LABORATOIRE

INFORMATIQUE, SIGNAUX ET SYSTÈMES
DE SOPHIA ANTIPOLIS
UMR 6070

AMBIGUITY OF INFINITE WORDS

Sandrine Julia

Equipe MC3

Rapport de recherche
ISRN I3S/RR–2008-07–FR

April 2008

Laboratoire d'Informatique de Signaux et Systèmes de Sophia Antipolis - UNSA-CNRS
2000, rte.des Lucioles – Les Algorithmes – Bât Euclide B – B.P. 121 – 06903 Sophia-Antipolis Cedex – France
Tél.: 33 (0)4 92 94 27 01 – Fax: 33 (0)4 92 94 28 98 – www.i3s.unice.fr
UMR6070

Ambiguity of infinite words

Sandrine JULIA

Université de Nice - Sophia Antipolis,
Laboratoire I3S - CNRS, B.P. 121,
06903 Sophia Antipolis Cedex, France.
`Sandrine.Julia@unice.fr`

Abstract. The ambiguous words in L^+ prevent the language L from being a code. Respectively, the ambiguous infinite words in L^ω prevent a language L from being an ω-code. This paper is an extensive study of these two sets. We clarify their structures and also their links. In addition, we propose a restricted notion of ambiguity, the strong ambiguity coupled to its infinitary version. The aim is to avoid synchronizing words which are not actually responsible of ambiguity or ω-ambiguity. Finding appropriate representatives for them constitutes a challenge to go further inside open decidability problems like deciding whether a rational ω-language is generated by a code or an ω-code.

Introduction

Our research deals with the classical theory of automata and languages. We particularly focus on the rational languages of infinite words, called *ω-languages* [10]. The precise topic is the ambiguity of words and especially the ambiguity of infinite words. A language L is a *code* if and only if every word in L^+ has a unique factorization over L [1]. A language L is an *ω-code*, or originally an *ifl-code*, if and only if every ω-word in L^ω has a unique factorization over L [11]. In a dual way, a language L is said to be *ambiguous* if it fails to be a code and an ω-language is said to be *ω-ambiguous* if it fails to be an ω-code.

For a language L, it is usual to represent ambiguity and ω-ambiguity of L by the respective sets $Amb(L^+)$ and $Amb(L^\omega)$ [6] [7]. The set $Amb(L^+)$ consists of words in L^+ that have at least two different factorizations with different first steps over L. Respectively, the set $Amb(L^\omega)$ is made up of words in L^ω that admit at least two different factorizations with different first steps over L. Hence, a language is a code as soon as $Amb(L^+)$ is empty and is an ω-code when $Amb(L^\omega)$ is empty.

Furthermore, our interest extends to open decidability problems concerning ambiguity, like how to decide, for a given rational language L, whether there exists a code or an ω-code C such that $C^\omega = L^\omega$. The operation $^\omega$ stands for the infinite concatenation. Our idea is firstly to deepen our knowledge of ambiguity and ω-ambiguity. The challenge is to find canonical representatives and to better identify them. Beyond a method to approach such problems, as yet, affirmative answers which are known concerns only languages whose ambiguity and ω-ambiguity seem to be minimal [8][5].

So we focus on languages highly representative of their ω-power L^ω. A rational ω-power is characterized by the language $\chi(L^\omega)$ [4]. The deepest way to decompose an ω-word in L^ω happens when L^+ coincides with the set $\chi(L^\omega)$. This is the reason why our study deals with languages L equal to the root of $\chi(L^\omega)$.

Our work is very gradual. After some basic definitions, the preliminary section introduces the concepts of *relations* and *ω-relations* which are based on the relabelling of the words in a language. The benefit is to handle more easily distinct factorizations of words and ω-words. Then Section 2 is dedicated to the investigation of the sets of ambiguous words and ω-words including their connections. Particularly, for finite languages, a link is done between ambiguity and ω-ambiguity by means of a partition of the set $Amb(L^\omega)$. Section 3 proposes restrictive definitions such as *strong ambiguity* and *strong ω-ambiguity*. The idea is to ignore steps in factorizations as far as they are not concerned with ambiguity since they are *synchronising*. More precisely, some words can be involved in ω-ambiguity without contributing at all to ambiguity. Therefore, a concise way to describe both ambiguities rises from these restrictive new definitions. In conclusion, we explain our interest in understanding *minimal* ambiguity and *minimal* ω-ambiguity as well. The goal is to identify cases for which the related decidability problems have reachable solutions.

