Novi Sad J. Math.
Vol. 30, No. 2, 2000, 159-163

n-WORDS OVER ANY ALPHABET WITH FORBIDDEN ANY 3-SUBWORDS

Rade Doroslovački[1], Olivera Marković[2]

Abstract. We determine the cardinal numbers of the sets of words of n-length over any m-member alphabet in which any fixed subword of length three is forbidden. Cardinal numbers can be calculated in two ways, which for accurately determined values of the parameter m consequently gives some new combination identities and some interesting limits.

AMS Mathematics Subject Classification (1991): 0840

Key words and phrases: Forbidden subwords, $m-$member alphabet,

1. Definitions and notations

The set of first k natural numbers is denoted by N_k i.e. $N_k = \{1, 2, .., k\}$. The set $\mathcal{A} = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ is called an alphabet if for all $i \in N_m$, α_i are arbitrary symbols. The elements of A are then the letters of the alphabet and $\mathcal{A}$ is m-letter alphabet. If $x \in \mathcal{A}^n$ i.e. if $x = (x_1, x_2, \dots, x_n)$ is an ordered n-tuple with components from $\mathcal{A}$, we say that x is a word of length n over the alphabet A. For the sake of brevity we shall write $(x_1, x_2, \dots, x_n)$ as $x_1x_2 \dots x_n$. A subword of length k of the word $x_1x_2 \dots x_n$ is any word $x_sx_{s+1} \dots x_{s+k-1}$ where $s \in N_{n-k+1}$ and $k \in N_n$. A subword $y_1y_2 \dots y_k$ is good iff $y_1y_2 \dots y_s \neq y_{k-s+1}y_{k-s+2} \dots y_k$ for each natural number $s < k$. The only element of $\mathcal{A}^0$ is the empty string i.e. the string of length 0. The set of all 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 $l_\alpha(x)$. 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 greatest integer $\leq x$, respectively. If $|\lfloor x \rfloor - x| \leq 0.5$ then $[x] = \lfloor x \rfloor$ and if $|\lceil x \rceil - x| < 0.5$ then $[x] = \lceil x \rceil$. i.e. $[x]$ is the nearest integer to x.

2. Results and discussion

We will observe all the words of n-length over the alphabet $\mathcal{A} = \{\alpha_1, \dots, \alpha_m\}$, $m \in N$ and $m > 2$ (for $m = 2$ see [3, 4, 5]) in which any fixed subword of length

[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

three is forbidden. Let us denote the set of all the forbidden words of length three with $M = \{abc|a,b,c \in \mathcal{A}\}$ and let us define the relation ρ in the set M in the following way: $a_1b_1c_1 \ \rho \ a_2b_2c_2 \Leftrightarrow |M_1| = |M_2|$ where
$M_1 = \{\boldsymbol{x_n}|\boldsymbol{x_n} = x_1, x_2 \dots x_n \in \mathcal{A}^n \wedge (\forall i \in N_{n-2})(x_ix_{i+1}x_{i+2} \neq a_1b_1c_1)\}$,
$M_2 = \{\boldsymbol{x_n}|\boldsymbol{x_n} = x_1, x_2 \dots x_n \in \mathcal{A}^n \wedge (\forall i \in N_{n-2})(x_ix_{i+1}x_{i+2} \neq a_2b_2c_2)\}$
We will notice that the relation ρ is an equivalence relation and that there are exactly three equivalence classes: $A = \{aaa|a \in \mathcal{A}\}$,
$B = \{aab, baa, abc|a,b,c \in \mathcal{A} \wedge a \neq b \neq c \neq a\}$, $C = \{aba|a,b \in \mathcal{A} \wedge a \neq b\}$.

