INFORMATIQUE THÉORIQUE ET APPLICATIONS

L. ROSAZ

Unavoidable languages, cuts and innocent sets of words

Informatique théorique et applications, tome 29, nº 5 (1995), p. 339-382

<http://www.numdam.org/item?id=ITA_1995__29_5_339_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 29, n° 5, 1995, p. 339 à 382)

UNAVOIDABLE LANGUAGES, CUTS AND INNOCENT SETS OF WORDS (*)

by L. ROSAZ ([1])

Communicated by C. CHOFFRUT

Abstract. – *A language X on an alphabet A is unavoidable iff all but finitely many words in A^* have a factor in X. In this paper, cuts (which are basically simplifications of languages in terms of avoidability) and innocent language (those which cannot be cut) are defined. Their properties, most of which are related to unavoidable languages, are studied.*

1. INTRODUCTION

A language X on the finite alphabet A (that is a subset X of the free monoid A^*, that is a set X of words on an alphabet A) is unavoidable iff all but finitely many words in A^* have a factor in X. This is not to be confused with unavoidable patterns, such as the square in squarefree words. See D. R. Bean, A. Ehrenfeucht and G. F. MacNulty [2] or Lothaire [13] for references on this latter topic.

Unavoidable languages appeared in 1964 in a paper by M. P. Schutzenberger (*see* [19]) where he gave a bound on the maximal length of a word that avoids a finite unavoidable language. This bound depends on the maximal length of the words in the unavoidable language. M. Crochemore, M. Lerest and P. Wender proved later (in 1983) in [5], that the bound given by M. P. Schutzenberger was the best possible.

Unavoidable languages were explicitly introduced in 1983 by A. Ehrenfeucht, D. Haussler and G. Rozenberg in [6] in a generalization of Higman's

(*) Received March 4, 1993, accepted April 15, 1993.

([1]) L.I.T.P. (Laboratoire d'Informatique Théorique et de Programmation), Université Paris-VII, 2, place Jussieu, 75251 Paris Cedex 05, France.
Visitor at: Department of Mathematics, Pennsylvania State University, McAllister building, University Park, Pa 16802, U.S.

result [8]. Higman's theorem states that if A is a finite alphabet, then in every infinite language $\{ u_i \mid i \in I \}$ on A, there is a pair (u_i, u_j) of words with $i \neq j$ such that u_i is a subword of u_j (a *subword* of a word u is a word v obtained by taking a subsequence of the letters of u. For example ac is a subword of abc). The generalization by A. Ehrenfeucht, D. Haussler and G. Rozenberg says that if the partial order relation $\leq$ is the transitive closure of:

$$\text{“}u \leq v \quad \text{iff} \quad \exists\, w,\, y,\, z \quad \text{with } w,\, z \in A^* \quad \text{and} \quad y \in X \\ \text{such that } u = wz \text{ and } v = wyz\text{”},$$

then X is unavoidable iff every infinite language on A contains two different words u and v such that $u \leq v$. One gets Higman's theorem from this result by considering $X = A$.

W. Bucher, A. Ehrenfeucht and D. Haussler generalized the latter result in [3]. Kruskal in [9] and L. Puel in [16] gave some similar results on trees instead of words.

It had been conjectured that if X is unavoidable, then there is a word w in X and a letter α such that $X - \{ w \} + \{ w\alpha \}$ is still unavoidable. This word-extension conjecture was often called Ehrenfeucht's conjecture, though it might be due to D. Haussler. A counter-example to this conjecture can be found in [17].

In 1984, C. Choffrut and K. Culik published [4] where they recalled some basic results (An unavoidable language always contains a finite sublanguage which is unavoidable, recall of the automaton of A. V. Aho and M. J. Corasick [1], and use of this automaton to decide whether a given language is unavoidable) and gave some interesting new ones (partial answer to the word-extension conjecture, uniqueness of the extention of a word when it exists, first use of some important tools such as bi-infinite periodic words...). This paper is the one to be read as an introduction to unavoidable languages.

There are other notions of "unavoidable" in theoretical computer science: Unavoidable patterns, such as the square in squarefree words, *see* D. R. Bean, A. Ehrenfeucht and G. F. MacNulty [2] or Lothaire [13] for references on this topic; unavoidable words with patterns, *see* [12]; unavoidable trees, which were studied by L. Puel in her thesis [16] where she generalized Kruskal's theorem [9]; unavoidable subset of an ordered set, *see* [15].

An unavoidable language X is minimal iff no proper sublanguage Y of X is unavoidable. This paper first gives a simple necessary and sufficient condition on a finite language Y for the existence of a finite minimal unavoidable language X containing Y. This result is used in the proof of proposition 6.11.

Then, I will introduce cuts: A language X cuts into a language Y in an elementary way iff $[Y = X - \{u\}$ where $u \in X$ and there is a proper factor v of u which belongs to X, or if $Y = X - \{ua\} + \{u\}$ where $ua \in X$ (a is a letter) and for every letter $b \neq a$, there is a suffix u_b of u such that $u_b b \in X$, or if $Y = X - \{au\} + \{u\}$ where $au \in X$ (a is a letter) and for every letter $b \neq a$, there is a prefix u_b of u such that $bu_b \in X]$. A language X cuts into a language Y iff a finite sequence of elementary cuts leads from X to Y.

The basic interest of cuts is that if X cuts into Y, then Y is "shorter" than X, but the set of bi-infinite words avoiding X (that is the set of bi-infinite words with no factor in X) is the same as the set of bi-infinite words avoiding Y, so a cut is a simplification when you are interested in the set of bi-infinite words avoiding a given set of finite words (For example, for unavoidable languages, you want this set of bi-infinite words to be empty, and if you deal with symbolic dynamic systems, this set of bi-infinite words is a system of finite type defined from the finite set of finite words they avoid.)

A language is innocent iff it cannot be cut. The main result for *finite* languages X is that there is a unique innocent language $\overline{X}$ such that the bi-infinite words avoiding X are the same as the bi-infinite words avoiding $\overline{X}$, and that X cuts into $\overline{X}$. Consequently, one gets an algorithm to decide whether a finite language is unavoidable (It seems that J. P. Duval discovered this algorithm before I did, but he did not publish it.) By defining eventual cuts, one gets similar results for infinite languages. One can define strong equivalence for languages as $X \sim_s Y$ iff $\overline{X} = \overline{Y}$ iff the bi-infinite words avoiding X are the same as the ones avoiding Y. Another definition, with not so clear properties, is the weak equivalence: $X \sim_w Y$ iff the *periodic* bi-infinite words avoiding X are the same as the ones avoiding Y.

Various results and properties on these notions will be given in this paper.

I first recall in section 2 some basic definitions on words: on finite words and languages (length of a word, word ε, concatenation, factors, prefixes, suffixes on finite words, operations $+$, product, $*$ and $^+$ on languages), and on infinite words (finite factor of a bi-infinite word, periodic bi-infinite words, notation $u^{\mathbb{Z}}$, equivalence $\equiv$ (equality up to a translation)). I define unavoidable languages and I give some examples in section 3. A necessary and sufficient condition for the existence of an enlargement of a language X into a minimal unavoidable language is given in section 4. In section 5, cuts are defined, as well as innocent and guilty languages, and their basic properties are given, which lead to an algorithm to decide whether a finite language is unavoidable. Section 6 gives further properties of cuts, which lead to the definition of acquitted X for finite languages X and to strong and weak equivalences. Section 7 restates the uniqueness of acquitted X for finite languages X, with a more natural, but longer and more technical proof than the one in section 6. This section also states that one can get acquitted X by acquitting on the left, then on the right. Section 8 generalizes the most important results of the previous sections to *infinite* languages. This requires the definition of eventual cuts. In section 9, I give a few open problems.

2. BASIC DEFINITIONS

To begin with, let me precise that I consider that $\mathbb{N} = \{0, 1, \ldots\}$, so that $0 \in \mathbb{N}$ (The set $\{1, 2, \ldots\} = \mathbb{N} - \{0\}$ will be denoted by $\mathbb{N}^*$). I also precise that whenever I write $X - Y$, where X and Y are two sets, I implicity assume that $Y \subset X$.

An *alphabet* is a finite set whose elements are called *letters*. The alphabet is usually denoted by A. A *finite word* (or for short, a *word*) on A is a finite sequence of elements of the alphabet. A word will be denoted by writing its letters one after the other. Unless otherwise stated, every word, and every set of words we will talk about, is implicity on an alphabet denoted by A. The *length* of a word u, denoted by $|u|$, is the number of its letters. There is a word of length 0 which is denoted by ε. The number of occurrences of a letter α in a word u is denoted by $|u|_\alpha$. It is clear that $\sum_{\alpha \in A} |u|_\alpha = |u|$.

The *concatenation* of two words u and v, denoted by uv, is the word obtained by writing the letters of u and then those of v. A *factor* of a word u is a word v such that there exist words w and z such that $u = wvz$. A factor v of u is *proper* if it is different from u.

A word v is a *prefix* of a word u iff there exists a word w such that $u = vw$. A word v is a *suffix* of a word u iff there exists a word w such that $u = wv$.

A *language* is a set of words. Then the set of all the words on an alphabet A, (which is denoted by A^*) with the concatenation product is the free monoid on A. The set of the words on A of length l is A^l and the set of the words of length less than or equal to l is denoted by $A^{\leq l}$. Sometimes, the language $\{u\}$ will be denoted simply by u.

A *bi-infinite word* is a $\mathbb{Z}$-sequence of elements in A (An *infinite word* is an $\mathbb{N}$-sequence). The set of all bi-infinite words on A is denoted by $A^{\mathbb{Z}}$. A bi-infinite word is denoted by $(a_i)_{i \in \mathbb{Z}}$ or by $\dots a_{-p}\, a_{-p+1} \dots a_{-2}\, a_{-1}\, a_0\, a_1\, a_2 \dots a_q \dots$

A *finite factor of a bi-infinite word* $\aleph = (a_i)_{i \in \mathbb{Z}}$ is a finite word u such that there are integers N and N' such that $u = a_N \dots a_{N'-1}$.

If the sequence is periodic of period T, then the bi-infinite word will be said to be *periodic of period* T. Such a word $\aleph = (a_i)_{i \in \mathbb{Z}}$ is given by $(a_i)_{i \in \{0, \dots, T-1\}}$. That word will be denoted by $u^{\mathbb{Z}}$ where $u = a_0 \dots a_{T-1}$.

Let $\aleph_1 = (a_{1,n})_{n \in \mathbb{Z}}$ and $\aleph_2 = (a_{2,n})_{n \in \mathbb{Z}}$ be two bi-infinite words on the alphabet A. They will be said to be *translates of each other* iff there is a p such that $\forall n \in \mathbb{Z}$, $a_{1,n} = a_{2,n+p}$. The notation $\aleph_1 \equiv \aleph_2$ will be used to say $\aleph_1$ and $\aleph_2$ are translates of each other, and $\aleph_1 \not\equiv \aleph_2$ to say they are not. The relation $\equiv$ is clearly an equivalence relation. From now on, bi-infinite words will always be considered up to translation. Note that if p, s are words (not both equal to ε), then $(ps)^{\mathbb{Z}} \equiv (sp)^{\mathbb{Z}}$.

An *infinite word* is a $\mathbb{N}$-sequence of elements in A. An infinite word can be denoted by $(a_i)_{i \in \mathbb{N}}$. If the sequence is periodic of period T, then the infinite word will be denoted by $u^{\mathbb{N}}$, where $u = a_0 \dots a_{T-1}$.

3. UNAVOIDABLE LANGUAGES

PROPOSITION 3.1: *Let X be a language, then the following two properties are equivalent:*

(i) *There is an integer N such that, for every word u in A^* of length at least N, there is a word v in X which is a factor of u.*

(ii) *For every word $\aleph$ in $A^{\mathbb{Z}}$, there is a word v in X which is a factor of $\aleph$.*

Moreover, if X is finite, then the above two properties are equivalent to the following one:

(iii) *For every periodic word $\aleph$ in $A^{\mathbb{Z}}$, there is a word v in X which is a factor of $\aleph$.*

Proof of proposition 3.1: (i) $\Rightarrow$ (ii): Assume $\neg$(ii): There is a bi-infinite word $\aleph$ such that no element in X is a factor of $\aleph$. Let E be the set of the finite factors of $\aleph$. Elements in X are factors of *no* word in E. This language E contains words of every length and therefore (i) is not satisfied. One has $\neg$(i).

(ii) $\Rightarrow$ (i): Assume $\neg$(i), then there is, for every $n \in \mathbb{N}$, a word u_n with no factor in X and which is of length $2n+1$. Let $(a_{n,i})_{n \in \mathbb{N}, i \in \mathbb{Z}, -n \leq i \leq n}$ be the letters such that $u_n = a_{n,-n}\, a_{n,-n+1} \dots a_{n,0} \dots a_{n,n}$ for every $n \in \mathbb{N}$. Define $(a_{n,i})_{n \in \mathbb{N}, i \in \mathbb{Z}, |i| > n}$ in an arbitrary way. Then let $\aleph_n = (a_{n,i})_{i \in \mathbb{Z}}$ for every $n \in \mathbb{N}$. Put the discrete topology on the alphabet A, which becomes a compact metric space, and the infinite-product topology on $A^{\mathbb{Z}}$ which becomes also a compact metric space. Thus, one can extract from $(\aleph_n)_{n \in \mathbb{N}}$ a subsequence $(\aleph_{\phi(n)})_{n \in \mathbb{N}}$ which converges to a bi-infinite word $\aleph = (a_i)_{i \in \mathbb{Z}}$. The convergence of $(\aleph_{\phi(n)})_{n \in \mathbb{N}}$ to $\aleph$ implies that each finite factor of $\aleph$ is a factor at the same position of all but finitely many $\aleph_{\phi(n)}$'s, therefore is a factor of a word u_n for some $n \in \mathbb{N}$, and therefore is not in X. So no element in X is a factor of $\aleph$. Consequently, (ii) is not satisfied: One has $\neg$(ii).

(ii) $\Rightarrow$ (iii) is obvious.

If X is finite, then (iii) $\Rightarrow$ (i):

Assume (iii). Let $l = \max\limits_{v \in X} |v|$, $K = (\operatorname{card} A)^l$ (K is the number of words on A of length l) and $N = (K+1)\,l$. Let u be a word of length at least N. The word u can be written $u = u_0\, u_1 \ldots u_K\, z$ where for every i in $[0, k]$, u_i is a word of length l, and where z is a word. Because there are $K+1$ u_k's and only K different words of length l, two u_k's must be equal, *i.e.* $\exists\, i < j$ such that $u_i = u_j$. Let $w = u_i\, u_{i+1} \ldots u_{j-1}$. Because (iii) is assumed to be true, there is an $x \in X$ which is a factor of $w^{\mathbb{Z}}$. We have now two cases:

◇ If x is a factor of w, then x is also a factor of u since w is a factor of u.

◇ If x is not a factor of w, then there are an $n \in \mathbb{N}$, a suffix s and a prefix p of w such that $x = sw^n\, p$. But $|w| = (j-i)\,l \geq l = \max\limits_{v \in X} |v| \geq |x|$, therefore n must be 0 and $x = sp$ (or $n = 1$ and $s = p = \varepsilon$, but then $x = w$ which cannot happen here since we have assumed that x is not a factor of w). But p is a prefix of $w = u_i \ldots u_{j-1}$ and (since $x = sp$), $|p| \leq |x| \leq l = |u_i|$, therefore p is a prefix of u_i, which is the same as u_j. Since s is a suffix of $w = u_i \ldots u_{j-1}$, since p is a prefix of u_j and since $x = sp$, one gets that x is a factor of $wu_j = u_i \ldots u_{j-1}\, u_j$ which is a factor of u. Therefore x is a factor of u.

In both case, x is found to be a factor of u, and (i) is proved.

Proposition 3.1 is proved. □

Notes 3.2

◇ The implication (iii) $\Rightarrow$ (i) is false for infinite languages, see for example $X = \{uu \mid u \in A^+\}$, the set of non-$\varepsilon$ squares on the alphabet $A = \{a, b, c\}$ with the help of [2].

◇ When X is finite, another way to prove proposition 3.1 is to build an automaton recognizing finite and infinite words with no factor in X and then to see that the above three conditions are equivalent to "there are no loops in the automaton".

DEFINITIONS 3.3: A language X is *unavoidable* iff it satisfies the first two conditions in proposition 3.1, it is *avoidable* iff it does not.

Equivalent definitions are:

Let X be a language, then X is unavoidable iff:

$\diamond$ $A^* - A^* X A^*$ is finite: all but finitely many finite words have a factor in X.

$\diamond$ $A^{\mathbb{Z}} - A^{-\mathbb{N}} X A^{\mathbb{N}}$ is empty: all bi-infinite words have a factor in X.

DEFINITION 3.4: Let X be a language and ω be a finite or a bi-infinite word, then ω *avoids* X if no element in X is a factor of ω.