1 Preliminaries

1.1 Basic definitions

Let Σ be a finite alphabet. A *word* is a finite concatenation of letters in Σ and an *ω-word* is an infinite one. We note ε the empty word. Σ^* is the set of words over Σ and Σ^ω is the set of ω-words. Σ^+ denotes $\Sigma^* \setminus \{\varepsilon\}$. A *language* is a subset of Σ^* and an *ω-language* is a subset of Σ^ω.
A word u is a *prefix* of v in $(\Sigma^* \cup \Sigma^\omega)$ if $v \in u(\Sigma^* \cup \Sigma^\omega)$. The *prefix order* is denoted by $<$. *Pref*(v) stands for the set of the prefixes of v. If L is a language or an ω-language, *Pref*(L) gathers the prefixes of the elements in L. Let L be a language, L^* is defined as $L^* = \{\varepsilon\} \cup \left(\bigcup_{n>0}\{a_1...a_n/\ \forall i\ 1 \leq i \leq n,\ a_i \in L\}\right)$ and the *ω-power* L^ω is defined as $L^\omega = \{w = a_1...a_n.../\ \forall i > 0,\ a_i \in L \setminus \{\varepsilon\}\}$.
The *root* of L^* is the language $Root(L^*) = (L^* \setminus \{\varepsilon\}) \setminus (L^* \setminus \{\varepsilon\})^2$. The *limit* $\overrightarrow{L}$ of L is the ω-language $\overrightarrow{L} = \{w \in \Sigma^\omega,\ Card(Pref(w) \cap L) \textit{ is infinite}\}$ [3].
A *L-factorization* of a word u in L^+ is a finite sequence of words in $L \setminus \{\varepsilon\}$: $(u_1, u_2, ..., u_n)$ such that $u = u_1u_2...u_n$. A L-factorization of an ω-word w in L^ω is an infinite sequence: $(w_1, w_2, ..., w_n, ...)$ such that $w = w_1w_2...w_n...$. We will say indifferently L-factorization or factorization over L.
Rational languages are recognized by finite automata. In a *Büchi automaton*, an ω-word is accepted if its run enters infinitely often in a recognition state. An ω-language is rational if it is recognized by a *Büchi automaton.* Moreover, a rational language is deterministic if it is recognized by a deterministic Büchi automaton [10].

For more convenience, rational languages and ω-languages will be denoted in the sequel by their regular expressions or ω-regular expressions.

1.2 Relabelling

We consider a relabelling of words in a given language L in order to handle more easily distinct factorizations of words and ω-words. Let A and Σ be two disjoint alphabets and let L be a language over Σ. A relabelling of the words in L is a one-to-one mapping $\tilde{}$ from $\Lambda \subseteq A$ over $L \subseteq \Sigma^+$ extended in the canonical one-to-one morphism from $(\Lambda^+, \cdot)$ over $(L^+, \cdot)$ where $\cdot$ denotes the concatenation operation.
Hence, a L-factorization $(u_1, u_2, ..., u_n)$ is associated to a word $\varphi = \varphi_1 ... \varphi_n$ in Λ^+ such that $\forall i,\ u_i = \tilde{\varphi}_i$ (idem for a factorization of an ω-word).

Definition 1. *Let L be a language associated to the relabelling Λ. The relation $\simeq$ is a binary relation over Λ^+ stating that $\varphi \simeq \psi$ if and only if there exists a word v in L^+ and two L-factorizations with different first steps corresponding respectively to φ and ψ such that: $v = \tilde{\varphi} = \tilde{\psi}$.*

By analogy, we define the relation concerning the factorizations of ω-words:

Definition 2. *Let L be a language associated to the relabelling Λ. The relation $\simeq_\omega$ is a binary relation over Λ^ω stating that $\varphi \simeq_\omega \psi$ if and only if there exists an ω-word w in L^ω and two L-factorizations with different first steps corresponding respectively to φ and ψ such that: $w = \tilde{\varphi} = \tilde{\psi}$.*

In the sequel, we call *a relation* an item $\varphi \simeq \psi$ over Λ^+ and an *ω-relation* an item $\varphi \simeq_\omega \psi$ over Λ^ω. The factorizations are required with different first steps and it makes the sets *Pref*(φ) and *Pref*(ψ) necessarily disjoint.
To make things absolutely clear, we give a definition for the intuitive notion of two factorizations that never *meet*. We call them *non-coinciding* factorizations.

