RAIRO. INFORMATIQUE THÉORIQUE ET APPLICATIONS

JEAN NERAUD

Elementariness of a finite set of words is co-NP-complete

RAIRO. Informatique théorique et applications, tome 24, n° 5 (1990), p. 459-470

<http://www.numdam.org/item?id=ITA_1990__24_5_459_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 24, n° 5, 1990, p. 459 à 470)

ELEMENTARINESS OF A FINITE SET OF WORDS IS co-NP-COMPLETE (*)

by Jean NERAUD (1)

Communicated by J.-E. PIN

Abstract. – *We define the rank of a finite set of words X as the integer: $r(X)=\min\{|Y| : X \subseteq Y^*\}$. X is said elementary iff $r(X)=|X|$ otherwise X is said simplifiable. We show that deciding whether X is simplifiable and deciding whether $r(X)$ is not greater than a given integer are NP-complete and we consider different related problems.*

Résumé. – *On définit le rang d'un ensemble fini de mots X comme l'entier $r(X)=\min\{|Y| : X \subseteq Y^*\}$. X est dit élémentaire ssi $r(X)=|X|$, sinon X est dit simplifiable. On montre que décider si X est simplifiable et décider si $r(X)$ est majoré par un entier donné sont des problèmes NP-complets, et on examine différents problèmes associés.*

1. INTRODUCTION

A finite set of words X in a free monoid is *simplifiable* if there exists another set of words Y of smaller cardinality such that every word of X can be factorized into words of Y, *i.e.* such that X is included in the submonoid Y^* generated by Y. Otherwise X is *elementary*. Thus $X_1=\{a, abc, bca\}$ is simplifiable (take $Y=\{a, bc\}$) and so is $X_2=\{aba, bba, abb, baa\}$ (take $Y=\{a, b\}$) but $X_3=\{a, abc, cba\}$ is elementary.

Here we prove the following:

THEOREM: *Deciding whether a given finite set of words is elementary is co-NP-complete.*

Historically the notion of elementariness applied to morphisms: a morphism of the free monoid into inself is *elementary* if the image of the alphabet is elementary (provided the images of two different letters are different). Its

(*) Received June 1988, revised in June 1989.
(1) L.I.R., Faculté des Sciences, Université de Rouen, 76130 Mont-Saint-Aignan, France.

Informatique théorique et Applications/Theoretical Informatics and Applications
0988-3754/90/05 459 12/$3.20/

introduction reduced the famous DOL-sequence equivalence problem to the case where the two morphisms were elementary, providing thus the major step toward the general solution (*cf.* [5] and [12]). In fact it captured the notion of bounded balance of two morphisms used by former approaches (*cf.* [4] and [14]), giving it a precise and understandable form. As discussed in [8] this is a beautiful illustration of how deep results usually entail as by-products rich new concepts. Ever since, this notion has proved particularly enlightning in numerous areas of combinatorics on words such as test sets, code theory, representation of formal languages etc. Let us just show why simplifiability is an important notion of the theory of equations in free monoids.

The theory is concerned with determining under which conditions certain words, say x, y, z satisfy a fixed non trivial equality such as $xy=yz$ for instance. The description of the solutions is impossible except in very special case (*cf. e.g.*, [9] Chap. IX) but it is possible to determine the *rank* of an equation i. e., the maximum number of "necessary" parameters for expressing the solutions. A solution is *cyclic* iff x, y, z are powers of a common word. In our example all non cyclic solutions can be expressed using only the two word parameters u and v: $x=(uv)^i$, $y=(uv)^j u$, $z=(vu)^i$. There are actually two different notions of rank. The first one is mentioned below and uses the concept of codes. The second, particularly popular in the soviet literature is related to the notion of simplifiability as it is discussed in the next paragraph. However, these different approaches lead to the same integer and Makanin's result on equations [10] yields an effective computation of the rank as was shown in [11].

