**Theoretical Informatics and Applications**
Theoret. Informatics Appl. **35** (2001) 453–475

# LINEAR SIZE TEST SETS FOR CERTAIN COMMUTATIVE LANGUAGES

Štěpán Holub[1] and Juha Kortelainen[1]

**Abstract**. We prove that for each positive integer $n$, the finite commutative language $E_n = c(a_1 a_2 \cdots a_n)$ possesses a test set of size at most $5n$. Moreover, it is shown that each test set for $E_n$ has at least $n - 1$ elements. The result is then generalized to commutative languages $L$ containing a word $w$ such that (i) $\mathrm{alph}(w) = \mathrm{alph}(L)$; and (ii) each symbol $a \in \mathrm{alph}(L)$ occurs at least twice in $w$ if it occurs at least twice in some word of $L$: each such $L$ possesses a test set of size $11n$, where $n = \mathrm{Card}(\mathrm{alph}(L))$. The considerations rest on the analysis of some basic types of word equations.

**Mathematics Subject Classification.** 68R15.

## Introduction

In this note we shall study the test sets of some commutative languages. By a test set for a language $L$ we mean any subset $L'$ of $L$ such that if any two morphisms agree on $L'$, then they agree also on $L$. By the famous Ehrenfeucht's Conjecture, each language has a finite test set. Since the proof of the conjecture ([2], see also [16]), the size of test sets for different types of languages has been under active investigation. The size of the test set with respect to the considered language can be measured in different ways. We shall measure it by the cardinality of the language alphabet. The choice is understandable: the structure of a commutative language is generally not determined by an automaton or a grammar.

The test sets for regular and context-free languages can be effectively determined and this subject has been studied in several papers. A survey of the results as well as a comprehensive list of references can be found in [4] and [10]. For context-sensitive languages, finite test sets can not in general be effectively constructed. The articles [1, 7–9, 12] and [14] contain results on restricted types of

[1] Turku Centre for Computer Science & Charles University, Prague, Czech Republic
Department of Information Processing Science, University of Oulu, P.O. Box 3000, 90014 Oulun Yliopisto, Finland.

context-sensitive languages. By [5], every language over a two-letter alphabet has a test set of size at most three. Test set research on commutative languages was started in [3] and the work was continued in [9] where it is shown that each commutative language over an alphabet of $n$ symbols possesses a test set of size $O(n^2)$.

Finite sets have an important role as a source of (counter)examples in test set considerations. Let $\Sigma$ be an alphabet of $n$ symbols. In [13] it is proved that there exists a finite (and thus regular) language over $\Sigma$ whose test set size is at least $\Omega(n^4)$. Furthermore, in [9] the existence of a finite commutative language $L \subseteq \Sigma^*$ with a test set of size at least $\Omega(n^2)$ is verified. Our central research subject can in terms of word equations be expressed as follows. For each positive integer $n$, determine a smallest possible set $\mathcal{T}_n \subseteq \mathcal{S}_n$ such that

$$x_{\rho(1)}x_{\rho(2)}\cdots x_{\rho(n)} = y_{\rho(1)}y_{\rho(2)}\cdots y_{\rho(n)} \quad \text{for each } \rho \in \mathcal{T}_n \tag{$*$}$$

implies

$$x_{\sigma(1)}x_{\sigma(2)}\cdots x_{\sigma(n)} = y_{\sigma(1)}y_{\sigma(2)}\cdots y_{\sigma(n)} \quad \text{for each } \sigma \in \mathcal{S}_n. \tag{$**$}$$

Above $\mathcal{S}_n$ is the set of all permutations of $1, 2, \ldots, n$ and $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$ are words. An equivalent expression of the same task is to find a test set for the language

$$E_n = \{a_{\sigma(1)}a_{\sigma(2)}\cdots a_{\sigma(n)} \mid \sigma \in \mathcal{S}_n\}\cdot$$

The paper at hand has the following contents. In the first section some basic results and concepts on formal language theory and combinatorics on words are given. In Section 2 a simple sufficient condition implying commutation of a sequence of words is derived. In the third section the new concepts of permutation, weak permutation, conjugacy and shuffle property are introduced. Their power and interrelations are studied up to certain extent. In Section 4 using weak permutation, conjugacy and shuffle we formulate two sufficient conditions that imply permutation of words. Applying the results obtained in the previous sections, a linear size test set for the language $c(a_1a_2\cdots a_n)$ is constructed in the fifth section. In other words, we build a set $\mathcal{T}_n$ of size $O(n)$ such that $(*)$ implies $(**)$. In the seventh section a linear size test set is found for languages containing a word $w$, such that the alphabet of $w$ is equal to the alphabet of $L$ and each symbol $a$ occurs at least twice in $w$ if it occurs at least twice in some word of $L$. Such are for example so called CLIP-languages, *i.e.*, commutative languages whose Parikh-map is a linear set. The final section contains some concluding remarks and further topics of research.

## 1. Preliminaries

We assume that the reader is familiar with the basic notations and results of formal language theory and word combinatorics as presented in [11] and [15].

Let $\Sigma$ be a (finite) alphabet. As usual, $\Sigma^*$ ($\Sigma^+$, resp.) is the free monoid (free semigroup, resp.) generated by $\Sigma$. The elements of $\Sigma^*$ are called *words.* Let $w \in \Sigma^*$. For each $a \in \Sigma$, $|w|_a$ is the number of occurrences of the symbol $a$ in $w$. The *length* of $w$, denoted by $|w|$, is the total number of symbols in $w$: $|w| = \sum_{a\in\Sigma} |w|_a \cdot$ Define the *powers* of $w$ inductively as follows: $w^0 = \epsilon$, $w^{k+1} = w^k \cdot w$ ($k \in \mathbb{N}$). Let $w^* = \{w^k \mid k \in \mathbb{N}\}$ and $w^+ = w^* \setminus \{\epsilon\}$. Let $b_1, b_2, \dots b_m \in \Sigma$ and $w = b_1 b_2 \dots b_m$. Denote

$$c(w) = \{b_{\sigma(1)} b_{\sigma(2)} \dots b_{\sigma(m)} \mid \sigma \in \mathcal{S}_m\} = \{u \in \Sigma^* \mid \ |u|_a = |w|_a \text{ for each } a \in \Sigma\}\cdot$$

A *factor* of $w$ is any word $z \in \Sigma^*$ such that $w = xzy$ for some $x, y \in \Sigma^*$.

The word $w$ is *primitive* if $w$ is nonempty and for each word $u$ and nonnegative integer $n$, the equality $w = u^n$ implies $w = u$ (and $n = 1$, of course). A basic result in word combinatorics says that for each nonempty word $x$ there exists a unique primitive word $t$, the *primitive root* of $x$, such that $x \in t^+$. The words $u$ and $w$ *commute* if $uw = wu$. It is again a well-known fact that two nonempty words commute if and only if they have the same primitive root.

The words $u$ and $w$ are *conjugate* (*words of each other*) if there exist $x$ and $y$ such that $u = xy$ and $w = yx$. Let $R$ be the relation of $\Sigma^*$ defined by: $uRw$ if $u$ and $w$ are conjugate. Then the relation $R$ is certainly an equivalence, and the concept of conjugacy can be generalized to more than two words.

Let $L \subseteq \Sigma^*$ be a language. The set of all symbols of $\Sigma$ occurring in words of $L$ is called *alphabet* of $L$, denoted by $\mathrm{alph}(L)$. Write $\mathrm{alph}(w) = \mathrm{alph}(\{w\})$ and call it the alphabet of the word $w$. The *commutative closure* $c(L)$ of the language $L$ is the set

$$c(L) = \{x \mid x \in c(w) \text{ for some } w \in L\}\cdot$$

We say that $L$ is *commutative* if $L = c(L)$. In this paper we will study in particular the finite commutative language $c(a_1 a_2 \cdots a_n)$, which we denote by $E_n$.

We say that morphisms $g$ and $h$ *agree on the word* $u$ if $g(u) = h(u)$ holds. Morphisms *agree on a language* $L$ if they agree on all $u \in L$. We say that $g$ and $h$ are *length–equivalent* on a language $L$ if $|g(w)| = |h(w)|$ for each $w \in L$.

The symbol $\mathbb{N}$ indicates, as usually, the set of all natural numbers and $\mathbb{N}_+ = \mathbb{N} \setminus \{0\}$. For each $n \in \mathbb{N}_+$, let $\Sigma_n = \{a_1, a_2, \dots, a_n\}$ be the alphabet consisting of $n$ distinct symbols $a_1, a_2, \dots, a_n$. The traditional *Parikh-map* $\Psi_n$ from $\Sigma_n^*$ onto $\mathbb{N}^n$ is defined by $\Psi_n(w) = (|w|_{a_1}, |w|_{a_2}, \dots, |w|_{a_n})$. The cardinality of a set $X$ is denoted by $\mathrm{Card}(X)$.

Assume now that $n \in \mathbb{N}_+$ and $L \subseteq \Sigma_n^*$. A *basis* of $L$ is any finite subset $F$ of $L$ such that:

(i) the set $\Psi_n(F)$ consists of exactly $\mathrm{Card}(F)$ linearly independent elements (over $\mathbb{Q}$, the rational numbers);

(ii) for each $w \in L$, the vector $\Psi_n(w)$ is a linear combination (over $\mathbb{Q}$) of some vectors in $\Psi_n(F)$.

The *dimension* of $L$, denoted by $\dim L$, is cardinality of any basis of $L$.

A set $S \subseteq \mathbb{N}^n$ is *linear* if there exist $m \in \mathbb{N}$ and vectors $\bar{v}, \bar{v}_1, \bar{v}_2, \ldots, \bar{v}_m \in \mathbb{N}^n$ such that $S = \{\bar{v} + k_1\bar{v}_1 + k_2\bar{v}_2 + \cdots + k_m\bar{v}_m \mid k_1, k_2, \ldots, k_m \in \mathbb{N}\}$. A commutative language $L \subseteq \Sigma_n^*$ is a CLIP-*language* if $\Psi_n(L)$ is a linear set.

For each nonnegative rational number $q$, let $\lceil q \rceil$ ($\lfloor q \rfloor$, resp.) denote the smallest (the greatest, resp.) integer $k$ such that $q \leq k$ ($k \leq q$, resp.).

A permutation $\sigma \in \mathcal{S}_n$ is a bijective mapping $\{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ and it can be simply represented by the queue $\sigma(1)\sigma(2)\cdots\sigma(n)$ or, in the case of possible confusion, by $(\sigma(1), \sigma(2), \ldots, \sigma(n))$.

