(1970b) Über eine exakte Wissenschaft von der Literatur. In: *Aspekte* 3, 34–37.

(1970c) Über formale Struktureigenschaften musikalischer Partituren. In: *Experimentelle Musik.* (Hrsg. F. Winckel). Berlin: Mann, 33–58.

(1971a) Possibilities of exact style analysis. In: *Patterns of literary style.* (Hrsg. J. Strelka). Pennsylvania State Univ. Press, 51–76.

(1971b) Über den Gesetzesbegriff einer exakten Literaturwissenschaft, erläutert an Sätzen und Satzfolgen. In: *Zeitschrift für Literaturwissenschaft und Linguistik* 1 (1/2), 113–137.

(1972) Maß und Zahl bei Dürer, Maß und Zahl heute. In: *Am Beispiel Dürers.* (Hrsg. H. Glaser). München: Bruckmann, 26–68.

(1975) Gesetze der Dichtung. In: *Bild der Wissenschaften,* 78–84.

(1978) *Mächte von morgen.* Stuttgart: Dt. Verl.-Anst.

7. Literatur (in Auswahl)

Bauer, Hans-Joachim (1976), Statistik, eine objektive Methode zur Analyse von Kunst? In: *International review of the aesthetics and sociology of music* 7, 249–263.

Best, Karl-Heinz/Altmann, Gabriel (1996), Zur Länge der Wörter in deutschen Texten. In: *Glottometrika 15.* (Hrsg. P. Schmidt). Trier: Wissenschaftlicher Verlag, 166–180.

Grotjahn, Rüdiger (1982), Ein statistisches Modell für die Verteilung der Wortlänge. In: *Zeitschrift für Sprachwissenschaft* 1, 44–75.

Piotrowski, Rajmund G./Bektaev, K. B./Piotrowskaja, A. A. (1985), *Mathematische Linguistik.* Bochum: Brockmeyer.

Wagner, Günther (1976), Exaktwissenschaftliche Musikanalyse und Informationsästhetik. In: *International review of the aesthetics and sociology of music* 7, 63–76.

Dieter Aichele, Worms (Deutschland)

11a. Mathematical aspects and modifications of Fucks' Generalized Poisson Distribution (GPD)

1. Historical context

German physicist Wilhelm Fucks (1902–1990) is well-known for his influential studies in the field of quantitative linguistics and stylistics in the 1950–70s (cf. chapter 11.). His inspiring works as to a scientific approach to text and language – motivated by the desire to find laws in the strict meaning of this word, not only in the realm of nature, but in the social and cultural spheres, as well – remain worth while being analyzed, still today. Actually, however, due to the alleged clash between natural and human sciences, only part of his suggestions and ideas have met sufficient attention in the scientific community. And even in those fields, where reference is made to his work, this is done only with regard to particular aspects of his general approach.

As to the study of language(s), for example, Fucks' studies on word length, or sentence length, have been extremely influential. Fucks (1955a; 1955b; 1956b) assumed that the 1-displaced Poisson distribution might serve as a general standard model for theoretically describing word length frequencies in syllable-based languages. Thus, Fucks considered the 1-displaced Poisson distribution to be the "mathematical law of the process of word-formation from syllables for all those languages, which form their words from syllables" (Fucks 1955b, 209).

Fucks' assumption was generally accepted in the 1960s and 70s; it inspired many follow-up studies all over the world, and the model suggested by him led to the fact that the 1-displaced Poisson distribution began to be termed "Fucks distribution" by the linguist community.

The general celebration of this model as the Fucks distribution remained unchanged, when scholars began to realize that a Russian military doctor from Sankt Petersburg, Sergej Grigor'evič Čebanov (1897–1966), had tried to find a general model for the distribution of words according to the number of syllables, as early as in the 1940s. In fact, Čebanov was one of the first to propose a theoretical model of word length frequencies, which he considered to be valid for various languages (cf. Grzybek 2005). Having observed a specific relation between the mean word length λ of a text and the relative frequencies p_i of the individual word length classes, Čebanov was the first to suggest the Poisson distribution as a general model for various languages. Since the texts studied by Čebanov contained no zero-syllable words, he suggested the 1-displaced Poisson distribution – see below, (22) – as an appropriate model for his data from various languages. Therefore, the 1-displaced Poisson distribution became to be referred to as the "Čebanov-Fucks distribution" by many linguists, thus adequately honoring the pioneering work of Čebanov. Despite this historical correction, it was rather Fucks than Čebanov, who would be credited for having established the 1-displaced Poisson distribution as a standard model for word length frequency distributions. In a way, this high estimation of Fucks' is justified, since Fucks based his ruminations on elaborate mathematical ideas and embedded them in a broader context. Still, the appreciation covers only part of Fucks's merits, since, in his concept, the 1-displaced Poisson distribution turns out to be one special case of a much more general model. This generalized model, which shall be termed Fucks' Generalized Poisson Distribution (GPD) throughout this chapter, is a specific generalization of the Poisson distribution; in fact, we are concerned with a sum of weighted Poisson probabilities from which, under particular conditions, various special cases may be derived. However, this generalization takes no prominent place in Fucks' linguistic analyses; rather, he mentioned it in some of his publications (e.g., Fucks 1956a; 1956d). Therefore, it is not surprising that Fucks' GPD has hardly ever been discussed in detail. Curiously enough, however, if at all, the Fucks GPD has been discussed more intensively by a Russian-reading audience, due to the Russian translation of one of Fucks' articles in 1957 (see below, Sec. 3). In fact, in that context, Fucks' theoretical assumptions were not only generally accepted, but also served as a starting point for new developments as to alternative ways of parameter estimation and even further generalizations.

The purpose of this chapter is to consider the concept of Fucks' GPD more precisely, as well as the above-mentioned modifications and generalizations. Since, in most cases, the relevant works are not systematic in their approaches, the corresponding derivations shall be calculated and presented, in detail.

2. Fucks' Generalized Poisson Distribution (GPD)

Assuming any text generation to be a stochastic process, Fucks arrived at a model, which later became to be known as the Fucks binomial distribution (cf. Fucks 1956a, 12). The derivation of the model need not be explained here, in detail. Ultimately, we are concerned with a generalization of the well-known binomial distribution as a sum of weighted binomial probabilities. Precisely, the Fucks binomial distribution is given as follows:

$$p_i = P(X = i) = \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) \cdot \binom{n-k}{i-k} q^{i-k}(1-q)^{n-i}. \qquad (1)$$

Here, the random variable X denotes the number of syllables per word, i.e. $X = i$, $i = 0, 1, 2, 3, \ldots, n$; $p_i = P(X = i)$ is the probability that a given word has i syllables, with $\sum_{i=0}^{n} p_i = 1$; $0 < p < 1$; $q = 1 - p$; the specific weights are denoted by ε_k, k indicating the number of components to be analyzed. The expected value or mean value is found to be $\mu = (n - \varepsilon')\ q + \varepsilon'$ with $\varepsilon' = \sum_{k=1}^{\infty} \varepsilon_k$.

Furthermore, for $n \to \infty$ and $q \to 0$, with the condition $\mu - \varepsilon' = (n - \varepsilon')\ q = constant$, the Fucks binomial distribution (1) converges to the Fucks GPD, which shall be focused upon in this section.

