The commutation with codes and ternary sets of words

Juhani Karhumäki[1], Michel Latteux[2], and Ion Petre[3⋆]

[1] Department of Mathematics, University of Turku and
Turku Centre for Computer Science (TUCS)
Turku 20014, Finland
karhumak@cs.utu.fi
[2] LIFL, URA CNRS 369, Université des Sciences et Technologie de Lille
F-59655 Villeneuve d'Ascq, France
michel.latteux@lifl.fr
[3] Department of Mathematics, University of Turku and
Turku Centre for Computer Science (TUCS)
Turku 20520, Finland
ipetre@cs.utu.fi

Abstract. We prove several results on the commutation of languages. First, we prove that the largest set commuting with a given code X, i.e., its *centralizer* $\mathcal{C}(X)$, is always $\rho(X)^*$, where $\rho(X)$ is the *primitive root* of X. Using this result, we characterize the commutation with codes similarly as for words, polynomials, and formal power series: a language commutes with X if and only if it is a union of powers of $\rho(X)$. This solves a conjecture of Ratoandromanana, 1989, and also gives an affirmative answer to a special case of an intriguing problem raised by Conway in 1971. Second, we prove that for any nonperiodic ternary set of words $F \subseteq \Sigma^+$, $\mathcal{C}(F) = F^*$, and moreover, a language commutes with F if and only if it is a union of powers of F, results previously known only for ternary codes. A boundary point is thus established, as these results do not hold for all languages with at least four words.

Topics: *Regular Languages, Combinatorics on Words.*

1 Introduction

The centralizer $\mathcal{C}(L)$ of a language L is the largest set of words commuting with L, i.e., the maximal solution of the language equation $XL = LX$. As it can be readily seen, the notion of centralizer of L is well defined for any language L; as a matter of fact, $\mathcal{C}(L)$ is the union of all languages commuting with L. The best known problem with respect to the notion of centralizer is the intriguing question raised by Conway [9] more than thirty years ago.

Conway's Problem: Is it true that for any rational language, its centralizer is rational?

⋆ Current address: Department of Computer Science, Åbo Akademi University, Turku 20520, Finland, ipetre@abo.fi

Surprisingly enough, very little is known on Conway's problem. In fact, a much weaker question than Conway's is unanswered up to date: *it is not known whether or not the centralizer of any finite language is even recursively enumerable !*

A closely related problem is that of characterizing the commutation of languages. It is an elementary result of Combinatorics on Words that two words commute if and only if they have the same primitive root, see, e.g., [4] and [17]. A similar result holds also for the commutation of two polynomials and of two formal power series in noncommuting variables, with coefficients in a commutative field: two polynomials/formal power series commute if and only if they are combinations of a third one; these results are due to Bergman and Cohn, respectively, see [2], [7], and [8]. Characterizing the commutation of two languages appears to be a very difficult problem in general and certainly a similar result as above does not hold. E.g., if $X = \{a, a^3, b, ab, ba, bb, aba\}$ and $Y = X \cup \{a^2\}$, then $XY = YX$, but X and Y cannot be written as unions of powers of a third set. Nevertheless, it has been conjectured by Ratoandromanana, [21], that the commutation with a code can be characterized as in free monoids:

Conjecture 1 ([21]). For any code X and any language Y commuting with X, there is a language $R \subseteq \Sigma^+$ such that $X = R^n$ and $Y = \cup_{i \in I} R^i$, for some $n \in \mathbb{N}$, $I \subseteq \mathbb{N}$.

The first major result on the conjecture was achieved by Ratoandromanana [21], in the case of prefix codes, using ingenious (and involved) techniques on codes and prefix sets. Conjecture 1 remained open however in its general form.

We say that a language X satisfies the *BTC-property*, i.e., the Bergman-type of characterization, if the commutation with X can be characterized as in the statement of Conjecture 1, that is, similarly as in Bergman's theorem, see [13]. Thus, Conjecture 1 proposes that all codes satisfy the BTC-property. This property has been established for all singletons, as well as for all two-word (or *binary*) languages, and it has been proved that it does not hold for four-word languages, see [6]. It is a conjecture of [13] and [14] that the BTC-property holds also for all ternary sets of words.

Conjecture 2 ([13],[14]). For any ternary language $F \subseteq \Sigma^+$, the following hold:

(i) If $F \subseteq u^+$, for some primitive word $u \in \Sigma^+$, then a language X commutes with F if and only if $X = \cup_{i \in I} u^i$, for some $I \subseteq \mathbb{N}$.
(ii) If $F \not\subseteq u^+$, for all $u \in \Sigma^+$, then a language X commutes with F if and only if $X = \cup_{i \in I} F^i$, for some $I \subseteq \mathbb{N}$.

A language L is called *periodic* if $L \subseteq u^*$, for some $u \in \Sigma^*$. The first part of Conjecture 2 has been proved already in [19]: all periodic sets satisfy the BTC-property. The second part however remained open.