Definition 3. *Let L be a language associated to the relabelling Λ. Consider two factorizations with different first steps of a word in L^+, respectively associated to φ and ψ in Λ^+. Those factorizations are said non-coinciding if and only if there exists no couple (μ, ν) in $(Pref(\varphi) \times Pref(\psi))$ verifying $\tilde{\mu} = \tilde{\nu}$. Idem for two factorizations of an ω-word in L^ω.*

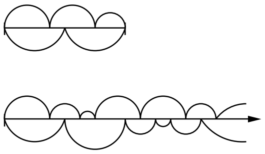

Fig. 1. Two couples of non-coinciding factorizations over L^+ and L^ω.

Such factorizations are illustrated in Figure 1 in the finitary and infinitary cases.

1.3 Codes and ω-codes

A language L is a *code* if every word $u \in \Sigma^*$ has at most one L-factorization [1]. Thus, no relations are verified between elements in L^+. A language L is an ω-*code* if every ω-word in Σ^ω has at most one L-factorization [11]. Consequently, no ω-relation holds over L^ω. Any ω-code is *a fortiori* a code.
A language, if it is not a code, is ambiguous. So, an infinity of relations and an infinity of ω-relations rise from a single relation $\varphi \simeq \psi$ over L^+. Intuitively, we perceive the most significant ones, even if the meaning of significant must be clarified (cf. Section 3).

Examples The ω-code $L = a + ab$ associated to $\Lambda = 0 + 1$ hasn't any relation nor ω-relation. The code $C = a^2 + ba + b$ associated to $\Lambda = 0 + 1 + 2$ has the ω-relation $10^\omega \simeq_\omega 20^\omega$. The language $L = a + ab + ba$ associated to $\Lambda = 0 + 1 + 2$ has the relation $02 \simeq 10$ and the ω-relations $1^\omega \simeq_\omega 02^\omega$ and $(02)^\omega \simeq_\omega (10)^\omega$.

1.4 The context

This study is partially motivated by two twins open decidability problems. Indeed, for a given rational language L, we wonder if there exists a code C such that $L^\omega = C^\omega$. The second problem is similar and related to ω-codes. Roughly speaking, the problems consist in decreasing ambiguity or ω-ambiguity of a language L. At the same time, the expressed ω-power must be preseved. A rational ω-power admits the following characteristic rational language [4]:

$$\chi(L^\omega) = \{u \in \Sigma^+ \mid uL^\omega \subseteq L^\omega \text{ and } u^\omega \in L^\omega\}$$

A suitable hypothesis may be that this language is, with respect to inclusion, the greatest element whose ω-power equals to L^ω. This hypothesis includes at least all ω-powers L^ω as soon as L is finite. Under this assumption, we know that $\chi(L^\omega)$ is a semigroup and that every word in a language G such that $G^\omega = L^\omega$ has a factorization over $\chi(L^\omega)$ [9].

2 Ambiguity

Consider L a language over the alphabet Σ. Originally, $Amb(L)$ was defined in [6] as the set of words w such that there exist two distinct words α and β in L verifying $w \in (Pref(\alpha L^\omega) \cap Pref(\beta L^\omega))$. In [2], the set of ω-words in L^ω having two L-factorizations with different first steps is studied. Here, we call $Amb(L^\omega)$ this ω-language and we distinguish $Amb(L^+)$ the set of words in L^+ with at least two L-factorizations with different first steps. These languages are rational as soon as L is rational [2].

Theorem 1. *Let L be a finite language with $L = Root(\chi(L^\omega))$. The following inclusion holds:*

$$\overrightarrow{Amb(L^+)} \subseteq Amb(L^\omega)$$

Proof Clearly, the set $Amb(L^+)$ is a subset of L^+ so, the limit of $Amb(L^+)$ is included in the limit of L^+. We know that $\overrightarrow{L^+} = L^\omega + L^* \overrightarrow{L}$ (cf. [10]). Since L is finite, we obtain that $\overrightarrow{L} = \varnothing$. We deduce that $\overrightarrow{L^+} = L^\omega$ and, consequently, that $\overrightarrow{Amb(L^+)} \subseteq L^\omega$.

