



HAL
open science

Reversible Regular Languages: Logical and Algebraic Characterisations

Paul Gustin, Amaldev Manuel, R Govind

► **To cite this version:**

Paul Gustin, Amaldev Manuel, R Govind. Reversible Regular Languages: Logical and Algebraic Characterisations. *Fundamenta Informaticae*, 2021, 180, pp.333 - 350. 10.3233/fi-2021-2045 . hal-03710241

HAL Id: hal-03710241

<https://hal.science/hal-03710241v1>

Submitted on 7 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Reversible Regular Languages: Logical and 2 Algebraic Characterisations

3 Paul Gastin ✉ 

4 Université Paris-Saclay, ENS Paris-Saclay, CNRS, LMF, 91190, Gif-sur-Yvette, France
5 CNRS, ReLaX, IRL 2000, Siruseri, India

6 Amaldev Manuel ✉

7 Indian Institute of Technology Goa, India

8 R. Govind ✉ 

9 Chennai Mathematical Institute, Chennai, India

10 LaBRI, University of Bordeaux, France

11 — Abstract —

12 We present first-order (FO) and monadic second-order (MSO) logics with predicates ‘between’ and
13 ‘neighbour’ that characterise the class of regular languages that are closed under the reverse operation
14 and its subclasses. The ternary between predicate $\text{bet}(x, y, z)$ is true if the position y is strictly
15 between the positions x and z . The binary neighbour predicate $\text{N}(x, y)$ is true when the the positions
16 x and y are adjacent. It is shown that the class of reversible regular languages is precisely the class
17 definable in the logics $\text{MSO}(\text{bet})$ and $\text{MSO}(\text{N})$. Moreover the class is definable by their existential
18 fragments $\text{EMSO}(\text{bet})$ and $\text{EMSO}(\text{N})$, yielding a normal form for MSO formulas. In the first-order
19 case, the logic $\text{FO}(\text{bet})$ corresponds precisely to the class of reversible languages definable in $\text{FO}(<)$.
20 Every formula in $\text{FO}(\text{bet})$ is equivalent to one that uses at most 3 variables. However the logic
21 $\text{FO}(\text{N})$ defines only a strict subset of reversible languages definable in $\text{FO}(+1)$. A language-theoretic
22 characterisation of the class of languages definable in $\text{FO}(\text{N})$, called locally-reversible threshold-
23 testable (LRTT), is given. In the second part of the paper we show that the standard connections
24 that exist between MSO and FO logics with order and successor predicates and varieties of finite
25 semigroups extend to the new setting with the semigroups extended with an involution operation
26 on its elements. The case is different for $\text{FO}(\text{N})$ where we show that one needs an additional
27 equation that uses the involution operator to characterise the class. While the general problem of
28 characterising $\text{FO}(\text{N})$ is open, an equational characterisation is shown for the case of neutral letter
29 languages.

30 **2012 ACM Subject Classification** Theory of computation → Formal languages and automata theory;
31 Theory of computation → Logic

32 **Keywords and phrases** Regular languages, reversible languages, first-order logic, automata, semig-
33 rroups

34 **1** Introduction

35 In this work we look closely at the class of regular languages that are closed under the
36 reverse operation. We fix a finite alphabet A for the rest of our discussion. The set A^*
37 (respectively A^+) denotes the set of all (resp. non-empty) finite words over the alphabet A .
38 If $w = a_1 \cdots a_k$ with $a_i \in A$ is a word then $w^r = a_k \cdots a_1$ denotes the reverse of w . This
39 notion is extended to sets of words pointwise, i.e., $L^r = \{w^r \mid w \in L\}$ and we can talk about
40 reverse of languages. A regular language $L \subseteq A^*$ is *closed under reverse* or simply *reversible*
41 if $L^r = L$. We let Rev denote the class of all reversible regular languages. Clearly Rev is a
42 strict subset of the class of all regular languages.

43 One way to look at a reversible language is as a collection of *undirected words*. When
44 seen as first-order structures, words are directed graphs with directed edges that constitute a
45 linear ordering on positions. If we forgo the direction then the resulting undirected graph

46 can be read either way and hence will correspond to both the word and its reverse. Hence a
 47 set of undirected words can be equated with a reversible language and by extension the class
 48 of undirected languages can be equated with Rev .

49 The class Rev is easily verified to be closed under union, intersection and complementa-
 50 tion. It is also closed under homomorphic images, and inverse homomorphic images under
 51 alphabetic (i.e., length preserving) morphisms. However it is not closed under quotients
 52 of the form $a^{-1}L := \{v \mid av \in L\}$, where a is a letter and L is a reversible language over
 53 A . For instance, the language $L = (abc)^* + (cba)^*$ is closed under reverse but the quotient
 54 $a^{-1}L = bc(abc)^*$ is not closed under reverse. Thus the class Rev fails to be a *variety* of
 55 languages — i.e., a class closed under Boolean operations, inverse homomorphic images
 56 and quotients. However reversible languages are closed under bidirectional quotients, i.e.,
 57 quotients of the form $u^{-1}Lv^{-1} \cup (v^r)^{-1}L(u^r)^{-1}$, given words u, v . Thus, to a good extent,
 58 Rev shares properties similar to that of regular languages. Hence it makes sense to ask the
 59 question

60 “are there good logical characterisations for the class Rev and its well behaved sub-
 61 classes?”.

62 **Our results.** We suggest a positive answer to the above question. We introduce two
 63 predicates *between* ($\text{bet}(x, y, z)$ is true if the position y is strictly between the positions x
 64 and z) and *neighbour* ($\text{N}(x, y)$ is true if the positions x and y are adjacent). The predicates
 65 *between* and *neighbour* are the natural analogues of the order relation $<$ and successor relation
 66 $+1$ in the undirected case. In fact this analogy extends to the case of logical definability. We
 67 show that Rev is the class of monadic second order (MSO) definable languages using either
 68 of the predicates, i.e., $\text{MSO}(\text{bet})$ or $\text{MSO}(\text{N})$. This is analogous to the classical Büchi-Elgot-
 69 Trakhtenbrot theorem relating regular languages and the MSO logic. Moreover, as in the
 70 Büchi-Elgot-Trakhtenbrot theorem Rev is definable in the existential MSO logics $\text{EMSO}(\text{bet})$
 71 and $\text{EMSO}(\text{N})$.

72 The above analogy extends to the case of first order logic as well. We show that $\text{FO}(\text{bet})$
 73 definable languages are precisely the reversible languages definable in $\text{FO}(<)$. Also, every
 74 formula in $\text{FO}(\text{bet})$ is equivalent to one that uses at most 3 variables.

75 However the case of FO with the neighbour relation is different. It turns out that the
 76 class of $\text{FO}(\text{N})$ definable languages is a strict subset of those reversible languages definable
 77 in $\text{FO}(+1)$. The precise characterisation of this class is one of our main contributions. A
 78 classical result on $\text{FO}(+1)$ -definable languages [1] states that a language is $\text{FO}(+1)$ definable
 79 if and only if it is a union of classes of an equivalence relation \approx_k^t for some $k, t \in \mathbb{N}$, whereby
 80 two words are \approx_k^t -equivalent if they have identical prefixes and suffixes of length $k - 1$ and
 81 have the same subwords of length k upto threshold t (see Definition 7). For characterising
 82 $\text{FO}(\text{N})$ -definable languages one needs an equivalence coarser than \approx_k^t . We say two words are
 83 \approx_k^r -equivalent if they have the same prefixes and suffixes upto reverse and have the same
 84 subwords of length k upto reverse and upto threshold t (see Definition 13). It is shown that
 85 a language is definable in $\text{FO}(\text{N})$ if and only if it is a union of equivalence classes of \approx_k^r
 86 for some $k, t \in \mathbb{N}$.

