Yugoslav Journal of Operations Research
4 (1994), Number 2, 167–178

MARKOV CHAINS CONNECTED TO THE PROCESS OF GENERATION OF THE WORDS BY A GENERATIVE SYSTEM

Gabriel V. ORMAN

Department of Mathematics, University of Brasov,
50 Iuliu Maniu St. 2200 Brasov, Romania

Abstract: In this paper we considered the process of generation of the words from a stochastic point of view. One of the generative systems constituting the Chomsky hierarchy in the Theory of Formal Languages, namely the so–called phrase–structure grammar, is taken into consideration. The random variable giving the number of derivations by which a word can be generated is introduced and its characteristics are determined. Also, the problem of Markov dependence is discussed and a new procedure for generating the words is proposed.

Keywords: Random variable, Markov chains, alternating generation.

1. INTRODUCTION

In the process of transmission of information a very important aspect is that of generation of the words by a generative system. In our tentative for finding new possibilities to characterize the process of generation of the words by sequences of intermediate words we have adopted a stochastic point of view involving Markov chains. Because such sequences of intermediate words (called "derivations") by which the words are generated are finite, it results that finite Markov chains will be connected to the process. In order that our discussion is as general as possible, the derivations are considered according to the most general class of formal grammars from the so–called "Chomsky hierarchy", namely those that are free of any restrictions and are called "phrase–structure grammars".

The novelty that we have proposed consists in the organization of the process of generation of the words by considering the set of all the derivations according to such a grammar split into equivalence classes, each of them containing derivations of the same length (here we are not interested in the internal structure of the intermediate

---

Paper presented at the 2nd Balkan Conference on OR, Thessaloniki, Greece, 18–21 October, 1993

words of a derivation but only in its length). Because the facts to be considered are based specially on our results obtained in some previous papers, there have been some situations when one said that they are rather difficult to be well understood. For this reason we shall pass over some details that are collateral to our main topic.

Thus, a finite nonempty set is called an *alphabet* and will be denoted by Σ. A *word* over Σ is a finite sequence $u = u_1 \dots u_k$ of elements in Σ. The integer $k \geq 0$ is the *length* of the word and is denoted by $| u |$. The word of length zero is called the *empty word* and is denoted by ε. If Σ is an alphabet let us denote by Σ^* the free semigroup, with identity, generated by Σ (Σ^* is considered in relation to the usual operation of concatenation).

A *phrase–structure grammar* (briefly psg) is a system $G = (V, \Sigma, P, \sigma)$ where:

(i) V is an alphabet called the *total* alphabet;

(ii) $\Sigma \subseteq V$ is an alphabet the elements of which are called *terminal* symbols (or *letters*);

(iii) P is a finite subset of the Cartesian product $[(V \setminus \Sigma)^* \setminus \{\varepsilon\}] \times V^*$. Its elements are called *productions* (they are the rules of the grammar);

(iv) $\sigma \in (V \setminus \Sigma)$ is said to be the initial symbol. The elements of $V \setminus \Sigma$ are called *variables* (or *nonterminals*).

Note. A production is often written as so: $\sigma \to a$, $\xi \to a\ \xi\ b$, etc.

For y and z in V^*, it is said that y *directly generates* z, and one writes $y \Rightarrow z$, if there exist the words t_1, t_2, u and v such that $y = t_1 u t_2$, $z = t_1 v t_2$ and $(u, v) \in P$ (alternatively written as $u \to v$. Then, y is said *to generate* z, and one writes $y \Rightarrow^* z$, if either $y = z$ or there exists a sequence $(w_0, w_1, \dots, w_j)$ of words in V^* such that $y = w_0$, $z = w_j$ and $w_i \Rightarrow w_i + 1$ for each i (we write $\Rightarrow^*$ for the reflexive–transitive closure of $\Rightarrow$). The sequence $(w_0, w_1, \dots, w_j)$ is called a *derivation* of length j and will be denoted by $D(j)$. Evidently, many derivations of the same length according to G may exist.