We now briefly comment our result. The *deficit* of a set X of words is the integer:

$$d(X)=|X|-\min\{|Y|: X\subseteq Y^*\} \tag{1}$$

where $|X|$ denotes the cardinality of X. Thus $d(X_1)=1$, $d(X_2)=2$ and $d(X_3)=0$ for the above defined subsets. Our theorem simply says that computing $d(X)$ is co-NP-complete. The deficiency is very close to the notion of defect introduced in the theory of codes (*cf.* [2]). Indeed, define the rank of a finite set X as the cardinality $r'(X)$ of the code generating the submonoid X^*. Then the *defect* of X is the integer:

$$d'(X)=|X|-r'(X) \tag{2}$$

In particular $d'(X_1)=1$ and $d'(X_2)=d'(X_3)=0$. It is well known that determining whether $d'(X)>0$ amounts to determining whether X is a code which

can easily be achieved in time $O(n^2)$ where stands for the sum of the lengths of the words of X using Sardinas and Paterson's algorithm. In [13], the problem of computing the base Y of the smallest free submonoid containing X is resolved.

It is interesting to observe how the complexity dramatically changes when passing from the first definition to the second and to compare our result with the fact that under say reasonable distribution, the probability of being elementary for a set whose cardinality is equal to the cardinality of the alphabet tends to 1 when the sum of the lengths of the words in the set tends to infinity.

The notion of simplifiability or its converse elementariness corresponds to a well known concept in other theories. E.g. the rank of a finite set of vectors in a vector space is the direct analog of our notion of rank. Actually we will observe that the set basis problem has a formulation in terms of simplifiability.

We now shortly describe the contents of our paper.

Section 2 is concerned with the basic definitions used in our paper. The terminology on free monoids is settled and different versions of factorizations of a set of words are proposed. A standard NP-complete problem is stated: the "vertex cover" which will be reduced to simplifiability.

Section 3 establishes the NP-completeness of what was introduced as the strong factorization problem in the preliminaries. It is the basic step towards the proof of our main result.

In section 4 we prove the main theorem and state the direct consequence that deciding whether the rank of a finite set of words is not greater than a given integer is NP-complete.

As another consequence, we prove in section 5 that elementariness is co-NP-complete.

Some remarks in the commutative case are considered at the end of the paper.

2. PRELIMINARIES

2.1. Free monoid

2.1.1. *Factorisation of a set of words*

Given a finite alphabet A, we denote by A^* the free monoid it generates. The elements of A are *letters*, the elements of A^* are *words*. The subset of non empty words is denoted by $A^+ : A^+ = A^* - \{1\}$.

Given a word $w \in A^*$, the following notations are standard: $|w|$ is the length of w, *i.e.* the number of occurrences of letters of A in w. The word of length 0 is the *emtpy* word, denoted by 1. The word w is said *primitive* iff $w = x^n \Rightarrow x = 1$. The *primitive root* of w is the unique primitive word x such that $w = x^n$. For any arbitrary subsets X, $Y \subseteq A^*$ we denote by XY their (concatenation) product: $XY = \{xy \in A^* : x \in X, y \in Y\}$ and by X^* the submonoid generated by X.

We denote by $|X|$ the cardinality of X.

We say that X is *prefix* (resp. *suffix*) iff $X \cap XA^+ = \varnothing$ (resp. $X \cap A^+ X = \varnothing$). A subset is *biprefix* iff it is prefix and suffix.

Given a word w *fact*(w) (resp. *pref*(w)) denotes the set of every *factors* (resp. *prefixes*) of w, *i.e.* the words u such that $w \in A^* u A^*$ (resp. $w \in uA^*$). We set $fact(X) = \bigcup_{w \in X} fact(w)$, and $alph(X) = fact(X) \cap A$.

Let X, $Y \subseteq A^*$ be two arbitrary subsets. Then X is *factorizable* on Y iff $X \subseteq Y^*$, *i.e.*, iff every word of X can be factorized into words of Y. If w belongs to X^*, we denote by $|w|_X$ the smallest integer p such that $w = w_1 \ldots w_p$, where each w_i belongs to X.

We need some precise notions of factorizations.

Let k be an integer belonging to $[1, 2|X| - 1]$ and let $X \subseteq A^*$. We say that:

(1) X is *k-strongly factorizable* iff there exists a set Y such that: $X \subseteq Y^*$, $|Y| \leqq k$ and $X \cap Y = \varnothing$; (as a consequence, words longer than all the words of X can be removed from Y).

(2) X is *k-simplifiable* iff there exists a set Y such that: $X \subseteq Y^*$ and $|Y| \leqq k$ (clearly this definition is usefull only if $k \leqq |X|$).