Finite unavoidable languages are quite representative of unavoidable languages thanks to the following proposition:

PROPOSITION 3.5: *Let X be an (infinite) unavoidable language. There is a finite sublanguage X' of X which is unavoidable.*

Proof of proposition 3.5: This proposition is proved by W. Bucher, A. Erhenfeucht and D. Haussler in [3] and by C. Choffrut and K. Culik in [4]. A short proof of this fact is: Let S_w be the set of bi-infinite words containing w as a factor, then $\cup_{w \in X} S_w = A^{\mathbb{Z}}$. But with the infinite-product topology, the S_w's are open and $A^{\mathbb{Z}}$ is compact, thus there is a finite sublanguage X' of X such that $\bigcup_{w \in X'} S_w = A^{\mathbb{Z}}$, that is, which is unavoidable.

Proposition 3.5 is proved. $\square$

Examples 3.6

$\diamond$ $X = A$ is unavoidable.

$\diamond$ $\forall n \in \mathbb{N}$, $X = A^n$ (The set of the words of length n) is unavoidable.

$\diamond$ If $A = \{a, b\}$, then $X = \{aa, bab, bbbbbbbbbb\}$ is unavoidable.

Indeed, try to construct a bi-infinite word $\aleph$ which avoids X: all a's must be preceded and followed by a b because $aa \in X$, thus must be included in a factor bab. But $bab \in X$, so there must be no a's in $\aleph$, so $\aleph$ has to be $b^{\mathbb{Z}}$, but then it contains $bbbbbbbbbb$ which is in X. Thus no bi-infinite word can avoid X, which is therefore unavoidable.

$\diamond$ If $A = \{a, b\}$, then

$$X = \{bb, bab, baab, baaab, \ldots, ba^i b, \ldots, ba^n b, ba^{n+1}, a^{n+2}\}$$

is unavoidable.

Indeed, assume $\aleph$ is a bi-infinite word which avoids X and contains a b. This b must be followed by an a, because bb is in X. Moreover, there cannot be more than $n+1$ consecutive a's after that b because ba^{n+1} is in X. Thus, after that b, there are k a's where $1 \leq k \leq n$, and those a's are followed by a b, so that $\aleph$ contains $ba^k b$, which is in X, there is a contradiction. Thus there cannot be any b in a word avoiding X, so there are only a's, but this is also forbidden, since a^{n+2} is in X. So, no bi-infinite word avoids X, which is therefore unavoidable.

◇ If $X = \{ l_1, \ldots, l_n \}$, $X' = \{ l'_1, \ldots, l'_n \}$, if l_i is a factor of l'_i for all i and if X' is unavoidable, then X is also unavoidable.

◇ If $X \subset X'$ and if X is unavoidable, then X' is also unavoidable.

By looking at the last example, one can see how unavoidable languages can be uselessly big. That leads to define *minimal* unavoidable languages.

DEFINITION 3.7: Let X be an unavoidable language, it is *minimal* iff no proper sublanguage of X is unavoidable, that is iff

$$[[Y \subset X, Y \neq X] \Rightarrow [Y \text{ is avoidable}]].$$

Note that because of proposition 3.5, every minimal unavoidable language is finite.

See [17] to see how to make systematic inventories of unavoidable languages.

Note 3.8: To know whether or not a given finite language X is unavoidable, one can build an automaton recognizing the words with no factor in X, and see if this automaton has loops. For more details, *see* A. V. Aho and M. J. Corasick [1], or C. Choffrut and K. Culik [4]. Another algorithm will be given in this article.

4. EXISTENCE OF AN ENLARGEMENT OF A FINITE LANGUAGE INTO A MINIMAL UNAVOIDABLE LANGUAGE

It is the aim of this section to study the following problem: Let Y be a finite language. Is there a finite language X which is a minimal unavoidable language such that $Y \subset X$?

Note that if one does not ask for a minimal, but for a plain unavoidable language X, then the problem becomes obvious: The language $X = Y \cup Z$, where Z is any unavoidable language, is an unavoidable language containing Y. Note also that if Y is infinite, then it is contained in no minimal unavoidable language X since such a language X cannot be infinite.

DEFINITION 4.1: Let Y be a finite language. The language Y is *separable* iff for every y in Y, there is a periodic bi-infinite word $\aleph_y$ such that y is the one and only one element in Y which is a factor of $\aleph_y$.

THEOREM 4.2: *Let Y be a finite language, then there is a minimal unavoidable language X such that $Y \subset X$ iff Y is separable.*

Proof of theorem 4.2:

The necessity of the condition

Assume Y is not separable, then there is a $y \in Y$ such that every periodic bi-infinite word containing the factor y contains also a factor in $Y - \{ y \}$. Let X be an unavoidable language with $Y \subset X$. Then every periodic bi-infinite word $\aleph$ has a factor in X which is:

- either an element in $X - Y$
- or an element in $Y - \{ y \}$
- or y.

But in the last case, $\aleph$ has also a factor in $Y - \{ y \}$, so that $\aleph$ always has a factor in $X - \{ y \}$. Therefore, $X - \{ y \}$ is unavoidable and X is not minimal. Thus, there is no minimal unavoidable language X which contains Y. □

The sufficiency of the condition

Assume Y is separable: For every $y \in Y$, there is a periodic bi-infinite word $\aleph_y$ such that y is the unique element of Y which is a factor of $\aleph_y$. Let $P = \{ \aleph_y \mid y \in Y \}$.

We will use the following lemma:

LEMMA 4.3: *Let $(\aleph_y)_{y \in Y}$ (at that point, Y can be any finite set of index) be a finite family of periodic bi-infinite words, then there is a finite language C such that a periodic bi-infinite word $\aleph$ avoids C iff there is $y \in Y$ such that $\aleph \equiv \aleph_y$ (that is, $\{ \aleph_y \mid y \in Y \}$ are the only bi-infinite periodic words avoiding C).*

Proof of lemma 4.3: Let T_y be the smallest period of $\aleph_y$ and $T = \max_{y \in Y} T_y$. Let B be the set of the words which are factor of no $\aleph_y$:

$$B = \{ w \in A^* \mid \forall y \in Y, w \text{ is not a factor of } \aleph_y \}.$$

Let C be the set of the words $w \in B$ with no proper factor in B : $C = \{ w \in B \mid [v \text{ is a factor of } w, v \neq w] \Rightarrow [v \notin B] \}$

(that is, C is the set of the minimal elements of B according to the order "is a factor of").

LEMMA 4.4: *If $\aleph$ and $\aleph'$ are two different periodic words of respective periods T and T', and if u is a factor of both $\aleph$ and $\aleph'$, then $|u| \leq T + T'$. (Here, "different" means "different up to translation", that is $\not\equiv$.)*

Proof: Left to the reader or *see* [7] or [13]. □

We show now that C is finite:

Let w be a word in C of length at least 2. One can write $w = \alpha\, v\, \beta$ where α, $\beta \in A$ and $v \in A^*$. The words $\alpha\, v$ and $v\, \beta$ are proper factors of w, so they are not in B (*see* the definition of C), therefore there are two bi-infinite words $\aleph_\alpha$ and $\aleph_\beta$ in $\{ \aleph_y \,|\, y \in Y \}$, such that $\alpha\, v$ is a factor of $\aleph_\alpha$ and $v\, \beta$ is a factor of $\aleph_\beta$.

If $\aleph_\alpha \not\equiv \aleph_\beta$: Let T_α and T_β be the respective periods of $\aleph_\alpha$ and $\aleph_\beta$. The word v is a factor of both $\aleph_\alpha$ and $\aleph_\beta$, therefore (*see* lemma 4.4), $|v| \leq T_\alpha + T_\beta$, hence $|w| \leq T_\alpha + T_\beta + 2$.

If $\aleph_\alpha \equiv \aleph_\beta$: Let $\overline{T}$ be the period of $\aleph_\alpha \equiv \aleph_\beta$ (denoted by $\aleph$ from now on), then [$\alpha\, v$ and $v\, \beta$ are some factors of $\aleph$, but $\alpha\, v\, \beta$ is not] implies $|v| < \overline{T}$. Indeed, assume $|v| \geq \overline{T}$. Let u be the prefix of v of length $\overline{T}$. The word u is a factor of $\aleph$ (because v is a factor of $\aleph$), of length the period of $\aleph$. Therefore $\aleph \equiv u^{\mathbb{Z}}$, and there are $k > 0$ and $p \in A^*$ such that $v\, \beta = u^k\, p\, \beta$ and such that $p\, \beta$ is a prefix of u. But $\alpha\, v\, \beta = \alpha\, u^k\, p\, \beta$ is not a factor of $\aleph$, therefore the last letter of u is NOT α. The word $\alpha\, v$ is a factor of $\aleph$, so $\alpha\, u$ is a factor of $\aleph$, so ($|\alpha\, u| = \overline{T} + 1$), $\alpha\, u$ is a factor of $u u$, but is not a suffix of $u u$ (because α is not the last letter of u). Therefore, u can be written $u = u_1\, u_2$ with $|u_1| > 0$, $|u_2| > 0$ and $u_2\, u_1 = u$. Then $|u_2|$ is a period of $\aleph$ which contradicts the fact that $\overline{T}$ is the smallest period of $\aleph'$.

In both cases, one finds that w must be of length less than $2\,T + 2$. The length of the words in C is bounded, *therefore C is finite.*

Let $\aleph$ be a periodic bi-infinite word of period $\overline{T}$.

If $\aleph \equiv \aleph_y$ for some $y \in Y$, then (by definition of B) no element in B is a factor of any $\aleph_y$, so $\aleph$ avoids B and therefore (because $C \subset B$) $\aleph$ avoids C.

If $\aleph \not\equiv \aleph_y$ for every $y \in Y$, then:

Let z be a factor of $\aleph$ of length $T + \overline{T} + 1$; by lemma 4.4, z cannot be a factor of any $\aleph_y$ (for $y \in Y$), so z is in B. There is a factor z' of z which is in C. Then z' is in C and is a factor of $\aleph$, therefore $\aleph$ does not avoid C.

Therefore, C satisfies the conditions of lemma 4.3.

Lemma 4.3 is proved. □

Let C be a finite language satisfying conditions of lemma 4.3. Since $Y \cup C$ is finite and since a *finite* language is unavoidable iff no periodic bi-infinite word avoids it, one gets that $\underline{Y \cup C}$ *is unavoidable.*

Moreover, every minimal unavoidable language contained in $Y \cup C$ contains Y.

Indeed, let X be an unavoidable language such that $X \subset (Y \cup C)$. Let $y \in Y$, then $\aleph_y$ has a factor in X, but y is the only factor of $\aleph_y$ in Y, so no element in $Y - \{y\}$ is a factor of $\aleph_y$ and $\aleph_y$ has no factor in C (because $\aleph_y$ avoids C), so the factor of $\aleph_y$ in X must be y itself, so $y \in X$. This is true for every $y \in Y$, so $Y \subset X$.

As a conclusion, $Y \cup C$ is finite and unavoidable, there is a sublanguage X of $Y \cup C$ which is a minimal unavoidable language. This language X contains Y, so it satisfies the required conditions.

Theorem 4.2 is proved. □

5. CUTS

In this section, I will introduce cuts, and give the properties which lead to an algorithm to decide whether a finite language is unavoidable. It seems that J. P. Duval discovered this algorithm before I did, but he never published it.

The idea of cuts is to remove what is obviously useless in terms of avoidability.

Let $X = \{ab, baba\}$, a bi-infinite word which contains a factor $baba$ contains also a factor ab (because ab is a factor of $baba$), therefore a bi-infinite word avoids X iff it avoids $\{ab\}$, so that $baba$ has no effect on the avoidability of the set.

Let $\{a, b\}$ be the alphabet and let X be $\{babba, bbb\}$. A bi-infinite word which contains a factor $babb$ contains either a factor $babba$ or a factor $babbb$. In the latter case, it contains also a factor bbb, since bbb is a factor of $babbb$. Therefore a bi-infinite word avoids $X = \{babba, bbb\}$ iff it avoids $\{babb, bbb\}$, so that, in terms of avoidability, the last a in $babba$ is useless.

Cuts are going to remove these useless parts. If X cuts into Y, then a bi-infinite word will avoid Y iff it avoids X.

Let X and Y be some languages on the alphabet A, we consider the following properties on X and Y:

(1) There are some words $u, v \in X$, such that $u \neq v$, u is a factor of v and $Y = X - \{v\}$

(2r) There is a letter $\alpha \in A$ and a word $u \in A^*$ such that:

$$\begin{cases} u \notin X, \\ u\alpha \text{ is in } X, \\ \text{for every } \beta \in A,\ \beta \neq \alpha, \text{ there is a suffix } u_\beta \text{ of } u \text{ such that } \beta u_\beta \in X \text{ and} \\ Y = X - \{u\alpha\} + \{u\} \end{cases}$$

(2l) There is a letter $\alpha \in A$ and a word $u \in A^*$ such that:

$$\begin{cases} u \notin X, \\ \alpha u \text{ is in } X, \\ \text{for every } \beta \in A,\ \beta \neq \alpha, \text{ there is a prefix } u_\beta \text{ of } u \text{ such that } u_\beta \beta \in X \text{ and} \\ Y = X - \{\alpha u\} + \{u\} \end{cases}$$

(3r) There is a word $u \in A^*$ such that:

$$\begin{cases} u \notin X, \\ \text{for every } \alpha \in A,\ u\alpha \in X \\ Y = X - uA + \{u\} \end{cases}$$

(3l) There is a word $u \in A^*$ such that:

$$\begin{cases} u \notin X, \\ \text{for every } \alpha \in A,\ \alpha u \in X \\ Y = X - Au + \{u\} \end{cases}$$

DEFINITIONS 5.1: Let X and Y be two languages, then:

The language X *cuts into* Y (resp. *cuts into* Y *on the right*, resp. *cuts into* Y *on the left*) *in an elementary way* iff (1), (2r) or (2l) is satisfied (resp. iff (1) or (2r) is satisfied, resp. iff (1) or (2l) is satisfied).

The notation $X \overset{\text{cut}}{\to} Y$ (resp. $X \overset{\text{cut}}{\to}_r Y$, resp. $X \overset{\text{cut}}{\to}_l Y$) will be used to say that X cuts into Y (resp. on the right, resp. on the left) in an elementary way.

The language X *cuts into* Y (resp. *on the right*, resp. *on the left*) *in an* <u>*almost*</u> *elementary way* iff (1), (2r), (2l), (3r) or (3l) is satisfied (resp. iff (1), (2r) or (3r) is satisfied, resp. iff (1), (2l) or (3l) is satisfied).

The language X *cuts into* Y (resp. *on the right*, resp. *on the left*) iff there is a finite sequence $X = X_0, X_1, \ldots, X_n = Y$ such that for all $i \in [1, n]$, X_{i-1} cuts into X_i (resp. on the right, resp. on the left) in an elementary way.

The notation $X \overset{\text{cut}^*}{\rightarrow} Y$ (resp. $X \overset{\text{cut}^*}{\rightarrow}_r Y$, resp. $X \overset{\text{cut}^*}{\rightarrow}_l Y$) will be used to say that X cuts into Y (resp. on the right, resp. on the left).

Note: If X cuts into Y in an *almost* elementary way, then X cuts into Y (proof left to the reader), and therefore (proof also left to the reader), a definition equivalent to the last of the definitions 5.1 is:

DEFINITION 5.1 *(bis)*: Let X and Y be two languages, the language X cuts into Y (resp. on the right, resp. on the left) iff there is a finite sequence $X = X_0, X_1, \ldots, X_m = Y$ such that for all $i \in [1, m]$, X_{i-1} cuts into X_i (resp. on the right, resp. on the left) in an almost elementary way.

Note 5.2: In $(2r)$ and in $(2l)$, the "$\beta \neq \alpha$" is not necessary since, by defining $u_\alpha = u$, one gets that u_α is a prefix of u and that $u_\alpha\, \alpha \in X$. Note also that the definition of cuts would not be changed if one removed the condition "$u \notin X$" in $(2r)$, $(2r)$, $(3r)$, $(3l)$. It is for the sake of convenience that they were included.