Let x be an element in $\overrightarrow{Amb(L^+)}$. Is it ω-ambiguous ? Let us call Λ the relabelling of L. There exists an infinite increasing sequence $(x_i)_{i \geq 0}$ of prefixes of x such that $\forall i \geq 0,\ x_i \in Amb(L^+)$. Then, for every $i \geq 0$, there exist two elements φ_i and ψ_i in Λ^+ with different first steps such that the relations $\varphi_i \simeq \psi_i$ and the equalities $x_i = \tilde{\varphi}_i = \tilde{\psi}_i$ hold. Since L is finite, Λ is finite and at least one element μ_1 in Λ is the first step of a Λ-factorization of each element in an infinite subsequence of $(\varphi_i)_{i \geq 0}$. Let I be the infinite subset of $\mathbb{N}$ such that μ_1 is the first step of a factorization of each word in $(\varphi_i)_{i \in I}$. Consider the corresponding sequence $(\psi_i)_{i \in I}$. Again since Λ is finite, at least one element ν_1 in Λ, different from μ_1, is the first step of a Λ-factorization of each word in an infinite subsequence of $(\psi_i)_{i \in I}$. Let J be the infinite subset of I such that ν_1 is the first step of a Λ-factorization of each word in $(\psi_i)_{i \in J}$ and recall that μ_1 is the first step of a Λ-factorization for each word in $(\varphi_i)_{i \in J}$. Whatever the first step is, the second respective steps μ_2 and ν_2 can be chosen using the same argument of finiteness of Λ, and so on, infinitely often. In conclusion, there exist two ω-words φ and ψ in Λ^ω, an infinite subset K of J and two sequences $(\mu_m)_{m \geq 1}$ and $(\nu_n)_{n \geq 1}$ in Λ such that $\varphi = \overrightarrow{(\varphi_i)_{i \in K}} = \mu_1 \mu_2 ... \mu_m ...$ and respectively $\psi = \overrightarrow{(\psi_i)_{i \in K}} = \nu_1 \nu_2 ... \nu_n ...$. μ_1 is the prefix of length 1 of φ while ν_1 is the prefix of length 1 of ψ and they are different. The ω-word $x = \tilde{\varphi} = \tilde{\psi}$ is in L^ω (as already shown). In addition, we've proved that $x \in Amb(L^\omega)$. □

The converse is false since, for a code non-ω-code C, $\overrightarrow{Amb(C^+)} = \varnothing$ although $Amb(C^\omega) \neq \varnothing$.

Example Consider the code non ω-code $C = a^2 + b + ba$. $Amb(C^\omega) = ba^\omega$ but $Amb(C^+) = \varnothing$.

Let us define the ω-language composed of ω-words in L^ω for which, given two factorizations with different first steps, those factorizations never coincide. This set is denoted by $Amb_v(L^\omega)$ and is rational if L is itself rational (we justify the name we give in Section 3).

Definition 4. *Let L be a language, $Amb_v(L^\omega)$ is defined as follows:*

$$Amb_v(L^\omega) = Amb(L^\omega) \setminus \big(Amb(L^+)\ L^\omega\big)$$

Now, we are able to explore the deep structure of the set $Amb(L^\omega)$, at least when L is finite.

Theorem 2. *Let L be a language with $L = Root(\chi(L^\omega))$. If L is finite, the following equality holds:*

$$Amb(L^\omega) = \overrightarrow{Amb(L^+)}\ \cup\ \big((Amb(L^+))^*\ Amb_v(L^\omega)\big)$$

Proof $\subseteq$: Let w be an element in $\mathit{Amb}(L^\omega)$ and consider the L-factorizations of w with different first steps. Two cases arise:
– either two factorizations meet each other infinitely often. This means that w has an infinity of prefixes in $\mathit{Amb}(L^+)$. Then $w \in \overrightarrow{\mathit{Amb}(L^+)}$.
– or, for every pair of factorizations, they meet each other finitely often. So, for a given factorization, there exist two elements u (of maximal length) and z such that $u \in \mathit{Amb}(L^+) \cup \{\varepsilon\} = (\mathit{Amb}(L^+))^*$ and $z \in \mathit{Amb}(L^\omega) \setminus ((\mathit{Amb}(L^+)\, L^\omega)$ such that $w = uz$. It means that $z \in \mathit{Amb}_v(L^\omega)$. Remark that two L-factorizations of z never meet, otherwise the length of u fails to be maximal.
So, the first inclusion is proved since $z \in \overrightarrow{\mathit{Amb}(L^+)} \cup ((\mathit{Amb}(L^+))^*\ \mathit{Amb}_v(L^\omega)))$.
$\supseteq$: Theorem 1 makes the reverse inclusion clear as far as it concerns $\overrightarrow{\mathit{Amb}(L^+)}$. For the other part, let us consider an ω-word z in $(\mathit{Amb}(L^+))^*\mathit{Amb}_v(L^\omega)$ and let us set $z = xw$ with x an element of maximal length in $(\mathit{Amb}(L^+))^*$ and w in $\mathit{Amb}_v(L^\omega)$. With $x = \varepsilon$ or not, w is clearly an element in $\mathit{Amb}(L^\omega)$. $\square$

However, the previous theorem doesn't necessarily induce a splitting of $\mathit{Amb}(L^\omega)$ into a partition. The trouble comes from the limit $\overrightarrow{\mathit{Amb}(L^+)}$. Due to its definition, the set $\mathit{Amb}(L^+)$ is a semigroup and it is rational as soon as L is rational [2]. Therefore, we can deduce the following corollary.

