Reversible Regular Languages and $\star$-Semigroups

Paul Gastin* Amaldev Manuel* R Govind*

LSV, ENS Paris-Saclay, CNRS
Indian Institute of Technology, Goa
LaBRI, University of Bordeaux
Chennai Mathematical Institute

7 August 2019
DLT 2019

Reversible regular languages

Reverse Operation (r)

If $w = a_1 a_2 \cdots a_n, \quad w^r = a_n \cdots a_2 a_1$

Reverse operation is an involution, $i.e., (w^r)^r = w$

L^r is the reverse operation extended to languages

> L is a reversible language if $L = L^r$
> e.g. $(abc)^* + (cba)^*$

Regular languages	Reversible regular languages
Directed words	Undirected words
$a \to b \to b \to a \to b$	$a - b - b - a - b$

Our Goals

Good logical characterizations for reversible regular languages and its well-behaved subclasses

Effective decision procedures for these logics

Classical results (on directed words)

$L \in \mathrm{MSO}(<) = \mathrm{MSO}(+1)$

L is regular

$M(L)$ is finite

$L \in \mathrm{FO}(<)$

L is star-free

$M(L)$ is finite and aperiodic

$L \in \mathrm{FO}(+1)$

L is locally threshold testable

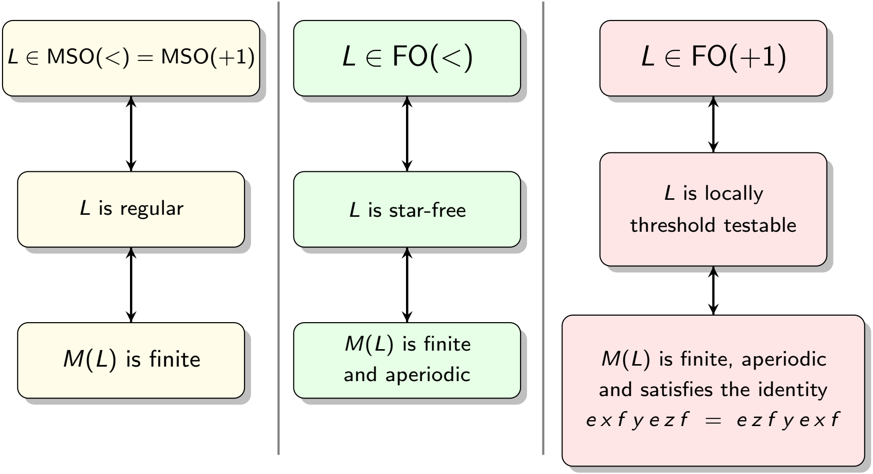
$M(L)$ is finite, aperiodic and satisfies the identity

$$e\,x\,f\,y\,e\,z\,f \;=\; e\,z\,f\,y\,e\,x\,f$$

Schützenberger'65, McNaughton-Papert'71

Brzozowski-Simon'71, Beauquier-Pin'91

Classical results (on directed words)



The monoid characterization gives the decision procedures for these logics

Schützenberger'65, McNaughton-Papert'71

Brzozowski-Simon'71, Beauquier-Pin'91

Our Predicates

Regular languages
Directed words

$a \to b \to b \to a \to b$

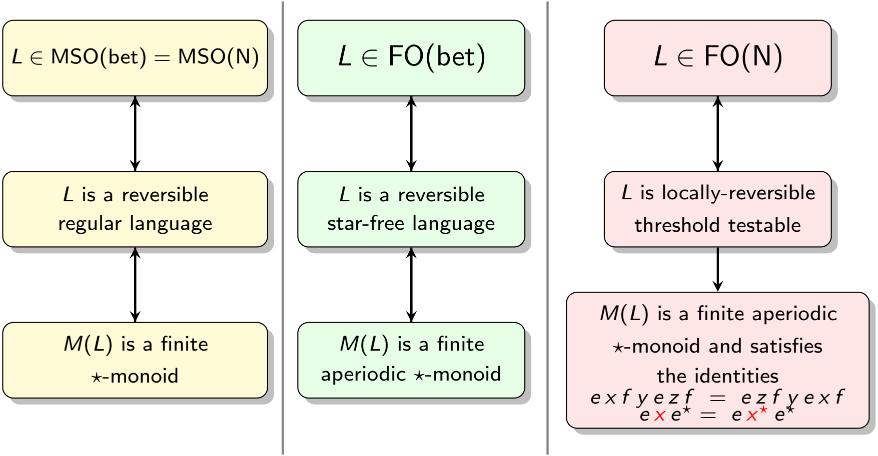
Reversible regular languages
Undirected words