These three problems recently received some well deserved attention and a handful of different approaches have been investigated: the combinatorial approach, the equational method, the branching point approach, and the multiplicity approach, see [12], [15], and [20] for some surveys. Thus, it has been proved

in [19] and [6] that Conway's problem has an affirmative answer for all periodic and binary languages, respectively, and in [13] and [14] an affirmative answer has been given for ternary sets, proving also that Conjectures 1 and 2 hold for ternary codes, see also [16] for a different approach. Using different techniques, it has been proved in [11] that Conjecture 1 holds for all ω-codes. Moreover, it has been established that any code has a unique *primitive root*, a notion of [1] and [23], see [20] for details, and that two codes commute if and only if they have the same primitive root. These results of [11] also led to an affirmative answer for Conway's problem for rational ω-codes.

In this paper, we solve both Conjectures 1 and 2, characterizing the commutation with codes and with ternary sets of words. We also give an affirmative answer for Conway's problem in the case of rational codes: for any rational code X, both $\rho(X)$ and $\mathcal{C}(X)$ are rational and moreover, $\mathcal{C}(X) = \rho(X)^*$, where $\rho(X)$ denotes the primitive root of X; this is the most general result known up to date on Conway's problem. Our results also lead to a much simpler proof than that of [14] for Conway's problem for ternary languages; as a matter of fact, we give here a sharper result, proving that for any ternary language $F \subseteq \Sigma^+$, either $\mathcal{C}(F) = F^*$, or $\mathcal{C}(F) = u^*$, for some primitive word $u \in \Sigma^+$ and thus, $\mathcal{C}(F)$ is rational.

2 Preliminaries

We recall here several notions and results needed throughout the paper. For basic notions and results of Combinatorics on Words we refer to [4], [17], and [18] and for those of Theory of Codes, we refer to [3]. For details on the notion of centralizer and the commutation of languages we refer to [14], [15], and [20].

In the sequel, Σ denotes a finite alphabet, Σ^* the set of all finite words over Σ and Σ^ω the set of all (right) infinite words over Σ. We denote by 1 the empty word and by $|u|$ the length of $u \in \Sigma^*$. For a word $u \in \Sigma^*$, u^ω denotes the infinite word $uuu\ldots$, while for $L \subseteq \Sigma^*$, $L^\omega = \{u_1u_2u_3\ldots \mid u_n \in L, n \geq 1\} \subseteq \Sigma^\omega$.

We say that a word u is a *prefix* of a word v, denoted as $u \leq v$, if $v = uw$, for some $w \in \Sigma^*$. We denote $u < v$ if $u \leq v$ and $u \neq v$. We say that u and v are *prefix comparable* if either $u \leq v$, or $v \leq u$. For a word $u \in \Sigma^+$, $\mathrm{pref}_1(u)$ denotes the first letter of u. For $L \subseteq \Sigma^+$, we denote $\mathrm{pref}_1(L) = \{\mathrm{pref}_1(u) \mid u \in L\}$. The word u is a *root* of v if $v = u^n$, for some $n \in \mathbb{N}$; v is *primitive* if it has no root other than itself.

A language L of cardinal two (three) is called *binary (ternary*, resp.). L is called *periodic* if $L \subseteq u^*$, for some $u \in \Sigma^*$.

For a word u and a language L, we say that $v_1 \ldots v_n$ is an *L-factorization* of u if $u = v_1 \ldots v_n$ and $v_i \in L$, for all $1 \leq i \leq n$. For an infinite word α, we say that $v_1v_2 \ldots v_n \ldots$ is an L-factorization of α if $\alpha = v_1v_2 \ldots v_n \ldots$ and $v_i \in L$, for all $i \geq 1$. A *relation* over L is an equality $u_1 \ldots u_m = v_1 \ldots v_n$, with $u_i, v_j \in L$, for all $1 \leq i \leq m$, $1 \leq j \leq n$; the relation is *trivial* if $m = n$ and $u_i = v_i$, for all $1 \leq i \leq m$.

We say that L is a *code* if any word of Σ^* has at most one L-factorization. Equivalently, L is a code if and only if all relations over L are trivial.

Let Σ be a finite alphabet, and Ξ a finite set of unknowns in a one-to-one correspondence with a set of nonempty words $X \subseteq \Sigma^*$, say $\xi_i \leftrightarrow x_i$, for some fixed enumeration of X. A (constant-free) *equation* over Σ with Ξ as the set of unknowns is a pair $(u, v) \in \Xi^\omega \times \Xi^\omega$, usually written as $u = v$. The subset X *satisfies* the equation $u = v$ if the morphism $h : \Xi^\omega \to \Sigma^\omega$, $h(\xi_i) = x_i$, for all $i \geq 0$, verifies $h(u) = h(v)$. These notions extend in a natural way to *systems of equations*.

