A logical characterization of various classes of regular languages

Argimiro Arratia
Departamento de Matemáticas,
Universidad Simón Bolívar
Apartado 89000, Caracas 1080-A,
Venezuela
E-mail: arratia@ma.usb.ve

A class of languages is defined from the operations of union, intersection, complement, concatenation and a new operation which, for two given languages A and B, and a fixed language V, called the context, builds the set of words whose number of possible factorizations as three factors, the prefix in A, the suffix in B, and the middle in the context, is congruent to an integer modulus. An appropriate family of generalized quantifiers is defined that, when added to first order logic, captures exactly the aforesaid class of languages with respect to finite linearly ordered structures. It is shown, using model theoretical tools, that our languages so constructed are regular and how particular cases of this construction corresponds to various classes of regular languages. The main tool that we have developed is a theorem that characterizes equivalence of formulas in our logics in terms of games, and in terms of satisfiability of suitable isomorphism types.

Key Words: regular languages, finite automata, generalized quantifiers, Ehrenfeucht games, isomorphism types.

1. INTRODUCTION

The logical characterization of different classes of regular languages has its origins in a paper by Büchi [1]. There, it is shown that a language is regular if, and only if, it is definable by a sentence in a monadic second order theory of linear order. Following that work, McNaughton and Papert showed that the star–free regular languages are exactly those definable by sentences in a first order theory of linear order (see [6]). It was natural then to believe that an increase in expressive power of first order logic plus ordering by means of adding some logical construct, definable in monadic second order logic but not in first order logic, would yield logics that corresponds to classes of regular languages above the class of star–free, in a way much the same as first order logic corresponds to star-free languages and monadic second order logic to all the regular languages.

One method of increasing the expressive power of logics (not just first order) was provided by Lindström in [5] (but see [3] for a more updated account), and has come to be known as *the method of adding generalized (or Lindström) quantifiers.* Such method was used in [8] to show that the class of regular languages with corresponding syntactic monoid containing only solvable groups is, precisely, the class of languages definable by sentences in a first order theory of linear order with generalized quantifiers that allow counting modulo a prime number. Specifically, the generalized quantifier introduced by the authors of [8] is denoted by $\exists^{(p,q)}$ (p and q positive integers) and interpreted as follows. For a formula $\phi(x)$ with free variable x, the formula $\exists^{(p,q)}x\phi(x)$ is well-formed and it is true if, and only if, the number of witnesses for x that makes $\phi(x)$ true is congruent to p modulo q. The idea for the definition of this counting quantifier comes naturally from previous work by Straubing [7] and Thérien [9], where the class of regular languages with solvable syntactic monoid was identified as the class of languages containing the empty and the single letter languages, and closed under the operations of union, intersection, concatenation, complementation and *cyclic counting concatenation.* The latter being defined as follows: if Σ is some finite alphabet, σ a letter in Σ, A and B languages over Σ, p and q positive integers with $p < q$, then

$$\begin{aligned}\langle A,\sigma,B,p,q\rangle \;=\; & \{w \in \Sigma^* : \text{ the number of factorizations of } w \\ & \text{in the form } w = v\sigma u \text{ with } v \in A \text{ and } u \in B \\ & \text{is congruent to } p \text{ modulo } q\}.\end{aligned}$$

Now, the problem of characterizing classes of regular languages whose syntactic monoid contains a non solvable group, seems yet open. Such regular languages exist, for one can take a finite non solvable group as S_5 (the group of permutations of five elements), and build a finite automata with transitions simulating the relations that a set of generators for this group satisfy. Looking at the known operations used to obtain new regular languages from known ones, and as described above, it is natural to conjecture that counting subwords within words is a possible extension to the operation of letter counting which might yield languages with non solvable syntactic monoid. This idea is also suggested in [9], and motivated by it we define, in this paper, an infinite family of classes of languages obtained from the basic boolean operations and a new operation, which we named cyclic counting concatenation with respect to a context (section 2). We define an appropriate family of generalized quantifiers that, when added to first order logic, give logics that defines such language constructs (section 3). We then show that our logics only define languages within the class of regular languages, and show as examples how to capture the classes of regular languages with corresponding syntactic monoid containing only solvable groups, or none (the aperiodic languages).

Our terminology and presentation is inspired, to a certain extend, by that in the book [3], and the reader can find there the basics on logic and finite models. We are mainly concerned here with developing the appropriate logical machinery for understanding the structural properties of regular languages, and to set the basis for classifying the non solvable regular languages. Our main logical tool (Theorem 3.2) follows the pattern of similar theorems in [3] where formula equivalence is shown equivalent to the existence of a winning strategy in a suitable (Ehrenfeucht's style) game, and equivalent to satisfiability of a certain isomorphism type (to be defined in section 3).

2. REGULAR LANGUAGES

Let Σ denote a finite alphabet and Σ^* the set of finite strings (or words) over Σ plus e (the empty word); $\Sigma^+ := \Sigma^* \setminus \{e\}$. If $w \in \Sigma^+$, $|w|$ is the length of w; $|e| = 0$. A subset of Σ^* is called a *language*. 2^{Σ^+} denotes the set of all subsets of Σ^+.

A *regular language* is a member of the smallest class of languages that contains $\emptyset$, $\{\sigma\}$ for every $\sigma \in \Sigma$, and is closed under finite union, concatenation, and the Kleene star.

According to a theorem by Kleene, a regular language is exactly the language computed by a deterministic finite automata (DFA). If M is a DFA we denote by $L(M)$ the language computed by M. The details on finite automata and regular languages can be found in any standard book on theoretical computer sciences (e.g., [4]). We will use Kleene's machine characterization of regular languages to show that the languages definable by the logics that we construct in this paper are all regular (this is Theorem 4.1).

We shall now define our new operation on laguages, which we have named *cyclic counting concatenation with respect to a context*. Fix V a language over Σ. For any pair of subsets of Σ^*, A and B, for any pair of positive integers p and q, with $p < q$, we define the language

$$\begin{aligned} \langle A, V, B, p, q\rangle \;=\; & \{w \in \Sigma^* : \text{ the number of factorizations of } w \\ & \text{in the form } w = vzu \text{ with } v \in A,\ u \in B, \text{ and} \\ & \text{some } z \in V \text{ is} > 0 \text{ and congruent to } p \text{ modulo } q\}. \end{aligned}$$

Definition 2.1. Let $P \subseteq \mathbb{Z}^+$ and $\Gamma \subseteq 2^{\Sigma^+}$. $\mathcal{AC}(P,\Gamma)$ is the smallest class of languages over Σ that contains $\emptyset$, $\{\sigma\}$ for all $\sigma \in \Sigma$, closed under finite union, finite concatenation and complement; and it is also closed under cyclic counting concatenation with context in Γ and modulus in P; that is, given $A, B \in \mathcal{AC}(P,\Gamma)$, $V \in \Gamma$, $q \in P$, and $p < q$, $\langle A, V, B, p, q\rangle \in \mathcal{AC}(P,\Gamma)$.