It is obvious that $\{A,B,C\}$ is partitions of the set $\mathcal{A}^3 = \{a_1,\dots,a_m\}^3$ and because of that $|A| + |B| + |C| = |\mathcal{A}^3|$ i.e.

$$m + \Big(2m(m-1) + m(m-1)(m-2)\Big) + m(m-1) = m^3.$$

First of all we will observe all n-length words over the $\mathcal{A}$-alphabet in which the forbidden subword is aaa, where a is an arbitrary but fixed letter from $\mathcal{A}$. If this set is marked with $P_A(m,n)$, it is known [5] that the following theorem is valid:

Theorem 1. $|P_A(m,n)| =$
$$= \sum_{i=0}^{\lceil \frac{2n}{3} \rceil} \sum_{j=0}^{\lfloor \frac{i}{2} \rfloor} \binom{n-i+1}{i-j}\binom{i-j}{j}(m-1)^{n-i} = \left[\frac{m\alpha^2+m\alpha+(m-1)}{(m-1)\alpha^2+2(m-1)\alpha+3(m-1)}\alpha^n\right]$$
where α is the real root of the equation $x^3-(m-1)x^2-(m-1)x-(m-1)=0$.

Corollary 1.
$$\lim_{n\to\infty} \frac{1}{n^\alpha} \sum_{i=0}^{\lceil \frac{2n}{3} \rceil} \sum_{j=0}^{\lfloor \frac{i}{2} \rfloor} \binom{n-i+1}{i-j}\binom{i-j}{j}(m-1)^{n-i} = \frac{m\alpha^2+m\alpha+(m-1)}{(m-1)\alpha^2+2(m-1)\alpha+3(m-1)}$$
where α is the real root of the equation $x^3-(m-1)x^2-(m-1)x-(m-1)=0$.

Let us then observe the words of n-length in which the forbidden subword belongs to the equivalence class B, i.e. the words in which a good subword is forbidden. If we denote that set with $P_B(m,n)$, it is known [3] that the following theorem 2 is valid :

Theorem 2. $|P_B(m,n)| = \sum_{i=0}^{\lfloor \frac{n}{3} \rfloor}(-1)^i\binom{n-2i}{i}m^{n-3i}$

On the other hand, now we shall prove the next theorem.

Theorem 3.
$$|P_B(m,n)| = \left[\frac{m\alpha^2-1}{m\alpha^2-3}\alpha^n\right]$$
where α is the real root of the equation $x^3 - mx^2 + 1 = 0$ from the interval $[m-1,m]$.

Proof.

We shall consider the alphabet $\mathcal{A} = \{\alpha_1,\dots,\alpha_m\}$ and a forbidden subword baa, where a,b are some fixed elements from the alphabet $\mathcal{A}$. We make the words $\boldsymbol{x_n} \in P_B(m,n)$ from the words $\boldsymbol{x_{n-1}} \in P_B(m,n-1)$ by adding one by one the elements $\alpha_1,\dots,\alpha_m$ in front of them. Let $\boldsymbol{x_{n-1}} \in P_B(m,n-1)$

and $\boldsymbol{x_{n-3}} \in P_B(m, n-3)$. If $\alpha_i \neq b$ then $\alpha_i \boldsymbol{x_{n-1}} \in P_B(m,n)$, but $b\boldsymbol{x_{n-1}} \in P_B(m,n)$ iff $\boldsymbol{x_{n-1}}$ does not begin with aa. Since $aa\boldsymbol{x_{n-3}} \in P_B(m, n-1)$ the recurrence relation $|P_B(m,n)| = m|P_B(m,n-1)| - |P_B(m,n-3)|$ follows, whose characteristic equation is $x^3 - mx^2 + 1 = 0$. This equation has all three real roots and $m-1 < \alpha < m$, $-1 < \beta < 0$, $0 < \gamma < 1$. Since $|\beta| < 1$ and $|\gamma| < 1$ we have $\lim_{n\to\infty} \beta^n = 0$ and $\lim_{n\to\infty} \gamma^n = 0$ so that $|P_B(m,n)| = \left[\frac{m\alpha^2-1}{m\alpha^2-3}\alpha^n\right]$ follows. □

Corollary 2. $\sum_{i=0}^{\lfloor \frac{n}{3} \rfloor} (-1)^i \binom{n-2i}{i} m^{n-3i} = \left[\frac{m\alpha^2-1}{m\alpha^2-3}\alpha^n\right]$.

