ROMANIAN JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY
Volume **13**, Number 3, 2010, 255–278

Elementary Alphabets of a Language

Daniel Claudian VOINESCU

MOCALC - Models of Computation, Algorithms and Cryptography,
University of Bucharest,
E-mail: voinescu.daniel@bcr.ro

Abstract. The paper pertains to the field of Combinatorics on words. We introduce the notion of elementary alphabet of a language as the natural generalization of the combinatorial alphabet introduced in [10]. By analogy with the famous Ehrenfeucht conjecture, we prove a compactness property of all words in a language, namely, that any language has a finite elementary test-set. Then we show that the class of all elementary sets of a regular language can be effectively constructed. As a consequence, it is decidable whether a given finite part of a regular language is an elementary test-set for that language.

Key-words: combinatorics on words, combinatorial dimension, combinatorial alphabet, elementary alphabet, elementary test-set.

1. Introduction

The terminology and notations used in this section can be found in the Preliminaries section below.

We briefly outline the importance and usefulness of the notion of combinatorial degree by two facts:

– this notion allows a formalization of the well known defect effect, which is considered to be folklore knowledge in mathematics [2].

– notions of elementary set and elementary morphism were introduced based on the combinatorial degree. It is worth mentioning that these notions are important ingredients of an elegant proof of the DoL equivalence problem in [3].

Unlike the case of other algebraic ranks (see [2] for example), the problem of computing the combinatorial degree of a given finite language is a very difficult problem. In [8] and [9] the following results were proven:

Theorem 1. *(i) The problem of deciding, for a finite set $L \subset \Sigma^*$ and for a given integer $k \geq 3$, whether $d(L) \leq k$ is NP-complete.*

(ii) The problem of deciding whether a given finite set is not elementary is NP-complete.

(iii) For $k = 2$ the problem (i) can be solved in time $\mathcal{O}(n \log^2 m)$ where $n = sizeof(L)$ and $m = \max\{|w| \mid w \in L\}$.

Due to these complexity results, algorithms for computing the combinatorial degree of a language or for finding a combinatorial alphabet of a language do not appear in the literature. The first paper that defines the class of all combinatorial alphabets of a language and investigates its properties is [10]. The same paper introduced the notion of combinatorial test-set (by analogy with the famous notion of test-set), proved a compactness property of words and tried to find some relations with the famous Ehrenfeucht conjecture.

The present paper has three main purposes:

- To introduce the notion of elementary alphabet of a language and to investigate the main properties of the class of all the elementary alphabets of a language.
- To prove that any language (not necessarily regular) has a finite elementary test-set (this result generalizes the result for combinatorial test-sets presented in [10]).
- To prove that for regular languages the class of all elementary sets can be effectively constructed and some problems regarding elementary alphabets are decidable.

2. Preliminaries

2.1. Basic Terminology, Notations and Definitions

Throughout the paper Σ will denote a finite nonempty set called *alphabet*, and Σ^+ and Σ^* will be the free semigroup and the free monoid generated by Σ. The elements of Σ^* will be called *words* and any set of words will be called a *language over* Σ. The empty word will be denoted by λ. For a word w we shall denote its length by $|w|$ and we shall write it by letters $w = w[1]w[2]\dots w[|w|]$, with $w[1], w[2], \dots, w[|w|] \in \Sigma$. For L, L_1, $L_2 \subseteq \Sigma^*$: $L_1L_2 = \{x_1x_2 \mid x_1 \in L_1,\ x_2 \in L_2\}$, $L^0 = \{\lambda\}$, and for any positive integer n: $L^n = LL^{n-1}$, $L^{\leq n} = \bigcup_{k=0}^{n} L^k$, $L^{\geq n} = \bigcup_{k=n}^{\infty} L^k$, $L^* = \bigcup_{n\geq 0} L^n$. For a finite language L we denote $sizeof(L) = \sum_{w\in L} |w|$.

The *combinatorial degree* or *rank* or *dimension* of a language L, [5], is the positive integer

$$d(L) = \min\{card(A) | A \subseteq \Sigma^*, L \subseteq A^*\}.$$

It follows immediately from the definition that $d(L) \leq card(\Sigma)$ and if $L_1 \subseteq L_2$ then $d(L_1) \leq d(L_2)$.

We say that *the elements of a set $A \subseteq \Sigma^*$ satisfy a nontrivial relation* iff there exists a word $w \in A^*$ which has two different factorizations over A.

A language $C \subseteq \Sigma^*$is called a *code* if any $w \in C^*$ has a unique factorization over C. Sometimes when we want to point out one or more of the factors in the unique factorization of w over C, we shall use "$\cdot$" before and after the respective factors. For example if we want to point out the factor v in $w = uvxy$ we shall write $w = u \cdot v \cdot xy$ where $u, xy \in C^*$ and $v \in C^+$.

If C is a code and $\alpha \in C^*$, the set of the words from C that appear in the unique factorization of α over C will be called *the alphabet of α over the code C* and will be denoted by $alph_C(\alpha)$. For $L \subseteq C^*$, the set $\bigcup_{\alpha \in L} alph_C(\alpha)$ will be called *the alphabet of the language L over the code C* and will be denoted by $alph_C(L)$. In the particular case $C = \Sigma$ we shall write simply $alph(\alpha)$ and $alph(L)$ instead of $alph_\Sigma(\alpha)$ and $alph_\Sigma(L)$.

An $A \subset \Sigma^*$ for which $d(A) = card(A)$ is called an *elementary set*. We shall denote the set of all the subsets of Σ^* which are elementary sets by $\mathcal{E}(\Sigma^*)$. It is obvious that each elementary set is a code.

For a set M we denote $\mathcal{P}(M)$ the set of all the subsets of M and $\mathcal{P}_{fin}(M)$ the set of all the finite subsets of M.

If X is a nonempty set and $\varphi : X \longrightarrow \Sigma^*$ we shall denote by φ^* the unique extension of φ to a morphism of monoids from X^* to Σ^*.

For a partially ordered set $(M, \leq)$ we shall denote with $Min(M)$ and $Max(M)$ the set of all minimal elements of M and the set of all maximal elements of M.

A subset T of a language $L \subseteq \Sigma^*$ is called *a test-set of L* if:

$$\forall \Delta \neq \emptyset \; \forall f, g : \Sigma \longrightarrow \Delta^* \; (f^*|_L = g^*|_L \Longleftrightarrow f^*|_T = g^*|_T)$$

that is if two morphisms defined on Σ^* agree on T, then they agree on L.

Theorem 2. *(Ehrenfeucht Conjecture) For any language there exists a finite test-set.*

A *deterministic finite automaton*, *DFA* for short, is a quintuple $M = (Q, \Sigma, \delta, q_0, F)$, where:

Q is the finite nonempty *set of states*;
Σ is the finite nonempty set of input symbols - the *input alphabet*;
$\delta : Q \times \Sigma \longrightarrow Q$ is the *state transition function*;
q_0 is the *starting/initial state*;
$F \subseteq Q$ is the *set of final states.*

For convenience, we define a natural extension of δ namely $\delta^* : Q \times \Sigma^* \longrightarrow Q$ inductively as follows $\delta^*(q, \lambda) = q$ and $\delta^*(q, wa) = \delta(\delta^*(q, w), a)$ for $q \in Q$, $a \in \Sigma$ and

$w \in \Sigma^*$. The *language accepted/recognized by the DFA* M, denoted $L(M)$ is defined as follows:

$$L(M) = \{w \mid w \in \Sigma^*,\ \delta^*(q_0, w) \in F\}\,.$$

Two DFAs are called *equivalent* iff they recognize the same language.

For all the notations and definitions above the reader is referred for example to [11], [2] and [4].

We shall also use in the present paper the following definitions introduced in [10]:

- *A combinatorial alphabet of the language* L is a subset A of Σ^* for which $L \subseteq A^*$ and $card(A) = d(L)$. The class of all the combinatorial alphabets of the language L is denoted by $\mathcal{CA}(L)$, therefore

$$\mathcal{CA}(L) = \{A \mid A \subseteq \Sigma^*,\ L \subseteq A^*,\ card(A) = d(L)\}.$$

- *A minimal elementary set* is any $A \subseteq \Sigma^*$ for which $\mathcal{CA}(A) = \{A\}$. It is obvious that a minimal elementary set is an elementary set and that Σ is a minimal elementary set.

- *A minimal combinatorial alphabet of the language* L is a combinatorial alphabet of L which is a minimal elementary set. *The class of all the minimal combinatorial alphabet*s of L is denoted by $Min\mathcal{CA}(L)$.

- *A maximal combinatorial alphabet of the language* L is a *combinatorial alphabet* A, for which

$$\forall B \in \mathcal{CA}(L)\ (L \subseteq B^* \subseteq A^* \Rightarrow A = B).$$

  The class of all the maximal combinatorial alphabets of L is denoted by $Max\mathcal{CA}(L)$.

- *A combinatorial test-set for the language* L is a subset L' of L for which $\mathcal{CA}(L) = \mathcal{CA}(L')$.

2.2. New Definitions and Notations

Besides the definitions and notations mentioned above, we shall introduce and use in the present paper the following ones:

Definition 1. *An elementary alphabet of the language* L is an elementary set E such that $L \subseteq E^*$ and $alph_E(L) = E$.

This is a natural generalization of the notion of combinatorial alphabet of a language. The idea is to allow/consider also other intermediary levels of atoms/bricks – indivisible units – in the construction of a language, not only the lowest level of the alphabet Σ and the highest level of the combinatorial alphabets. We materialized this idea by relaxing the tough condition $card(E) = d(L)$ (which, provided $L \subseteq E^*$, implies that E is an elementary set, that $alph_E(L) = E$ and that $d(E) = d(L)$) by replacing it only with: E elementary set and $alph_E(L) = E$. Therefore, an elementary

alphabet E of the language L over Σ can have any combinatorial dimension between $d(L)$ and $card(\Sigma)$.

The class of all the elementary alphabets of the language L will be denoted by $\mathcal{EA}(L)$, therefore

$$\mathcal{EA}(L) = \left\{ E \;\middle|\; \begin{array}{l} E \subseteq \Sigma^*,\ card(E) = d(E), \\ L \subseteq E^*,\ alph_E(L) = E \end{array} \right\}$$

For $\diamond \in \{<, \leq, =, \geq, >\}$ and $k \in \mathbf{N}$ we shall use the following notation:

$$\mathcal{EA}_{\diamond k}(L) = \{E \mid E \in \mathcal{EA}(L),\ card(E) \diamond k\}$$

If $\diamond$ is " = " then we shall write simply $\mathcal{EA}_k(L)$ instead of $\mathcal{EA}_{=k}(L)$.