Examples 5.3: (The alphabet is $\{a, b, c\}$)

$$\begin{Bmatrix} aaaa \\ aaab \\ ac \\ bc \\ ccc \end{Bmatrix} \text{ cuts into } \begin{Bmatrix} aaa \\ aaab \\ ac \\ bc \\ ccc \end{Bmatrix} \text{ on the right in an elementary way by property } (2r) \text{ with } \alpha = a \text{ and } u = aaa,$$

$$\begin{Bmatrix} aaa \\ aaab \\ ac \\ bc \\ ccc \end{Bmatrix} \text{ cuts into } \begin{Bmatrix} aaa \\ ac \\ bc \\ ccc \end{Bmatrix} \text{ on the right in an elementary way by property } (1) \text{ with } u = aaa \text{ and } v = aaab,$$

$$\begin{Bmatrix} aaa \\ ac \\ bc \\ ccc \end{Bmatrix} \text{ cuts into } \begin{Bmatrix} aaa \\ ac \\ bc \\ cc \end{Bmatrix} \text{ on the left in an elementary way by property } (2l) \text{ with } \alpha = c \text{ and } u = cc,$$

$$\begin{Bmatrix} aaa \\ ac \\ bc \\ cc \end{Bmatrix} \text{ cuts into } \begin{Bmatrix} aaa \\ c \end{Bmatrix} \text{ on the left in an \textit{almost} elementary way by property } (3l) \text{ with } u = c.$$

Therefore, $\left\{\begin{matrix} aaaa \\ aaab \\ ac \\ bc \\ ccc \end{matrix}\right\}$ cuts into $\left\{\begin{matrix} aaa \\ ac \\ bc \\ ccc \end{matrix}\right\}$ on the right,

$\left\{\begin{matrix} aaa \\ ac \\ bc \\ ccc \end{matrix}\right\}$ cuts into $\left\{\begin{matrix} aaa \\ c \end{matrix}\right\}$ on the left and $\left\{\begin{matrix} aaaa \\ aaab \\ ac \\ bc \\ ccc \end{matrix}\right\}$ cuts into $\left\{\begin{matrix} aaa \\ c \end{matrix}\right\}$.

DEFINITIONS 5.4: Let X be a language, X is *innocent* (resp. *innocent on the left*, resp. *innocent on the right*) iff it cannot be cut (resp. cut on the left, resp. cut on the right) except in the obvious way, that is iff [X cuts into Y (resp. on the left, resp. on the right)] implies [$X = Y$ and the sequence of elementary cuts is trivial: $X = X_0 = Y$].

A language is *guilty* (resp. *guilty on the left*, resp. *guilty on the right*) iff it is not innocent (resp. on the left, resp. on the right).

Examples 5.5: (Still with alphabet $\{a, b, c\}$)

$\left\{\begin{matrix} aaaa \\ aaab \\ ac \\ bc \\ ccc \end{matrix}\right\}$ is guilty on both sides and is therefore guilty,

$\left\{\begin{matrix} aaa \\ ac \\ bc \\ ccc \end{matrix}\right\}$ is innocent on the right, guilty on the left and is therefore guilty,

$\left\{\begin{matrix} aaa \\ c \end{matrix}\right\}$ is innocent on both sides and is therefore innocent.

Cuts are a simplification in terms of unavoidability and a convenient tool. We study now the basic properties of cuts we will use for unavoidable languages:

PROPOSITION 5.6: *Let X and Y be two languages such that X cuts into Y, then a bi-infinite word $\aleph$ avoids X iff it avoids Y.*

Proof of proposition 5.6: It is enough to prove the proposition in the case where X cuts into Y in an elementary way (and then one gets the general case by an easy induction on the length of the sequence). There are three cases:

◇ *case* 1: the language Y is $X - \{v\}$ with $u, v \in X$, $u \neq v$, u is a factor of v

If $\aleph$ avoids X, it avoids Y since $Y \subset X$.

If $\aleph$ avoids Y, it avoids u (since $u \in Y$), therefore it avoids v (because u is a factor of v) and therefore it avoids X (it avoids Y and v, and $X = Y \cup \{v\}$).

◇ *case* 2*r*: $Y = X - \{u\,\alpha\} + \{u\}$ with $u \in A^*$, $\alpha \in A$, $u\,\alpha \in X$ and [for every $\beta \in A$, $\beta \neq \alpha$, there is a suffix u_β of u such that $u_\beta\,\beta \in X$].

If $\aleph$ avoids Y, then $\aleph$ avoids $X - \{u\,\alpha\}$ (because $X - \{u\,\alpha\} \subset Y$) and $u\,\alpha$ (because $\aleph$ avoids Y and $u \in Y$, so $\aleph$ avoids u, but u is a factor of $u\,\alpha$, so $\aleph$ avoids also $u\,\alpha$), therefore $\aleph$ avoids X (because $X = (X - \{u\,\alpha\}) + \{u\,\alpha\}$).

If $\aleph$ avoids X, it avoids $X - \{u\,\alpha\}$ (because $X - \{u\,\alpha\} \subset X$). Assume $\aleph$ does not avoid u, this means that u is a factor of $\aleph$. The (or one of the) occurrence of u in $\aleph$ is followed by a letter $\beta \in A$, so that $u\,\beta$ is a factor of $\aleph$. Now, there is u_β suffix of u such that $u_\beta\,\beta$ is in X (If $\beta = \alpha$, then take $u_\alpha = u$), but $u_\beta\,\beta$ is a factor of $\aleph$ (because $u_\beta\,\beta$ is a factor of $u\,\beta$ which is a factor of $\aleph$), which contradicts the fact that $\aleph$ avoids X.

Therefore $\aleph$ avoids $\{u\}$, so $\aleph$ avoids Y (because $\aleph$ avoids $X - \{u\,\alpha\}$ and $\{u\}$, and $Y = (X - \{u\,\alpha\}) + \{u\}$).

◇ *case* 2*l*: $Y = X - \{u\,\alpha\} + \{u\}$ with $u \in A^*$, $\alpha \in A$, $\alpha\,u \in X$ and [for every $\beta \in A$, $\beta \neq \alpha$, there is a prefix u_β of u such that $\beta\,u_\beta \in X$].

The proof of this case is the symmetric of the previous one.

Proposition 5.6 is proved. □

PROPOSITION 5.7: *Let X and Y be two languages. If X cuts into Y, then, for every language Z, [$X \cup Z$ is unavoidable] iff [$Y \cup Z$ is unavoidable].*

Proof: This is an easy corollary of proposition 5.6. □

PROPOSITION 5.8: *Let X and Y be two languages. If X cuts into Y, then X is unavoidable iff Y is unavoidable.*

Proof: This is proposition 5.7 with $Z = \emptyset$. □

PROPOSITION 5.9: *Let X be a language. If X is innocent on the right (resp. innocent on the left, resp. innocent) and unavoidable, then $X = \{\varepsilon\}$.*

Proof of proposition 5.9: Assume X is innocent on the right and $X \neq \{\varepsilon\}$. We show that X is avoidable.

If $X = \emptyset$, then X is avoidable, so we assume from now on that $X \neq \emptyset$. We prove by induction on $L \in \mathbb{N}$ that there is a word $u \in A^*$ of length L, such that u avoids X:

$L = 0$: $X \neq \emptyset$ and $X \neq \{\varepsilon\}$, so there is a word $v \in X$ with $v \neq \varepsilon$, therefore $\varepsilon \notin X$ (because if $\varepsilon \in X$, then the facts that ε is a factor of v, that $\varepsilon \neq v$, and that ε, $v \in X$ contradict the innocence of X) and therefore ε (which is of length 0) avoids X (The only factor of ε is itself).

Assume there is a word v of length L which avoids X, we will find a letter α such that $v\,\alpha$ (which is of length $L+1$) avoids X. Assume that for every $\beta \in A$, $v\,\beta$ does *not* avoid X, this means that there is $x_\beta \in X$ which is a factor of $v\,\beta$. The word x_β is not a factor of v (because v avoids X and $x_\beta \in X$), so x_β is a suffix of $v\,\beta$, and since $x_\beta \neq \varepsilon$ (because $x_\beta \in X$ and $\varepsilon \notin X$), there is a suffix u_β of v such that $x_\beta = u_\beta\,\beta$. Now let α be such that $|\,u_\alpha\,| = \max\limits_{\beta \in A} |\,u_\beta\,|$ and let $u = u_\alpha$. Then $u\,\alpha = x_\alpha \in X$ and for every $\beta \in A$, $\beta \neq \alpha$, u_β is a suffix of u (because u_β and $u = u_\alpha$ are suffix of the same word v and $|\,u_\beta\,| \leq |\,u\,|$) such that $u_\beta\,\beta\,(= x_\beta) \in X$. This contradicts the fact that X is innocent. Therefore, there is a letter $\beta \in A$ such that $v\,\beta$ avoids X, and the induction hypothesis is true for $L+1$.

By induction, there are words of every length which avoid X. So X is avoidable.

The result is the same if X is innocent on the left (by symmetry) and if X is innocent (because if X is innocent, then it is innocent on both side).

Proposition 5.9 is proved. $\square$

THEOREM 5.10: *Let X be a* finite *language, then the following properties are equivalent:*

(i): *X is unavoidable*

(ii): *X cuts into $\{\varepsilon\}$*

(iii): *X cuts into $\{\varepsilon\}$ on the right*

(iv): *X cuts into $\{\varepsilon\}$ on the left*

Proof of theorem 5.10: (iii) $\Rightarrow$ (ii) is obvious.

(ii) $\Rightarrow$ (i) because of proposition 5.8 and because $\{\varepsilon\}$ is unavoidable.

(i) $\Rightarrow$ (iii) is proved by induction on $L = \sum\limits_{w \in X} |\,w\,|$:

$L = 0$: If $\sum_{w \in X} |w| = 0$ and X is unavoidable, then $X = \{\varepsilon\}$ and the result is obvious.

Assume that $[[Y$ is unavoidable and $\sum_{w \in Y} |w| \leq L] \Rightarrow Y$ cuts into $\{\varepsilon\}$ on the right], and that X is an unavoidable language with $\sum_{w \in X} |w| = L+1$, then: X is guilty on the right (because $\{\varepsilon\}$ is the only unavoidable, innocent on the right language, *see* proposition 5.9), therefore, there is Y such that X cuts into Y on the right in an elementary way. But Y is unavoidable (because of proposition 5.8) and $\sum_{w \in Y} |w| < \sum_{w \in X} |w| = L + 1$, so, by the induction hypothesis, Y cuts into ε on the right. Therefore X cuts into $\{\varepsilon\}$ on the right.

By induction, the property is true for every L.

Therefore (i), (ii) and (iii) are equivalent. By symmetry, (i), (ii) and (iv) are equivalent and therefore (i), (ii), (iii) and (iv) are equivalent.

Theorem 5.10 is proved. $\square$

Note 5.11: When X is unavoidable, then it cuts into $\{\varepsilon\}$ whichever way you take, because if you cut X into Y, then Y is unavoidable and therefore it cuts into $\{\varepsilon\}$.

Note 5.12: Proposition 5.10 and note 5.11 give an algorithm to know whether a given finite language X is unavoidable: write down the elements in X and cut as long as you can. The language X is unavoidable iff you can reach $\{\varepsilon\}$ and is not if you get stuck before reaching $\{\varepsilon\}$.

6. FURTHER STUDIES OF CUTS

This section is devoted to natural questions one can ask oneself about cuts.

The main results of this section are, first that for every finite language X, there is a unique *innocent* language $\overline{X}$ such that X cuts into $\overline{X}$, second the equivalence between (where X and Y are two finite languages):

(1) $\overline{X} = \overline{Y}$.

(2) for every bi-infinite word $\aleph$, ([$\aleph$ avoids X] iff [$\aleph$ avoids Y])

and third the equivalence between:

(3) For every *finite* language Z, $X \cup Z$ is unavoidable iff $Y \cup Z$ is unavoidable.

(3′) For every language Z, $X \cup Z$ is unavoidable iff $Y \cup Z$ is unavoidable.

(4) For every *periodic* bi-infinite word $\aleph$, ([$\aleph$ avoids X] iff [$\aleph$ avoids Y]).

The languages satisfying (1) and (2) will be defined to be strongly equivalent.

The languages satisfying (3) and (4) will be defined to be weakly equivalent.

LEMMA 6.1: *Let Y be an innocent language, and v be a finite word which avoids Y, then there is a bi-infinite word $\aleph$ such that $\aleph$ avoids Y and such that v is a factor of $\aleph$.*

Proof of lemma 6.1: If $\varepsilon \in Y$ (which in fact would imply $Y = \{\varepsilon\}$), then v cannot exist, because every word contains ε as a factor, so one can assume from now on that $\varepsilon \notin Y$.

One builds by induction some letters $a_1, \ldots, a_L, \ldots$ and $b_1, \ldots, b_L, \ldots$ such that the word defined by $u_L = a_L \ldots a_1 v b_1 \ldots b_L$ avoids Y:

$L = 0$: No letter needs to be built and $u_0 = v$ avoids Y.

Assume $(a_i)_{i \leq L}$ and $(b_i)_{i \leq L}$ are built so that $u_L = a_L \ldots a_1 v b_1 \ldots b_L$ avoids Y.

Assume that for every letter $\beta \in A$, there is a word $x_\beta \in Y$ which is a factor of $u_L \beta$. Because u_L avoids Y, x_β is not a factor of u_L, and therefore x_β must be a suffix of $u_L \beta$. Since $x_\beta \neq \varepsilon$ (because $x_\beta \in Y$ and $\varepsilon \notin Y$), there must be a suffix u_β of u_L such that $x_\beta = u_\beta \beta$. Let α be such that $|u_\alpha| = \max_{\beta \in A} |u_\beta|$ and let $u = u_\alpha$, then $u \alpha \in Y$ and for every letter $\beta \neq \alpha$, $u_\beta \beta \in Y$ and u_β is a suffix of u (because they both are suffixes of u_L and $|u| \geq |u_\beta|$). This contradicts the innocence of Y, so there is a letter b_{L+1} such that no word in Y is a factor of $u_L b_{L+1}$, that is, such that $u_L b_{L+1}$ avoids Y.

By a symmetric reasoning (on the left, using $u_L b_{L+1}$ instead of u_L), there is a_{L+1} such that $a_{L+1} (u_L b_{L+1})$ avoids Y.

Therefore, $(a_i)_{i \leq L+1}$ and $(b_i)_{i \leq L+1}$ are built with the required property.

By induction, $(a_i)_{i \in \mathbb{N}^*}$ and $(b_i)_{i \in \mathbb{N}^*}$ are built and it is easy to see that the bi-infinite word $\aleph = \ldots a_p \ldots a_2 a_1 v b_1 b_2 \ldots b_q \ldots$ avoids Y.

Lemma 6.1 is proved. $\square$

COROLLARY 6.2: *Let Y be an innocent language. For every $y \in Y$, there is a bi-infinite word $\aleph_y$ which avoids $Y - \{y\}$, but contains y as a factor.*

Proof of corollary 6.2: Because Y is innocent, no $y' \in Y - \{y\}$ is a factor of y, so y avoids $Y - \{y\}$. The language Y is innocent, so $Y - \{y\}$

is also innocent. Therefore corollary 6.2 is a consequence of proposition 6.1 with $(y, Y - \{y\})$ used instead of (v, Y).

Corollary 6.2 is proved. $\square$

Note 6.3 on corollary 6.2: The words $\aleph_y$ are not stated to be periodic and in fact cannot be. Consider for example $Y = \{ab, ba\}$.

DEFINITIONS 6.4: Let S be a set of bi-infinite words, then $\tilde{S}$ and $\widehat{S}$ are defined as:

$\tilde{S} = \{u \in A^* \mid \forall \aleph \in S,\ u$ is *not* a factor of $\aleph\}$.

$\widehat{S} = \{u \in \tilde{S} \mid \forall v \in A^*,\ [v \in \tilde{S}$ and v is a factor of $u] \Rightarrow [v = u]\}$.

$\tilde{S}$ is the set of the words that never appear as a factor of a word in S, and $\widehat{S}$ is the set of the minimal elements of $\tilde{S}$ for the order "is a factor of" ($\widehat{S}$ is the subset of $\tilde{S}$ consisting of all the words u such that no factor of u (except u itself) is in $\tilde{S}$).

PROPOSITION 6.5: *Let Y be an innocent language and S be the set of the bi-infinite words avoiding Y, then $Y = \widehat{S}$.*

Proof of proposition 6.5: S is the set of the bi-infinite words which avoid Y, so $\forall y \in Y$, $\forall \aleph \in S$, y is not a factor of $\aleph$, and therefore $y \in \tilde{S}$. So $\underline{Y \subset \tilde{S}}$.

Let $s \in \widehat{S}$, assume $s \notin Y$, then s avoids Y (Indeed, assume it doesn't, then there is $y \in Y$ which is a factor of s. Since $y \in Y$ and $s \notin Y$, y cannot be s, so y is a proper factor of s, but since $y \in \tilde{S}$, this implies that $s \in \widehat{S}$ has a proper factor in $\tilde{S}$. This contradicts the definition of $\widehat{S}$). So thanks to lemma 6.1, there is a bi-infinite word $\aleph$ which contains s as a factor and avoids Y. Because $\aleph$ avoids Y, $\aleph$ is in S. Because $s \in \widehat{S} \subset \tilde{S}$ and $\tilde{S}$ is the set of the words which are factor of no word in S, s is not a factor of $\aleph$, but $\aleph$ was assumed to have s as a factor. There is a contradiction, therefore if $s \in \widehat{S}$, then $s \in Y$, so that $\underline{\widehat{S} \subset Y}$.