$a - b - b - a - b$

For undirected words, we introduce analogues of successor relation and order relation:

- **Neighbour** - $\mathrm{N}(x, y)$ is true when x and y are neighbours
- **Between** - $\mathrm{bet}(x, y, z)$ is true when y is in between x and z

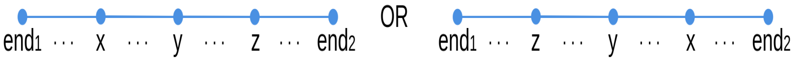
Our Results



$L \in \mathrm{MSO}(\mathrm{bet}) = \mathrm{MSO}(\mathrm{N})$
L is a reversible regular language
$M(L)$ is a finite $\star$-monoid
$L \in \mathrm{FO}(\mathrm{bet})$
L is a reversible star-free language
$M(L)$ is a finite aperiodic $\star$-monoid
$L \in \mathrm{FO}(\mathrm{N})$
L is locally-reversible threshold testable
$M(L)$ is a finite aperiodic $\star$-monoid and satisfies the identities
$e\,x\,f\,y\,e\,z\,f = e\,z\,f\,y\,e\,x\,f$
$e\,x\,e^\star = e\,x^\star\,e^\star$

Between predicate[1]

$$\text{bet}(x, y, z) \quad := \quad x < y < z \text{ or } z < y < x.$$



[1] A variant of the between predicate was studied by Andreas Krebs, Kamal Lodaya, Paritosh Pandya, Howard Straubing (2016)

Between predicate: Examples

Occurrence of subword or its reverse

- Contain the subword "*abc*" or "*cba*"

$$\exists x \exists y \exists z \ \mathrm{bet}(x, y, z) \wedge a(x) \wedge b(y) \wedge c(z)$$

Between predicate: Examples

Occurrence of subword or its reverse

- Contain the subword "*abc*" or "*cba*"

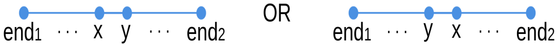
$$\exists x \exists y \exists z \ \mathrm{bet}(x, y, z) \wedge a(x) \wedge b(y) \wedge c(z)$$

- Contain the subword $a_1 a_2 \cdots a_n$ or $a_n a_{n-1} \cdots a_2 a_1$.

$$\exists x_1 \exists x_2 \cdots \exists x_n \ \bigwedge_{i=1}^{n} a_i(x_i) \ \wedge \ \bigwedge_{i=2}^{n-1} \mathrm{bet}(x_{i-1}, x_i, x_{i+1})$$

Neighbour predicate

$$N(x, y) \quad := \quad x + 1 = y \text{ or } y + 1 = x.$$



Neighbour predicate: Examples

Occurrence of a factor or its reverse

- "*ab*" or "*ba*" occurs as factor

$$\exists x \exists y \; \mathsf{N}(x, y) \wedge a(x) \wedge b(y)$$

Neighbour predicate: Examples

Occurrence of a factor or its reverse

- "*ab*" or "*ba*" occurs as factor

$$\exists x \exists y \; \mathrm{N}(x, y) \wedge a(x) \wedge b(y)$$

- Contain the factor $a_1 a_2 \cdots a_n$ or $a_n a_{n-1} \cdots a_1$

$$\exists x_1 x_2 \cdots x_n \bigwedge_{i=1}^{n} a_i(x_i) \wedge \bigwedge_{i=1}^{n-1} \mathrm{N}(x_i, x_{i+1}) \wedge \bigwedge_{i=2}^{n-1} (x_{i-1} \neq x_{i+1})$$