The subset of Σ^*, written $L(G)$, such that

$$L(G) = \{ w \in \Sigma^* \mid \sigma \Rightarrow^* w \}$$

is called the *phrase–structure language* (briefly psl) generated by G. It is known that the family of psl's coincides with the family of recursively enumerable sets studied in mathematical logic.

Because a derivation of length equal to 1 is just a production we shall suppose, from now on, that the length of a derivation is $j \geq 2$.

Let us denote by $\wp$ the family of all the derivations according to G and let D_j be the class of all the derivations of length j. Evidently, $\wp$ splits into equivalence classes, each of them being represented by one of its elements arbitrarily chosen. Thus, $D(j)$ is the representative of the class D_j.

2. THE PROBABILITY OF GENERATION OF A WORD

Now let ν_x be the number of derivations into equivalence class in $\wp$ of $D(x)$, $x \geq 2$, by which the word w is generated. Obviously, w can or cannot be generated into equivalence class of $D(x)$. Thus, if w is generated into D_x, then the probability that it will be generated again into D_{x+1} is denoted by γ; but given that w is not generated into D_x, the probability that w should be generated into D_{x+1} is denoted by β. Hence we are in the case when the equivalence classes of the derivations are connected into a simple Markov chain. We assume that both γ and β are different from 0 and 1 (these cases are of no particular interest). But each term of a derivation is the result of application of an *only* production on the precedent term. Furthermore, w can or cannot be generated by a production (that is P can or cannot contains an element (σ, w), such that the probability that w should or should not be generated by a production are unknown to us). In other words if we agree to denote by D_1 the class of productions, it means that there is not an equivalence class preceding the class D_1. Thus, we denote by p_1 the probability that w should be generated into D_1 and by $q_1 = 1 - p_1$ the probability that w should not be generated into D_1.

We now propose to determine the probability that w should be generated into the class D_x. Let p_x be the probability that w should be generated into the class D_x and then we have $q_x = 1 - p_x$. Clearly w can be generated into D_x in two mutually exclusive ways: (1) w will be generated into D_{x-1} and will be generated again into D_x; (2) w was not generated into D_{x-1} but it will be generated into D_x. Hence we have

$$p_x = p_{x-1}\gamma + q_{x-1}\beta$$

or

$$p_x = p_{x-1}(\gamma - \beta) + \beta$$

Denoting $\gamma - \beta = \delta$ we obtain

$$p_x = p_{x-1}\delta + \beta$$

Generally it results (after some transformations) is

$$p_x = (p_1 - \frac{\beta}{1-\delta})\,\delta^{x-1} + \frac{\beta}{1-\delta} \qquad (1)$$

But from the previous conditions imposed on γ and β it results $|\delta| < 1$ so that it follows that as $x \to \infty$, $p_x \to \beta/(1-\delta)$. On the other hand x being the length of a derivation it must be finite such that this situation corresponds to the case when the word cannot be generated into D_x. Thus, we have:

THEOREM 1. The word $w \in L(G)$ is generated into the class D_x if

$$p_x \neq \frac{\beta}{1-\delta}$$

We observe that the constant to which p_x tends does not depend on the probability p_1. Since it plays the role of a limiting probability it is natural to introduce the notation

$$p_x = \frac{\beta}{1-\delta} \quad \text{and} \quad q = 1 - p = \frac{1-\gamma}{1-\delta}$$

Now, the relation (1) becomes

$$p_x = (p_1 - p)\delta^{x-1} + p \tag{2}$$

3. THE CONDITIONAL PROBABILITY OF GENERATION OF A WORD

Let us denote by $p^{(i)}_j$ the probability that w should be generated into the class D_j if it was generated into the class D_i, $i < j$. Following as in the precedent Section 2, we get

$$p^{(i)}_j = p^{(i)}_{j-1}\gamma + q^{(i)}_{j-1}\beta = p^{(i)}_{j-1}\gamma + (1 - p^{(i)}_{j-1})\beta$$

or

$$p^{(i)}_j = p^{(i)}_{j-1}\delta + \beta, \text{ for all } j > i + 1 \tag{3}$$

But the probability that w should be generated into D_{i+1} if it was already generated into D_i is γ such that we have $p^{(i)}_{i+1} = \gamma$.

