NOVI SAD J. MATH.
VOL. 30, NO. 2, 2000, 165-170

n-WORDS OVER ANY ALPHABET WITHOUT SUBWORD $a\underbrace{b\ldots b}_{k-1}a$ FOR FIXED a, b AND k

Rade Doroslovački[1], **Olivera Marković**[2]

Abstract. The paper gives a special construction of those words of length n over any alphabet $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ in which the subword $a\underbrace{b\ldots b}_{k-1}a$ is forbidden for some natural number k where the letters a and b are different and fixed from the alphabet $\mathcal{A}$. This construction gives the number of all these words. The number of such words is counted in two diferent ways, which gives some new combination identities.

AMS Mathematics Subject Classification (1991): 0840

Key words and phrases: subword

1. Definitions and notations

The set of first m natural numbers is denoted by N_m. Hence $N_m = \{1, 2, .., m\}$. The set $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ is called an alphabet if for all $i \in N_m$, α_i is an arbitrary symbol. The elements of $\mathcal{A}$ are then the letters of the alphabet and $\mathcal{A}$ is an m-letter alphabet.

If $x \in \mathcal{A}^n$ i.e. if $x = (x_1, x_2, \ldots, x_n)$ is an ordered n-tuple with components from $\mathcal{A}$, we say that x is a word of length n over the alphabet $\mathcal{A}$. For the sake of brevity we shall write $(x_1, x_2, \ldots, x_n)$ as $x_1 x_2 \ldots x_n$.

A subword of length k of the word $x_1 x_2 \ldots x_n$ is any word $x_s x_{s+1} \ldots x_{s+k-1}$ where $s \in N_{n-k+1}$ and $k \in N_n$.

The only element of $\mathcal{A}^0$ is the empty string i.e. the string of length 0.

The set of all the words of finite length over the alphabet $\mathcal{A}$ will be denoted by $\mathcal{A}^*$ i.e.

$$\mathcal{A}^* = \bigcup_{i \geq 0} \mathcal{A}^i$$

The number of ways in which a subword y occurs in a word $x \in \mathcal{A}^*$ is denoted by $l_y(x)$. In particular, the number of ways in which the letter $\alpha \in \mathcal{A}$ occurs in the word $x \in \mathcal{A}^*$ is $l_\alpha(x)$.

[1]Department of Mathematics, Faculty of Engineering University of Novi Sad, 21000 Novi Sad, Trg Dositeja Obradovića 6, Yugoslavia

[2]Faculty of Education University of Kragujevac, 31000 Užice, Trg Svetog Save 36, Yugoslavia

If S is a set, then $|S|$ is the cardinality of S. By $\lceil x\rceil$ and $\lfloor x\rfloor$ we denote the smallest integer $\geq x$ and the largest integer $\leq x$, respectively and

$$[x] = \begin{cases} \lfloor x\rfloor & if\ |\lfloor x\rfloor - x| \leq 0.5 \\ \lceil x\rceil & if\ |\lceil x\rceil - x| < 0.5 \end{cases}$$

i.e. $[x]$ is the nearest integer to x.

2 . Results and discussion

We will construct and determine the cardinal numbers of the set of words

$$Q(k,m,n) = \{\boldsymbol{x_n} = x_1 \ldots x_n \in \mathcal{A}^n |\ (\forall s \in N_{n-k})\ x_s x_{s+1} \ldots x_{s+k} \neq a\underbrace{b\ldots b}_{k-1}a\}$$

for each natural number k and fixed letters a and b from the alphabet $\mathcal{A}$. Let $m > 2$ (for $m = 2$ see [2]).

For $k = 1$ it is known [3] that

$$Card\,Q(1,m,n) = \sum_{i=0}^{\lfloor \frac{n}{2}\rfloor} \binom{n-i+1}{i} (m-1)^{n-i}$$

where

$$Q(1,m,n) = \{\boldsymbol{x_n} | \boldsymbol{x_n} = x_1 \ldots x_n \in \mathcal{A}^n,\ (\forall s \in N_{n-1})\ x_s x_{s+1} \neq aa\}$$

For $k = 2$ is shown in [4] that the following theorem is valid.