We define the *dependence graph* of a system of equations S, as the nondirected graph G, whose vertices are the elements of Ξ, and whose edges are the pairs $(\xi_i, \xi_j) \in \Xi \times \Xi$, with ξ_i and ξ_j appearing as the first letters of the left and right handsides of some equation of S, respectively. The following basic result on combinatorics of words ([4]), known as *Graph Lemma*, is very useful and efficient in our later considerations. Note that in Graph Lemma it is crucial that *all words are nonempty*.

Lemma 1 ([4], Graph Lemma). *Let S be a system and let $X \subset \Sigma^+$ be a subset satisfying it. If the dependence graph of S has p connected components, then there exists a subset F of cardinality p such that $X \subseteq F^*$.*

For languages L, R, we say that R is a *root* of L if $L = R^n$, for some $n \in \mathbb{N}$. We say that L is *primitive* if for any R such that $L = R^n$, $n \in \mathbb{N}$, we have $L = R$ and $n = 1$. The following is a result of [11], extending central properties of words to codes.

Theorem 1 ([11]). *Any code has a unique primitive root. Moreover, two codes commute if and only if they have the same primitive root.*

3 Ternary sets of words

As it is well-known, two words commute if and only if they have the same primitive root, or equivalently, if and only if they are powers of another word. Based on this, it is not difficult to prove, see [19], that a set of words X commutes with a word $u \in \Sigma^+$ if and only if $X \subseteq \rho(u)^*$, where $\rho(u)$ denotes the primitive root of u. Consequently, for any word $u \in \Sigma^+$, $\mathcal{C}(\{u\}) = \rho(u)^*$.

If instead of a singleton $\{u\}$, we consider a language $L \subseteq u^+$, i.e., a periodic set, then the situation is not much different than that of a singleton: a language X commutes with L if and only if $X \subseteq \rho(u)^*$ and moreover, $\mathcal{C}(L) = \rho(u)^*$.

The above results extend to binary sets of words as well. As it is well-known, a binary set F is either a periodic set, or a code. If F is a code, then it is proved in [6], see also [16] for a simpler proof, that $\mathcal{C}(F) = F^*$ and any set X commuting with F is of the form $X = \cup_{i \in I} F^i$, for some $I \subseteq \mathbb{N}$. We recall these results in the following theorem.

Theorem 2 ([6]). *Let F be a language over the alphabet Σ.*

(i) If $1 \in F$, *then* $\mathcal{C}(F) = \Sigma^*$.
(ii) If $F \subseteq u^+$, *for some primitive word* $u \in \Sigma^+$, *then* $\mathcal{C}(F) = u^*$. *Thus, a language* X *commutes with* F *if and only if* $X = \cup_{i \in I} u^i$, *for some* $I \subseteq \mathbb{N}$.
(iii) If F *is a binary code, then* $\mathcal{C}(F) = F^*$. *Moreover, a language* X *commutes with* F *if and only if* $X = \cup_{i \in I} F^i$, *for some* $I \subseteq \mathbb{N}$.

Consequently, for any binary language $F \subseteq \Sigma^+$, either $\mathcal{C}(F) = F^*$, or $\mathcal{C}(F) = u^*$, for some primitive word u. On the other hand, the above results do not hold for sets with at least four words. E.g., for $F = \{a, ab, ba, bb\}$, the set $X = F \cup F^2 \cup \{bab, bbb\}$ commutes with F, but it is not a union of powers of F, see [6]. Thus, the case of ternary sets is a boundary point for the commutation of languages. We prove in Section 3.2 that the commutation with ternary sets can be characterized as for words, polynomials, and power series, thus solving Conjecture 2. To this aim, a central result is achieved in Section 3.1, regarding the centralizer of a ternary language $F \subseteq \Sigma^+$. It has been proved in [13] and [14], using some involved techniques of equations on languages that the centralizer of any ternary set is rational. However, the general form of the centralizer is known only in the case of codes: for a ternary code F, $\mathcal{C}(F) = F^*$. It has been conjectured in [13], [14] that $\mathcal{C}(F) = F^*$, for all nonperiodic ternary sets $F \subseteq \Sigma^+$. We solve in Section 3.1 this problem, using only elementary techniques of Combinatorics on Words. In particular, this gives a new, elementary solution for Conway's problem in the ternary case, much simpler than the original solution of [14].

3.1 Conway's problem for ternary sets of words

The next result ([16]) shows that we can always reduce Conway's problem to the so-called branching sets of words. We say that a language L is *branching* if there are two words $u, v \in L$ such that u and v start with different letters. This simplification turns out to be essential in our results.The intuitive idea behind this result is that having a language L and a letter $a \in \Sigma$, Conway's problem has the same answer for languages aL and La. Thus, if all words in a language start with the same letter, we can "shift" the letter in the end, without essentially changing the problem. For any nonperiodic language, repeating this procedure a finite number of times will lead to a branching language.