We have already noticed the natural $1-1$ correspondence between sets of permutations and subsets of $E_n$. We say that the set $\mathcal{R} \subseteq \mathcal{S}_n$ *produces* the set

$$R = \{a_{\sigma(1)} \cdots a_{\sigma(n)} \mid \sigma \in \mathcal{R}\}\cdot$$

The construction of a test set $S$ for the language $E_n$ is equivalent to the construction of the corresponding set of permutations $\mathcal{T}_n$. Another equivalent characterisation of the sought-after set $\mathcal{T}_n$ is that $(*)$ implies $(**)$ for any pair of morphisms $g, h$, such that $g(a_i) = x_i$ and $h(a_i) = y_i$ ($i = 1, 2, \ldots, n$). These facts are obvious but quite important for the future exposition, since they allow us to make use of both word equation and morphisms agreement notation, as well as to switch, if convenient, between languages and sets of permutations.

We shall need some results from the rudiments of combinatorics on words. For the proofs of the first two see for instance [15]. The first is the famous Periodicity Lemma of Fine and Wilf.

**Theorem 1.** *If two powers $u^m$ and $v^n$ of nonempty words $u$ and $v$ have a common factor of length at least $|u| + |v| - d$, where $d$ is the greatest common divisor of $|u|$ and $|v|$, then the primitive roots of $u$ and $v$ are conjugate.*

The conjugacy, the second important property between two words (commutativity is the first) can be characterized as follows.

**Theorem 2.** *Let $x$ and $y$ be nonempty words. The following three conditions are equivalent:*

(i) *the words $x$ and $y$ are conjugate;*

(ii) *the words $x$ and $y$ are of equal length and there exist unique words $t_1$, and $t_2$, with $t_2$ nonempty, such that $t = t_1t_2$ is primitive and $x \in (t_1t_2)^+$ and $y \in (t_2t_1)^+$;*

(iii) *there exists a word $z$ such that $xz = zy$.*

*Furthermore, if* (ii) *holds, then for each word $w$, we have $xw = wy$ if and only if $w \in (t_1t_2)^*t_1$.*

By the next theorem and its corollary (for the easy proof, see [9]), given distinct words $x_1, y_1$, the structure of any solution $\alpha, \beta$ of the system of equations

$$x_1\alpha = y_1\beta, \qquad \alpha x_1 = \beta y_1$$

is unique.

**Theorem 3.** *Let $x_1$ and $y_1$ be distinct words. The following two conditions are equivalent:*

(i) *there exist words $x_2$ and $y_2$ such that*

$$x_1x_2 = y_1y_2 \qquad x_2x_1 = y_2y_1;$$

(ii) *there exist a unique word $t_1$ and a unique nonempty word $t_2$ such that $t_1t_2$ is primitive and $x_1, y_1 \in (t_1t_2)^*t_1$.*

*Furthermore, if* (ii) *holds, then for each pair of words $x_3, y_3$ we have*

$$x_1x_3 = y_1y_3 \qquad x_3x_1 = y_3y_1$$

*if and only if $|x_1x_3| = |y_1y_3|$ and $x_3, y_3 \in (t_2t_1)^*t_2 \cup \{\epsilon\}$. Moreover $x_1x_3 \in (t_1t_2)^+$ and $x_3x_1 \in (t_2t_1)^+$.*

We can write the following usable:

**Corollary.** *Let $x_1, x_2, x_3, y_1, y_2, y_3$ be words such that $|x_1| \neq |y_1|$, $|x_2| = |x_3|$ and*

$$\begin{cases} x_1x_2 = y_1y_2 \\ x_2x_1 = y_2y_1 \end{cases} \qquad \begin{matrix} x_1x_3 = y_1y_3 \\ x_3x_1 = y_3y_1. \end{matrix}$$

*Then $x_2 = x_3$ and $y_2 = y_3$.*

## 2. On commutation of words

Let us generalize the concept of commutation to arbitrary many words. For each $n \in \mathbb{N}_+$ we say that the words $x_1, x_2, \ldots, x_n$ *commute* if

$$x_1x_2 \cdots x_n = x_{\sigma(1)}x_{\sigma(2)} \cdots x_{\sigma(n)} \tag{$\diamond$}$$

for each permutation $\sigma \in \mathcal{S}_n$. Certainly, if the words $x_1, x_2, \ldots, x_n$ are all nonempty, they commute if and only if they have the same primitive root.

Let $n \in \mathbb{N}_+$. For how many permutation $\sigma \in \mathcal{S}_n$ the equality ($\diamond$) has to be valid to guarantee that the words $x_1, x_2, \ldots, x_n$ commute? In the following we shall see that a number depending logarithmically on $n$ is sufficient (Th. 4), but, in general, a constant number is not (Th. 5). All logarithms are of course in the base 2.

For each $m \in \mathbb{N}, n \in \mathbb{N}_+$ define the permutation $\delta_m^n$ of $1, 2, \ldots, n$ inductively as follows. Let

$\delta_0^n = (1, 2, \ldots, n)$;
$\delta_m^1 = (1)$; and
$\delta_1^n = (r+1, r+2, \ldots, n, 1, 2, \ldots, r)$, where $r = \lceil n/2 \rceil$.

Let now $m \in \mathbb{N}_+, n \in \{2, 3, \ldots\}$ and assume that $\delta_j^k$ is given for $j = 0, 1, 2, \ldots, m$ and $k = 1, 2, \ldots, n-1$. Then (denoting again $r = \lceil n/2 \rceil$), define

$$\delta_{m+1}^n = (\delta_m^r(1), \delta_m^r(2), \ldots, \delta_m^r(r), r + \delta_m^{n-r}(1), r + \delta_m^{n-r}(2), \ldots, r + \delta_m^{n-r}(n-r)).$$

It should be clear that $\delta_m^n = (1, 2, \ldots, n)$ for each $m > \lceil \log n \rceil$.

For each $n \in \mathbb{N}_+$, let $\Delta_n = \{\delta_m^n \mid m = 1, 2, \ldots, \lceil \log n \rceil\}$.

The definition of $\delta_m^n$ is easy to understand when $n = 2^k$ for some $k$. For general $n$, some work with integer parts of fractions is inevitable. The reader, who wants to grasp the main idea of our construction and avoid the exercise of counting with ceilings and floors, can simply forget them and confine oneself to the case $n = 2^k$.

**Example 1.** For $n = 4$, $n = 8$ and $n = 11$ we have
$\Delta_4 = \{(3, 4, 1, 2), (2, 1, 4, 3)\}$
$\Delta_8 = \{(5, 6, 7, 8, 1, 2, 3, 4), (3, 4, 1, 2, 7, 8, 5, 6), (2, 1, 4, 3, 6, 5, 8, 7)\}$, and
$\Delta_{11} = \{(7, 8, 9, 10, 11, 1, 2, 3, 4, 5, 6), (4, 5, 6, 1, 2, 3, 10, 11, 7, 8, 9),$
$(3, 1, 2, 6, 4, 5, 9, 7, 8, 11, 10), (2, 1, 3, 5, 4, 6, 8, 7, 9, 10, 11)\}\cdot$

The following result is a slight modification of Theorem 9 in [9].

**Theorem 4.** *Let $n \in \mathbb{N}_+$ be a number and $x_1, x_2, \ldots, x_n$ be words. If*

$$x_1 x_2 \cdots x_n = x_{\delta(1)} x_{\delta(2)} \cdots x_{\delta(n)} \tag{1}$$

*for each $\delta \in \Delta_n$, then the words $x_1, x_2, \ldots, x_n$ commute.*

*Proof.* By induction on $n$. The cases $n = 1$ and $n = 2$ are not difficult.

Let $n \geq 3$ and and assume that the theorem holds for each $k \in \{1, 2, \ldots, n-1\}$. Let $r = \lceil n/2 \rceil$. By the equality (1) we have $x_1 \ldots x_r x_{r+1} \ldots x_n = x_{r+1} \ldots x_n x_1 \ldots x_r$, so the words $x_1 x_2 \cdots x_r$ and $x_{r+1} x_{r+2} \cdots x_n$ commute. Also, by the remaining equalities in (1), we have

$$x_1 x_2 \cdots x_r = x_{\delta(1)} x_{\delta(2)} \ldots x_{\delta(r)}$$

for all permutations $\delta \in \Delta_r$ and

$$x_{r+1} x_{r+2} \ldots x_n = x_{r+\rho(1)} x_{r+\rho(2)} \cdots x_{r+\rho(n-r)}$$

for each $\rho \in \Delta_{n-r}$. By the induction hypothesis, the words $x_1, x_2, \ldots, x_r$ commute, as well as the words $x_{r+1}, x_{r+2}, \ldots, x_n$. This extends the induction. □

**Theorem 5.** *For each $m \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that for any $m$ permutations $\sigma_1, \sigma_2, \ldots, \sigma_m \in \mathcal{S}_n$ we can find words $x_1, x_2, \ldots, x_n$ which do not commute and satisfy*

$$x_1 x_2 \cdots x_n \;=\; x_{\sigma_i(1)} x_{\sigma_i(2)} \cdots x_{\sigma_i(n)} \quad (i = 1, 2, \ldots, m). \tag{2}$$

*Proof.* Assume that $m$ is in $\mathbb{N}$ and choose $n \geq 3^{2^m}$. Let $\sigma_1, \sigma_2, \ldots, \sigma_m$ be any permutations of $1, 2, \ldots, n$. We show that there exist three distinct elements $p, q, r \in \{1, 2, \ldots, n\}$ which in each sequence $\sigma_j(1), \sigma_j(2), \ldots \sigma_j(n)$, $j = 1, 2, \ldots, m$, form either an increasing or a decreasing (*i.e.* monotone) subsequence.

It is a well known fact (dating back to [6]) that for each $s \in \mathbb{N}$, any sequence of $s^2$ distinct real numbers contains a subsequence of $s$ numbers which is either increasing or decreasing. Thus there exist integers $i_1, i_2, \ldots, i_{3^{2^{m-1}}}$ in $\{1, 2, \ldots, n\}$ which in $\sigma_1(1), \sigma_1(2), \ldots, \sigma_1(n)$ appear in a monotone order.