Corollary 1. *Let L be a language with $L = \mathit{Root}(\chi(L^\omega))$. If L is finite and if $(\mathit{Amb}(L^+))^\omega$ is a deterministic ω-language, the following equality holds:*

$$\mathit{Amb}(L^\omega) = (\mathit{Amb}(L^+))^* \big((\mathit{Amb}(L^+))^\omega \ \oplus\ \mathit{Amb}_v(L^\omega)\big)$$

Proof The rational semigroup $\mathit{Amb}(L^+)$ verifies $\overrightarrow{\mathit{Amb}(L^+)} = (\mathit{Amb}(L^+))^\omega \cup (\mathit{Amb}(L^+))^*\ \overrightarrow{\mathit{Root}(\mathit{Amb}(L^+))}$. If $(\mathit{Amb}(L^+))^\omega$ is a deterministic ω-language, the equality $\overrightarrow{\mathit{Amb}(L^+)} = (\mathit{Amb}(L^+))^\omega$ holds. In Theorem 2, we can substitute the limit of $\mathit{Amb}(L^+)$ with its ω-power. Moreover, the obtained union is a partition: for every element in $(\mathit{Amb}(L^+))^\omega$, there exist two L-factorizations with different first steps that coincide infinitely often. For every element in $\mathit{Amb}_v(L^\omega)$, such a couple of factorizations doesn't exist. $\square$

We give a simple example to show that, even if L is finite, $\overrightarrow{\mathit{Amb}(L^+)}$ may strictly contain $(\mathit{Amb}(L^+))^\omega$, preventing the latter from being deterministic.

Example Let $L = a + ab + ba$ be the language associated to $\Lambda = 0 + 1 + 2$. Two different factorizations of the ω-word $a(ba)^\omega$ are modelized by $1^\omega \simeq_\omega 02^\omega$ so it belongs to $\mathit{Amb}(L^\omega)$. Clearly, $a(ba)^\omega$ is in $\overrightarrow{\mathit{Amb}(L^+)} = \overrightarrow{a(ba)^+}$ but it doesn't belong to $(\mathit{Amb}(L^+))^\omega = (a(ba)^+))^\omega$, contrarily to $(aba)^\omega$ another element in $\mathit{Amb}(L^\omega)$.

Yet, a criterion exists to split $\mathit{Amb}(L^\omega)$ into two disjoint subsets. It consists in the potential existence of two factorizations with different first steps that coincide infinitely often. If such a case happens, a sub-criterion is the possible belongingness to $(Amb(L^+))^\omega$. Consequently, even if $(Amb(L^+))^\omega$ fails to be deterministic, a partition of $Amb(L^\omega)$ has been found.

Theorem 3. *Let L be a language with $L = Root(\chi(L^\omega))$. If L is finite, the set $Amb(L^\omega)$ verifies the following equality:*

$$Amb(L^\omega) = (Amb(L^+))^* \big(((Amb(L^+))^\omega \oplus (Amb(L^+)\,(L^\omega \setminus Amb(L^\omega)))) \;\oplus\; Amb_v(L^\omega)\big)$$