Neighbour predicate: Examples

Occurrence of a factor or its reverse

- "*ab*" or "*ba*" occurs as factor

$$\exists x \exists y\ \mathrm{N}(x, y) \wedge a(x) \wedge b(y)$$

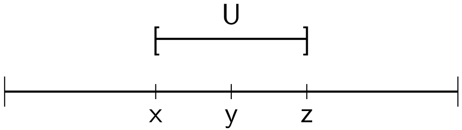
- Contain the factor $a_1 a_2 \cdots a_n$ or $a_n a_{n-1} \cdots a_1$

$$\exists x_1 x_2 \cdots x_n \bigwedge_{i=1}^{n} a_i(x_i) \wedge \bigwedge_{i=1}^{n-1} \mathrm{N}(x_i, x_{i+1}) \wedge \bigwedge_{i=2}^{n-1} (x_{i-1} \neq x_{i+1})$$

Cannot express subword relation in FO using the 'Neighbour' predicate

MSO(bet) = MSO(N)

'Between' predicate can be defined using 'Neighbour' predicate (and second order quantification).



Any subset U of positions that satisfies the conditions (1) and (2), contains y

1. U contains x, z and some other position.
2. any position in U, except for x and z has exactly two neighbours in U.

MSO(bet) = MSO(N)

'Neighbour' predicate can be defined in terms of the 'Between' predicate

$$\mathsf{N}(x, y) \;\equiv\; (x \neq y) \wedge \forall z\, \neg \mathsf{bet}(x, z, y)$$

- 'Neighbour' using 'Between' - can be expressed using this first order macro
- 'Between' using 'Neighbour' - requires second order quantification

MSO(bet) = MSO(N) = *Rev-Reg*

Theorem

The following are equivalent:

1. *L is a reversible regular language*
2. *L is definable in* MSO(bet)
3. *L is definable in* MSO(N)

Proof sketch

(1) $\implies$ (2)

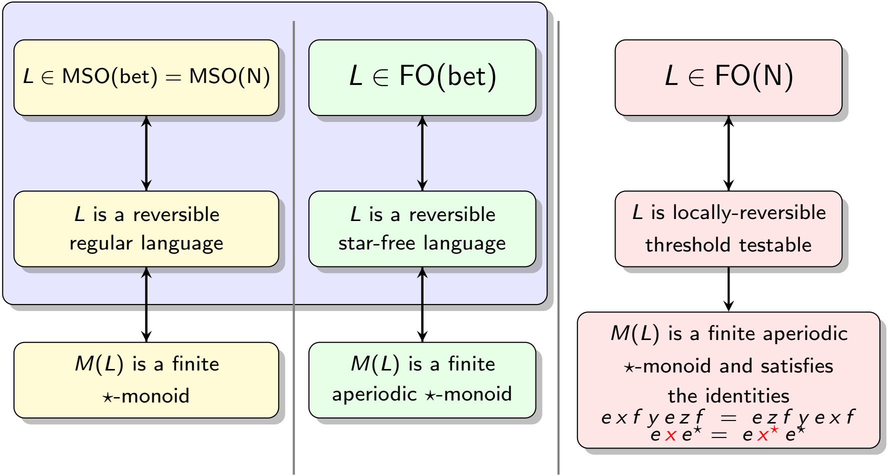
Let $\varphi \in$ MSO($<$) defining L.