Proceed by induction. Let $k \in \{1, 2, \ldots, m-1\}$. Suppose that there exist integers $j_1, j_2, \ldots, j_{3^{2^{m-k}}}$ in $\{1, 2, \ldots, n\}$ such that the these integers form a monotone subsequence in $\sigma_s(1), \sigma_s(2), \ldots, \sigma_s(n)$ for each $s \in \{1, 2, \ldots, k\}$. Consider the permutation $\sigma_{k+1}$. By the facts above, there exist integers $s_1, s_2, \ldots, s_{3^{2^{m-k-1}}}$ in $j_1, j_2, \ldots, j_{3^{2^{m-k}}}$ which in $\sigma_s(1), \sigma_s(2), \ldots, \sigma_s(n)$ appear in either increasing or decreasing order for each $s \in \{1, 2, \ldots, k+1\}$. This extends the induction.

Let $p < q < r \in \{1, 2, \ldots, n\}$ be integers which in the sequence $\sigma_j(1), \sigma_j(2), \ldots, \sigma_j(n)$ form a monotone subsequence for each $j = 1, 2, \ldots, m$. Let $a$ and $b$ be distinct symbols. Choose $x_p = x_r = a$, $x_q = b$ and $x_i = \epsilon$ for each $i \in \{1, 2, \ldots, n\} \setminus \{p, q, r\}$. For any $j \in \{1, 2, \ldots, m\}$ the equality (2) looks like $aba = aba$, but the words $x_p, x_q, x_r$ do not commute and thus neither do the words $x_1, x_2, \ldots, x_n$.

□

## 3. Permutation, shuffle and conjugacy

In the rest of the paper $g$ and $h$ will be arbitrary morphisms defined on $\Sigma_n$ and we put

$$x_i = g(a_i), \quad y_i = h(a_i) \quad \text{for} \;\; i = 1, \ldots, n. \tag{†}$$

We now introduce four conditions on the basis of which our vital test set and word equation problem can be solved.

**Definition.** The morphisms $g$ and $h$ satisfy the *permutation condition* if they agree on $E_n$, *i.e.* if they agree on

$$a_{\sigma(1)} a_{\sigma(2)} \cdots a_{\sigma(n)}$$

for each $\sigma \in \mathcal{S}_n$.

We shall generalize the above definition to subsets of $\Sigma_n$. Let $I = \{a_{i_1}, a_{i_2}, \ldots, a_{i_k} \mid 1 \leq i_1 < i_2 < \cdots < i_k \leq n\}$. The morphisms $g$ and $h$ *satisfy the permutation condition on the set $I$* if they agree on

$$a_{i_{\sigma(1)}} a_{i_{\sigma(2)}} \cdots a_{i_{\sigma(k)}}$$

for each $\sigma \in \mathcal{S}_k$. The permutation condition is very restrictive. The first anticipation is that it, in the nontrivial case, implies commutativity. The next theorem says that exactly this is not the case.

**Theorem 6.** *Let $g$ and $h$ satisfy the permutation condition. Then one of the following statements holds:*

(i) *$x_i = y_i$ for each $i \in \{1, 2, \ldots, n\}$;*

(ii) *there exist $p, q \in \{1, 2, \ldots, n\}$, $p < q$, such that $x_p \neq y_p$, $x_q \neq y_q$ and $x_i y_i = \epsilon$ for each $i \in \{1, 2, \ldots, n\} \setminus \{p, q\}$. Then there exist unique word $t_1$ and a unique nonempty word $t_2$ such that $t = t_1 t_2$ is the primitive root of $x_1 x_2 \cdots x_n$, $x_p, y_p \in (t_1 t_2)^* t_1$ and $x_q, y_q \in (t_2 t_1)^* t_2$;*

(iii) *there exist three indices $p, q, r \in \{1, 2, \ldots, n\}$ such that $x_p \neq y_p$, $x_q \neq y_q$ and $x_r y_r \neq \epsilon$. Then the words $x_1, x_2, \ldots, x_n$, $y_1, y_2, \ldots, y_n$ commute, i.e., if $t$ is the primitive root of $x_1 x_2 \cdots x_n$, we have $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n \in t^*$.*

*Proof.* Assume that (i) does not hold. There then exist at least two indices $j \in \{1, 2, \ldots, n\}$ such that $x_j \neq y_j$. Let $p, q \in \{1, 2, \ldots, n\}$, $p < q$, be such that $x_p \neq y_p$ and $x_q \neq y_q$. Two possibilities arise. Either $x_i y_i = \epsilon$ for each $i \in \{1, 2, \ldots, n\} \setminus \{p, q\}$ or there exist $r \in \{1, 2, \ldots, n\} \setminus \{p, q\}$ such that $x_r y_r \neq \epsilon$.

**1.** Consider the first possibility. By Theorem 3, the case (ii) holds.

**2.** Assume that the second possibility holds. Suppose without loss of generality, that $x_r \neq \epsilon$. Let

$$z_i = \begin{cases} x_i, & \text{for } i = 1, 2, \ldots, p-1, \\ x_{i+1} & \text{for } i = p, p+1, \ldots, n-1; \end{cases} \quad u_i = \begin{cases} y_i, & \text{for } i = 1, 2, \ldots, p-1, \\ y_{i+1} & \text{for } i = p, p+1, \ldots, n-1. \end{cases}$$

Then

$$x_p z_{\sigma(1)} z_{\sigma(2)} \cdots z_{\sigma(n-1)} = y_p u_{\sigma(1)} u_{\sigma(2)} \cdots u_{\sigma(n-1)}$$

and

$$z_{\sigma(1)} z_{\sigma(2)} \cdots z_{\sigma(n-1)} x_p = u_{\sigma(1)} u_{\sigma(2)} \cdots u_{\sigma(n-1)} y_p$$

for each permutation $\sigma \in \mathcal{S}_{n-1}$.

Since $x_p \neq y_p$, we deduce from the corollary of Theorem 3, that

$$z_1 z_2 \cdots z_{n-1} = z_{\sigma(1)} z_{\sigma(2)} \cdots z_{\sigma(n-1)}$$

for each $\sigma \in \mathcal{S}_{n-1}$. This certainly means that the words $x_1, \ldots, x_{p-1}, x_{p+1}, \ldots, x_n$ commute. Similarly it can be shown that the words $x_1, \ldots, x_{q-1}$, $x_{q+1}, \ldots, x_n$ commute. Since $p, q$, and $r$ are all distinct and $x_r \neq \epsilon$, we deduce that the words

$x_1, x_2, \ldots, x_n$ commute, which finally implies that $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$ commute. □

Note that under the adopted assignment the permutation condition is equivalent to $(**)$. Let us recall that we are interested in the smallest possible subsets of $\mathcal{S}_n$ that produce a test set for $E_n$. It is shown that there exist a test set of size $O(n)$ and also that this order of magnitude is the best possible.

**Example 2.** Consider the language $E_3$. Define morphisms $g, h$ by

$$\begin{array}{lll} g(a_1) = a & g(a_2) = ab & g(a_3) = bab \\ h(a_1) = aba & h(a_2) = ab & h(a_3) = b \end{array}$$

and verify that they agree on all elements of $E_3$ except of $a_1a_2a_3$. By the symmetry of letters, it shows that no proper subset of $E_3$ is its test set.

**Definition.** The morphisms $g$ and $h$ satisfy the *conjugacy condition* if they agree on each conjugate word of $a_1a_2 \ldots a_n$.

The concept "conjugacy condition" refers to the fact that

$$\{a_ia_{i+1} \ldots a_na_1a_2 \ldots a_{i-1} \mid i = 1, 2, \ldots, n\}$$

is exactly the set of all conjugate words of $a_1a_2 \ldots a_n$. Along our conventions $g$ and $h$ satisfy the conjugacy condition if the equality

$$x_ix_{i+1} \cdots x_nx_1x_2 \cdots x_{i-1} \;=\; y_iy_{i+1} \cdots y_ny_1y_2 \cdots y_{i-1}$$

holds for each $i \in \{1, 2, \ldots, n\}$.

Denote by $\mathcal{CON}_n$ the subset of $\mathcal{S}_n$ that produces the set of all conjugates of the word $a_1a_2 \cdots a_n$.

Words satisfying the conjugacy condition have some remarkable properties.

**Theorem 7.** *Let $g$ and $h$ be morphisms satisfying conjugacy condition and suppose that $x_1x_2 \cdots x_n$ is nonempty. For each $i \in \{1, 2, \ldots, n\}$, let $t_i$ be the primitive root of $x_ix_{i+1} \cdots x_nx_1x_2 \cdots x_{i-1}$ and $d_i$ be the word such that either $x_i = d_iy_i$ or $y_i = d_ix_i$. Then $t_1, t_2, \ldots, t_n$ are conjugate words of each other and, for each $i \in \{1, 2, \ldots, n\}$, we have $d_i \in t_i^*$.*

*Proof.* Since $x_1x_2 \cdots x_n$, $x_2x_3 \cdots x_nx_1$, … ,$x_nx_1x_2 \cdots x_{n-1}$ are conjugate, their primitive roots $t_1, t_2, \ldots, t_n$ are also conjugate, by the basic results in combinatorics of words. Let $i \in \{1, 2, \ldots, n\}$. If $d_i \neq \epsilon$, Theorem 3 implies that both $d_i$ and $x_ix_{i+1} \cdots x_nx_1x_2 \cdots x_{i-1}$ are $\in t_i^+$. The case $d_i = \epsilon$ is clear. □