Generally speaking, Fucks' GPD can be understood to be a sum of weighted Poisson probabilities. The corresponding weights are denoted by $(\varepsilon_k - \varepsilon_{k+1})$, k indicating the number of components to be analyzed. The Fucks GPD distribution is given by

$$p_i = P(X = i) = e^{-\lambda} \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) \cdot \frac{\lambda^{i-k}}{(i-k)!}. \qquad (2)$$

Here, the random variable X denotes the number of syllables per word, i. e. $X = i, i = 0, 1, 2, 3, ..., I$; $p_i = P(X = i)$ is the probability that a given word has i syllables, with $\sum_{i=1}^{I} p_i = 1$; $\lambda = \mu - \varepsilon'$, $\varepsilon' = \sum_{k=1}^{\infty} \varepsilon_k$ and $\mu = E(X)$, i. e. μ is the expected number of syllables per word. The parameters of the distribution $\{\varepsilon_k\}$ are called the ε-spectrum. For (2), the following conditions were postulated by Fucks:

(a) the condition $\varepsilon_k - \varepsilon_{k+1} \geq 0$ implies that $\varepsilon_{k+1} \leq \varepsilon_k$;
(b) since the sum of all weights equals 1, we have
$$1 = \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) = \sum_{k=0}^{\infty} \varepsilon_k - \sum_{k=0}^{\infty} \varepsilon_{k+1} = \varepsilon_0;$$
(c) from (a) and (b), we obtain $1 = \varepsilon_0 \geq \varepsilon_1 \geq \varepsilon_2 \geq \varepsilon_3 \geq ... \geq \varepsilon_k \geq \varepsilon_{k+1} ...$

It should be noted that (2) remains a valid probability distribution even when condition (a) is dropped. This more general case is not the aim of this chapter, but it will be discussed in subsequent papers.

In the following section, we go more into details, concentrating on the derivation of the Fucks' GPD and its generating function. Our aim will be to determine the unknown ε_k values which characterize the Fucks' GPD as given in equation (2).

2.1. The probability generating function and moments

Let us assume that the frequency distribution of some linguistic element is given; furthermore, we expect Fucks GPD (2) to fit the empirical data. Therefore, we know the theoretical representation of the observed distribution, if we find the estimated values of ε_k and λ. As to the estimation of the ε_k values, Fucks (1956d, 165) suggested to apply the method of moments. Since the moments are represented as polynomials in ε_k, one obtains algebraic equations for the ε_k, and the estimates by way of equating the theoretical with the empirical moments.

The estimation process is rather complex because many equations with unknown parameters ε_k ask for their solution. For many linguistic problems, it is sufficient, however, to indicate only a few moments of a given distribution. Below, it will be shown that the estimation process is easier for simple special cases of the Fucks GPD.

2.1.1. The probability generating function

In a first step, it seems reasonable to determine the probability generating function of the Fucks GPD. Generally speaking, the probability generating function is useful in simplifying mathematical proofs for discrete distributions. Given the generating function of the Fucks GPD, and knowing the statistical relations between the probability generating function and the moments of a distribution, the factorial, initial and central moments of Fucks' GPD will easily be derived.

Let X be a random variable with values in $\{0, 1, 2, ...\}$. The probability generating function is then defined by

$$G(t) = \sum_{i=0}^{\infty} p_i t^i = E(t^X), \tag{3}$$

where $p_i = P(X = i)$, $i = 0, 1, 2, ...$ Having $G(t)$ at our disposal, it is easy to find all factorial moments $\mu_{(k)}$ by differentiating function $G(t)$ k-times, and setting $t = 1$, i. e.

$$\mu_{(k)} = \frac{\partial^k G(t)}{\partial t^k}\bigg|_{t=1} = \sum_{i=0}^{\infty} i(i-1) \dots (i-k+1) p_i. \tag{4}$$

Multiplying the expressions in parenthesis, we can represent the factorial moments in terms of the initial moments. For our purposes, we need the first three initial moments, which are given as

$$\mu_{(1)} = \sum_{i=0}^{\infty} i\, p_i = \mu'_1 = E(X)$$
$$\mu_{(2)} = \sum_{i=0}^{\infty} i(i-1)\, p_i = \sum_{i=0}^{\infty} i^2 p_i - \sum_{i=0}^{\infty} i\, p_i = \mu'_2 - \mu'_1$$
$$\mu_{(3)} = \sum_{i=0}^{\infty} i(i-1)(i-2)\, p_i = \sum_{i=0}^{\infty} (i^3 - 3i^2 + 2i)\, p_i = \mu'_3 - 3\mu'_2 + 2\mu'_1. \tag{5}$$

The central moments, too, can be represented in terms of the factorial moments which are given as:

$$\mu_2 = Var(X) = \mu'_2 - (\mu'_1)^2 = \mu_{(2)} + \mu_{(1)} - \mu_{(1)}^2$$
$$\mu_3 = \mu_{(3)} + 3\mu_{(2)} + \mu_{(1)} - 3\mu_{(1)}\mu_{(2)} - 3(\mu_{(1)})^2 + 2(\mu_{(1)})^3. \tag{6}$$

Knowing the probability generating function $G(t)$, one can find all probabilities p_i as the coefficients of t^i by differentiating function $G(t)$ i-times and setting $t = 0$, i.e.

$$p_i = \frac{1}{i!} \cdot \frac{\partial^i G(t)}{\partial t^i} \bigg|_{t=0} \qquad (7)$$

Using equation (3), we can easily find the probability generating function of the Fucks GPD as

$$G(t) = e^{\lambda(t-1)} \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) t^k. \qquad (8)$$

Fucks did not present the relevant calculation, in detail; rather, he discussed the results of these calculations and gave an idea how to arrive at the moments of the distribution, as a result of taking the probability generating function into consideration (cf. Fucks 1956c, 523). Before going into details as to the problem of the moments, it is necessary to note that there are three important relations, which are useful for finding the factorial moments:

$$\sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) = \varepsilon_0 = 1$$

$$\sum_{k=0}^{\infty} k(\varepsilon_k - \varepsilon_{k+1}) = \sum_{k=1}^{\infty} \varepsilon_k$$

$$\sum_{k=0}^{\infty} k^2 (\varepsilon_k - \varepsilon_{k+1}) = \sum_{k=1}^{\infty} (2k-1)\varepsilon_k. \qquad (9)$$

2.1.2. Factorial moments

Let us now concentrate on the factorial moments of the Fucks GPD. In doing so, we assume that the probability function, given in (2), is known, except for its parameters ε_k. Hence, our further investigation has to be concentrated on the estimation of the parameters ε_k, by way of recourse to the moments of the sample. As was mentioned above, the simplest method, also recommended by Fucks, is the method of moments; substituting the theoretical moments by the empirical ones, we obtain the estimates of the unknown parameters.

Each text contains N words ($w_1, w_2, \ldots, w_N$). Word length is measured in syllables and can be different for each word; $x_j = i$ denotes the word length of word w_j, where $j = 1, 2, \ldots, N$; $i = 0, 1, 2, \ldots, I$. Actually, we are concerned with words of zero, one, two, three, ..., or I syllables. The whole number of words are divided into $I+1$ frequency classes; f_i refers to the number of elements that belong to the class i (absolute frequencies).