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{\lfloor \frac{n}{3} \rfloor} (-1)^i \binom{n-2i}{i} m^{n-3i}}{\alpha^n} = \frac{m\alpha^2-1}{m\alpha^2-3}$$

where $\alpha \in [m-1, m]$ is the real root of the equation $x^3 - mx^2 + 1 = 0$ □

Finally, we will determine the cardinal number of the set $P_C(m,n)$ whose elements, the words of length n over the alphabet $\mathcal{A}$, are characterized in such a way as to have the forbidden fixed subword of length three that belongs to the equivalence class C.

Theorem 4. *For $m \geq 3$ valid $|P_C(m,n)| =$*
$(m-1)^n + \sum_{i=1}^{n} \sum_{j=0}^{i-1} \sum_{k=0}^{\lfloor \frac{n-i-j}{2} \rfloor} \binom{i-1}{j} \binom{i-1-j}{k} \binom{n-i-j-k+1}{k+1} (m-1)^{n-i-j} (m-2)^j$

Proof.
$P_C(m,n) = \{\boldsymbol{x_n} | \boldsymbol{x_n} = x_1 \ldots x_n \in \mathcal{A}^n \ \wedge\ \forall s \in N_{n-2}\ (x_s x_{s+1} x_{s+2} \neq aba)\}$. We will partition the set $P_C(m,n)$ into the subsets $P_C^i(m,n)$, which contain words with exactly i a-elements i.e.
$P_C^i(m,n) = \{\boldsymbol{x_n} = x_1 \ldots x_n \in \mathcal{A}^n | \forall s \in N_{n-2}\ x_s x_{s+1} x_{s+2} \neq aba,\ l_a(\boldsymbol{x_n}) = i\}$

Let us calculate $|P_C^i(m,n)|$ for $i \in \{1, \ldots, n\}$. Between these a-elements in the total number of i, in the $i-1$ places we have to write down one of the elements from the set $\{p, q, \lambda\}$, where p is any letter from the set $\mathcal{A} \backslash \{a, b\}$ and q is any word of length two over the alphabet $\mathcal{A} \backslash \{a\}$. The letter λ represents a blank letter, i.e. if there is a letter λ between two a-elements, that means that nothing has been written down.

Let j be the number of occurrences of the letter p, and k the number of occurrences of the letter q. We need to choose j of $i-1$ places for the letter p, and then k of $i-1-j$ places for the letter q. This can be done in

$$\binom{i-1}{j}\binom{i-1-j}{k}$$

different ways.

Now we only have $n-i-j-2k$ letters from the set $A \backslash \{a\}$ which we can put in k places where we have the q-element, as well as in front of and behind the words, and which makes a total of $k+2$ places. This is done by placing $k+1$ partitions between these $n-i-j-2k$ letters. The total number of all the permutations of these $k+1$ partitions and $n-i-j-2k$ letters is

$$\binom{n-i-j-k+1}{k+1}$$

Since the arrangement of $m-2$ letters in j places can be done in $(m-2)^j$ ways, and the arrangement of $m-1$ letters in $n-i-j$ places can be one in $(m-1)^{n-i-j}$ ways, it follows that:
$P_C^i(m,n)| = \sum_{j=0}^{i-1}\sum_{k=0}^{\lfloor\frac{n-i-j}{2}\rfloor}\binom{i-1}{j}\binom{i-1-j}{k}\binom{n-i-j-k+1}{k+1}(m-1)^{n-i-j}(m-2)^j$.

When $i=0$, the total number of words that do not contain the forbidden subword aba equals the number of all the words of n-length of $m-1$ elements, i.e. $(m-1)^n$.

Since $|P_C(m,n)| = (m-1)^n + \sum_{i=1}^{n}|P_C^i(m,n)|$, we have $|P_C(m,n)| = (m-1)^n + \sum_{i=1}^{n}\sum_{j=0}^{i-1}\sum_{k=0}^{\lfloor\frac{n-i-j}{2}\rfloor}\binom{i-1}{j}\binom{i-1-j}{k}\binom{n-i-j-k+1}{k+1}(m-1)^{n-i-j}(m-2)^j$, which we had to prove in the first place. □

Theorem 5.

$$|P_C(m,n)| = \left[\frac{m\alpha^2+m-1}{m\alpha^2-2\alpha+3(m-1)}\alpha^n\right]$$

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

Proof.