Remark 1. The following properties are obvious or follow immediately from the above definitions:

1. $\mathcal{EA}_{d(L)}(L) = \mathcal{CA}(L)$
2. $(\mathcal{CA}(L), \preceq) \subseteq (\mathcal{EA}(L), \preceq)$ as partially ordered sets, with " $\preceq$ " $\subseteq \mathcal{E}(\Sigma^*) \times \mathcal{E}(\Sigma^*)$ defined as in [10] : $A \preceq B \overset{def}{\Longleftrightarrow} B \subseteq A^*$.
3. $\forall k \leq d(L)\ \ (\mathcal{EA}_k(L) = \emptyset)$
4. $\forall k > card(alph(L))\ \ (\mathcal{EA}_k(L) = \emptyset)$
5. $\forall E \in \mathcal{EA}(L)\ \ (d(L) \leq card(E) = d(E) \leq card(alph(L)))$
6. $\forall k \neq l \in \mathbb{N}\ \ (\mathcal{EA}_k(L) \cap \mathcal{EA}_l(L) = \emptyset)$
7. $\mathcal{EA}(L) = \bigcup\limits_{k=d(L)}^{card(alph(L))} \mathcal{EA}_k(L)$ is a finite union of finite and pairwise disjoint sets.
8. if: $L_1,\ L_2 \subseteq \Sigma^*,\ d(L_1) = d(\ L_2) = d,\ card(alph(L_1)) = card(alph(L_2)) = \sigma$ then: for each $k = \overline{d, \sigma}$

$$\begin{array}{ll} (i) & \mathcal{EA}_{\leq k}(L_1) = \mathcal{EA}_{\leq k}(L_2) \Longleftrightarrow \forall i \leq k \quad (\mathcal{EA}_i(L_1) = \mathcal{EA}_i(L_2)) \\ (ii) & \mathcal{EA}_{\leq k}(L_1) \subseteq \mathcal{EA}_{\leq k}(L_2) \Longleftrightarrow \forall i \leq k \quad (\mathcal{EA}_i(L_1) \subseteq \mathcal{EA}_i(L_2)) \end{array}$$

Definition 2. *An elementary test-set for the language* L is a subset L' of L for which $\mathcal{EA}(L) = \mathcal{EA}(L')$.

3. Some useful lemmas

The well known defect effect will be extensively used in this paper. It can be formalized in many interesting ways, see for example [2]. To be more specific, for our purpose, we shall state the most common formalization of the defect effect:

Theorem 3. *If a finite set $X \subset \Sigma^*$ has the property that its elements satisfy a nontrivial relation, then there exists a set $Y \subset \Sigma^*$ such that $card(Y) < card(X)$ and $X \subset Y^*$.*

In this section we shall prove some lemmas that will be used later. Another reason for mentioning them as standalone results is that we think they may be useful by themselves as they are fundamental properties of elementary sets.

Lemma 1. *If E is an elementary set with $n \geq 2$ elements and $u, v \in E$, then $E' = (E - \{u, v\}) \cup \{uv\}$ is an elementary set with $n-1$ elements and $E \preceq E'$.*

Proof. It is obvious that $E' \subseteq E^*$. Suppose, to the contrary, that E' is not elementary. Then, according to the defect effect, $\exists X \subseteq \Sigma^*$ ($E' \subseteq X^*$, $card(X) \leq n-2$). Let $A \in \mathcal{CA}(X)$ and $uv = \alpha_1\alpha_2\cdots\alpha_k$ the unique factorization of uv over A. If $\exists l < k$ ($u = \alpha_1\alpha_2\cdots\alpha_l$, $v = \alpha_{l+1}\cdots\alpha_k$), then $d(E) \leq n-2$ and we have got a contradiction. If $\exists l \leq k$ ($u = \alpha_1\alpha_2\cdots\alpha_{l-1}\alpha'_l$, $v = \alpha''_l\alpha_{l+1}\cdots\alpha_k$, $\alpha'_l\alpha''_l = \alpha_l$ and $\alpha'_l, \alpha''_l \neq \lambda$), then $E \subseteq ((A - \{\alpha_l\}) \cup \{\alpha'_l, \alpha''_l\})^*$ and therefore $d(E) \leq n-1$ and again we have got a contradiction, hence we have to conclude that E' is elementary. □

Lemma 2. *If E is an elementary set and $u, uv \in E$, then $E' = (E - \{uv\}) \cup \{v\}$ is an elementary set and $E' \preceq E$, moreover $E' \in \mathcal{CA}(E)$.*

Proof. It is obvious that $E \subseteq (E')^*$. Suppose, to the contrary, that E' is not elementary. Then, according to the defect effect, $\exists X \subseteq \Sigma^*$ ($E' \subseteq X^*$, $card(X) < card(E')$). Therefore we have $card(X) < card(E') = card(E)$ and $E \subseteq (E')^* \subseteq X^*$ and we contradicted the elementarity of E.

Finally, let us prove that $alph_{E'}(E) = E'$. Assuming, to the contrary, that $alph_{E'}(E) \subset E'$ it follows that $uv \in (E - \{uv\})^*$ which contradicts the elementarity of E. □

Lemma 3. *If E is a minimal elementary set and $card(E) < card(alph(E))$, then for each $w = uv \in E$ with $u, v \neq \lambda$ the set $E' = (E - \{w\}) \cup \{u, v\}$ is an elementary set, $E' \preceq E$ and $alph_{E'}(E) = E'$ and therefore $E' \in \mathcal{EA}_{card(E)+1}(E)$.*

Proof. First let us notice that $card(alph(E)) > card(E) \Longrightarrow \exists w$ ($|w| \geq 2$). Let $w = uv \in E$ with $u, v \neq \lambda$. Obviously, $E \subseteq (E')^2$ and $card(E') = card(E)+1$ (indeed if $u = v$, then $E \subseteq ((E - \{w\}) \cup \{u\})^*$ therefore E is not minimal and we have got a contradiction). Let us prove now that E' is elementary. Suppose, to the contrary, that it is not, then, according to the defect effect, $\exists A$ ($E' \subseteq A^*$, $card(A) \leq card(E)$). If $card(A) < card(E)$, then $d(E) < card(E)$, which is a contradiction. If $card(A) = card(E)$, and then from $E \subseteq A^*$, $card(A) = card(E)$ and $d(E) = card(E)$ it follows that $A \in \mathcal{CA}(E)$ but $A \neq E$ (because obviously $sizeof(A) < sizeof(E)$) therefore E is not minimal, contradiction.

Finally, let us prove that $alph_{E'}(E) = E'$. Supposing, to the contrary, that $alph_{E'}(E) \subset E'$ it follows that at least one of the following statements: (i) $\exists x \in E'$ ($x \notin alph_{E'}(E)$) or (ii) ($u \notin alph_{E'}(E)$) $\vee$ ($v \notin alph_{E'}(E)$) is true. We shall prove that both of them contradict the minimality of E. If (i) holds, let $F = E' - \{x\}$. We have: F elementary, $x \in F^+$, $E \subseteq F^+$, $F \neq E$ and $card(F) = card(E)$ therefore $F \in \mathcal{CA}(E)$ and $F \neq E$ which tells us that E is not minimal. If (ii) holds then let us suppose that

$u \notin alph_{E'}(E)$ (the case $v \notin alph_{E'}(E)$ can be treated similarly). Let $F = E' - \{u\}$. We have: F elementary, $w \in F^+$, $E \subseteq F^+$, $F \neq E$ and $card(F) = card(E)$ therefore $F \in \mathcal{CA}(E)$ and $F \neq E$ which tells us again that E is not minimal. □

Lemma 4. *If* $card(L) < \infty$ *then* $card(\mathcal{EA}(L)) < \infty$. *Implicitly* $\mathcal{EA}(L)$ *and, for any* k, $\mathcal{EA}_k(L)$ *are finite.*

Proof. Obviously $card(\mathcal{P}_{\leq sizeof(L)}(\Sigma^{\leq \max\{|w||w\in L\}}))$ is a rough majorant for $card(\mathcal{EA}(L))$. □

Lemma 5. *If* $L_1 \subseteq L_2$ *then:*

1. $\forall E \in \mathcal{EA}(L_2)\ \exists A \in \mathcal{EA}(L_1)\ (A \subseteq E)$
2. $Max(\mathcal{EA}(L_1) \cup \mathcal{EA}(L_2)) \subseteq Max(\mathcal{EA}(L_1))$
3. $Min(\mathcal{EA}(L_1) \cup \mathcal{EA}(L_2)) = Min(\mathcal{EA}(L_2)) = \{alph(L_2)\}$.

Proof. We shall prove the first claim; the second follows immediately from the first; the third one is obvious. Let $E \in \mathcal{EA}(L_2)$ i.e. $L_2 \subseteq E^*$, E is an elementary set and $alph_E(L_2) = E$. We have $L_1 \subseteq L_2 \subseteq E^*$ therefore $L_1 \subseteq E^*$. Let $A = alph_E(L_1)$, obviously we have $A \subseteq E$, A elementary set (as a subset of an elementary set), $L_1 \subseteq A^*$ and $alph_A(L_1) = A$ therefore we have also $A \in \mathcal{EA}(L_1)$. □

Lemma 6. *If* $L_1 \subseteq L_2$, $d(L_1) = d(L_2) = d$ *and* $\sigma = card(alph(L_1)) = card(alph(L_2))$ *then for each* $k = \overline{d, \sigma - 1}$

$$\mathcal{EA}_{\leq k}(L_1) = \mathcal{EA}_{\leq k}(L_2) \implies \mathcal{EA}_{k+1}(L_1) \supseteq \mathcal{EA}_{k+1}(L_2).$$

Proof. Let $k \in \{d, d+1, \ldots, \sigma - 1\}$ such that $\mathcal{EA}_{\leq k}(L_1) = \mathcal{EA}_{\leq k}(L_2)$ and let $E \in .\mathcal{EA}_{\leq k+1}(L_2)$. Let $A \in \mathcal{EA}(L_1)$ with $A \subseteq E$ (there exist such an A according to the previous lemma). If $card(A) < card(E)$ then $A \in \mathcal{EA}_{\leq k}(L_1)$ (because $card(E) = k+1$) therefore $A \in \mathcal{EA}_{\leq k}(L_2)$ so $alph_E(L_2) = A \subsetneq E$ which is in contradiction with $E \in \mathcal{EA}(L_2)$ and therefore we must have $card(A) = card(E)$ therefore $E = A \in \mathcal{EA}_{\leq k+1}(L_1)$ and therefore $\mathcal{EA}_{\leq k+1}(L_2) \subseteq \mathcal{EA}_{\leq k+1}(L_1)$. □

Lemma 7. *If* $L_1 \subseteq L_2 \subseteq L_3$ *and* $\mathcal{EA}(L_1) = \mathcal{EA}(L_3)$, *then* $\mathcal{EA}(L_1) = \mathcal{EA}(L_2) = \mathcal{EA}(L_3)$.