Let us first concentrate on the definition of the theoretical factorial moments. Given that

$$\mu_{(k)} = \frac{\partial^k G(t)}{\partial t^k} \bigg|_{t=1} = E(X(X-1)\ldots(X-k+1)), \qquad (10)$$

it is possible to determine the factorial moments by differentiating the generating function $G(t)$ k times, as was mentioned in Sec. 2.1.1. The first derivative of $G(t)$ is given as

$$\frac{\partial G(t)}{\partial t} = e^{\lambda(t-1)} \lambda \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) t^k + e^{\lambda(t-1)} \sum_{k=0}^{\infty} k t^{k-1} (\varepsilon_k - \varepsilon_{k+1}).$$

Substituting $t = 1$ in the first derivative of $G(t)$, one obtains the first factorial moment:

$$\mu_{(1)} = \frac{\partial^k G(t)}{\partial t} \bigg|_{t=1} = \lambda \underbrace{\sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1})}_{1} + \underbrace{\sum_{k=0}^{\infty} k(\varepsilon_k - \varepsilon_{k+1})}_{\varepsilon'} = \lambda + \varepsilon' = \mu. \qquad (11)$$

Since the second derivative of $G(t)$ is given as

$$\frac{\partial^2 G(t)}{\partial t^2} = e^{\lambda(t-1)} \lambda^2 \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) t^k + 2e^{\lambda(t-1)} \lambda \sum_{k=0}^{\infty} k(\varepsilon_k - \varepsilon_{k+1}) t^{k-1} + e^{\lambda(t-1)} \sum_{k=0}^{\infty} k(k-1)(\varepsilon_k - \varepsilon_{k+1}) t^{k-2},$$

the second derivative of $G(t)$ for $t = 1$ results in the second factorial moment:

$$\mu_{(2)} = \frac{\partial^2 G(t)}{\partial t^2} \bigg|_{t=1} = \lambda^2 \underbrace{\sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1})}_{1} + 2\lambda \underbrace{\sum_{k=0}^{\infty} k(\varepsilon_k - \varepsilon_{k+1})}_{\varepsilon'} + \underbrace{\sum_{k=0}^{\infty} k(k-1)(\varepsilon_k - \varepsilon_{k+1})}_{\sum_{k=1}^{\infty} (2k-1)\varepsilon_k - \sum_{k=1}^{\infty} \varepsilon_k}$$

$$= \lambda^2 + 2\lambda\varepsilon' + 2 \cdot \sum_{k=1}^{\infty} k \cdot \varepsilon_k - 2\varepsilon'.$$

Written in a different way, with $\mu = \lambda + \varepsilon'$, we thus have:

$$\mu_{(2)} = \mu^2 - \varepsilon'^2 - 2\varepsilon' + 2\sum_{k=1}^{\infty} k\varepsilon_k. \quad (12)$$

Analogically, the third derivative of $G(t)$ is given as

$$\begin{aligned}\frac{\partial^3 G(t)}{\partial t^3} &= e^{\lambda(t-1)}\lambda^3 \sum_{k=0}^{\infty}(\varepsilon_k - \varepsilon_{k+1})\, t^k + \\ &+ 3e^{\lambda(t-1)}\lambda^2 \sum_{k=0}^{\infty} k(\varepsilon_k - \varepsilon_{k+1})\, t^{k-1} + \\ &+ 3e^{\lambda(t-1)}\lambda \sum_{k=0}^{\infty} k(k-1)(\varepsilon_k - \varepsilon_{k+1})\, t^{k-2} + \\ &+ e^{\lambda(t-1)} \sum_{k=0}^{\infty} k(k-1)(k-2)(\varepsilon_k - \varepsilon_{k+1})\, t^{k-3}.\end{aligned}$$

For $t = 1$, the third derivative of $G(t)$ thus provides the third factorial moment:

$$\begin{aligned}\mu_{(3)} &= \lambda^3 \underbrace{\sum_{k=0}^{\infty}(\varepsilon_k - \varepsilon_{k+1})}_{1} + 3\lambda^2 \underbrace{\sum_{k=0}^{\infty} k(\varepsilon_k - \varepsilon_{k+1})}_{\varepsilon'} + \\ &+ 3\lambda \underbrace{\sum_{k=0}^{\infty} k(k-1)(\varepsilon_k - \varepsilon_{k+1})}_{\sum_{k=1}^{\infty}(2k-1)\varepsilon_k - \sum_{k=1}^{\infty}\varepsilon_k} + \sum_{k=0}^{\infty} k(k-1)(k-2)(\varepsilon_k - \varepsilon_{k+1}) \\ &= \lambda^3 + 3\lambda^2\varepsilon' + 3\lambda\left(2\sum_{k=1}^{\infty} k\varepsilon_k - 2\varepsilon'\right) + \sum_{k=1}^{\infty} k^3(\varepsilon_k - \varepsilon_{k+1}) - 6\sum_{k=0}^{\infty} k\varepsilon_k + 5\varepsilon'.\end{aligned}$$

Again, written in a different way, with $\mu = \lambda + \varepsilon'$, we have:

$$\begin{aligned}\mu_{(3)} &= \sum_{k=0}^{\infty} k^3(\varepsilon_k - \varepsilon_{k+1}) + \\ &+ 6(\mu - \varepsilon' - 1)\sum_{k=1}^{\infty} k\varepsilon_k + 5\varepsilon' + \\ &+ \mu^3 - 3\mu\varepsilon'^2 + 2\varepsilon'^3 - \\ &- 6\varepsilon'\mu + 6\varepsilon'^2. \quad (13)\end{aligned}$$

For an empirical distribution, the k-th factorial moment will consequently be calculated as

$$m_{(k)} = \frac{1}{N} \cdot \sum_{i=0}^{I} [i(i-1)(i-2) \ldots (i-k+1) f_i], \quad (14)$$

which serves as an estimate for the k-th theoretical factorial moment $\mu_{(k)}$

2.1.3. Central moments

As to the estimation of the unknown parameters ε_k, Fucks (1956a, 12) suggested their calculation by reference to the central moments μ_k. In analogy to the procedure discussed above (cf. Sec. 2.1.1.), the second and third central moments of the Fucks GPD can be determined, using their relations to the factorial moments, given as

$$\begin{aligned}\mu_2 &= \mu - \varepsilon'^2 - 2\varepsilon' + 2\sum_{k=1}^{\infty} k\varepsilon_k \\ &= \mu - \varepsilon'^2 + 2\sum_{k=1}^{\infty}(k-1)\,\varepsilon_k \\ \mu_3 &= \mu + 2\varepsilon'^3 + 3\varepsilon'^2 - \varepsilon' - \\ &- 6\varepsilon'\sum_{k=1}^{\infty} k\varepsilon_k + \sum_{k=0}^{\infty} k^3(\varepsilon_k - \varepsilon_{k+1}). \quad (15)\end{aligned}$$

The empirical central moment of the order k

$$m_k = \frac{1}{N-1} \cdot \sum_i (i - \bar{x})^k \cdot f_i \quad (16)$$

is an estimate of the k-th theoretical central moment μ_k. For example, the second central moment (estimate for the variance σ^2) will be calculated as:

$$\begin{aligned}m_2 &= \frac{1}{N-1} \cdot \sum_i (i - \bar{x})^2 \cdot f_i \\ &= \frac{1}{N-1} \cdot \left(\sum_i f_i \cdot i^2 - N \cdot \bar{x}^2\right). \quad (17)\end{aligned}$$

2.1.4. Initial Moments

The estimation of the unknown parameters can also be based on the initial moments of a given distribution, as will be demonstrated below (cf. Sec. 3.). In this case, the parameters ε_k will be estimated not with recourse to the central moments, but to the initial moments of the empirical distribution. Now, knowing the theoretical factorial moments and the relations between factorial and initial moments, one can easily determine the theoretical initial moments for the Fucks GPD. The first three initial moments of the Fucks generalized distribution can be derived from equation (5):

$$\mu'_1 = \mu_{(1)} = \mu$$
$$\mu'_2 = \mu_{(2)} + \mu_{(1)} = \mu^2 + \mu - \varepsilon'^2 - 2\varepsilon' + 2\sum_{k=1}^{\infty} k\varepsilon_k$$
$$\mu'_3 = \mu^3 + 3\mu^2 + \mu + 2\varepsilon'^3 + 3\varepsilon'^2 - \varepsilon' - 3\mu\varepsilon'^2 - 6\mu\varepsilon' + \sum_{k=0}^{\infty} k^3 (\varepsilon_k - \varepsilon_{k+1}) + 6(\mu - \varepsilon')\sum_{k=1}^{\infty} k\varepsilon_k. \quad (18)$$

By the method of moments, $E(X) = \mu$ is estimated as the arithmetical mean. We thus obtain (19) as the estimate of the first initial moment μ:

$$m'_1 = \bar{x} = \frac{1}{N} \cdot \sum_{i=1}^{I} i \cdot f_i. \quad (19)$$

Generally, the k-th initial moment will be estimated as

$$m'_k = \frac{1}{N} \cdot \sum_{i=1}^{I} i^k \cdot f_i. \quad (20)$$

2.2. Special cases of the Fucks GPD

As was mentioned above, the Fucks GPD has hardly ever been applied to linguistic material. As to this question, Fucks favored the 1-displaced Poisson distribution, which he considered to be the "mathematical law of the process of word-formation from syllables for all those languages, which form their words from syllables" (cf. Fucks 1955b, 209). It seems to be this particular focus on the 1-displaced Poisson distribution, why this model, though being only one special case of this GPD, has often been assumed to be "the Fucks distribution".