87 The immediate question that arises from the above characterisations is one of definability
 88 in a logic: *Given a reversible language is it definable in the logic?*. The case of $\text{FO}(\text{bet})$ is
 89 decidable due to Schützenberger-McNaughton-Papert theorem that states that syntactic
 90 monoids of $\text{FO}(<)$ definable languages are aperiodic (equivalent to the condition that the
 91 monoid contains no groups as subsemigroups) [2, 3]. However in the case of $\text{FO}(\text{N})$ one
 92 needs to consider additional restrictions on the syntactic semigroups apart from those needed

93 to characterise $\text{FO}(+1)$. This is done by means of an additional involution operation (an
 94 involution \star is a unary operation satisfying the laws $a^{\star\star} = a$ and $(ab)^{\star} = b^{\star}a^{\star}$). It is shown
 95 that syntactic semigroups of languages definable in $\text{FO}(\mathbb{N})$ satisfies the equation $exe^{\star} = ex^{\star}e^{\star}$
 96 where x, e are elements the semigroup and e is furthermore an idempotent. The converse
 97 direction is open in the general case. But we prove it in the restricted case of neutral letter
 98 languages. It is to be noted that the characterisation of $\text{FO}(+1)$ is a tedious one that goes
 99 via categories [4].

100 **Related work.** A different but related *between* predicate (namely $a(x, y)$, for $a \in A$, is true
 101 if there is an a -labelled position between positions x and y) was introduced and studied in
 102 [5, 6, 7]. Such a predicate is not definable in $\text{FO}^2(<)$, the two variable fragment of first-order
 103 logic (which corresponds to the well known semigroup variety DA [8]). The authors of [5, 6, 7]
 104 study the expressive power of $\text{FO}^2(<)$ enriched with the between predicates $a(x, y)$ for $a \in A$,
 105 and show an algebraic characterisation of the resulting family of languages. The between
 106 predicate (predicates rather) in [5] is strictly less expressive than the between predicate
 107 introduced in this paper. However the logics considered in [5] have the between predicates in
 108 conjunction with order predicates $<$ and $+1$. Hence their results are orthogonal to ours.

109 Another line of work that has close parallels with the one in this paper is the variety
 110 theory of involution semigroups (also called \star -semigroups) (see [9] for a survey). Most
 111 investigations along these lines have been on subvarieties of *regular* \star -semigroups (i.e., \star -
 112 semigroups satisfying the equation $xx^{\star}x = x$). As far as we are aware the equation introduced
 113 in this paper has not been studied before.

114 **Structure of the paper.** In Section 2 we introduce the predicates and present our logical
 115 characterisations. This is followed by a characterisation of $\text{FO}(\mathbb{N})$. In Section 3 we discuss
 116 semigroups with involution, a natural notion of syntactic semigroups for reversible languages.
 117 In Section 4 we conclude.

118 An extended abstract of this work appeared in [10].

119 **2** Logics with *Between* and *Neighbour*

120 As usual we represent a word $w = a_1 \cdots a_n$ as a structure containing positions $\{1, \dots, n\}$,
 121 and unary predicates P_a for each letter a in the alphabet. The predicate P_a is precisely true
 122 at those positions labelled by letter a . The atomic predicate $x < y$ (resp. $x + 1 = y$) is true if
 123 position y is after (resp. immediately after) position x . The logic FO is the logic containing
 124 atomic predicates, boolean combinations ($\phi \vee \psi, \phi \wedge \psi, \neg\psi$ whenever ϕ, ψ are formulas of the
 125 logic), and first order quantifications ($\exists x \psi, \forall x \psi$ if ψ is a formula of the logic). The logic
 126 MSO in addition contains second order quantification as well ($\exists X \psi, \forall X \psi$ if ψ is a formula
 127 of the logic) — i.e., quantification over sets of positions. By $\text{FO}(\tau)$ or $\text{MSO}(\tau)$ we mean the
 128 corresponding logic with atomic predicates τ in addition to the unary predicates P_a . The
 129 classical result relating MSO and regular languages states that $\text{MSO}(<) = \text{MSO}(+1)$ (in
 130 terms of expressiveness) defines all regular languages. We introduce two analogous predicates
 131 for the class Rev of reversible regular languages.

132 **2.1 MSO(bet), MSO(N) and FO(bet)**

The ternary *between* predicate $\text{bet}(x, y, z)$ is true for positions x, y, z when y is strictly in between x and z , i.e.,

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

► **Example 1.** The set of all words containing $a_1 a_2 \cdots a_k$ or $a_k a_{k-1} \cdots a_1$ as subword is defined by the formula

$$\exists x_1 \exists x_2 \cdots \exists x_k \bigwedge_{i=1}^k P_{a_i}(x_i) \wedge \bigwedge_{i=2}^{k-1} \text{bet}(x_{i-1}, x_i, x_{i+1}).$$

The ‘successor’ relation of bet is the binary predicate *neighbour* $\text{N}(x, y)$ that holds true when x and y are neighbours, i.e.

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

133 ► **Example 2.** A position in a word is an endpoint if it has exactly one neighbour. The
134 following formula defines endpoints.

$$135 \quad \varphi(x) := \forall y \forall z (\text{N}(x, y) \wedge \text{N}(x, z) \rightarrow y = z)$$

The set of words of even length is defined by the formula

$$\exists e_1 e_2 \exists X (\varphi(e_1) \wedge \varphi(e_2) \wedge X(e_1) \wedge \neg X(e_2) \wedge \forall x \forall y (\text{N}(x, y) \rightarrow (X(x) \leftrightarrow \neg X(y)))) .$$

136 The relation $\text{N}(x, y)$ can be defined in terms of bet using first-order quantifiers as $x \neq$
137 $y \wedge \forall z \neg \text{bet}(x, z, y)$. One can also define $\text{bet}(x, y, z)$ in terms of N , but using second-order set
138 quantification. To do this we assert that x, y, z are distinct positions and any subset X of
139 positions

140 ■ that contains x, z and at least some other position

141 ■ and such that any position in X , except for x and z , has exactly two neighbours in X ,
142 contains the position y .

143 ► **Proposition 3.** For definable languages, $\text{MSO}(\text{bet}) = \text{MSO}(\text{N}) = \text{Rev}$.

Proof. Clearly from the discussion above, $\text{MSO}(\text{bet}) = \text{MSO}(\text{N}) \subseteq \text{Rev}$. To show the other inclusion, let L be a reversible regular language and let φ be a formula in $\text{MSO}(<)$ defining it. Pick an endpoint e of the given word; an endpoint is a position with exactly one neighbour, a property expressible in $\text{FO}(\text{N}) \subseteq \text{FO}(\text{bet})$. We relativize the formula φ with respect to e by replacing all occurrences of $x < y$ in the formula by $(e = x \neq y) \vee \text{bet}(e, x, y)$. Let $\varphi'(e)$ be the formula obtained in this way and let $\psi(e) = \neg \exists x, y (x \neq y \wedge \text{N}(e, x) \wedge \text{N}(e, y))$ be the $\text{FO}(\text{N})$ formula asserting that e is an endpoint, then we claim that

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

144 defines the language L . Let w be a word of length $k \geq 1$ then,

$$\begin{aligned} 145 \quad w \models \chi &\Leftrightarrow w, 1 \models \varphi'(e) \text{ or } w, k \models \varphi'(e) \\ 146 \quad &\Leftrightarrow w \models \varphi \text{ or } w^r \models \varphi \\ 147 \quad &\Leftrightarrow w \models \varphi \text{ (since } L \text{ is reversible).} \end{aligned}$$

149 Hence $L(\chi) = L(\varphi) = L$. ◀

150 An MSO(τ) formula is in the *existential MSO fragment*, denoted as EMSO(τ), if it is of
 151 the form $\exists X_1 \cdots \exists X_n \varphi$ where φ is a first-order formula over τ . In the case of words every
 152 MSO($<$) as well as MSO(+1) formula is equivalent to a formula in EMSO. This extends to
 153 the case of EMSO(bet) and EMSO(N) as well.

154 ► **Proposition 4.** $\text{Rev} = \text{EMSO}(\text{bet}) = \text{EMSO}(\text{N})$.

155 **Proof.** Because of Proposition 3 it suffices to show that $\text{Rev} \subseteq \text{EMSO}(\text{bet})$ and $\text{Rev} \subseteq$
 156 $\text{EMSO}(\text{N})$ in terms of languages accepted.