(The notation $\mathcal{AC}$ is taken from [8], where it is justified as follows: $\mathcal{A}$ stands for aperiodic and $\mathcal{C}$ for counting. We found this symbology appropriate and thus use it.)

3. THE LOGIC FRAMEWORK

Let Σ be a finite alphabet. Let $\tau_\Sigma = \{<\} \cup \{P_\sigma : \sigma \in \Sigma\}$, where $<$ is a binary relation symbol and each P_σ a unary relation symbol. To each word $w \in \Sigma^+$, one associates the (finite) τ_Σ–structure

$$\mathcal{A}_w = \langle |\mathcal{A}_w|, <^{\mathcal{A}_w}, P_\sigma^{\mathcal{A}_w} : \sigma \in \Sigma\rangle$$

where $|\mathcal{A}_w| = \{1, 2, \ldots, |w|\}$, $<^{\mathcal{A}_w}$ is interpret as linear order on the naturals, and

$$P_\sigma^{\mathcal{A}_w} = \{j \in |\mathcal{A}_w| : \text{ the } j\text{-th letter of } w \text{ (from left to right) is } \sigma\}.$$

Consider the extended vocabulary $\tau = \tau_\Sigma \cup \{U_1, U_2\}$, where U_1 and U_2 are new unary relation symbols. For $V \subseteq \Sigma^+$, p and q positive integers with $p < q$, define the class of τ-structures

$$\begin{aligned}\mathcal{Q}_V^{(p,q)} \;=\; \{\, & \langle \mathcal{A}_w, U_1^{\mathcal{A}_w}, U_2^{\mathcal{A}_w}\rangle : w \in \Sigma^+, U_i^{\mathcal{A}_w} \subseteq |\mathcal{A}_w|, \text{ and the number of} \\ & \text{intervals } [i,j] \subseteq |\mathcal{A}_w| \text{ with } i \in U_1^{\mathcal{A}_w} \text{ and } j \in U_2^{\mathcal{A}_w} \text{ such that} \\ & \langle [i,j], <^{\mathcal{A}_w}, P_\sigma^{\mathcal{A}_w} : \sigma \in \Sigma\rangle \cong \mathcal{A}_z \text{ for some } z \in V \text{ is } > 0 \text{ and} \\ & \text{congruent to } p \text{ modulo } q\}.\end{aligned}$$

From the class of τ-structures $\mathcal{Q}_V^{(p,q)}$ one defines a quantifier (following Lindström's idea), which we denote also by $\mathcal{Q}_V^{(p,q)}$, and add it to FO to enrich its expressive power. FO is first order logic with equality, and FO(τ) is first order logic with equality plus the predicate symbols in the vocabulary τ. This is the set of formulas constructed from the basic set of boolean connectives ($\wedge$, $\vee$, $\neg$), first order quantifiers, equality, and the elements in τ.

For $P \subseteq \mathbb{Z}^+$, $\Gamma \subseteq 2^{\Sigma^+}$, define the logic over the vocabulary τ, $\mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$, as follows.

Syntax: (1) $\mathrm{FO}(\tau) \subseteq \mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$ (i.e. every formula in FO(τ) is a formula in $\mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$);

(2) If $\phi, \psi \in \mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$ then so are $\neg\phi$, $\phi \wedge \psi$ and $\exists x \phi$;

(3) If $\phi(x), \psi(y)$ are formulas in $\mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$, $V \in \Gamma$, $q \in P$, $p < q$, then the formula

$$\mathcal{Q}_V^{(p,q)}[x\,\phi(x), y\,\psi(y)]$$

is in $\mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$.

(The quantifier $\mathcal{Q}_V^{(p,q)}$ binds the variables x and y; any other variable distinct from x and y that shows free in ϕ or ψ is a free variable in $\mathcal{Q}_V^{(p,q)}[x\,\phi(x), y\,\psi(y)]$.)

Semantics: for (1) and (2) is as usual; for (3) consider $w \in \Sigma^+$ and $\mathcal{A}_w$ its associated τ_Σ–structure, $U_1^{\mathcal{A}_w} = \{j \in |\mathcal{A}_w| : \mathcal{A}_w \models \phi(j)\}$ and $U_2^{\mathcal{A}_w} = \{j \in |\mathcal{A}_w| : \mathcal{A}_w \models \psi(j)\}$, then

$$\mathcal{A}_w \models \mathcal{Q}_V^{(p,q)}[x\,\phi(x), y\,\psi(y)] \iff \langle \mathcal{A}_w, U_1^{\mathcal{A}_w}, U_2^{\mathcal{A}_w}\rangle \in \mathcal{Q}_V^{(p,q)}.$$

In the theorem below, and elsewhere, we will use the following "logical shorthands":

$$\begin{gathered} x \le y \text{ for } x = y \vee x < y \\ min.is(x) \text{ for } \forall y(x \le y) \text{ (“}x\text{ is the minimum”)} \\ max.is(x) \text{ for } \forall y(y \le x) \text{ (“}x\text{ is the maximum”)} \\ s(x,y) \text{ for } x < y \wedge \forall z(x < z \longrightarrow y \le z) \text{ (“}y\text{ is the successor of }x\text{”)} \\ [x,y] \longrightarrow \phi \text{ for } \forall z(x \le z \wedge z \le y \longrightarrow \phi) \end{gathered}$$

THEOREM 3.1. *Let* $P \subseteq \mathbb{Z}^+$, $\Gamma \subseteq 2^{\Sigma^+}$, *and* $L \subseteq \Sigma^+$. *If* $L \in \mathcal{AC}(P,\Gamma)$ *then* L *is definable in* $\mathcal{Q}^P_\Gamma[\mathrm{FO}](\tau)$.