Proof. From $\mathcal{EA}(L_1) = \mathcal{EA}(L_3)$ it follows $alph(L_1) = alph(L_3)$, $d(L_1) = d(L_3)$ and $\forall k\ (\mathcal{EA}_k(L_1) = \mathcal{EA}_k(L_3))$. From $L_1 \subseteq L_2 \subseteq L_3$, $d(L_1) = d(L_3)$ and $alph(L_1) = alph(L_3)$ it results $d(L_1) = d(L_2) = d(L_3)$ and $alph(L_1) = alph(L_2) = alph(L_3)$; let us denote by d and by m the common combinatorial dimension of the languages L_1, L_2 and L_3 and respectively the cardinality of their common alphabet over Σ. Now let us prove inductively that $\forall k = \overline{d, m}\ (\mathcal{EA}_k(L_1) = \mathcal{EA}_k(L_2))$:

Basis: "$k = d$"

$L_1 \subseteq L_2$ and $d(L_1) = d(L_2)$ implies $\mathcal{CA}(L_1) \supseteq \mathcal{CA}(L_1)$ (see [10]).
$L_2 \subseteq L_3$ and $d(L_2) = d(L_3)$ implies $\mathcal{CA}(L_2) \supseteq \mathcal{CA}(L_3)$ (see again [10]).
Since we have also $\mathcal{CA}(L_1) = \mathcal{CA}(L_3)$ it follows $\mathcal{CA}(L_1) = \mathcal{CA}(L_2)$.

Inductive step: "$k \longrightarrow k+1$" Using lemma 6, statement 2:
$L_1 \subseteq L_2$ and $\mathcal{EA}_k(L_1) = \mathcal{EA}_k(L_2)$ implies $\mathcal{EA}_{k+1}(L_1) \supseteq \mathcal{EA}_{k+1}(L_1)$.
$L_2 \subseteq L_3$ and $\mathcal{EA}_k(L_2) = \mathcal{EA}_k(L_3)$ implies $\mathcal{EA}_{k+1}(L_2) \supseteq \mathcal{EA}_{k+1}(L_3)$.
Since we have also $\mathcal{EA}_{k+1}(L_1) \supseteq \mathcal{EA}_{k+1}(L_3)$ it follows $\mathcal{EA}_{k+1}(L_1) = \mathcal{EA}_{k+1}(L_2)$. □

Lemma 8. *"A kind of transitivity".*
1. $E_1 \in \mathcal{EA}(L)$ & $E_2 \in \mathcal{EA}(E_1) \Longrightarrow E_2 \in \mathcal{EA}(L)$ *(or obviously equivalent* $\forall E \in \mathcal{EA}(L)$ $(\mathcal{EA}(E) \subseteq \mathcal{EA}(L))$*).*
2. $E_1 \in \mathcal{EA}_k(L)$ & $E_2 \in \mathcal{EA}_l(E_1) \Longrightarrow E_2 \in \mathcal{EA}_l(L)$ *for each* $k = \overline{d(L), card(alph(L))}$ *and* $l = \overline{k, card(alph(L))}$.
3. For any k *the following relations hold:*

$$\begin{aligned} Min(\mathcal{EA}_k(L)) &= \{E \mid E \in \mathcal{EA}_k(L) \ \&\ E \text{ minimal elementary set}\} = \\ &= \bigcup_{E \in \mathcal{EA}_k(L)} Min(\mathcal{CA}(E)) \end{aligned}$$

Proof. 1. Obviously E_2 is an elementary set and $L \subseteq E_1^* \subseteq E_2^*$ therefore it suffices to prove that $alph_{E_2}(L) \supseteq E_2$. Let $\alpha \in E_2$, since $alph_{E_2}(E_1) = E_2$ it follows that there exists a $u \in E_1$ such that α occurs as a factor in the unique factorization of u over E_2. Since $alph_{E_1}(L) = E_1$ it follows that there exists a $v \in L$ such that u occurs as a factor in the unique factorization of v over E_1. Finally, since the unique factorization of v over E_2 is a refinement of the unique factorization of v over E_1 it follows that α occurs as a factor in the unique factorization of v over E_2 therefore $\alpha \in alph_{E_2}(L)$.
2. The second statement results immediately using claim 1.
3. The third statement follows immediately from the definitions. □

Lemma 9. *The "nontriviality of* $\mathcal{EA}_k(L)$*".*

1. $k < card(alph(L))$ & $\mathcal{EA}_k(L) \neq \emptyset \Longrightarrow \mathcal{EA}_{k+1}(L) \neq \emptyset$.
2. $\forall k = \overline{d(L), card(alph(L))}$ $(\mathcal{EA}_k(L) \neq \emptyset)$ *and therefore* $\mathcal{EA}(L) = \bigcup_{k=d(L)}^{card(alph(L))} \mathcal{EA}_k(L)$ *is a finite union of nonempty, finite and pairwise disjoint sets.*

Proof. 1. Let $E \in \mathcal{EA}_k(L)$ and $A \in Min\mathcal{CA}(E)$. We have $card(A) = k < card(alph(L))$ and $alph(A) = alph(E) = alph(L)$ therefore $card(A) < card(alph(A))$ and therefore $\exists \alpha \in A$ $(|\alpha| \geq 2)$. Let $\alpha = uv$ with $u, v \neq \lambda$. Let $X = (A - \{\alpha\}) \cup \{u, v\}$. According to lemma 3: X is an elementary set, $A \subseteq X^+$, $card(X) = card(A) + 1$ and $alph_X(A) = X$. So we have: $X \in \mathcal{EA}_{k+1}(A)$ therefore $X \in \mathcal{EA}(A)$ and from $A \in Min\mathcal{CA}(E) \subseteq \mathcal{EA}(E)$, according to lemma 8, it follows $X \in \mathcal{EA}(E)$, but since $E \in \mathcal{EA}(L)$, again according to lemma 8, it follows $X \in \mathcal{EA}_{k+1}(L)$.
2. It results inductively from $\emptyset \neq \mathcal{CA}(L) \subseteq \mathcal{EA}(L)$ and claim 1. □

4. Any language has a finite elementary test-set

The notion of compactness is fundamental in Topology and generally in Mathematics. Let $(X, \mathcal{T})$ be a topological space. An open covering of X is a collection/family of open subsets of X which union is equal to X. The space X is said to be compact iff every open covering of X contains a finite sub-collection/sub-family still covering X.

In a more general context, a compactness-type property works as follows: if usually a certain characteristic of an object Y can be described in an "infinite manner", then, when Y has an appropriate compactness property, this property grants the existence of a finite manner of describing that characteristic of Y.

At the beginning of 1970s, A. Ehrenfeucht stated a famous conjecture namely that each language has a finite test-set. This compactness property of words remained open for more than a decade and was solved affirmatively by M. H. Albert and J. Lawrence, and independently, by V. S. Guba in 1985. This is one of the most important compactness properties in Formal Languages Theory.

We shall prove in this section a compactness property of words, which seems interesting, namely that any language, not necessarily regular, has a finite elementary test-set.

Theorem 4. *Each language has a finite elementary test-set.*

Proof. It is obvious that for a finite language L a finite elementary test-set is L itself.

Let L be an infinite language with $d = d(L)$ and $\sigma = card(alph(L))$. Let $(L_n)_{n\in\mathbb{N}} \subseteq \mathcal{P}_{fin}(\Sigma^*)$ such that $L_n \nearrow L$ *i.e.*

$$L_1 \subseteq L_2 \subseteq \ldots \subseteq L_n \subseteq L_{n+1} \subseteq \ldots \subseteq \bigcup_{n\in\mathbb{N}} L_n = L. \tag{1}$$

Obviously there exists such a sequence, for example $L_n = L \cap \Sigma^{\leq n}$. We shall find $\sigma - d + 1$ ranks, $n_1 \leq n_2 \leq \ldots \leq n_{\sigma-d+1}$, such that

$$\forall i = \overline{1, \sigma - d + 1} \quad \forall n \geq n_i \ (\mathcal{EA}_{\leq d+i-1}(L_n) = \mathcal{EA}_{\leq d+i-1}(L))$$

and for concluding the proof we must notice that each finite part L_f of L which contains $L_{n_{\sigma-d+1}}$ verifies $\mathcal{EA}(L_f) = \mathcal{EA}(L)$.

Let us find these ranks inductively as follows:

– following the proof of Theorem 9 from [10] we find n_1 such that

$$\forall n \geq n_1 \ (\mathcal{EA}_{\leq d}(L_n) = \mathcal{CA}(L_n) = \mathcal{CA}(L) = \mathcal{EA}_{\leq d}(L))$$

– supposing we found n_i with $i < \sigma - d + 1$, let us find n_{i+1}. So we assumed that

$$\mathcal{EA}_{\leq d+i-1}(L_{n_i}) = \ldots = \mathcal{EA}_{\leq d+i-1}(L_n) = \\ = \mathcal{EA}_{\leq d+i-1}(L_{n+1}) = \ldots = \mathcal{EA}_{\leq d+i-1}(L) \ .$$

According to lemma 6 we have:

$$\mathcal{EA}_{d+i}(L_{n_i}) \supseteq \ldots \supseteq \mathcal{EA}_{d+i}(L_n) \supseteq$$
$$\supseteq \mathcal{EA}_{d+i}(L_{n+1}) \supseteq \ldots \supseteq \mathcal{EA}_{d+i}(L)$$

but, according to lemma 4, $card(\mathcal{EA}_{d+i}(L_{n_i})) < \infty$ therefore the decreasing sequence of positive integers $(card(\mathcal{EA}_{d+i}(L_n)))_{n \geq n_i}$ is constant beginning with a certain rank $n_{i+1} \geq n_i$ and therefore

$$\forall n \geq n_{i+1} \ \left(\mathcal{EA}_{d+i}(L_n) = \mathcal{EA}_{d+i}(L_{n_{i+1}}) \supseteq \mathcal{EA}_{d+i}(L)\right). \tag{2}$$

Let's prove now that we have also $\mathcal{EA}_{d+i}(L_{n_{i+1}}) \subseteq \mathcal{EA}_{d+i}(L)$.

Let $E \in \mathcal{EA}_{d+i}(L_{n_{i+1}})$. We shall prove that $E \in \mathcal{EA}_{d+i}(L)$ *i.e.* (i) E elementary set and $card(E) = d + i$; (ii) $L \subseteq E^*$; (iii) $alph_E(L) = E$.