157 ($\text{Rev} \subseteq \text{EMSO}(\text{bet})$)

We modify the proof of Proposition 3. We observe that in the proof the formula φ can
 be assumed to be in EMSO($<$). Therefore the formula $\varphi'(e)$ is in EMSO(bet). Let us assume
 $\varphi'(e) = \exists X_1 \cdots \exists X_n \varphi''(e)$ then

$$\chi = \exists e (\psi(e) \wedge \exists X_1 \cdots \exists X_n \varphi''(e)) \equiv \exists X_1 \cdots \exists X_n \exists e (\psi(e) \wedge \varphi''(e)) .$$

158 Hence $\chi \in \text{EMSO}(\text{bet})$ is a formula accepting the language L .

159 ($\text{Rev} \subseteq \text{EMSO}(\text{N})$)

160 Let L be a language in Rev and let $\chi = \exists X_1 \cdots \exists X_n \varphi$ be a formula in EMSO(+1)
 161 defining L such that $\varphi \in \text{FO}(\text{+1})$. Let $\psi(e)$ be a formula in FO(N) that expresses the
 162 following properties:

- 163 ■ Every position in the word is labelled with exactly one element from the set $\{0, 1, 2\}$
 164 indicated by the monadic predicates Y_0, Y_1, Y_2 .
- 165 ■ Position e is an endpoint that is labelled by 0 and its neighbour is labelled by 1.
- 166 ■ Let x, y, z be any three consecutive positions in the word such that x and z are the
 167 neighbours of y . Then x, y, z are labelled by $i, (i+1) \bmod 3, (i+2) \bmod 3$ or $(i+2) \bmod 3, (i+1) \bmod 3, i$
 168 in the respective order, for some $i \in \{0, 1, 2\}$.

169 Let χ' be the formula

$$170 \chi' = \exists Y_0 \exists Y_1 \exists Y_2 \exists X_1 \cdots \exists X_n (\varphi' \wedge \exists e \psi(e))$$

171 where φ' is obtained by replacing each occurrence of $x + 1 = y$ by the formula

$$172 \sigma(x, y) = \mathbf{N}(x, y) \wedge \bigvee_{i \in \{0, 1, 2\}} Y_i(x) \wedge Y_{(i+1) \bmod 3}(y).$$

We claim that L is recognised by χ' . Clearly if $w \models \chi$ then

$$w, Y_0 = \{1, 4, \dots\}, Y_1 = \{2, 5, \dots\}, Y_2 = \{3, 6, \dots\}, e = 1 \models \psi(e) \wedge \exists X_1 \cdots \exists X_n \varphi'.$$

173 Hence $w \models \chi'$.

174 Next we claim that if $w \models \chi'$ then $w \models \chi$. Assume $w \models \chi'$ and it has length n . The only
 175 interpretations for the predicates Y_0, Y_1, Y_2 that satisfy $\psi(e)$ are either $\{1, 4, \dots\}, \{2, 5, \dots\},$
 176 $\{3, 6, \dots\}$ (when $e = 1$) or $\{n, n-3, \dots\}, \{n-1, n-4, \dots\}, \{n-2, n-5, \dots\}$ (when $e = n$).
 177 We have two cases. When e is taken to be 1 then $x + 1 = y$ if and only if $\sigma(x, y)$ is true,
 178 and hence $w \models \chi$. When e is taken to be n , then $\sigma(x, y)$ is true if and only if $y + 1 = x$ is
 179 true. This implies that $w \models \chi''$ where χ'' is the formula obtained from χ by replacing all
 180 atomic formulas of the form $x + 1 = y$ by $y + 1 = x$. It is easy to show by induction on the
 181 structure of the formula that $w \models \chi''$ if and only if $w^r \models \chi$. Since L is closed under reverse,
 182 we deduce that $w \models \chi$. Hence the claim is proved. ◀

183 Proposition 3 says that $\text{MSO}(\text{bet}) = \text{MSO}(<) \cap \text{Rev}$. This carries down to the first-order
 184 case using the same relativization idea. In fact the result holds for the prefix class Σ_i
 185 (first-order formulas in prenex normal form with i blocks of alternating quantifiers starting
 186 with a \exists -block).

187 ► **Proposition 5.** *The following is true for definable languages.*

- 188 1. $\text{FO}(\text{bet}) = \text{FO}(<) \cap \text{Rev}$.
- 189 2. $\Sigma_i(\text{bet}) = \Sigma_i(<) \cap \text{Rev}$.

190 **Proof.** Given an $\text{FO}(<)$ formula in prenex form defining a language in Rev , we replace every
 191 occurrence of $x < y$ by $(e = x \neq y) \vee \text{bet}(e, x, y)$ as before, where e is asserted to be an
 192 endpoint with $\psi(e) = \forall x, y \neg \text{bet}(x, e, y)$. For every formula in $\Sigma_i(<)$, $i \geq 2$ this results in an
 193 equivalent formula in $\Sigma_i(\text{bet})$. For the case of Σ_1 , let us note that every formula in $\Sigma_1(<)$
 194 defines a union of languages of the form $A^*a_1A^*a_2A^*\dots A^*a_kA^*$. Such a language can be
 195 written as a disjunction of formulas like the one in Example 1. ◀

196 We noted that $\text{FO}(\text{bet}) = \text{FO}(<) \cap \text{Rev}$. Kamp [11] established that linear time temporal
 197 logic $\text{LTL}(\text{X}, \text{U})$ has the same expressive power as $\text{FO}(<)$. This is used below to establish that
 198 $\text{FO}(\text{bet})$ has the three variable property. Formulas in LTL are built from atomic propositions
 199 using boolean connectives and the two modalities *next* (X) and *until* (U). Each formula
 200 $\varphi \in \text{LTL}(\text{X}, \text{U})$ has an implicit free variable and is evaluated with respect to a word $w \in A^+$
 201 and a position i in w , we write $w, i \models \varphi$ when w at position i satisfies φ . $\text{X}\varphi$ means that φ
 202 holds at the next position, and $\varphi_1 \text{U} \varphi_2$ means that φ_2 holds at some future position and φ_1
 203 holds between the current position and this future position.

204 ► **Proposition 6.** $\text{FO}(\text{bet}) = \text{FO}^3(\text{bet})$, *i.e., for each sentence in $\text{FO}(\text{bet})$, there is an*
 205 *equivalent sentence in $\text{FO}^3(\text{bet})$ using at most three variable names.*

206 **Proof.** It suffices to show that for every reversible language $L \subseteq A^+$ definable in $\text{LTL}(\text{X}, \text{U})$
 207 there is a corresponding $\text{FO}^3(\text{bet})$ formula defining L .

208 In the proof below, we use the following macros, definable in $\text{FO}^3(\text{bet})$:

$$\begin{aligned} 209 \quad \text{E}(z) &= \neg \exists x, y \text{bet}(x, z, y) \\ 210 \quad \text{N}(x, y) &= \neg(x = y) \wedge \neg \exists z \text{bet}(x, z, y) \end{aligned}$$

212 For each formula $\varphi \in \text{LTL}(\text{X}, \text{U})$, we construct inductively an $\text{FO}^3(\text{bet})$ formula $\overline{\varphi}(x, y)$
 213 with (at most) two free variables x and y such that for all words $w \in A^+$ and position
 214 $1 \leq i \leq |w|$ we have

$$215 \quad w, i \models \varphi \quad \text{iff} \quad w, x \mapsto 1, y \mapsto i \models \overline{\varphi}(x, y) \quad (1)$$

216 The base case is when $\varphi = a \in A$ and we let $\overline{a} = P_a(y)$. For boolean connective, we define

$$217 \quad \overline{\neg \varphi}(x, y) = \neg \overline{\varphi}(x, y) \quad \overline{\varphi_1 \vee \varphi_2}(x, y) = \overline{\varphi_1}(x, y) \vee \overline{\varphi_2}(x, y).$$

219 The interesting cases are when the top connective of the formula is a modality X or U . We
 220 give the translation for the strict version of until, defined as $\varphi_1 \text{SU} \varphi_2 = \text{X}(\varphi_1 \text{U} \varphi_2)$. This is
 221 sufficient since we have $\text{X}\varphi = \perp \text{SU} \varphi$ and $\varphi_1 \text{U} \varphi_2 = \varphi_2 \vee (\varphi_1 \wedge (\varphi_1 \text{SU} \varphi_2))$. We set

$$\begin{aligned} 222 \quad \overline{\varphi_1 \text{SU} \varphi_2}(x, y) &= \exists z \left((\text{bet}(x, y, z) \vee x = y) \wedge \overline{\varphi_2}(x, z) \wedge \forall x \right. \\ 223 \quad &\quad \left. \text{bet}(y, x, z) \implies \exists y (\text{bet}(y, x, z) \wedge \text{E}(y) \wedge \overline{\varphi_1}(y, x)) \right) \end{aligned}$$

224

225 We should prove that (1) holds. We apply induction on the structure of the formula φ . The
 226 base case is when φ is $a \in A$. Then $w, i \models a$ iff $w, x \mapsto 1, y \mapsto i \models P_a(y)$. When φ is of the
 227 form $\neg\varphi'$ or $\varphi_1 \vee \varphi_2$ the claim (1) follows from induction hypothesis.