$$\chi = \exists e \ (\psi(e) \wedge \varphi'(e))$$

$\psi(e)$ says that e is an endpoint: $\neg\exists x, y \ \mathrm{bet}(x, e, y)$

φ' is φ with $x < y$ replaced by $(e = x \neq y) \vee \mathrm{bet}(e, x, y)$

Our Results


$L \in \mathrm{MSO(bet)} = \mathrm{MSO(N)}$
$L \in \mathrm{FO(bet)}$
$L \in \mathrm{FO(N)}$
L is a reversible regular language
L is a reversible star-free language
L is locally-reversible threshold testable
$M(L)$ is a finite $\star$-monoid
$M(L)$ is a finite aperiodic $\star$-monoid
$M(L)$ is a finite aperiodic $\star$-monoid and satisfies the identities
$e\,x\,f\,y\,e\,z\,f = e\,z\,f\,y\,e\,x\,f$
$e\,x\,e^\star = e\,x^\star\,e^\star$

FO(+1) = Locally Threshold Testable languages

$u \approx_k^t v$ **equivalence**

- same prefix of length $k-1$.
- same suffix of length $k-1$.
- same number of factors of size k up to threshold t.

Example: $abababab \approx_2^2 ababab$

FO(+1) = Locally Threshold Testable languages

$u \approx_k^t v$ **equivalence**

- same prefix of length $k-1$.
- same suffix of length $k-1$.
- same number of factors of size k up to threshold t.

Example: $abababab \approx_2^2 ababab \not\approx_2^2 abbab$

FO(+1) = Locally Threshold Testable languages

$u \approx_k^t v$ **equivalence**

- same prefix of length $k-1$.
- same suffix of length $k-1$.
- same number of factors of size k up to threshold t.

Example: $abababab \approx_2^2 ababab \not\approx_2^2 abbab$
$abababab \not\approx_2^3 ababab$

FO(+1) = Locally Threshold Testable languages

$u \approx_k^t v$ **equivalence**

- same prefix of length $k-1$.
- same suffix of length $k-1$.
- same number of factors of size k up to threshold t.

Example: $abababab \approx_2^2 ababab \not\approx_2^2 abbab$
$abababab \not\approx_2^3 ababab$

L is locally threshold testable if L is a union of $\approx_k^t$ classes, for some $t, k > 0$.

Examples:

- Locally Threshold Testable language: $(ab)^*$
- Not locally threshold testable : $c^* \, a \, c^* \, b \, c^*$

Is FO(N) = FO(+1) ∩ *Rev*?

Clearly, FO(N) ⊆ FO(+1) ∩ *Rev*
But this inclusion is strict.

Example:

$$L = \{w \mid \#(ab) = 2 \text{ , } \#(ba) = 1 \text{ OR } \#(ab) = 1 \text{ , } \#(ba) = 2 \text{ }\}$$

$L \in$ FO(+1) ∩ *Rev*, but $L \notin$ FO(N)

$c^k\, ab\, c^k\, ba\, c^k\, ab\, c^k$ $\qquad$ $c^k\, ab\, c^k\, ab\, c^k\, ab\, c^k$

FO(N) = Locally-Reversible Threshold Testable Languages

$u \overset{r}{\approx}{}^{t}_{k} v$ **equivalence**

- same undirected ends of size $k-1$ $\{\mathit{prefix}, \mathit{suffix}^r\}$
- same undirected factors of size k up to threshold t

Examples:

- $abbabab \overset{r}{\approx}{}^{2}_{2} babba$
- $aba \overset{r}{\not\approx}{}^{1}_{2} baa$

FO(N) = Locally-Reversible Threshold Testable Languages

$u \overset{r}{\approx}{}_k^t v$ **equivalence**

- same undirected ends of size $k-1$ $\{prefix, suffix^r\}$
- same undirected factors of size k up to threshold t

Examples:

- $abbabab \overset{r}{\approx}{}_2^2 babba$
- $aba \overset{r}{\not\approx}{}_2^1 baa$

L is locally-reversible threshold testable if L is a union of $\overset{r}{\approx}{}_k^t$ classes, for some $t, k > 0$.

