

Characterization of ω -Regular Languages

by First-Order Formulas

(An Extended Abstract)

Kojiro Kobayashi (小林 寿次郎)

Masako Takahashi (高橋 正子)

Hideki Yamasaki (山崎 秀記)

(Tokyo Institute of Technology
Department of Information Science)

1. Introduction

Finite automata have been used to define not only languages but also ω -languages, i.e., sets of ω -words (or ω -sequences) $a_0a_1a_2\dots$ over some alphabet. In the case of usual languages (of finite words), there is the standard notion of acceptance; the one to specify the set of final states. On the other hand, in the case of ω -languages there is a variety of notions of acceptance.

Suppose that $M = (S, \Sigma, \delta, s_0)$ is a (deterministic) finite automaton (S is the set of states, Σ is the input alphabet, δ is the transition function from $S \times \Sigma$ into S , and $s_0 (\in S)$ is the initial state). For an ω -word $\alpha = a_0a_1a_2\dots$ over Σ let $\text{Run}^M(\alpha)$ denote the set of states which M visits in reading α and let $\underline{\text{Run}}^M(\alpha)$ denote the set of states which M visits infinitely often in reading α . Then, in the literature, following ω -languages are used as "the ω -languages accepted by M ":

$$O_1(M, F) = \{\alpha \mid \underline{\text{Run}}^M(\alpha) \cap F \neq \emptyset\},$$

$$O_2(M, F) = \{\alpha \mid \underline{\text{Run}}^M(\alpha) \subseteq F\},$$

$$O_3(M, F) = \{\alpha \mid \underline{\underline{\text{Run}}}^M(\alpha) \cap F \neq \emptyset\},$$

$$O_4(M, F) = \{\alpha \mid \underline{\underline{\text{Run}}}^M(\alpha) \subseteq F\},$$

$$O(M, \mathcal{F}) = \{\alpha \mid \underline{\underline{\text{Run}}}^M(\alpha) \in \mathcal{F}\}.$$

Here $F (\subseteq S)$ is a set of states and \mathcal{F} is a class of sets of states. Let O_1, \dots, O_4, R denote the classes of ω -languages of the forms $O_1(M, F), \dots, O_4(M, F), O(M, \mathcal{F})$ respectively.

There are some other definitions of acceptance (see, for example, [2] - [5]) but the resulting classes of ω -languages coincide with some of O_1, \dots, O_4, R or their Boolean combinations.

Now a question arises: Can we think of any other kind of acceptance to define a new interesting class of ω -languages? Our work towards this question is motivated by the observation that all the conditions mentioned above and some others can be expressed by certain first-order formulas. For example, the condition $\underline{\text{Run}}^M(\alpha) \cap F \neq \emptyset$ used to define $O_3(M, F)$ can be expressed as $\bigvee_{s \in F} \forall i \exists j (i < j \wedge P_s(j))$, where $P_s(j)$ means "M is in state s at time j ."

The first problem we are concerned with in this paper is: When we define the acceptance of finite automata with these first-order formulas, what do we get as the ω -languages accepted thereby? After providing necessary definitions in the next section, in section 3 we show that the class of such ω -languages is precisely the class R .

The next problem we are interested in is a hierarchy of the ω -languages based on the complexity of the formulas. We define

classes of ω -languages Σ_n^{fa} and Π_n^{fa} for $n \geq 0$ as the arithmetic hierarchy in the theory of recursive functions, based on the alternation of the quantifiers \exists and \forall in the defining formulas in prenex normal form. We show in sections 3 and 4 that $\Sigma_1^{\text{fa}} = \mathbf{0}_1$, $\Pi_1^{\text{fa}} = \mathbf{0}_2$, $\Sigma_2^{\text{fa}} = \mathbf{0}_4$, $\Pi_2^{\text{fa}} = \mathbf{0}_3$, $\Sigma_3^{\text{fa}} = \Sigma_4^{\text{fa}} = \dots = \Pi_3^{\text{fa}} = \Pi_4^{\text{fa}} = \dots = \mathbf{R}$.

From these results it seems reasonable to say that those (and possibly their Boolean combinations) exhaust the natural ways of acceptance of ω -languages by (deterministic) finite automata. In section 5, we remark on a "machine-independent" version of the first-order description of ω -languages.