228 For the final case, let φ be $\varphi_1 \text{ SU } \varphi_2$. Then, $w, i \models \varphi_1 \text{ SU } \varphi_2$ iff there exists $k > i$ such
 229 that $w, k \models \varphi_2$ and $w, j \models \varphi_1$ for all $i < j < k$. By induction hypothesis, the latter is true
 230 if and only if $w, x \mapsto 1, z \mapsto k \models \overline{\varphi_2}(x, z)$ and $w, y \mapsto 1, x \mapsto j \models \overline{\varphi_1}(y, x)$ for all $i < j < k$.
 231 This is precisely true when $w, x \mapsto 1, y \mapsto i \models \varphi_1 \text{ SU } \varphi_2(x, y)$. This finishes the proof of (1).

Let L be a reversible language defined by the formula $\varphi \in \text{LTL}(\mathbf{X}, \mathbf{U})$. This means that L
 is reversible and $L = \{w \in A^+ \mid w, 1 \models \varphi\}$. Define the $\text{FO}^3(\text{bet})$ sentence

$$\Phi = \exists x(\mathbf{E}(x) \wedge \overline{\varphi}(x, x)).$$

232 We claim that Φ defines the language L , which concludes the proof. Since L is the set of all
 233 words w such that $w, 1 \models \varphi$. By (1), this is precisely when $w, x \mapsto 1, y \mapsto 1 \models \overline{\varphi}(x, y)$ and by
 234 renaming when $w, x \mapsto 1 \models \overline{\varphi}(x, x)$. Since L is closed under reverse $w, x \mapsto 1 \models \overline{\varphi}(x, x)$ iff
 235 $w^r, x \mapsto 1 \models \overline{\varphi}(x, x)$ iff $w, x \mapsto n \models \overline{\varphi}(x, x)$ where n is the last position of w . Therefore L is
 236 precisely the set of all words satisfying the formula Φ . ◀

237 2.2 FO(N)

238 Next we address the expressive power of FO with the neighbour predicate.

239 We start by detailing the class of *locally threshold testable languages*. Recall that word y
 240 is a *factor* of word u if $u = xyz$ for some x, z in A^* . We use $\#(u, y)$ to denote the number of
 241 times the factor y appears in u . For $t > 0$, we define the equality with threshold t on the set
 242 \mathbb{N} of natural numbers by $i =^t j$ if $i = j$ or $i, j \geq t$.

243 ► **Definition 7.** Let \approx_k^t , for $k, t > 0$, be the equivalence on A^* , whereby two words u and v
 244 are equivalent if either they both have length at most $k - 1$ and $u = v$, or otherwise they have

- 245 1. the same prefix of length $k - 1$,
- 246 2. the same suffix of length $k - 1$,
- 247 3. and the same number of occurrences, up to threshold t , for all factors of length $\leq k$, i.e.,
 248 for each word $y \in A^*$ of length at most k , $\#(u, y) =^t \#(v, y)$.

249 ► **Example 8.** We have $ababab \approx_2^1 abab \not\approx_2^1 abbab$. Indeed, all the words start and end with
 250 the same letter. In the first two words the factors ab as well as ba appear at least once.
 251 While in the last word the factor bb appears once while it is not present in the word $abab$.
 252 Notice also that $ababab \not\approx_2^2 abab$ due to the factor ba .

253 A language is *locally threshold testable* (or LTT for short) if it is a union of \approx_k^t classes,
 254 for some $k, t > 0$.

255 ► **Example 9.** The language $(ab)^*$ is LTT. In fact it is *locally testable* (the special case of
 256 locally threshold testable with $t = 1$). Indeed, $(ab)^*$ is the union of three classes: $\{\varepsilon\}$, $\{ab\}$
 257 and $abab(ab)^*$ which is precisely the set of words that begin with a , end with b , and whose
 258 only factors of length 2 are ab and ba .

259 A language that is definable in $\text{FO}(<)$ and not LTT is $c^*ac^*bc^*$. In this language if
 260 a and b are sufficiently separated by c -blocks then the order between a and b cannot be
 261 differentiated. It can be proved that for any t, k there is a sufficiently large n such that
 262 $c^n ac^n bc^n \approx_k^t c^n bc^n ac^n$.

263 Locally threshold testable languages are precisely the class of languages definable in
 264 $\text{FO}(+1)$ [12, 1]. Since we can define the neighbour predicate N using $+1$, clearly $\text{FO}(\text{N}) \subseteq$
 265 $\text{FO}(+1) \cap \text{Rev} = \text{LTT} \cap \text{Rev}$. But this inclusion is strict as shown in Example 11.

266 ► **Example 10.** Consider the language $L = ua^* + a^*u^r$ of words that have either u as prefix
 267 and followed by an arbitrary number of a 's, or u^r as suffix and preceded by an arbitrary
 268 number of a 's. The language L is in $\text{FO}(\text{N})$. When $u = a_1 \cdots a_n$, it can be defined by a
 269 formula of the form $\exists x_1, \dots, x_n \psi$ where ψ states that x_1 is an endpoint, $\bigwedge_{1 \leq i < n} \text{N}(x_i, x_{i+1})$,
 270 $\bigwedge_{1 < i < n} x_{i-1} \neq x_{i+1}$, $\bigwedge_{1 \leq i \leq n} P_{a_i}(x_i)$, and all other positions are labeled a .

► **Example 11.** Consider the language L over the alphabet $\{a, b, c\}$,

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

271 Since L is locally threshold testable and reverse closed, $L \in \text{FO}(+1) \cap \text{Rev}$.

272 We can show that $L \notin \text{FO}(\text{N})$ by showing that the words,

$$273 \quad c^k ab c^k ba c^k ab c^k \in L \qquad c^k ab c^k ab c^k ab c^k \notin L$$

274 for $k > 0$ are indistinguishable by an $\text{FO}(\text{N})$ formula of quantifier depth k . For showing the
 275 latter claim, one uses Ehrenfeucht-Fraissé games and argues that in the k -round EF-game
 276 the duplicator has a winning strategy. The strategy is roughly described below:

$$277 \quad \underline{c^k abc^k b} ac^k abc^k \qquad c^k abc^k a \underline{bc^k abc^k}$$

278 Any move of the spoiler is mimicked by the duplicator in the corresponding underlined or
 279 non-underlined part of the other word, while maintaining the neighbourhood relation between
 280 positions. For instance, if the spoiler plays the first b on the underlined part of the first word,
 281 then the duplicator chooses the last b on the underlined portion of the word on the right.
 282 Similarly, if the spoiler plays the first a on the non-underlined part of the first word, the
 283 duplicator chooses the last a on the non-underlined portion of the word on the right. Note
 284 that, since no order on positions in the words can be checked with the neighbour predicate,
 285 there is no way to distinguish between these words, if the duplicator plays in the above way
 286 ensuring that the position played has the same neighbourhood relation as the position played
 287 by the spoiler. Therefore, the Neighbour predicate will not be able to distinguish between ab
 288 and ba when they are sufficiently separated by c 's.

289 From the above example, we get,

290 ► **Proposition 12.** For definable languages, $\text{FO}(\text{N}) \subsetneq \text{FO}(+1) \cap \text{Rev} = \text{LTT} \cap \text{Rev}$.