**Definition.** Let $n \geq 2$ be an integer and $r = \lceil n/2 \rceil$. The morphisms $g, h$ satisfy the *shuffle condition* if they agree on following $n$ words:

$$\left\{\begin{array}{l}
a_1\ a_2 \cdots a_{r-3}\ a_{r-2}\ a_{r-1}\ a_r\ \boldsymbol{a_{r+1}}\ \boldsymbol{a_{r+2}}\ \boldsymbol{a_{r+3}}\ \boldsymbol{a_{r+4}} \cdots \boldsymbol{a_n}\\
a_1\ a_2 \cdots a_{r-3}\ a_{r-2}\ a_{r-1}\ \boldsymbol{a_{r+1}}\ a_r\ \boldsymbol{a_{r+2}}\ \boldsymbol{a_{r+3}}\ \boldsymbol{a_{r+4}} \cdots \boldsymbol{a_n}\\
a_1\ a_2 \cdots a_{r-3}\ a_{r-2}\ \boldsymbol{a_{r+1}}\ a_{r-1}\ \boldsymbol{a_{r+2}}\ a_r\ \boldsymbol{a_{r+3}}\ \boldsymbol{a_{r+4}} \cdots \boldsymbol{a_n}\\
a_1\ a_2 \cdots a_{r-3}\ \boldsymbol{a_{r+1}}\ a_{r-2}\ \boldsymbol{a_{r+2}}\ a_{r-1}\ \boldsymbol{a_{r+3}}\ a_r\ \boldsymbol{a_{r+4}} \cdots \boldsymbol{a_n}\\
\qquad\qquad\qquad\qquad\vdots\\
\boldsymbol{a_{r+1}}\ \boldsymbol{a_{r+2}} \cdots \boldsymbol{a_{n-3}}\ a_1\ \boldsymbol{a_{n-2}}\ a_2\ \boldsymbol{a_{n-1}}\ a_3\ \boldsymbol{a_n}\ a_4 \cdots a_r\\
\boldsymbol{a_{r+1}}\ \boldsymbol{a_{r+2}} \cdots \boldsymbol{a_{n-3}}\ \boldsymbol{a_{n-2}}\ a_1\ \boldsymbol{a_{n-1}}\ a_2\ \boldsymbol{a_n}\ a_3\ a_4 \cdots a_r\\
\boldsymbol{a_{r+1}}\ \boldsymbol{a_{r+2}} \cdots \boldsymbol{a_{n-3}}\ \boldsymbol{a_{n-2}}\ \boldsymbol{a_{n-1}}\ a_1\ \boldsymbol{a_n}\ a_2\ a_3\ a_4 \cdots a_r\\
\boldsymbol{a_{r+1}}\ \boldsymbol{a_{r+2}} \cdots \boldsymbol{a_{n-3}}\ \boldsymbol{a_{n-2}}\ \boldsymbol{a_{n-1}}\ \boldsymbol{a_n}\ a_1\ a_2\ a_3\ a_4 \cdots a_r.
\end{array}\right. \tag{$\mathcal{SH}$}$$

The bold typeface in the definition above helps to grasp the structure of the set $\mathcal{SH}$ and has no semantic relevance.

We give also a more formal definition of the set $\mathcal{SH}$. For all integers $i \in \mathbb{Z}$ define words $c_i$, $d_i$ by

$$c_i = \begin{cases} a_i, & i = 1, \ldots, r \\ \epsilon, & \text{otherwise;} \end{cases} \qquad d_i = \begin{cases} a_i, & i = r+1, \ldots, n \\ \epsilon, & \text{otherwise.} \end{cases}$$

Then

$$\mathcal{SH} = \left\{ \prod_{i \in \mathbb{Z}} c_i d_{i+k} \mid k \in \mathbb{Z} \right\} = \left\{ \prod_{i \in \mathbb{Z}} c_i d_{i+k} \mid k = 1, \ldots, n \right\}.$$

Again, denote by $\mathcal{SHU}_n$ the set of $n$ permutations that produces the words in $\mathcal{SH}$. There is not much to say about the structure of morphisms satisfying the shuffle condition. It certainly does not alone imply the permutation condition. In fact Example shows that even together the shuffle and the conjugacy conditions are not as strong as the permutation condition. We are going to introduce one more tool.

**Definition.** Let $n \geq 2$ be an integer and $r = \lceil n/2 \rceil$. The morphisms $g$ and $h$ satisfy the *weak permutation condition* if they agree on words

$$\left\{\begin{array}{l}
a_1 a_2 \cdots a_r a_{r+\delta(1)} a_{r+\delta(2)} \cdots a_{r+\delta(n-r)}\\
a_{r+\delta(1)} a_{r+\delta(2)} \cdots a_{r+\delta(n-r)} a_1 a_2 \cdots a_r\\
a_{r+1} a_{r+2} \cdots a_n a_{\rho(1)} a_{\rho(2)} \cdots a_{\rho(r)}\\
a_{\rho(1)} a_{\rho(2)} \cdots a_{\rho(r)} a_{r+1} a_{r+2} \cdots a_n
\end{array}\right. \tag{$\mathcal{WP}$}$$

for all $\delta \in \Delta_{n-r}$ and $\rho \in \Delta_r$.

Let $\mathcal{WPE}_n$ be the subset of $\mathcal{S}_n$ producing the words in $\mathcal{WP}$. Certainly $\mathcal{WPE}_n$ contains $2\lceil \log\ r \rceil + 2\lceil \log(n-r) \rceil$ permutations.

## 4. Sufficient conditions for the permutation property

**Theorem 8.** *Let $n \geq 4$ be an integer and $r = \lceil n/2 \rceil$. Suppose that morphisms $g$ and $h$ satisfy both the conjugacy and the weak permutation condition and furthermore $|x_1x_2\cdots x_r| \neq |y_1y_2\cdots y_r|$. Then the morphisms satisfy the permutation condition.*

*Proof.* By the corollary of Theorem 3,

$$x_1x_2\cdots x_r = x_{\delta(1)}x_{\delta(2)}\cdots x_{\delta(r)}$$
$$y_1y_2\cdots y_r = y_{\delta(1)}y_{\delta(2)}\cdots y_{\delta(r)}$$

for each $\delta \in \Delta_r$, and

$$x_{r+1}x_{r+2}\cdots x_n = x_{r+\rho(1)}x_{r+\rho(2)}\cdots x_{r+\rho(n-r)}$$
$$y_{r+1}y_{r+2}\cdots y_n = y_{r+\rho(1)}y_{r+\rho(2)}\cdots y_{r+\rho(n-r)}$$

for each $\rho \in \Delta_{n-r}$. Theorem 4 now implies that the words $x_1, x_2, \ldots, x_r$ commute and so do also the words $y_1, y_2, \ldots, y_r$ as well as $x_{r+1}, x_{r+2}, \ldots, x_n$ and the words $y_{r+1}, y_{r+2}, \ldots, y_n$.

We shall now show that the morphisms $g$ and $h$ satisfy the permutation property.

If there are exactly two distinct indices $i, j \in \{1, 2, \ldots, n\}$ such that $x_iy_i$ and $x_jy_j$ are nonempty, there is nothing to prove: the total system of equations collapses to $x_ix_j = y_iy_j$, $x_jx_i = y_jy_i$.

Assume thus, without loss of generality, that there exist indices $p, q \in \{1, 2, \ldots, r\}$, $p < q$, and $s \in \{r+1, r+2, \ldots, n\}$ such that $x_py_p$,$x_qy_q$ and $x_sy_s$ are all nonempty and $|x_1 \ldots x_r| > |y_1 \ldots y_r|$. Let $t$, $u$ and $v$ be the primitive roots of $x_1x_2\cdots x_r$, $y_1y_2\cdots y_r$, and $x_1x_2\cdots x_n$, respectively. (If $y_1y_2\cdots y_r$ is empty, put $u = v$.) Certainly $x_p, x_q$ are in $t^*$, $y_p, y_q$ are in $u^*$ and, by Theorem 7, the difference of $x_1\cdots x_r$ and $y_1\cdots y_r$ is a conjugate of $v$.

Now, if $x_p = y_p$ (or $x_q = y_q$), we have necessarily $t = u = v$ since $x_1 \ldots x_r \in t^+$ is a prefix of $x_1 \ldots x_n \in v^+$. Then also $x_{r+1}x_{r+2}\cdots x_n$ and $y_{r+1}y_{r+2}\cdots y_n$ are in $v^*$ and we are through: all the words commute. Assume thus that $x_p \neq y_p$ and $x_q \neq y_q$. If any of the words $x_p, x_q, y_p, y_q$ is empty, we again have, by Theorem 7, $t = u = v$. Let thus $x_p, x_q, y_p, y_q$ all be nonempty. Then $x_1x_2\cdots x_r$ is longer than $2|v|$ and, by the Periodicity Lemma, $t = v$. Again we see that the words commute. □

**Theorem 9.** *Assume $n \geq 4$ is an integer, $r = \lceil n/2 \rceil$, and the morphisms $g$ and $h$ satisfy the*

(i) *conjugacy and shuffle condition;*
(ii) *permutation condition on $\{a_1, a_2, \ldots, a_r\}$ and on the set $\{a_{r+1}, a_{r+2}, \ldots, a_n\}$.*

*Then they satisfy the permutation condition.*

*Proof.*

**1.** If $x_i = y_i$ for $i = 1, 2, \dots, n$, we are through.

**2.** Assume then, without loss of generality, that $q$ is the greatest number $i \in \{1, 2, \dots, r\}$ such that $x_i \neq y_i$. By Theorem 7, the words $x_1 x_2 \cdots x_r$ and $x_1 x_2 \cdots x_n$ and therefore also $x_{r+1} x_{r+2} \cdots x_n$ are powers of the same primitive word $t$, say. By Theorem 6 two cases appear. Either

1° all the words $x_1, x_2, \dots, x_r \in t^*$

or

2° there exist exactly two distinct indices $p$ and $q$ in $\{1, 2, \dots, r\}$ such that $x_p y_p$ and $x_q y_q$ are nonempty and $x_p$ and $x_q$ do not commute.

**2.1.** Consider first Case 2°. By the corollary of Theorem 3, there exist nonempty words $t_1, t_2$ and integers $r_1, r_2, s_1, s_2 \in \mathbb{N}$ such that $t = t_1 t_2$, $r_1 \neq s_1$, $r_1 + r_2 = s_1 + s_2$, $x_p = t^{r_1} t_1$, $y_p = t^{s_1} t_1$, $x_q = t_2 t^{r_2}$, and $y_q = t_2 t^{s_2}$.

**2.1.1.** If $x_{r+1} \cdots x_n = \epsilon$, we are done.

**2.1.2.** Let $s$ be the smallest number $i \in \{r+1, r+2, \dots, n\}$ such that $x_i y_i \neq \epsilon$. By Theorem 3 there exist words $u_1, u_2$ and integers $r_3, r_4, s_3, s_4 \in \mathbb{N}$ such that $u_1 u_2 = t = t_1 t_2$, $x_s = t^{r_3} u_1$, $y_s = t^{s_3} u_1$, $x_{s+1} x_{s+2} \cdots x_n = u_2 t^{r_4}$ and $y_{s+1} y_{s+2} \cdots y_n = u_2 t^{s_4}$. By the shuffle condition we have the following identity.

$$t^{r_1} t_1 t^{r_3} u_1 t_2 t^{r_2} u_2 t^{r_4} \;\; = \;\; t^{s_1} t_1 t^{s_3} u_1 t_2 t^{s_2} u_2 t^{s_4}. \qquad (3)$$

Assume without loss of generality that $r_1 > s_1$ (and thus $s_2 > 0$). Certainly (3) implies

$$(t_2 t_1)^{r_1 - s_1} t^{r_3} u_1 t_2 t^{r_2} u_2 t^{r_4} = t^{s_3} u_1 t_2 t^{s_2} u_2 t^{s_4}. \qquad (4)$$

**2.1.2.1.** If $s_3 > 0$, then $t = t_1 t_2 = t_2 t_1$, a contradiction, since $t$ is primitive and $t_1, t_2$ are both nonempty.