Then, as in the first case we have successively

$$p^{(i)}_{i+2} = p^{(i)}_{i+1}\delta + \beta = \gamma\delta + \beta,$$

$$p^{(i)}_{i+3} = p^{(i)}_{i+2}\delta + \beta = \gamma\delta^2 + \beta(1 + \delta),$$

....

$$p^{(i)}_j = p^{(i)}_{i+(j-i-1)}\delta + \beta = \gamma\delta^{j-i-1} + \beta(1+\delta+\delta^2+...+\delta^{j-i-2}),$$

or

$$p^{(i)}_j = (\gamma - \frac{\beta}{1-\delta})\delta^{j-i-1} + \frac{\beta}{1-\delta}.$$

Substituting now $\gamma - \beta$ by δ it results

$$p^{(i)}_j = p + q\delta^{j-i} \tag{4}$$

Therefore we have obtained the following result:

THEOREM 2. The probability that the word $w \in L(G)$ should be generated into the class D_j if it was generated into the class D_i, $i < j$, $i \geq 2$, $j > 2$, is given by the formula (4).

4. A MARKOV DEPENDENCE CHARACTERIZATION OF THE AMBIGUOUS PS LANGUAGES

As it was easily seen the words able to be generated by more derivations have been considered. This is a specific property of the ambiguous languages that we shall characterize below in terms of ν_x defined in Section 2. Obviously it is a random variable that takes the values 1 and 0 with the probabilities p_x and $q_x = 1 - p_x$ respectively. Then, the number of derivations in $n-1$ equivalence classes by which the word w is generated is

$$\nu = \sum_{x=2}^{n} \nu_x$$

Now, the expectation of ν is

$$\mathrm{E}\,\nu = \sum_{x=2}^{n} \mathrm{E}\,\nu_x \tag{5}$$

where $\mathrm{E}\,\nu_x = 1 \cdot p_x + 0 \cdot q_x = p_x$, so that we have

$$\mathrm{E}\,\nu = \sum_{x=2}^{n} p_x$$

However the equivalence classes are connected into a homogeneous Markov chain, so that p_x is given by the relation (2) and we get

$$\mathrm{E}\,\nu = (n-1)p + \sum_{x=2}^{n} (p_1 - p)\delta^{x-1} = (n-1)p + (p_1 - p)\frac{\delta - \delta^n}{1-\delta} \tag{6}$$

Further on the variance of ν is

$$\mathrm{D}\,\nu = \mathrm{E}[\sum_{x=2}^{n} (\nu_x - p_x)]^2$$

$$= \sum_{x=2}^{n} \mathrm{E}(\nu_x - p_x)^2 + 2 \sum_{i<j,\, i\geq 2} \mathrm{E}(\nu_i - p_i)(\nu_j - p_j) \tag{7}$$

But

$$\mathrm{E}(\nu_x - p_x)^2 = \mathrm{D}\,\nu_x = p_x q_x = [(p_1 - p)\delta^{x-1} + p][1 - p_x]$$

$$= pq + (p_1 - p)(q - p)\delta^{x-1} - (p_1 - p)^2 \delta^{2x-2}$$

and the second term in (7) becomes

$$\mathrm{E}(\nu_i - p_i)(\nu_j - p_j) = \mathrm{E}\,\nu_i \nu_j - p_i p_j$$

where $\nu_i \nu_j$ is a random variable taking the values 1 and 0. The value 1 is taken with the probability $p_i p^{(i)}{}_j$ so that we have

$$\mathrm{E}(\nu_i - p_i)(\nu_j - p_j) = p_i p^{(i)}{}_j - p_i p_j = p_i (p^{(i)}{}_j - p_j)$$