Proof. Induction on L. According to the form that L has, we must write a sentence Φ_L in $\mathcal{Q}^P_\Gamma[\mathrm{FO}](\tau)$ such that, for any word $w \in \Sigma^+$,

$$w \in L \iff \mathcal{A}_w \models \Phi_L.$$

If $L = \emptyset$, the corresponding Φ_L is $\exists x \neg(x = x)$.
If $L = \{\sigma\}$, for $\sigma \in \Sigma$, $\Phi_L := \exists x \forall y (x = y \wedge P_\sigma(x))$.
Inductively, assume A and B are languages in $\mathcal{AC}(P,\Gamma)$, and Φ_A and Φ_B are their respective defining sentences.
If $L = A \cup B$, $\Phi_L := \Phi_A \vee \Phi_B$. If $L = \Sigma^+ \setminus A$, $\Phi_L := \neg\Phi_A$.
If $L = AB$,

$$\begin{aligned}\Phi_L \;:=\;& \exists x \exists y \exists z (\mathit{min.is}(x) \wedge \mathit{max.is}(z) \wedge \\ & \forall v (x \leq v \wedge v < y \;\longrightarrow\; \Phi_A) \wedge \\ & \forall v (y \leq v \wedge v \leq z \;\longrightarrow\; \Phi_B)).\end{aligned}$$

Finally, if $L = \langle A, V, B, p, q\rangle$, for some $V \in \Gamma$, $q \in P$ and $p < q$, then

$$\begin{aligned}\Phi_L \;:=\;& \mathcal{Q}^{(p,q)}_V [y_1\,(\exists x_1 \exists x_2 \exists x_3 \exists x_4 (\psi_1(x_1,x_2,x_3,x_4) \wedge s(x_2,y_1))), \\ & \quad y_2\,(\exists x_1 \exists x_2 \exists x_3 \exists x_4 (\psi_1(x_1,x_2,x_3,x_4) \wedge s(y_2,x_3)))] \\ \vee\;& \mathcal{Q}^{(p,q)}_V [y_1\,(\exists x_1 \exists x_2 (\psi_2(x_1,x_2) \wedge \mathit{min.is}(y_1))), \\ & \quad y_2\,(\exists x_1 \exists x_2 (\psi_2(x_1,x_2) \wedge s(y_2,x_1)))] \\ \vee\;& \mathcal{Q}^{(p,q)}_V [y_1\,(\exists x_1 \exists x_2 (\psi_3(x_1,x_2) \wedge s(x_2,y_1))), \\ & \quad y_2\,(\exists x_1 \exists x_2 (\psi_3(x_1,x_2) \wedge \mathit{max.is}(y_2)))]\end{aligned}$$

where

$$\begin{aligned}\psi_1(x_1,x_2,x_3,x_4) \;:=\;& x_1 \leq x_2 \wedge x_2 \leq x_3 \wedge x_3 \leq x_4 \wedge \mathit{min.is}(x_1) \wedge \mathit{max.is}(x_4) \wedge \\ & [x_1,x_2] \;\longrightarrow\; \Phi_A \wedge [x_3,x_4] \;\longrightarrow\; \Phi_B, \\ \psi_2(x_1,x_2) \;:=\;& x_1 \leq x_2 \wedge \mathit{max.is}(x_2) \wedge [x_1,x_2] \;\longrightarrow\; \Phi_B, \text{ and} \\ \psi_3(x_1,x_2) \;:=\;& x_1 \leq x_2 \wedge \mathit{min.is}(x_1) \wedge [x_1,x_2] \;\longrightarrow\; \Phi_A.\end{aligned}$$

The correcteness of the sentence Φ_L can be seen from the following argument:

$$\begin{aligned} w \in L \iff\;& w = vzu = v_1 \cdots v_n z_{n+1} \cdots z_m u_{m+1} \cdots u_k, \text{ with } v \in A, u \in B, z \in V \\ & \text{and the number of these factorizations is congruent to } p \text{ modulo } q.\end{aligned}$$

The sentence Φ_L expresses precisely the above situation:

$$\mathcal{A}_w \models \Phi_L \iff \text{the universe } |\mathcal{A}_w| \text{ is divided in 3 intervals } [\mathit{min}, n],$$

$[r,m],[t,max]$ such that $r=n+1$ and $t=m+1$,
$[min,n] \longrightarrow \Phi_A$ and $[t,max] \longrightarrow \Phi_B$, and
$\langle [r,m], <^{\mathcal{A}_w}, P_\sigma^{\mathcal{A}_w} : \sigma \in \Sigma\rangle \cong \mathcal{A}_z$, with the number of such intervals $[r,m]$ being congruent to p modulo q.

(The sentence Φ_L also takes into consideration the cases of $r = min$ or $m = max$, where min - resp. max - stands for the minimal - resp. maximal - element of the universe. Observe that these elements are definable in FO(τ).) ■

For the converse we need to develop more model theory.

3.1. Isomorphism types

Let $\Gamma \subseteq 2^{\Sigma^+}$ and $P \subseteq \mathbb{Z}^+$. For the remainder of the paper Δ denotes the following alphabet of quantifiers:

$$\Delta := \{\exists, \neg\exists\} \cup \{\mathcal{Q}_V^{(p,q)}, \neg\mathcal{Q}_V^{(p,q)} : V \in \Gamma, q \in P, p < q\}.$$

Fix s a positive integer. We define sets of formulas in s variables in $\mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$ according to the kind of quantifiers that appears in these formulas. Below α denotes a finite word over Δ (i.e., $\alpha \in \Delta^*$) and $\overline{x} = (x_1,\dots,x_s)$ a tuple of s variables.

Let $\alpha = e$ (the empty word). Then $\mathcal{F}^e(\overline{x})$ is the set of all quantifier free formulas in the variables $x_1,\dots,x_s$ in FO(τ).

Let $\alpha = Q\beta$, with $Q \in \Delta$ and $\beta \in \Delta^*$. We have two cases:

Case 1. If $Q = \exists$ or $Q = \neg\exists$, then $\mathcal{F}^\alpha(\overline{x})$ is the closure with respect to the connectives $\vee$ and $\wedge$ of formulas of the form $\exists x_{s+1}\psi(\overline{x},x_{s+1})$ with $\psi \in \mathcal{F}^\beta(\overline{x},x_{s+1})$, or formulas of the form $\neg\exists x_{s+1}\psi(\overline{x},x_{s+1})$ with $\neg\psi \in \mathcal{F}^\beta(\overline{x},x_{s+1})$, or formulas in $\mathcal{F}^\beta(\overline{x})$.

Case 2. If $Q = \mathcal{Q}_V^{(p,q)}$ or $Q = \neg\mathcal{Q}_V^{(p,q)}$, for some $V \in \Gamma$, $q \in P$ and $p < q$, then $\mathcal{F}^\alpha(\overline{x})$ is the closure with respect to the connectives $\vee$ and $\wedge$ of formulas of the form $\mathcal{Q}_V^{(p,q)}[x_{s+1}\,\psi_1(\overline{x},x_{s+1}), y_{s+1}\,\psi_2(\overline{x},y_{s+1})]$ with $\psi_1 \in \mathcal{F}^\beta(\overline{x},x_{s+1})$ and $\psi_2 \in \mathcal{F}^\beta(\overline{x},y_{s+1})$, or formulas of the form $\neg\mathcal{Q}_V^{(p,q)}[x_{s+1}\,\psi_1(\overline{x},x_{s+1}), y_{s+1}\,\psi_2(\overline{x},y_{s+1})]$ with $\neg\psi_1 \in \mathcal{F}^\beta(\overline{x},x_{s+1})$ and $\neg\psi_2 \in \mathcal{F}^\beta(\overline{x},y_{s+1})$, or formulas in $\mathcal{F}^\beta(\overline{x})$.