Theorem 1.

$$Card\,Q(2,m,n) = (m-1)^n + \sum_{i=1}^{n}\sum_{j=0}^{i-1}\sum_{l=0}^{\lfloor\frac{n-i-j}{2}\rfloor} \binom{i-1}{j}\binom{i-1-j}{l}$$

$$\binom{n-i-j-l+1}{l+1}(m-1)^{n-i-j}(m-2)^j = \left[\frac{m\alpha^2+m-1}{m\alpha^2-2\alpha+3(m-1)}\alpha^n\right]$$

where α is the unique real root of the equation $x^3 - mx^2 + x - (m-1) = 0$ and

$$Q(2,m,n) = \{\boldsymbol{x_n} | \boldsymbol{x_n} = x_1 \ldots x_n \in \mathcal{A}^n,\ (\forall s \in N_{n-2})\ x_s x_{s+1} x_{s+2} \neq aba\}$$

Now let $k \geq 3$.

Theorem 2.

$$Card\,Q(k,m,n)=(m-1)^n+\sum_{i=1}^{n}\sum_{j_1=0}^{i-1}\cdots\sum_{j_{k-1}=0}^{\lfloor\frac{i-1-s_{k-2}}{k-1}\rfloor}\sum_{l=0}^{\lfloor\frac{i-1-s_{k-1}}{k}\rfloor}\prod_{p=0}^{k-2}\binom{i-1-s_p}{j_{p+1}}$$

$$\binom{i-1-s_{k-1}}{l}\binom{n-i-S_{k-1}-(k-1)l+1}{l+1}(m-1)^{n-i-(k-1)j_{k-1}}((m-1)^{k-1}-1)^{j_{k-1}}$$

where $s_p=j_1+j_2+\ldots+j_p,\ \ s_0=0,\ \ S_p=j_1+2j_2+\ldots+pj_p$ *and*

$$Q(k,m,n)=\{\boldsymbol{x_n}|\boldsymbol{x_n}=x_1\ldots x_n\in\mathcal{A}^n,(\forall s\in N_{n-k})\ x_sx_{s+1}\ldots x_{s+k}\neq a\underbrace{b\ldots b}_{k-1}a\}$$

Proof.

We partition the set $Q(k,m,n)$ into the subsets $Q^i(k,m,n)$ which contain exactly i elements a i.e. $Q^i(k,m,n)=$

$$=\{\boldsymbol{x_n}|\boldsymbol{x_n}\in\mathcal{A}^n\ \wedge\ (\forall s\in N_{n-k})\ x_sx_{s+1}\ldots x_{s+k}\neq a\underbrace{b\ldots b}_{k-1}a\ \wedge\ l_a(\boldsymbol{x_n})=i\}$$

We will construct words from $Q^i(k,m,n)$ in the following way. First we write i elements a and then we write one of the elements from the set $\{p_1,p_2,\ldots,p_{k-2},q,p_k,\lambda\}$ on $i-1$ places between i elements a, where $p_s, s\in\{1,2,\ldots,k-2,k\}$ arbitrary words of length s over the alphabet $\mathcal{A}$, q arbitrary words of length $k-1$ (except $\underbrace{b\ldots b}_{k-1}$) over $\mathcal{A}$, and λ is an empty letter. The letter λ is a letter with the property that if λ is written between two elements a then actually nothing is written.

Let $j_m, m\in\{1,\ldots,k-2\}$ be the number of elements p_m, j_{k-1} the number of elements q and l number of elements p_k. We choose j_1 places from $i-1$ places for the element p_1 , j_2 places from $i-1-j_1$ places for the element $p_2,\ldots,j_{k-2}$ places from $i-1-s_{k-3}$ places for the element p_{k-2}, j_{k-1} places from $i-1-s_{k-2}$ places for the element q and l places from $i-1-s_{k-1}$ places for p_k. It can be done in

$$\prod_{p=0}^{k-2}\binom{i-1-s_p}{j_{p+1}}\binom{i-1-s_{k-1}}{l}$$

different ways, where $s_p=j_1+j_2+\ldots+j_p$ and $s_0=0$.

There remains to write $n-i-S_{k-1}-kl$ letters from the alphabet $\mathcal{A}$ we put on l places which already contain elements p_k, as well as in the places in front of and behind the word, that is into $l+2$ places in all. It can be done in

$$\binom{n-i-S_{k-1}-(k-1)l+1}{l+1}$$

different ways, where $S_k = j_1 + 2j_2 + \ldots + kj_k$.