(i) is obvious.

(ii) Let $w \in L$, from (1) it results that $\exists n_w \ \forall n \geq n_w \ \ (w \in L_n)$, from (2) it follows $\forall n \geq n_{i+1} \ \ (E \in \mathcal{EA}_{d+i}(L_n))$ therefore $w \in L_{\max(n_{i+1}, n_w)} \subseteq E^*$ and therefore $L \subseteq E^*$.

(iii) Supposing, to the contrary, that $alph_E(L) = A \subsetneq E$ then from 1 it results $L_{n_{i+1}} \subseteq L \subseteq A^*$ which is in contradiction with $E \in \mathcal{EA}_{d+i}(L_{n_{i+1}})$ because A is a proper subset of E such that $L_{n_{i+1}} \subseteq A^*$.

Consequently we obtained $\forall n \geq n_{i+1} \ \ (\mathcal{EA}_{d+i}(L_n) = \mathcal{EA}_{d+i}(L))$. □

5. Computing the class of elementary alphabets for regular languages

Lemma 10. *If A is an elementary set over Σ and $u\alpha^* v \subset A^*$ where $u, y \in \Sigma^*$; $\alpha \in \Sigma^+$ then $(\exists \beta, \gamma \in \Sigma^*)$ $(\alpha = \beta\gamma, \ \gamma\beta \in A^+)$ and moreover if one of u and v is λ then $\alpha \in A^+$.*

Proof. (i) Case $u = v = \lambda$. Obvious.
(ii) Case $u = \lambda$ and $v \neq \lambda$. Let us assume $\alpha \notin A^+$.
(ii.1) Case $\alpha^2 \in A^+$. Let

$$\begin{cases} v = x_1 \cdot x_2 \cdot \ldots \cdot x_k \\ \alpha v = \alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_r y_1 \cdot y_2 \cdot \ldots \cdot y_l \\ \alpha^2 = \beta_1 \cdot \beta_2 \cdot \ldots \cdot \beta_s \\ k, l, r, s \geq 1; \ \alpha_r, y_1 \neq \lambda \end{cases} \tag{3}$$

be the unique factorizations of v, αv and α^2 over A. We define $B = \{x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_l\}$ and $D = \{\alpha_1, \alpha_2, \ldots, \alpha_r, \beta_1, \beta_2, \ldots, \beta_s\}$. Let us notice that since $\alpha \notin A^+$ the relation $v = x_1 x_2 \ldots x_k = y_1 y_2 \ldots y_l$ is a nontrivial one, therefore, according to the defect effect there exists $B' \subset \Sigma^+$ such that $B \subset (B')^+$ and $card(B') < card(B)$.

Let us also notice that, again since $\alpha \notin A^+$, the relation $\alpha^2 = \alpha_1\alpha_2 \ldots \alpha_r\alpha_1\alpha_2 \ldots \alpha_r = \beta_1\beta_2 \ldots \beta_s$ is nontrivial. Therefore, according to the defect effect there exists $D' \subset \Sigma^+$ such that $D \subset (D')^+$ and $card(D') < card(D)$. We define

$$A' = [A - ((B - \{y_1\}) \cup (D - \{\alpha_r\}) \cup \{\alpha_r y_1\})] \cup (B' \cup D')\,.$$

Obviously we have $A \subset (A')^+$ and $card(A') < card(A)$, which contradicts the fact that A is elementary.

(ii.2) Case $\alpha^2 \notin A^+$. Let

$$\begin{cases} v = x_1 \cdot x_2 \cdot \ldots \cdot x_k \\ \alpha v = \alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_r y_1 \cdot y_2 \cdot \ldots \cdot y_l \\ \alpha^2 v = \beta_1 \cdot \beta_2 \cdot \ldots \cdot \beta_s z_1 \cdot z_2 \cdot \ldots \cdot z_m \\ k, l, m, r, s \geq 1;\ \alpha_r, y_1, \beta_s, z_1 \neq \lambda \end{cases} \tag{4}$$

be the unique factorizations of v, αv and $\alpha^2 v$ over A. As above, considering $B = \{x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_l\}$, there exists $B' \subset \Sigma^+$ such that $B \subset (B')^+$ and $card(B') < card(B)$. Let $v = t_1 t_2 \ldots t_n$ be the factorization of v over B' resulting from the merger of the delimitation points of the factorizations of v and αv from (4). We define $C = \{t_1, t_2, \ldots, t_n, z_1, z_2, \ldots, z_m\}$. If the relation $t_1 t_2 \ldots t_n = z_1 z_2 \ldots z_m$ is a nontrivial one then, according to the defect effect there exists $C' \subset \Sigma^+$ such that $C \subset (C')^+$ and $card(C') < card(C)$. If the relation $t_1 t_2 \ldots t_n = z_1 z_2 \ldots z_m$ is a trivial one then we put $C' = C$. We define $D_1 = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$, $D_2 = \{\beta_1, \beta_2, \ldots, \beta_s\}$ and $D = D_1 \cup D_2$. If the relation $\alpha^2 = \alpha_1\alpha_2 \ldots \alpha_r\alpha_1\alpha_2 \ldots \alpha_r = \beta_1\beta_2 \ldots \beta_s$ is:

- nontrivial, then according to the defect effect there exists $D' \subset \Sigma^+$ such that $D \subset (D')^+$ and $card(D') < card(D)$, and we define

$$A' = [A - ((C - \{z_1\}) \cup (D - \{\alpha_r, \beta_s\}) \cup \{\beta_s z_1, \alpha_r y_1\})] \cup (C' \cup D')\,;$$

- trivial, then if $\beta_s \in (D_1)^+$we define

$$A' = [A - ((C - \{z_1\}) \cup \{\beta_s z_1, \alpha_r y_1\})] \cup (C' \cup \{\alpha_r\})\,;$$

  and if $\alpha_r \in (D_2)^+$ we define

$$A' = [A - ((C - \{z_1\}) \cup \{\beta_s z_1, \alpha_r y_1\})] \cup (C' \cup \{\beta_s\})\,.$$

We have $A \subset (A')^+$ and $card(A') < card(A)$, which contradicts the property of A to be elementary.

The case $v = \lambda$ and $u \neq \lambda$ can be dealt with analogously.

(iii) Case $u, v \neq \lambda$.

Let us notice that there exists $i \geq 1$ and $\beta, \gamma \in \Sigma^*$ such that $\alpha = \beta\gamma$ and $(\gamma\beta)^i \in A^+$. Indeed, let us consider the word $w = u\alpha^h \underbrace{\alpha^{k|\alpha|}}_{=\omega} \alpha^h v$, where $k = \max_{w \in A} |w|$ and h minimal such that $\left|u\alpha^h\right|, \left|\alpha^h v\right| > k$. Within each factor α, from the $k\,|\alpha|$ such factors of ω there exist only $|\alpha|-1$ places (between the $|\alpha|$ letters) where a delimitation

of the words in the factorization over A can occur. By choosing the length of ω we can be sure that there exist at least $|\alpha|$ such delimitation points, therefore at least one occur twice. Obviously the word delimited by two such points has the form $(\gamma\beta)^i$, with $\alpha = \beta\gamma$; $\beta, \gamma \in \Sigma^*$ and $i \geq 1$. Therefore we have $(\gamma\beta)^i \in A^+$. Let $j \geq 1$ be the smallest such that $(\gamma\beta)^j \in A^+$. We will prove now that $j = 1$ by assuming $j > 1$ and contradicting the elementarity of A.

We write u and v as follows: $u = u_1 (\gamma\beta)^p$ and $v = (\gamma\beta)^q v_1$ where $p, q \geq 0$, $\gamma\beta \notin suf(u_1)$, $\gamma\beta \notin pref(v_1)$. If u_1 and v_1 are both λ or only one of them is, then one can proceed as in the cases (i) and (ii) respectively. Let us deal now with the case $u_1, v_1 \neq \lambda$. Let us consider the words:

$$\begin{cases} \omega_1 = u (\gamma\beta)^{kj} v \\ \omega_2 = u (\gamma\beta)^{kj+1} v \end{cases} \tag{5}$$

with k big enough to insure the occurrence of $(\gamma\beta)^j$ as a factor in their unique factorizations over A in the "middle zone" *i.e.* within $(\gamma\beta)^{kj}$ and $(\gamma\beta)^{kj+1}$ respectively.

We can write the following factorization of ω_1 over A:
$\omega_1 = u (\gamma\beta)^{kj} v = u_1 (\gamma\beta)^p (\gamma\beta)^{kj} (\gamma\beta)^q v_1$
as

$$\omega_1 = \underbrace{u_1 (\gamma\beta)^{l_1 j + r_1}}_{=\omega_{11} \in A^+} \cdot \underbrace{(\gamma\beta)^{(k - l_1 - m_1 - 1)j}}_{\in ((\gamma\beta)^j)^+} \cdot \underbrace{(\gamma\beta)^{m_1 j + (j - r_1)} v_1}_{=\omega_{12} \in A^+}$$

which is unique with the following properties:

- $l_1, m_1 \geq 0$; $r_1 \leq j$;
- the last factor of the unique factorization of ω_{11} over A is not $(\gamma\beta)^j$;
- the first factor of the unique factorization of ω_{12} over A is not $(\gamma\beta)^j$;
- Analogously we write the unique factorization of ω_2, $\omega_2 = \omega_{21} \cdot (\gamma\beta)^{mj} \cdot \omega_{22}$, having the following properties:
- $\omega_{21} \in A^+$ and the last factor of its unique factorization over A is not $(\gamma\beta)^j$;
- $\omega_{22} \in A^+$ and the first factor of its unique factorization over A is not $(\gamma\beta)^j$.