**2.1.2.2.** Suppose that $s_3 = 0$. Then $r_3 \neq 0$ and $(t_2 t_1) u_1 = u_1 (t_2 t_1)$, so $u_1 = \epsilon$ since $t_2 t_1$ is primitive. Now we have $u_2 = t$ and (4) yields $t_1 t_2 = t_2 t_1$, again a contradiction.

**2.2.** Consider then the case 1°. Apply Theorem 6 to the words $x_{r+1}, x_{r+2}, \dots, x_n$ and $y_{r+1}, y_{r+2}, \dots, y_n$.

**2.2.1.** If $x_{r+1}, x_{r+2}, \dots, x_n \in t^*$, the proposition obviously is true: all the words commute.

**2.2.2.** The case (ii) of Theorem 6 was studied above (**2.1**).

**2.2.3.** Suppose that $x_j = y_j$ for $j = r+1, r+2, \dots, n$. The shuffle condition now leads to equalities $(x_1 x_2 \cdots x_{q-1}) x_{r+1} x_q = (y_1 y_2 \cdots y_{q-1}) y_{r+1} y_q$. Since $x_1, x_2, \dots, x_{q-1}, x_q$ as well as $y_1, y_2, \dots, y_{q-1}$, $y_q$ are all in $t^*$, it is not difficult to see that the word $x_{r+1} = y_{r+1}$ is also in $t^*$. Similarly, using further shuffle equalities, we prove that $x_j$ is in $t^*$ for all $j = r+1, r+2, \dots, n$. This completes the proof. □

## 5. A linear size test set for the language $E_n$

In this section we construct recursively a sequence $(\mathcal{T}_n)_{n\in\mathbb{N}_+}$, with $\mathcal{T}_n \subseteq \mathcal{S}_n$, which will determine our test sets. Let us first explain the idea of the recursivity in our construction.

Let $\mathcal{R}$ be a subset of $\mathcal{S}_n$. Put $r = \lceil n/2 \rceil$ and for any $\tau \in \mathcal{R}$ define mappings

$$\sigma_\tau : \{1,\dots,r\} \to \{1,\dots,n\} \quad \text{by} \quad \sigma_\tau(i) = \tau(i),$$

and

$$\rho_\tau : \{1,\dots,n-r\} \to \{1,\dots,n\} \quad \text{by} \quad \rho_\tau(i) = \tau(r+i) - r.$$

If $\{\tau(i) \mid i = 1,\dots,r\} = \{1,\dots,r\}$, then both $\sigma_\tau, \rho_\tau$ are permutations and we say that $\tau$ is *per partes.* Put

$$\begin{aligned} \mathcal{P}_{\text{LEFT}}(\mathcal{R}) &= \{\sigma_\tau \mid \tau \in \mathcal{R},\ \ \tau \text{ is } \textit{per partes}\}; \\ \mathcal{P}_{\text{RIGHT}}(\mathcal{R}) &= \{\rho_\tau \mid \tau \in \mathcal{R},\ \ \tau \text{ is } \textit{per partes}\} \cdot \end{aligned}$$

We say that the set $\mathcal{R} \subseteq \mathcal{S}_n$ is *founded* if

(i) $\mathcal{P}_{\text{LEFT}}(\mathcal{R})$ produces a test set for $E_r$;
(ii) $\mathcal{P}_{\text{RIGHT}}(\mathcal{R})$ produces a test set for $E_{n-r}$.

Our construction of the sequence $(\mathcal{T}_n)_{n\in\mathbb{N}_+}$ is based on the following

**Theorem 10.** *Let $\mathcal{R}$ be a subset of $\mathcal{S}_n$ such that*

(i) *$\mathcal{CON}_n \cup \mathcal{SHU}_n \cup \mathcal{WPE}_n \subseteq \mathcal{R}$; and*
(ii) *$\mathcal{R}$ is founded.*

*Then $R$ (the set $\mathcal{R}$ produces) is a test set for $E_n$.*

*Proof.* Suppose that the morphisms $g$ and $h$ agree on $R$. We want to show that $g$ and $h$ satisfy the permutation condition.

Let $r = \lceil n/2 \rceil$ and suppose first that $|x_1x_2\cdots x_r| \neq |y_1y_2\cdots y_r|$ . By (i), the morphisms $g$ and $h$ satisfy both the conjugacy and the weak permutation condition and therefore, by Theorem 8, we are through.

Consider next the case $|x_1x_2\cdots x_r| = |y_1y_2\cdots y_r|$. The fact that $\mathcal{R}$ is founded guarantees that the equalities

$$x_{\sigma(1)}x_{\sigma(2)}\cdots x_{\sigma(r)} = y_{\sigma(1)}y_{\sigma(2)}\cdots y_{\sigma(r)}$$

for each $\sigma \in \mathcal{S}_r$, and

$$x_{r+\rho(1)}x_{r+\rho(2)}\cdots x_{r+\rho(n-r)} = y_{r+\rho(1)}y_{r+\rho(2)}\cdots y_{r+\rho(n-r)}$$

hold for each $\rho \in \mathcal{S}_{n-r}$. Theorem 9 completes the proof. □

Now we are prepared to construct the desired test sets. For $n = 1, 2, 3$ the set $\mathcal{T}_n$ has to be equal to $\mathcal{S}_n$. Consider the case $n = 4$. We have

$$\mathcal{CON}_4 \left\{\begin{array}{l}\mathbf{1\,2\,3\,4}\\ \mathbf{2\,3\,4\,1}\\ \mathbf{3\,4\,1\,2}\\ \mathbf{4\,1\,2\,3}\end{array}\right. \qquad \mathcal{SHU}_4 \left\{\begin{array}{l}(\mathbf{1\,2\,3\,4})\\ \ \mathbf{1\,3\,2\,4}\\ \ \mathbf{3\,1\,4\,2}\\ (\mathbf{3\,4\,1\,2})\end{array}\right. \qquad \mathcal{WPE}_4 \left\{\begin{array}{l}\underline{\mathbf{1\,2}}\,\underline{\mathbf{4\,3}}\\ \mathbf{4\,3\,1\,2}\\ \mathbf{3\,4\,2\,1}\\ \underline{\mathbf{2\,1}}\,\underline{\mathbf{3\,4}}.\end{array}\right.$$

The underlined elements show that $\mathcal{CON}_4 \cup \mathcal{SHU}_4 \cup \mathcal{WPE}_4$ is founded and repeated elements are in brackets. By Theorem 10, the set

$$\mathcal{T}_4 = \{\mathbf{1234},\ \mathbf{2341},\ \mathbf{3412},\ \mathbf{4123},\ \mathbf{1324},\ \mathbf{3142},\ \mathbf{1243},\ \mathbf{4312},\ \mathbf{3421},\ \mathbf{2134}\}$$

produces a test set for $E_4$.

Before we give the general construction formula, let us still consider separately cases $n = 5, 6$. For $n = 5$ the definitions yield

$$\mathcal{CON}_5 \left\{\begin{array}{l}\mathbf{1\,2\,3\,4\,5}\\ \mathbf{2\,3\,4\,5\,1}\\ \mathbf{3\,4\,5\,1\,2}\\ \mathbf{4\,5\,1\,2\,3}\\ \mathbf{5\,1\,2\,3\,4}\end{array}\right. \qquad \mathcal{SHU}_5 \left\{\begin{array}{l}(\mathbf{1\,2\,3\,4\,5})\\ \ \mathbf{1\,2\,4\,3\,5}\\ \ \mathbf{1\,4\,2\,5\,3}\\ \ \mathbf{4\,1\,5\,2\,3}\\ (\mathbf{4\,5\,1\,2\,3})\end{array}\right. \qquad \mathcal{WPE}_5 \left\{\begin{array}{l}\underline{\mathbf{1\,2\,3}}\,\underline{\mathbf{5\,4}}\\ \mathbf{5\,4\,1\,2\,3}\\ \underline{\mathbf{3\,1\,2}}\,\underline{\mathbf{4\,5}}\\ \underline{\mathbf{2\,1\,3}}\,\mathbf{4\,5}\\ \mathbf{4\,5\,3\,1\,2}\\ \mathbf{4\,5\,2\,1\,3}.\end{array}\right.$$

It is not difficult to verify that $\mathcal{CON}_5 \cup \mathcal{SHU}_5 \cup \mathcal{WPE}_5$ is not founded. For any $\sigma \in \mathcal{S}_3$ the set $\mathcal{T}_5$ has to contain a permutation starting by $\sigma$. The underlined elements show that we have to add for example $\underline{\mathbf{132}}\mathbf{45}$, $\underline{\mathbf{231}}\mathbf{45}$, $\underline{\mathbf{321}}\mathbf{45}$ and we get

$$\begin{aligned}\mathcal{T}_5 = \{&\mathbf{12345},\ \mathbf{23451},\ \mathbf{34512},\ \mathbf{45123},\ \mathbf{51234},\ \mathbf{12435},\ \mathbf{14253},\ \mathbf{41523},\ \mathbf{12354},\\ &\mathbf{54123},\ \mathbf{31245},\ \mathbf{21345},\ \mathbf{45312},\ \mathbf{45213},\ \mathbf{13245},\ \mathbf{23145},\ \mathbf{32145}\}\cdot\end{aligned}$$