Lemma 2 ([16]). *For any nonperiodic set of words* $L \subseteq \Sigma^+$, *there is a branching set of words* L' *such that* $\mathcal{C}(L)$ *is rational if and only if* $\mathcal{C}(L')$ *is rational. Moreover,* $\mathcal{C}(L) = L^*$ *if and only if* $\mathcal{C}(L') = L'^*$.

We describe in the next result the centralizer of any ternary set. This provides a tool to solve Conjecture 2 and it also gives a very simple proof for Conway's problem in the ternary case, cf. [14].

Theorem 3. *Let* F *be a ternary set of words over the alphabet* Σ.

(i) If $1 \in F$, *then* $\mathcal{C}(F) = \Sigma^*$.

(ii) If $F \subseteq u^+$, for some $u \in \Sigma^+$, then $\mathcal{C}(F) = \rho(u)^*$. Thus, a language X commutes with F if and only if $X = \cup_{i \in I} \rho(u)^i$, for some $I \subseteq \mathbb{N}$.
(iii) If F is a nonperiodic ternary set, $F \subseteq \Sigma^+$, then $\mathcal{C}(F) = F^*$.

Proof. The statements (i) and (ii) follow from Theorem 2 above. For (iii), we can assume by Lemma 2 that F is branching. Thus, let $F = \{u, u', v\}$, where $\mathrm{pref}_1(v) \notin \{\mathrm{pref}_1(u), \mathrm{pref}_1(u')\}$. The following two claims are straightforward to prove.

Claim 1. For any $1 < v' < v$ there is $\alpha \in F$ such that $v'\alpha$ is prefix incomparable with the words of F. Moreover, $v' \notin \mathcal{C}(F)$.
Claim 2. If $1 < v' < v$, then $\mathcal{C}(F) \cap v^+v' = \emptyset$.

Assume now that there is $z \in \mathcal{C}(F) \setminus F^*$. Thus, since $F^*\mathcal{C}(F) \subseteq \mathcal{C}(F)$ and $\mathrm{pref}_1(v) \notin \{\mathrm{pref}_1(u), \mathrm{pref}_1(u')\}$, it follows that $v^*z \subseteq \mathcal{C}(F) \setminus F^*$. Consequently, there is a shortest nonempty word $x \in \Sigma^+$ such that $v^*x \cap (\mathcal{C}(F) \setminus F^*) \neq \emptyset$. In particular, note that $x \notin F^*$.

Let $n \geq 0$ be such that $v^n x \in \mathcal{C}(F) \setminus F^*$. Thus, $v^{n+1}x \in F\mathcal{C}(F) = \mathcal{C}(F)F$ and so, $v^{n+1}x = \alpha\beta$, with $\alpha \in \mathcal{C}(F)$ and $\beta \in F$. Thus, either $\alpha = v^i v'$, $i \leq n$, $v' \leq v$, or $\alpha = v^{n+1}x'$, $x' \leq x$. A simple case analysis based on Claims 1 and 2 shows that this contradicts the choice of x as the shortest word $\delta \in \Sigma^+$ such that $v^*\delta \cap (\mathcal{C}(F) \setminus F^*) \neq \emptyset$. Due to space limitations, we omit here the details.

Consequently, $\mathcal{C}(F) \setminus F^* = \emptyset$, i.e., $\mathcal{C}(F) \subseteq F^*$. Since for all languages L, $L^* \subseteq \mathcal{C}(L)$, it follows that $\mathcal{C}(F) = F^*$. ■

Theorem 3 implies an affirmative answer for Conway's problem in the ternary case.

Corollary 1. *Conway's problem has an affirmative answer for all ternary sets: for any ternary set F, $\mathcal{C}(F)$ is a rational set.*

3.2 The commutation with ternary sets of words

In this section, we characterize all sets commuting with a given ternary set $F \subseteq \Sigma^+$. For periodic sets $F \subseteq u^+$, with $u \in \Sigma^+$, a language X commutes with F if and only if $X \subseteq \rho(u)^*$. For nonperiodic ternary sets $F \subseteq \Sigma^+$, we prove here that $XF = FX$ if and only if $X = \cup_{i \in I} F^i$, for some $I \subseteq \mathbb{N}$, a result previously known only for ternary codes, see [13] and [14]. In particular, based on Theorem 3, our proof for the general case turns out to be simpler than that of [13] and [14] for ternary codes, see [20] for a detailed discussion.

The case of codes can be easily solved using Theorem 3 and the following result of [6].

Theorem 4 ([6]). *For any ternary code F, the BTC-property holds for F if and only if $\mathcal{C}(F) = F^*$.*

Using different arguments, we prove next that a similar result holds for all ternary sets. Note that the ternary hypothesis is essential in this result.