In this section, it will be shown that the 1-displaced Poisson distribution is a special case of the Fucks GPD; additionally, two more special cases which can be derived from the Fucks GPD, will be discussed in detail: depending on the number of parameters taken into consideration, one may distinguish one-, two-, and three-parameter special cases of Fucks' GPD, which will be considered in this order.

2.2.1. A one-parameter special case (Fucks-Čebanov Distribution)

Let us first direct our attention to the one-parameter model as the simplest of all special cases mentioned. As can be seen from equation (2), the Fucks GPD includes the Poisson distribution and the 1-displaced Poisson distribution, as two of its special cases. Assuming that $\varepsilon_0 = 1$ and $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \ldots = 0$, the Fucks GPD (2) leads to the Poisson distribution:

$$p_i = \frac{e^{-\lambda} \cdot \lambda^i}{i!}, \; i = 0, 1, 2 \ldots \quad (21)$$

where $\hat{\lambda} = \bar{x}$. If we choose $\varepsilon_0 = \varepsilon_1 = 1$ and $\varepsilon_2 = \varepsilon_3 = \ldots = 0$ (i.e. assuming our sample has no zero-syllable words), we obtain the 1-displaced Poisson distribution:

$$p_i = e^{-\lambda} \cdot \frac{\lambda^{i-1}}{(i-1)!}, \; i = 1, 2, \ldots \quad (22)$$

Here, $\lambda = \mu - 1$, $\hat{\mu} = \bar{x}$. Fucks repeatedly applied the 1-displaced Poisson distribution to linguistic data from various languages. Unfortunately, he did not, however, calculate any tests to check the significance of the goodness of his fits, as was quite usual at his time. In order to compensate this lack, one of the possibilities would be to calculate Pearson's χ^2, as defined by formula (23):

$$\chi^2 = \sum_{i=1}^{I} \frac{(f_i - E_i)^2}{E_i}. \quad (23)$$

In formula (23), I denotes the number of classes, f_i is the observed absolute frequency of a given class, and E_i is the theoretical absolute frequency. The statistical significance of this χ^2 value depends on the degrees of freedom ($d.f.$) of the corresponding χ^2 distribution, which are calculated with regard to the number of classes I minus 1, on the one hand, and the number of parameters k, which have to be estimated, on the other hand: $d.f. = I - k - 1$. Being convinced that his data were not particularly adequate for the application of the χ^2 test, Fucks went a different way and tried to empirically prove the adequacy of this model, using graphical techniques, only. One of the reasons for Fucks' rejection of the the χ^2 test probably is the fact that the χ^2 value linearly increases with an increase of the sample size – and linguistic samples tend to be rather large. From a contemporary point of view, to avoid this problem, it has become a common practice to calculate a standardization of χ^2 scores. Thus, in contemporary linguistics, the so-called discrepancy coefficient (C) meets broad acceptance, being defined as $C = \chi^2 / N$. The discrepancy coefficient has the advantage that it is not dependent on the degrees of freedom. One speaks of a good fit for $C < 0.02$, and of a very good fit for $C < 0.01$.

It seems reasonable to run a re-analysis of linguistic data given by Fucks, and to statistically test the goodness of fit of the 1-displaced Poisson distribution, including the χ^2 value as well as the discrepancy coefficient C. Unfortunately, Fucks never presented any raw data (what was quite usual at his times); rather, he confined himself to presenting relative instead of absolute frequencies. This fact renders it almost impossible to control the results at which he arrived; the only way to do a re-analysis, is to create artificial samples of ca. 10.000 each, by way of multiplying the relative frequencies given by him with 10.000.

Table 11a.1 represents data from various languages presented by Fucks (1956a, 10; 1956d, 157). In addition to the relative frequencies of x-syllable words, also mean word length for each of the nine languages are contained.

Table 11a.1: Relative frequencies and mean word length for different languages (Fucks 1956a)

	English	German	Esperanto	Arabic	Greek
1	0.7152	0.5560	0.4040	0.2270	0.3760
2	0.1940	0.3080	0.3610	0.4970	0.3210
3	0.0680	0.0938	0.1770	0.2239	0.1680
4	0.0160	0.0335	0.0476	0.0506	0.0889
5	0.0056	0.0071	0.0082	0.0017	0.0346
6	0.0012	0.0014	0.0011	–	0.0083
7	–	0.0002	–	–	0.0007
8	–	0.0001	–	–	–
$\bar{x}$	1.4064	1.6333	1.8971	2.1032	2.1106

	Japanese	Russian	Latin	Turkish
1	0.3620	0.3390	0.2420	0.1880
2	0.3440	0.3030	0.3210	0.3784
3	0.1780	0.2140	0.2870	0.2704
4	0.0868	0.0975	0.1168	0.1208
5	0.0232	0.0358	0.0282	0.0360
6	0.0124	0.0101	0.0055	0.0056
7	0.004	0.0015	0.0007	0.0004
8	0.0004	0.0003	0.0002	0.0004
9	0.0004	–	–	–
$\bar{x}$	2.1325	2.2268	2.3894	2.4588

Table 11a.2 represents the results of the goodness-of-fit test, giving the C values for each language. It is obvious that the 1-displaced Poisson distribution is not equally appropriate for the linguistic data given by Fucks: the model is appropriate only for Esperanto (best fit), Latin, and German, but inappropriate for the remaining six languages. One reason for this failure might be the fact that the data for each of the languages originated from text mixtures, not from individual texts. Therefore, the material might be characterized by an internal heterogeneity, violating the statistical principle of data homogeneity. Since this point cannot be pursued in detail, here, it will be ignored, and Fucks' data shall be used throughout this text, understanding them as exemplary linguistic data.

Table 11a.2: Discrepancy coefficient C as a result of fitting the 1-displaced Poisson distribution to different languages (Fucks 1956a)

English	German	Esperanto	Arabic	Greek
0.0903	0.0186	0.0023	0.1071	0.0328

Japanese	Russian	Latin	Turkish
0.0380	0.0208	0.0181	0.0231

Another way to test the 1-displaced Poisson distribution has been presented by Grotjahn (1982). Grotjahn discussed the 1-displaced Poisson distribution, particularly focusing the question, under which empirical conditions this model may turn out to be adequate for word length frequencies. Again, faced with the problem of the χ^2 test for the analysis of linguistic data, Grotjahn (1982, 53) suggested to calculate the dispersion quotient (δ), defined as the quotient of the theoretical values for the variance (σ^2) and the mean (μ):

$$\delta = \frac{\sigma^2}{\mu}. \qquad (24)$$

For r-displaced distributions, the corresponding equation is

$$\delta = \frac{\sigma^2}{\mu - r}, \qquad (25)$$

r being the displacement parameter.