Now, fix $w \in \Sigma^+$ and let $\mathcal{A}_w$ be its associated τ-structure. Fix s a positive integer, and $\overline{a} = (a_1,\dots,a_s) \in |\mathcal{A}_w|^s$. We look at the maximal consistent set of τ-formulas, with s free variables and given quantifier sequence $\alpha \in \Delta^*$, satisfied by $\overline{a}$. This is the type of $\overline{a}$, denoted by $T^\alpha_{w,\overline{a}}(x_1,\dots,x_s)$, and defined below. As before, $\overline{x}$ stands for the tuple of variables $(x_1,\dots,x_s)$.

If $\alpha = e$,

$$T^e_{w,\overline{a}}(x_1,\dots,x_s) := \{\psi(x_1,\dots,x_s) : \psi \text{ is atomic or negated atomic and } \mathcal{A}_w \models \psi(\overline{a})\}$$

If $\alpha = Q\beta$, with $Q \in \Delta$ and $\beta \in \Delta^*$, we have two cases.

Case 1. $Q \in \{\exists, \neg\exists\}$. Then

$$\begin{aligned} T^{\alpha}_{w,\overline{a}}(x_1,\ldots,x_s) \;:=\; & \{\exists x_{s+1}\psi(x_1,\ldots,x_s,x_{s+1}) : \text{ for some } a \in |\mathcal{A}_w|, \\ & \psi(\overline{x},x_{s+1}) \in T^{\beta}_{w,\overline{a}a}(\overline{x},x_{s+1})\} \cup \\ & \{\neg\exists x_{s+1}\psi(x_1,\ldots,x_s,x_{s+1}) : \text{ for all } a \in |\mathcal{A}_w|, \\ & \neg\psi(\overline{x},x_{s+1}) \in T^{\beta}_{w,\overline{a}a}(\overline{x},x_{s+1})\} \cup \\ & T^{\beta}_{w,\overline{a}}(\overline{x}) \end{aligned}$$

Case 2. $Q \in \{\mathcal{Q}^{(p,q)}_V, \neg\mathcal{Q}^{(p,q)}_V : V \in \Gamma, q \in P, p < q\}$. Then

$$\begin{aligned} T^{\alpha}_{w,\overline{a}}(x_1,\ldots,x_s) \;:=\; & \{\mathcal{Q}^{(p,q)}_V[x_{s+1}\psi_1(\overline{x},x_{s+1}), y_{s+1}\psi_2(\overline{x},y_{s+1})] : \text{ for some set} \\ & S \subseteq \{(i,j) \in |\mathcal{A}_w|^2 : i \leq j\}, |S| = p \,mod\, q \text{ and, for all } (i,j) \\ & \text{in } S,\ \psi_1 \in T^{\beta}_{w,\overline{a}i}(\overline{x},x_{s+1}) \text{ and } \psi_2 \in T^{\beta}_{w,\overline{a}j}(\overline{x},y_{s+1}) \\ & \text{and } \langle [i,j], <^{\mathcal{A}_w}, P^{\mathcal{A}_w}_{\sigma} : \sigma \in \Sigma\rangle \cong \mathcal{A}_z \text{ for some } z \in V\} \cup \\ & \{\neg\mathcal{Q}^{(p,q)}_V[x_{s+1}\psi_1(\overline{x},x_{s+1}), y_{s+1}\psi_2(\overline{x},y_{s+1})] : \text{ for every set} \\ & S \subseteq \{(i,j) \in |\mathcal{A}_w|^2 : i \leq j\}, |S| \neq p \,mod\, q \text{ or, for some } (i,j) \\ & \text{in } S,\ \neg\psi_1 \in T^{\beta}_{w,\overline{a}i}(\overline{x},x_{s+1}) \text{ or } \neg\psi_2 \in T^{\beta}_{w,\overline{a}j}(\overline{x},y_{s+1}) \\ & \text{or } \langle [i,j], <^{\mathcal{A}_w}, P^{\mathcal{A}_w}_{\sigma} : \sigma \in \Sigma\rangle \ncong \mathcal{A}_z \text{ for any } z \in V\} \cup \\ & T^{\beta}_{w,\overline{a}}(\overline{x}) \end{aligned}$$

For fixed $s \geq 0$, there are only finitely many atomic or negated atomic formulas in s many variables. Therefore, inductively, for fixed $s \geq 0$ and quantifier sequence $\alpha \in \Delta^*$, each $T^{\alpha}_{w,\overline{a}}(x_1,\ldots,x_s)$ is a finite set; hence, the conjunction of all formulas in $T^{\alpha}_{w,\overline{a}}(x_1,\ldots,x_s)$ is a well defined formula in $\mathcal{Q}^P_{\Gamma}[\mathrm{FO}](\tau)$. Furthermore,

Proposition 3.1. *For fixed $s \geq 0$ and quantifier sequence $\alpha \in \Delta^*$, the set of formulas $\{\bigwedge T^{\alpha}_{w,\overline{a}}(x_1,\ldots,x_s) : w \in \Sigma^+$ and $\overline{a} \in |\mathcal{A}_w|^s\}$ is finite, up to logical equivalence.*

The proof is a counting argument on all the possible different ways of writing a syntactically correct formula (or well formed formula) using s variables, the quantifiers in α and the finitely many relations in τ.

3.2. Games

Let $P \subseteq \mathbb{Z}^+$, $\Gamma \subseteq 2^{\Sigma^+}$, $\Delta := \{\exists, \neg\exists\} \cup \{\mathcal{Q}^{(p,q)}_V, \neg\mathcal{Q}^{(p,q)}_V : V \in \Gamma, q \in P, p < q\}$, $w, u \in \Sigma^+$, α a finite sequence of quantifiers from Δ. The game on the structures $\mathcal{A}_w$ and $\mathcal{A}_u$ of $|\alpha|$ moves with context in Γ and modulus in P, denoted $G(w,u,\alpha,\Gamma,P)$, is played by Spoiler and Duplicator, who alternate in pebbling elements in each of these structures according to the following rules. Spoiler is always the first to play. A move consists of a play by Spoiler followed by a play by Duplicator. The moves are determined by the letters in the word α, and these are performed by going through α from left to right; hence, say $\alpha = \alpha_1\alpha_2\cdots\alpha_k$ (and, so, $|\alpha| = k$), where each $\alpha_i \in \Delta$, then the first move of the game is α_1, the second move α_2, and so on through α_k. Each move is as follows:

$\exists$ *move:* Spoiler places an unused pebble on an element a in $\mathcal{A}_w$. Duplicator responds by placing an unused pebble on an element b in $\mathcal{A}_u$.