Let F be a ternary set of words and X an arbitrary subset of $\mathcal{C}(F)$. We say that a word $x \in X$ satisfies the property $\mathcal{P}_F^X(x)$ if for all $n \in \mathbb{N}$, $x \in F^n$ implies $F^n \not\subseteq X$. Whenever F and X are clearly understood from the context, we simply write $\mathcal{P}(x)$ instead of $\mathcal{P}_F^X(x)$. Note that for any F and for any X, $\mathcal{P}_F^X(1)$ does not hold.

For a finite set of words F, we denote by l_F (L_F, resp.) the length of a shortest (longest, resp.) word in F.

Theorem 5. *Let $F \subseteq \Sigma^+$ be a nonperiodic three word noncode. Then the BTC-property holds for F if and only if $\mathcal{C}(F) = F^*$.*

Proof. If the BTC-property holds for F, then, since $F\mathcal{C}(F) = \mathcal{C}(F)F$, it follows that $F = R^i$ and $\mathcal{C}(F) = \cup_{j\in J} R^j$, for some set of words R and $i \in \mathbb{N}$, $J \subseteq \mathbb{N}$. It is straightforward to prove, based on cardinality arguments, that any nonperiodic ternary set of words is primitive, and so, $i = 1$ and $R = F$. Also, since $\mathcal{C}(F)$ is the largest set commuting with F, it follows that $J = \mathbb{N}$, and thus, $\mathcal{C}(F) = F^*$.

To prove the reverse implication, assume that $\mathcal{C}(F) = F^*$, and let X be a language commuting with F. We prove that in this case, $X = \cup_{i\in I} F^i$, for some $I \subseteq \mathbb{N}$. If $X = \emptyset$, then the claim is trivially true. So, let us assume that $X \neq \emptyset$.

Since F is not a code, there is a nontrivial relation over F:

$$u_1 u_{i_2} \dots u_{i_m} = u_2 u_{i_{m+1}} \dots u_{i_n}, \tag{1}$$

with $u_1 \neq u_2$ and $u_1, u_2, u_{i_k} \in F$, for all $2 \leq k \leq n$. Let u_3 be the third element of F: $F = \{u_1, u_2, u_3\}$.

Since $\mathcal{C}(F)$ includes any set commuting with F, it follows that $X \subseteq \mathcal{C}(F) = F^*$. The following claim is straightforward to prove, based on Graph Lemma.

Claim 1. Let $x \in X$ be such that $\mathcal{P}_F^X(x)$ holds. If $y \in X$, $v \in F$ are such that $u_3 x = yv$, then $\mathcal{P}_F^X(y)$ holds.

Using Claim 1, we can prove the following claim.

Claim 2. If there is $x_1 \in X$ such that $\mathcal{P}_F^X(x_1)$ holds, then $\mathcal{P}_F^X(u_3^q)$ holds, for some positive integer q.

Proof of Claim 2. Since $FX = XF$, there exist for all $n \geq 1$, $v_n \in F$ and $x_{n+1} \in X$ such that

$$u_3 x_n = x_{n+1} v_n. \tag{2}$$

Moreover, by Claim 1, $\mathcal{P}_F^X(x_n)$ holds, for all $n \geq 1$. Since $\mathcal{P}(1)$ never holds, it follows that $x_n \neq 1$, for all $n \geq 1$.

Assume now that $x_n \notin u_3^+$, for all $n \geq 1$. Using the Graph Lemma, it can be easily proved by induction on n that for all $n \geq 1$, $x_n \in u_3^{n-1} F^+$. Consequently,

$$|x_n| \geq |u_3^{n-1}| = (n-1)|u_3|,$$

for all $n \geq 1$. On the other hand, by (2),

$$|x_n| = |x_{n-1}| + |u_3| - |v_{n-1}| \leq |x_{n-1}| + |u_3| - l_F,$$

where $l_F = \min_{u \in F} |u| \geq 1$. Thus, $|x_n| \leq |x_1| + (n-1)(|u_3| - l_F)$. Altogether, we obtain that

$$(n-1)|u_3| \leq |x_n| \leq |x_1| + (n-1)(|u_3| - l_F),$$

for all $n \geq 1$. This further implies that $n \leq 1 + \frac{|x_1|}{l_F}$, for all $n \geq 1$, which is impossible.

Our assumption is thus false: there is $n \geq 1$ such that $x_n \in u_3^+$, i.e., $x_n = u_3^q$, for some positive integer q. Consequently, by Claim 1, $\mathcal{P}(u_3^q)$ holds, proving Claim 2.

Assume now that there is an $x \in X$ such that $\mathcal{P}_F^X(x)$ holds. Then, by Claim 2, there is a positive integer q such that $\mathcal{P}(u_3^q)$ holds. For such a positive integer q, consider arbitrary words $v_1, \ldots, v_q \in F$. We prove that $u_3^{q-i} v_1 \ldots v_i \in X$, for all $0 \leq i \leq q$.