Taking into account the difference between ω_1 and ω_2 (see (5)) we can have the following two cases:

$$(a)\ \ \omega_2 = \underbrace{u_1 (\gamma\beta)^{l_2 j + r_2}}_{=\omega_{21} \in A^+} \cdot \underbrace{(\gamma\beta)^{(k - l_2 - m_2 - 1)j}}_{\in ((\gamma\beta)^j)^+} \cdot \underbrace{(\gamma\beta)^{m_2 j + (j - r_2) + 1} v_1}_{=\omega_{22} \in A^+}$$

$$(b)\ \ \omega_2 = \underbrace{u_1 (\gamma\beta)^{l_2 j + r_2 + 1}}_{=\omega_{21} \in A^+} \cdot \underbrace{(\gamma\beta)^{(k - l_2 - m_2 - 1)j}}_{\in ((\gamma\beta)^j)^+} \cdot \underbrace{(\gamma\beta)^{m_2 j + (j - r_2)} v_1}_{=\omega_{22} \in A^+}$$

We shall consider the case (a), the other can be dealt with analogously. Let us consider now the following four words from A^*:

$$\begin{cases} w_{11} = x_1 (\gamma\beta)^{l_1 j + r_1}, \text{ the last factor of the unique factorization} \\ \qquad \text{of } \omega_{11} \text{ over } A \text{ with } x_1 \in suf(u_1), x_1 \neq \lambda \\ w_{12} = (\gamma\beta)^{m_1 j + (j - r_1)} y_1, \text{ the first factor of the unique factorization} \\ \qquad \text{of } \omega_{12} \text{ over } A \text{ with } y_1 \in pref(v_1), y_1 \neq \lambda \\ w_{21} = x_2 (\gamma\beta)^{l_2 j + r_2}, \text{ the last factor of the unique factorization} \\ \qquad \text{of } \omega_{21} \text{ over } A \text{ with } x_2 \in suf(u_1), x_2 \neq \lambda \\ w_{22} = (\gamma\beta)^{m_2 j + (j - r_2) + 1} y_2, \text{ the first factor of the unique factorization} \\ \qquad \text{of } \omega_{22} \text{ over } A \text{ with } y_2 \in pref(v_1), y_2 \neq \lambda \end{cases} \tag{6}$$

We shall prove now that $w_{11} \neq w_{12}$ by assuming $w_{11} = w_{12}$ and contradicting the elementarity of A. Let us use the following simplifying notations $w = w_{11} = w_{12} = x_1 (\gamma\beta)^{l_1 j + r_1} = (\gamma\beta)^{m_1 j + (j - r_1)} y_1$, $l_1 j + r_1 = r$, $m_1 j + (j - r_1) = s$. Therefore we can write:

$$\begin{aligned} x_1 &= (\gamma\beta)^p z (\gamma\beta)^{r'}; \; y_1 = (\gamma\beta)^{s'} z (\gamma\beta)^q; \; w = (\gamma\beta)^p z (\gamma\beta)^q \\ p &= s + s'; \; q = r + r'; \; s', r' \geq 0; \\ z &\neq \lambda; \; \gamma\beta \notin pref(z); \; \gamma\beta \notin suf(z) \end{aligned}$$

We have $w \in A$, $(\gamma\beta)^j \in A$ and $w\gamma\beta w \in A^+$, therefore $(\gamma\beta)^p z (\gamma\beta)^q, (\gamma\beta)^j \in A$ and $(\gamma\beta)^p z (\gamma\beta)^q \gamma\beta (\gamma\beta)^p z (\gamma\beta)^q \in A^+$. Regarding the unique factorization of $w\gamma\beta w$ over A we point out the following:

- within the "middle zone" $(\gamma\beta)^q \gamma\beta (\gamma\beta)^p$ of $w\gamma\beta w$ the only factor from A that can occur is $(\gamma\beta)^j$;
- within the prefix $(\gamma\beta)^p$ and within the suffix $(\gamma\beta)^q$ of $w\gamma\beta w$ the only factor from A that can occur is $(\gamma\beta)^j$;
- the factor w can appear at most once in the first position or in the last one.

Therefore, after eliminating from the unique factorization of $w\gamma\beta w$ over A the factor w and all the factors $(\gamma\beta)^j$ from the beginning and the end, we get one of the following two situations:

- $(\gamma\beta)^{i_1} z (\gamma\beta)^{i_2} \in A^+$;
- $(\gamma\beta)^{i_1} z (\gamma\beta)^{p+q+1} z (\gamma\beta)^{i_2} \in A^+$.

In the second case, if $(\gamma\beta)^j$ appears as factor in the unique factorization over A in the "middle zone" $(\gamma\beta)^{p+q+1}$, then we break the word in two parts and keeping one of them we are again in the first case, otherwise the second case cannot be reduced to the first one. Finally we got the following two cases:

(a) $(\gamma\beta)^{i_1} z (\gamma\beta)^{i_2} \in A^+$ and $(\gamma\beta)^j$ does not appear as first or last factor in the unique factorization of $(\gamma\beta)^{i_1} z (\gamma\beta)^{i_2}$ over A.

(b) $(\gamma\beta)^{i_1} z (\gamma\beta)^{p+q+1} z (\gamma\beta)^{i_2} \in A^+$ and $(\gamma\beta)^j$ does not appear as first or last factor or in the "middle zone" $(\gamma\beta)^{p+q+1}$ in the unique factorization of $(\gamma\beta)^{i_1} z (\gamma\beta)^{p+q+1} z (\gamma\beta)^{i_2}$ over A.

Now let us contradict the elementarity of A in each case.

(a) if $(\gamma\beta)^{i_1} z (\gamma\beta)^{i_2} \in A$ then $\left\{w, (\gamma\beta)^j, (\gamma\beta)^{i_1} z (\gamma\beta)^{i_2}\right\} \subset \{\gamma\beta, z\}^+$ and the elementarity of A is contradicted, therefore the factorization of $(\gamma\beta)^{i_1} z (\gamma\beta)^{i_2}$ over A must be of the form $(\gamma\beta)^{i_1} z (\gamma\beta)^{i_2} = (\gamma\beta)^{i_1} z_1 \cdot z_2 \cdot \ldots \cdot z_{n-1} \cdot z_n (\gamma\beta)^{i_2}$ where $n \geq 2$; $z_1, z_n \neq \lambda$. Let us define

$$A' = \left(A - \{w, (\gamma\beta)^j, (\gamma\beta)^{i_1} z_1, z_n (\gamma\beta)^{i_2}\}\right) \cup \{\gamma\beta, z_1, z_n\}$$

We have $A \subset (A')^+$ and $card(A') < card(A)$, which contradicts the elementarity of A.

(b) If a delimitation point appears in the "middle zone" $(\gamma\beta)^{p+q+1}$ then one can proceed as in the case **a.**, otherwise the factorization of $(\gamma\beta)^{i_1} z (\gamma\beta)^{p+q+1} z (\gamma\beta)^{i_2}$ over A is of the form:

$$(\gamma\beta)^{i_1} z (\gamma\beta)^{p+q+1} z (\gamma\beta)^{i_2} =$$
$$= (\gamma\beta)^{i_1} z_1 \cdot z_2 \cdot \ldots \cdot z_{m-1} \cdot z_m (\gamma\beta)^{p+q+1} t_1 \cdot t_2 \cdot \ldots \cdot t_{n-1} \cdot t_n (\gamma\beta)^{i_2}$$

where $m, n \geq 2$. Let us notice that

$$z_1 z_2 \ldots z_m = t_1 t_2 \ldots t_n = z \tag{7}$$

For $T = \{z_1, z_2, \ldots, z_m, t_1, t_2, \ldots, t_n\}$ we define T' as follows: if the relation (7) is nontrivial then there exists $T' \subset \Sigma^+$ such that $T \subset (T')^+$ and $card(T') < card(T)$, otherwise we put $T = T'$. We define:

$$\begin{aligned} A' \quad = \quad & (A - (\{w, (\gamma\beta)^j, (\gamma\beta)^{i_1} z_1, z_m (\gamma\beta)^{p+q+1} t_1, t_n (\gamma\beta)^{i_2}\} \cup \\ & \cup (T - \{z_1, z_m, t_1, t_n\}))) \cup (T' \cup \{\gamma\beta\}) \end{aligned}$$

and we have $A \subset (A')^+$ and $card(A') < card(A)$, which contradicts the elementarity of A.

Therefore we proved that $w_{11} \neq w_{12}$.

One can prove analogously that $w_{21} \neq w_{22}$.

Let us notice now that:

– if $w_{11} = w_{21}$ then $w_{12} \neq w_{22}$, and

– if $w_{12} = w_{22}$ then $w_{11} \neq w_{21}$.

This is true because of the way the four words are defined (see (6)). Therefore the set $D = \{w_{11}, w_{12}, w_{21}, w_{22}\}$ has at least 3 elements.

Let also notice that if $w_{11} = w_{21}$ then $x_1 = x_2$ and also that if $w_{12} = w_{22}$ then $y_1 = y_2$.

We define:

$$A' = \left(A - \left(D \cup \{(\gamma\beta)^j\}\right)\right) \cup \{x_1, x_2, y_1, y_2\}$$

We have $A \subset (A')^+$ and $card(A') < card(A)$, which contradicts the elementarity of A. □

Lemma 11. *If A is an elementary set over Σ and $w_1, y, w_2 \in \Sigma^*$, $w_1 w_2 \neq \lambda$, $y \neq \lambda$ such that $w_1 y^* w_2 \in A^*$, then no word from A can be a "proper cover" of y, in the factorization of $w_1 y w_2$ over A, i.e. it is not possible to have the following situation: $w_1 = ux$, $w_2 = zv$ with $u, x, z, v \in \Sigma^*$, $xz \neq \lambda$ and $u \cdot xyz \cdot v$ with $xyz \in A$.*

Proof. We shall use the proof by contradiction. More precisely we shall prove the following proposition: **if** Σ is an arbitrary alphabet, $u, x, y, z, v \in \Sigma^*$, A is an elementary set over Σ such that $uxy^*zv \in A^*$, $xz \neq \lambda$ and $\alpha = xyz \in A$, **then** there exists A' such that $A \subset (A')^+$ and $card(A') < card(A)$, which contradicts the elementarity of A.

We shall prove this last proposition by induction on $|y|$:

$$P(n): \left.\begin{array}{l} u, x, y, z, v \in \Sigma^* \\ uxy^*zv \in A^* \\ \alpha = xyz \in A \\ u \cdot \alpha \cdot v \in A^+ \\ y \neq \lambda;\ xz \neq \lambda \\ |y| \leq n \end{array}\right\} \Longrightarrow \exists A' \left(\begin{array}{l} card(A') < card(A) \\ A \subset (A')^+ \end{array}\right)$$

Basis: "$n = 1$". We have $y = a \in \Sigma$ and

$$\left\{\begin{array}{l} \forall k \geq 0 \ \left(w_k = uxa^k zv \in A^*\right) \\ \alpha = xaz \in A \\ u \cdot \alpha \cdot v \in A^+ \end{array}\right.$$

According to lemma 10 we have $a \in A$.

(case 1) $u = v = \lambda$.

(case 1.1) $x \neq \lambda,\ z = \lambda$.

We have $x \in A^+$ and $\alpha = xa \in A$ but also $a \in A$, therefore we contradicted the elementarity of A. Analogously results that it is not possible to have $x = \lambda,\ z \neq \lambda$.

(case 1.2) $x, z \neq \lambda$.