$\neg\exists$ *move:* similar to the $\exists$ move, but Spoiler plays on $\mathcal{A}_u$ and Duplicator plays on $\mathcal{A}_w$.

$\mathcal{Q}_V^{(p,q)}$ *move:* Spoiler selects a set of pairs $S \subseteq \{(x_1, x_2) \in |\mathcal{A}_w|^2 : x_1 \leq^{\mathcal{A}_w} x_2\}$, such that $|S| = p \, mod \, q$. Duplicator selects a set $D \subseteq \{(x_1, x_2) \in |\mathcal{A}_u|^2 : x_1 \leq^{\mathcal{A}_u} x_2\}$, such that $|D| = p \, mod \, q$. Then Spoiler places a pair of pebbles on two elements b_1 and b_2 in $\mathcal{A}_u$. Duplicator responds by placing a pair of pebbles on two elements a_1 and a_2 in $\mathcal{A}_w$, and such that

$$(a_1, a_2) \in S \text{ and } \langle [a_1, a_2], <^{\mathcal{A}_w}, P_\sigma^{\mathcal{A}_w} : \sigma \in \Sigma \rangle \cong \mathcal{A}_z \text{ for some } z \in V$$
$$\text{if, and only if,}$$
$$(b_1, b_2) \in D \text{ and } \langle [b_1, b_2], <^{\mathcal{A}_u}, P_\sigma^{\mathcal{A}_u} : \sigma \in \Sigma \rangle \cong \mathcal{A}_z \text{ for some } z \in V.$$

$\neg\mathcal{Q}_V^{(p,q)}$ *move:* similar to the $\mathcal{Q}_V^{(p,q)}$ move, but with the roles of $\mathcal{A}_w$ and $\mathcal{A}_u$ interchange.

As usual, Duplicator wins if, and only if, at the end of the game (i.e., after the $|\alpha|$-th move), the elements in $\mathcal{A}_w$ and in $\mathcal{A}_u$ where pebbles were placed on, determined a partial isomorphism of $\mathcal{A}_w$ into $\mathcal{A}_u$. Duplicator has a winning strategy if, for every move of Spoiler, he has a corresponding move that will lead him to a win.

Let s be a positive integer. If $\overline{a} \in |\mathcal{A}_w|^s$ and $\overline{b} \in |\mathcal{A}_u|^s$, then $G((w, \overline{a}), (u, \overline{b}), \alpha, \Gamma, P)$ denotes the game on the extended structures $\langle \mathcal{A}_w, \overline{a} \rangle$ and $\langle \mathcal{A}_u, \overline{b} \rangle$ of $|\alpha|$ moves with context in Γ and modulus in P.

3.3. The main logical tool and consequences

THEOREM 3.2. *Let s be a positive integer. Let $w, u \in \Sigma^+$, $\overline{a} = (a_1, \ldots, a_s) \in |\mathcal{A}_w|^s$, $\overline{b} = (b_1, \ldots, b_s) \in |\mathcal{A}_u|^s$, $\Gamma \subseteq 2^{\Sigma^+}$, $P \subseteq \mathbb{Z}^+$ and α a finite sequence of quantifiers from $\Delta := \{\exists, \neg\exists\} \cup \{\mathcal{Q}_V^{(p,q)}, \neg\mathcal{Q}_V^{(p,q)} : V \in \Gamma, q \in P, p < q\}$. The following are equivalent:*

(i)Duplicator has a winning strategy in the game $G((w, \overline{a}), (u, \overline{b}), \alpha, \Gamma, P)$;
(ii)for every formula $\phi(x_1, \ldots, x_s) \in \mathcal{F}^\alpha(x_1, \ldots, x_s)$,

$$\mathcal{A}_w \models \phi(\overline{a}) \iff \mathcal{A}_u \models \phi(\overline{b}) \;;$$

(iii)$\mathcal{A}_u \models \bigwedge T^\alpha_{w,\overline{a}}(\overline{b})$ and $\mathcal{A}_w \models \bigwedge T^\alpha_{u,\overline{b}}(\overline{a})$.

Proof. $[(i) \longrightarrow (ii)]$: By induction on α, being the case $\alpha = e$ trivial. Suppose $|\alpha| > 0$, and without loss of generality we can assume $\alpha = \mathcal{Q}_V^{(p,q)}\beta$ with $\beta \in \Delta^*$ since the other cases have analogous proof. Suppose there exists a formula $\Phi(x_1, \ldots, x_s)$ in $\mathcal{F}^\alpha(x_1, \ldots, x_s)$, such that

$$\mathcal{A}_w \models \Phi(\overline{a}) \text{ and } \mathcal{A}_u \models \neg\Phi(\overline{b}). \tag{1}$$

We will arrive to a contradiction of the fact that Duplicator has a winning strategy in the game $G((w,\overline{a}),(u,\overline{b}),\alpha,\Gamma,P)$. We have that $\Phi(\overline{x})$ is a boolean combination of formulas that are either of the form $\mathcal{Q}_V^{(p,q)}[x_{s+1}\,\psi_1(\overline{x},x_{s+1}),\ y_{s+1}\,\psi_2(\overline{x},y_{s+1})]$, or $\neg\mathcal{Q}_V^{(p,q)}[x_{s+1}\,\theta_1(\overline{x},x_{s+1}),\ y_{s+1}\,\theta_2(\overline{x},y_{s+1})]$, or of the form $\psi(\overline{x})$, with ψ in $\mathcal{F}^\beta(\overline{x})$. By (1), at least one of these formulas that compose $\Phi(\overline{x})$ is true in $\langle\mathcal{A}_w,\overline{a}\rangle$ and false in $\langle\mathcal{A}_u,\overline{b}\rangle$, or vice versa. Suppose it is one of the form $\mathcal{Q}_V^{(p,q)}[x_{s+1}\,\psi_1(\overline{x},x_{s+1}),$ $y_{s+1}\,\psi_2(\overline{x},y_{s+1})]$, and let us denote it by $\Theta(\overline{x})$. Suppose also that

$$\mathcal{A}_w \models \Theta(\overline{a}) \ \text{ and } \ \mathcal{A}_u \models \neg\Theta(\overline{b}).$$

Since $\mathcal{A}_w \models \Theta(\overline{a})$, we have $\langle A_w, U_1^{\mathcal{A}_w}, U_2^{\mathcal{A}_w}\rangle \in \mathcal{Q}_V^{(p,q)}$, where

$$U_1^{\mathcal{A}_w} = \{a_1 \in |\mathcal{A}_w| : \mathcal{A}_w \models \psi_1(\overline{a},a_1)\} \text{ and } U_2^{\mathcal{A}_w} = \{a_2 \in |\mathcal{A}_w| : \mathcal{A}_w \models \psi_2(\overline{a},a_2)\}$$