2. Preliminaries

In the set Σ^* of all ω -words over Σ we can define a topology in a natural way, that is the product topology of the discrete topology on Σ . Let \mathbf{G} (\mathbf{F} , resp.) be the class of open sets (closed sets), \mathbf{G}' (\mathbf{F}' , resp.) be the class of denumerable intersections (unions) of open sets (closed sets), and \mathbf{G}'' (\mathbf{F}'' , resp.) be the class of denumerable unions (intersections) of sets in \mathbf{G}' (\mathbf{F}').

Then $\mathbf{0}_1, \dots, \mathbf{0}_4$ are characterized as $\mathbf{0}_1 = \mathbf{G} \cap \mathbf{R}$, $\mathbf{0}_2 = \mathbf{F} \cap \mathbf{R}$, $\mathbf{0}_3 = \mathbf{G}' \cap \mathbf{R}$, $\mathbf{0}_4 = \mathbf{F}' \cap \mathbf{R}$, and \mathbf{R} is included in $\mathbf{G}'' \cap \mathbf{F}''$. The members of \mathbf{R} are called ω -regular languages. We refer the reader to [1], [5], [6] for other properties of classes $\mathbf{0}_1, \dots, \mathbf{0}_4, \mathbf{R}$.

With each finite automaton $M = (S, \Sigma, \delta, s_0)$ we associate a class of first-order formulas, called M -formulas, in the following way. Terms are expressions of the forms $i + n$ or n ,

where i is a variable and n is a natural number. Atomic M-formulas are expressions of the forms $P_s(t)$, $t = t'$, $t < t'$, where s is a state of M and t, t' are terms. Finally, M-formulas are constructed from atomic M-formulas using propositional connectives \vee (or), \wedge (and), \neg (not), \rightarrow (implication), \leftrightarrow (equivalence), and quantifiers \exists, \forall . We may simply say "formulas" instead of M-formulas provided there is no fear of confusion.

Suppose ϕ is an M-formula, and α is an ω -word over Σ . We define the α -interpretation of ϕ as follows: The domain of the interpretation is the set $N = \{0, 1, 2, \dots\}$ of natural numbers. The terms $i + n, n$ and atomic M-formulas $t = t', t < t'$ are interpreted as usual. An atomic M-formula $P_s(t)$ is interpreted as "reading α , M is in state s at time t ."

For any closed M-formula (i.e., an M-formula without free variables) ϕ , we say ϕ defines the ω -language

$$A(M, \phi) = \{\alpha \in \Sigma^\omega \mid$$

ϕ is true under the α -interpretation\}.

Finally we define types $\Sigma_n^{fa}, \Pi_n^{fa}$ ($n = 0, 1, 2, \dots$) of M-formulas. Suppose ϕ is an M-formula of the form

$$\exists i_{1,1} \exists i_{1,2} \dots \exists i_{1,e_1} \forall i_{2,1} \forall i_{2,2} \dots \forall i_{2,e_2} \\ \exists i_{3,1} \exists i_{3,2} \dots \exists i_{3,e_3} \dots Q i_{n,1} Q i_{n,2} \dots Q i_{n,e_n} \psi$$

where $e_p \geq 1$ ($1 \leq p \leq n$), $i_{1,1}, \dots, i_{n,e_n}$ are different variables, Q is \forall or \exists according as n is even or odd, and ψ is an open M-formula (that is, an M-formula which contains no quantifiers). Then we say ϕ is an M-formula of type Σ_n^{fa} . An M-formula of type Π_n^{fa} is similarly defined with \exists and \forall replaced each other. In particular, we say open M-formulas are

of type Σ_0^{fa} and Π_0^{fa} . For each $n \geq 0$, we denote by Σ_n^{fa} (or Π_n^{fa} , resp.) the class of ω -languages of the form $A(M, \phi)$ with M -formulas ϕ of type Σ_n^{fa} (or Π_n^{fa}), and denote the class $\Sigma_n^{fa} \cap \Pi_n^{fa}$ by Δ_n^{fa} .

3. Characterization of \mathbf{R}

In this section we show that ω -regular languages are precisely the ω -languages defined by M -formulas, and the class of them is equal to Σ_3^{fa} and Π_3^{fa} .

Theorem 1. For any closed M -formula ϕ , $A(M, \phi)$ is an ω -regular language.