By means of the relations (2) and (4) it results

$$\mathrm{E}(\nu_i - p_i)(\nu_j - p_j) = pq\delta^{j-i} - (p_1 - p)^2 \delta^{i+j-2} + (p_1 - p)(q - p)\delta^{j-1}$$

Thus, the first term in (7) is obtained from the equality

$$\sum_{x=2}^{n} \mathrm{E}(\nu_x - p_x)^2 = \sum_{x=2}^{n} \mathrm{D}\,\nu_x =$$

$$(n-1)pq + (p_1 - p)(q - p)\frac{\delta(1 - \delta^{n-1})}{1-\delta} - (p_1 - p)^2 \frac{\delta^2(1 - \delta^{2n-2})}{1 - \delta^2} \tag{8}$$

and the second from the equality

$$\sum_{i<j,\, i\geq 2} \mathrm{E}(\nu_i - p_i)(\nu_j - p_j) =$$

$$\sum_{i<j,\, i\geq 2} [pq\delta^{j-i} - (p_1 - p)^2 \delta^{i+j-2} + (p_1 - p)(q - p)\delta^{j-1}] \tag{9}$$

Now for each term of (9) we obtain as follows

$$\sum_{i<j,i\geq 2} pq\delta^{j-i} = npq\frac{\delta}{1-\delta} - pq\frac{\delta}{1-\delta}(2+\frac{\delta-\delta^{n-1}}{1-\delta}) \tag{10}$$

$$-\sum_{i<j,i\geq 2}(p_1-p)^2\delta^{i+j-2} = -\frac{(p_1-p)^2}{1-\delta}\cdot\frac{\delta^3(1-\delta^{n-2})(1-\delta^{n-1})}{1-\delta^2} \tag{11}$$

$$\sum_{i<j,i\geq 2}(p_1-p)(q-p)\delta^{j-1} = (p_1-p)(q-p)\frac{\delta^2-\delta^n}{1-\delta} \tag{12}$$

Let us return to (6) and we observe that excepting the term $(n-1)p$ the second is bounded as n increases (we say that $L(G)$ tends to become an ambiguous language of order n).

Thus, the expectation of ν is

$$\mathrm{E}\,\nu = (n-1)p + u_n$$

where u_n is a certain quantity that remains bounded as n increases.

As regards the variance of ν, excepting the terms $(n-1)pq$ in (8) and $npq\,\delta/(1-\delta)$ in (1), the other all are also bounded as n increases, so that we have

$$\mathrm{D}\,\nu = (n-1)pq + 2npq\frac{\delta}{1-\delta} + \nu_n$$

where ν_n is a certain quantity that remains bounded as n increases. In this way we have obtained the following result:

THEOREM 3. If among the equivalence classes of the derivations according to a psg, G, a Markov dependence exists, then $L(G)$ tends to become an ambiguous language of order n if there exists a word $w \in L(G)$ such that the expectation and the variance of the random variable giving the number of derivations by which w is generated verify the following relations

$$\mathrm{E}\,\nu = (n-1)p + u_n$$

$$\mathrm{D}\,\nu = pq[\,n\frac{1+\delta}{1-\delta} - 1\,] + \nu_n$$

5. THE PROBABILITY OF ALTERNATING GENERATION OF THE WORDS

Now we propose to examine the process of generation of the words in the special case when a word can be generated into the equivalence class of a derivation on the following conditions:

1. it can be generated into the class D_x, $x \geq 2$, by more elements of it (as it was the case until now);
2. if it is generated into the class D_x then, it is not generated into the preceding and the next class.

We refer to such a procedure for generating the words as being an *alternating generation*. From now on we shall use the notation w for the case when this word is generated into an equivalence class and the notation $\overline{w}$ otherwise.

We now propose to determine the probability $P_n(m)$ that the word w should be generated by m $(m < n)$ derivations in the following ways:

(i) it will be generated by the derivations from the first class and the last and there is a production $(\sigma, w) \in P$ (for shortening we shall express simple: "w will be generated by the first class and the last");

(ii) it will be generated by the first class and will not be generated by the last and there is a production $(\sigma, w) \in P$;

(iii) it will not be generated by the first class but will be generated by the last and there is not a production $(\sigma, w) \in P$;

(iv) it will not be generated both by the first class and the last and there is not a production $(\sigma, w) \in P$.