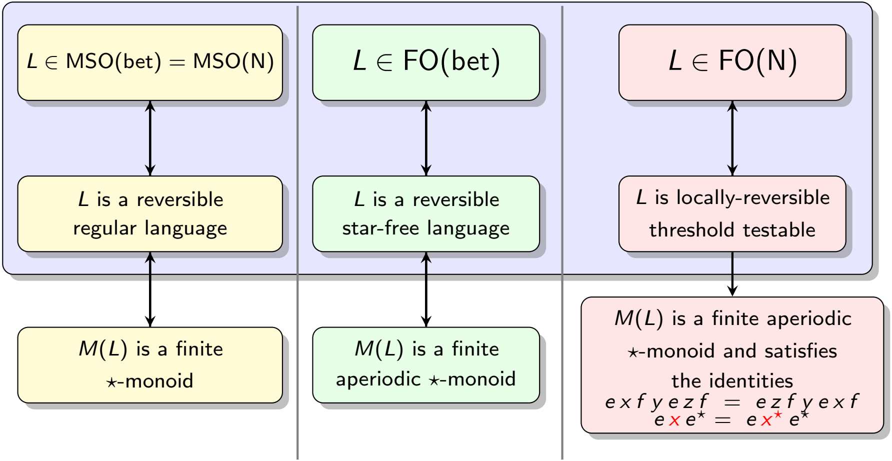
- $L = \{w \mid ab$ or ba occurs as factor 3 times in $w\}$ is LRTT
- A non-example

FO(N) = Locally-reversible Threshold Testable Languages

Proof sketch
($\Leftarrow$) Since L is a union of $\overset{r}{\approx}{}_k^t$-classes, we write an FO(N) formula for each $\overset{r}{\approx}{}_k^t$-class.
Recall : occurrence of a factor or it's reverse $\rightarrow$ expressible in FO(N).
Similarly, we can say that x or x^r occurs at least m times in w.

($\Rightarrow$) Hanf's theorem

Our Results


$L \in \mathrm{MSO}(\mathrm{bet}) = \mathrm{MSO}(\mathrm{N})$
L is a reversible regular language
$M(L)$ is a finite $\star$-monoid
$L \in \mathrm{FO}(\mathrm{bet})$
L is a reversible star-free language
$M(L)$ is a finite aperiodic $\star$-monoid
$L \in \mathrm{FO}(\mathrm{N})$
L is locally-reversible threshold testable
$M(L)$ is a finite aperiodic $\star$-monoid and satisfies the identities
$e\,x\,f\,y\,e\,z\,f = e\,z\,f\,y\,e\,x\,f$
$e\,x\,e^\star = e\,x^\star\,e^\star$

Are there effective decision procedures for these logics?

Given a regular language L, is it decidable
if L is definable in the logic?

Are there effective decision procedures for these logics?

Given a regular language L, is it decidable
if L is definable in the logic?

$\mathrm{MSO}(\mathrm{bet}) = \mathrm{MSO}(\mathrm{N}) = \mathrm{MSO}(<) \cap \mathit{Rev}$ → Check if L is reversible

$\mathrm{FO}(\mathrm{bet}) = \mathrm{FO}(<) \cap \mathit{Rev}$ → Check if L is star-free
Check if L is reversible

$\mathrm{FO}(\mathrm{N})$ → Not yet known

$\star$-semigroup or Semigroup with involution

A $\star$-semigroup is a triple $(S, \cdot, \star)$, where

- $(S, \cdot)$ is a semigroup
- $\star : S \to S$ is an involution on S, i.e, $\forall x \in S$,

$$(x^\star)^\star = x$$

- $\star$ is an anti-automorphism on S, i.e, $\forall x, y \in S$,

$$(x \cdot y)^\star = y^\star \cdot x^\star$$

Example: Free monoid A^*
The reverse operation is an involution on the free monoid that is an anti-automorphism, since