291 Next we will characterise the class of languages accepted by $\text{FO}(\text{N})$. Recall that $\#(w, v)$
 292 denotes the number of occurrences of v in w , i.e., the number of pairs (x, y) such that
 293 $w = xvy$. We extend this to $\#^r(w, v)$ that counts the number of occurrences of v or
 294 v^r in w , i.e., the number of pairs (x, y) such that $w = xvy$ or $w = xv^ry$. Notice that
 295 $\#^r(w, v) = \#^r(w, v^r) = \#^r(w^r, v) = \#^r(w^r, v^r)$.

296 ► **Definition 13.** We define now the locally-reversible threshold testable (LRTT) equivalence
 297 relation. Let $k, t > 0$. Two words $w, w' \in A^*$ are (k, t) -LRTT equivalent, denoted $w \overset{r}{\approx}_k^t w'$ if
 298 $|w| < k$ and $w' \in \{w, w^r\}$, or

- 299 ■ w, w' are both of length at least k , and
- 300 ■ $\#^r(w, v) = \#^r(w', v)$ for all $v \in A^{\leq k}$, and
- 301 ■ if x, x' are the prefixes of w, w' of length $k - 1$ and y, y' are the suffixes of w, w' of length
 302 $k - 1$ then $\{x, y^r\} = \{x', y'^r\}$.

303 Notice that $w \overset{r}{\approx}_k^t w^r$ for all $w \in A^*$ and $w \overset{r}{\approx}_k^t w'$ implies $w \overset{r}{\approx}_k^t w'$ for all $w, w' \in A^*$.
 304 Notice also that $\overset{r}{\approx}_k^t$ is not a congruence. Indeed, we have $ab \overset{r}{\approx}_k^t ba$ but $aba \not\overset{r}{\approx}_k^t baa$. On the
 305 other hand, if $v \overset{r}{\approx}_k^t w$ then for all $u \in A^*$ we have $uv \overset{r}{\approx}_k^t uw$ or $uv \overset{r}{\approx}_k^t uw^r$, and similarly
 306 $vu \overset{r}{\approx}_k^t wu$ or $vu \overset{r}{\approx}_k^t w^r u$.

307 ► **Definition 14** (Locally-Reversible Threshold Testable Languages). *A language L is locally-*
 308 *reversible threshold testable, LRTT for short, if it is a union of equivalence classes of $\overset{r}{\approx}_k^t$ for*
 309 *some $k, t > 0$.*

310 ► **Theorem 15.** *Languages defined by $\text{FO}(\mathbb{N})$ are precisely the class of locally-reversible*
 311 *threshold testable languages.*

312 **Proof.** (\Leftarrow) Assume we are given an LRTT language, i.e., a union of $\overset{r}{\approx}_k^t$ -classes for some
 313 $k, t > 0$. We explain how to write an $\text{FO}(\mathbb{N})$ formula for each $\overset{r}{\approx}_k^t$ -class. Consider a word
 314 $v = a_1 a_2 \cdots a_n \in A^+$. For $m \in \mathbb{N}$, we can say that v or its reverse occurs at least m times in
 315 a word $w \in A^*$, i.e., $\sharp^r(w, v) \geq m$, by the formula

$$\begin{aligned} \varphi_v^{\geq m} = & \exists x_{1,1} \cdots \exists x_{1,n} \cdots \exists x_{m,1} \cdots \exists x_{m,n} \\ & \bigwedge_{i=1}^m \left(\bigwedge_{j=1}^{n-1} \mathbf{N}(x_{i,j}, x_{i,j+1}) \wedge \bigwedge_{j=2}^{n-1} (x_{i,j-1} \neq x_{i,j+1}) \wedge \bigwedge_{j=1}^n P_{a_j}(x_{i,j}) \right) \\ & \wedge \bigwedge_{1 \leq i < j \leq m} \neg((x_{i,1} = x_{j,1} \wedge x_{i,n} = x_{j,n}) \vee (x_{i,1} = x_{j,n} \wedge x_{i,n} = x_{j,1})). \end{aligned}$$

317 Similarly, we can write a formula $\psi_v \in \text{FO}(\mathbb{N})$ that says that a word belongs to $\{v, v^r\}$.
 318 Finally, given two words of same length $u, v \in A^n$, we can write a formula $\chi_{u,v} \in \text{FO}(\mathbb{N})$ that
 319 says that u, v occur at two different end points of a word w , i.e., that $\{x, y^r\} = \{u, v\}$ where
 320 x, y are the prefix and suffix of w of length n .

321 (\Rightarrow) Hanf's theorem ([13], Theorem 2.4.1) states that two first-order structures \mathfrak{A} and \mathfrak{B}
 322 are m -equivalent (i.e., indistinguishable by any FO formula of quantifier rank at most m),
 323 for some $m \in \mathbb{N}$ if for each 3^m ball type S , both \mathfrak{A} and \mathfrak{B} have the same number of 3^m balls
 324 of type S up to a threshold $m \times e$, where $e \in \mathbb{N}$. Models of $\text{FO}(\mathbb{N})$ formulas are first-order
 325 structures of the form $(\{1, \dots, n\}, (P_a)_{a \in \Sigma}, \mathbb{N})$ that are labelled undirected path graphs.
 326 Balls in such a graph are nothing but factors of the corresponding undirected word. Applying
 327 Hanf's theorem to undirected path graphs, we obtain that given an $\text{FO}(\mathbb{N})$ formula Φ , there
 328 exist $k, t > 0$ such that if two words w and w' are $\overset{r}{\approx}_k^t$ -equivalent, then w satisfies Φ if and
 329 only if w' satisfies Φ . Therefore, the set of all words satisfying Φ is an LRTT language. ◀

330 **3 The Membership problem for the Logics**

331 In this section we address the question of definability of a language — “is the given reversible
 332 regular language definable by a formula in the logic?” — in the previously defined logics. We
 333 show that in the case of $\text{FO}(\text{bet})$ the existing theorems provide an algorithm for the problem,
 334 while for $\text{FO}(\mathbb{N})$ the answer is not yet known.

335 **3.1 Membership in $\text{MSO}(\text{bet})$, $\text{MSO}(\mathbb{N})$, $\text{FO}(\text{bet})$**

336 By Proposition 3, to check if a regular language is definable in $\text{MSO}(\text{bet})$ or in $\text{MSO}(\mathbb{N})$ it
 337 suffices to check if it is reversible. Next we look at the membership problem for $\text{FO}(\text{bet})$.

338 First we recall the notion of recognisability by a finite semigroup. A finite semigroup
 339 (S, \cdot) is a finite set S with an associative binary operation $\cdot : S \times S \rightarrow S$. If the semigroup
 340 operation has an identity, then it is necessarily unique and is denoted by 1. In this case S
 341 is called a monoid. A semigroup morphism from (S, \cdot) to $(T, +)$ is a map $h : S \rightarrow T$ that
 342 preserves the semigroup operation, i.e., $h(a \cdot b) = h(a) + h(b)$ for a, b in S . Further if S and T
 343 are monoids the map is a monoid morphism if h maps the identity of S to the identity of T .

344 The set A^+ under concatenation forms a free semigroup while A^* under concatenation
 345 forms a free monoid with the empty word ε as the identity. A language $L \subseteq A^*$ is *recognised*
 346 by a monoid (M, \cdot) , if there is a morphism $h : A^* \rightarrow M$ and a set $P \subseteq M$, such that
 347 $L = h^{-1}(P)$.

348 Given a language L , the *syntactic congruence* of L , denoted as \sim_L is the congruence on
 349 A^* ,

$$350 \quad x \sim_L y \quad \text{if} \quad uxv \in L \Leftrightarrow uyv \in L \text{ for all } u, v \in A^*. \quad (2)$$

351 The quotient A^*/\sim_L , denoted as $M(L)$, is called the *syntactic monoid*. It recognises L and is
 352 the unique minimal object with the following canonicity property: any monoid M recognising
 353 L has a *surjective* morphism from a submonoid of M to $M(L)$ [4].