Then the probability $P_n(m)$ is computed by the relation

$$P_n(m) = P_n(m, ww) + P_n(m, w\overline{w}) + P_n(m, \overline{w}w) + P_n(m, \overline{w}\overline{w}) \qquad (13)$$

where $Pn(m, ww)$ is the notation for (i), then $P_n(m, w\overline{w})$ is the notation for (ii) a.s.o.

To compute these probabilities we denote by r_i the cardinal of the class D_i so that the expression "w is generated by the class D_i" means that it is generated by the r_i derivations. Consequently when we refer to a word as being generated by m derivations, this is nothing but than m can be equal to r_i or that it can be equal to the number of the derivations from more equivalence classes.

A. Now to compute $P_n(m, ww)$ we consider that the word w is generated by the first class and the last. In this case the maximum length of a derivation must be an even integer, say $2k$. Thus the word is generated in the following manner: it is generated by the class D_2, then it is not generated by the class D_3, then it is generated by the class D_4, ... then it is not generated by the class D_{2k-1} but it is generated by the class D_{2k}, such that

$$\sum_{i=1}^{k} r_{2i} = m \quad \text{and} \quad \sum_{i=1}^{k-1} r_{2i+1} = n - m$$

Therefore, the probability of generation of the word w in such a way is

$$p_1 \gamma^{r_2}(1-\gamma)(1-\beta)^{r_3-1}\beta\gamma^{r_4-1}(1-\gamma)(1-\beta)^{r_5-1}\beta \ldots$$
$$\ldots \cdot(1-\beta)^{r_{2k-1}-1}\beta\gamma^{r_{2k}-1}$$

or

$$p_1 \gamma^{m-k+1}(1-\gamma)^{k-1}(1-\beta)^{n-m-k+1}\beta^{k-1} \qquad (14)$$

But this probability is dependent only on n, m and k and it is independent on the values of r_i. Also the number m can be decomposed into k positive summands in

$\binom{m-1}{k-1}$ ways and the number $n-m$ can be decomposed into $k-1$ groups in $\binom{n-m-1}{k-2}$ ways such that from (14) we obtain

$$\binom{m-1}{k-1}\binom{n-m-1}{k-2}p_1\gamma^{m-k+1}(1-\gamma)^{k-1}(1-\beta)^{n-m-k+1}\beta^{k-1}$$

Since in our hypothesis k can take on any value from 1 to m, we obtain that

$$P_n(m, w\omega) = p_1\sum_{k=2}^{m}\binom{m-1}{k-1}\gamma^{m-k+1}(1-\gamma)^{k-1}\binom{n-m-1}{k-2}(1-\beta)^{n-m-k+1}\beta^{k-1} \quad (15)$$

B. Now the word is generated by the first class but not by the last. The maximum length of a derivation must be an odd integer, say $2_k + 1$. The word is generated by the class D_2, then it is not generated by the class D_3, ... then it is generated by the class D_{2k} and is not generated by the class D_{2k+1} such that

$$\sum_{i=1}^{k} r_{2i} = m \quad \text{and} \quad \sum_{i=1}^{k} \mathrm{r}_{2\mathrm{i}+1} = n-m$$

The probability of generation of the word in this case is

$$p_1\gamma^{r_2}(1-\gamma)(1-\beta)^{r_3-1}\beta\gamma^{r_4-1}(1-\gamma)(1-\beta)^{r_5-1}\beta \ldots$$
$$\ldots \cdot\gamma^{r_{2k}-1}(1-\gamma)(1-\beta)^{r_{2k+1}-1}$$

or

$$p_1\gamma^{m-k+1}(1-\gamma)^k(1-\beta)^{n-m-k}\beta^{k-1} \quad (16)$$

In this case m and $n-m$ can be decomposed into k groups in $\binom{m-1}{k-1}$ and respective $\binom{n-m-1}{k-1}$ ways such that we have

$$P_n(m, w\overline{w}) = p_1\sum_{k=1}^{m}\binom{m-1}{k-1}\gamma^{m-k+1}(1-\gamma)^k\binom{n-m-1}{k-1}(1-\beta)^{n-m-k}\beta^{k-1} \quad (17)$$