It goes without saying, that the coefficient δ can be estimated from the data as

$$d = \frac{m_2}{\bar{x} - r}. \tag{26}$$

Table 11a.3: Values of the dispersion quotient d for different languages

English	German	Esperanto	Arabic	Greek
1.3890	1.1751	0.9511	0.5964	1.2179
Japanese	Russian	Latin	Turkish	
1.2319	1.1591	0.8704	0.8015	

Now, it is easy to check the goodness of fit of the Poisson model to empirical data. For this purpose, it is necessary to calculate the empirical value d, and to compare it with the theoretical value δ. Since, for the 1-displaced Poisson distribution, the variance $Var(X) = \sigma^2 = \mu - 1$ and $E(X) = \mu$, we have

$$\delta = \frac{\mu - 1}{\mu - 1} = 1.$$

There is an important consequence to be drawn from the fact that, for the 1-displaced Poisson distribution, $\delta = 1$: it can be an adequate model only as long as an empirical sample delivers $d \approx 1$. Therefore we calculate the d values for each of the nine languages represented in Table 11a.1 above; the results are given in Table 11a.3.

Now checking again the adequacy of the 1-displaced Poisson distribution, as represented in Table 11a.2, and comparing it with the d values in Table 11a.3, it can easily be seen that in fact, there is no good fit ($C > 0.02$) for those cases where d is significantly different from $d = 1$.

Summarizingly, we can thus say that, although Fucks himself did not apply any test to check the goodness-of-fit for the 1-displaced Poisson distribution, and although the graphical illustrations of Fucks' fittings force us to believe that the 1-displaced Poisson distribution can be expected to be adequate, statistical tests show that is the case only for samples with $d \approx 1$. In other words: the 1-displaced Poisson distribution must be rejected as an acceptable, overall valid standard model for word length frequencies. With this conclusion in mind, the next logical step is an analysis of Fucks' two- and three-parameter special cases of his GPD.

2.2.2. A two-parameter special case (Dacey-Poisson Distribution)

In the previous section, the Poisson distribution was discussed, both in its 'ordinary' and in its 1-displaced form, as a one-parameter special case of the Fucks GPD. In either case, only one parameter (λ) has to be estimated. In this section, we will concentrate on another special case of the Fucks GPD, with two parameters to be estimated.

Setting $\varepsilon_0 = 1$, $\varepsilon_1 = \alpha$ and $\varepsilon_k = 0$, $k \geq 2$ in Fucks GPD (2), yields a two-parameter distribution. This distribution, which tends to be termed Dacey-Poisson distribution in contemporary research (cf. Wimmer/Altmann 1999, 111), has been discussed by Fucks (1955b, 207) as another special case of his GPD, though not by this name, and only in its 1-displaced form – see below, (29). In its ordinary form, it takes the following shape:

$$p_i = (1-\alpha)\frac{e^{-\lambda}\lambda^i}{i!} + \alpha\frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!}, \quad i = 0, 1, 2, \ldots \tag{27}$$

with $\lambda = \mu - \alpha$. In addition to λ, a second parameter (α) has to be estimated, which, referring to (15), can be estimated as $\hat{a} = \sqrt{\bar{x} - m_2}$.

Before analyzing this model (in its 1-displaced form) and testing its adequacy, it seems worthwhile mentioning, that on particular conditions, namely for $\mu = 2\alpha$, $\varepsilon_0 = 1$, $\varepsilon_1 = \alpha$ and $\varepsilon_k = 0$, $k \geq 2$, the Fucks GPD implies the so-called Kemp-Kemp-Poisson distribution (cf. Wimmer/Altmann 1999, 344) with:

$$\begin{aligned} p_i &= (1-\alpha)\frac{e^{-\lambda}\cdot\lambda^i}{i!} + \alpha\frac{e^{-\lambda}\cdot\lambda^{i-1}}{(i-1)!} \\ &= (1-\alpha)\frac{e^{-\alpha}\cdot\alpha^i}{i!} + \alpha\frac{e^{-\alpha}\cdot\alpha^{i-1}}{(i-1)!} \\ &= \frac{e^{-\alpha}\cdot\alpha^i}{i!}\cdot(i-\alpha+1), \quad i = 0, 1, 2, \ldots \end{aligned} \tag{28}$$

Here, we have $\lambda = \mu - \alpha = \alpha$.

Similarly, *for* $\varepsilon_0 = \varepsilon_1 = 1$, $\varepsilon_2 = \alpha$ and $\varepsilon_k = 0$, $k \geq 3$, the 1-displaced Dacey-Poisson model results from the Fucks GPD (2) (cf. Wimmer/Altmann 1999, 111) as

$$p_i = (1-\alpha)\cdot e^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!} + \alpha\cdot e^{-\lambda}\frac{\lambda^{i-2}}{(i-2)!}, \quad i = 1, 2, \ldots, \tag{29}$$

with $\lambda = \mu - \alpha - 1$. In this case, α can be estimated as $\hat{\alpha} = \sqrt{\bar{x} - 1 - m_2}$, what can be concluded from equation (15).

Now, turning to the two-parameter Dacey-Poisson distribution (29), discussed by Fucks (though not by this name), it might be interesting to apply this model to Fucks' data (cf. Table 11a.1) and relate the results to the values of the dispersion quotient (cf. Table 11a.3).

For this purpose, a re-analysis of Fucks' data is necessary. The results are represented in Table 11a.4, indicating the values of the discrepancy coefficient C for each language, as a result of fitting the 1-displaced Dacey-Poisson model.

Table 11a.4: Discrepancy coefficient C as a result of fitting the 1-displaced (two-parameter) Dacey-Poisson distribution to different languages

	English	German	Esperanto	Arabic	Greek
C	–	–	0.0019	0.0077	–
d	1.3890	1.1751	0.9511	0.5964	1.2179
	Japanese	Russian	Latin	Turkish	
C	–	–	0.0149	0.0021	
d	1.2319	1.1591	0.8704	0.8015	

As can be seen from Table 11a.4, in some cases, the results are slightly better than those obtained from fitting the 1-displaced Poisson distribution (cf. Table 11a.2). Again, the best fit is obtained for Esperanto. The 1-displaced Dacey-Poisson model provides a very good fit for Arabic, in contrast to the 1-displaced Poisson model. In some cases, however, no valid results can be obtained; this is due to the fact that the estimate $\hat{\alpha} = \sqrt{\bar{x} - 1 - m_2}$ of α is not defined if $\bar{x} - 1 \leq m_2$.

In order to arrive at an explanation of this problem, we refer again to Grotjahn's work, analyzing the theoretical scope of Fucks' two-parameter model. Let us, in analogy to the procedure above, discuss the theoretical dispersion quotient δ for Fucks' two-parameter distribution (29).

Since in this case, $Var(X) = \mu - 1 - \varepsilon_2^2$ and $E(X) = \mu$, it turns out that $\delta \leq 1$. This means that the two-parameter model is likely to be adequate as a theoretical model for empirical samples with $d \leq 1$. Now, once again checking the dispersion quotient d for the specific languages (cf. Table 11a.4), it becomes clear, why the results for the two-parameter model are appropriate only in case of Esperanto, Arabic, Latin and Turkish: in all these cases, the d value is smaller than 1.

This rather poor result leads to the question whether the three-parameter special case of the Fucks GPD is more adequate as an overall model for his data.