(3) X is *simplifiable* iff X is k-simplifiable, with $k = |X| - 1$.

The *rank* of X is the integer: $r(X) = \min\{|Y| : X \subseteq Y^*\}$.

2.1.2. *A biprefixity property*

Given a finite subset $Z \subseteq A^*$, we say that Z is *biprefix primitive* iff it satisfies the two following properties:

- Z is biprefix.
- Every word in Z is primitive.

If $X \subseteq A^*$, we denotes by $P[X]$ the set of the primitive roots of X. Clearly $|P[X]| \leqq |X|$ and $X \subseteq P[X]^*$. Moreover, as a direct consequence of [3], we claim that if X is not biprefix then there exists a biprefix set Y such that $|Y| \leqq |X|$ and $X \subseteq Y^*$.

Refering once again to [3], consider the sequence $X=Z_0, Z_1, \ldots$ where $Z_{2i+1}=P[Z_{2i}]$ and Z_{2i} is the basis of the minimal unitary submonoid containing Z_{2i-1}. Clearly $\sum_{w\in Z_n} |w| \geqq \sum_{w\in Z_{n+1}} |w|$ thus there exists an integer n such that $Z_n=Z_{n+1}$. Since the cardinality of Z_i is decreasing and since $X\subseteq Z_i^*$, we may state the following result:

PROPOSITION 2.1: *Let X be a subset of A^*. If X is factorizable on Y then there exists a biprefix primitive set Z such that $|Z|\leqq|Y|$ and $X\subseteq Z^*$.*

2.2. Basic results on NP-completeness

2.2.1. *The Vertex Cover and two of its restrictions*

We assume the reader familiar with the basic notions of NP-completeness (*cf.* [6] or [7]). We recall that the Vertex Cover problem (denoted by VC) can be described in the following way:

Instance: Graph $G=(V,E)$, positive integer $k\leqq|V|$;

Question: Is there a subset $V'\subseteq V$ with $|V'|\leqq k$ such that for each edge $\{\alpha,\beta\}\in E$, at least one of α and β belongs to V'?

Let VC_1 be the restriction of VC to the class of non oriented graphs (V,E), without isolated vertices and such that the set of vertices V, and the set of edges E satisfy $|V|\leqq|E|$.

In [7] the *NP*-completeness of VC is proved by reducing the satisfiability problem 3-SAT to it. This reduction actually assigns to every instance of 3-SAT an instance of VC which meets the condition $|V|\leqq|E|$, *i.e.*, which is in VC_1. This proves that VC_1 is *NP*-complete.

Now let VC_r be the restriction of VC_1 to the class of the graphs (V,E) such that $|V|=|E|$. A simple linear reduction from VC_1 permits to state:

PROPOSITION 2.2: *The problem VC_r is NP-complete.*

Proof: VC_r is the intersection of VC_1 with the set of the graphs having exactly as many edges as vertices. Thus $VC_r\in NP$ because it is the intersection of a set in *NP* with a polynomially decidable set. To show that VC_r is *NP*-hard we establish that it is the linear reduction of VC_1.

Reduction: Let $G=(V,E)$ be an instance of VC_1. Set $\delta=|E|-|V|$. Let A and A' be two sets such that A, A', V are pairwise disjoint, and $|A|=|A'|=\delta$. Let $a\to a'$ be a bijective mapping: $A\to A'$. We set $E'=E\cup\bigcup_{a\in A}\{a,a'\}$ and $V'=V\cup A\cup A'$. Clearly $|E'|=|E|+\delta=|V|+2\delta=|V'|$. Consequently

$G'=(V,E)$ is an instance of VC_r. Moreover, constructing G' requires linear time.

We can easily prove that G has a vertex cover K with $|K|\leqq k$ iff G' has a vertex cover K' with $|K'|\leqq K+\delta$. ■

2.2.2. *NP-ness of the different factorization problems*

Let X be a finite subset of A^+ and k an integer of $[1, 2|X|-1]$.

We define the *strong factorization problem SF* (resp. the *simplifiability problem S*) *as follows*:

Instance: (X,k);

Question: is X k-strongly factorizable? (resp. is X k-simplifiable?).