Assume now that $\widehat{S} \neq Y$, then there is $x \in Y - \widehat{S}$. But $x \in Y$, so $x \in \tilde{S}$, therefore x has a factor y in $\widehat{S}$ (It is straightforward to see by induction on $|s|$ that every word s in $\tilde{S}$ has a factor s' in $\widehat{S}$), but $\widehat{S} \subset Y$, so $y \in Y$. So $x, y \in Y$, $x \neq y$ (because $x \notin \widehat{S}$ and $y \in \widehat{S}$) and y is a factor of x. This contradicts the innocence of Y, so $\widehat{S} = Y$.

Proposition 6.5 is proved. $\square$

THEOREM 6.6: *Let Y be a finite language, then there is a unique* innocent *language $\overline{Y}$ such that Y cuts into $\overline{Y}$.*

Proof of theorem 6.6: An easy induction on $\sum_{x \in Y} |x|$ shows that there is an innocent set $\overline{Y}$ such that Y cuts into $\overline{Y}$. Assume now that $\overline{Y}$ and $\overline{Y}'$ are two innocent languages such that Y cuts into $\overline{Y}$ and into $\overline{Y}'$. Then thanks to proposition 5.6 (and because Y cuts into $\overline{Y}$ and into $\overline{Y}'$), for every bi-infinite word $\aleph$, ([$\aleph$ avoids Y] iff [$\aleph$ avoids $\overline{Y}$]) and ([$\aleph$ avoids Y] iff [$\aleph$ avoids $\overline{Y}'$]), so that ([$\aleph$ avoids $\overline{Y}$] iff [$\aleph$ avoids $\overline{Y}'$]). Therefore S and S', the sets of bi-infinite words avoiding respectively $\overline{Y}$ and $\overline{Y}'$, are the same. So thanks to proposition 6.5: $\overline{Y} = \widehat{S} = \widehat{S'} = \overline{Y'}$. Therefore, there is a unique innocent language $\overline{Y}$ such that Y cuts into $\overline{Y}$.

Theorem 6.6 is proved. □

DEFINITION 6.7: Let X be a finite language, we will call *acquitted* X the unique innocent language given by theorem 6.6. Acquitted X will be denoted by $\overline{X}$.

PROPOSITION 6.8: *Let X and Y be two finite languages, the following two propositions are equivalent:*

(1) $\overline{X} = \overline{Y}$.

(2) *for every bi-infinite word $\aleph$, ([$\aleph$ avoids X] iff [$\aleph$ avoids Y]).*

Proof of proposition 6.8: (1) $\Rightarrow$ (2): According to proposition 5.6 (using that X cuts into $\overline{X}$), [$\aleph$ avoids X] iff [$\aleph$ avoids $\overline{X}$] and similarly [$\aleph$ avoids Y] iff [$\aleph$ avoids $\overline{Y}$], so that if $\overline{X} = \overline{Y}$, then [$\aleph$ avoids X] iff [$\aleph$ avoids Y].

(2) $\Rightarrow$ (1): Let S_X, $S_{\overline{X}}$, S_Y, $S_{\overline{Y}}$ be the sets of the sets of the bi-infinite words avoiding respectively X, $\overline{X}$, Y, $\overline{Y}$. Property (2) means that $S_X = S_Y$. Thanks to proposition 5.6, ([$\aleph$ avoids X] iff [$\aleph$ avoids $\overline{X}$]) and ([$\aleph$ avoids Y] iff [$\aleph$ avoids $\overline{Y}$]), so that $S_X = S_{\overline{X}}$ and $S_Y = S_{\overline{Y}}$. Therefore $S_{\overline{X}} = S_{\overline{Y}}$. But $\overline{X}$ and $\overline{Y}$ are innocent, so thanks to proposition 6.5, $\overline{X} = \widehat{S_{\overline{X}}} = \widehat{S_{\overline{Y}}} = \overline{Y}$.

Proposition 6.8 is proved. □

DEFINITION 6.9: Two finite languages X and Y are *strongly equivalent* iff they satisfy the properties of proposition 6.8. The fact that X and Y are strongly equivalent will be denoted by $X \sim_s Y$.

Note 6.10: One could get properties "on the right", which are similar to the previous ones, by considering infinite words instead of bi-infinite ones.

PROPOSITION 6.11: *Let X and Y be two languages, then the following two properties are equivalent:*

(3) *For every* finite *language* Z, $X \cup Z$ *is unavoidable iff* $Y \cup Z$ *is unavoidable.*

(3′) *For every language* Z, $X \cup Z$ *is unavoidable iff* $Y \cup Z$ *is unavoidable.*

If X *and* Y *are finite, then the above two properties are equivalent to the following one:*

(4) *For every* periodic *bi-infinite word* $\aleph$, ([$\aleph$ *avoids* X] *iff* [$\aleph$ *avoids* Y]).

Proof of proposition 6.11: (3′) $\Rightarrow$ (3) is obvious.

(3) $\Rightarrow$ (3′): Assume (3). Let Z be a language. If $X \cup Z$ is unavoidable, then (*see* proposition 3.5), there is a *finite* sublanguage of $X \cup Z$ which is unavoidable, so there are some *finite* sublanguages X' and Z' of X and Z such that $X' \cup Z'$ is unavoidable. Since $X' \subset X$, $X \cup Z'$ is unavoidable. Since Z' is finite and (3) is assumed to be satisfied, $Y \cup Z'$ is unavoidable. Since $Z' \subset Z$, $Y \cup Z$ is unavoidable. By a symmetric argument, if $Y \cup Z$ is unavoidable, then $X \cup Z$ is unavoidable. So (3′) is satisfied.

(4) $\Rightarrow$ (3) is straightforward since a *finite* language is unavoidable iff no periodic bi-infinite word avoids it.

$\neg$(4) $\Rightarrow$ $\neg$(3): Assume for example that there is a periodic bi-infinite word $\aleph$ which avoids X, but not Y, then according to proposition 4.3, there is a *finite* language Z such that every periodic bi-infinite word, *but* $\aleph$, contains a factor in Z. Since $\aleph$ avoids X and avoids Z, it avoids $X \cup Z$ which therefore is *not* unavoidable. Since $\aleph$ does not avoid Y, there is $y \in Y$ which is a factor of $\aleph$. Every periodic bi-infinite word contains a factor in $Z \cup \{y\}$ ($\aleph$ contains y and every other word contains an element in Z) and $Z \cup \{y\}$ is finite (since Z is finite), so $Z \cup \{y\}$ is unavoidable, and therefore $Z \cup Y$ is unavoidable. So (3) is not satisfied.

Proposition 6.11 is proved. $\square$

Note 6.12: (4) $\Rightarrow$ (3) is not true if X or Y can be infinite. *See* for example $X = \{\varepsilon\}$ and $Y = \{uu \mid u \in A^+\}$ the set of the non-ε squares on the alphabet $\{a, b, c\}$, with the help of [2]. On the other hand, $\neg$(4) $\Rightarrow$ $\neg$(3) is valid even if X or Y is infinite.

Definition 6.13: Two finite languages X and Y are *weakly equivalent* iff they satisfy the properties of proposition 6.11. The fact that X and Y are weakly equivalent will be denoted by $X \sim_w Y$.

Proposition 6.14: *Let* X *and* Y *be two finite languages. If* X *and* Y *are strongly equivalent, then there are weakly equivalent.*

Proof: Obvious. □

Example 6.15: Let $A = \{ a, b \}$ and n be an integer. Let $X = \{ a^n b \}$ and $Y = \{ ba^n \}$, then X and Y are weakly, but not strongly equivalent.

Proof of the statement of example 6.15: To prove that $a^n b \sim_w ba^n$, one can use any definition of weak equivalence.

Proof 1: Let $\aleph$ be a bi-infinite *periodic* word, then $\aleph$ avoids $a^n b$ iff $\aleph = a^{\mathbb{Z}}$ or $\aleph$ avoids a^n. Indeed, if $\aleph = a^{\mathbb{Z}}$ or $\aleph$ avoids a^n, then $\aleph$ avoids $a^n b$. If $\aleph \neq a^{\mathbb{Z}}$ and $\aleph$ contains a^n as a factor, then ($\aleph$ contains a^n as a factor and is periodic) there is a word u such that $\aleph = (a^n u)^{\mathbb{Z}}$. Since $\aleph \not\equiv a^{\mathbb{Z}}$, u contains the letter b, so can be written $u = a^k bv$, so $\aleph = (a^n a^k bv)^{\mathbb{Z}} = (a^n (a^n b) v)^{\mathbb{Z}}$ contains $a^n b$ as a factor. By symmetry, $\aleph$ avoids ba^n iff $\aleph = a^{\mathbb{Z}}$ or $\aleph$ avoids a^n, so $\aleph$ avoids $a^n b$ iff $\aleph$ avoids ba^n.

Proof 2: Let Z be a finite language.

If Z contains no power of a, then $a^{\mathbb{Z}}$ avoids both $Z \cup \{ a^n b \}$ and $Z \cup \{ ba^n \}$.

If $a^k \in Z$ and a^k is the smallest power of a in Z, then let $Z' = Z - \{ a^k \}$.

If $k \leq n$, then (a^k is a factor of $a^n b$) $Z \cup \{ a^n b \} = Z' \cup \{ a^k, a^n b \}$ cuts into $Z' \cup \{ a^k \} = Z$, and similarly, $Z \cup \{ ba^n \}$ cuts into $Z' \cup \{ a^k \} = Z$. Therefore, $Z \cup \{ a^n b \}$ is unavoidable iff Z is unavoidable iff $Z \cup \{ ba^n \}$ is unavoidable.

If $k \geq n$, then $Z \cup \{ a^n b \} = Z' \cup \{ a^k, a^n b \}$ cuts into $Z' \cup \{ a^n \}$ (By induction on k : $k = n$: because a^n is a factor of $a^n b$, if $k > n$, because $a^k = a^{(k-n-1)} a^n a$, so $Z' \cup \{ a^k, a^n b \}$ cuts into $Z' \cup \{ a^{k-1}, a^n b \}$ by using the cut ($2r$) with $\alpha = a$, $u = a^{k-1}$ and $u_b = a^n$), and by symmetry, $Z \cup \{ ba^n \}$ cuts into $Z' \cup \{ a^n \}$, so $Z \cup \{ a^n b \}$ is unavoidable iff $Z' \cup \{ a^n \}$ is unavoidable iff $Z \cup \{ ba^n \}$ is unavoidable.

Now $a^n b \not\sim_s ba^n$ because $\aleph = (a_i)_{i \in \mathbb{Z}}$ with $a_i = a$ for $i \leq 0$ and $a_i = b$ for $i > 0$ avoids ba^n, but not $a^n b$ (or because they are distinct singletons and every singleton is innocent). **The statement of example 6.15 is proved.** □

THEOREM 6.16: *Let $A = \{ a, b \}$, let u and v be two words on A, the singletons $\{ u \}$ and $\{ v \}$ are weakly equivalent iff $[u = v]$ or if [there is an integer n such that $\{ u, v \} = \{ a^n b, ba^n \}$ or such that $\{ u, v \} = \{ b^n a, ab^n \}$].*

Proof of theorem 6.16: It is obvious that if $u = v$, then $\{ u \}$ and $\{ v \}$ are strongly and therefore weakly equivalent, and it is easy to see from example 6.15 and some symmetry considerations that

$$\text{if } \{u, v\} = \{a^n b, ba^n\} \quad \text{or} \quad \text{if } \{u, v\} = \{b^n a, ab^n\},$$

then u and v are weakly equivalent.

Assume now that u and v are weakly equivalent, but different. We have to show that $\{u, v\} = \{a^n b, ba^n\}$ or $\{u, v\} = \{b^n a, ab^n\}$ for some n.

Let T be an integer, then there is at most, up to translation, one bi-infinite word, periodic of period T, containing u as a factor iff $[T \leq |u|]$. Indeed:

Assume $T \leq |u|$. Let u_T be the prefix of u with length T. If $\aleph$ is of period T and contains u as a factor, then it contains u_T, and therefore ($|u_T| = T$ the period of $\aleph$) $\aleph \equiv (u_T)^{\mathbb{Z}}$. So there is only one (up to translation) periodic bi-infinite word of period T that might contain u as a factor.

Assume $T = |u| + k$, with $k > 0$, then $\aleph_a = (ua^k)^{\mathbb{Z}}$ and $\aleph_b = (ub^k)^{\mathbb{Z}}$ are two words of period T containing u. Furthermore, $\aleph_a \not\equiv \aleph_b$ (Indeed, if $u_a^{\mathbb{Z}} \equiv u_b^{\mathbb{Z}}$ with $|u_a| = |u_b|$, then there are as many a's in u_a as in u_b, but $\aleph_a = (ua^k)^{\mathbb{Z}}$, $\aleph_b = (ub^k)^{\mathbb{Z}}$, $|ua^k| = |ub^k| = T$ and $|ua^k|_a = |u|_a + k \neq |u|_a = |ub^k|_a$). So there are at least two words of period T containing u as a factor and which are different up to translation.

One has the similar property for v and therefore

$$\begin{aligned}
|u| &= \max\{k \in \mathbb{N} \mid \text{. There is at most, up to translation,}\\
&\qquad \text{one bi-infinite word which is periodic of period } T,\\
&\qquad \text{and which is a factor of } u\}\\
&= \max\{k \in \mathbb{N} \mid \text{. There is at most, up to translation,}\\
&\qquad \text{one bi-infinite word which is periodic of period } T,\\
&\qquad \text{and which is a factor of } v\}\\
&\quad (\text{because } u \sim_w v, \text{ so a periodic bi-infinite word } \aleph\\
&\qquad \text{contains } u \text{ as a factor iff it contains } v \text{ as a factor})\\
&= |v|.
\end{aligned}$$

So u and v have the same length.

Now $v^{\mathbb{Z}}$ is periodic of period $|v|$ and contains v as a factor, therefore ($|u| = |v|$ and $u \sim_w v$), $v^{\mathbb{Z}}$ is periodic of period $|u|$ and contains u as a factor, but $u^{\mathbb{Z}}$ is also periodic of period $|u|$ with u as a factor, but there is at most one bi-infinite word which is periodic of period $|u|$ and which contains u as a factor, therefore $u^{\mathbb{Z}} \equiv v^{\mathbb{Z}}$, and since $|u| = |v|$, this implies that there are some words t and s such that $\underline{u = st}$ and $\underline{v = ts}$.

Moreover, $s \neq \varepsilon$ and $t \neq \varepsilon$ because $u \neq v$.

Let $L = |u| = |v|$. The word $\aleph = (ua^L)^{\mathbb{Z}}$ contains u as a factor and is periodic, therefore (since $u \sim_w v$), $\aleph$ contains v and ($\aleph$ is periodic of period

$2L$ and $|v| = L$) there is a word w of length L such that $\aleph \equiv (vw)^{\mathbb{Z}}$. Now $|ua^L| = |vw| (= 2L)$ and $(ua^L)^{\mathbb{Z}} \equiv (vw)^{\mathbb{Z}}$, therefore there are as many a's in ua^L as in vw, that is $|ua^L|_a = |vw|_a$. Since $|ua^L|_a = |s|_a + |t|_a + L$ and $|vw|_a = |tsw|_a = |t|_a + |s|_a + |w|_a$, one has $|w|_a = L = |w|$ so $w = a^L$ and $(ua^L)^{\mathbb{Z}} \equiv (va^L)^{\mathbb{Z}}$.

Therefore, there is an integer k_a such that $ua^{k_a} = a^{k_a} v$ or such that $va^{k_a} = a^{k_a} u$ (Indeed $(ua^L)^{\mathbb{Z}} \equiv (va^L)^{\mathbb{Z}}$ implies that there are words s' and t' such that $ua^L = t' s'$ and $va^L = s' t'$. Because $|t' s'| = 2L$, one has $|s'| \geq L$ or $|t'| \geq L$. If $|s'| \geq L$, then let $k_a = |t'| (\leq L)$.

From $va^L = s' t'$, one gets that $t' = a^{k_a}$ and $s' = va^{L-k_a}$. From $ua^L = t' s'$, one gets that $ua^L = a^{k_a} va^{L-k_a}$, so that $ua^{k_a} = a^{k_a} v$. If $|t'| \geq L$, one gets a number k_a such that $va^{k_a} = a^{k_a} u$ in a similarly way). Note that $k_a \neq 0$ (because $k_a = 0$ would imply that $u = v$).

The latter result can be rewritten: there is an integer $k_a \neq 0$ such that $sta^{k_a} = a^{k_a} ts$ or such that $tsa^{k_a} = a^{k_a} st$. A similar reasoning shows that there is an integer $k_b \neq 0$ such that $stb^{k_b} = b^{k_b} ts$ or such that $tsb^{k_b} = b^{k_b} st$.