The Spoiler selects

$$\begin{aligned} S \ = \ & (U_1^{\mathcal{A}_w} \times U_2^{\mathcal{A}_w}) \cap \{(i,j) \in |\mathcal{A}_w|^2 : i \leq^{\mathcal{A}_w} j \text{ and} \\ & \langle [i,j], <^{\mathcal{A}_w}, P_\sigma^{\mathcal{A}_w} : \sigma \in \Sigma\rangle \cong \mathcal{A}_z \text{ for some } z \in V\}. \end{aligned}$$

$S \neq \emptyset$ and $|S| = p \bmod q$. Since $\mathcal{A}_u \models \neg\Theta(\overline{b})$, whatever $D \subseteq \{(y_1,y_2) \in |\mathcal{A}_u|^2 : y_1 \leq^{\mathcal{A}_u} y_2\}$ Duplicator selects (following his winning strategy), there will be a pair $(b_1,b_2) \in D$ such that

$$\mathcal{A}_u \models \neg(\psi_1(\overline{b},b_1) \wedge \psi_2(\overline{b},b_2)) \quad \text{or} \quad \langle [b_1,b_2], <^{\mathcal{A}_u}, P_\sigma^{\mathcal{A}_u} : \sigma \in \Sigma\rangle \not\cong \mathcal{A}_z \text{ for any } z \in V.$$

Spoiler places pebbles on (b_1,b_2). Duplicator responds with some pair (a_1,a_2) in S, according to his winning strategy. Then Duplicator has a winning strategy in the game $G((w,\overline{a},a_1,a_2),(u,\overline{b},b_1,b_2),\beta,\Gamma,P)$, and by inductive hypothesis,

$$\mathcal{A}_w \models \phi(\overline{a},a_1,a_2) \iff \mathcal{A}_u \models \phi(\overline{b},b_1,b_2),$$

for every formula $\phi(\overline{x},x_1,x_2) \in \mathcal{F}^\beta(\overline{x},x_1,x_2)$. Taking $\phi(\overline{x},x_1,x_2) := \psi_1(\overline{x},x_1) \wedge \psi_2(\overline{x},x_2)$ we get the desired contradiction.
Similar reasoning resolves the other cases of $\Theta(\overline{x})$.

$[(ii) \longrightarrow (iii)]$: $\bigwedge T^\alpha_{w,\overline{a}}(x_1,\ldots,x_s)$ is a formula in $\mathcal{F}^\alpha(\overline{x})$ and $\mathcal{A}_w \models \bigwedge T^\alpha_{w,\overline{a}}(\overline{a})$. By (ii), $\mathcal{A}_u \models \bigwedge T^\alpha_{w,\overline{a}}(\overline{b})$. Similarly, $\mathcal{A}_u \models \bigwedge T^\alpha_{u,\overline{b}}(\overline{b})$ and, by (ii), $\mathcal{A}_w \models \bigwedge T^\alpha_{u,\overline{b}}(\overline{a})$.

$[(iii) \longrightarrow (i)]$: By induction on length of α. If $|\alpha| = 0$, we get from the hypothesis that the application $\overline{a} \mapsto \overline{b}$ is a partial isomorphism from $\mathcal{A}_w$ to $\mathcal{A}_u$. Hence, Duplicator wins the game of 0 move. If $|\alpha| > 0$, Duplicator's strategy is described by the equations $\mathcal{A}_u \models \bigwedge T^\alpha_{w,\overline{a}}(\overline{b})$ and $\mathcal{A}_w \models \bigwedge T^\alpha_{u,\overline{b}}(\overline{a})$. Indeed, for whatever elements Spoiler pebbles, Duplicator can compute their type and play accordingly to satisfiability in the structures $\mathcal{A}_w$ and $\mathcal{A}_u$; hence, guaranteeing partial isomorphism among the substructures induced by the pebbled elements. ∎

A consequence of the previous theorem is the following normal form for formulas in the logic $\mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$.

Proposition 3.2. *Let $\phi(x_1,\dots,x_s) \in \mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$ a formula with s free variables and quantifier sequence $\alpha \in \Delta^*$. Then, for all $u \in \Sigma^+$ and $\overline{b} \in |\mathcal{A}_u|^s$,*

$$\mathcal{A}_u \models \phi(\overline{b}) \iff \mathcal{A}_u \models \bigvee\{\bigwedge T^\alpha_{w,\overline{a}}(\overline{b}) : w \in \Sigma^+, \overline{a} \in |\mathcal{A}_w|^s \textit{ and } \mathcal{A}_w \models \phi(\overline{a})\}.$$

Note that by Proposition 3.1 the disjunction on the right of the above equivalence is finite. The proof of Proposition 3.2 is straightforward. Nonetheless, we would like to point out the following connection to lattices and boolean algebras, since we believe that this algebraic point of view enhances the understanding of the already laid logical framework. For the needed background the reader can consult [2].
For fixed $s \geq 0$ and $\alpha \in \Delta^*$, the set of $\mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$–formulas $\mathcal{F}^\alpha(x_1,\dots,x_s)$ is a finite distributive lattice with partial order $\psi \leq \phi$ defined by

$$\begin{gathered}\text{for all } w \in \Sigma^+ \text{ and } \overline{a} \in |\mathcal{A}_w|^s \\ \mathcal{A}_w \models \psi(\overline{a}) \wedge \phi(\overline{a}) \longleftrightarrow \psi(\overline{a})\end{gathered}$$

Then, for each $\phi \in \mathcal{F}^\alpha(x_1,\dots,x_s)$, the set

$$\begin{aligned}T^\alpha_s(\phi) := \{&\bigwedge T^\alpha_{w,\overline{a}}(x_1,\dots,x_s) : w \in \Sigma^+, \\ &\overline{a} \in |\mathcal{A}_w|^s \text{ and } \mathcal{A}_w \models \phi(\overline{a})\}\end{aligned}$$

is the set of all join irreducible elements $\leq \phi$ (because of the equivalence $(ii) \longleftrightarrow (iii)$ in Theorem 3.2). Consequently, Proposition 3.2 is just a particular application of a theorem of representation of elements in any finite lattice.

We recall from [3] that the *ordered sum* of two τ-structures $\mathcal{A}$ and $\mathcal{B}$, is the structure $\mathcal{A} \lhd \mathcal{B}$ with universe $|\mathcal{A}| \cup |\mathcal{B}|$ and relation $R^{\mathcal{A}\lhd\mathcal{B}} := R^\mathcal{A} \cup R^\mathcal{B}$, for all relation symbol $R \in \tau$, except for the order which is set to $<^{\mathcal{A}\lhd\mathcal{B}} := <^\mathcal{A} \cup <^\mathcal{B} \cup \{(a,b) : a \in |\mathcal{A}|, b \in |\mathcal{B}|\}$. An important fact is that, for two words w and v, the concatenation wv corresponds to the ordered sum $\mathcal{A}_w \lhd \mathcal{A}_v$.