2.2.3. A three-parameter special case

In case of the three-parameter model, in addition to μ, two more parameters (ε_2 and ε_3) from the whole ε-spectrum have to be estimated. Setting $\varepsilon_0 = \varepsilon_1 = 1$ and $\varepsilon_k = 0$, $k \geq 4$, and $\varepsilon_2 = \alpha$, $\varepsilon_3 = \beta$, results in the three-parameter special case of the Fucks GPD:

$$\begin{aligned} p_i &= P(X = i) \\ &= e^{-(\mu-1-\alpha-\beta)} \sum_{k=1}^{3} (\varepsilon_k - \varepsilon_{k+1}) \cdot \\ &\quad \cdot \frac{(\mu - 1 - \alpha - \beta)^{i-k}}{(i-k)!}. \end{aligned} \quad (30)$$

Replacing $\lambda = \mu - 1 - \alpha - \beta$, the probability function has the following form:

$$\begin{aligned} p_1 &= e^{-\lambda} \cdot (1 - \alpha) \\ p_2 &= e^{-\lambda} \cdot [(1 - \alpha) \cdot \lambda + (\alpha - \beta)] \\ p_i &= e^{-\lambda} \left[(1 - \alpha) \frac{\lambda^{i-1}}{(i-1)!} + \right. \\ &\quad \left. + (\alpha - \beta) \frac{\lambda^{i-2}}{(i-1)!} + \beta \frac{\lambda^{i-3}}{(i-3)!} \right], \\ &\qquad i \geq 3. \end{aligned} \quad (31)$$

In the next step, α and β have to be estimated; in order to do so, we can use the second and third theoretical central moment of the Fucks GPD; according to (15), we thus obtain equation (32)

$$\begin{aligned} \mu_2 &= \mu - (1 + \alpha + \beta)^2 + 2(\alpha + 2\beta) \\ &= \mu - 1 - (\alpha + \beta)^2 + 2\beta \\ \mu_3 &= \mu + 2(1 + \alpha + \beta)^3 - 3(1 + \alpha + \beta)^2 - \\ &\quad - 6(\alpha + \beta)(\alpha + 2\beta) + 6\beta. \end{aligned} \quad (32)$$

Simplifying the system of equations (32) with $\alpha_+ = \alpha + \beta$, we get the following formula:

$$\begin{aligned} \mu_2 &= \mu - 1 - \alpha_+^2 + 2\beta \\ \mu_3 &= \mu + 2(1 + \alpha_+)^3 - 3(1 + \alpha_+)^2 - \\ &\quad - 6\alpha_+(\alpha_+ + \beta) + 6\beta \\ &= \mu - 1 + 2\alpha_+^3 - 3\alpha_+^2 - 6\alpha_+\beta + 6\beta. \end{aligned} \quad (33)$$

The solution of the 2×2 system (33) is a cubic equation with regard to parameter α_+:

$$\alpha_+^3 - 3\alpha_+(\mu - 1 - \mu_2) + \mu_3 - \\ - 3\mu_2 + 2\mu - 2 = 0. \tag{34}$$

In other words: substituting the theoretical moments by the empirical ones and setting $a = \hat{a}_+$, we get

$$a^3 - 3a(\bar{x} - 1 - m_2) + m_3 - \\ - 3m_2 + 2\bar{x} - 2 = 0. \tag{35}$$

As a result of this equation, three solutions are obtained, not all of which are necessarily real ones. For each real solution a (possibly a_i, $i = 1, 2, 3$), the values for $\varepsilon_2 = \alpha$ and $\varepsilon_3 = \beta$ have to be estimated; this can easily be done by computer programs. Before further analyzing this estimation, let us remind that there are two important conditions:

(a) $\varepsilon_2 = \alpha \leq 1$ and $\varepsilon_3 = \beta \leq 1$,
(b) $\varepsilon_2 = \alpha \geq \beta = \varepsilon_3$.

With these two conditions, we can now analyze the data presented in Table 11a.1, this time fitting Fucks' three-parameter model. The results are listed in Table 11a.5; results not satisfying the two conditions above, are marked as ∅.

Table 11a.5: Discrepancy coefficient C as a result of fitting Fucks' three-parameter Poisson distribution to different languages

	English	German	Esperanto	Arabic	Greek
C	∅	∅	0.00004	0.0021	∅
$\hat{\varepsilon}_2$	–	–	0.3933	0.5463	–
$\hat{\varepsilon}_3$	–	–	0.0995	−0.1402	–

	Japanese	Russian	Latin	Turkish
C	∅	0.0005	0.0003	0.0023
$\hat{\varepsilon}_2$	–	0.2083	0.5728	0.6164
$\hat{\varepsilon}_3$	–	0.1686	0.2416	0.1452

It can be observed that, quite reasonably, in some cases the results for the three-parameter model are better than those of the two models discussed above (cf. the results represented in Tables 11a.2 and 11a.4).

As to the result for Arabic, it should be noted that the value for $\hat{\varepsilon}_3$ is negative; this is due to the fact that we confine ourselves to Fucks's conditions (a)–(c) outlined above (cf. Sec. 2). The introduction of the additional condition, $0 < \varepsilon_k < 1$, $k = 2, 3$, results in another solution which is slightly worse, with $\hat{\varepsilon}_2 = 0.7174$, $\hat{\varepsilon}_3 = 0.1805$, $C = 0.0058$.

It can also be seen that there are no solutions for four of the languages, and it seems worth while trying to find an explanation for this finding.

Relating these results to the d values of the individual languages (cf. Table 11a.3), one can also see that the three-parameter model may be an appropriate model for empirical distributions, in which $d > 1$ (what was a crucial problem for the two models described above). Thus, in the Russian sample, for example, where $d = 1.1590$, the discrepancy coefficient is $C = 0.00054$. However, as the results for the German and Japanese data (with $d = 1.175$ and $d = 1.2319$, respectively) show, d does not seem to play the decisive role in case of the three-parameter model. Obviously, there seem to be other limitations responsible for the possible inadequacy of this model. In fact, in each of these examples where there is no solution for Fucks' three-parameter distribution, condition $\varepsilon_2 = \alpha \geq \beta = \varepsilon_3$ is not fulfilled.

Using the fact that $a = \hat{\alpha} + \hat{\beta}$ implies $\hat{\alpha} = a - \hat{\beta}$, as well as $m_2 = \bar{x} - 1 - a^2 + 2\hat{\beta}$ implies $\hat{\beta} = \dfrac{a^2 + m_2 - \bar{x} + 1}{2}$, condition $\hat{\alpha} \geq \hat{\beta}$ can be written as:

$$\frac{2a - a^2 - m_2 + \bar{x} - 1}{2} \geq \\ \geq \frac{a^2 + m_2 - \bar{x} + 1}{2}. \tag{36}$$

We thus define the difference $M = \bar{x} - m_2$, i. e. the difference between the mean of the empirical distribution ($\bar{x}$) and its variance (m_2). As a result, equation (36) can be simplified to

$$2a - a^2 + M - 1 \geq a^2 - M + 1,$$

Likewise, it can be written as:

$$a^2 - a + (-M + 1) \leq 0.$$

As a consequence, one obtains the following two conditions:

(a) The sum $a = \hat{\varepsilon}_2 + \hat{\varepsilon}_3 = \hat{\alpha} + \hat{\beta}$ must be in a particular interval for each of three possible solutions of a:

$$a_i \in \left[\frac{1 - \sqrt{4M - 3}}{2}, \frac{1 + \sqrt{4M - 3}}{2}\right] \qquad i = 1, 2, 3.$$