C. In the case of $P_n(m, \overline{w}w)$ the word is not generated by the first class but it is generated by the last. The maximum length of a derivation is again and odd integer $2k + 1$ as in the previous case. Now the word is not generated by the class D_2, then it is generated by the class D_3, ... then it is not generated by the class D_{2k} and is generated by the class D_{2k+1}, such that

$$\sum_{i=1}^{k} r_{2i+1} = m \quad \text{and} \quad \sum_{i=1}^{k} \mathrm{r}_{2\mathrm{i}} = n-m$$

The probability of generation of the word is

$$q_1(1-\beta)^{r_2}\beta\gamma^{r_3-1}(1-\gamma)(1-\beta)^{r_4-1}\beta\gamma^{r_5-1}(1-\gamma) \ldots$$
$$\ldots \cdot(1-\beta)^{r_{2k}-1}\beta\gamma^{r_{2k+1}-1}$$

that is to say

$$q_1(1-\beta)^{n-m-k+1}\beta^k\gamma^{m-k}(1-\gamma)^{k-1} \tag{18}$$

Now we get the same number of decompositions for m and $n-m$ as in the precedent, hence

$$P_n(m,\overline{w}w) = q_1\sum_{k=1}^{m}\binom{m-1}{k-1}\gamma^{m-k}(1-\gamma)^{k-1}\binom{n-m-1}{k-1}(1-\beta)^{n-m-k+1}\beta^k \tag{19}$$

D. As regards $P_n(m,\overline{ww})$ we have the following arrangement of the equivalence classes (in this case the maximum length of a derivation must be an even integer $2k$ as in the first case): the word is not generated by the class D_2, then it is generated by the class D_3, ... then it is generated by the class D_{2k-1} and is not generated by the class D_{2k}, so that

$$\sum_{i=1}^{k-1} r_{2i+1} = m \quad \text{and} \quad \sum_{i=1}^{k} r_{2i} = n-m$$

We find that the probability of generation of the word is

$$q_1(1-\beta)^{r_2}\beta\gamma^{r_3-1}(1-\gamma)(1-\beta)^{r_4-1}\beta\gamma^{r_5-1}(1-\gamma)\ \ldots$$
$$\ldots\cdot\gamma^{r_{2k-1}-1}(1-\gamma)(1-\beta)^{r_{2k}-1}$$

that means

$$q_1(1-\beta)^{n-m-k+1}\beta^{k-1}\gamma^{m-k+1}(1-\gamma)^{k-1} \tag{20}$$

In this case m and $n-m$ can be decomposed as in the first case but in a reverse order. Thus is results

$$P_n(m,\overline{ww}) = q_1\sum_{k=2}^{m}\binom{m-1}{k-2}\gamma^{m-k+1}(1-\gamma)^{k-1}\binom{n-m-1}{k-1}(1-\beta)^{n-m-k+1}\beta^{k-1} \tag{21}$$

For computing these probabilities we consider the sum

$$F_{mn} = \sum_{k=1}^{m}\binom{m}{k}\gamma^{m-k}(1-\gamma)^{k}\binom{n-m}{k}(1-\beta)^{n-m-k}\beta^k$$

and we put

$$x = \frac{k-m(1-\gamma)}{\sqrt{m\gamma(1-\gamma)}} \qquad \text{and}$$

$$\mathrm{m} = \mathrm{np} + \mathrm{z}\sqrt{\frac{\mathrm{m}\gamma(1-\gamma)+(\mathrm{n}-\mathrm{m})\beta(1-\beta)}{(1-\gamma+\beta)^2}} \tag{22}$$