We shall consider the alphabet $\mathcal{A}=\{\alpha_1,\alpha_2,\dots,\alpha_m\}$ and the forbidden subword aba, where a and b are some fixed elements from the alphabet $\mathcal{A}$. The words $\boldsymbol{x_n}\in P_C(m,n)$ are made from the words $\boldsymbol{x_{n-1}}\in P_C(m,n-1)$ by adding one by one of the elements $\alpha_1,\alpha_2,\dots,\alpha_m$ in front of them. Let $\boldsymbol{x_{n-1}}\in P_C(m,n-1)$, $\boldsymbol{x_{n-2}}\in P_C(m,n-2)$ and $\boldsymbol{x_{n-3}}\in P_C(m,n-3)$. If $\alpha_i\neq a$ then $\alpha_i\boldsymbol{x_{n-1}}\in P_C(m,n)$ and $a\boldsymbol{x_{n-1}}\in P_C(m,n)$ iff $\boldsymbol{x_{n-1}}$ does not begin with ba i.e. $ab\alpha_i\boldsymbol{x_{n-3}}\in P_C(m,n)$ for all $\alpha_i\neq a$ and $aba\boldsymbol{x_{n-3}}\notin P_C(m,n)$, which means that $ab\boldsymbol{x_{n-2}}\in P_C(m,n)$ iff $\boldsymbol{x_{n-2}}$ begins with the letter $\alpha_i\neq a$. This implies the recurrence relation

$$|P_C(m,n)| = m|P_C(m,n-1)| - |P_C(m,n-2)| + (m-1)|P_C(m,n-3)|$$

whose characteristic equation is $x^3-mx^2+x-(m-1)=0$. This equation has the real root α $(m-1<\alpha<m)$, and two non-real roots β i γ whose modules are less than 1 i.e. $|\beta|=|\gamma|=\sqrt{\frac{m-1}{\alpha}}<1$. Since $\lim_{n\to\infty}\beta^n=0$ and $\lim_{n\to\infty}\gamma^n=0$ it follows that

$$|P_C(m,n)| = \left[\frac{m\alpha^2+m-1}{m\alpha^2-2\alpha+3(m-1)}\alpha^n\right]$$

Corollary 3.

$$2^n+\sum_{i=1}^{n}\sum_{j=0}^{i-1}\sum_{k=0}^{\lfloor\frac{n-i-j}{2}\rfloor}\binom{i-1}{j}\binom{i-1-j}{k}\binom{n-i-j-k+1}{k+1}2^{n-i-j}=\left[\frac{3\alpha^2+2}{3\alpha^2-2\alpha+6}\alpha^n\right] \text{ where}$$

$$\alpha=\sqrt[3]{\frac{\sqrt{633}}{18}+\frac{3}{2}}+\sqrt[3]{\frac{3}{2}-\frac{\sqrt{633}}{18}}+1\approx 2.893289196305. \quad \square$$

Corollary 4.

$$3^n+\sum_{i=1}^{n}\sum_{j=0}^{i-1}\sum_{k=0}^{\lfloor\frac{n-i-j}{2}\rfloor}\binom{i-1}{j}\binom{i-1-j}{k}\binom{n-i-j-k+1}{k+1}3^{n-i-j}\cdot 2^j=\left[\frac{4\alpha^2+3}{4\alpha^2-2\alpha+9}\alpha^n\right]$$

$$\text{where } \alpha=\sqrt[3]{\frac{\sqrt{29}}{2}+\frac{173}{54}}+\sqrt[3]{\frac{173}{54}-\frac{\sqrt{29}}{2}}+\frac{4}{3}\approx 3.939465058587. \quad \square$$

References

[1] Austin Richard, Guy Richard, Binary sequences without isolated ones, The Fibonacci Quarterly, Volume 16, Number 1, 1978, 84–86.

[2] Cvetković, D., The generating function for variations with restrictions and paths of the graph and self complementary graphs, Univ. Beograd, Publ. Elektrotehnički fakultet, serija Mat. Fiz. No320–No328 (1970), 27–34.

[3] Doroslovački, R., The set of all words of length n over any alphabet with a forbidden good subword, Rev. of Res., Fac. of Sci. math. ser.23, 2 (1993), 239-244 Novi Sad.

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

[5] 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).

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

Received by the editors December 10, 1999.