Proof The proof is based on Theorem 2. It remains to prove that elements in $\overrightarrow{Amb(L^+)}$ that don't belong to $(Amb(L^+))^\omega$ are necessarily either in $Amb(L^+)$ $(L^\omega \setminus Amb(L^\omega))$ or in $(Amb(L^+))^* Amb_v(L^\omega)$. Let w be an element in $\overrightarrow{Amb(L^+)} \setminus (Amb(L^+))^\omega$. According to the equality about the limit of a semigroup, w is in $(Amb(L^+))^* \; \overrightarrow{Root(Amb(L^+))}$. There exist two words u in $(Amb(L^+))^*$ and $y \in \overrightarrow{Root(Amb(L^+))} \setminus \left(Amb(L^+)\,\overrightarrow{Root(Amb(L^+))}\right)$ such that $w = uy$. Note that y exists otherwise y would be in $(Amb(L^+))^\omega$. Since $\overrightarrow{Root(Amb(L^+))} \subseteq \overrightarrow{Amb(L^+)}$, we deduce from Theorem 1 that $y \in Amb(L^\omega)$. Two cases arises:
– either two factorizations with different first steps of y meet each other infinitely often. So, y doesn't belong to $Amb_v(L^\omega)$. There exists a sequence $(u_i)_{i\geq 0}$ of elements in L^+ such that $y = u_0 u_1 \ldots$. Only a finite and non-empty subsequence of $(u_i)_{i\geq 0}$ contains elements in $Amb(L^+)$, otherwise, y would be in $(Amb(L^+))^\omega$. Hence, there exist v in $Amb(L^+)$ and z in L^ω such that $y = vz$. By definition of y, z belongs neither to $\overrightarrow{Root(Amb(L^+))}$ nor to $Amb_v(L^\omega)$. Thus, z is necessarily in $L^\omega \setminus Amb(L^\omega)$.
– or, for every pair of factorizations with different first steps of y, they meet each other finitely often. We get that $w \in (Amb(L^+))^* Amb_v(L^\omega)$.
Moreover, the sets $(Amb(L^+))^\omega$, $Amb(L^+)\,(L^\omega \setminus Amb(L^\omega))$ and $Amb_v(L^\omega)$ are necessarily disjoint. For words in the first two sets, pairs of factorizations that coincide infinitely often exist, in opposition to these in the latter set. Finally, the partition over $((Amb(L^+))^\omega) \cup (Amb(L^+)\,(L^\omega \setminus Amb(L^\omega)))$ is induced by the fact that an element belongs to $(Amb(L^+))^\omega$ or not. $\square$

Such a partition for the set $Amb(L^\omega)$ is illustrated in the following example.

Example Consider the language $L = a^2 + a^3 + b + ba$. The set $Amb(L^+)$ equals to $(a^5 + a^6 + ba^3a^*)L^*$. The set $Amb(L^\omega)$ gathers ω-words in $(Amb(L^+))^\omega$. It also contains the elements in $(Amb(L^+)\,(L^\omega \setminus Amb(L^\omega)))$ (for instance, a^5b^ω which is not in $(Amb(L^+))^\omega$). Finally, its other elements belong to $(Amb(L^+))^*\,Amb_v(L^\omega)$ that is equal to $((a^5 + a^6 + ba^3a^*)L^*)^*ba^\omega$.

By definition, L is a code if and only if $Amb(L^+)$ is empty. If $Amb(L^\omega) = Amb_v(L^\omega)$ then L is a code non ω-code. We can wonder which are the languages L such that $Amb(L^\omega) = (Amb(L^+))^\omega$. Usually but not very formally such languages are referred to as *"languages that would be ω-codes if they were codes"*. Indeed, like in the following example, every ω-ambiguous word is an infinite concatenation of words in $Amb(L^+)$.

Example Consider the language $L = a + bc + ab + c$. $Amb(L^+) = (abc)^+$. In addition, $Amb(L^\omega) = (Amb(L^+))^\omega = (abc)^\omega$.

Another question could be, if a language is not a code, whether $Amb_v(L^\omega)$ is included in $\overrightarrow{Amb(L^+)}$ or not. We see on the next example that the answer is negative.

Example The language $L = ab+c+a+bc+d+de+ee$ admits de^ω in $Amb_v(L^\omega)$ although $Amb(L^+) = abc\,L^*$.

3 Strong ambiguity

We propose here an attempt to refine the usual notions of ambiguity and ω-ambiguity. Indeed, we remark that some elements in L are not fully involved in ambiguity because they *synchronise* any beginning of ambiguous factorizations. They may be omitted in every expression of finitary ambiguity although they may be needed to express ω-ambiguity. If so, they only appear as first steps of factorizations.

Example Consider the language $L = a^2 + a^3 + b + bc + cc$. The ambiguity of words in L^+ is due to the elements a^2 and a^3. Indeed, $Amb(L^+) = (a^5 + a^6)L^*$. In addition, $Amb(L^\omega) = (a^5 + a^6)L^\omega + \big((a^5 + a^6)L^*\big)^* bc^\omega$. Yet, only the two ω-words a^ω and bc^ω are fully responsible of ω-ambiguity.

So, we propose to simplify the usual definitions of the sets $Amb(L^+)$ and $Amb(L^\omega)$. The aim is to make more concise the descriptions of ambiguity and ω-ambiguity and beyond, to quantify them. To reach a lightened definition of ambiguity, we need to identify useless words with respect to ambiguity of words. We propose the following definition of synchronising words.