- $(w^r)^r = w$
- $(w_1 \cdot w_2)^r = w_2^r \cdot w_1^r$

Acceptance by $\star$-semigroups

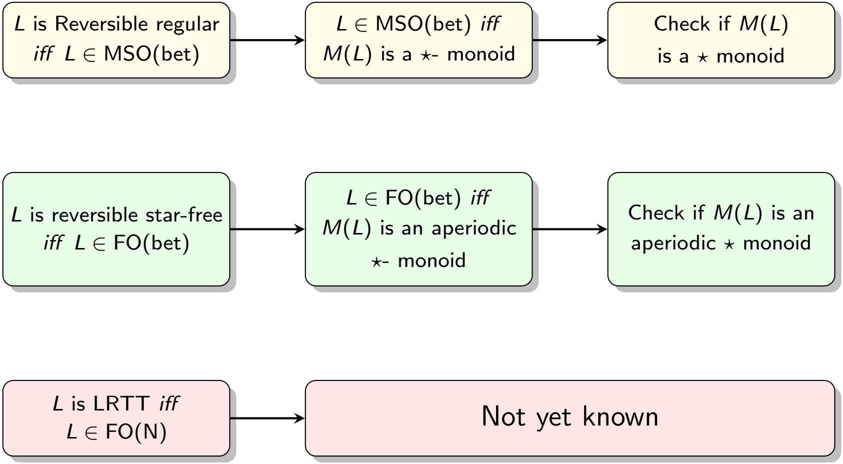
A language $L \subseteq \Sigma^*$ is said to be recognized by a $\star$-semigroup $(S, \cdot, \star)$, if there is a morphism $\phi : \Sigma^* \rightarrow S$ and a set $P \subseteq S$, such that the following conditions are satisfied:

1. $L = \phi^{-1}(P)$
2. ϕ is a $\star$-semigroup morphism *i.e*,

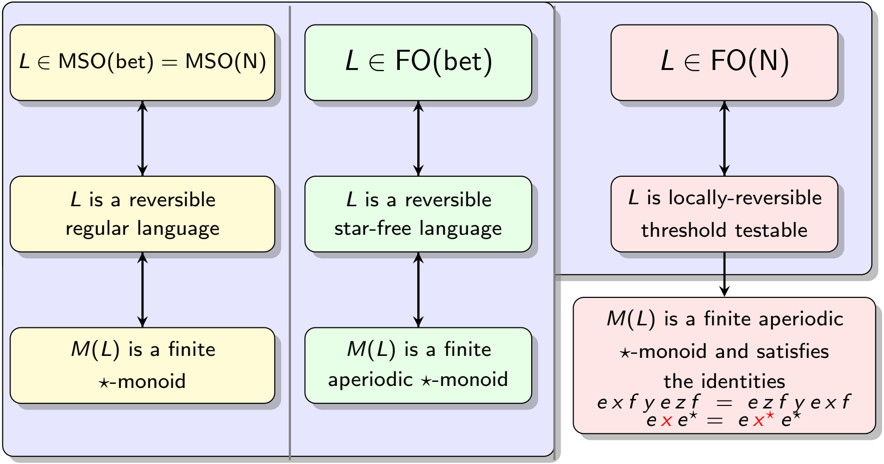
$$\phi(x \cdot y) = \phi(x) \cdot \phi(y) \text{ and } \phi(x^r) = (\phi(x))^\star$$

3. $P^\star = P$

Alternate decision procedures for MSO(*bet*) and FO(bet)



Our Results


$L \in \mathrm{MSO(bet)} = \mathrm{MSO(N)}$
L is a reversible regular language
$M(L)$ is a finite $\star$-monoid
$L \in \mathrm{FO(bet)}$
L is a reversible star-free language
$M(L)$ is a finite aperiodic $\star$-monoid
$L \in \mathrm{FO(N)}$
L is locally-reversible threshold testable
$M(L)$ is a finite aperiodic $\star$-monoid and satisfies the identities
$e\,x\,f\,y\,e\,z\,f = e\,z\,f\,y\,e\,x\,f$
$e\,x\,e^\star = e\,x^\star\,e^\star$

Identity for FO(+1) languages

Theorem (Brzozowski-Simon 1973, Beauquier-Pin 1991)

The following are equivalent:

