a_1L_1 \ldots a_sL_s$, where each $L_i$ is a commutative star-free language and the concatenation is unambiguous iff $M(L)$ is in $\text{DA}$.

**Theorem:** (TW) $L \in FO_2[<]$ iff $M(L)$ is in $\text{DA}$.
ex: $A^*ac^*aA^*$ is in $FO[<]$ but not in $FO_2[<]$.

ex: $(c^*ac^*bc^*)^*$ is in $FO[<]$ but not in $FO_2[<]$.

ex: $\{b, c, d\}^*bd^*aA^*$ is in $FO_2[<]$.

$$
\exists x[Q_ax \land \forall y[y < x \Rightarrow \neg Q_ax] \land \exists y[y < x \land Q_by] \land \\
\forall x[y < x \rightarrow Q_dx \lor \exists y[y < x \land Q_ax]]]]
$$
Theorem: (ST) \( L \in FO_2[<] \iff M(L) \) divides a block product of semilattices bracketed from left to right.

The variety \( A \) of group-free monoids is the smallest class such that:

- every semilattice is in \( A \), and
- if \( M \in A \) and \( S \) is a semilattice, then \( S \uplus M \) is in \( A \).

The variety \( DA \) is the smallest class such that:

- every semilattice is in \( DA \), and
- if \( M \in DA \) and \( S \) is a semilattice then \( M \boxplus S \) is in \( DA \).
Theorem: (ST) \( L \in FO + MOD_2[<] \) iff \( M(L) \) divides \( M \triangleleft G \) where \( M \) is in DA and \( G \) is a solvable group.

Crucial step:

Any formula in \( FO + MOD_2[<] \) is equivalent to a formula in \( FO + MOD_2[<] \) such that no existential or universal quantifier appears in the scope of a modular quantifier.

(i.e. we can always push the modular quantifiers inside)
\( V = M_{\text{sol}} \) is the smallest variety such that:

- every commutative monoid is in \( V \), and
- if \( M \in V \) and \( C \) is a commutative monoid, then \( C \sqcap M \) is in \( V \).

\( V = DA \sqcap G_{\text{sol}} \) is the smallest variety such that:

- every commutative monoid is in \( V \), and
- \( M \) is in \( V \) and \( C \) is a commutative monoid, then \( M \sqcap C \) is in \( V \).
ex:

- $A^*ac^*aA^*$ is in $FO[<]$ but not in $FO + MOD_2[<]$. 

- $(c^*ac^*bc^*)^*$ is in $FO[<]$, not in $FO_2[<]$ but it is in $FO + MOD_2[<]$. 
Theorem: (Lautemann-T) \( L \in FO_2[\text{arbitrary}] \) iff \( L \) can be recognized by an \( AC^0 \) circuit with \( O(n) \) gates.

Theorem: (Lautemann-T) \( L \in FO + MOD_2[\text{arbitrary}] \) iff \( L \) can be recognized by an \( ACC^0 \) circuit with \( O(n) \) gates.
Conjectured:  
- Any $ACC^0$ circuit for $A^*ac^*aA^*$ requires a superlinear number of gates.  
- Any $AC^0$ circuit for $(c^*ac^*bc^*)^*$ requires a superlinear number of gates.  
- A regular language $L$ (with a neutral letter) can be decided by an $AC^0$ circuit with $O(n)$ gates iff $M(L)$ is in DA.

**Theorem:** *(TT)* If the product in $M$ can be computed in $ACC^0$ with $O(n)$ gates then so can the product of $M \square T$, where $T$ is a commutative monoid.
Theorem: (KPT) A regular language (with neutral letter) \( L \) can be recognized by an \( AC^0 \) circuit with \( O(n) \) wires iff \( M(L) \) is in \( DA \).

A regular language (with neutral letter) \( L \) can be recognized by an \( ACC^0 \)-circuit with \( O(n) \) wires iff \( M(L) \) is orthodox and all groups in \( M(L) \) are commutative.
Theorem:  

(\text{TT}) \quad L \text{ can be denoted by a 2-variable formula in which no modular quantifier appears in the scope of another quantifier iff } M(L) \text{ is orthodox and all groups in } M(L) \text{ are commutative.}

A^*ac^*aA^* \text{ is not doable with } O(n) \text{ wires.}

(c^*ac^*bc^*)^* \text{ is not doable with } O(n) \text{ wires.}
Key step for the lower bound: Adapt results on superconcentrators to our situation.

Key step for the upper bound: If multiplication in $M$ can be done in constant-depth with $O(n)$ wires, then so can be the multiplication in $M \square S$ for any semilattices $S$.

Idea: On input $x_1, \ldots x_n$ the circuit needs to compute $OR(x_1, \ldots x_i)$ and $OR(x_i, \ldots x_n)$ for each $i$. This is possible with $O(n)$ wires!
The algebra and the logic of regular languages are deeply intertwined.

The results are often non-trivial and always elegant.

The results give solid intuition on what to expect e.g. for Boolean circuits. Of course, jacking up theorems from automata land to circuit land is a big challenge...