The arrangement of $m-1$ letters in j_1 places can be done in $(m-1)^{j_1}$ ways, the arrangement of all words of length two over the alphabet $\mathcal{A}$ in j_2 places can be done in $(m-1)^{2j_2}$ ways,..., the arrangement of all words of length $k-2$ over the same alphabet in j_{k-2} places can be done in $(m-1)^{(k-2)j_{k-2}}$ ways.

The arrangement of words of length $k-1$ over the alphabet $\mathcal{A}$ in j_{k-1} places (except $\underbrace{b\ldots b}_{k-1}$) can be done in $((m-1)^{k-1}-1)^{j_{k-1}}$ ways, and finally, the arrangement of $m-1$ letters in $n-i-j_1-2j_2-\ldots-(k-1)j_{k-1}$ places can be done in $(m-1)^{n-i-j_1-2j_2-\ldots-(k-1)j_{k-1}}$ ways.

Since,

$$(m-1)^{j_1}(m-1)^{2j_2}(m-1)^{3j_3}\ldots(m-1)^{(k-2)j_{k-2}}\left((m-1)^{k-1}-1\right)^{j_{k-1}}\cdot$$

$$\cdot(m-1)^{n-i-j_1-2j_2-\ldots-(k-1)j_{k-1}} = (m-1)^{n-i-(k-1)j_{k-1}}((m-1)^{k-1}-1)^{j_{k-1}}$$

we have

$$Card^i\, Q(k,m,n) = \sum_{j_1=0}^{i-1}\cdots\sum_{j_{k-1}=0}^{\lfloor\frac{i-1-S_{k-2}}{k-1}\rfloor}\sum_{l=0}^{\lfloor\frac{i-1-S_{k-1}}{k}\rfloor}\prod_{p=0}^{k-2}\binom{i-1-s_p}{j_{p+1}}\binom{i-1-s_{k-1}}{l}$$

$$\binom{n-i-S_{k-1}-(k-1)l+1}{l+1}(m-1)^{n-i-(k-1)j_{k-1}}((m-1)^{k-1}-1)^{j_{k-1}}$$

When $i=0$, the total number of words that do not contain the forbidden subword $a\underbrace{b\ldots b}_{k-1}a$ equals the number of all the words of n-length of $m-1$ elements, i.e. $(m-1)^n$.

Since,

$$Card\, Q(k,m,n) = (m-1)^n + \sum_{i=1}^{n} Card\, Q^i(k,m,n)$$

the proof is completed. □

Corollary 1.

$$Card\, Q(4,3,n) = 2^n + \sum_{i=1}^{n}\sum_{j_1=0}^{i-1}\sum_{j_2=0}^{i-1-j_1}\sum_{j_3=0}^{i-1-j_1-2j_2}\sum_{l=0}^{\lfloor\frac{i-1-j_1-2j_2-3j_3}{4}\rfloor}$$

$$\binom{i-1}{j_1}\binom{i-1-j_1}{j_2}\binom{i-1-j_1-j_2}{j_3}\binom{i-1-j_1-j_2-j_3}{l}$$

$$\binom{n-i-j_1-2j_2-3j_3-3l+1}{l+1}\cdot 2^{n-i-3j_3}\cdot 7^{j_3} \quad \square$$

Theorem 3.

$$Card\,Q(4,3,n) = \left[\frac{3\alpha^4+2}{3\alpha^4-4\alpha+10}\alpha^n\right]$$

where $\alpha = 2,9876$, is the real root of the equation $x^5 - 3x^4 + x - 2 = 0$.

Proof.

We will observe the alphabet $\mathcal{A} = \{a, b, c\}$ and the forbidden subword $abbba$ Words $\boldsymbol{x_n} \in Q(4,3,n)$ are obtained from the words $\boldsymbol{x_{n-1}} \in Q(4,3,n-1)$ by adding a, b or c in front of them. Let $\boldsymbol{x_{n-1}} \in Q(4,3,n-1)$, $\boldsymbol{x_{n-4}} \in Q(4,3,n-4)$ and $\boldsymbol{x_{n-5}} \in Q(4,3,n-5)$. Then $b\boldsymbol{x_{n-1}} \in Q(4,3,n)$, $c\boldsymbol{x_{n-1}} \in Q(4,3,n)$, $abbbb\boldsymbol{x_{n-5}} \in Q(4,3,n)$, $abbbc\boldsymbol{x_{n-5}} \in Q(4,3,n)$, $abbba\boldsymbol{x_{n-5}} \notin Q(4,3,n)$, which means that $abbb\boldsymbol{x_{n-4}} \in Q(4,3,n)$ iff $\boldsymbol{x_{n-4}}$ begins either in the letter b or in the letter c.