Thus, there are two interval limits a_{i1} and a_{i2}:

$$a_{i1} = \frac{1 - \sqrt{4M - 3}}{2}$$

and

$$a_{i2} = \frac{1 + \sqrt{4M - 3}}{2}.$$

Table 11a.6: Violations of the conditions for Fucks' three-parameter model

	English	German	Esperanto	Arabic	Greek
C	∅	∅	<0.01	<0.01	∅
$\hat{\varepsilon}_2$	–	–	0.3933	0.5463	–
$\hat{\varepsilon}_3$	–	–	0.0995	−0.1402	–
$a = \hat{\varepsilon}_2 + \hat{\varepsilon}_3$	−0.0882	−0.1037	0.4929	0.4061	0.2799
a_{i1}	0.1968	0.1270	−0.0421	−0.3338	0.4108
a_{i2}	0.8032	0.8730	1.0421	1.3338	0.5892
$a_{i1} < a < a_{i2}$	–	–	✓	✓	–
$\bar{x}$	1.4064	1.6333	1.8971	2.1032	2.1106
m_2	0.5645	0.7442	0.8532	0.6579	1.3526
$M = \bar{x} - m_2$	0.8420	0.8891	1.0438	1.4453	0.7580
$M \geq 0.75$	✓	✓	✓	✓	✓
	Japanese	Russian	Latin	Turkish	
C	∅	<0.01	<0.01	<0.01	
$\hat{\varepsilon}_2$	–	0.2083	0.5728	0.6164	
$\hat{\varepsilon}_3$	–	0.1686	0.2416	0.1452	
$a = \hat{\varepsilon}_2 + \hat{\varepsilon}_3$	−0.1798	0.3769	0.8144	0.7616	
a_{i1}	$\mathbb{C}$	0.2659	−0.1558	−0.2346	
a_{i1}	$\mathbb{C}$	0.7341	1.1558	1.2346	
$a_{i1} < a < a_{i2}$	–	✓	✓	✓	
$\bar{x}$	2.1325	2.2268	2.3894	2.4588	
m_2	1.3952	1.4220	1.2093	1.1692	
$M = \bar{x} - m_2$	0.7374	0.8048	1.1800	1.2896	
$M \geq 0.75$	–	✓	✓	✓	

(b) In order to be $a \in \mathbb{R}$, the root $4M - 3$ must be positive, i. e. $4M - 3 \geq 0$; therefore, $M = \bar{x} - m_2 \geq 0.75$.

Inspecting the results in Table 11a.6, it can clearly be seen why, in four of the nine cases, the results are not satisfying: there are a number of violations, which are responsible for the failure of Fucks' three-parameter model. These violations can be caused by two facts:

(a) As soon as $M < 0.75$, the definition of the interval limits of a_{i1} and a_{i2} involves a negative root – this is the case with the Japanese data, for example;
(b) Even if the first condition is fulfilled with $M \geq 0.75$, fitting Fucks' threeparameter model may fail, if $a < a_{i1}$ – this is the case for English, German, and Greek – or if $a > a_{i2}$.

The three-parameter Fucks' model thus is adequate only for particular types of empirical distributions, and it can not serve as an overall model for language, not even for languages which form their words from syllables, as Fucks himself claimed.

However, some of the problems faced might have their foundation in related issues. As was mentioned above, one possible explanation might be the heterogeneity of the data material, inherent in any linguistic corpus; another reason might be motivated by the specific manner of estimating the parameters, suggested by Fucks – and this, in turn, might be the cause why some authors, though generally following Fucks' line of thinking, tried to find alternative ways to estimate the parameters of the Fucks GPD.

3. A generalization of the Fucks GPD (Fucks-Gačečiladze Distribution)

As to the reception of Fucks' ideas, it is strange enough that they were relatively soon well-known among scholars from Eastern European and the former Soviet Union. Quite early, for example, three Georgian scholars, G. N. Cercvadze, G. B. Čikoidze, and T. G. Gačečiladze (1959), applied Fucks' ideas to Georgian linguistic material, mainly concentrating on phoneme frequencies and word length frequencies. Their study, which was translated into German in 1962, was originally inspired by the Russian translation of Fucks' English-written article "Mathematical Theory of Word Formation" (Fucks 1956d).

In fact, Cercvadze, Čikoidze, and Gačečiladze (1959) based their analyses on Fucks's generalization of the Poisson distribution. These authors, in turn, once more generalized Fucks' model. This additional generalization is not explicitly discussed in the early 1959 paper; rather, these authors' extension is represented in subsequent papers – cf. Gačečiladze/Cercvadze/Čikoidze (1961); Bokučava/Gačečiladze (1965), Bokučava et al. (1965), and Gačečiladze/Cilosani (1971).

Basically, this extension contains an additional factor (φ_k), which, in turn, depends on three parameters: (a) the sum of all $\varepsilon_k = A$ (termed ε' by Fucks), (b) the mean of the sample (μ), and (c) the relevant class i. As a result, the individual weights of the Fucks GPD, defined as $(\varepsilon_k - \varepsilon_{k+1})$, are multiplied by the function φ_k.

Unfortunately, Gačečiladze/Cilosani (1971, 114) do not exactly explain the process by which φ_k may be theoretically derived; they only present the final formula (39). In the earlier papers mentioned above (cf. Gačečiladze/Cercvadze/Čikoidze 1961, 5; Bokučava/Gačečiladze 1965, 174), some of the relevant ideas are presented, but not the exact procedure how to arrive at formula (39). The same lack of information characterizes subsequent reports on the Georgians' work as, e.g., by Piotrovskij/Bektaev/Piotrovskaja (1977, 195). These Russian scholars, too, who term formula (39) the "Fucks-Gacečiladze distribution", give no derivation of φ_k.

Basically, the Gačečiladze generalization, resulting in the Fucks-Gačečiladze distribution (39), is based on the assumption that the process of text generation is a stochastic process; elaborating this assumption, Gačečiladze/Cercvadze/Čikoidze (1961, 5) arrive at the following formula, representing a sum of weighted binomial probabilities:

$$V(i; p, q) = \sum_{k=0}^{n} (\varepsilon_k - \varepsilon_{k+1}) \cdot \binom{n-k}{i-k} p^{i-k} q^{i-k} [1 - p + p(1-q)]^{n-i}. \quad (37)$$

Here, the difference $(\varepsilon_k - \varepsilon_{k+1})$ represents the statistical weight of the system's status before the beginning of the distribution process. If $p \in [0, 1]$, equation (37) takes the following form:

$$\begin{aligned} P(i; q) &= \sum_{k=0}^{n} (\varepsilon_k - \varepsilon_{k+1}) \cdot \binom{n-k}{i-k} \int_0^1 p^{i-k} q^{i-k} [1 - p + p(1-q)]^{n-i} dp \\ &= \sum_{k=0}^{n} (\varepsilon_k - \varepsilon_{k+1}) \cdot \binom{n-k}{i-k} \int_0^1 (p \cdot q)^{i-k} (1 - pq)^{n-i} dp. \quad (38) \end{aligned}$$

For $q \to 0, n \to \infty$, as the limit of (38), we obtain the generalized Fucks-Gačečiladze distribution (39):

$$p_i = P(X = i) = e^{-(\mu - A)} \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) \cdot \frac{(\mu - A)^{i-k}}{(i-k)!} \varphi_k(A, \mu, i) \quad (39)$$

with

$$\varphi_k(A, \mu, i) = \frac{1}{2} \int_{-1}^{+1} (t+1)^{i-k} e^{-(\mu - A) \cdot t} dt.$$

Here, μ denotes the mean word length:

$$\begin{aligned} \mu &= \sum_{i=0}^{n} i P(X = i) \\ &= \frac{q}{2} \left(n - \sum_{k=1}^{n-1} \varepsilon_k \right) + \sum_{k=1}^{n-1} \varepsilon_k \end{aligned}$$

with A being defined as

$$A = \sum_{k=1}^{\infty} \varepsilon_k < +\infty \text{ (the series converges).}$$

In the next step, we want to represent formula (39) similarly as formula (2); denoting $\lambda = (\mu - A)$, we thus obtain:

$$p_i = e^{-\lambda} \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) \frac{\lambda^{i-k}}{(i-k)!} \varphi_k(\lambda, i)$$

$$\varphi_k(\lambda, i) = \frac{1}{2} \int_{-1}^{+1} (t+1)^{i-k} e^{-\lambda \cdot t} dt. \quad (40)$$

Again, in (40) as well as in case of the Fucks GPD (2), the conditions $\varepsilon_0 = 1 \geq \varepsilon_k \geq \varepsilon_{k+1}, k \geq 1$ are stated. In principle, distribution (40) is the already known Fucks GPD, multiplied, however, by the function φ_k – cf. formula (2). As to φ_k, it can be shown that the following recurrent relation holds:

$$\varphi_k = -\frac{2^{k-1} e^{-\lambda}}{\lambda} + \frac{k}{\lambda} \cdot \varphi_{k-1}. \qquad (41)$$