Then it results

$$k = m(1-\gamma)+x\sqrt{m\gamma(1-\gamma)} \qquad \text{and}$$

$$\mathrm{m}-\mathrm{k} = \mathrm{m}\gamma - x\sqrt{m\gamma(1-\gamma)}$$

Now, after some transformation, the first part in F_{mn} becomes

$$\binom{m}{k}\gamma^{m-k}(1-\gamma)^k = \frac{m!}{k!(m-k)!}\gamma^{m-k}(1-\gamma)^k =$$

$$\frac{1}{\sqrt{2\pi m\gamma(1-\gamma)}}e^{-\frac{x^2}{2}}(1+a_n) \tag{23}$$

where a_n vanishes uniformly when x lies in some finite interval.

Analogously, the second becomes

$$\binom{n-m}{k}(1-\beta)^{n-m-k}\beta^k = \frac{(n-m)!}{k!(n-m-k)!}(1-\beta)^{n-m-k}\beta^k =$$

$$\frac{1}{\sqrt{2\pi(n-m)\beta(1-\beta)}}e^{-\frac{y^2}{2}}(1+b_n) \tag{24}$$

where

$$y = \frac{k-(n-m)\beta}{\sqrt{(n-m)\beta(1-\beta)}}$$

and b_n vanishes uniformly when x lies in some finite interval.

Therefore we have now

$$k = (n-m)\beta + y\sqrt{(n-m)\beta(1-\beta)} \qquad \text{and}$$

$$n-m-k = (n-m)(1-b) - y\sqrt{(n-m)\beta(1-\beta)}$$

assuming that $x = O(m^{1/6})$ and $y = O(n^{\delta})$, where δ is a number satisfying the inequalities $0 < \delta < 1/6$. Now we decompose F_{mn} so

$$F_{mn} = \sum_{k=1}^{m(1-\gamma)-t\sqrt{m\gamma(1-\gamma)}} + \sum_{m(1-\gamma)-t\sqrt{m\gamma(1-\gamma)}}^{m(1-\gamma)+t\sqrt{m\gamma(1-\gamma)}} + \sum_{m(1-\gamma)+t\sqrt{m\gamma(1-\gamma)}}^{m}$$

First, we consider the middle sum (we denote it F_2) and by virtue of the relations (23) and (24) we get

$$F_2 = \frac{1}{2\pi\sqrt{m\gamma(1-\gamma)(n-m)\beta(1-\beta)}}\sum_{x=-t}^{t} e^{-\frac{x^2+y^2}{2}}(1+a_n)(1+b_n)$$

Taking into consideration that x and y are connected by the relation

$$m(1-\gamma) + x\sqrt{m\gamma(1-\gamma)} = (n-m)\beta + y\sqrt{(n-m)\beta(1-\beta)}$$

and substituting the value of m of (22) it follows

$$y\sqrt{(n-m)\beta(1-\beta)} = z\sqrt{m\gamma(1-\gamma)+(n-m)\beta(1-\beta)} + x\sqrt{m\gamma(1-\gamma)}$$

from which we have

$$y = \frac{1}{\sqrt{(n-m)\beta(1-\beta)}}\left[\, z\sqrt{m\gamma(1-\gamma)+(n-m)\beta(1-\beta)} + x\sqrt{m\gamma(1-\gamma)}\,\right]$$

The sum $x^2 + y^2$ can be obtained now

$$x^2 + y^2 = z^2 + \frac{m\gamma(1-\gamma)+(n-m)\beta(1-\beta)}{(n-m)\beta(1-\beta)}\left[x + z\sqrt{\frac{m\gamma(1-\gamma)}{m\gamma(1-\gamma)+(n-m)\beta(1-\beta)}}\right]^2$$

After some transformation F_2 becomes

$$F_2 = \frac{1}{\sqrt{2\pi[\, m\gamma(1-\gamma)+(n-m)\beta(1-\beta)\,]}} e^{-\frac{z^2}{2}}(1+c_n)$$