We denote by S_0 the restriction of S to instances (X,k) such that $k=|X|-1$ and by RANK the problem:

Instance: (X,k);

Question: is the rank of X not greater than k?

PROPOSITION 2.4: *The problems SF, S, S_0 and RANK are in the class NP.*

3. NP-COMPLETENESS OF THE STRONG FACTORIZATION PROBLEM

We denote by 3-*SF* the restriction of *SF* to instances where all words have length equal to 3. We prove here the major step towards our main result, *i.e.*:

PROPOSITION 3.1: *The problem* 3-*SF is NP-complete.*

Proof: We show that 3-*SF* is *NP*-hard by reducing VC_r to it.

1. *Reduction:* Let (G,k) be an instance of VC_r, where G is a graph (V,E) and k an integer belonging to $[1,|E|-1]$. Let $n=|V|=|E|$, and set $A=V\cup T$, where T is disjoint with V and in one-one correspondence with E.

With every edge $\alpha\beta$ of E we associate the word $\alpha a\beta$ belonging to VTV, such that:

(1) If $\alpha\beta$ and $\alpha'\beta'$ are two distinct edges of G then the corresponding letters of T in the words $\alpha a\beta$ and $\alpha' a'\beta'$ are different.

Let X be the subset of A^3 thus obtained. Then X can be constructed in polynomial time. Moreover, G has a vertex cover of cardinality of size k or less iff X is $(k+n)$-strongly factorizable. Indeed:

2. Suppose G has a vertex cover K, of size k or less. Consider the edge $\alpha\beta$.

If α belongs to K then the corresponding word $\alpha a \beta$ is factorized as $\alpha(a\beta)$. Otherwise it is factorized as $(\alpha a)\beta$. According to (1), the factors of length 2 are distinct. Then X is factorizable on $K \cup B$, where $B \subseteq A^2$ and $|B| = n$. Consequently, X is $(k+n)$-strongly factorizable.

3. Suppose now that X is $(k+n)$-strongly factorizable on Y.

Then X can be partitioned in two subsets X_1 and X_2 such that:

- if w belongs to X_1 then w is factorizable as $(\alpha a)\beta$ or $\alpha(a\beta)$;
- if w belongs to X_2 then w is factorizable as $(\alpha)(a)(\beta)$.

For $i = 1, 2$ let V_i (resp. T_i) be the subset of V (resp. $Y - V$) whose elements appear in the factorization of some word in X_i.

For each word w belonging to X_2 let t be an arbitrary letter of w in V_2 and let V_2' be the set of letters thus chosen. Set $K = V_1 \cup V_2'$.

It is a direct consequence of (1) that:

$|T_1| = |X_1|$, $|T_2| = |X_2|$, and T_1, T_2, $V_1 \cup V_2$ are pairwise disjoint.

Then we have

$$|Y| = |V_1 \cup V_2 \cup T_1 \cup T_2| = |V_1 \cup V_2| + |T_1| + |T_2| \geqq |K| + |X|$$

i.e. $|K| \leqq k$. ■

As a corollary we have:

PROPOSITION 3.2: *The problem SF is NP-complete.*

4. NP-COMPLETENESS OF THE RANK

4.1. The case where all the words have length 2

In the special case where $X \subseteq A^2$, solving the factorization problem is particularly simple: it can be easily seen to be polynomial.

PROPOSITION 4.1: *Let A be a finite alphabet, and let X be a finite subset of A^2, with $alph(X) = A$. Computing the rank of X requires time $O(|X|^2)$.*

Proof: With every set $X \subseteq A^2$ we associate the graph $G = (V, E)$ as follows:

The set of vertices is A. The edges are all pairs (a, b) such that $ab \in X$. It is known that determining the (finite) family of connected components of G requires time $O(|E|^2)$ *i.e.* $O(|X|^2)$ (*cf.* [1]).

Let $(V_1, E_1)_{i \in I}$ be this family. Then $r(X) = \sum_{i \in I} \min\{|V_i|, |E_i|\}$, and determining this integer requires time $O(\max\{|V|, |E|\})$. ■

4.2. The general case

We proceed to the proof that simplifiability is *NP*-complete.

THEOREM 4.2: *The problem S is NP-complete.*

Proof: To show that S is *NP*-hard, we establish that it is the polynomial reduction of 3-*SF*.