354 A semigroup (or monoid) is aperiodic if there is some $n \in \mathbb{N}$ such that $a^n = a^{n+1}$ for
 355 each element a of the semigroup. Schützenberger-McNaughton-Papert theorem [2, 3] states
 356 that a language L is definable in $\text{FO}(<)$ if and only if the syntactic monoid of L is aperiodic.
 357 This theorem in conjunction with Proposition 5 gives that,

358 **► Corollary 16.** *A reversible language L is definable in $\text{FO}(\text{bet})$ if and only if $M(L)$ is*
 359 *aperiodic.*

360 The above result hence yields an algorithm for definability of a language in $\text{FO}(\text{bet})$,
 361 i.e., check if the language is reversible, if so compute the syntactic monoid and test for
 362 aperiodicity.

363 3.2 Membership in $\text{FO}(\mathbb{N})$

364 Next we look at the membership problem for the logic $\text{FO}(\mathbb{N})$. The corresponding problem
 365 for $\text{FO}(+1)$ is known only in terms of syntactic semigroups that we recall now. A language
 366 $L \subseteq A^+$ is *recognised* by a semigroup (S, \cdot) , if there is a morphism $h : A^+ \rightarrow S$ and a set
 367 $P \subseteq S$, such that $L = h^{-1}(P)$.

368 The *syntactic congruence* of $L \subseteq A^+$, denoted as \sim_L , is the congruence on A^+ given
 369 by Equation (2). The quotient A^+/\sim_L , denoted as $S(L)$, is called the *syntactic semigroup*.
 370 It shares the canonicity property of syntactic monoids, namely it recognises L and is the
 371 unique minimal object that has a surjective morphism from a subsemigroup of any semigroup
 372 recognising L [14].

373 The characterisation theorem for $\text{FO}(+1)$ due to Brzowski and Simon [15], and Beauquier
 374 and Pin [12], is stated below. Recall that an element of a semigroup e is an *idempotent* if
 375 $e \cdot e = e$.

376 **► Theorem 17** (Brzowski-Simon, Beauquier-Pin). *The following are equivalent for a language*
 377 *$L \subseteq A^+$.*

- 378 1. *L is locally threshold testable.*
- 379 2. *L is definable in $\text{FO}(+1)$.*
- 380 3. *The syntactic semigroup of L is finite, aperiodic and satisfies the identity $exfy e z f =$*
 381 *$e z f y e x f$ for all $e, f, x, y, z \in S(L)$ with e, f idempotents.*

382 Because of Proposition 12 it is clear that we need to add more identities to characterise
383 the logic FO(N).

384 In the particular case of reversible languages the syntactic semigroups described above
385 admit further properties. The observation is that the reverse operation extends to congruence
386 classes of the syntactic congruence as shown next. Fix a reversible language L . Let $[x] \in S(L)$
387 denote the equivalence class of a word $x \in A^+$ under the syntactic congruence. Then we let
388 $[x]^r = [x^r]$. This is well defined since $x \sim_L y$ if and only if $x^r \sim_L y^r$. Furthermore this map
389 admits two properties — it is an involution (a map that is its own inverse), since

$$390 \quad ([x]^r)^r = ([x^r])^r = [(x^r)^r] = [x], \quad (3)$$

391 and it is an anti-automorphism on the semigroup $S(L)$ since

$$392 \quad ([x] \cdot [y])^r = ([x \cdot y])^r = [(x \cdot y)^r] = [y^r \cdot x^r] = [y^r] \cdot [x^r]. \quad (4)$$

393 Thus $S(L)$ is a semigroup with an involution operation, namely the reverse. Formally, a
394 *semigroup with involution* (also called a \star -semigroup) (S, \cdot, \star) is a semigroup (S, \cdot) extended
395 with an operation $\star: S \rightarrow S$ (called the involution) such that

- 396 1. the operation \star is an involution on S , i.e., $(a^\star)^\star = a$ for all elements a of S ,
- 397 2. the operation \star is an anti-automorphism on S (isomorphism between S and opposite of
398 S), i.e., $(a \cdot b)^\star = b^\star \cdot a^\star$ for any a, b in S .

399 It is a \star -monoid if S is a monoid. An element x in a \star -semigroup is called *hermitian* if it is
400 its own involution, i.e. $x^\star = x$. It is easy to see that in the case of \star -monoids, necessarily
401 $1^\star = 1$, i.e. identity is hermitian. Similarly if the semigroup has a zero it is hermitian as well.

402 In the light of this definition we call $S(L)$ the *syntactic \star -semigroup* of a reversible
403 language L . Next we show that syntactic \star -semigroups of FO(N)-definable languages obey
404 the identity $exe^\star = ex^\star e^\star$, where x is an element of the semigroup and e is an idempotent of
405 the semigroup. Before we prove it, we look at a couple of examples.

406 ► **Example 18.** Fix the alphabet $\{c, x, y\}$ for the example below. Consider the reversible
407 languages $L_1 = c^\star xyc^\star xyc^\star + c^\star yxc^\star yxc^\star$, $L_2 = c^\star(xy + yx)c^\star(xy + yx)c^\star$. It is easy to verify
408 that both languages are definable in FO(+1). Their syntactic semigroups are shown in Figure
409 1. These semigroups were computed using the online tool of Charles Paperman [16].

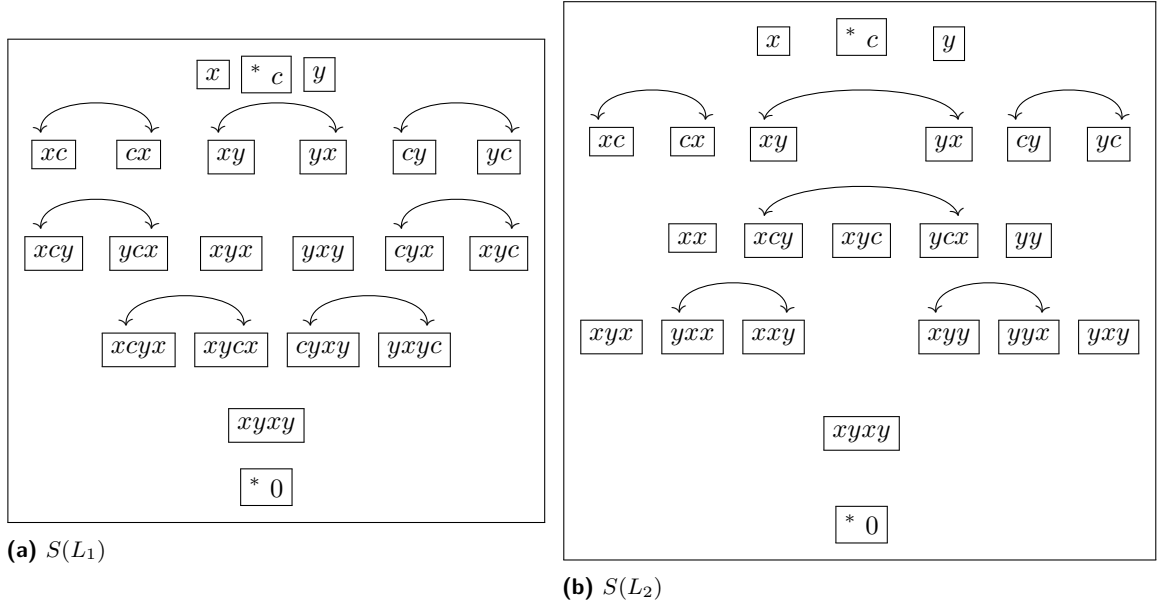
410 Let x, y, c also denote the images of the corresponding letters in the syntactic semigroups,
411 which are indeed hermitian. Clearly the syntactic semigroups are generated by these elements.
412 It is easy to deduce that c is an idempotent while x, y are not. Also both semigroups have
413 zeros (for instance any product x^k involving more than two occurrences of x is a zero). Next
414 we have a closer look at them.