But according to (22) we have

$$m\gamma(1-\gamma)+(n-m)\beta(1-\beta) = np\gamma(1-\gamma)+nq\beta(1-\beta)+O(z\sqrt{n}) =$$

$$npq(1+\gamma-\beta)(1-\gamma+\beta)+O(z\sqrt{n})$$

so that asymptotically,

$$m\gamma(1-\gamma)+(n-m)\beta(1-\beta) = npq(1+\gamma-\beta)(1-\gamma+\beta) \qquad (25)$$

Therefore

$$F_2 = \frac{1}{\sqrt{2\pi npq(1+\gamma-\beta)(1-\gamma+\beta)}} e^{-\frac{z^2}{2}}(1+d_n)$$

As regards the first and the third terms of F_{mn} it is not difficult to observe that for sufficiently large n they tend to zero: $F_1 = O(1)$ and $F_3 = O(1)$. Therefore F_2 is the main term of F_{mn} such that we have

$$P_n(m, ww) = \frac{p_1\beta}{\sqrt{2\pi[\, m\gamma(1-\gamma)+(n-m)\beta(1-\beta)\,]}} e^{-\frac{z^2}{2}}(1+c_n)$$

Analogously we come to the conclusion that

$$P_n(m, w\overline{w}) = \frac{p_1(1-\gamma)}{\sqrt{2\pi[\, m\gamma(1-\gamma)+(n-m)\beta(1-\beta)\,]}} e^{-\frac{z^2}{2}}(1+c_n),$$

$$P_n(m, \overline{w}w) = \frac{q_1\beta}{\sqrt{2\pi[\, m\gamma(1-\gamma)+(n-m)\beta(1-\beta)\,]}} e^{-\frac{z^2}{2}}(1+c_n),$$

$$P_n(m, \overline{w}\overline{w}) = \frac{q_1(1-\gamma)}{\sqrt{2\pi[\, m\gamma(1-\gamma)+(n-m)\beta(1-\beta)\,]}} e^{-\frac{z^2}{2}}(1+c_n).$$

Now $P_n(m)$ is obtained by adding these probabilities. We get

$$P_n(m) = \frac{p_1\beta + p_1(1-\gamma)+q_1\beta+q_1(1-\gamma)}{\sqrt{2\pi[\, m\gamma(1-\gamma)+(n-m)\beta(1-\beta)\,]}} e^{-\frac{z^2}{2}}(1+d_n)$$

or, by taking into consideration the equality (25), it results

$$P_n(m) = \frac{1-\gamma+\beta}{\sqrt{2\pi npq(1+\gamma-\beta)(1-\gamma+\beta)}} e^{-\frac{z^2}{2}}(1+d_n)$$

Thus, we have the following main result:

THEOREM 4. If the word w is generated by an alternating generation procedure according to a psg, the derivations of which belonging to n equivalence classes then, the probability that it should be generated by m classes out of n is given by the relation

$$P_n(m) = \frac{1}{\sqrt{2\pi npq}} \sqrt{\frac{1-\gamma+\beta}{1+\gamma-\beta}} e^{-\frac{z^2}{2}} (1+d_n)$$

REFERENCES

[1] Ginsburg,S., *Algebraic and Automata–Theoretic Properties of Formal Languages*, North–Holland Publ. Co., Amsterdam, 1975.

[2] Gnedenko,V.B., *The Theory of Probability*, Mir Publisher, Moscow, 1976.

[3] Orman,V.G., *Limbaje formale*, ed. Tehnica, Bucuresti (in Romanian), 1982.

[4] Orman,V.G., "A random variable connected with the process of generation of the words", Bul. Univ. Brasov Issue C Vol.XXX, 1988.

[5] Orman,V.G., "Alternating generation of the words by a phrase–structure grammar; conceptual aspects", Bul. Univ. Brasov Issue C Vol.XXXI, 1989.

[6] Salomaa,A., and Soittola,M., *Automata–Theoretic Aspects of Formal Power Series*, Springer–Verlag New York Inc., 1978.