Definition 5. *Let L be a language. The set of synchronising words in L is defined as follows:*

$$Sync(L) = \{u \in L,\ \big(Amb(L^+) \cap (L^*\,u\,L^*)\big) \subseteq \big(Amb(L^+)\,u\,L^*\big)\}$$

Clearly, the language $Sync(L)$ is rational whenever L is rational, using the rationality of $Amb(L^+)$.

Continuating example Consider once again the language $L = a^2+a^3+b+bc+cc$. Clearly, $Sync(L) = b + bc + cc$.

Remark that the words in $Sync(L)$ cannot appear in any element of a couple of non-coinciding factorizations of a word in L^+. Hence, we give refined definitions of ambiguity on words and ω-words.

Definition 6. *Let L be a language, $Amb_s(L^+)$ is the set of strongly ambiguous words in L^+ defined as follows:*

$$Amb_s(L^+) = Amb\left((L \setminus Sync(L))^+\right)$$

$Amb_s(L^+)$ is not a semigroup anymore and the equality $Amb(L^+) = Amb_s(L^+)L^*$ allows to retrieve $Amb(L^+)$ from $Amb_s(L^+)$.

Continuating example The language $L = a^2+a^3+b+bc+cc$. The set $Amb_s(L^+) = (a^5 + a^6)(a^2 + a^3)^*$.

To introduce the section, we said that words involved in ambiguity or in ω-ambiguity are not necessary the same, precisely when considering first steps. So, we've extended strong ambiguity to ω-words carefully.

Definition 7. *Let L be a language, $Amb_s(L^\omega)$ is the set of strongly ω-ambiguous ω-words in L^ω defined as follows:*

$$Amb_s(L^\omega) = Amb\left((L \setminus Sync(L))^\omega\right) \cup \left((Amb_s(L^+))^* Amb_v(L^\omega)\right)$$

Hence, except as first steps of ω-words in $Amb_v(L^\omega)$, words in $Sync(L)$ disappear from factorizations of elements when comparing $Amb(L^\omega)$ to $Amb_s(L^\omega)$.

Continuating example The language $L = a^2+a^3+b+bc+cc$. The set $Amb_s(L^\omega) = a^\omega + \left((a^5 + a^6)(a^2 + a^3)^* bc^\omega\right)$.

Proposition 1. *Let L be a language. L is a code if and only if $Amb_s(L^+)$ is empty. L is an ω-code if and only if $Amb_s(L^\omega)$ is empty.*

Proof It is easy to verify that $(Amb_s(L^+) = \varnothing) \Leftrightarrow (Amb(L^+) = \varnothing)$ and the corresponding equivalence for ω-words. □

In addition, $Amb_s(L^\omega)$ has a decomposition similar to that of $Amb(L^\omega)$ provided by Theorem 2.

Theorem 4. *Let L be a language with $L = Root(\chi(L^\omega))$. If L is finite, the following equality holds:*

$$Amb_s(L^\omega) = \overrightarrow{(Amb_s(L^+))^+} \;\cup\; \left((Amb_s(L^+))^* \; Amb_v(L^\omega)\right)$$

Proof The restriction to strong ambiguity has no influence on demonstrations for Theorem 1 and Theorem 2. □

Corollary 2. *Let L be a language with $L = Root(\chi(L^\omega))$. If L is finite and if $(Amb_s(L^+))^\omega$ is deterministic, the following equality holds:*

$$Amb_s(L^\omega) = (Amb_s(L^+))^* \;\left((Amb_s(L^+))^\omega \;\oplus\; Amb_v(L^\omega)\right)$$

By analogy, we can define the sets of ambiguous words and ω-ambiguous ω-words for which two factorizations with different first steps never coincide. Among them, we've already used $Amb_v(L^\omega)$.

Definition 8. *Let L be a language, the sets $Amb_v(L^+)$ of very strongly ambiguous words and $Amb_v(L^\omega)$ of very strongly ambiguous ω-words in L^ω are defined as:*

$$Amb_v(L^+) = Amb(L^+) \setminus \big(Amb(L^+)\ L^+\big)$$

$$Amb_v(L^\omega) = Amb(L^\omega) \setminus \big(Amb(L^+)\ L^\omega\big)$$

Actually, $Amb_v(L^+)$ may be a strict subset of $Root(Amb(L^+))$.

Continuating example Consider again the language $L = a^2+a^3+b+bc+cc$. The set $Amb_s(L^+) = (a^5+a^6)(a^2+a^3)^*$ and $Amb_s(L^\omega) = a^\omega+(a^5+a^6)(a^2+a^3)^*bc^\omega$. Besides, the set $Amb_v(L^+) = a^5 + a^6$ and the set $Amb_v(L^\omega) = bc^\omega$.