Since $\mathcal{P}_F^X(u_3^q)$ holds, necessarily $u_3^q \in X$, proving the claim for $i = 0$. Let now $i \geq 0$ so that $u_3^{q-i} v_1 \ldots v_i \in X$, $0 \leq i < q$. We prove that $u_3^{q-(i+1)} v_1 \ldots v_i v_{i+1} \in X$. Since $XF = FX$, there exist $w \in F$ and $y \in X$ such that

$$u_3^{q-i} v_1 \ldots v_i \cdot v_{i+1} = w \cdot y. \tag{3}$$

If $w \neq u_3$, then by Graph Lemma on (1) and (3) we obtain that F is periodic, a contradiction. Thus, $w = u_3$ and so, $y = u_3^{q-(i+1)} v_1 \ldots v_i v_{i+1} \in X$.

Consequently, for $i = q$, we obtain that $v_1 \ldots v_q \in X$, for all $v_1, \ldots, v_q \in F$. Thus, $F^q \subseteq X$, contradicting $\mathcal{P}_F^X(u_3^q)$.

The conclusion is that there is no $x \in X$ such that $\mathcal{P}_F^X(x)$ holds. Equivalently, for any $x \in X$, there is an $m \in \mathbb{N}$ such that $x \in F^m \subseteq X$. In other words, X is of the form $X = \bigcup_{i \in I} F^i$, with $I = \{i \in \mathbb{N} \mid \exists x \in X : x \in F^i \subseteq X\}$. Thus, the BTC-property holds for F. ■

The following result is a simple consequence of Theorem 2, Theorem 5 and Theorem 3.

Theorem 6. *Let $F \subseteq \Sigma^+$ be a ternary set of words.*

(i) If F is periodic, $F \subseteq u^+$, for some $u \in \Sigma^+$, then a language X commutes with F if and only if $X \subseteq \rho(u)^*$.
(ii) If F is nonperiodic, then a language X commutes with F if and only if $X = \cup_{i \in I} F^i$, for some $I \subseteq \mathbb{N}$.

Note that the cases of ternary and periodic sets are the only known cases of non-codes for which the BTC-property holds. Moreover, the ternary case is the boundary point for the validity of the BTC-property with respect to the cardinality of a finite set, as this property does not hold for languages with at least four words.

4 The commutation with codes

The most general result known on Conjecture 1 is that of [11] where it is proved that all ω-codes satisfy the BTC-property. Using the so-called *multiplicity approach*, developed in [11], we give a complete solution for Conjecture 1, proving that all codes satisfy the BTC-property.

In the multiplicity approach, we consider an equation on languages, in this case the commutation equation, and we translate it into the corresponding equation on formal power series. We then solve the problem in terms of formal power series and finally translate the result back into sets of words. The main result used in this respect is Cohn's theorem characterizing the commutation of two formal power series.

Theorem 7 (Cohn's Theorem, [7, 8]). *Let K be a commutative field and Σ a finite alphabet. Two formal power series $p, q \in K\langle\langle\Sigma^*\rangle\rangle$ commute if and only if there are some formal power series $r \in K\langle\langle\Sigma^*\rangle\rangle$ and $p', q' \in K[[t]]$ such that $p = p'(r)$ and $q = q'(r)$, for a single variable t.*

Note, however, that in general the commutation of two sets of words does not necessarily imply the commutation of their characteristic formal power series. E.g., the sets of words $X = \{aa, ab, ba, bb, aaa\}$ and $Y = \{a, b, aa, ab, ba, bb, aaa\}$ commute, while their characteristic power series do not.

The following result of [21] is instrumental in our multiplicity approach solution to Conjecture 1.

Lemma 3 ([21]). *Let X be a code and Y a language commuting with X. For any $x \in X$ and $y \in Y$, there exist $k > 0$ and $\alpha \in X^+$ such that $(xy)^k\alpha \in X^+$.*

Recall that for any language X, its centralizer $\mathcal{C}(X)$ is the largest set commuting with X: $X\mathcal{C}(X) = \mathcal{C}(X)X$. Using Lemma 3 we can prove that for any code X, the products $X\mathcal{C}(X)$ and $\mathcal{C}(X)X$ are unambiguous. This implies that the characteristic formal power series of X and $\mathcal{C}(X)$ commute, thus effectively translating the commutation of two languages into the commutation of two formal power series. Based on this result and on Cohn's theorem, we can then characterize the commutation with codes.

Lemma 4. *For any code X, the product $X\mathcal{C}(X)$ is unambiguous.*

Proof. Assume that $X\mathcal{C}(X)$ is ambiguous, i.e., there are $x, y \in X$, $u, v \in \mathcal{C}(X)$ such that $xu = yv$ and $x \neq y$. By Lemma 3, there exists $\alpha \in X^+$ such that $(xu)^k\alpha \in X^+$. Let $z = (xu)^k\alpha$. Then

$$z^\omega = ((xu)^k\alpha)^\omega = (x(ux)^{k-1}u\alpha)^\omega = x((ux)^{k-1}u\alpha x)^\omega = x(wx)^\omega,$$

where $w = (ux)^{k-1}u\alpha \in \mathcal{C}(X)$. As it is easy to see, for any $\delta \in \mathcal{C}(X)$ and any $t \in X$, $(\delta t)^\omega \in X^\omega$ and so, $(wx)^\omega \in X^\omega$. Consequently, $z^\omega \in xX^\omega$.