Since $sta^{k_a} = a^{k_a} ts$ forces the first letter of s to be an a, while $stb^{k_b} = b^{k_b} ts$ forces the first letter of s to be a b, those two conditions are incompatible, so there are integers k_a and k_b with $k_a \neq 0$, $k_b \neq 0$, such that $sta^{k_a} = a^{k_a} ts$ and $tsb^{k_b} = b^{k_b} st$ or such that $tsa^{k_a} = a^{k_a} st$ and $stb^{k_b} = b^{k_b} ts$. For symmetry reasons, we can assume that $\underline{sta^{k_a} = a^{k_a} ts}$ and $\underline{tsb^{k_b} = b^{k_b} st}$.

If s is not a power of a, then a^{k_a} is a prefix of s (because $sta^{k_a} = a^{k_a} ts$, so s is a prefix of a^{k_a} or a^{k_a} is a prefix of s, but s prefix of a^{k_a} would contradict the hypothesis that s is not a power of a) and also a suffix of a (by a symmetric argument). Since s is not a power of a, this implies that there is a word z_a such that $s = a^{k_a} z_a a^{k_a}$. So, either there is an integer m_a such that $s = a^{m_a}$ (and $m_a \neq 0$ because $s \neq \varepsilon$) or there is a word z_a such that $s = a^{k_a} z_a a^{k_a}$. Similarly, either there is an integer $m_b \neq 0$ such that $t = b^{m_b}$ or there is a word z_b such that $t = b^{k_b} z_b b^{k_b}$.

◇ If $s = a^{m_a}$ and b^{m_b}, then $u (= st) = a^{m_a} b^{m_b}$ and $v (= ts) = b^{m_b} a^{m_a}$.

◇ If $s = a^{m_a}$ and $t = b^{k_b} z_b b^{k_b}$, then the equation $tsb^{k_b} = b^{k_b} st$ becomes $b^{k_b} z_b b^{k_b} a^{m_a} b^{k_b} = b^{k_b} a^{m_a} b^{k_b} z_b b^{k_b}$, that is $z_b b^{k_b} a^{m_a} = a^{m_a} b^{k_b} z_b$. This last equation forces a^{m_a} to be a prefix of z_b, so there is a word w_b such that $z_b = a^{m_a} w_b$ and the equation $z_b b^{k_b} a^{m_a} = a^{m_a} b^{k_b} z_b$ becomes $a^{m_a} w_b b^{k_b} a^{m_a} = a^{m_a} b^{k_b} a^{m_a} w_b$, that is $w_b b^{k_b} a^{m_a} = b^{k_b} a^{m_a} w_b$. So there is an integer p such that $w_b = (b^{k_b} a^{m_a})^p$ (Because, by induction on $|w|$, the relation $wb^{k_b} a^{m_a} = b^{k_b} a^{m_a} w$ forces $w = (a^{m_a} b^{k_b})^p$

for some integer p: This is true if $|w| = 0$, and if $|w| \neq 0$, then the relation $wb^{k_b} a^{m_a} = b^{k_b} a^{m_a} w$ forces $b^{k_b} a^{m_a}$ to be a prefix of w, so $w = b^{k_b} a^{m_a} w'$ and the relation $wb^{k_b} a^{m_a} = b^{k_b} a^{m_a} w$ become $w' b^{k_b} a^{m_a} = b^{k_b} a^{m_a} w'$, so by an induction, $w' = (b^{k_b} a^{m_a})^{p'}$ and therefore $w = b^{k_b} a^{m_a} w' = (b^{k_b} a^{m_a})^{p'+1})$. Therefore

$$z_b = a^{m_a} w_b = a^{m_a} (b^{k_b} a^{m_a})^p,$$

$$t = b^{k_b} z_b b^{k_b} = b^{k_b} a^{k_a} (b^{k_b} a^{m_a})^p b^{k_b} = (b^{k_b} a^{m_a})^{p+1} b^{k_b},$$

$$u = st = a^{m_a} (b^{k_b} a^{m_a})^{p+1} b^{k_b} = (a^{m_a} b^{k_b})^{p+2}$$

and

$$v = ts = (b^{k_b} a^{m_a})^{p+1} b^{k_b} a^{m_a} = (b^{k_b} a^{m_a})^{p+2}.$$

◇ If $s = a^{k_a} z_a a^{k_a}$ and $t = b^{m_b}$, one concludes that there is an integer p such that $u = (a^{k_a} b^{m_b})^{p+2}$ and $v = (b^{m_b} a^{k_a})^{p+2}$ by a reasoning symmetric to the previous one.

◇ If $s = a^{k_a} z_a a^{k_a}$ and $t = b^{k_b} z_b b^{k_b}$, then the equations $sta^{k_a} = a^{k_a} ts$ and $tsb^{k_b} = b^{k_b} st$ become

$$a^{k_a} z_a a^{k_a} b^{k_b} z_b b^{k_b} a^{k_a} = a^{k_a} b^{k_b} z_b b^{k_b} a^{k_a} z_a a^{k_a}$$

and

$$b^{k_b} z_b b^{k_b} a^{k_a} z_a a^{k_a} b^{k_b} = b^{k_b} a^{k_a} z_a a^{k_a} b^{k_b} z_b b^{k_b},$$

that is:

$$z_a a^{k_a} b^{k_b} z_b b^{k_b} = b^{k_b} z_b b^{k_b} a^{k_a} z_a \tag{1}$$

and

$$z_b b^{k_b} a^{k_a} z_a a^{k_a} = a^{k_a} z_a a^{k_a} b^{k_b} z_b \tag{2}$$

One shows by induction on N that

$$\begin{cases} \text{Either } b^{k_b} (a^{k_a} b^{k_b})^N \text{ is a prefix of } z_a \\ \text{or } \exists n_a \text{ such that } z_a = b^{k_b} (a^{k_a} b^{k_b})^{n_a} \end{cases}$$

and

$$\begin{cases} \text{Either } a^{k_a} (b^{k_b} a^{k_a})^N \text{ is a prefix of } z_b \\ \text{or } \exists n_b \text{ such that } z_b = a^{k_a} (b^{k_b} a^{k_a})^{n_b} \end{cases}$$

$N = 0$: The equation (1) forces b^{k_b} to be a prefix of z_a while the equation (2) forces a^{k_a} to be a prefix of z_b.

Assume the proposition is true for N.

If $z_a = b^{k_b} (a^{k_a} b^{k_b})^{n_a}$ for some integer n_a (rank N) then $z_a = b^{k_b} (a^{k_a} b^{k_b})^{n_a}$ for the same integer n_a (rank $N + 1$).

If $z_b = a^{k_a} (b^{k_b} a^{k_a})^{n_b}$ for some integer n_b (rank N) then $z_b = a^{k_a} (b^{k_b} a^{k_a})^{n_n}$ for the same integer n_b (rank $N + 1$).

If $b^{k_b} (a^{k_a} b^{k_b})^N$ is a prefix of z_a and $z_b = a^{k_a} (b^{k_b} a^{k_a})^{n_b}$, then there is a word w_a such that $z_a = b^{k_b} (a^{k_a} b^{k_b})^N w_a$ and the equation (1) induces that $z_a a^{k_a} b^{k_b} = b^{k_b} (a^{k_a} b^{k_b})^N w_a a^{k_a} b^{k_b}$ is a prefix of $b^{k_b} a^{k_a} (b^{k_b} a^{k_a})^{n_b} b^{k_b} a^{k_a} b^{k_b} (a^{k_a} b^{k_b})^N w_a$, so that $w_a a^{k_a} b^{k_b}$ is a prefix of $(a^{k_a} b^{k_b})^{n_b+2} w_a$. Therefore either $w_a = \varepsilon$ and one has $z_a = b^{k_b} (a^{k_a} b^{k_b})^{n_a}$ with $n_a = N$ or $a^{k_a} b^{k_b}$ is a prefix of w_a and $b^{k_b} (a^{k_a} b^{k_b})^{N+1}$ is a prefix of $z_a = b^{k_b} (a^{k_a} b^{k_b}) w_a$.

If $a^{k_a} (b^{k_b} a^{k_a})^N$ is a prefix of z_b and $z_a = b^{k_b} (a^{k_a} b^{k_b})^{n_a}$, then one can conclude by an argument symmetric to the one in the previous case.

If $b^{k_b} (a^{k_a} b^{k_b})^N$ and $a^{k_a} (b^{k_b} a^{k_a})^N$ are respectively prefixes of z_a and z_b, then there are words w_a and w_b such that $z_a = b^{k_b} (a^{k_a} b^{k_b})^N w_a$ and $z_b = a^{k_a} (b^{k_b} a^{k_a})^N w_b$. If $w_a = \varepsilon$ or $w_b = \varepsilon$, then $z_a = b^{k_b} (a^{k_a} b^{k_b})^{n_a}$ with $n_a = N$ or $z_b = a^{k_a} (b^{k_b} a^{k_a})^{n_b}$ with $n_b = N$, and the situation has already been considered, so we can assume that $w_a \neq \varepsilon$ and that $w_b \neq \varepsilon$. The equation (1) induces that $z_a a^{k_a} b^{k_b} = b^{k_b} (a^{k_a} b^{k_b})^N w_a a^{k_a} b^{k_b}$ is a prefix of $b^{k_b} a^{k_a} (b^{k_b} a^{k_a})^N w_b b^{k_b} a^{k_a} z_a$, so that

$$w_a a^{k_a} b^{k_b} \text{ is a prefix of } a^{k_a} w_b b^{k_b} a^{k_a} \tag{3}$$

and the equation (2) induces that

$$w_b b^{k_b} a^{k_a} \text{ is a prefix of } b^{k_b} w_a a^{k_a} b^{k_b} \tag{4}$$

From (3), one gets that $w_a = a^{l_a}$, $l_a < k_a$ or that a^{k_a} is a prefix of w_a.

From (4), one gets that $w_b = b^{l_b}$, $l_b < k_b$ or that b^{k_b} is a prefix of w_b.

But $w_a = a^{l_a}$, $l_a < k_a$ is impossible because of (3) and because w_b begins with a b. For similar reason, $w_b = b^{l_b}$, $l_b < k_b$ is impossible. Therefore a^{k_a} is a prefix of w_a and b^{k_b} is a prefix of w_b, now (3) again shows that $a^{k_a} b^{k_b}$ is a prefix of w_a and (4) shows that $b^{k_b} a^{k_a}$ is a prefix of w_b, so $b^{k_b} (a^{k_a} b^{k_b})^{N+1}$ is a prefix of $z_a = b^{k_b} (a^{k_a} b^{k_b})^N w_a$ and $a^{k_a} (b^{k_b} a^{k_a})^{N+1}$ is a prefix of z_b.

The proposition is true for $N + 1$.

So the proposition is true for every N.

For large N's (for example, $N > |z_a|$, $N > |z_b|$), $b^{k_b} (a^{k_a} b^{k_b})^N$ cannot be a prefix of z_a, nor $a^{k_a} (b^{k_b} a^{k_a})^N$ be prefix of z_b, therefore there are some integers n_a and n_b such that $z_a = b^{k_b}(a^{k_a}b^{k_b})^{n_a}$ and $z_b = a^{k_a}(b^{k_b}a^{k_a})^{n_b}$, therefore $s = a^{k_a}z_a a^{k_a} = (a^{k_a}b^{k_b})^{n_a+1}a^{k_a}$ and $t = b^{k_b}z_b b^{k_b} = (b^{k_b}a^{k_a})^{n_b+1}b^{k_b}$, so that $u = st = (a^{k_a}b^{k_b})^n$ and $v = ts = (b^{k_b}a^{k_a})^n$ with $n = n_a + n_b + 3$.

In every case, one finds that there are some positive integers n_a, n_b, n such that $u = (a^{n_a}b^{n_b})^n$ and $v = (b^{n_b}a^{n_a})^n$.

Assume $n \geq 2$, then $(a^{n_a+1}b^{n_b}(a^{n_a}b^{n_b})^{n-2}a^{n_a}b^{n_b+1})^{\mathbb{Z}}$ avoids v but not u, this contradicts the assumption that $u \sim_w v$, so $n = 1$, $u = a^{n_a}b^{n_b}$ and $v = b^{n_b}a^{n_a}$.

Assume now that $n_a \geq 2$ and that $n_b \geq 2$, then $(a^{n_a}b^{n_b}ab)^{\mathbb{Z}}$ avoids v, but not u, so $n_a = 1$ or $n_b = 1$, that is $\begin{cases} u = ab^{n_b} \\ v = b^{n_b}a \end{cases}$ or $\begin{cases} u = a^{n_a}b \\ v = ba^{n_a} \end{cases}$ which is what we wanted to prove.

Theorem 6.16 is proved. $\square$

7. SOME DIRECT PROOFS ON ACQUITTED X

The first aim of this section is to give a direct proof of proposition 6.6, stating the uniqueness of acquitted X. This proof is more natural than the one given in the previous section, but it turns out to be quite tedious and bulky because of the high number of cases which are to be considered, even though it is not very difficult.

The second part of this section states that one can get acquitted X by acquitting first on the left, then on the right (or conversely).

PROPOSITION 7.1: *Let X be a finite language, then there is one and only one language $\overline{X}$ (resp. $\overline{X}^r$, resp. $\overline{X}^l$) of words which is innocent (resp. on the right, resp. on the left) and such that X cuts into $\overline{X}$ (resp. into $\overline{X}^r$ on the right, resp. into $\overline{X}^l$ on the left).*

Proof of proposition 7.1: The following lemma is the core of the proof of proposition 7.1:

LEMMA 7.2: *Let X be a finite language. Assume Y and Y' are two languages such that X cuts into Y and into Y' (resp. on the right, resp. on the left) in an elementary way, then there is a language Z such that both Y and Y' cut (resp. on the right, resp. on the left) into Z.*

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ Y \arrow[dr, "*"] & & Y' \arrow[dl, "*"'] \\ & Z & \end{tikzcd}$$

Proof of lemma 7.2: Only the skeleton of the proof will be given. I will just indicate the different cases to be considered, and for each of them, what the set Z is. Readers interested in details can look in [18] (It is not very difficult to make out those details, it is just very long and tedious).

Assume X cuts into Y and into Y' on the right in an elementary way. Should Y and Y' be interchanged, we can split the problem in three main cases.

Case 1

There are u, v in X such that $u \neq v$, u is a factor of v, and $Y = X - \{ v \}$, and there are u', v' in X such that $u' \neq v'$, u' is a factor of v', and $Y = X - \{ v' \}$, and we can assume that $| v' | \leq | v |$. This case subdivides in three cases:

Case 1.1: $v = v'$: The conclusion is obvious with $Z = Y = Y'$.

Case 1.2: $u = v'$: $Z = X - \{ v, v' \} = Y - \{ v' \} = Y' - \{ v \}$.

Case 1.3: $v' \neq v$ and $v' \neq u$: $Z = X - \{ v, v' \} = Y - \{ v' \} = Y' - \{ v \}$.

Case 2

There are α in A and u in A^* such that $u\,\alpha \in X$, such that for every letter $\beta \neq \alpha$, there is a suffix u_β of u such that $u_\beta\,\beta \in X$, and such that $Y = X - \{ u\,\alpha \} + \{ u \}$, and there are u', v' in X such that $u' \neq v'$, u' is a factor of v', and $Y = X - \{ v' \}$.

If $u \in X$, then let $v = u\,\alpha$: u, $v \in X$, u is a factor of v and $Y = X - \{ u\,\alpha \} + \{ u \} =_{u \in X} X - \{ u\,\alpha \} = X - \{ v \}$. This situation has already been studied in case 1. Therefore, one can assume from now on that $u \notin X$. This case subdivides in five cases:

Case 2.1: $u\,\alpha = v'$ and u' is a factor of u :

$$Z = X - \{ v' \} = Y - \{ u \} = Y'.$$

Case 2.2: $u\,\alpha = v'$ and u' is *not* a factor of u: The word u' is a factor of $v' = u\,\alpha$, but not a factor of u, therefore there is a suffix u_α of u such that $u' = u_\alpha\,\alpha$. Because $| u_\alpha | + 1 = | u_\alpha\,\alpha | = | u' | < | v' | = | u\,\alpha | = | u | + 1$, one has $u_\alpha \neq u$. Now u_β is defined for every $\beta \in A$, including for $\beta = \alpha$. Let $\overline{\alpha}$ be such that $|u_{\overline{\alpha}}| = \max_{\beta \in A} |u_\beta|$.