1. L is locally threshold testable.
2. $L \in \mathrm{FO}(+1)$.
3. The syntactic semigroup of L is finite and aperiodic and satisfies the identity, for all $e, f, x, y, z \in M(L)$ with e, f idempotents,

$$e\,x\,f\,y\,e\,z\,f \;=\; e\,z\,f\,y\,e\,x\,f$$

Identity for FO(N) languages

The syntactic $\star$-semigroup of a $FO(N)$-definable language satisifes the identities, for any elements e, f, x, y, z of the semigroup with e, f idempotents,

$$e\,x\,f\,y\,e\,z\,f \;=\; e\,z\,f\,y\,e\,x\,f$$

$$e\,x\,e^\star = \; e\,x^\star\,e^\star$$

Proof sketch

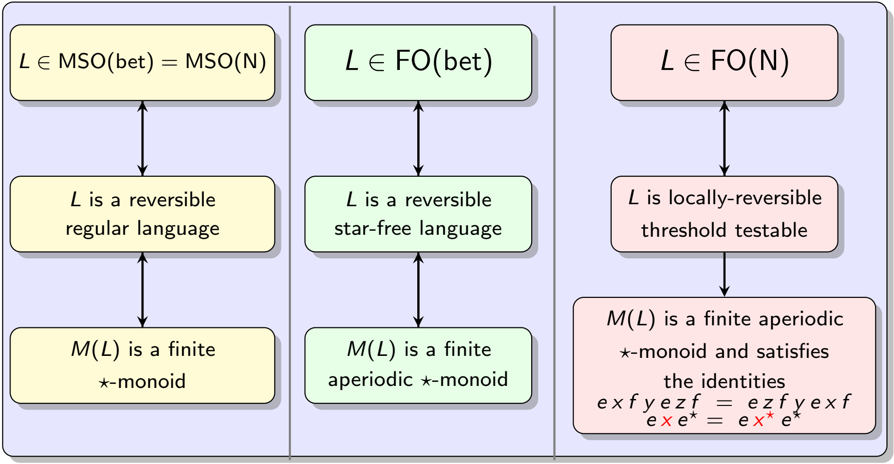
Let $h\colon \Sigma^+ \mapsto M = (\Sigma^+/\sim_L, \cdot, \star)$ be the canonical morphism recognising L.
Let $u, s \in \Sigma^+$ s.t. $h(u) = e$ and $h(s) = x \implies h(u^r) = e^\star$ and $h(s^r) = x^\star$
Let $w = (u^k)s(u^k)^r$ and $w^r = (u^k)s^r(u^k)^r$
$h(w) = h(usu^r) = exe^\star$ and $h(w^r) = h(us^r u^r) = ex^\star e^\star$
For all contexts $\alpha, \beta \in \Sigma^*$, we can show that $\alpha w \beta \overset{r}{\approx}{}_k^t\, \alpha w^r \beta$,

$$\alpha w \beta \in L \text{ iff } \alpha w^r \beta \in L$$

$$h(w) = h(w^r)$$

$$exe^\star = ex^\star e^\star$$

Our Results


$L \in \mathrm{MSO(bet)} = \mathrm{MSO(N)}$
L is a reversible regular language
$M(L)$ is a finite $\star$-monoid
$L \in \mathrm{FO(bet)}$
L is a reversible star-free language
$M(L)$ is a finite aperiodic $\star$-monoid
$L \in \mathrm{FO(N)}$
L is locally-reversible threshold testable
$M(L)$ is a finite aperiodic $\star$-monoid and satisfies the identities
$e\,x\,f\,y\,e\,z\,f = e\,z\,f\,y\,e\,x\,f$
$e\,x\,e^\star = e\,x^\star\,e^\star$

Conclusion

Reversible regular languages and their well-defined subclasses are characterized by logics using the 'between' and 'neighbour' predicates

MSO(bet), MSO(N) and FO(bet) behave like MSO($<$), MSO($+1$) and FO($<$), respectively.

FO(N) corresponds to the class locally-reversible threshold testable languages.

$\star$-semigroups are the algebraic structures that recognize reversible regular languages.