- 415 1. We claim that $S(L_1)$ does not satisfy the identity $exe^\star = ex^\star e^\star$. Consider the two words
416 cxy and $cyxc$. Note that $(xy)^\star = y^\star x^\star = yx$. It suffices to show that $cxy \not\sim_{L_1} cyxc$,
417 that is evident since $cxy \cdot cxy \cdot \varepsilon \in L_1$ while $cyxc \cdot cyxc \cdot \varepsilon \notin L_1$.
- 418 2. Next we verify that $S(L_2)$ satisfies the identity $exe^\star = ex^\star e^\star$. Since the only two
419 idempotents are c and 0 , it suffices to show that $cuc \sim_{L_2} cu^r c$ for all nonempty words
420 u . It is easy to verify that $pcucq \in L$ if and only if $pcu^r cq \in L$ for all words p, q and
421 hence $cuc \sim_{L_2} cu^r c$. Since u was arbitrary it follows that $S(L_2)$ satisfies the identity
422 $exe^\star = ex^\star e^\star$.

► **Theorem 19.** *The syntactic \star -semigroup of an FO(N)-definable language satisfies the identity*

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

423 where e is an idempotent, and x is any element of the semigroup.



■ **Figure 1** Syntactic involution semigroups of $L_1 = c^*xyc^*xyc^* + c^*yxc^*yxc^*$ and $L_2 = c^*(xy + yx)c^*(xy + yx)c^*$ over the alphabet $\{c, x, y\}$. Idempotents are indicated by $*$. Involutions on elements are indicated by arrows, unless the element is hermitian.

424 **Proof.** Assume we are given an $\text{FO}(\mathbb{N})$ -language L , with its syntactic \star -semigroup $S(L) =$
 425 $(A^+/\sim_L, \cdot, \star)$, and $h: A^+ \rightarrow S(L)$ the canonical morphism recognising L . Let e be an
 426 idempotent of $S(L)$, and let x be an element of $S(L)$. Pick nonempty words u and s such
 427 that $h(u) = e$ and $h(s) = x$.

428 By definition of the involution, $h(u^r) = e^*$ and $h(s^r) = x^*$. We are going to show
 429 that $usu^r \sim_L us^r u^r$ and hence they will correspond to the same element in the syntactic
 430 \star -semigroup, proving that $exe^* = ex^*e^*$.

431 Since L is $\text{FO}(\mathbb{N})$ definable, we know by Theorem 15 that L is a union of \approx_k^r equivalence
 432 classes for some $k, t > 0$. Consider the words $w = (u^k)s(u^r)^k$ and $w^r = (u^k)s^r(u^r)^k$, obtained
 433 by pumping the words corresponding to e and e^* . Since e, e^* are idempotents, it is clear that
 434 $h(w) = h(usu^r) = exe^*$ and $h(w^r) = h(us^r u^r) = ex^*e^*$.

435 For all contexts $\alpha, \beta \in A^*$, we show below that $\alpha w \beta \approx_k^r \alpha w^r \beta$, which implies $\alpha w \beta \in L$
 436 iff $\alpha w^r \beta \in L$ since L is a union of \approx_k^r classes. It follows that $w \sim_L w^r$ and therefore
 437 $h(w) = h(w^r)$, that will conclude the proof.

438 Fix some contexts $\alpha, \beta \in A^*$. Since $u \neq \varepsilon$, the words $\alpha w \beta$ and $\alpha w^r \beta$ have the same prefix
 439 of length $k - 1$ and the same suffix of length $k - 1$. Now, consider $v \in A^k$. If an occurrence
 440 of v (resp. v^r) in $\alpha w \beta$ overlaps with α or β then we have the very same occurrence in $\alpha w^r \beta$.
 441 Using $w \approx_k^r w^r$, we deduce that $\#^r(\alpha w \beta, v) = \#^r(\alpha w^r \beta, v)$. Therefore, $\alpha w \beta \approx_k^r \alpha w^r \beta$. ◀

442 The converse direction is open. The similar direction in the case of $\text{FO}(+1)$ goes via categories
 443 [17] and uses the Delay theorem of Straubing [18, 4]. However in the special case when the
 444 syntactic \star -semigroups are monoids (i.e. contains an identity) we can get an easy converse.

445 Let A be an alphabet and let $L \subseteq A^+$ be a language over A . A letter $c \in A$ is *neutral* in
 446 the language L if $xy \in L \Leftrightarrow xcy \in L$ for all $x, y \in A^*$ such that $|xy| \geq 1$, i.e. membership
 447 in L is invariant under insertion or deletion of the letter c . By definition, it is easy to see
 448 that if L has a neutral letter then that maps to an element that is identity in the syntactic

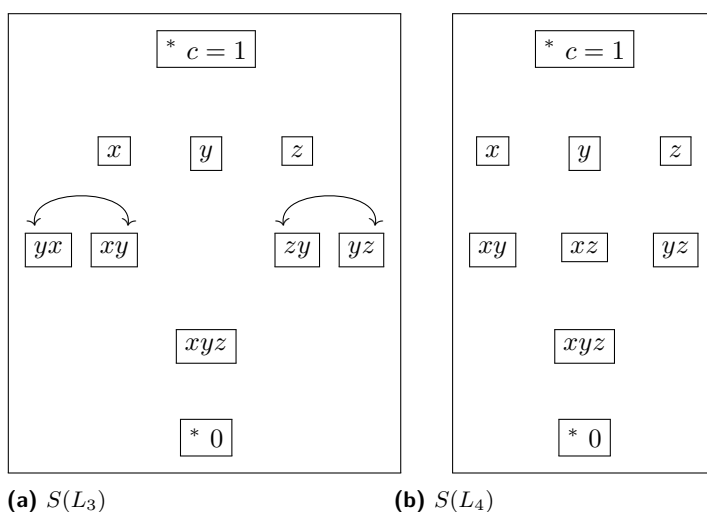


Figure 2 Syntactic involution semigroups of $L_3 = c^*xc^*yc^*zc^* + c^*zc^*yc^*xc^*$ and $L_4 = \text{Permutations}(L_3)$ over the alphabet $\{c, x, y, z\}$. Idempotents are indicated by *. Involutions on elements are indicated by arrows, unless the element is hermitian.

semigroup of L . For aperiodic semigroups the converse is also true.

► **Lemma 20.** *Let L be an aperiodic language. Then the syntactic semigroup $S(L)$ contains an identity if and only if L has a neutral letter.*

Proof. (\Leftarrow) By definition.

(\Rightarrow) Assume $S(L)$ has an identity and let $cu \in A^+$, where $c \in A$, be a word mapping to it. If u is empty then c is a neutral letter and we are done. Otherwise let $a = \varphi(c)$ and $b = \varphi(u)$. Then $aab = a$ and by repeated substitution $a^nab^n = a$ and since $S(L)$ is aperiodic there is some n such that $a^n a = a^n$ and hence $a = a^n ab^n = a^n b^n = ab$. Therefore c is a neutral letter.

► **Example 21.** Fix the alphabet $\{c, x, y, z\}$ for the example below. Let $L_3 = c^*xc^*yc^*zc^* + c^*zc^*yc^*xc^*$, and L_4 be the set of all permutations of words in L_3 , ie. the commutative closure of L_3 . Both the languages are closed under reverse. Moreover it is easy to verify that both are definable in $\text{FO}(<)$ and by extension in $\text{FO}(\text{bet})$. Hence their syntactic semigroups are aperiodic. They are shown in Figure 2.

Let x, y, z, c denote the images of the corresponding letters in the syntactic semigroups. Clearly the syntactic semigroups are generated by these elements. Since letter c is neutral in both L_3 and L_4 , we deduce that c is the identity. Also any product involving at least two occurrences of x (or y , or z) is a (non-accepting) zero element denoted as 0. These are the only idempotents in the syntactic semigroups. Next we have a closer look at them.

1. Consider the language L_3 . The semigroup $S(L_3)$ is a monoid with identity c . It satisfies the additional rules $xz = zx = 0$ and $xyz = zyx$. Since $cxcccyc = xy \neq yx = cyccxc$, $S(L_3)$ does not satisfy the condition on syntactic semigroups given by Theorem 17 and hence L_3 is not LTT, and by extension not in LRTT either.
2. Next consider the language L_4 that is all the permutations of words in L_1 . The semigroup $S(L_4)$ is commutative and it has an identity (the element c) and a zero. Since all elements

475 in $S(L_4)$ are hermitian, it is clear that $S(L_4)$ satisfies the identity $exe^* = ex^*e^*$ in
 476 addition to those corresponding to $\text{FO}(+1)$.