There are now two subsubcases:

Case 2.2.1: $u_{\overline{\alpha}} = u$:

$$Z = X - \{u\alpha,\ u\overline{\alpha}\} + \{u\} = Y - \{u\overline{\alpha}\} = Y' - \{u\overline{\alpha}\} + \{u\}.$$

Case 2.2.2: $u_{\overline{\alpha}} \neq u$:

$$Z = X - \{u\alpha,\ u_{\overline{\alpha}}\overline{\alpha}\} + \{u_{\overline{\alpha}}\} = (Y - \{u_{\overline{\alpha}}\overline{\alpha}\} + u_{\overline{\alpha}}) - \{u\} = Y' - \{u_{\overline{\alpha}}\overline{\alpha}\} + u_{\overline{\alpha}}.$$

Case 2.3: $u\alpha = u'$:

$$Z = X - \{u\alpha,\ v'\} + \{u\} = Y - \{v'\} = Y' - \{u\alpha\} + \{u\}.$$

Case 2.4: There is a letter $\beta \neq \alpha$ such that $u_\beta\beta = v'$. There are two subcases:

Case 2.4.1: u' is a suffix of v' and $u' \neq \varepsilon$:

$$Z = X - \{u\alpha,\ v'\} + \{u\} = Y - \{v'\} = Y' - \{u\alpha\} + \{u\}.$$

Case 2.4.2: u' is not a suffix of v' (or $u' = \varepsilon$):

$$Z = X - \{u\alpha,\ v'\} = Y - \{u,\ v'\} = Y' - \{u\alpha\}.$$

Case 2.5: $v' \neq u\alpha$, $u' \neq u\alpha$ and for every letter $\beta \neq \alpha$, $u_\beta\beta \neq v'$:

$$Z = X - \{u\alpha,\ v'\} + \{u\} = Y - \{v'\} = Y' - \{u\alpha\} + \{u\}.$$

Case 3

There are α in A and u in A^* such that $u\alpha \in X$, such that for every letter $\beta \neq \alpha$, there is a suffix u_β of u such that $u_\beta\beta \in X$ and such that $Y = X - \{u\alpha\} + \{u\}$, and there are α' in A and u' in A^* such that $u'\alpha' \in X$, such that for every letter $\beta \neq \alpha'$, there is a suffix u'_β of u' such that $u'_\beta\beta \in X$ and such that $Y' = X - \{u'\alpha'\} + \{u'\}$, and one can assume that $|u| \geq |u'|$.

The situation when $u \in X$ or $u' \in X$ or both, has already been studied in case 1 or 2. Therefore one can assume from now on that $u \notin X$ and $u' \notin X$. This case subdivides into four cases:

Case 3.1: $u\alpha = u'\alpha'$: $Z = Y = Y'$.

Case 3.2: $u = u'$ but $u\alpha \neq u'\alpha'$:

$$Z = X - \{u\alpha,\ u'\alpha'\} + \{u\} = Y - \{u'\alpha'\} = Y' - \{u\alpha\}.$$

Case 3.3: u' is a suffix of u, but $u' \neq u$ (so that $|u'| < |u|$) :

$$Z = X - \{u\alpha,\ u'\alpha'\} + \{u'\} = Y - \{u,\ u'\alpha'\} + \{u'\} = Y' - \{u\alpha\}.$$

Case 3.4: u' is not a suffix of u :

$$Z = X - \{u\alpha, u'\alpha'\} + \{u, u'\} = Y - \{u'\alpha'\} + \{u'\} = Y' - \{u\alpha\} + \{u\}.$$

So, if X cuts into Y and into Y' on the right in an elementary way, then one can find in every case a language Z such that both Y and Y' cut into Z on the right. By symmetry, if X cuts into Y and into Y' on the left in an elementary way, then one can find a language Z such that both Y and Y' cut into Z on the left. If X cuts into Y and into Y' in an elementary way, then

either X cuts into Y and into Y' both on the right, or both on the left in an elementary way, but then we already know that there is a language Z such that Y and Y' cut into Z both on the right, or both on the left, so that Y and Y' cut into Z,

or (should one interchange Y and Y'), X cuts into Y on the right, but not on the left, and X cuts into Y' on the left, but not on the right. Therefore one is in the

Case 4

There are α in A and u in A^*, such that $u\,\alpha \in X$ and for every letter $\beta \neq \alpha$, there is a suffix u_β of u such that $u_\beta\,\beta \in X$, and $Y = X - \{u\,\alpha\} + \{u\}$, and there are $\alpha' \in A$, $u' \in A^*$, such that $\alpha'\,u' \in X$ and for every letter $\beta \neq \alpha'$, there is a prefix u'_β of u such that $\beta\,u'_\beta \in X$, and $Y = X - \{\alpha'\,u'\} + \{u'\}$.

If $u \in X$, then u, $u\,\alpha \in X$, u is a factor of $u\,\alpha$, $Y = X - \{u\,\alpha\}$, so X cuts into Y on both sides, which contradicts "X cuts into Y on the right, but not on the left", so $u \notin X$. For similar reasons, $u' \notin X$. This case subdivides in six cases:

Case 4.1: $u = u'$ and $u\,\alpha = \alpha'\,u'$: $Z = Y = Y'$.

Case 4.2: $u = u'$ and $u\,\alpha \neq \alpha'\,u'$:

$$Z = X - \{u\,\alpha,\, \alpha'\,u'\} + \{u\} = Y - \{\alpha'\,u'\} = Y' - \{u\,\alpha\}.$$

Case 4.3: $\varepsilon = u' \neq u$:

$$Z = X - \{u\,\alpha,\, \alpha'\} + \{\varepsilon\} = Y - \{u,\, \alpha'\} + \{\varepsilon\} = Y' - \{u\,\alpha\}.$$

Case 4.4: $\varepsilon = u \neq u'$: This case is the symmetric of the previous one.

Case 4.5: $u \neq \varepsilon$, $u' \neq \varepsilon$, $u \neq u'$, $u\,\alpha \neq \alpha'\,u'$:

$$\begin{aligned} Z &= X - \{\alpha'\,u',\, u\,\alpha\} + \{u',\, u\} \\ &= Y - \{\alpha'\,u'\} + \{u'\} = Y' - \{u\,\alpha\}\{u\}. \end{aligned}$$

Case 4.6: $u \neq \varepsilon$, $u' \neq \varepsilon$, $u \neq u'$, $u\,\alpha = \alpha'\,u'$:

From $u \neq \varepsilon$, $u' \neq \varepsilon$ and $u\,\alpha = \alpha'\,u'$, one can deduce that there is a word v such that $u' = v\,\alpha$ and $u = \alpha'\,v$, so that $u\,\alpha = \alpha'\,u' = \alpha'\,v\,\alpha$. For every letter $\beta \neq \alpha$, u_β is a suffix of $u = \alpha\,v'$, so either $u_\beta = \alpha'\,v$, or u_β is a suffix of v', and for every letter $\beta \neq \alpha'$, u'_β is a prefix of $u' = v\,\alpha$, so either $u'_\beta = v\,\alpha$, or u'_β is a prefix of v. There are now four subsubcases.

Case 4.6.1: For every letter $\beta \neq \alpha$, u_β is a suffix of v and for every letter $\beta \neq \alpha'$, u'_β is a prefix of v :

$$\begin{aligned} Z &= X - \{\,\alpha'\,v\,\alpha\,\} + \{\,v\,\} \\ &= Y - \{\,\alpha'\,v\,\} + \{\,v\,\} = Y' - \{\,v\,\alpha\,\} + \{\,v\,\}. \end{aligned}$$

Case 4.6.2: There is $\gamma \neq \alpha$ such that $u_\gamma = \alpha'\,v\,(= u)$ and for every letter $\beta \neq \alpha'$, u'_β is a prefix of v :

$$\begin{aligned} Z &= X - \{\,\alpha'\,v\,\alpha,\ \alpha'\,v\,\gamma\,\} + \{\,v\,\} \\ &= Y - \{\,\alpha'\,v\,\} + \{\,v\,\}) - \{\,\alpha'\,v\,\gamma\,\} \\ &= ((Y' - \{\,\alpha'\,v\,\gamma\,\} + \{\,\alpha'\,v\,\}) - \{\,\alpha'\,v\,\} + \{\,v\,\}) - \{\,v\,\alpha\,\}. \end{aligned}$$

Case 4.6.3: There $\gamma \neq \alpha'$, such that $u'_\gamma = v\,\alpha\,(= u')$ and for every letter $\beta \neq \alpha$, u_β is a suffix of v'. This case is just the symmetric of the previous one.

Case 4.6.4: There is $\gamma \neq \alpha$ such that $u_\gamma = u\,(= \alpha'\,v)$ and there is $\gamma' \neq \alpha'$ such that $u'_{\gamma'} = u'\,(= v\,\alpha)$.

$$\begin{aligned} Z &= X - \{\,\gamma'\,v\,\alpha,\ \alpha'\,v\,\alpha,\ \alpha'\,v\,\gamma\,\} + \{\,\alpha'\,v,\ v\,\alpha\,\} \\ &= (Y - \{\,\gamma'\,v\,\alpha\,\} + \{\,v\,\alpha\,\}) - \{\,\alpha'\,v\,\gamma\,\} \\ &= (Y' - \{\,\alpha'\,v\,\gamma\,\} + \{\,\alpha'\,v\,\}) - \{\,\gamma'\,v\,\alpha\,\}. \end{aligned}$$

So, in every case again, one can find a language Z such that both Y and Y' cut into Z, so **lemma 7.2 is proved.** □

LEMMA 7.3: *Let X be a finite language. If Y and Y' are two languages such that X cuts into Y and into Y' (resp. on the right, resp. on the left), then there is a language Z such that both Y and Y' cut (resp. on the right, resp. on the left) into Z.*

Proof of lemma 7.3: Lemma 7.3 is proved by induction on $\sum_{x \in X} |\,x\,|$:

◇ If $\sum_{x \in X} |\,x\,| = 0$, then $X = \emptyset$ or $X = \{\,\varepsilon\,\}$ and the result is obvious.

◇ Assume the result is true for every language W with $\sum_{w \in W} |\,w\,| \leq L$, assume that X, Y and Y' are some languages such that $\sum_{x \in X} |\,x\,| = L + 1$, and such that X cuts into Y and into Y' (resp. on the left, resp. on the right). We prove the result is true for X.

If $Y = X$, the result is obvious with $Z = Y'$.

If $Y' = X$, the result is obvious with $Z = Y$.

If $Y \neq X$ and $Y' \neq X$, then there are some languages Y_0, Y_0' such that X cuts into Y_0 and into Y_0' (resp. on the left, resp. on the right) in an elementary way, such that Y_0 cuts into Y (resp. on the left, resp. on the right) and such that Y_0' cuts into Y' (resp. on the left, resp. on the right). Then, thanks to lemma 7.2, there is a language $\overline{Z}$ such that both Y_0 and Y_0' cut into $\overline{Z}$ (resp. on the left, resp. on the right).

Now, $\sum_{y \in Y_0} |y| < \sum_{x \in X} |x| \leq L + 1$, and Y_0 cuts both into Y and into $\overline{Z}$ (resp. on the left, resp. on the right), so thanks to the induction hypothesis, there is a language Z_0 such that both Y and $\overline{Z}$ cut into Z_0 (resp. on the left, resp. on the right). For symmetric reasons, there is a language Z_0' such that both Y' and $\overline{Z}$ cut into Z_0' (resp. on the left, resp. on the right). But $\sum_{z \in \overline{Z}} |z| < \sum_{x \in X} |x| \leq L + 1$, and $\overline{Z}$ cuts into Z_0 and into Z_0', so thanks to the induction hypothesis again, there is a language Z such that both Z_0 and Z_0' cut into Z (resp. on the left, resp. on the right). Now Y cuts into Z_0 which cuts into Z, so Y cuts into Z, and Y' cuts into Z_0' which cuts into Z, so Y' cuts into Z too. The language Z satisfies the required property, so the induction hypothesis is true for $L + 1$.

By induction, the result is true for every finite language X, whatever $\sum_{x \in X} |x|$ is.

Lemma 7.3 is proved. $\square$

$$
\begin{tikzcd}
 & & X \arrow[dl] \arrow[dr] & & \\
 & Y_0 \arrow[dl, "*"'] \arrow[dr, "*"] & & Y_0' \arrow[dl, "*"'] \arrow[dr, "*"] & \\
Y \arrow[dr, "*"] & & \overline{Z} \arrow[dl, "*"'] \arrow[dr, "*"] & & Y' \arrow[dl, "*"'] \\
 & Z_0 \arrow[dr, "*"] & & Z_0' \arrow[dl, "*"'] & \\
 & & Z & &
\end{tikzcd}
$$

By an obvious induction on $\sum_{x \in X} |x|$, there is an innocent languages $\overline{X}$ such that X cuts into $\overline{X}$. Assume $\overline{X'}$ is another one. Then X cuts into $\overline{X}$ and into $\overline{X'}$, so by lemma 7.2, there is a language $\overline{X''}$ such that both $\overline{X}$ and $\overline{X'}$ cut into $\overline{X''}$, but $\overline{X}$ and $\overline{X'}$ are innocent, therefore $\overline{X} = \overline{X''} = \overline{X'}$, and therefore there is a *unique* innocent language $\overline{X}$, such that X cuts into $\overline{X}$.

The same proof can be made with $\overline{X}^r$ and with $\overline{X}^l$.

Proposition 7.1 is proved. □

PROPOSITION 7.4: *Let X be a finite language, then $\overline{X} = \overline{\overline{X}^r}^{\,l} = \overline{\overline{X}^l}^{\,r}$.*

Proof of proposition 7.4:

LEMMA 7.5: *Let X be a finite language which is innocent on the left, guilty on the right, then there is an innocent on the left language Y such that $Y \neq X$ and such that X cuts into Y on the right.*

Proof of proposition 7.5: X is guilty on the right. The existence of words $u, v \in X$ such that $u \neq v$ and u is a factor of v would contradict the innocence of X on the left, so there is a letter $\alpha \in A$ and a word $u \in A^*$ such that $u\alpha$ is in X and such that for every letter $\beta \in A$, $\beta \neq \alpha$, there is a suffix u_β of u such that $u_\beta \beta \in X$.

Let $X' = X - \{u\alpha\} + \{u\}$, then X cuts into X' on the right. Let $X_u = \{v \in X' \mid v \neq u$ and u is a factor of $v\}$ and let $Y = X' - X_u$, then X' cuts into Y on the right (by a sequence of deletion using the cut (1)), so X cuts into Y on the right, and it is clear that $X \neq Y$.

Note that if a word z is in Y and is different from u, then it is in X.

We have to show now that Y is innocent on the left. Assume Y is guilty on the left, then:

◇ Either there are distinct words u', $v' \in Y$ such that u' is factor of v'.

If $u' \neq u$, $v' \neq u$, then u, $u' \in X$ and this contradicts the innocent on the left of X.

If $v' = u$, then $u' \neq u$ so $u' \in X$, and u' is a factor of $u\alpha$ (because u' is a factor of $u = v'$ which is a factor of $u\alpha$) and $u\alpha \in X$, this contradicts the innocence of X on the left.

If $u' = u$, then $v' \in Y$ is a contradiction because u is a factor of v' and $Y = X' - X_u = X' - \{v \in X' \mid v \neq u,\ u$ is factor of $v\}$.

◇ Or there is a letter $\overline{\alpha} \in A$ and a word $\overline{u} \in A^*$ such that $\overline{\alpha}\,\overline{u} \in Y$ and for every letter $\overline{\beta} \neq \overline{\alpha}$, there is a prefix $\overline{u}_{\overline{\beta}}$ of $\overline{u}$ such that $\overline{\beta}\,\overline{u}_{\overline{\beta}} \in Y$.

$\diamond\diamond$ If $\overline{\alpha}\,\overline{u} \neq u$, and for every letter $\overline{\beta} \neq \overline{\alpha}$, $\overline{\beta}\,\overline{u}_{\overline{\beta}} \neq u$, then $\overline{\alpha}\,\overline{u} \in X$ and for every letter $\overline{\beta} \neq \overline{\alpha}$, $\overline{\beta}\,\overline{u}_{\overline{\beta}} \in X$. This contradicts the innocence of X on the left.

$\diamond\diamond$ If $\overline{\alpha}\,\overline{u} = u$, then let $\hat{u} = \overline{u}\,\alpha$. Then $\overline{\alpha}\,\hat{u} = \overline{\alpha}\,\overline{u}\,\alpha = u\,\alpha \in X$, for every letter $\overline{\beta} \neq \overline{\alpha}$, $\overline{u}_{\overline{\beta}}$ is a prefix of $\hat{u}$ (because $\overline{u}_{\overline{\beta}}$ is a prefix of $\overline{u}$ which is a prefix of $\hat{u}$) and $\overline{\beta}\,\overline{u}_{\overline{\beta}} \in Y$, and since $\overline{\beta}\,\overline{u}_{\overline{\beta}} \neq u\,(= \overline{\alpha}\,\overline{u})$, one has $\overline{\beta}\,\overline{u}_{\overline{\beta}} \in X$. This contradicts the innocence of X on the left.