We have $a, xaz \in A$ and $xa^*z \in A^+$. In each of the cases: $x \in a^+$ or $z \in a^+$ or $x, z \in a^+$ we can immediately contradict the elementarity of A, therefore we assume $x, z \notin a^*$. We write:

$$\left\{\begin{array}{l} x = x'a^m;\ m \geq 0;\ a \notin suf(x') \\ z = a^n z';\ n \geq 0;\ a \notin pref(z') \end{array}\right.$$

(case 1.2.1) $xz \in A$.

From $xa^2z \in A^+$ it follows that we can have the following two subcases:

(case 1.2.1.1) $xa^2z = x'a^m a^2 a^n z' = \underbrace{x'_1}_{\in A^*} \cdot \underbrace{x'_2 a^m a}_{\in A} \cdot \underbrace{a a^n z'_1}_{\in A} \cdot \underbrace{z'_2}_{\in A^*}$ with $x'_1 x'_2 = x'$; $x'_2 \neq \lambda$; $a \notin suf(x'_2)$; $z'_1 z'_2 = z'$; $z'_1 \neq \lambda$; $a \notin pref(z'_1)$.

We define

$$A' = \left(A - \left\{xz, xaz, x'_2 a^{m+1}, a^{n+1} z'_1\right\}\right) \cup \{x'_2, z'_1\}$$

We have $A \subset (A')^+$ and $card(A') < card(A)$.

(case 1.2.1.2) $xa^2z = x'a^m a^2 a^n z' = \underbrace{x'_1}_{\in A^*} \cdot \underbrace{x'_2 a^m a^2 a^n z'_1}_{\in A} \cdot \underbrace{z'_2}_{\in A^*}$ with $x'_1 x'_2 = x'$; $x'_2 \neq \lambda$; $a \notin suf(x'_2)$; $z'_1 z'_2 = z'$; $z'_1 \neq \lambda$; $a \notin pref(z'_1)$.

We define

$$A' = \left(A - \left\{xz, xaz, x'_2 a^{m+n+2} z'_1\right\}\right) \cup \{x'_2, z'_1\}$$

We have $A \subset (A')^+$ and $card(A') < card(A)$.

(case 1.2.2) $xz \in A^{\geq 2}$.

Let us write the unique factorization of $xz = x'a^m a^n z'$ over A $xz = x'_1 \cdot x'_2 \cdot ... \cdot x'_k a^m a^n z'_1 \cdot z'_2 \cdot ... \cdot z'_l$. From $xa^2z \in A^+$ it follows that we can have the following two possibilities for the unique factorization of $xa^2z = x'a^m a^2 a^n z'$ over A :

(case 1.2.2.1) $xa^2z = x''_1 \cdot x''_2 \cdot ... \cdot x''_i a^m a \cdot a a^n z''_1 \cdot z''_2 \cdot ... \cdot z''_j$.

We have:

$$x' = x'_1 x'_2 ... x'_k = x''_1 x''_2 ... x''_i \tag{8}$$

and

$$z' = z'_1 z'_2 ... z'_l = z''_1 z''_2 ... z''_j \tag{9}$$

We define: $X' = \{x'_1, x'_2, ..., x'_k\}$, $X'' = \{x''_1, x''_2, ..., x''_i\}$ and $Z' = \{z'_1, z'_2, ..., z'_l\}$, $Z'' = \left\{z''_1, z''_2, ..., z''_j\right\}$ and $A'' = \{xaz, x'_k a^{m+n} z'_1, x''_i a^{m+1}, a^{n+1} z''_1\}$. Let us notice that at least one of the relations (8) and (9) is nontrivial. Indeed, if we assume both of them are trivial, then we can have the following four cases:

(a) $x''_i \in (X')^+$ and $z''_1 \in (Z')^+$, we define $A' = (A - A'') \cup \{x'_k, z'_1\}$

(b) $x''_i \in (X')^+$ and $z'_1 \in (Z'')^+$, we define $A' = (A - A'') \cup \{x'_k, z''_1\}$

(c) $x'_k \in (X'')^+$ and $z''_1 \in (Z')^+$, we define $A' = (A - A'') \cup \{x''_i, z'_1\}$

(d) $x'_k \in (X'')^+$ and $z'_1 \in (Z'')^+$, we define $A' = (A - A'') \cup \{x''_i, z''_1\}$

We have $A \subset (A')^+$ and $card(A') < card(A)$.

Therefore at least one of the relations (8) and (9) is nontrivial.

If (8) is nontrivial, then there exists B such that $X' \cup X'' \subset B^+$ and $card(B) < card(X' \cup X'')$ and we define

$$A' = (A - ((X' - \{x'_k\}) \cup (X'' - \{x''_i\}) \cup A'')) \cup (B \cup \{z'_1, z''_1\}) .$$

If (9) is nontrivial, then there exists C such that $Z' \cup Z'' \subset C^+$ and $card(C) < card(Z' \cup Z'')$ and we define

$$A' = (A - ((Z' - \{z'_1\}) \cup (Z'' - \{z''_1\}) \cup A'')) \cup (C \cup \{x'_k, x''_i\}) .$$

In both cases we have $A \subset (A')^+$ and $card(A') < card(A)$.

(case 1.2.2.2) $xa^2z = x''_1 \cdot x''_2 \cdot ... \cdot x''_i a^{m+n+2} z''_1 \cdot z''_2 \cdot ... \cdot z''_j$.

We have the same two relations (8) and (9) as above. We define the same X', X'' and Z', Z'', but this time $A'' = \{xaz, x'_k a^{m+n} z'_1, x''_i a^{m+n+2} z''_1\}$.

If both relations (8) and (9) are trivial, then we can have the same four cases as above and we define A' in the same way but with the new A''. Again we have $A \subset (A')^+$ and $card(A') < card(A)$.

If (8) is nontrivial and (9) is trivial, then there exists B such that $X' \cup X'' \subset B^+$ and $card(B) < card(X' \cup X'')$ and we have two subcases resulting from (9) trivial:

(e) $z_1'' \in (Z')^+$ and we define

$$A' = (A - ((X' - \{x_k'\}) \cup (X'' - \{x_i''\}) \cup A'')) \cup (B \cup \{z_1'\}) .$$

(f) $z_1' \in (Z'')^+$ and we define

$$A' = (A - ((X' - \{x_k'\}) \cup (X'' - \{x_i''\}) \cup A'')) \cup (B \cup \{z_1''\}) .$$

If (9) is nontrivial and (8) is trivial, then there exists C such that $Z' \cup Z'' \subset C^+$ and $card(C) < card(Z' \cup Z'')$ and we have two subcases resulting from (8) trivial:

(g) $x_i'' \in (X')^+$ and we define

$$A' = (A - ((Z' - \{z_1'\}) \cup (Z'' - \{z_1''\}) \cup A'')) \cup (C \cup \{x_k'\}) .$$

(h) $x_k' \in (X'')^+$ and we define

$$A' = (A - ((Z' - \{z_1'\}) \cup (Z'' - \{z_1''\}) \cup A'')) \cup (C \cup \{x_i''\}) .$$

If both (8) and (9) are nontrivial, then there exist B and C such that $X' \cup X'' \subset B^+$, $Z' \cup Z'' \subset C^+$ and $card(B) < card(X' \cup X'')$, $card(C) < card(Z' \cup Z'')$ and we define

$$\begin{aligned} A' \quad = \quad & (A - ((X' - \{x_k'\}) \cup (X'' - \{x_i''\}) \cup (Z' - \{z_1'\}) \cup \\ & \cup (Z'' - \{z_1''\}) \cup A'')) \cup (B \cup C) \end{aligned}$$

In all the cases we have $A \subset (A')^+$ and $card(A') < card(A)$.

(case 2) $u \neq \lambda,\ v = \lambda$.

The simplifying idea is to use $u \cdot xaz \in A^+$ to reduce (case 2) to (case 1) by defining/considering an elementary set $\overline{A}$ such that $A \subset (\overline{A})^+$, $card(\overline{A}) \leq card(A)$ and

$$\begin{cases} \forall k \leq 2 \ \left(w_k = xa^k z \in \overline{A}^*\right) \\ \alpha = xaz \in \overline{A} \end{cases}$$

Let us notice that it suffices to have only $\forall k \leq 2 \ \left(w_k = xa^k z \in \overline{A}^*\right)$ and not $xa^* z \in \overline{A}^*$ because in the proof of (case 1) we used only w_0, w_1 and w_2. After defining $\overline{A}$, in order to complete the proof, one can repeat the proof of (case 1) with $\overline{A}$ instead of A and one will get an A' such that $A \subset (\overline{A})^+ \subset (A')^+$ and $card(A') < card(\overline{A}) \leq card(A)$.

Let us now define $\overline{A}$. We have $u \in A^+$, therefore $u = u_1' \cdot u_2' \cdot \ldots \cdot u_i'$ with $i \geq 1$.

(case 2.1) If $u \cdot xz \in A^+$ and $u \cdot xa^2 z \in A^+$ then $\overline{A} = A$.

(case 2.2) If $u \cdot xz \in A^+$, but the unique factorization of $uxa^2 z$ over A is not of the form $u \cdot xa^2 z$ then

$$uxa^2 z = \underbrace{u_1'' \cdot u_2'' \cdot \ldots \cdot u_j''}_{=u} \underbrace{\omega_1 \cdot \omega_2 \cdot \ldots \cdot \omega_k}_{=xa^2 z} , \ u_j'' \omega_1 \in A$$

therefore we have the relation

$$u = u'_1 u'_2 \dots u'_i = u''_1 u''_2 \dots u''_j \tag{10}$$

We define $U' = \{u'_1, u'_2, ..., u'_i\}$, $U'' = \{u''_1, u''_2, ..., u''_j\}$.

If the relation (10) is trivial, then we can have two subcases:

– $u''_j \in (U')^+$ and we define

$$\overline{A} = (A - \{u''_j \omega_1\}) \cup \{\omega_1\}$$

– $u'_i \in (U'')^+$ and we define

$$\overline{A} = (A - \{u''_j \omega_1, u'_i\}) \cup \{\omega_1, u''_j\}$$

In both subcases we have $A \subset \left(\overline{A}\right)^+$ and $card(\overline{A}) \leq card(A)$. Let us notice that, according to lemma 2, $\overline{A}$ is elementary.

If the relation (10) is nontrivial, then there exists B such that $U' \cup U'' \subset B^+$ and $card(B) < card(U' \cup U'')$ and we define

$$\overline{A} = \left(A - \left(U' \cup \left(U'' - \{u''_j\}\right) \cup \{u''_j \omega_1\}\right)\right) \cup (B \cup \{\omega_1\})$$

and again we have $A \subset (\overline{A})^+$ and $card(\overline{A}) \leq card(A)$ and also $\overline{A}$ is elementary.