Analogously, $z^\omega = ((yv)^k\alpha)^\omega \in yX^\omega$ and so, z^ω has two different X-factorizations. Now, a result of [10] states that X is a code if and only if for any $\gamma \in X^+$, γ^ω has exactly one X-factorization. This leads to a contradiction. ∎

By symmetry, it follows from Lemma 3 and Lemma 4 that for any code X, the product $\mathcal{C}(X)X$ is also unambiguous.

The following result gives the exact form of the centralizer in case of codes, the last step in solving Conjecture 1.

Theorem 8. *For any code X, $\mathcal{C}(X) = \rho(X)^*$, where $\rho(X)$ denotes the primitive root of X.*

Proof. By Lemma 4, both products $X\mathcal{C}(X)$ and $\mathcal{C}(X)X$ are unambiguous. Thus, if r_X is the characteristic formal power series of X and $r_{\mathcal{C}(X)}$ that of $\mathcal{C}(X)$, it follows that $r_X r_{\mathcal{C}(X)} = r_{\mathcal{C}(X)} r_X$. By Cohn's theorem, this implies that both r_X and $r_{\mathcal{C}(X)}$ can be expressed as combinations of another series r. If R is the support of r, then we obtain that $X = \cup_{i\in I} R^i$ and $\mathcal{C}(X) = \cup_{j\in J} R^j$, for some $I, J \subseteq \mathbb{N}$. However, X is a code, and so, I is a singleton: $X = R^i$, $i \in N$. It then follows from [11] and [21] that R is a code commuting with X and so, $R = \rho(X)^k$, for some $k \in \mathbb{N} \setminus \{0\}$. Thus, $\mathcal{C}(X) = \cup_{j'\in J'} \rho(X)^{j'}$, $J' \subseteq \mathbb{N}$. Since $\mathcal{C}(X)$ is the largest set commuting with X, it follows then that $\mathcal{C}(X) = \rho(X)^*$. ■

Corollary 2. *Conway's problem has affirmative answer for all rational codes: if X is a rational code, then both $\rho(X)$ and $\mathcal{C}(X)$ are rational, and $\mathcal{C}(X) = \rho(X)^*$.*

Proof. As proved in [22], see also [5], for any rational language R, if $R = L^n$, for some $L \subseteq \Sigma^*$, then there is a rational language S such that $L \subseteq S$ and $R = S^n$. Thus, there is a rational language S such that $\rho(X) \subseteq S$ and $X = \rho(X)^n = S^n$, for some $n \geq 1$. But then, S is a code and $XS = SX$. By Theorem 1, X and S have the same primitive root: $S = \rho(X)^m$ and so, $\rho(X)^n = \rho(X)^{mn}$, i.e., $m = 1$. Consequently, $\rho(X)$ is rational and so is $\mathcal{C}(X) = \rho(X)^*$. ■

Based on Theorem 8, we can solve now Conjecture 1.

Theorem 9. *The BTC-property holds for all codes. Equivalently, for any code X, a language Y commutes with X if and only if there is $I \subseteq \mathbb{N}$ such that $Y = \cup_{i\in I} \rho(X)^i$.*

Proof. Since $XY = YX$, it follows that $Y \subseteq \mathcal{C}(X) = \rho(X)^*$. To prove the claim of the theorem, it is enough to prove that for any $n \geq 0$, if $Y \cap \rho(X)^n \neq \emptyset$, then $\rho(X)^n \subseteq Y$.

Let $u_1, \ldots, u_n \in \rho(X)$ such that $u_1 \ldots u_n \in Y$ and let $\alpha_1, \ldots, \alpha_n$ be arbitrary elements of $\rho(X)$. Let also $X = \rho(X)^k$, $k \geq 1$. Then, since $X^n Y = YX^n$ and $(\alpha_1 \ldots \alpha_n)^k \in \rho(X)^{nk} = X^n$, it follows that $u_1 \ldots u_n (\alpha_1 \ldots \alpha_n)^k \in X^n Y = \rho(X)^{kn} Y$. Since $Y \subseteq \rho(X)^*$ and $\rho(X)$ is a code, this can only lead to a trivial $\rho(X)$-relation, i.e., $\alpha_1 \ldots \alpha_n \in Y$. Thus, $\rho(X)^n \subseteq Y$, proving the claim. ■

5 Conclusions

The commutation of languages turns out to be a very challenging problem in general and it is difficult to even conjecture a possible general characterization.