1. *Reduction.* Let (X, k) be an instance of 3-*SF* ($|X|=n$), and let A_1 and A_2 be two disjoint alphabets in one-one correspondence with X. Set $\Sigma = A_1 \cup A_2 \cup A$. With every word $w = abc \in X$ we associate the *slice* of four words in Σ^3:

$$T_w = \{ abc, aba_1, a_2 bc, a_2 ba_1 \}$$

such that:

(1) $a_1 \in A_1$, $a_2 \in A_2$

(2) if abc, a' b' c' are two distinct words of X then the corresponding letters a_1 and a_1', a_2 and a_2' are pairwise different.

Clearly the subsets T_w are pairwise disjoint. Set: $Z = \bigcup_{w \in X} T_w$. Constructing Z requires linear time.

2. We shall verify that if X is k-strongly factorizable then Z is $(k+2n)$-simplifiable. Indeed, if X is factorizable on Y where $|Y|=k$, then every word $w = abc \in X$ is factorized in one of the following ways:

- $w = (ab)\ (c)$, and then:

$$T_w \subseteq \{ ab, c, a_2 b, a_1 \}^* \qquad \text{with} \quad ab, c \subseteq Y; \tag{3}$$

- $w = (a)\ (bc)$, and then:

$$T_w \subseteq \{ a, bc, ba_1, a_2 \}^* \qquad \text{with} \quad a, bc \subseteq Y; \tag{4}$$

- $w = (a)\ (b)\ (c)$, and then:

$$T_w \subseteq (\{ a \} \cup \{ b \} \cup \{ c \} \cup \{ a_1, a_2 \})^* \tag{5}$$

with a, b, $c \in Y$ (a, b, c are not necessarily different).

In all cases Z can be factorized on $Y \cup B$ where $|B| = 2n$. Thus $|Y \cup B| = 2n+k$.

3. Before proving the converse of (2) let us establish the following result:

LEMMA 4.3: *If Z is $(k+2n)$-simplifiable then there exists a $(k+2n)$-strong factorization of Z such that every slice T_w is factorizable according to the scheme* (3), (4) *or* (5).

Proof: Assume $Z \subseteq Y^*$ with $|Y| \leqq k+2n$. According to Proposition 2.1, Y can be supposed biprefix. For a given $w=abc \in X \subseteq Z$, different factorizations may occur:

– suppose first w that belongs to Y. According to the biprefixity of Y, we have $T_w \subseteq Y$. (The proof is straighforward.) Note that no element of T_w belongs to $Z-Y$. Then:

. if $c \neq a$ and $c \neq b$ then the set $\{ab, c, a_1, a_2 b\} \cup (Y - T_w)$ is biprefix and can be substituted to Y. Its cardinality is equal to $|Y|$.

. otherwise the biprefix set $\{a\} \cup \{b\} \cup \{c\} \cup \{a_1, a_2\} \cup (Y - T_w)$ can be substituted to Y. Again its cardinality is $|Y|$ or less.

– According to these results, it can be assumed that Z is $(k+2n)$-strongly factorizable on a biprefix set Y. The proof is completed by examining the different ways of factorizing of each word $w \in X$, and by using again the argument of biprefixity of Y. □

4. Now, it is a direct consequence of Lemma 4.3 and claim (2) that X is k-strongly factorizable on $Y \cap A^*$. But $|Y \cap A^*| \leqq |Y| - 2n \leqq k$. ■

As a consequence we have the following result:

THEOREM 4.5: *Given a finite set of words it is NP-complete to decide whether its rank is not greater than a given integer r.*

5. CO-NP-COMPLETENESS OF ELEMENTARINESS

We are now able to prove our main theorem.

THEOREM 5.1: *Given a finite set of words X, deciding whether X is elementary is co-NP-complete.*

Proof: We show that the problem S_0 is *NP*-hard by reducing S to it.

1. *Reduction*: Let (X, d) be an instance of S. Let $n=|X|$ and $A=\text{alph}(X)$. Let B be an alphabet such that $A \cap B = \varnothing$ and $|B| = n-d-1$. Set $k=|B|$, $B=\{b_1, \ldots, b_k\}$, and $X=\{x_1, \ldots, x_n\}$. With every word $x_j \in X$ we associate the word $w_j = x_j b_1^j \ldots b_k^j$.