(case 2.3) If $u \cdot xa^2 z \in A^+$, but the unique factorization of uxz over A is not of the form $u \cdot xz$ then

$$uxz = \underbrace{u'''_1 \cdot u'''_2 \cdot \ldots \cdot u'''_l \omega'_1}_{=u} \cdot \underbrace{\omega'_2 \cdot \ldots \cdot \omega'_m}_{=xz}, \; u'''_l \omega'_1 \in A$$

therefore we have the relation

$$u = u'_1 u'_2 \dots u'_i = u'''_1 u'''_2 \dots u'''_l \tag{11}$$

We define $U' = \{u'_1, u'_2, ..., u'_i\}$, $U''' = \{u'''_1, u'''_2, \dots, u'''_l\}$.

If the relation (11) is trivial, then we can have two subcases:

– $u'''_l \in (U')^+$ and we define

$$\overline{A} = (A - \{u'''_l \omega'_1\}) \cup \{\omega'_1\}$$

– $u'_i \in (U''')^+$ and we define

$$\overline{A} = (A - \{u'''_l \omega'_1, u'_i\}) \cup \{\omega'_1, u'''_l\}$$

In both subcases we have $A \subset \left(\overline{A}\right)^+$ and $card(\overline{A}) \leq card(A)$. Let us notice that, according to lemma 2, $\overline{A}$ is elementary.

If the relation (11) is nontrivial, then there exists C such that $U' \cup U''' \subset C^+$ and $card(C) < card(U' \cup U''')$ and we define

$$\overline{A} = (A - (U' \cup (U''' - \{u'''_l\}) \cup \{u'''_l \omega'_1\})) \cup (C \cup \{\omega'_1\})$$

and again we have $A \subset (\overline{A})^+$ and $card(\overline{A}) \leq card(A)$ and also $\overline{A}$ is elementary.

(case 2.4) If the unique factorizations of both uxz and uxa^2z over A are not of the form $u \cdot xz$ and $u \cdot xa^2z$ respectively, then we can write:

$$uxz = \underbrace{u''_1 \cdot u''_2 \cdot \ldots \cdot u''_j}_{=u} \underbrace{\omega_1 \cdot \omega_2 \cdot \ldots \cdot \omega_k}_{=xz}, \; u''_j\omega_1 \in A$$

and we have the relation:

$$u = u'_1u'_2 \ldots u'_i = u''_1u''_2 \ldots u''_j \tag{12}$$

As above, we define $U' = \{u'_1, u'_2, ..., u'_i\}$, $U'' = \{u''_1, u''_2, ..., u''_j\}$.

If the relation (12) is trivial, then we can have two subcases:

– $u''_j \in (U')^+$ and we define

$$C = (A - \{u''_j\omega_1\}) \cup \{\omega_1\}$$

– $u'_i \in (U'')^+$ and we define

$$C = (A - \{u''_j\omega_1, u'_i\}) \cup \{\omega_1, u''_j\}$$

In both subcases we have $A \subset (C)^+$ and $card(C) \leq card(A)$. Let us notice that, according to lemma 2, C is elementary.

If the relation (12) is nontrivial, then there exists B such that $U' \cup U'' \subset B^+$ and $card(B) < card(U' \cup U'')$ and we define

$$C = \left(A - \left(U' \cup \left(U'' - \{u''_j\}\right) \cup \{u''_j\omega_1\}\right)\right) \cup (B \cup \{\omega_1\})$$

and again we have $A \subset (C)^+$ and $card(C) \leq card(A)$ and also C elementary.

(case 2.4.1) If the unique factorization of uxa^2z over C is of the form $u \cdot xa^2z$, then $\overline{A} = C$.

(case 2.4.2) If the unique factorization of uxa^2z over C is not of the form $u \cdot xa^2z$, then we can write:

$$\begin{aligned} u &= c'_1 \cdot c'_2 \cdot \ldots \cdot c'_p \\ uxa^2z &= \underbrace{c''_1 \cdot c''_2 \cdot \ldots \cdot c''_q}_{=u} \underbrace{\omega''_1 \cdot \omega''_2 \cdot \ldots \cdot \omega''_m}_{=xa^2z}, \; c''_q\omega''_1 \in A \end{aligned}$$

and we have the relation:

$$u = c'_1c'_2 \ldots c'_p = c''_1c''_2 \ldots c''_q \tag{13}$$

we define $C' = \{c'_1, c'_2, \ldots, c'_p\}$, $C'' = \{c''_1, c''_2, \ldots, c''_q\}$.

If the relation (13) is trivial, then we can have two subcases:

– $c''_q \in (C')^+$ and we define

$$\overline{A} = (C - \{c''_q\omega''_1\}) \cup \{\omega''_1\}$$

– $c'_p \in (C'')^+$ and we define

$$\overline{A} = (C - \{c''_q\omega''_1, c'_p\}) \cup \{\omega''_1, c''_q\}$$

In both subcases we have $A \subset (C)^+ \subset \left(\overline{A}\right)^+$ and $card(\overline{A}) \leq card(C) \leq card(A)$, and also $\overline{A}$ elementary.

If the relation (13) is nontrivial, then there exists B such that $C' \cup C'' \subset B^+$ and $card(B) < card(C' \cup C'')$ and we define

$$\overline{A} = \left(A - \left(C' \cup \left(C'' - \{c''_q\}\right) \cup \{c''_q\omega''_1\}\right)\right) \cup (B \cup \{\omega''_1\})$$

and again we have $A \subset (C)^+ \subset \left(\overline{A}\right)^+$ and $card(\overline{A}) \leq card(C) \leq card(A)$ and also $\overline{A}$ elementary.

(case 3) $u = \lambda,\ v \neq \lambda$. Obviously his case can be dealt with analogously.

(case 4) $u, v \neq \lambda$. Now it is clear how one can deal with this case. One will proceed first with "the isolation of u" as in the (case 2) and then with "the isolation of v" as in the (case 3).

Inductive step: "$n \longrightarrow n+1$"

According to lemma 10: $\exists \beta, \gamma \in \Sigma^*\ \ (y = \beta\gamma,\ \gamma\beta \in A^+)$ and moreover if one of $u = \lambda$ or $v = \lambda$ then $\alpha \in A^+$.

(case 1) $alph_A(\gamma\beta) \nsubseteq \Sigma$ i.e. if within the unique factorization of $\gamma\beta$ over A there exists a factor, let's say ω_0, with length at least 2, then we consider $\Delta = \Sigma \cup \{\$\}$ where \$ is a new symbol, and the mapping $\varphi : A^* \longrightarrow \Delta^*$, defined as follows:

$$\varphi(\omega) = \begin{cases} \omega & \text{for } \omega \neq \omega_0 \\ \$ & \text{for } \omega = \omega_0 \end{cases}$$

Let us notice that $\Omega = \varphi(A) = (A - \{\omega_0\}) \cup \{\$\}$, therefore a factorization of w over A can be transformed into a factorization of $\varphi(w)$ over Ω by simply replacing the factors ω_0 with \$ and vice versa. Using φ we "move" the problem from Σ^* to Δ^*:

$$\begin{cases} \varphi(u), \varphi(x), \varphi(y), \varphi(z), \varphi(v) \in \Delta^* \\ \forall k \geq 0\ \ \left(\varphi(u)\varphi(x)\varphi(y)^k\varphi(z)\varphi(v) \in \Omega^*\right) \\ \varphi(\alpha) = \varphi(x)\varphi(y)\varphi(z) \in \varphi(A) = \Omega \end{cases}$$

Let us notice that $|\varphi(y)| \leq n$ therefore, according to the inductive assumption, there exists $\Omega' \subset \Delta^*$ such that $\Omega \subset (\Omega')^+$ and $card(\Omega') < card(\Omega)$. Obviously $\$ \in \Omega'$. Let us define now $A' = (\Omega' - \{\$\}) \cup \{\omega_0\}$. We have $A = (\Omega - \{\$\}) \cup \{\omega_0\} \subset ((\Omega' - \{\$\}) \cup \{\omega_0\})^* = (A')^+$ and $card(A') < card(A)$.

(case 2) $alph_A(\gamma\beta) \subseteq \Sigma$. Let us denote $B = alph_A(\gamma\beta) \subseteq A \subseteq \Sigma$ and let l be big enough to insure the unique factorization of $w_l = ux\,(\beta\gamma)^l\, zv$ over A is of the form

$$w_l = \underbrace{ux\,(\beta\gamma)^i\,\beta}_{=\omega_1 \in A^*} \cdot \underbrace{(\gamma\beta)^j}_{\in A^*} \cdot \underbrace{\gamma\,(\beta\gamma)^k\,zv}_{=\omega_2 \in A^*}$$

Let us notice that since $xz \neq \lambda$ it follows that $\omega_1\omega_2 \neq \lambda$.

(case 2.1) $\omega_1 \neq \lambda$, $\omega_2 = \lambda$. We have $\alpha \in A^+$ and $x \neq \lambda$. Let δ be the right most factor of the unique factorization of ω_1 over A that is not from B.

(case 2.1.1) $\delta = \delta_1\delta_2$ with $\delta_1 \in suf(x)$, $\delta_1 \neq \lambda$ and $\delta_2 \in pref((\beta\gamma)^i\,\beta)$ therefore $\delta_2 \in B$. Then, since $x\alpha \in A$, we define $A' = (A - \{\delta, x\alpha\}) \cup \{\delta_1\}$ and we have $A \subset (A')^+$ and $card(A') < card(A)$.

(case 2.1.2) $\delta = \delta_1 x \delta_2$ with $\delta_1 \in suf(u)$, $\delta_1 \neq \lambda$ and $\delta_2 \in pref((\beta\gamma)^i\,\beta)$ therefore $\delta_2 \in B$. We have the following unique factorizations over A:

$$u = u'_1 \cdot u'_2 \cdot \ldots \cdot u'_i$$

and

$$ux\delta_2 = u''_1 \cdot u''_2 \cdot \ldots \cdot u''_j \cdot \delta_1 x \delta_2$$

therefore we have the following relation

$$u = u'_1 u'_2 \ldots u'_i = u''_1 u''_2 \ldots u''_j \delta_1 \tag{14}$$

We define $U' = \{u'_1, u'_2, ..., u'_i\}$, $U'' = \{u''_1, u''_2, ..., u''_j, \delta_1\}$.