Proof. We first extend the definition of $A(M, \phi)$ to all M -formulas ϕ (rather than just for closed M -formulas).

Suppose that $\phi(i_1, \dots, i_n)$ is an M -formula that has n free variables i_1, \dots, i_n ($n \geq 0$). Let $\Gamma = \{0, 1\}$ and let Γ_n be the alphabet $\Sigma \times \Gamma^n = \{(a, b_1, \dots, b_n) \mid a \in \Sigma, b_1, \dots, b_n \in \Gamma\}$. For an ω -word α on Σ and ω -words β_1, \dots, β_n on Γ let $\langle \alpha, \beta_1, \dots, \beta_n \rangle$ denote the ω -word γ on Γ_n such that $\gamma(m) = (\alpha(m), \beta_1(m), \dots, \beta_n(m))$ ($0 \leq m$). Here, if $\alpha = a_0 a_1 a_2 \dots$ then $\alpha(m)$ denotes a_m (that is, $\alpha(m)$ denotes the m th symbol of α) and similarly for $\beta_1(m), \dots, \beta_n(m), \gamma(m)$.

$A(M, \phi)$ denotes the ω -language

$$\{\langle \alpha, 0^{m_1} \beta_1, \dots, 0^{m_n} \beta_n \rangle \mid$$

$$\alpha \in \Sigma^\omega, 0 \leq m_1, \dots, 0 \leq m_n, \beta_1, \dots, \beta_n \in \Gamma^\omega,$$

$$\phi(m_1, \dots, m_n) \text{ is true under the } \alpha\text{-interpretation}\}.$$

Then we prove the theorem for M -formulas ϕ that are not necessarily closed. The proof is by induction on the structure of ϕ . It uses the fact that (1) ω -languages of the form $L\Sigma^\omega$

with (usual) regular languages $L (\subseteq \Sigma^*)$ are in \mathbf{R} , and that (2) \mathbf{R} is closed under Boolean operations and projections (that is, length preserving homomorphisms). We omit the details of the proof. \square

The converse of theorem 1 is also true.

Theorem 2. For any ω -regular language A , there exists a closed M-formula ϕ such that $A(M, \phi) = A$.

Proof. Suppose that $A = A(M, \phi)$. Then we have

$$\begin{aligned} A &= \{ \alpha \in \Sigma^\omega \mid \underline{\text{Run}}^M(\alpha) = F \text{ for some } F \in \mathcal{F} \} \\ &= A(M, \bigvee_{F \in \mathcal{F}} (\exists i \forall j (j \geq i \rightarrow \bigvee_{s \in F} P_s(j)) \\ &\quad \wedge \bigwedge_{s \in F} \forall j \exists k (k \geq j \wedge P_s(k)))). \quad \square \end{aligned}$$

We can refine the above result as follows.

Theorem 3. For any $A \in \mathbf{R}$ there are finite automata M, M' and open formulas ψ, ψ' having free variables i, j, k such that

$$\begin{aligned} A &= A(M, \exists i \forall j \exists k \psi(i, j, k)) \\ &= A(M', \forall i \exists j \forall k \psi'(i, j, k)). \end{aligned}$$

Proof. We will prove only for $\exists i \forall j \exists k \psi(i, j, k)$ because \mathbf{R} is closed under the operation of complement.

Suppose that $A = O(M_1, \mathcal{F})$ with $M_1 = (S_1, \Sigma, \delta, s_{01})$. Let n be $\max\{|\mathcal{F}|, |S_1|\}$, where $||$ denotes the cardinality of sets.

We construct another finite automaton $M_2 = (S_2, \Sigma, \delta_2, s_{02})$. Intuitively, M_2 simulates M_1 , and at the same time counts the time with modulo n . Formally, S_2 is the set $\{(s, h) \mid s \in S_1, 1 \leq h \leq n\}$, s_{02} is $(s_{01}, 1)$, and δ_2 is defined by $\delta_2((s, h), a) = (\delta_1(s, a), h')$, where $h' = h + 1$ if $h \leq n - 1$ and $h' = 1$ if $h = n$.