Similarly for $n = 6$ we construct

$$\mathcal{CON}_6 \left\{\begin{array}{l}\mathbf{1\,2\,3\,4\,5\,6}\\ \mathbf{2\,3\,4\,5\,6\,1}\\ \mathbf{3\,4\,5\,6\,1\,2}\\ \mathbf{4\,5\,6\,1\,2\,3}\\ \mathbf{5\,6\,1\,2\,3\,4}\\ \mathbf{6\,1\,2\,3\,4\,5}\end{array}\right. \qquad \mathcal{SHU}_6 \left\{\begin{array}{l}(\mathbf{1\,2\,3\,4\,5\,6})\\ \ \mathbf{1\,2\,4\,3\,5\,6}\\ \ \mathbf{1\,4\,2\,5\,3\,6}\\ \ \mathbf{4\,1\,5\,2\,6\,3}\\ \ \mathbf{4\,5\,1\,6\,2\,3}\\ (\mathbf{4\,5\,6\,1\,2\,3})\end{array}\right. \qquad \mathcal{WPE}_6 \left\{\begin{array}{l}\underline{\mathbf{1\,2\,3}}\,\underline{\mathbf{6\,4\,5}}\\ \mathbf{1\,2\,3}\,\underline{\mathbf{5\,4\,6}}\\ \mathbf{6\,4\,5\,1\,2\,3}\\ \mathbf{5\,4\,6\,1\,2\,3}\\ \underline{\mathbf{3\,1\,2}}\,\underline{\mathbf{4\,5\,6}}\\ \underline{\mathbf{2\,1\,3}}\,\mathbf{4\,5\,6}\\ \mathbf{4\,5\,6\,3\,1\,2}\\ \mathbf{4\,5\,6\,2\,1\,3}.\end{array}\right.$$

One can easily see that adding elements **132 465**, **231 564** and **321 654**, we obtain a founded set

$$\mathcal{T}_6 = \{\mathbf{123456},\ \mathbf{234561},\ \mathbf{345612},\ \mathbf{456123},\ \mathbf{561234},\ \mathbf{612345},\ \mathbf{124356},\ \mathbf{142536},\ \mathbf{415263},\ \mathbf{451623},\ \mathbf{123645}, \mathbf{123546}\ \mathbf{645123},\ \mathbf{546123},\ \mathbf{312456},\ \mathbf{213456},\ \mathbf{456312},\ \mathbf{456213},\ \mathbf{132465},\ \mathbf{231564},\ \mathbf{321654}\}\cdot$$

For $n \geq 7$ define

$$\mathcal{T}_n \;=\; \mathcal{CON}_n \cup \mathcal{SHU}_n \cup \mathcal{WPE}_n \cup \mathcal{FUN}_n$$

where $\mathcal{FUN}_n$ is the recursive part of $\mathcal{T}_n$, which guarantees that $\mathcal{T}_n$ is founded. The set $\mathcal{FUN}_n$ is constructed as follows. Assume that $\mathcal{T}_k$ is given for $k = 2, 3, \ldots, n-1$. Let $\sigma_1, \sigma_2, \ldots, \sigma_p$ be a sequence of all distinct elements of

$$\mathcal{T}_{\lceil n/2\rceil} \setminus \mathcal{P}_{\,\mathrm{LEFT}}(\mathcal{CON}_n \cup \mathcal{SHU}_n \cup \mathcal{WPE}_n),$$

and similarly, let $\rho_1, \rho_2, \ldots, \rho_q$ be a sequence of all distinct elements in

$$\mathcal{T}_{\lfloor n/2\rfloor} \setminus \mathcal{P}_{\,\mathrm{RIGHT}}(\mathcal{CON}_n \cup \mathcal{SHU}_n \cup \mathcal{WPE}_n).$$

Denote $m = \max\{p, q\}$ and put $\rho_k = \rho_q$ for all $k = q+1, \ldots, m$, $\sigma_k = \sigma_p$ for all $k = p+1, \ldots, m$. For each $i \in \{1, 2, \ldots, m\}$ define $\tau_i \in \mathcal{S}_n$ by

$$\tau_i(j) = \begin{cases} \sigma_i(j), & \text{for } j = 1, 2, \ldots, \lceil n/2\rceil \\ \lceil n/2\rceil + \rho_i(j - \lceil n/2\rceil), & \text{for } j = \lceil n/2\rceil + 1, \ldots, n, \end{cases}$$

and put $\mathcal{FUN}_n = \{\tau_i \mid i = 1, \ldots, m\}$.

**Theorem 11.** *For all $n \in \mathbb{N}_+$, the set $\mathcal{T}_n$ produces a test set for $E_n$.*

*Proof.* The construction of $\mathcal{T}_n$ shows that it satisfies both conditions of Theorem 10. □

In the following we investigate the size of our test sets. To avoid confusion, write the permutations temporarily in parentheses.

Let $r = \lceil n/2\rceil$, $s = n - r = \lfloor n/2\rfloor$, $r' = \lceil r/2\rceil$ and $s' = \lceil s/2\rceil$. For $n \geq 7$ one easily sees that

$$\begin{aligned} \mathrm{Card}(\mathcal{CON}_n \cup \mathcal{SHU}_n) &= 2n - 2 \\ \mathrm{Card}(\mathcal{WPE}_n) &= 2 \cdot \lceil log\lceil n/2\rceil\rceil \;+\; 2 \cdot \lceil log\lfloor n/2\rfloor\rceil. \end{aligned} \tag{5}$$

Note that

$$(1, 2, \ldots r) \;\in\; \mathcal{CON}_{\lceil n/2\rceil} \cap \mathcal{P}_{\,\mathrm{LEFT}}(\mathcal{CON}_n)$$

and

$$(r', r'+1, \ldots, r, 1, 2, \ldots, r'-1) \;\in\; \mathcal{CON}_{\lceil n/2\rceil} \cap \mathcal{P}_{\,\mathrm{LEFT}}(\mathcal{WPE}_n).$$

Similarly

$$(1,2,\dots,s) \in \mathcal{CON}_{\lfloor n/2 \rfloor} \cap \mathcal{P}_{\text{RIGHT}}(\mathcal{CON}_n)$$

and

$$(s', s'+1, \dots, s, 1, 2, \dots, s'-1) \in \mathcal{CON}_{\lfloor n/2 \rfloor} \cap \mathcal{P}_{\text{RIGHT}}(\mathcal{WPE}_n).$$

This implies that

$$\operatorname{Card}(\mathcal{FUN}_n) \le \max\{\operatorname{Card}(\mathcal{T}_{\lfloor n/2 \rfloor}), \operatorname{Card}(\mathcal{T}_{\lceil n/2 \rceil})\} - 2. \tag{6}$$

We estimate the size of $\mathcal{T}_n$ by a function $\mathbb{F} : \mathbb{N}_+ \to \mathbb{N}_+$. Let $\mathbb{F}(1) = 1$, $\mathbb{F}(2) = 2$, $\mathbb{F}(3) = 6$, $\mathbb{F}(4) = 10$, $\mathbb{F}(5) = 17$, $\mathbb{F}(6) = 21$, and for $n \ge 7$ put

$$\mathbb{F}(n) = 2n - 2 \;+\; 2 \cdot \lceil \log \lceil n/2 \rceil \rceil + 2 \cdot \lceil \log \lfloor n/2 \rfloor \rceil \;+\; \max\{\mathbb{F}(\lfloor n/2 \rfloor), \mathbb{F}(\lceil n/2 \rceil)\} - 2.$$

From the construction of test sets for $n \le 6$, and from (5) and (6) we deduce, by induction on $n$, that

$$\operatorname{Card}(\mathcal{T}_n) \le \mathbb{F}(n)$$

and that $\mathbb{F}$ is strictly increasing:

$$\mathbb{F}(n) < \mathbb{F}(n+1)$$

for all $n \in \mathbb{N}_+$. The monotony of $\mathbb{F}$ implies

$$\max\{\mathbb{F}(\lfloor n/2 \rfloor), \mathbb{F}(\lceil n/2 \rceil)\} = \mathbb{F}(\lceil n/2 \rceil).$$

Let

$$r(n) = 2 \cdot \lceil \log \lceil n/2 \rceil \rceil + 2 \cdot \lceil \log \lfloor n/2 \rfloor \rceil - 4.$$

Then we can write

$$\mathbb{F}(n) = \; 2 \cdot n \;+\; r(n) \;+\; \mathbb{F}(\lceil n/2 \rceil)$$

for $n \ge 7$.

Let $a > 4$ be a real number. As a polynomial of $n$, the function given by

$$f(n) = \frac{(a-4) \cdot n - a}{2}$$

grows faster than $r(n)$, so there exist $n_a \in \mathbb{N}$, $n_a \ge 7$, such that for each $n \ge n_a$, we have $f(n) \ge r(n)$. This implies that

$$\mathbb{F}(n) \le 2 \cdot n + f(n) + \mathbb{F}(\lceil n/2 \rceil) = \frac{a \cdot (n-1)}{2} + \mathbb{F}(\lceil n/2 \rceil) \tag{7}$$

for each $n \ge n_a$. Let now $b_a = \mathbb{F}(n_a)$. We prove by induction that $\mathbb{F}(n) \le a \cdot n + b_a$ for each $n \in \mathbb{N}_+$. The assertion certainly holds if $n \le n_a$. Suppose $n > n_a$. Then

$$\begin{aligned} \mathbb{F}(n) \;&\le\; \frac{a \cdot (n-1)}{2} + \; \mathbb{F}(\lceil n/2 \rceil) \le \; \frac{a \cdot (n-1)}{2} + a \cdot \lceil \frac{n}{2} \rceil + b_a \\ &\le\; \frac{a \cdot (n-1)}{2} + \frac{a \cdot (n+1)}{2} + b \;=\; a \cdot n + b_a. \end{aligned}$$

Also for any real number $a'$ such that $4 < a' < a$ there is an integer $b_{a'}$, for which $\mathbb{F}(n) \leq a'n + b_{a'}$ for all $n \in \mathbb{N}_+$. This implies the existence of a number $m_a \in \mathbb{N}$ such that

$$\mathbb{F}(n) \leq a \cdot n$$

for all $n \geq m_a$. We can summarize:

**Theorem 12.** *For any real number $a > 4$ there exist integers $b_a, m_a \in \mathbb{N}$ such that*

(i) $\mathrm{Card}(\mathcal{T}_n) \;\leq\; a\, n + b_a$ *for each $n \in \mathbb{N}$;*
(ii) $\mathrm{Card}(\mathcal{T}_n) \;\leq\; a\, n$ *for each integer $n \geq m_a$.*