Let Z be the subset of $(A \cup B)^*$ thus obtained. Because constructing the word w_j requires time $O(nk)$, constructing Z requires time $O(n^3)$.

2. Trivially, if $r(X) \leqq d$, Z is simplifiable.

Before proving the converse, let us establish the following result:

LEMMA 5.2: *Let T be a set of words $w_{ij} = b_1^i b_2^j \ldots b_k^j$ satisfying the following properties:*

1. $0 \leqq i \leqq j \leqq n$.
2. *T contains a word w_{ij} with $i > 0$.*
3. *If $w_{ij} \in T$ and w_{pj} then $i = p$.*

Then $r(T) = min\{|T|, k\}$.

Proof: Clearly $r(T) \leqq |T|$. Assume $r(T) < |T|$ and let Y be a biprefix primitive set such that $T \subseteq Y^*$ and $r(T) = |Y|$. We shall first prove that $B - \{b_1\} \subseteq Y$.

Suppose $(B - \{b_1\}) - Y \neq \varnothing$; we can assume without loss of generality that $k = \max\{m : b_m \notin Y\}$. Necessarily, all the elements of T need in their factorization a word belonging to $B^* b_{k-1} b_k^+$. Since $r(T) < |T|$ there exists a word $y \in Y \cap B^* b_{k-1} b_k^+$ used to factorize two distinct words of T. By construction that means that $T \cap Y^* y b_k^+ \neq \varnothing$. Since Y is biprefix primitive, we have $b_k \in Y$, a contradiction with $b_k \notin Y$. Hence $B - b_1 \subseteq Y$.

Let $w_{ij} \in T$, with $i > 0$. We have $w_{ij} \in b_1^+ (B - b_1)^* \subseteq b_1^+ Y^*$, and since Y is biprefix primitive, we obtain $b_1 \in Y$. Thus $B \subseteq Y$ and the result follows. □

3. Suppose Z is simplifiable. Let U be a biprefix set such that $Z \subseteq U^*$ and $|U| \leqq n - 1$.

With each word $w \in Z$, we associate two words y, t as follows:

– t is the longest suffix of w which belongs to $U^* \cap B^*$.

– If $wt^{-1} \in A^*$, we set $y = 1$ else y is the suffix of wt^{-1} which belongs to U.

Let T, T' and Y be the sets of the words t, wt^{-1} and y thus constructed. We now define the sets Y_i, Z_i, T_i, T_i', U_i, U_i' $(i = 1, 2)$ as follows:

– Z_1 is the subset of Z whose corresponding words y belong to $A^+ b_1^+ (B - b_1)^+$. Let $Z_2 = Z - Z_1$.

– Y_i, T_i and T_i' are the subsets of Y, T and T' corresponding to the words of $Z_i (i = 1, 2)$.

– U_i (resp. U_i') is the minimal subset of U such that $Z_i \subseteq U_i^*$ (resp. $T_i' \subseteq U_i'^*$).

By construction of Z, we have $Y_i \cap U_2 = \varnothing$. Hence:

$$|U| \geqq |Y_1| + |U_2| = |Z_1| + |U_2|.$$

Since $|Z| - 1 \geqq |U|$, we obtain: $|Z_1| + |Z_2| - |1\, Z|Z_1| + |U_2|$, thus:

$$|U_2| \leqq |Z_2| - 1. \tag{1}$$

By construction, T_2 satisfies hypothesis 1 and 3 of Lemma 5.2.

If $1 \in Y_2$ then hypothesis 2 is trivially satisfied. In the other case, according to (1) we have $|Y_2| < |Z_2|$, thus there exists two words w, $w' \in Z_2$ whose corresponding words y, $y' \in Y_2$ satisfy $y = y'$. By construction, one of the corresponding words t, $t' \in T_2$ belongs to $b_1^+ B^*$ thus hypothesis 2 is satisfied.

As a consequence we have $r(T_2) = \min\{|T_2|, k\}$.

Suppose we have $r(T_2) = |T_2| < k$. Since if $k = 1$ we have $r(T_2) = 1 = k$. It follows that $k \geqq 2$ and $|Z_2| = |T_2|$. Since $U_2' \subseteq A^* b_1^*$ and $U_2 - U_2' \subseteq B^+$, we have $U_2 \cap U_2' = \varnothing$, hence:

$|U_2| \geqq |U_2'| + r(T_2) = |U_2'| + Z_2|$, which contradicts (1).