If the relation (14) is trivial, then we can have two subcases:

- $\delta_1 \in (U')^+$ and we define

$$A' = A - \{\delta, x\alpha\}) \cup \{x\}$$

- $u'_i \in (U'')^+$ and we define

$$A' = (A - \{\delta, x\alpha, u'_i\}) \cup \{x, \delta_1\}$$

In both subcases we have $A \subset (A')^+$ and $card(A') < card(A)$.

If the relation (14) is nontrivial, then there exists B such that $U' \cup U'' \subset B^+$ and $card(B) < card(U' \cup U'')$ and we define

$$A' = (A - (U' \cup (U'' - \{\delta_1\}) \cup \{\delta, x\alpha\})) \cup (B \cup \{x\})$$

and again we have $A \subset (A')^+$ and $card(A') < card(A)$.

(case 2.1.3) $\delta = \delta_1\delta_2$ with $\delta_1 \in suf(u)$, $\delta_1 \neq \lambda$ and $\delta_2 \in pref(x)$.

If $\delta_2 \in B^*$ then we define $A' = A - \{x\alpha\}$ and we have $A \subset (A')^+$ and $card(A') < card(A)$.

Otherwise w have the following unique factorizations over A:

$$u = u'_1 \cdot u'_2 \cdot \ldots \cdot u'_i$$

and

$$u\delta_2 = u_1'' \cdot u_2'' \cdot \ldots \cdot u_j'' \cdot \delta_1\delta_2$$

therefore we have the following relation

$$u = u_1'u_2' \ldots u_i' = u_1''u_2'' \ldots u_j''\delta_1 \tag{15}$$

We define $U' = \{u_1', u_2', ..., u_i'\}$, $U'' = \{u_1'', u_2'', ..., u_j'', \delta_1\}$.

If the relation (15) is trivial, then we can have two subcases:

– $\delta_1 \in (U')^+$ and we define

$$A' = A - \{\delta, x\alpha\}) \cup \{\delta_2\}$$

– $u_i' \in (U'')^+$ and we define

$$A' = (A - \{\delta, x\alpha, u_i'\}) \cup \{\delta_1, \delta_2\}$$

In both subcases we have $A \subset (A')^+$ and $card(A') < card(A)$.

If the relation (15) is nontrivial, then there exists B such that $U' \cup U'' \subset B^+$ and $card(B) < card(U' \cup U'')$ and we define

$$A' = (A - (U' \cup (U'' - \{\delta_1\}) \cup \{\delta, x\alpha\})) \cup (B \cup \{\delta_2\})$$

and again we have $A \subset (A')^+$ and $card(A') < card(A)$..

(case 2.2) $\omega_1 = \lambda$, $\omega_2 \neq \lambda$. This case can obviously be treated just like (case 2.1).

(case 2.3) $\omega_1, \omega_2 \neq \lambda$. Now it is clear how one can deal with this case. One will deal first with ω_1 as in the (case 2.1) and then with ω_2 as in the (case 2.2). □

Theorem 5. *If L is a regular language, $M = (Q, \Sigma, \delta, q_0, F)$ a DFA recognizing L and $A \in \mathcal{EA}(L)$ then $\forall \alpha \in A$ $(|\alpha| \leq card(Q))$.*

Proof. We shall use the proof by contradiction. Let us assume that the conclusion is not true and let α be a word from A with $|\alpha| > card(Q)$ and $w = u\alpha v$ be a word in L having α as a factor in its unique factorization over A. Let $p_0 = \delta^*(q_0, u)$, $p_1 = \delta(p_0, \alpha(1))$, $p_2 = \delta(p_1, \alpha(1)\alpha(2))$,..., $p_{|\alpha|-1} = \delta(p_{|\alpha|-2}, \alpha(1)\alpha(2)\ldots\alpha(|\alpha|-1))$, $p_{|\alpha|} = \delta(p_{|\alpha|-1}, \alpha)$ be the sequence of states entered when reading α starting from $p_0 = \delta^*(q_0, u)$. Let (i, j) be the unique leftmost pair of positions for which we have

$$\begin{cases} 0 \leq i \leq j \leq |\alpha| \\ p_i = p_j \\ card(\{p_k \mid i < k \leq j\}) = j - i \end{cases}$$

This pair obviously exists because $|\alpha| > card(Q)$. Let us consider the following words: $x, y, z \in \Sigma^*$ defined by $p_i = \delta^*(p_0, x)$, $p_j = p_i = \delta^*(p_i, y)$, $p_{|\alpha|} = \delta(p_j, z)$ and $\alpha = xyz$. Let us notice that $xz \neq \lambda$ because $|y| \leq card(Q) < |\alpha|$. We have the following situation:

$$\begin{cases} uxy^*zv \subseteq L \subseteq A^* \\ \alpha = xyz \in A \\ x, y, z \in \Sigma^*,\ xz \neq \lambda,\ y \neq \lambda \\ u \cdot \alpha \cdot v \in A^+ \end{cases}$$

as in the figure below:

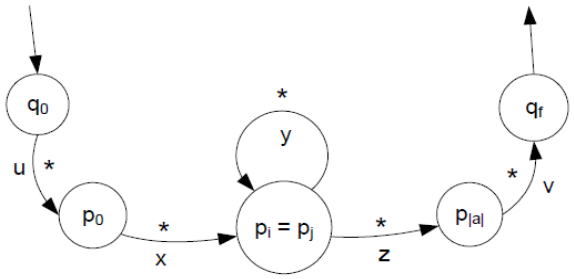

Fig. 1.

Now, using lemma 11, it follows that there exists A' such that $A \subset (A')^+$ and $card(A') < card(A)$, which contradicts the elementarity of A. □

Remark 2. Let us point out that if we relax the assumption $A \in \mathcal{EA}(L)$ in the theorem above by replacing it with: A code, $L \subseteq A^*$ and $alph_A(L) = A$, then the conclusion is not true any more. Indeed, let us consider the following simple example: $L = a^*b$ and $A = \{b, ab, ab^2, ab^3, a^4\}$.Obviously A is a code, $L \subseteq A^*$, $alph_A(L) = A$, the number of states of the minimal DFA recognizing L is 3 and A contains a word of length 4.

Moreover, for each $n \geq 2$ there exists a code A_n such that

$$L \subseteq A_n^*,\ alph_{A_n}(L) = A_n \text{ and } card(A_n) = n$$

Indeed $A_n = \{b, ab, a^2b, ..., a^{n-2}b, a^{n-1}\}$.

Remark 3. If L is a regular language and $M(L)$ its minimal DFA, then:

1. The lengths of all the words in its elementary alphabets are smaller than or equal to the number of states of $M(L)$.
2. The inequality at no. 1 above can be both strict and equality *i.e.*

   $$\exists A \in \mathcal{EA}(L)\ \exists w \in A\ (|w| = \text{ the number of states of } M(L))$$

   Indeed, for $L = (a^n)^*$ the minimal DFA has n states and $A = \{a^n\} \in \mathcal{EA}(L)$.
3. If its minimal DFA has a "black hole state" (*i.e.* a state from each no final state can be reached; it is obvious that if such a state exists then it is unique) then the lengths of all the words in its elementary alphabets are smaller than or equal to the number of states of its minimal DFA minus 1.
4. The inequality at no. 3 above can be both strict and equality *i.e.*

   $$\exists A \in \mathcal{EA}(L)\ \exists w \in A\ (|w| = \text{ the number of states of } M(L) \text{ minus } 1)$$

   Indeed, for $L = ((ab)^n)^*$ the minimal DFA has $2n+1$ states and $A = \{(ab)^n\} \in \mathcal{EA}(L)$.

Theorem 6. *For each regular language L:*

1. *$d(L)$ can be effectively calculated;*
2. *$\mathcal{EA}(L)$, and implicitly $\mathcal{CA}(L)$, can be effectively found;*
3. *given a finite part L_f of L it is decidable whether L_f is an elementary test-set, and implicitly a finite combinatorial test-set, for L.*

Proof. Claim 1. is an immediate consequence of claim 2.

2. According to theorem 5, if $M = (Q, \Sigma, \delta, q_0, F)$ a DFA recognizing L then $\forall A \in \mathcal{EA}(L)\ \forall \alpha \in A\ (|\alpha| < card(Q))$, therefore $\forall A \in \mathcal{EA}(L)\ (A \subset \Sigma^{\leq card(Q)})$ and therefore one can try for example all the elements A of $\wp_{\leq card(\Sigma)}(\Sigma^{\leq card(Q)})$ to see if A is elementary set, if $L \subseteq A^*$ and if these conditions are true, finally to see if $alph_A(L) = A$.

3. According to claim 2. one can find $\mathcal{EA}(L)$ and $\mathcal{EA}(L_f)$ and check whether they are equal or not. □

We mention that it is not in the scope of this paper to provide an efficient algorithm for computing $\mathcal{EA}(L)$.

We shall revert to this subject in the near future.

References

[1] Berstel J., Karhumäki J., *Combinatorics on Words – A Tutorial*, Bull. EATCS, **no. 79**, pp. 178–228, 2003.

[2] Choffrut C., Karhumäki J., *Combinatorics of words*, in Rozenberg G., Salomaa A. (eds.), *Handbook of Formal Languages*, Vol. 1, Springer-Verlag, 1997.

[3] Ehrenfeucht A., Rozenberg G., *Elementary Homomorphisms and a Solution to the DoL Sequence Equivalence Problem*, Theoretical Computer Science, **7**, pp. 169–183, 1978.

[4] Harju T., Karhumäki J., *Morphisms*, in G. Rozenberg, A. Salomaa (eds.), *Handbook of Formal Languages*, Vol. 1, Springer-Verlag, 1997.

[5] Harju T., Karhumäki J., *On the Defect Theorem and Simplifiability*, *Semigroup Forum*, **33**, pp. 199–217, 1986.

[6] Néraud J., *On the Deficit of a Finite Set of Words*, Semigroup Forum, **41**, pp. 1–21, 1990.

[7] Néraud J., *On the Rank of the Subsets of a Free Monoid*, Theoretical Computer Science, **99**, pp. 231–241, 1992.

[8] Néraud J., *Deciding Whether a Finite Set of Words Has Rank at Most Two*, Theoretical Computer Science, **112**, pp. 311–337, 1993.

[9] Néraud J., *Elementariness of a Finite Set of Words is co-NP-complete*, RAIRO Theoretical Informatics and Applications, **24**(5), pp. 459–470, 1990.

[10] Voinescu D. C., *On the Combinatorial Alphabets of a Language*, Journal of Automata, Languages and Combinatorics, **7**3, pp. 377–394, 2002.

[11] Yu S., *Regular Languages*, in G. Rozenberg, A. Salomaa (eds.), *Handbook of Formal Languages*, Vol. 1, Springer-Verlag, 1997.