Let $a = 5$. It is not difficult to verify that in such a case $f(n) \geq r(n)$ for all $n \geq 37$ and by (7)

$$\mathbb{F}(n) \leq 5 \cdot \frac{(n-1)}{2} + \mathbb{F}(\lceil n/2 \rceil)$$

for $n \geq 37$. A direct computation yields following list:

| | | | | |
|---|---|---|---|---|
| $\mathbb{F}(1) = 1$ | $\mathbb{F}(2) = 2$ | $\mathbb{F}(3) = 6$ | $\mathbb{F}(4) = 10$ | $\mathbb{F}(5) = 17$ |
| $\mathbb{F}(6) = 21$ | $\mathbb{F}(7) = 28$ | $\mathbb{F}(8) = 30$ | $\mathbb{F}(9) = 41$ | $\mathbb{F}(10) = 45$ |
| $\mathbb{F}(11) = 51$ | $\mathbb{F}(12) = 53$ | $\mathbb{F}(13) = 62$ | $\mathbb{F}(14) = 64$ | $\mathbb{F}(15) = 68$ |
| $\mathbb{F}(16) = 70$ | $\underline{\mathbb{F}(17) = 85}$ | $\mathbb{F}(18) = 89$ | $\underline{\mathbb{F}(19) = 95}$ | $\mathbb{F}(20) = 97$ |
| $\underline{\mathbb{F}(21) = 105}$ | $\mathbb{F}(22) = 107$ | $\mathbb{F}(23) = 111$ | $\mathbb{F}(24) = 113$ | $\mathbb{F}(25) = 124$ |
| $\mathbb{F}(26) = 126$ | $\mathbb{F}(27) = 130$ | $\mathbb{F}(28) = 132$ | $\mathbb{F}(29) = 138$ | $\mathbb{F}(30) = 140$ |
| $\mathbb{F}(31) = 144$ | $\mathbb{F}(32) = 146$ | $\underline{\mathbb{F}(33) = 165}$ | $\mathbb{F}(34) = 169$ | $\underline{\mathbb{F}(35) = 175}$ |
| $\mathbb{F}(36) = 177$ | $\underline{\mathbb{F}(37) = 185}$ | $\mathbb{F}(38) = 187$ | $\mathbb{F}(39) = 191$ | $\mathbb{F}(40) = 193$ |

so that $\mathbb{F}(n) \leq 5n$, when $n = 1, 2, \ldots, 36$. For $n \geq 37$ proceed by induction to obtain

$$\mathbb{F}(n) \;\leq\; 5 \cdot \frac{(n-1)}{2} \;+\; \mathbb{F}(\lceil n/2 \rceil) \;\leq\; 5 \cdot \frac{(n-1)}{2} \;+\; 5 \cdot \frac{(n+1)}{2} = 5n.$$

Observe that for $n = 37$, as well as for other underlined values, the estimate is sharp: $\mathbb{F}(n) = 5n$. We can now answer a question stated in [9].

**Theorem 13.** *For each $n \in \mathbb{N}_+$, the language $E_n = c(a_1 a_2 \cdots a_n)$ possesses a test set (produced by $\mathcal{T}_n$) with the size at most $5 \cdot n$.*

The following result gives a lower bound for the size of a test set:

**Theorem 14.** *Each test set for the language $E_n = c(a_1 a_2 \cdots a_n)$ contains at least $n - 1$ elements.*

*Proof.* The assertion is certainly true for $n = 1, 2, 3$. Assume that $n > 3$. Let $S = \{w_1, w_2, \cdots, w_{n-2}\}$ be any subset of $E_n$ with cardinality $n - 2$. Suppose, without loss of generality, that $a_1$ is the last letter of $w_1$. We construct two (nonerasing) morphisms that agree on $S$, but not on $E_n$.

For each $i \in \{1, 2, \dots, n-2\}$, let

$$M_i = \{k \mid w_i = xa_kya_1z \quad \text{for some words} \quad x, y, z \in \Sigma_n^*\}\cdot$$

Thus $M_i$ is the set of all numbers $k \in \{2, 3, \dots, n\}$ such that $a_k$ precedes the symbol $a_1$ in the word $w_i$. By assumption, $M_1 = \{2, \dots, n\}$. For each $i \in \{1, 2, \dots, n-2\}$ and each $j \in \{2, 3, \dots, n\}$, let

$$r_{ij} \;=\; \begin{cases} 1, & \text{if } j \in M_i \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$\bar{v}_j = (r_{1j}, r_{2j}, \dots, r_{n-2,j})$$

for each $j \in \{2, 3, \dots, n\}$. The vectors $\bar{v}_2, \bar{v}_3, \dots, \bar{v}_n$, having only $n-2$ coordinates, are linearly dependent over $\mathbb{Q}$, the rationals. There thus exist integers $d_2, d_3, \dots, d_n$, not all zero, such that

$$d_2\bar{v}_2 + d_3\bar{v}_3 + \cdots + d_n\bar{v}_n = \bar{0}$$

with $\bar{0}$ the zero vector. Since each $r_{ij} \in \{0, 1\}$, the equality

$$\sum_{j \in M_i} d_j = 0$$

holds for each $i \in \{1, 2, \dots, n-2\}$. For each $j \in \{2, 3, \dots, n\}$, we state

$$\begin{cases} k_j \;=\; d_j + 1, \;\; l_j = 1 & \text{if } d_j > 0 \\ k_j \;=\; l_j \;\; = 1 & \text{if } d_j = 0 \\ k_j \;=\; 1, l_j \;\; = \;\; -d_j + 1 & \text{if } d_j < 0. \end{cases}$$

Then $k_2, k_2, \dots, k_n$, $l_2, l_3, \dots, l_n$ are all strictly positive integers and

$$\sum_{j \in M_i} (k_j - l_j) = 0$$

for each $i \in \{1, 2, \dots, n-2\}$. In particular, $i = 1$ implies that

$$k_2 + k_3 + \cdots + k_n = l_2 + l_3 + \cdots + l_n.$$

Let $a$, $b$ be distinct symbols and $g$ and $h$ nonerasing morphisms: $\Sigma_n^* \to \{a, b\}^*$ defined by

$$\begin{cases} g(a_1) = b \\ g(a_j) = a^{k_j} & \text{for } j = 2, 3, \dots, n; \end{cases} \qquad \begin{cases} h(a_1) = b \\ h(a_j) = b^{l_j}, & \text{for } j = 2, 3, \dots, n. \end{cases}$$

Let $i \in \{1, 2, \ldots, n-2\}$. We have

$$g(w_i) = a^{r_1} b a^{r_2}, \qquad h(w_i) = a^{s_1} b a^{s_2}$$

where

$$r_1 = \sum_{j \in M_i} k_j = \sum_{j \in M_i} l_j = s_1$$

and, since $k_2 + k_3 + \cdots + k_n = l_2 + l_3 + \cdots + l_n$, also

$$r_2 = \sum_{j \notin M_i} k_j = \sum_{j \notin M_i} l_j = s_2.$$

Thus $g(w_i) = h(w_i)$. On the other hand,

$$g(a_i a_1 \cdots a_{i-1} a_{i+1} \cdots a_n) \neq h(a_i a_1 \cdots a_{i-1} a_{i+1} \cdots a_n)$$

as soon as $d_i \neq 0$. This completes the proof. □

## 6. General commutative languages

We first contemplate the correlation between two concepts: “basis” and “test set” of a language.

**Lemma 15.** *Let $P$ be any subset of $L \subseteq \Sigma_n^*$. Length-equivalence on $P$ guarantees length–equivalence on $L$ if and only if the set $P$ contains a basis of $L$.*

The sufficiency of the condition is evident. That it is necessary is not difficult to verify either. For the proof we refer to [3] (see also [HaKo2]). The fact implies that **each test set for $L$ contains a basis of $L$.**

If the basis of $\Psi(L)$ contains the maximal possible number $n$ of vectors, then, by simple length consideration, any two morphisms, which agree lengthwise on the basis, agree at the same time lengthwise on every letter. Consequently, any basis of $L$ is also a test set. In general this certainly is not true: the set $\{a_1 a_2\}$ is one basis of $E_2 = c(a_1 a_2)$ (the other possibility is $\{a_2 a_1\}$), but the only test set for the language $E_2$ is $E_2$ itself.

In this section we show:

**Theorem 16.** *Let $L \subseteq \Sigma_n^*$ be a language and $w \in L$ a word such that* $\mathrm{alph}(w) = \mathrm{alph}(L)$ *and, for each $i$, symbol $a_i$ occurs at least twice in $w$ if it occurs at least twice in some word of $L$. Then $c(L)$ possesses a test set of the size at most* $11 \cdot n$.

*Proof.* The proof is given by the construction of the test set. Let

$$\Psi_n(w) = (d_1, d_2, \ldots, d_n),$$

with $d_1, d_2, \ldots, d_n \in \mathbb{N}$. For each $i \in \{1, 2, \ldots, n\}$ we state

$$r_i = \lceil d_i/2 \rceil, \quad \text{and} \quad r_{n+i} = d_i - r_i.$$

Let

$$v \;=\; a_1^{r_1}\, a_2^{r_2} \;\cdots\; a_n^{r_n}\, a_{n+1}^{r_{n+1}}\, a_{n+2}^{r_{n+2}} \;\cdots\; a_{2n}^{r_{2n}} \;\in\; \Sigma_{2n}^*$$

and define a projection $\pi : \Sigma_{2n}^* \to \Sigma_n^*$ by

$$\pi(a_i) = \pi(a_{n+i}) = a_i \quad \text{for} \quad i = 1, 2, \ldots, n.$$

Note that

$$\pi(c(v)) = c(w).$$

Let $B$ be a basis of $L$. Our test set will consist of the set $B$ and of projection of some permutations of the word $v$. Namely, we claim that

$$T_L \;=\; \pi\; \left( \left\{ a_{\sigma(1)}^{r_{\sigma(1)}}\; a_{\sigma(2)}^{r_{\sigma(2)}} \;\cdots\; a_{\sigma(2n)}^{r_{\sigma(2n)}} \mid \sigma \in \mathcal{T}_{2n} \right\} \right) \;\cup\; B$$

is a test set for $L$. Obviously

$$\mathrm{Card}(T_L) \leq \mathrm{Card}(\mathcal{T}_{2n}) + \mathrm{Card}(B) \leq 11 \cdot n.$$

To prove this assertion, let $g$ and $h$ be morphisms defined on $\Sigma_n^*$ that agree on $T_L$. Define another morphism $\alpha : \Sigma_{2n}^* \to \Sigma_n^*$ by

$$\alpha(a_i) \;=\; \pi(a_i)^{r_i} \quad \text{for } i = 1, 2, \ldots, 2n$$

and morphisms

$$g_1 = g \circ \alpha, \quad h_1 = h \circ \alpha$$

with the domain $\Sigma_{2n}^*$. Since $g, h$ agree on $T_L$, the morphisms $g_1, h_1$ agree on

$$T_{2n} = \{a_{\sigma(1)}\; a_{\sigma(2)} \;\cdots\; a_{\sigma(2n)} \mid \sigma \in \mathcal{T}_{2n}\}.$$

From Theorem 11 we deduce that $g_1$ and $h_1$ agree on $E_{2n}$, as $T_{2n}$ is a test set for $E_{2n}$.