As a consequence we obtain $r(T_2) = k$. Therefore we have:

$$|U_2'| \leqq |U_2| - k. \tag{2}$$

For a given word $y \in Y - 1$, let $p(y)$ be the longest prefix of y belonging to A^*. We have $X \subseteq ((U_1' \cup U_2' - Y) \cup \{p(y) : y \in Y\})^*$, and since $|(U_1' \cup U_2' - Y) \cup \{p(y) : y \in Y\}| \leqq |U_1' \cup U_2'|$, we have $r(X) \leqq r(T')$.

But $T' \subseteq ((U_1' - U_2') \cup U_2')^*$ thus $r(T') \leqq |U_1' - U_2'| + |U_2'|$.

If $|T_1'| \leqq |U_1' - U_2'|$ then, since $T' \subseteq (T_1' \cup T_2')^*$, we can substitute the set T_1' to $U_1' - U_2'$, hence we may assume that $|U_1' - U_2'| \leqq |T_1'|$.

Consequently:

$r(T') \leqq |T_1'| + |U_2'| = |Z_1| + |U_2'|$, hence according to (2):

$r(T') \leqq |Z_1| + |U_2| - k = |Z| - |Z_2| + |U_2| - k$, and according to (1):

$r(T') \leqq n - k - 1 = d$, thus $r(X) \leqq d$, which achieves the proof. ■

Remarks: It is of interest to examine the connections with other similar problems over commutative monoids and commutative idempotent monoids.

In fact, the "set basis problem" (*cf.* [6], Appendix A 3) can be reformulated in terms of monoids as:

Deciding whether a finite subset of the idempotent commutative free monoid is simplifiable is NP-complete.

Denoting by $C\mathscr{P}$ the extension of a problem $\mathscr{P}$ to the free commutative monoids, the proofs of the non commutative case may be modified to obtain the following result:

The problems CSF, CS, CRANK are NP-complete.

REFERENCES

1. S. Baase, Introduction to Design and Analysis. *Addison Wesley*, 1982.
2. J. Berstel and D. Perrin, Theory of codes, *Academic Press*, 1985.
3. J. Berstel, D. Perrin, J. F. Perrot and A. Restivo, Sur le théorème du défaut, Journal of Algebra, September 1979, *60*, n° 1, pp. 169-180.
4. K. Culik II and I. Fris, The decidability of equivalence problem for DOL-systems, *Inform. Control*, 1977, *33*, pp. 20-39.
5. A. Ehrenfeucht and G. Rosenberg, Elementary homomorphisms and a solution to DOL sequence equivalence problem, *Theoret. Comput. Sci.*, 1978, *7*, pp. 76-85.
6. M. Garey and D. Johnson, Computers and intractability, *W. H. Freeman and Company*, 1979.
7. J. Hopcroft and J. Ullman, Introduction to automata theory, languages, and computations, *Addison-Wesley Publishing Company*, 1979.
8. J. Karhumäki, On Recent Trends in Formal Language Theory, in the Procedings of the 14th ICALP, 1987, pp. 136-161.
9. M. Lothaire, Combinatorics on Words, Encyclopedia of Mathematics and its Applications, *Addison-Wesley*, 1983.
10. G. S. Makanin, The problem of solvability of equations in a free semigroup, Math. Sb., 1977, *103*, pp. 147-236 (English translation) in *Math USSR Sb.*, 1979, *32*, pp. 129-19.
11. J. P. Pécuchet, Sur la détermination du rang d'une équation dans le monoïde libre, *Theoret. Comput. Sci.*, 1981, *16*, pp. 337-340.
12. G. Rozenberg and A. Salomaa, The Mathematical Theory of L Systems, *Academic Press*, 1980.
13. J. C. Spehner, Quelques problèmes d'extension, de conjugaison et de présentation des sous-monoïdes d'un monoïde libre, Thèse, Université Paris-VII (France), 1976.
14. L. Valiant, The equivalence problem for DOL systems, in the Proceedings of the 3rd ICALP, 1976, pp. 31-37.