This implies the recurrence relation

$$Card\,Q(4,3,n) = 3\,Card\,Q(4,3,n-1) - Card\,(4,3,n-4) + 2\,Card\,Q(4,3,n-5)$$

whose characteristic equation is

$$x^5 - 3x^4 + x - 2 = 0$$

This equation has one real root $\alpha \approx 2,9876$ and four non-real roots β, γ, δ and ϵ whose modules are less than 1.

Now we have the explicite formula

$$Card\,Q(4,3,n) = C_1\alpha^n + C_2\beta^n + C_3\gamma^n + C_4\delta^n + C_5\epsilon^n$$

where the constants C_i, $1 \le i \le 5$ are determined according to the initial conditions $Card\,Q(4,3,0) = 1$, $Card\,Q(4,3,1) = 3$, $Card\,Q(4,3,2) = 9$, $Card\,Q(4,3,3) = 27$, $Card\,Q(4,3,4) = 81$.

By solving the corresponding system of equation we get

$$C_1 = \frac{3\alpha^4+2}{3\alpha^4-4\alpha+10}$$

Since $|\beta| = |\gamma| < 1$ and $|\delta| = |\epsilon| < 1$ we have $\lim_{n\to\infty}\beta^n = 0$, $\lim_{n\to\infty}\gamma^n = 0$, $\lim_{n\to\infty}\delta^n = 0$, $\lim_{n\to\infty}\epsilon^n = 0$. Because of that it follows that

$$Card\,Q(4,3,n) = \left[\frac{3\alpha^4+2}{3\alpha^4-4\alpha+10}\alpha^n\right] \quad \square$$

Using a procedure similar to the one applied in the previous theorem, we get the recurrence formula for $Card\,Q(k,m,n)$:

$$Card\,Q(k,m,n) = m\,Card\,Q(k,m,n-1) - Card\,(k,m,n-k) +$$
$$+(m-1)\,Card\,Q(k,m,n-k-1)$$

whose characteristic equation is

$$x^{k+1} - mx^k + x - (m-1) = 0$$

Corollary 2.

$$Card\,Q(3,5,n) = 4^n + \sum_{i=1}^{n}\sum_{j_1=0}^{i-1}\sum_{j_2=0}^{i-1-j_1}\sum_{l=0}^{\lfloor\frac{n-i-j_1-2j_2}{3}\rfloor}\binom{i-1}{j_1}\binom{i-1-j_1}{j_2}$$

$$\binom{i-1-j_1-j_2}{l}\binom{n-i-j_1-2j_2-2l+1}{l+1}\cdot 4^{n-i-2j_2}\cdot 15^{j_2}$$

$$= \left[\frac{5\alpha^3+4}{5\alpha^3-3\alpha+16}\alpha^n\right]$$

where $\alpha \approx 4,99203$ is one of two real roots of the equation $x^4-5x^3+x-4=0$.

References

[1] Doroslovački, R., The set of all words over alphabet $\{0,1\}$ of length n with the forbidden subword $11\ldots1$, Rev. of Res., Fac. of Sci. math. ser., Novi Sad, Vol. 14, Num. 2 (1984), 167–173.

[2] Doroslovački, R., Binary sequences without $0\overbrace{11\ldots11}^{k-1}0$ for fixed k, Matematički vesnik 46 (1994),93-98, Beograd.

[3] Doroslovački, R., The set of all the words of length n over any alphabet with the forbidden subword $a\ldots a$ where the letter a is fixed, Univ. u Novom Sadu Zb. Rad. Prirod.-Mat. Fak. Ser. Mat. 23, 1, 227-234, (1993).

[4] Marković O., M.Sc. thesis, University of Novi Sad (1998).

Received by the editors December 22, 1999.