Theorem 6 gives three possibilities to $g_1, h_1$.

**1.** If $g_1(a_i) = h_1(a_i)$ for each $i = 1, 2, \ldots, 2n$, then also $g(a_i) = h(a_i)$ for each $i = 1, 2, \ldots, n$, and there remains nothing to prove.

**2.** If $g_1(a_j)h_1(a_j) \neq \epsilon$ for at least three indices, then, by Theorem 6, there exist a primitive word $t$ such that $g_1(a_i), h_1(a_i) \in t^*$ for each $i = 1, 2, \ldots, 2n$. Then $g(a_i), h(a_i) \in t^*$ for $j = 1, 2, \ldots, n$, and we are through, since $g, h$ are length-equivalent on $L$, due to $B \subseteq T_L$.

**3.** Assume finally that there are indices $p, q \in \{1, 2, \ldots, 2n\}$, $p < q$ such that $g_1(a_p) \neq h_1(a_p)$, $g_1(a_q) \neq h_1(a_q)$ and $g_1(a_i)h_1(a_i) = \epsilon$ for each $i \in \{1, 2, \ldots, 2n\} \setminus \{p, q\}$.

**3.1.** If $q = n + p$, then $\pi(a_p) = \pi(a_q) = a_p$. For arbitrary $u \in E_{2n}$ we have

$$g(a_p)^{d_p} = g_1(u) = h_1(u) = h(a_p)^{d_p}$$

and $g(a_p) = h(a_p)$, a contradiction with $g_1(a_p) \neq h_1(a_p)$.

**3.2.** Therefore $q \neq n + p$ and $p, q \leq n$. By definition of $v$, the letters $a_p$ and $a_q$ occur both exactly once in $w$ and thus they have at most one occurrence in any word from $L$. Consequently $\alpha(a_p) = a_p$, $\alpha(a_q) = a_q$ and

$$g(u) = g_1(u) = h_1(u) = h(u)$$

for all $u \in L$. □

Assume in the following that each CLIP-language $c(w_0w_1^*w_2^* \cdots w_m^*)$ is effectively given (for instance, either through the vectors of its Parikh-map, or by giving the sequence of words $w_0, w_1, w_2, \dots w_n$). Next corollary is the solution of a problem left open in [9]:

**Corollary.** *Each* CLIP-*language over an alphabet of* $n$ *symbols possesses a test set of size at most* 11 $n$. *The test set can be effectively constructed.*

*Proof.* Let $L$ be a CLIP-language with $\text{alph}(L) = \Sigma_n$, $n \in \mathbb{N}_+$. Thus $L = w_0w_1^*w_2^* \cdots w_m^*$ where $m \in \mathbb{N}$ and $w_0, w_1, \dots, w_m$ are in $\Sigma_n^*$. Now it suffices to choose $w = w_0w_1^2w_2^2 \cdots w_m^2$ and use Theorem 16. The effectiveness is guaranteed by the construction of the test set for $E_n$ and by the fact that a subset of

$$\{w_0, w_0w_1, w_0w_2, \dots, w_0w_m\}$$

is the basis of $c(L)$. □

Given, for a language $L'$, (i) a word $w$, which is enough "representative" for the alphabet of $L'$; and (ii) a basis $B$ of $L'$, our result yields a straightforward although rough method to construct a test set for the commutative language $c(L')$. The test set is in fact a subset of $c(w)$ augmented by a basis of $L'$. As we have seen, the method can be applied to any CLIP-language $L = c(w_0w_1^*w_2^* \cdots w_m^*)$. If, moreover, each symbol of $\text{alph}(L)$ appears in also $w_1w_2 \cdots w_m$, we can construct a test set of size at most the dimension $\dim L$ of $L$. However, the small cardinality of the test set is recouped by the length of one of its elements.

**Theorem 17.** *Let* $L = c(w_0w_1^*w_2^* \cdots w_m^*)$ *be a* CLIP-*language such that* $\text{alph}(L) = \text{alph}(w_1w_2 \cdots w_m)$. *Then* $L$ *possesses an effectively constructible test set of size* $\dim L$.

*Proof.* Assume, without loss of generality, that $\text{alph}(L) = \Sigma_n$. By the corollary of Theorem 16, the language $L_1 = c(w_1^*w_2^* \cdots w_m^*)$ has a finite (effectively constructible) test set $T_1 = \{u_1, u_2, \dots, u_r\}$ such that $r \leq 11\,n$. Let $u = u_1u_2 \cdots u_r$ and

$$U = \{w_0, w_0u, w_0w_1, w_0w_2, \dots w_0w_m\}\cdot$$

It should be clear that $U$ is a subset of $L$ and that from $U$ we can effectively form a basis $T$ of $L$ such that $w_0u \in T$. We now verify that $T$ is a test set for $L$. Let $g$ and $h$ be morphisms that agree on $T$. Since $T$ is a basis of $L$, the morphisms $g, h$ are length-equivalent on $L$. This means that $g$ and $h$ agree on $\{w_0, u_1, u_2, \ldots, u_r\}$ and thereby also on $T_1$. Since $T_1$ is a test set of $L_1$, we deduce that the morphisms agree on $L_1$. Let $x \in L_1$ be such that $|x|_{a_i} \geq 2$ for each $i \in \{1, 2, , \ldots, n\}$. The morphisms $g$ and $h$ agree on $c(x)$. A slight modification of Theorem 6 (not all letters in $x$ are distinct) gives two possibilities: either

(i) $g(a_i) = f(a_i)$ for $i = 1, 2, \ldots, n$; or

(ii) there exists a primitive word $t$ such that $g(a_i), f(a_i) \in t^*$ for $i = 1, 2, \ldots, n$.

Because of the length-equivalence on $L$, the morphisms $g$ and $h$ agree on $L$. □

## 7. Conclusions and topics of further research

Hopefully the above considerations have given some information not only about test sets for finite commutative languages but also of the impact of conjugacy- and shuffle-like qualifications to word equations. In general, the assumption that a language is commutative (or bounded) is very restrictive; most of the languages are of neither type. One way to carry on the research is to study test sets for language families generated by some well-known set of commutative languages. Let $\mathcal{R}$ the set of all regular languages and $c(\mathcal{R}) = \{c(R) \mid R \in \mathcal{R}\}$. Furthermore, denote by $\mathcal{C}(c(\mathcal{R}))$ the smallest trio generated by the family $c(\mathcal{R})$. We end the discussion to the following:

**Research Problem.** Does there exist an effectively constructible finite test set for languages in the family $\mathcal{C}(c(\mathcal{R}))$?

It should be remembered that a trio is a family of languages closed under union, $\epsilon$-free morphism image and intersection with regular sets. The intuition says that the answer to the problem is affirmative. The construction of the test set is probably difficult.

*Acknowledgements.* The authors want to thank the anonymous referees for helpful comments and pointing out a mistake in the first version of Example 2.

## References

[1] J. Albert, *On test sets, checking sets, maximal extensions and their effective constructions.* Habilitationsschrift, Fakultät für Wirtschaftswissenschaften der Universität Karlsruhe (1968).

[2] M.H. Albert and J. Lawrence, A proof of Ehrenfeucht‘s Conjecture. *Theoret. Comput. Sci.* **41** (1985) 121-123.

[3] J. Albert and D. Wood, Checking sets, test sets, rich languages and commutatively closed languages. *J. Comput. System Sci.* **26** (1983) 82-91.

[4] Ch. Choffrut and J. Karhumäki, *Combinatorics on words*, in Handbook of Formal Languages, Vol. I, edited by G. Rosenberg and A. Salomaa. Springer-Verlag, Berlin (1997) 329-438.

[5] A. Ehrenfeucht, J. Karhumäki and G. Rosenberg, On binary equality sets and a solution to the test set conjecture in the binary case. *J. Algebra* **85** (1983) 76-85.

[6] P. Erdös and G. Szekeres, A combinatorial problem in geometry. *Compositio Math.* **2** (1935) 464-470.

[7] I. Hakala, *On word equations and the morphism equivqalence problem for loop languages.* Academic dissertation, Faculty of Science, University of Oulu (1997).

[8] I. Hakala and J. Kortelainen, On the system of word equations $x_0u_1^ix_1u_2^ix_2u_3^ix_3 = y_0v_1^iy_1^iv_2^iy_2v_3^iy_3$ $(i = 0, 1, 2, \dots)$ in a free monoid. *Theoret. Comput. Sci.* **225** (1999) 149-161.

[9] I. Hakala and J. Kortelainen, Linear size test sets for commutative languages. *RAIRO: Theoret. Informatics Appl.* **31** (1997) 291-304.

[10] T. Harju and J. Karhumäki, *Morphisms*, in Handbook of Formal Languages, Vol. I, edited by G. Rosenberg and A. Salomaa. Springer-Verlag, Berlin (1997) 439-510.

[11] M.A. Harrison, *Introduction to Formal Language Theory.* Addison-Wesley, Reading, Reading Massachusetts (1978).

[12] Š. Holub, Local and global cyclicity in free monoids. *Theoret. Comput. Sci.* **262** (2001) 25-36.

[13] J. Karhumäki and W. Plandowski, On the size of independent systems of equations in semigroups. *Theoret. Comput. Sci.* **168** (1996) 105-119.

[14] J. Kortelainen, On the system of word equations $x_0u_1^ix_1u_2^ix_2\cdots u_m^ix_m = y_0v_1^iy_1v_2^iy_2\cdots v_n^iy_n$ $(i = 1, 2, \dots)$ in a free monoid. *J. Autom. Lang. Comb.* **3** (1998) 43-57.

[15] M. Lothaire, *Combinatorics on Words.* Addison-Wesley, Reading Massachusetts (1983).

[16] A. Salomaa, The Ehrenfeucht conjecture: A proof for language theorists. *Bull. EATCS* **27** (1985) 71-82.

Communicated by J. Berstel.
Received April, 2001. Accepted December, 2001.