Definition 3.1. Let $w, u \in \Sigma^+$, $\Gamma \subseteq 2^{\Sigma^+}$, $P \subseteq \mathbb{Z}^+$ and α a finite sequence of quantifiers from $\Delta = \{\exists, \neg\exists\} \cup \{\mathcal{Q}_V^{(p,q)}, \neg\mathcal{Q}_V^{(p,q)} : V \in \Gamma, q \in P, p < q\}$. We say w *is* (Γ, P, α)*–equivalent to* v, and denote it as $w \equiv_\alpha^{(\Gamma,P)} v$, if and only if for every sentence $\Phi \in \mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$ with quantifier sequence α,

$$\mathcal{A}_w \models \Phi \iff \mathcal{A}_u \models \Phi.$$

(By Theorem 3.2, this is equivalent to verifying if Duplicator has a winning strategy in the game $G(w,u,\alpha,\Gamma,P)$.)

Properties of $\equiv_\alpha^{(\Gamma,P)}$.
- $\equiv_\alpha^{(\Gamma,P)}$ is an equivalence relation.

Follows from definition.
- $\equiv_\alpha^{(\Gamma,P)}$ is right invariant, i.e., if $w \equiv_\alpha^{(\Gamma,P)} v$ then, for any $z \in \Sigma^*$, $wz \equiv_\alpha^{(\Gamma,P)} vz$

Immediate, using the game characterization of $\equiv_\alpha^{(\Gamma,P)}$, and that wz corresponds to the ordered sum $\mathcal{A}_w \lhd \mathcal{A}_z$. Thus, a winning strategy for Duplicator in the game $G(wz, vz, \alpha, \Gamma, P)$ consists on playing his winning strategy for the game $G(w, v, \alpha, \Gamma, P)$, given by hypothesis, whenever Spoiler plays in $\mathcal{A}_w$ or in $\mathcal{A}_v$, and play identical over $\mathcal{A}_z$.

• $\equiv_\alpha^{(\Gamma,P)}$ has finite index.

This is proved using a counting argument similar to the one for Proposition 3.1.

We denote the $\equiv_\alpha^{(\Gamma,P)}$–equivalence classes by $[w]_\alpha^{(\Gamma,P)}$, but usually α and (Γ, P) are clear from context, and in those cases we will just write $[w]$.

We are ready for the converse of Theorem 3.1.

Theorem 3.3. *Let $P \subseteq \mathbb{Z}^+$, $\Gamma \subseteq 2^{\Sigma^+}$, and $L \subseteq \Sigma^+$. If L is definable in $\mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$ then $L \in \mathcal{AC}(P, \Gamma)$.*

Proof. By hypothesis $L = \{w : \mathcal{A}_w \models \Phi\}$ for some $\Phi \in \mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$. If Φ is in $\mathrm{FO}(\tau)$ we proceed as in the literature ([3],[8]); hence, we assume $\Phi := \mathcal{Q}_V^{(p,q)}[x\,\psi_1(x), y\,\psi_2(y)]$ with $\psi_1 \in \mathcal{F}^\beta(x)$ and $\psi_2 \in \mathcal{F}^\gamma(y)$; so, $\Phi \in \mathcal{F}^\alpha$ with $\alpha = \mathcal{Q}_V^{(p,q)}\beta\gamma$. By Proposition 3.2, for all $w \in \Sigma^+$ and $a \in |\mathcal{A}_w|$,

$$\mathcal{A}_w \models \psi_1(a) \iff \mathcal{A}_w \models \widehat{T}^\beta_{w_1,a_1}(a) \vee \widehat{T}^\beta_{w_2,a_2}(a) \vee \ldots \vee \widehat{T}^\beta_{w_k,a_k}(a),$$

where $\widehat{T}^\beta_{w_i,a_i}(x) = \bigwedge T^\beta_{w_i,a_i}(x)$ and $\mathcal{A}_{w_i} \models \psi_1(a_i)$, for $i = 1, 2, \ldots, k$. Intuitively, ψ_1 is in the class of formulas satisfied by some of the structures $\langle \mathcal{A}_{w_1}, a_1\rangle$, $\langle \mathcal{A}_{w_2}, a_2\rangle$, $\ldots$, $\langle \mathcal{A}_{w_k}, a_k\rangle$, in the sense that any other pair (w, a), whose associated structure $\langle \mathcal{A}_w, a\rangle$ satisfies $\psi_1(x)$, is $\equiv_\beta^{(\Gamma,P)}$–equivalent to one of the pairs (w_i, a_i), $i = 1, 2, \ldots, k$ (Theorem 3.2). Similarly, for all $u \in \Sigma^+$ and $b \in |\mathcal{A}_u|$,

$$\mathcal{A}_u \models \psi_2(b) \iff \mathcal{A}_u \models \widehat{T}^\gamma_{u_1,b_1}(b) \vee \widehat{T}^\gamma_{u_2,b_2}(b) \vee \ldots \vee \widehat{T}^\gamma_{u_r,b_r}(b),$$

where $\widehat{T}^\gamma_{u_i,b_i}(x) = \bigwedge T^\gamma_{u_i,b_i}(x)$ and $\mathcal{A}_{u_i} \models \psi_2(b_i)$, for $i = 1, 2, \ldots, r$. So, ψ_2 is in the class of formulas satisfied by some of the structures $\langle \mathcal{A}_{u_1}, b_1\rangle$, $\langle \mathcal{A}_{u_2}, b_2\rangle$, $\ldots$, $\langle \mathcal{A}_{u_r}, b_r\rangle$. Let $W := \{(w_1, a_1), (w_2, a_2), \ldots, (w_k, a_k)\}$ and $U := \{(u_1, b_1), (u_2, b_2), \ldots, (u_r, b_r)\}$. Let

$$\begin{aligned}\widehat{W} &:= \{(w, a) : (w, a) \equiv_\beta^{(\Gamma,P)} (w_i, a_i), \text{ for some } (w_i, a_i) \in W\} \textit{ and}\\ \widehat{U} &:= \{(u, b) : (u, b) \equiv_\gamma^{(\Gamma,P)} (u_i, b_i), \text{ for some } (u_i, b_i) \in U\}.\end{aligned}$$