$\diamond\diamond$ If there is $\overline{\gamma}$ such that $\overline{\gamma}\,\overline{u}_{\overline{\gamma}} = u$, then:

If $\overline{u}_{\overline{\gamma}} = \overline{u}$, then let $\overline{u}_{\overline{\alpha}} = \overline{u}$, then $\overline{\gamma}\,\overline{u}\,\alpha = \overline{\gamma}\,\overline{u}_{\overline{\gamma}}\,\alpha = u\,\alpha \in X$, and for every $\overline{\beta} \neq \overline{\gamma}$, $\overline{u}_{\overline{\beta}}$ is a prefix of $\overline{u}$ (It is true if $\overline{\beta} \neq \overline{\alpha}$ and it is true if $\overline{\beta} = \overline{\alpha}$). Since $\overline{\gamma}\,\overline{u}\,\alpha$ and for every letter $\overline{\beta} \neq \overline{\gamma}\;\overline{\beta}\,\overline{u}_{\overline{\beta}}$, belongs to X (because $\overline{\beta}\,\overline{u}_{\overline{\beta}} \in Y$, $\overline{\beta}\,\overline{u}_{\overline{\beta}} \neq \overline{\gamma}\,\overline{u}_{\overline{\gamma}} = u$ so $\overline{\beta}\,\overline{u}_{\overline{\beta}} \in X$), the innocence of X on the left is contradicted.

So one can assume that $\overline{u}_{\overline{\gamma}} \neq \overline{u}$. Since $\overline{u}_{\overline{\gamma}}$ is a prefix of $\overline{u}$, there is a letter γ and a word w such that $\overline{u} = \overline{u}_{\overline{\gamma}}\,\gamma\,w$.

$\diamond\ \diamond\ \diamond$ If $\gamma = \alpha$, then let $\hat{u}_{\overline{\gamma}} = \overline{u}_{\overline{\gamma}}\,\alpha = \overline{u}_{\overline{\gamma}}\,\gamma$, then $\overline{\alpha}\,\overline{u} \in X$ (because $\overline{\alpha}\,\overline{u} \in Y$ and $\overline{\alpha}\,\overline{u} \neq \overline{\gamma}\,\overline{u}_{\overline{\gamma}} = u$), for every $\overline{\beta} \notin \{\overline{\alpha}, \overline{\gamma}\}$, $\overline{u}_{\overline{\beta}}$ is a prefix of $\overline{u}$ and $\overline{\beta}\,\overline{u}_{\overline{\beta}} \in X$ (because $\overline{\beta}\,\overline{u}_{\overline{\beta}} \in Y$ and $\overline{\beta}\,\overline{u}_{\overline{\beta}} \neq \overline{\gamma}\,\overline{u}_{\overline{\gamma}} = u$) and $\hat{u}_{\overline{\gamma}} = \overline{u}_{\overline{\gamma}}\,\gamma$ is a prefix of $\overline{u} = \overline{u}_{\overline{\gamma}}\,\gamma\,w$, and $\overline{\gamma}\,\hat{u}_{\overline{\gamma}} = \overline{\gamma}\,\overline{u}_{\overline{\gamma}}\,\gamma = u\,\gamma = u\,\alpha \in X$. This contradicts the innocence of X on the left.

$\diamond\ \diamond\ \diamond$ If $\gamma \neq \alpha$, then recall that there is u_γ suffix of u such that $u_\gamma\,\gamma \in X$.

$\diamond\ \diamond\ \diamond\diamond$ If $u_\gamma \neq u$, then (because u_γ is a suffix of $u = \overline{\gamma}\,\overline{u}_{\overline{\gamma}}$), u_γ is a suffix of $\overline{u}_{\overline{\gamma}}$, therefore $u_\gamma\,\gamma$ is a factor of $\overline{u}_{\overline{\gamma}}\,\gamma$ and therefore of $\overline{\alpha}\,\overline{u}_{\overline{\gamma}}\,\gamma\,w = \overline{\alpha}\,\overline{u}$. But $u_\gamma\,\gamma \in X$, $\overline{\alpha}\,\overline{u} \in X$ (because $\overline{\alpha}\,\overline{u} \in Y$, $\overline{\alpha}\,\overline{u} \neq \overline{\gamma}\,\overline{u}_{\overline{\gamma}}\,\alpha = u$), $u_\gamma\,\gamma \neq \overline{\alpha}\,\overline{u}$ (because of their length) and this contradicts the innocence of X on the left.

$\diamond\ \diamond\ \diamond\ \diamond$ If $u_\gamma = u$, then let $\hat{u}_{\overline{\gamma}} = \overline{u}_{\overline{\gamma}}\,\gamma$, then $\overline{\alpha}\,\overline{u} \in X$ (because $\overline{\alpha}\,\overline{u} \neq \overline{\gamma}\,\overline{u}_{\overline{\gamma}} = u$), for every $\overline{\beta} \notin \{\overline{\alpha}, \overline{\gamma}\}$, $\overline{u}_{\overline{\beta}}$ is a prefix of $\overline{u}$ and $\overline{\beta}\,\overline{u}_{\overline{\beta}} \in X$ (because $\overline{\beta}\,\overline{u}_{\overline{\beta}} \in Y$ and $\overline{\beta}\,\overline{u}_{\overline{\beta}} \neq \overline{\gamma}\,\overline{u}_{\overline{\gamma}} = u$) and $\hat{u}_{\overline{\gamma}} = \overline{u}_{\overline{\gamma}}\,\gamma$ is a prefix of $\overline{u} = \overline{u}_{\overline{\gamma}}\,\gamma\,w$, and $\overline{\gamma}\,\hat{u}_{\overline{\gamma}} = \overline{\gamma}\,\overline{u}_{\overline{\gamma}}\,\gamma = u\,\gamma = u_\gamma\,\gamma \in X$. This contradicts the innocence of X on the left.

Lemma 7.5 is proved. $\square$

Now, we can prove proposition 7.4 by induction on $\sum_{x \in X} |x|$.

If $\sum_{x \in X} |x| = 0$, then $X = \emptyset$ or $X = \{\varepsilon\}$ and the result is obvious.

Let us assume that $\overline{\overline{Y}^l}^r = Y$ for every language Y such that $\sum_{y \in Y} |y| \leq L$.

Let X be a finite language satisfying $\sum_{x \in X} |x| = L + 1$, then:

If X is guilty on the left, let $Y = \overline{X}^l$, then $\sum_{y \in Y} |y| \leq L$ (because X is guilty on the left, so $\sum_{x \in \overline{X}^l} |x| < \sum_{x \in X} |x| = L + 1$), and therefore, by induction hypothesis $\overline{\overline{Y}^l}^r = \overline{Y}$. Since Y is innocent on the left, $\overline{Y}^l = Y$. As a conclusion, $\overline{\overline{X}^l}^r = \overline{Y}^r = \overline{\overline{Y}^l}^r = \overline{Y} = \overline{\overline{X}^l} = \overline{X}$.

If X is innocent on the left and on the right, then $\overline{\overline{X}^l}^r = X = \overline{X}$.

If X is innocent on the left and guilty on the right, then thanks to lemma 7.5, there is an innocent on the left language Y which is distinct from X, such that X cuts into Y on the right. Since $\sum_{y \in Y} |y| \leq L$, one has $\overline{\overline{Y}^l}^r = \overline{Y}$. The language X is innocent on the left so $X = \overline{X}^l$ and therefore $\overline{\overline{X}^l}^r = \overline{X}^r$. The language X cuts into Y on the right, therefore $\overline{X}^r = \overline{Y}^r$. The language Y is innocent on the left, so $Y = \overline{Y}^l$ and therefore $\overline{Y}^r = \overline{\overline{Y}^l}^r$. Recall that $\overline{\overline{Y}^l}^r = \overline{Y}$. The language X cuts into Y, therefore $\overline{Y} = \overline{X}$. Therefore $\overline{\overline{X}^l}^r = \overline{X}^r = \overline{Y}^r = \overline{\overline{Y}^l}^r = \overline{Y} = \overline{X}$.

By symmetry, $\overline{\overline{X}^l} = \overline{X}$.

Proposition 7.4 is proved. $\square$

8. CUTS AND INFINITE LANGUAGES

The aim of this section is to study cuts on infinite languages, and to try to give a definition of "X eventually cuts into Y", and of "acquitted X" for infinite languages, which generalizes in a satisfactory way the results we have for finite languages.

To begin with, recall that the definitions of elementary cuts, of cuts, of guilty and of innocent languages (*see* section 5) are valid for infinite languages, and that the following propositions are valid for infinite languages:

PROPOSITION 5.6: *Let X and Y be some languages such that X cuts into Y, then a bi-infinite word $\aleph$ avoids X iff it avoids Y.*

PROPOSITION 6.5: *Let X be an innocent language and S be the set of the words which avoid X, then $X = \hat{S}$:*

Recall (*see* definition 6.4) that if S is a set of bi-infinite words, then $\tilde{S}$ is the set of the words that never appear as a factor of a word in S, and that $\hat{S}$ is the set of the minimal elements in $\tilde{S}$ for the order "is a factor of".

We can state now the following proposition which is the beginning of an attempt of a generalization of proposition 6.6:

PROPOSITION 8.1: *Let X be a language, then there is a unique language, denoted by $\overline{X}$ and called acquitted X, which is innocent and is such that a bi-infinite word $\aleph$ avoids X iff it avoids $\overline{X}$.*

Proof of proposition 8.1: Let S be the set of bi-infinite words avoiding X, then:

– The language $\hat{S}$ is innocent, indeed:

The assumption that there are $s, s' \in \hat{S}$, $s \neq s'$ such that s is a factor of s', is absurd because of the definition of $\hat{S}$.

Assume that there is $sa \in \hat{S}$ (s is a word, a is a letter) such that for every letter $b \neq a$, there is a suffix s_b of s such that $s_b b \in \hat{S}$, then s cannot be a factor of a bi-infinite word in S (because if it is, this factor has to be followed by a letter α, so that sa, if $\alpha = a$, or $s_\alpha \alpha$ if $\alpha \neq a$, is a factor of that bi-infinite word in S. This contradicts the fact that $sa \in \tilde{S}$ or that $s_b b \in \tilde{S}$ for every $b \neq a$, and the definition of $\tilde{S}$). Therefore $s \in \tilde{S}$, and there is a factor s' of s which is in $\hat{S}$, then $s', sa \in \hat{S}$, $s' \neq sa$ and s' is a factor of sa. This contradicts the definition of $\hat{S}$.

– Let $\aleph$ be a bi-infinite word, then $\aleph$ avoids $\hat{S}$ iff $\aleph \in S$, that is iff $\aleph$ avoids X, indeed:

If $\aleph \in S$, then according to the definition of $\tilde{S}$, no word of $\tilde{S}$ and therefore no word of $\hat{S}$ is a factor of $\aleph$, so the word $\aleph$ avoids $\hat{S}$.

If $\aleph \notin S$, then $\aleph$ does not avoid X, so there is $x \in X$ which is a factor of $\aleph$. Because S is the set of bi-infinite words avoiding X, and $\tilde{S}$ the set of the words which are avoided by S, one has $X \subset \tilde{S}$, and therefore there is $x' \in \hat{S}$ such that x' is a factor of x. The word x' is a factor of $\aleph$ (since x is a factor of $\aleph$), so the bi-infinite word $\aleph$ does not avoid $\hat{S}$.

So $\hat{S}$ is innocent and a bi-infinite word $\aleph$ avoids X iff it avoids $\hat{S}$. We have to prove now that $\hat{S}$ is the only language satisfying these properties.

If $\hat{X}$ is an innocent language such that for every bi-infinite word $\aleph$, $\aleph$ avoids X iff it avoids $\hat{X}$, then S is also the set of bi-infinite words avoiding $\hat{X}$, so thanks to proposition 6.5, $\hat{X} = \hat{S}$, so $\hat{S}$ is the only language satisfying the required properties, so by defining $\overline{X} = \hat{S}$.

Proposition 8.1 is proved. □

We would like now to say that X cuts into $\overline{X}$ in some sense. Regular cuts will not work because one might need infinitely many cuts to go from X to $\overline{X}$, so there is a need to define eventual cuts.

DEFINITIONS 8.2: Let X be a language, a *sequence of cuts from* X is a sequence of languages $\mathcal{S} = (X_n)_{n \in \mathbb{N}}$ such that $X = X_0$, and such that for every $n \in \mathbb{N}$, X_n cuts into X_{n+1}.

Let X be a language and $\mathcal{S} = (X_i)_{i \in \mathbb{N}}$ be a sequence of cuts from X, then $X_{\mathcal{S},\sup}$ and $X_{\mathcal{S},\inf}$ are defined as follows:

$$X_{\mathcal{S},\sup} = \{ y \in A^* \mid \forall N \in \mathbb{N}, \exists n \geq N, y \in X_n \}$$

and

$$X_{\mathcal{S},\inf} = \{ y \in A^* \mid \exists N \in \mathbb{N}, \forall n \geq N, y \in X_n \}$$

Let X and Y be two languages, then, X *eventually cuts into* Y iff there is a sequence $\mathcal{S} = (X_i)_{i \in \mathbb{N}}$ of cuts from X, such that $X_{\mathcal{S},\sup} = X_{\mathcal{S},\inf} = Y$.

PROPOSITION 8.3: *Let X be a language, $\mathcal{S} = (X_i)_{i \in \mathbb{N}}$ be a sequence of cuts from X, and $\aleph$ be a bi-infinite word, then the following three propositions are equivalent:*

(1) $\aleph$ *avoids* X.

(2) $\aleph$ *avoids* $X_{\mathcal{S},\sup}$.

(3) $\aleph$ *avoids* $X_{\mathcal{S},\inf}$.

Proof of proposition 8.3: (2) $\Rightarrow$ (3): one has $X_{\mathcal{S},\inf} \subset X_{\mathcal{S},\sup}$, so if $\aleph$ avoids $X_{\mathcal{S},\sup}$, it avoids $X_{\mathcal{S},\inf}$.

(1) $\Rightarrow$ (2): If $\aleph$ avoids X, then let $x \in X_{\mathcal{S},\sup}$. Because of the definition of $X_{\mathcal{S},\sup}$, there is an integer n such that $x \in X_n$. The language X cuts into X_n and $\aleph$ avoids X, so thanks to proposition 5.6, $\aleph$ avoids X_n, so that x is not a factor of $\aleph$. Since this is true for every $x \in X_{\mathcal{S},\sup}$, the word $\aleph$ avoids $X_{\mathcal{S},\sup}$.

(3) $\Rightarrow$ (1): If $\aleph$ does not avoid X, then there is $x \in X$ which is a factor of $\aleph$. There is a sequence of words $(x_n)_{n \in \mathbb{N}}$ such that $x_0 = x$ and such that for every $n \in \mathbb{N}$, one has $x_n \in X_n$ and x_{n+1} is a factor of x_n (This sequence can be built by induction, using the fact that if $y \in Y$ and Y cuts

into Z, then there is $z \in Z$ which is a factor of y, this statement itself being proved by induction on the number of elementary cuts necessary to go from Y to Z). Now by an argument on the length of the x_n's (and using that each x_{n+1} is a factor of x_n), one can see that there is an integer N such that for every $n \geq N$, one has $x_n = x_N$, so that $x_N \in X_n$ for all $n \geq N$, therefore $x_N \in X_{\mathcal{S},\inf}$, and since x_N is a factor of x which is a factor of $\aleph$, one has that x_N is a factor of $\aleph$. Therefore, $\aleph$ does not avoid $X_{\mathcal{S},\inf}$.

Proposition 8.3 is proved. $\square$

PROPOSITION 8.4: *Let X be a language, then X eventually cuts into $\overline{X}$.*

Proof of proposition 8.4: Let us define $(X_n)_{n \in \mathbb{N}}$ by induction:

– Let $X_0 = X$,

– Assume X_n is built. If X_n is innocent, then let $X_{n+1} = X_n$. If X_n is guilty, then X_{n+1} is the language obtained by cutting one of the shortest word which can be possibly cut, and by removing a word rather than shortening one whenever there is a choice. More precisely:

Let us define the word cut by the elementary cut $Y \overset{\text{cut}}{\to} Z$ by u if $Z = Y - \{u\}$, by ua if $Z = Y - \{ua\} + \{u\}$ and by au if $Z = Y - \{au\} + \{u\}$. Let l be the minimum length of the words in X_n which can be cut. If there is a word u satisfying $|u| = l$ and such that X_n cuts into $X_n - \{u\}$ in elementary way, then let $X_{n+1} = X_n - \{u\}$. Otherwise, there are words u, v and a letter a, satisfying $|v| = l$, such that $v = au$ or $v = ua$ and such that X_n cuts into $X_n - \{v\} + \{u\}$. In that latter case, let $X_{n+1} = X_n - \{v\} + \{u\}$.

We prove now that $X_{\mathcal{S},\sup} = X_{\mathcal{S},\inf}$ and that it is innocent. If there is N such that X_N is innocent, then for every $n \geq N$, $X_n = X_N$, and therefore $X_{\mathcal{S},\sup} = X_{\mathcal{S},\inf} = X_N$ and is innocent. So we can assume from now on that X_n is guilty for every n, so that X_n cuts into X_{n+1} in an elementary way for every $n \in \mathbb{N}$. Let us prove now that for every $L \in \mathbb{N}$, there is a subset U_L of $A^{\leq L}$, and an integer n_L such that for every $n \geq n_L$, one has $X_n \cap A^{\leq L} = U_L$ (Note that this implies that $U_L = U_{L'} \cap A^{\leq L}$ if $L \leq L'$).