For each $F \in \mathcal{F}$ and $s \in S_1$ we can construct oper

M_2 -formulas $\chi_{1,F}(i)$, $\chi_{2,s}(j)$, $\chi_{3,s}(k)$ having the following intuitive meaning:

$\chi_{1,F}(i) \iff$ at time i , the second component h of the state (s, h) of M_2 is such that $h \leq |\mathcal{J}|$ and F is the h th element of \mathcal{J} ;

$\chi_{2,s}(j) \iff$ at time j , the second component h of the state (s, h) of M_2 is such that $h \leq |S_1|$ and s is the h th element of S_1 ;

$\chi_{3,s}(k) \iff$ at time k , s is the first component of the state (s, h) of M_2 .

Then, for each α , we have the following equivalence.

$$\begin{aligned}
& A \\
&= A(M_1, \bigvee_{F \in \mathcal{J}} (\exists i \forall j (j \geq i \rightarrow \bigvee_{S \in F} P_S(j)) \\
&\quad \wedge \bigwedge_{S' \in F} \forall j \exists k (k \geq j \wedge P_{S'}(k))) \\
&= A(M_2, \bigvee_{F \in \mathcal{J}} (\exists i \forall j (j \geq i \rightarrow \bigvee_{S \in F} \chi_{3,S}(j)) \\
&\quad \wedge \bigwedge_{S' \in F} \forall j \exists k (k \geq j \wedge \chi_{3,S'}(k))) \\
&= A(M_2, \bigvee_{F \in \mathcal{J}} \exists i (\forall j (j \geq i \rightarrow \bigvee_{S \in F} \chi_{3,S}(j)) \\
&\quad \wedge \bigwedge_{S' \in F} \forall j \exists k (k \geq j \wedge \chi_{3,S'}(k)))) \\
&= A(M_2, \exists i (\forall j (j \geq i \rightarrow \bigvee_{F \in \mathcal{J}} (\chi_{1,F}(i) \wedge \bigvee_{S \in F} \chi_{3,S}(j)) \\
&\quad \wedge \forall j \exists k (k \geq j \wedge \\
&\quad \bigvee_{F \in \mathcal{J}} (\chi_{1,F}(i) \wedge \bigwedge_{S' \in F} (\chi_{2,S'}(j) \rightarrow \chi_{3,S'}(k))))) \\
&= A(M_2, \exists i \forall j \exists k ((j \geq i \rightarrow \bigvee_{F \in \mathcal{J}} (\chi_{1,F}(i) \wedge \bigvee_{S \in F} \chi_{3,S}(j)) \\
&\quad \wedge (k \geq j \wedge \\
&\quad \bigvee_{F \in \mathcal{J}} (\chi_{1,F}(i) \wedge \bigwedge_{S' \in F} (\chi_{2,S'}(j) \rightarrow \chi_{3,S'}(k))))) .
\end{aligned}$$

The last M_2 -formula is of the desired form. \square

Corollary 4. For any ω -language A the following six conditions are equivalent.

$$(1) A \in \mathbf{R}.$$

(2) $A \in \Lambda_n^{fa}$ ($n \geq 3$).

(3) $A \in \Sigma_n^{fa}$ ($n \geq 3$).

(4) $A \in \Pi_n^{fa}$ ($n \geq 3$).

(5) $A = A(M, \exists i \forall j \exists k \psi)$ for a finite automaton M and an open M -formula ψ .

(6) $A = A(M, \forall i \exists j \forall k \psi)$ for a finite automaton M and an open M -formula ψ .

In closing this section, we mention a characterization of \mathbf{R} which is a direct consequence of the proof of Theorem 1. Corollary 5. The class of ω -regular languages is the closure of \mathbf{O}_1 (or \mathbf{O}_2) under Boolean operations and projections.

Compare the corollary with the fact that \mathbf{R} is the class of projective images of the sets in \mathbf{O}_3 , and also is the Boolean closure of \mathbf{O}_4 (see e.g., [1]).

4. Characterization of $\mathbf{O}_1, \mathbf{O}_2, \mathbf{O}_3, \mathbf{O}_4$

In this section we characterize $\mathbf{O}_1, \mathbf{O}_2, \mathbf{O}_3, \mathbf{O}_4$ by means of forms of M -formulas.

Theorem 6. For any ω -language A the following three conditions are equivalent.

(1) $A \in \mathbf{O}_1$.

(2) $A \in \Sigma_1^{fa}$.