477 ► **Lemma 22.** *If all elements of an involution semigroup S are hermitian, then S is commutative. Conversely if S is commutative and generated by a subset of its hermitian elements, then all elements of S are hermitian.*

480 **Proof.** If all elements of S are hermitian, then $ab = (ab)^* = b^*a^* = ba$, for all elements
 481 $a, b \in S$, i.e. S is commutative. Conversely assume S is commutative and generated by
 482 a subset of hermitian elements. Then every element can be written as a product $x_1 \cdots x_n$
 483 where each x_i is hermitian. Then $(x_1 \cdots x_n)^* = x_n^* \cdots x_1^* = x_n \cdots x_1 = x_1 \cdots x_n$. Hence all
 484 elements are hermitian. ◀

485 Since syntactic semigroups are generated by images of letters (that are clearly hermitian),
 486 we obtain the following.

487 ► **Proposition 23.** *Let $L \subseteq A^+$ be a reversible language with a neutral letter. Then the following are equivalent.*

- 489 1. All elements of $S(L)$ are hermitian.
- 490 2. $S(L)$ satisfies the identity $exe^* = ex^*e^*$.
- 491 3. $S(L)$ is locally idempotent i.e., it satisfies the identity $exeye = eyexe$ for all idempotents
 492 e and elements $x \in S(L)$.
- 493 4. $S(L)$ is commutative.

494 **Proof.** Equivalence of (1) and (4) is from Lemma 22. Now, (1) implies (2) is clear and the
 495 converse (2) implies (1) is because $S(L)$ has an identity. Similarly (4) implies (3) is clear
 496 and (3) implies (4) is due to the presence of an identity. ◀

497 ► **Corollary 24.** *Let L be a reversible language with a neutral letter. The following are equivalent.*

- 499 1. $L \in \text{LTT}$, equivalently, L is definable in $\text{FO}(+1)$.
- 500 2. $L \in \text{LR TT}$, equivalently, L is definable in $\text{FO}(\text{N})$.
- 501 3. $L \in \text{ACom}$ (the class of aperiodic and commutative languages), equivalently, L is definable
 502 in $\text{FO}(=)$ [4].

503 **Proof.** (1 \Rightarrow 3) and (2 \Rightarrow 3) follows from Proposition 23. The converse inclusions are by
 504 definition. ◀

505 4 Conclusion

506 The logics $\text{MSO}(\text{bet})$, $\text{MSO}(\text{N})$ and $\text{FO}(\text{bet})$ behave analogously to their classical counterparts
 507 $\text{MSO}(<)$, $\text{MSO}(+1)$ and $\text{FO}(<)$. But the logic $\text{FO}(\text{N})$ gives rise to a new class of languages,
 508 locally-reversible threshold testable languages. The quest for characterising the new class
 509 takes us to the formalism of involution semigroups. The full characterisation of the new
 510 class is the main question we leave open. It would also be interesting to know what are
 511 the natural analogues of standard fragments of FO and their expressive power, for instance
 512 classes defined by bounded number of variables, in the reversible world. Another line of
 513 investigation is to study the equationally-defined classes that arise naturally from automata
 514 theory.

515 ——— **References** ———

- 516 1 Thomas W. Classifying Regular Events in Symbolic Logic. *J. Comput. Syst. Sci.*,
517 1982. **25**(3):360–376. doi:10.1016/0022-0000(82)90016-2. URL [https://doi.org/10.1016/](https://doi.org/10.1016/0022-0000(82)90016-2)
518 0022-0000(82)90016-2.
- 519 2 Schützenberger MP. On Finite Monoids Having Only Trivial Subgroups. *Information and*
520 *Control*, 1965. **8**(2):190–194. doi:10.1016/S0019-9958(65)90108-7. URL [https://doi.org/10.](https://doi.org/10.1016/S0019-9958(65)90108-7)
521 1016/S0019-9958(65)90108-7.
- 522 3 McNaughton R, Papert SA. Counter-Free Automata (M.I.T. Research Monograph No. 65).
523 The MIT Press, 1971. ISBN 0262130769.
- 524 4 Straubing H. Finite Automata, Formal Logic, and Circuit Complexity. Birkhäuser Verlag,
525 Basel, Switzerland, 1994. ISBN 3-7643-3719-2.
- 526 5 Krebs A, Lodaya K, Pandya PK, Straubing H. Two-variable Logic with a Between Relation. In:
527 Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS
528 '16, New York, NY, USA, July 5-8, 2016. 2016 pp. 106–115. doi:10.1145/2933575.2935308.
529 URL <https://doi.org/10.1145/2933575.2935308>.
- 530 6 Krebs A, Lodaya K, Pandya PK, Straubing H. An Algebraic Decision Procedure for Two-
531 Variable Logic with a Between Relation. In: 27th EACSL Annual Conference on Computer
532 Science Logic, CSL 2018, September 4-7, 2018, Birmingham, UK. 2018 pp. 28:1–28:17. doi:
533 10.4230/LIPIcs.CSL.2018.28. URL <https://doi.org/10.4230/LIPIcs.CSL.2018.28>.
- 534 7 Krebs A, Lodaya K, Pandya PK, Straubing H. Two-variable logics with some betweenness
535 relations: Expressiveness, satisfiability and membership. *arXiv preprint arXiv:1902.05905*,
536 2019.
- 537 8 Tesson P, Therien D. Diamonds Are Forever: The Variety **DA**. In: Semigroups, Algorithms,
538 Automata and Languages, Coimbra (Portugal) 2001. World Scientific, 2002 pp. 475–500.
- 539 9 Crvenković S, Dolinka I. Varieties of involution semigroups and involution semirings: a survey.
540 In: Proceedings of the International Conference “Contemporary Developments in Mathematics”
541 (Banja Luka, 2000), Bulletin of Society of Mathematicians of Banja Luka. 2000 pp. 7–47.
- 542 10 Gastin P, Manuel A, Govind R. Logics for Reversible Regular Languages and Semigroups
543 with Involution. In: Hofman P, Skrzypczak M (eds.), Proceedings of the 23th International
544 Conference on Developments in Language Theory (DLT'19), volume 11647 of *Lecture Notes in*
545 *Computer Science*. Springer, Warsaw, Poland, 2019 pp. 182–191.
- 546 11 Kamp JAW. Tense Logic and the Theory of Linear Order. Ph.D. thesis, University of California,
547 1968.
- 548 12 Beauquier D, Pin JÉ. Languages and Scanners. *Theor. Comput. Sci.*, 1991. **84**(1):3–21.
549 doi:10.1016/0304-3975(91)90258-4. URL [https://doi.org/10.1016/0304-3975\(91\)90258-4](https://doi.org/10.1016/0304-3975(91)90258-4).
- 550 13 Ebbinghaus HD, Flum J. Finite Model Theory: Second Edition. Springer Monographs in
551 Mathematics. Springer Berlin Heidelberg, 2005. ISBN 9783540287889.
- 552 14 Pin JÉ. Mathematical Foundations of Automata Theory. URL [https://www.irif.fr/~jep/](https://www.irif.fr/~jep/PDF/MPRI/MPRI.pdf)
553 PDF/MPRI/MPRI.pdf.
- 554 15 Brzozowski JA, Simon I. Characterizations of Locally Testable Events. In: 12th Annual
555 Symposium on Switching and Automata Theory, East Lansing, Michigan, USA, October 13-15,
556 1971. 1971 pp. 166–176. doi:10.1109/SWAT.1971.6. URL [https://doi.org/10.1109/SWAT.](https://doi.org/10.1109/SWAT.1971.6)
557 1971.6.
- 558 16 Paperman C. Semigroup Online. URL <https://www.paperman.name/semigroup/>.
- 559 17 Tilson B. Categories as algebra: An essential ingredient in the theory of monoids. *Journal of*
560 *Pure and Applied Algebra*, 1987. **48**(1-2):83–198.
- 561 18 Straubing H. Finite semigroup varieties of the form $\mathbf{V} \star \mathbf{D}$. *Journal of Pure and Applied*
562 *Algebra*, 1985. **36**:53 – 94.