This is proved by induction on L:

$L = 0$: The only word of length 0 is ε. If ε is in no X_n, then the proposition is true for $L = 0$ with $n_0 = 0$ and $U_0 = \emptyset$. If $\varepsilon \in X_N$ for some integer N, then an induction on n shows that $\varepsilon \in X_n$ for every $n \geq N$ (you need to use that if Y cuts into Z in an elementary way and $\varepsilon \in Y$, then

$\varepsilon \in Z$, which is pretty clear from what the cuts are), so that the proposition is true with $n_0 = N$ and $U_0 = \{\varepsilon\}$.

Assume $L \geq 0$ and that there are a sublanguage U_L of $A^{\leq L}$ and an integer n_L such that for every $n \geq n_L$, one has $X_n \cap A^{\leq L} = U_L$. Let $Z_n = X_n \cap A^{L+1}$ for every n, then:

(a) For every $n \geq n_L$, the word cut by $X_n \overset{\text{cut}}{\to} X_{n+1}$ is of length at least $L+1$ (because if $X_n \overset{\text{cut}}{\to} X_{n+1}$ cuts a word of length less than L, then $X_n \cap A^{\leq L} \neq X_{n+1} \cap A^{\leq L}$, which contradicts the fact that there are both equal to U_L).

(b) There is $\overline{n} \geq n_L$ such that no word in U_L is a factor of $Z_{\overline{n}}$ (Indeed, if $u \in U_L$ is a factor of a word $z \in Z_n$ for $n \geq n_L$, then, since u, $z \in X_n$, the language X_n cuts into $X_n - \{z\}$. Since there is no elementary cut $X_n \overset{\text{cut}}{\to} Y$ cutting a word of length less than L (because of (a)), since $|z| = L+1$ (so that $X_n \overset{\text{cut}}{\to} X_n - \{z\}$ is a cut removing a word of length less than $L+1$) and because of the way $(X_n)_{n \in \mathbb{N}}$ has been defined (The word cut is among the shortest which could be cut, and a word is removed if possible), there is $z' \in Z_n$ such that $X_{n+1} = X_n - \{z'\}$, therefore $Z_{n+1} = Z_n - \{z'\}$. Therefore, if $n \geq n_L$ and if there is a word in U_L which is factor of a word in Z_n, then one has $\operatorname{card}(Z_{n+1}) = \operatorname{card}(Z_n) - 1$, but this cannot happen for every $n \geq n_L$ since the Z_n are finite, so that the $\operatorname{card}(Z_n)$'s are finite integers. Therefore, there is $\overline{n} \geq n_L$ such that no word in U_L is a factor of $Z_{\overline{n}}$).

(c) For every $n \geq \overline{n}$, no word in U_L is a factor of Z_n (By induction on n: this is true for $n = \overline{n}$ thanks to (b). Assume it is true for n, but not for $n+1$, so there is a word z which is in Z_{n+1}, but not in Z_n and a word u in U_L which is a factor of z. Since $z \in Z_{n+1} - Z_n$, one has $z \in X_{n+1} - X_n$, and since X_n cuts into X_{n+1} in an elementary way, this means that there is a word w and a letter a such that $w = za$ or $w = az$ and such that $X_{n+1} = X_n - \{w\} + \{z\}$. Now u and w are in X_n, and u is a factor of w (because u is a factor of z which is a factor of w), so X_n also cuts into $X_n - \{w\}$. The cut $X_n \overset{\text{cut}}{\to} X_n - \{w\} + \{z\}$ and the cut $X_n \overset{\text{cut}}{\to} X_n - \{w\}$ cut the same word, but the first one replaces it by z, while the second one removes it. Therefore, the fact that $X_{n+1} = X_n - \{w\} + \{z\}$ contradicts the way the $(X_n)_{n \in \mathbb{N}}$ have been built (It contradicts the fact that preference was given to removals, rather than to shortenings), so no element in U_L is a factor of a word in Z_n for $n \geq \overline{n}$).

(*d*) For $n \geq \overline{n}$, one has $Z_n \subset Z_{n+1}$ (Indeed, if $Z_n \not\subset Z_{n+1}$, then the cut $X_n \xrightarrow{\text{cut}} X_{n+1}$ cuts a word z in Z_n.

If $X_{n+1} = X_n - \{ z \}$, then there must be u in X_n which is a factor of z. Since $|u| < |z| = L + 1$, one has $u \in U_L$. This contradicts (*c*).

If $X_{n+1} = X_n - \{ z \} + \{ v \}$, for a word v such that there is a letter a with $z = av$ or $z = va$, then $|v| \leq n$ (because $|v| < |z| = L + 1$), $v \in X_{n+1}$, $v \notin X_n$ (otherwise, the cut would have been $X_n \xrightarrow{\text{cut}} X_{n+1} - \{ z \}$) and $n \geq n_L$, so $v \in X_{n+1} \cap A^{\leq L} = U_L$ and $v \notin X_n \cap A^{\leq L} = U_L$. There is a contradiction.

(*e*) Therefore Z_n is increasing, and since $\operatorname{card}(Z_n)$ is bounded (There is only a finite number of words of length $L+1$), there is a number $n_{L+1} \geq \overline{n}$, from which it is constant, so for every $n \geq n_{L+1}$, one has

$$\begin{aligned} X_n \cap A^{\leq L+1} &= X_n \cap (A^{\leq L} \cup A^{L+1}) \\ &= (X_n \cap A^{\leq L}) \cup (X_n \cap A^{L+1}) = U_L \cap Z_n = U_L \cap Z_{n_{L+1}}, \end{aligned}$$

so the proposition is true at rank $L + 1$ with $U_{L+1} \cup Z_{n_{L+1}}$.

Now $X_{\mathcal{S},\sup} = X_{\mathcal{S},\inf} = \overline{X}$. Indeed:

For every L, there is n_L such that for every $n \geq n_L$, one has $X_n \cap A^{\leq L} = U_L$, and therefore $X_{\mathcal{S},\sup} \cap A^{\leq L} = X_{\mathcal{S},\inf} \cap A^{\leq L} (= U_L)$, and since this is true for every L, one has $X_{\mathcal{S},\sup} = X_{\mathcal{S},\inf}$.

Now $X_{\mathcal{S},\sup} = X_{\mathcal{S},\inf}$ is innocent. Indeed, assume it is not, then there is a *finite* sublanguage Y of $X_{\mathcal{S},\sup}$ which is guilty, let $L = \max_{y \in Y} |y|$, then there is a language Z such that Y cuts into Z in an elementary way, and this cut cuts a word of length less than L. Let $W = X_{n_L} - Y$, then $X_{n_L} = Y \cup W$ cuts into $Z \cup W$, and this cut cuts the same word as $Y \xrightarrow{\text{cut}} Z$ does, that is, a word of length less than L. On the other hand, $X_{n_L} \xrightarrow{\text{cut}} X_{n_L+1}$ cuts a word of length at least $L + 1$ (because $X_{n_L} \cap A^{\leq L} = X_{n_L+1} \cap A^{\leq L} (= U_L)$), but this contradicts the way the $(X_n)_{n \in \mathbb{N}}$ have been built (It contradicts the fact that one always cuts one of the shortest word which can be cut).

So $X_{\mathcal{S},\sup} = X_{\mathcal{S},\inf}$ is innocent, and thanks to proposition 8.3, a bi-infinite word avoids X iff it avoids $X_{\mathcal{S},\sup}$, therefore, thanks to proposition 8.1, $X_{\mathcal{S},\sup} = X_{\mathcal{S},\inf} = \overline{X}$, and **proposition 8.4 is proved.** □

Remark 8.5: It is tempting to define simultaneous cuts, so that X simultaneously cuts into $\overline{X}$ in a finite number of steps. The first idea would be to say that X simultaneously cuts into Y in an elementary way if there is a set I of indices, some disjoint languages $(X_i)_{i \in I}$, and some disjoint languages $(Y_i)_{i \in I}$, such that $X = \bigcup_{i \in I} X_i$, such that $Y = \bigcup_{i \in I} Y_i$ and such that X_i cuts into Y_i in an elementary way for every $i \in I$, then to say that X simultaneously cuts into Y iff there a finite sequence of elementary simultaneous cuts leading from X to Y. Unfortunately, X might fail to simultaneously cut into $\overline{X}$ with that definition. Consider for example $X = \bigcup_{n \in \mathbb{N},\, n \geq 2} (\{ a^n b^n a^n b^{n+1} \} \cup (\bigcup_{0 \leq m \leq n} \{ a^n b^n a^n b^m a \}))$. One can see that for every $n \in \mathbb{N}$, $\{ a^n b^n a^n b^{n+1} \} \cup (\bigcup_{0 \leq m \leq n} \{ a^n b^n a^n b^m a \})$ cuts into $\{ a^n b^n a^n \}$, and that therefore X eventually cuts into $\bigcup_{n \in \mathbb{N},\, n \geq 2} \{ a^n b^n a^n \}$. Since the latter language is innocent, this is $\overline{X}$. Now the reader can check that no other cut that the ones inside $\{ a^n b^n a^n b^{n+1} \} \cup (\bigcup_{0 \leq m \leq n} \{ a^n b^n a^n b^m a \})$ are possible, so, to go from X to $\overline{X}$, one needs to cut each

$$\{ a^n b^n a^n b^{n+1} \} \cup (\bigcup_{0 \leq m \leq n} \{ a^n b^n a^n b^m a \})$$

into $\{ a^n b^n a^n \}$, this requires $n + 1$ steps for a given n, so the whole thing requires $\sup \{ n + 1 \mid n \in \mathbb{N},\, n \geq 2 \}$, that is infinitely many steps to go from X to $\overline{X}$.

To avoid such phenomena, one could thing to replace "X_i cuts into Y_i in an elementary way for every $i \in I$" in the above definition by "X_i cuts into Y_i for every $i \in I$". Then the X in the above example would simultaneously cut into $\overline{X}$. Unfortunately, there are still examples of languages Y which would not cut into $\overline{Y}$, such as $Y = \bigcup_{n \in \mathbb{N}} \{ ab^n ab^n \}$ (Note: $0 \in \mathbb{N}$). The reader can check that $\overline{Y} = \bigcup_{n \in \mathbb{N}} \{ ab^n a \}$, and that Y fails to cut simultaneously into $\overline{Y}$ with the latter definition.

9. OPEN PROBLEMS

Finiteness of the set of the enlargements of a language

See section 4.

Let Y be a finite language. Let us call *complements* of Y, the finite languages Z such that $Y \cup Z$ is minimal unavoidable. There are usually infinitely many different complements of a language Y, they might be infinitely many different innocent complements (For example, if $Y = \{ aa, bb \}$, the innocent complements are the $Z = \{ u \}$ with

$u \in (b+\varepsilon)(ab)^\star(a+\varepsilon)$, $|u| \geq 2$. They are infinite in number and they are all innocent.)

Open problem 9.1: Let Y be a finite language. Is it possible to give a finite number of complements of Y which would represent (in some sense to be defined) all of them?

I would look for a property such as: Let Y be a finite language, then there is a finite number of complements $(Z_i)_{i \in [1, n_Y]}$ of Y, such that if Z is a complement of Y, then there is an integer $i \in [1, n_Y]$ such that $Y \cup Z$ cuts into $\overline{Y} \cup Z_i$. (Consider $Y = \{aaa, aab, bbb, bba\}$ to see that one cannot replace $\overline{Y} \cup Z_i$ by $Y \cup Z_i$).

Enlargement into a non-extendible unavoidable language

An extendible unavoidable language is an unavoidable language X such that there is $x \in X$ and some letters $a_1, a_2, \ldots, a_n, \ldots$ such that $X - \{x\} + \{x a_1 a_2 \ldots a_n\}$ is unavoidable for every integer n. It was believed that every unavoidable language, but $\{\varepsilon\}$, was extendible. It is known now that it is not the case (*see* [17]).

Open problem 9.2: Let Y be a finite language. What is a necessary and sufficient condition on Y for there exists an enlargement of Y into a *non*-extendible unavoidable language?

A necessary condition is: For every $y \in Y$, there are at least two different periodic bi-infinite words $\aleph_y$ and $\aleph'_y$ such that y is a factor of both $\aleph_y$ and $\aleph'_y$, and such that y is the only element in Y to be a factor of $\aleph_y$ and the only one to be a factor of $\aleph'_y$.

This condition is not sufficient as one can see by considering the example $Y = \{ab\}$.

Weak equivalence

See section 6.

It is not clear what the weak equivalence is. I would like first to see some examples of weakly equivalent languages whose weak equivalence cannot be deduced from the strong equivalence and the equivalence between $a^n b$ and ba^n (For example, $\{aaab, aabaab\} \sim_w \{baaa, baabaa\}$ is not interesting because one can deduce it from strong equivalence and one-word equivalences, by

$$\{aaab, aabaab\} \sim_s \{aaab, baabaab\} \sim_w \{baaa, baabaab\}$$
$$\sim_s \{baaa, baabaa\} \sim_w \{aaab, baabaa\}).$$

Open problem 9.3: Is it possible to describe the weak equivalence in general, the weak equivalence for languages of bounded cardinality? Is there something finite in the set of the languages weakly equivalent to a given language X?

I am also surprised not to be able to find a short proof of the fact that $u \sim_w v$ iff $u = v$ or $\{u, v\} = \{a^n b, ba^n\}$ or $\{u, v\} = \{b^n a, ab^n\}$ (*see* theorem 6.16).

REFERENCES

1. A. V. AHO and M. J. CORASICK, Efficient string machines, an aid to bibliographic research, *Comm. ACM*, 1975, *18 (6)*, pp. 333-340.
2. D. R. BEAN, A. EHRENFEUCHT and G. F. MCNULTY, Avoidable patterns in strings of symbols, *Pacific J. Math.*, 1979, *85, 2*, pp. 261-294.
3. W. BUCHER, A. EHRENFEUCHT and D. HAUSSLER, On Total Regulators Generated by Derivation Relations, *Proc. of 12th International Colloquium on Automata, Languages and Programming; Lectures notes in Computer Science, 194, Springer-Verlag*, 1985, pp. 71-79.
4. C. CHOFFRUT and K. CULIK, On extendibility of unavoidable sets, *Discrete Applied Mathematics*, 1984, pp. 125-137.
5. M. CROCHEMORE, M. LEREST and P. WENDER, An optimal test on finite unavoidable sets of words, *Information Processing Letter*, 1983, *16*, pp. 179-180.
6. A. EHRENFEUCHT, D. HAUSSLER and G. ROZENBERG, On regularity of context-free languages, *Theor. Comp. Sci.*, 1983, *27*, pp. 311-332.
7. N. J. FINE and H. S. WILF, Uniqueness Theorem for Periodic functions, *Proc. Am. Math. Soc.*, 1965, *16*, pp. 109-114.
8. G. HIGMAN, Ordering by divisibility in abstract algebras, *Proc. London Math. Soc.*, 1952, *2*, pp. 326-336.
9. KRUSKAL, Well-quasi ordering, The Tree Theorem, and Vazsonyi's conjecture, *Trans. Amer. Math. Soc.*, 1960, *95*, pp. 210-225.
10. KRUSKAL, The theory of well-quasi ordering: a frequently discovered concept, *J. Comb. Theory Ser.*, 1972, *A 13 (3)*, pp. 297-305.
11. G. KUCHEROV and M. RUSINOVITCH, On Ground-Reducibility Problem for Word Rewriting Systems with Variables, In E. DEATON and R. WILKERSON, eds., *Proceedings 1994 ACM/SIGAPP Symposium on Applied Computing, Phoenix*, March 1994, ACM-Press.
12. G. KUCHEROV and M. RUSINOVITCH, Complexity of Testing Ground Reducibility for Linear Word Rewriting Systems with Variables, *preprint.*
13. LOTHAIRE, *Combinatorics on words*, chapter 1, Add.-Wesley, by D. PERRIN, 1983.
14. *Id*, chapter 6, by J. SAKAROVITCH.
15. E. OCHMANSKI, Inevitability in concurrent systems, *IPL 25*, 1987, pp. 221-225.
16. L. PUEL, Using unavoidable sets of trees to generalize Kruskal's theorem, *J.S.C.*, 1989, Vol. *8*, No. 4, pp. 335-382.

17. L. ROSAZ, Inventories of unavoidable languages and the word-extension conjecture, to appear in *Theoretical Computer Science*.
18. L. ROSAZ, Unavoidable languages, Mémoire de thèse, LITP Report.
19. M. P. SCHUTZENBERGER, On the synchronizing properties of certain prefix codes, *Information and Control*, 1964, *7*, pp. 23-36.