(3) $A = A(M, \exists i \psi)$ for a finite automaton M and an open M -formula ψ .

A similar result holds when $\mathbf{O}_1, \Sigma_1^{fa}, \exists$ are replaced by $\mathbf{O}_2, \Pi_1^{fa}, \forall$, respectively.

Proof. We only prove the part (2) \Rightarrow (1) for \mathbf{O}_1 . Let $\exists i_1 \dots$

$\exists i_p \psi(i_1, \dots, i_p)$ be a closed M-formula defining A, where ψ is open. We may assume that ψ does not contain \neg because we can eliminate \neg by $\neg P_S(t) \Leftrightarrow \bigvee_{s' \neq s} P_{S'}(t)$, $\neg t < t' \Leftrightarrow t' < t \vee t = t'$, $\neg t = t' \Leftrightarrow t' < t \vee t < t'$.

The proof of Theorem 1 shows that if ϕ is an atomic M-formula then $A(M, \phi)$ is in \mathbf{G} (that is, open). Hence, for any numbers m_1, \dots, m_p , $A(M, \psi(m_1, \dots, m_p))$ is obtained from sets in \mathbf{G} by finite union and finite intersection. Hence it is in \mathbf{G} . Therefore we have

$$\begin{aligned} A &= A(M, \exists i_1 \dots \exists i_p \psi(i_1, \dots, i_p)) \\ &= \cup A(M, \psi(m_1, \dots, m_p)) \\ &\in \mathbf{G} \cap \mathbf{R} = \mathbf{O}_1 \end{aligned}$$

(\cup ranges over all m_1, \dots, m_p). \square

Corollary 7. For any ω -language A in Σ^ω the following four conditions are equivalent.

- (1) $A \in \mathbf{O}_1 \cap \mathbf{O}_2$.
- (2) $A \in \mathbf{\Lambda}_1^{fa}$.
- (3) $A \in \mathbf{\Lambda}_0^{fa} (= \mathbf{\Sigma}_0^{fa} = \mathbf{\Pi}_0^{fa})$.
- (4) $A = L\Sigma^\omega$ for a finite subset L of Σ^* .

Proof. For the equivalence of (1) and (4), see, e.g., [6]. \square

Theorem 8. For any ω -language A the following three conditions are equivalent.

- (1) $A \in \mathbf{O}_3$.
- (2) $A \in \mathbf{\Pi}_2^{fa}$.
- (3) $A = A(M, \forall i \exists j \psi)$ for a finite automaton M and an open M-formula ψ .

A similar result holds when $\mathbf{0}_3$, Π_2^{fa} , \forall , \exists are replaced with $\mathbf{0}_4$, Σ_2^{fa} , \exists , \forall respectively.

Proof of (2) \Rightarrow (1) for $\mathbf{0}_3$. The proof is the same as that of (2) \Rightarrow (1) of Theorem 6 except that we use

$$\begin{aligned} A &= A(M, \forall i_1 \dots \forall i_p \exists j_1 \dots \exists j_q \\ &\quad \psi(i_1, \dots, i_p, j_1, \dots, j_q)) \\ &= \cap \cup A(M, \psi(m_1, \dots, m_p, n_1, \dots, n_q)) \\ &\in \mathbf{G}' \cap \mathbf{R} = \mathbf{0}_3 \end{aligned}$$

(\cap is over all m_1, \dots, m_p and \cup is over all n_1, \dots, n_q). \square

Corollary 9. For any ω -language A the following three conditions are equivalent.

(1) $A \in \mathbf{0}_3 \cap \mathbf{0}_4$.

(2) $A \in \mathbf{A}_2^{fa}$.

(3) $A = \{\alpha \in \Sigma^\omega \mid \text{Run}^M(\alpha) \in \mathcal{F}\}$ for a finite automaton $M = (S, \Sigma, \delta, s_0)$ and a class \mathcal{F} of subsets of S .

Proof. The equivalence of (1) and (3) was shown by Staiger and Wagner ([5]). \square

5. Concluding Remarks

In this paper we introduced the M -formulas associated with a finite automaton M , and studied classes of ω -languages accepted in the ways specified by the M -formulas. The result can be restated in terms of "machine-independent" first-order description of ω -languages.