Let R be the set of triples $((w, a), v, (u, b))$ such that $\mathcal{A}_w \lhd \mathcal{A}_v \lhd \mathcal{A}_u \models \Phi$, $(w, a) \in \widehat{W}$, $(u, b) \in \widehat{U}$, $v \in V$, and $\langle [i, j], <, P_\sigma : \sigma \in \Sigma\rangle \cong \mathcal{A}_v$, where i (resp. j) is the successor of a (resp. predecessor of b) in the order $<^{\mathcal{A}_w \lhd \mathcal{A}_v \lhd \mathcal{A}_u}$. The cardinality of R is at least p modulo q. We then choose p modulo q many pairwise non $\equiv_\alpha^{(\Gamma,P)}$–equivalent triples from R to form the set $\widehat{R}$. Let

$$\Phi_1 := \bigvee\{\bigwedge T^\beta_{w,a}(a+1) : ((w, a), v, (u, b)) \in \widehat{R} \text{ for some } v \in V, (u, b) \in \Sigma^+ \times \mathbb{N}\}$$

and

$$\Phi_2 := \bigvee\{\bigwedge T^\gamma_{u,b}(b-1) : ((w, a), v, (u, b)) \in \widehat{R} \text{ for some } v \in V, (w, a) \in \Sigma^+ \times \mathbb{N}\}$$

($a+1$ and $b-1$ abbreviates, respectively, the successor element of a and the predecessor element of b, in the order $<^{\mathcal{A}_w \lhd \mathcal{A}_v \lhd \mathcal{A}_u}$. Note that these elements are definable in $\mathrm{FO}(\tau)$).

For $i = 1, 2$, let $L_{\Phi_i} = \{w : \mathcal{A}_w \models \Phi_i\}$. Then, for any $z \in \Sigma^+$,

$$\mathcal{A}_z \models \Phi \iff z \in \langle L_{\Phi_1}, V, L_{\Phi_2}, p, q\rangle .$$

This completes the proof of the theorem. ■

4. FROM LOGIC TO REGULAR EXPRESIONS

Theorem 4.1. *Let $P \subseteq \mathbb{Z}^+$, $L \subseteq \Sigma^+$, and $\Gamma \subseteq 2^{\Sigma^+}$. If L is definable in $\mathcal{Q}_\Gamma^P[FO](\tau)$ then L is regular.*

Proof. Let a sentence $\Phi \in \mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$ be such that

$$w \in L \text{ if, and only if, } \mathcal{A}_w \models \Phi.$$

Let α be the quantifier sequence that shows in Φ, and for $w \in \Sigma^*$ let $[w]$ denote the equivalence class of w with respect to the relation $\equiv_\alpha^{(\Gamma,P)}$. Define the DFA

$$\begin{aligned} M_\Phi &:= \langle \Sigma, Q, \delta, s, F\rangle , \text{ where} \\ Q &:= \{[w] : w \in \Sigma^+\} \cup \{s\} \\ s &:= [e] \\ F &:= \{[w] : \mathcal{A}_w \models \Phi\}, \text{ and} \end{aligned}$$

$$\delta : Q \times \Sigma \to Q \text{ is defined as}$$

$$\begin{aligned} \delta(s,\sigma) &:= [\sigma], \text{ for all } \sigma \in \Sigma \\ \delta([w],\sigma) &:= [w\sigma], \text{ for all } w \in \Sigma^+ \text{ and } \sigma \in \Sigma \end{aligned}$$

Then $L = L(M_\Phi)$, so L is regular. ■

An interesting application of Theorem 4.1 is the following. If V is not regular language then, taking $\Gamma = \{V\}$, we get that the language $\langle A, V, B, p, q\rangle$ is regular, for A and B regular languages and p and q arbitrary integers with $p < q$. Thus, one can say that cyclic counting is a "regularizing" operation.

Other particular cases that can be derived from our work, and which were previously known, are the following:

If $\Gamma = \emptyset$ and $P = \emptyset$, then $\mathcal{AC}(P,\Gamma)$ is the class of star-free languages and its associated logic is just $\mathrm{FO}(\tau_\Sigma)$ (see [3]).

If $\Gamma = \{\{\sigma\} : \sigma \in \Sigma\}$ and P is the set of all prime numbers, then $\mathcal{AC}(P,\Gamma)$ is the class of regular languages with corresponding syntactic monoid finite and solvable. The associated logic $\mathcal{Q}_\Gamma^P[\mathrm{FO}](\tau)$ is the extension of first-order logic with cyclic counting quantifiers studied in [8].

Final Remarks. Since, for $\Gamma_1 := \{\{\sigma\} : \sigma \in \Sigma\}$ and P the set of prime integers, we have that the logic $\mathcal{Q}_{\Gamma_1}^P[\mathrm{FO}](\tau)$ defines all regular languages with corresponding syntactic monoid containing only solvable groups, a non solvable regular language would be a set $L \subseteq \Sigma^*$, such that for some $\Gamma \subseteq 2^{\Sigma^+}$ with $\Gamma \neq \Gamma_1$, and some $Q \subseteq \mathbb{Z}^+$, L is definable in $\mathcal{Q}_\Gamma^Q[\mathrm{FO}](\tau)$ but L is not definable in $\mathcal{Q}_{\Gamma_1}^P[\mathrm{FO}](\tau)$. From Theorem 3.2 we get the following characterization for non–definability (Q and Γ are arbitrary):

$$\begin{gathered} L \text{ is not definable in } \mathcal{Q}_\Gamma^Q[\mathrm{FO}](\tau) \text{ if, and only if,} \\ \text{for every } \alpha \in \Delta^* \text{ there exists words } w_\alpha \in L,\ v_\alpha \notin L \text{ with } w_\alpha \equiv_\alpha^{(\Gamma,Q)} v_\alpha. \end{gathered}$$

Further work is to study the structural properties of the operation $\langle A, V, B, p, q\rangle$ for various kinds of context V. In particular to study the algebraic structure of the syntactic monoid associated to $\langle A, V, B, p, q\rangle$ for various V. These ideas are the subject of forth coming papers.

REFERENCES

1. J.R. Büchi, Weak second-order arithmetic and finite automata, *Z. Math. Logik Grundlagen Math.* **6** (1960) 66–92.
2. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, MacMillan (1977).
3. H.D. Ebbinghaus and J. Flum, *Finite Model Theory*, Springer-Verlag (1995).
4. J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages and Computation*, Addison–Wesley, Reading, Massachusetts (1979).
5. P. Lindström, First order predicate logic with generalized quantifiers, *Theoria* **32** (1966) 186–195.
6. R. McNaughton and S. Papert, *Counter–free Automata*, MIT Press (1971).
7. H. Straubing, Families of recognizable sets corresponding to certain varieties of finite monoids, *J. Pure and App. Algebra* **15** (1979) 305–318.
8. H. Straubing, D. Thérien and W. Thomas, Regular languages defined with generalized quantifiers, *in Proc. 15th ICALP, LNCS 317* (1988) 561–575.
9. D. Thérien, Classification of finite monoids: the language approach, *Theoretical Comp. Sci.* **14** (1981) 195–208.