We proved however, that the commutation with a code can be characterized similarly as for words: for any code X, a language L commutes with X if and only if $L = \cup_{i \in I} \rho(X)^i$, for some $I \subseteq \mathbb{N}$, where $\rho(X)$ is the primitive root of X, thus solving an old conjecture of Ratoandromanana [21]. A similar characterization holds also for the commutation with periodic, binary, and - as we proved here - ternary sets of words, but not for languages with at least four words; this solves a conjecture of [13].

The intriguing problem of Conway [9], asking if the centralizer of a rational language is rational, still remains far from being solved. We proved here that the centralizer of any rational code X is rational and in fact, $\mathcal{C}(X) = \rho(X)^*$, an interesting connection between the notions of centralizer and primitive root. Except the "simple" cases of periodic, binary, and ternary languages - solved here with elementary arguments -, nothing else is known on Conway's problem. In fact, it turns out that despite the various techniques we have developed for this problem, see [20], the only cases where we could solve it are those when the centralizer is "trivial", i.e., $\mathcal{C}(X) = \rho(X)^*$ - the proofs however are quite involved in many case, see, e.g., [14] and [21]. We conclude by recalling once again this remarkable problem, as well as some of its possibly simpler variants.

Problem 1 (Conway's Problem). Is it true that for any rational language, its centralizer is rational ?

Problem 2. Is the centralizer of a rational set always: a) recursively enumerable, b) recursive, c) rational ?

Problem 3. Is the centralizer of a finite set always: a) recursively enumerable, b) recursive, c) rational, d) finitely generated ?

References

1. Autebert, J.M., Boasson, L., Latteux, M.: Motifs et bases de langages, *RAIRO Inform. Theor.*, 23(4) (1989) 379-393.
2. Bergman, G.: Centralizers in free associative algebras, *Transactions of the American Mathematical Society* 137 (1969) 327–344.
3. Berstel, J., Perrin, D.: *Theory of Codes*, Academic Press, New York (1985).
4. Choffrut, C., Karhumäki, J.: Combinatorics of Words. In Rozenberg, G., Salomaa, A. (eds.), *Handbook of Formal Languages*, Vol. 1, Springer-Verlag (1997) 329-438.
5. Choffrut, C., Karhumäki, J.: On Fatou properties of rational languages, in Martin-Vide, C., Mitrana, V. (eds.), *Where mathematics, Computer Science, Linguistics and Biology Meet*, Kluwer, Dordrecht (2000).
6. Choffrut, C., Karhumäki, J., Ollinger, N.: The commutation of finite sets: a challenging problem, *Theoret. Comput. Sci.*, 273 (1-2) (2002) 69–79.
7. Cohn, P.M.: Factorization in noncommuting power series rings, *Proc. Cambridge Philos. Soc.* 58 (1962) 452–464.
8. Cohn, P.M.: Centralisateurs dans les corps libres, in Berstel, J. (ed.), *Séries formelles*, Paris, (1978) 45–54.
9. Conway, J.H.: *Regular Algebra and Finite Machines*, Chapman Hall (1971).

10. Devolder, J., Latteux, M., Litovsky, I., Staiger, L.: Codes and infinite words, *Acta Cybernetica* 11 (1994) 241–256.
11. Harju, T., Petre, I.: On commutation and primitive roots of codes, submitted. A preliminary version of this paper has been presented at WORDS 2001, Palermo, Italy.
12. Karhumäki, J.: Challenges of commutation: an advertisement, in *Proc. of FCT 2001*, LNCS 2138, Springer (2001) 15–23.
13. Karhumäki, J., Petre, I.: On the centralizer of a finite set, in *Proc. of ICALP 2000*, LNCS 1853, Springer (2000) 536–546.
14. Karhumäki, J., Petre, I.: Conway's Problem for three-word sets, *Theoret. Comput. Sci.*, 289/1 (2002) 705–725.
15. Karhumäki, J., Petre, I.: Conway's problem and the commutation of languages, *Bulletin of EATCS* 74 (2001) 171–177.
16. Karhumäki, J., Petre, I.: The branching point approach to Conway's problem, LNCS 2300, Springer (2002) 69–76.
17. Lothaire, M.: *Combinatorics on Words* (Addison-Wesley, Reading, MA., (1983).
18. Lothaire, M.: *Algebraic Combinatorics on Words* (Cambridge University Press), (2002).
19. Mateescu, A., Salomaa, A., Yu, S.: On the decomposition of finite languages, TUCS Technical Report 222, http://www.tucs.fi/ (1998).
20. Petre, I.: *Commutation Problems on Sets of Words and Formal Power Series*, PhD Thesis, University of Turku (2002).
21. Ratoandromanana, B.: Codes et motifs, *RAIRO Inform. Theor.*, 23(4) (1989) 425-444.
22. Restivo, A.: Some decision results for recognizable sets in arbitrary monoids, in *Proc. of ICALP 1978*, LNCS 62 Springer (1978) 363–371.
23. Shyr, H.J.: *Free monoids and languages*, Hon Min Book Company, (1991).