For a given sequence $\mathbf{L} = (L_1, L_2, \dots, L_m)$ of regular languages ($\subseteq \Sigma^*$, $m \geq 1$), let us define \mathbf{L} -formulas as the first-order formulas which are constructed from the atomic

formulas of the forms $\alpha[t] \in L_p$ ($1 \leq p \leq m$), $t = t'$, and $t < t'$ where t and t' are either $i + n$ or n for a variable i and a natural number n . In other words, **L**-formulas are same as the **M**-formulas except that atomic formulas $P_s(t)$ ($s \in S$) are now replaced by $\alpha[t] \in L_p$ ($1 \leq p \leq m$). For a given ω -word α over Σ , we define the α -interpretation in the same way as for **M**-formulas except that the meaning of $\alpha[t] \in L$ is "the prefix of α of length t is in L ."

For a closed **L**-formula ϕ we define

$$A(\mathbf{L}, \phi) = \{ \alpha \in \Sigma^\omega \mid \alpha \text{ is true under the } \alpha\text{-interpretation} \}.$$

Lastly we define the types Σ_n^{reg} and Π_n^{reg} ($n \geq 0$) of **L**-formulas in prenex normal form as in the case of Σ_n^{fa} and Π_n^{fa} (except that "**M**-formulas" are now replaced by "**L**-formulas"). Under these definitions we can easily see the followings.

(1) For a given finite automaton M and a closed **M**-formula ϕ of type Σ_n^{fa} (or Π_n^{fa}), there exists a sequence $\mathbf{L} = (L_1, L_2, \dots, L_m)$ of regular languages and an **L**-formula ϕ' of type Σ_n^{reg} (or Π_n^{reg}) such that $A(M, \phi) = A(\mathbf{L}, \phi')$.

(2) Conversely for a given sequence $\mathbf{L} = (L_1, L_2, \dots, L_m)$ of regular languages and a closed **L**-formula ϕ of type Σ_n^{reg} (or Π_n^{reg}) one can construct a finite automaton M and an **M**-formula ϕ' of type Σ_n^{fa} (or Π_n^{fa}) such that $A(\mathbf{L}, \phi) = A(M, \phi')$.

Therefore we can get characterizations of **R** and its subclasses $\mathbf{O}_1 \cap \mathbf{O}_2$, \mathbf{O}_1 , \mathbf{O}_2 , $\mathbf{O}_3 \cap \mathbf{O}_4$, \mathbf{O}_3 , \mathbf{O}_4 by means of **L**-formulas of the types Σ_n^{reg} and Π_n^{reg} exactly as before. This can be seen as a generalization of the characterization (or definition) of $\mathbf{O}_1 - \mathbf{O}_4$ stated in terms of regular languages;

$$\begin{aligned}
 A \in \mathbf{O}_1 &\iff A = \{\alpha \in \Sigma^\omega \mid \exists i (\alpha[i] \in L)\}, \\
 A \in \mathbf{O}_2 &\iff A = \{\alpha \in \Sigma^\omega \mid \forall i (\alpha[i] \in L)\}, \\
 A \in \mathbf{O}_3 &\iff A = \{\alpha \in \Sigma^\omega \mid \forall i \exists j (i < j \wedge \alpha[j] \in L)\}, \\
 A \in \mathbf{O}_4 &\iff A = \{\alpha \in \Sigma^\omega \mid \exists i \forall j (i < j \rightarrow \alpha[j] \in L)\}.
 \end{aligned}$$

References

- [1] S. Eilenberg, Automata, Languages, and Machines, Vol. A, Academic Press, 1974.
- [2] L. Landweber, Decision problems for ω -automata, Mathematical Systems Theory 3 (1969) 376 - 384.
- [3] R. McNaughton, Testing and generating infinite sequences by a finite automaton, Information and Control 9 (1966) 521 - 530.
- [4] M. O. Rabin, Automata on Infinite Objects and Church's Problem, Regional Conference Series in Mathematics, No. 13, AMS, 1969.
- [5] L. Staiger and K. Wagner, Automatentheoretische und automatenfreie Charakterisierungen topologischer Klassen regularer Folgenmengen, Elektronische Informationsverarbeitung und Kybernetik 10 (1974) 379 - 392.
- [6] M. Takahashi and H. Yamasaki, A note on ω -regular languages, Theoretical Computer Science 23 (1983), to appear.