The following table illustrates these notions on simple but representative languages.

Λ	L	$Amb_s(L^+)$	$Amb_v(L^+)$	$\simeq$	restr. $Amb_s(L^\omega)$	$\simeq_\omega$
$0+1$	$a+ab$	$\varnothing$	$\varnothing$	–	$\varnothing$	–
$0+1+2$	a^2+b+ba	$\varnothing$	$\varnothing$	–	ba^ω	$10^\omega \simeq_\omega 20^\omega$
$0+1+2$	a^2+a^3+b	$a^5(a^2+a^3)^*$ $a^6(a^2+a^3)^*$	a^5 a^6	$01 \simeq 10$ $000 \simeq 11$	a^ω	$0^\omega \simeq_\omega 1^\omega$
$0+1+2$	$a+ab+ba$	$a(ba)^+$	aba	$02 \simeq 10$	$a(ba)^\omega$ $(aba)^\omega$	$1^\omega \simeq_\omega 02^\omega$ $(02)^\omega \simeq_\omega (10)^\omega$
$0+1+2$	$ab+abc+cabab$	$(abcab)^+$	$abcab$	$02 \simeq 100$	$(abcab)^\omega$	$02^\omega \simeq_\omega (10)^\omega$
$0+1+2+3$	$a+ab+bc+c$	$(abc)^+$	abc	$02 \simeq 13$	$(abc)^\omega$	$(02)^\omega \simeq_\omega (13)^\omega$
$0+1+2+3$	a^2+a^3+b+ba	a^5L^* a^6L^* $ba^3a^*L^*$	a^5 a^6 ba^3 ba^4	$01 \simeq 10$ $000 \simeq 11$ $21 \simeq 30$ $200 \simeq 31$	a^ω ba^ω	$0^\omega \simeq_\omega 1^\omega$ $20^\omega \simeq_\omega 30^\omega$
$0+1+2+3+4$	$a^2+a^3+b+bc+cc$	$a^5(a^2+a^3)^*$ $a^6(a^2+a^3)^*$	a^5 a^6	$01 \simeq 10$ $000 \simeq 11$	a^ω bc^ω	$0^\omega \simeq_\omega 1^\omega$ $23^\omega \simeq_\omega 24^\omega$

So, an idea to decrease the combinatorial spread of ambiguity or ω-ambiguity representatives is to use sets, relations and ω-relations restricted to strong ambiguity or even to very strong ambiguity, but the study must be pursued.

4 Conclusion

The study of ambiguity and ω-ambiguity, once done, will be completed by a study of minimal ambiguity and minimal ω-ambiguity. They appear as dual notions for maximal codes and maximal ω-codes. Considering the related problems of decidability, we know that we have to identify cases in which ambiguity or ω-ambiguity is minimal to find an affirmative answer and cases where ambiguity or ω-ambiguity are tractable to prove that no positive solution exists.

References

1. J. Berstel and D. Perrin. *Theory of codes.* Academic Press, 1985.
2. J. Devolder, M. Latteux, I. Litovsky, and L. Staiger. Codes and infinite words. *Acta Cybernetica*, 11(4):241–256, 1994.
3. S. Eilenberg. *Automata, Languages and Machines*, volume A. Academic Press, 1974.
4. S. Julia. A characteristic language for rational ω-powers. In *Proc.* 3^{rd} *Int. Conf. Developments in Language Theory*, pages 299–308, Thessaloniki, 1997.
5. S. Julia and Tran Vinh Duc. Reduced languages as ω-powers. In *Proc.* 11^{th} *Int. Conf. Developments in Language Theory*, Lecture Notes in Computer Science, pages 266–277, Turku, 2007. Springer.
6. J. Karhumäki. On three-element codes. *Theoretical Computer Science*, 40:3–11, 1985.
7. J. Karhumäki. A property of three-element codes. *Theoretical Computer Science*, 41:215–222, 1985.
8. I. Litovsky. Prefix-free languages as ω-generators. *Information Processing Letters*, 37:61–65, 1991.
9. I. Litovsky and E. Timmerman. On generators of rational ω-power languages. *Theoretical Computer Science*, 53:187–200, 1987.
10. D. Perrin and J.É. Pin. *Infinite words.* Elsevier, 2004.
11. L. Staiger. On infinitary finite length codes. *Theoretical Informatics and Applications*, 20(4):483–494, 1986.