In order to prove relation (41), it is necessary to re-write function φ_k in a different form. Therefore denoting $i - k = l$, and assuming that in the function φ_k, $z = (t + 1)\lambda$, we can write:

$$\varphi_l = \frac{1}{2}\int_{-1}^{+1} (t+1)^l e^{-\lambda \cdot t}\, dt$$
$$= \frac{1}{2}\int_0^{2\lambda} \left(\frac{z}{\lambda}\right)^l e^{\lambda - z} \frac{dz}{\lambda}$$
$$= \frac{1}{2} \cdot \frac{e^\lambda}{\lambda^{l+1}} \int_0^{2\lambda} z^l e^{-z}\, dz. \qquad (42)$$

Let us prove (41), transforming the right-hand side of the relation (41) and using (42):

$$\varphi_k = -\frac{2^{k-1} e^{-\lambda}}{\lambda} + \frac{k}{\lambda} \cdot \frac{1}{2} \cdot \frac{e^\lambda}{\lambda^k} \int_0^{2\lambda} z^{k-1} e^{-z}\, dz$$
$$= -\frac{2^{k-1} e^{-\lambda}}{\lambda} + \frac{1}{2} \cdot \frac{e^\lambda}{\lambda^{k+1}} \int_0^{2\lambda} k \cdot z^{k-1} e^{-z}\, dz$$
$$= \frac{1}{2} \cdot \frac{e^\lambda}{\lambda^{k+1}} \left(\int_0^{2\lambda} k \cdot z^{k-1} e^{-z}\, dz - \right.$$
$$\left. - (2\lambda)^k \cdot e^{-2\lambda} \right)$$
$$= \frac{1}{2} \cdot \frac{e^\lambda}{\lambda^{k+1}} \left(\int_0^{2\lambda} k \cdot z^{k-1} e^{-z}\, dz - \right.$$
$$\left. - \int_0^{2\lambda} (z^k e^{-z})'\, dz \right)$$
$$= \frac{1}{2} \cdot \frac{e^\lambda}{\lambda^{k+1}} \left(\int_0^{2\lambda} k \cdot z^{k-1} e^{-z}\, dz - \right.$$
$$\left. - \int_0^{2\lambda} (k \cdot z^{k-1} e^{-z} - z^k e^{-z})\, dz \right)$$
$$= \frac{1}{2} \cdot \frac{e^\lambda}{\lambda^{k+1}} \int_0^{2\lambda} z^k e^{-z}\, dz.$$

Another way to express function φ_k would imply a recourse to the Γ function, defined as

$$\Gamma(a) = \int_0^\infty z^{a-1} e^{-z}\, dz, \; a > 0.$$

Substituting $z = a + 2\lambda$ in the transformed form of equation (42), and using the fact that

$$(a + 2\lambda)^k = \binom{k}{0} (2\lambda)^k a^0 + \ldots + \binom{k}{k} (2\lambda)^0 a^k$$
$$= \sum_{s=0}^{k} \binom{k}{s} (2\lambda)^{k-s} a^s$$

the following equations hold:

$$\varphi_k = \frac{1}{2} \cdot \frac{e^\lambda}{\lambda^{k+1}} \int_0^{2\lambda} z^k e^{-z}\, dz$$
$$= \frac{1}{2} \cdot \frac{e^\lambda}{\lambda^{k+1}} \left(\int_0^\infty z^k e^{-z}\, dz - \int_{2\lambda}^\infty z^k e^{-z}\, dz \right)$$
$$= \frac{1}{2} \cdot \frac{e^\lambda}{\lambda^{k+1}} \left(\int_0^\infty z^k e^{-z} dz - \right.$$
$$\left. - e^{-2\lambda} \int_0^\infty (a + 2\lambda)^k e^{-a}\, da \right)$$
$$= \frac{e^\lambda}{2 \cdot \lambda^{k+1}} \left\{ \Gamma(k+1) - \right.$$
$$\left. - e^{-2\lambda} \sum_{s=0}^{k} \binom{k}{s} (2\lambda)^{k-s} \cdot \int_0^\infty a^s e^{-a}\, da \right\}$$
$$= \frac{e^\lambda}{2 \cdot \lambda^{k+1}} \left\{ \Gamma(k+1) - \right.$$
$$\left. - e^{-2\lambda} \sum_{s=0}^{k} \binom{k}{s} (2\lambda)^{k-s} \Gamma(s+1) \right\}. \qquad (43)$$

Comparing the two distributions (2) and (40), it can easily be seen that (40) yields the Fucks GPD, for $\varphi_k \to 1$. Hence, the Fucks GPD turns out to be a special case of the Fucks-Gačečiladze distribution.

Yet, as mentioned above, the Georgian authors only presented their generalization of Fucks' GPD, without mathematical derivation. Also, they applied this generalized model to specific linguistic data (letter frequencies), but without empirically testing its goodness of fit (cf. Gačečiladze/Cercvadze/Čikoidze, 1961, Bokučava/Gačečiladze 1965, Gačečiladze/Cilosani 1971). In the next step, our aim is to estimate parameters ε_k which characterize distribution (40). Therefore, it is a tempting step to apply their generalized model and to observe in how far it improves the results obtained for the ordinary Fucks GPD.

In a corresponding re-analysis, the value μ will be estimated by the sample mean, and the weight coefficients ε_k will be determined by the first three initial moments of the empirical distribution. The generating function corresponding to distribution (40) is given as:

$$G(t) = \frac{1}{2} \cdot \frac{e^{2\lambda(t-1)} - 1}{\lambda(t-1)} \cdot \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) t^k. \qquad (44)$$

Using the fact that

$$\mu_{(k)} = \left. \frac{\partial^k G(t)}{\partial t^k} \right|_{t=1},$$

it is easy to get the factorial moments of the distribution (40). The moments, derived from (44), are functions of μ and $\{\varepsilon_k\}$. Now, similarly as in case of the Fucks GPD, we will first determine the factorial moments, and then the initial moments of the Fucks-Gačečiladze distribution. The set of relations (9) implies the following results. The first derivative of $G(t)$ for $t = 1$ provides the first factorial moment

$$\mu_{(1)} = \sum_{k=0}^{\infty} (k+\lambda)(\varepsilon_k - \varepsilon_{k+1}) = A + \lambda = \mu.$$

The second derivative of $G(t)$ for $t = 1$ provides the second factorial moment:

$$\mu_{(2)} = \frac{1}{3} \sum_{k=0}^{\infty} (6k\lambda + 4\lambda^2 + 3k^2 + 3k) \cdot (\varepsilon_k - \varepsilon_{k+1}) = (2\lambda - 2)A + \frac{4}{3}\lambda^2 + 2\sum_{k=0}^{\infty} k\varepsilon_k$$

or, substituting $\lambda = \mu - A$,

$$\mu_{(2)} = \frac{4}{3}\mu^2 - \frac{2}{3}\mu A - \frac{2}{3}A^2 - 2A + 2\sum_{k=1}^{\infty} k\varepsilon_k.$$

Likewise, the third derivative of $G(t)$ for $t = 1$ provides the third factorial moment:

$$\mu_{(3)} = \sum_{k=0}^{\infty} (-3k\lambda - 3k^2 + 3k^2\lambda + 4\lambda^2 k + 2\lambda^3 + 2k + k^3)(\varepsilon_k - \varepsilon_{k+1})$$

or, substituting, $\lambda = \mu - A$,

$$\mu_{(3)} = 2\mu^3 - 2\mu^2 A - 2\mu A^2 - 6\mu A + 2A^3 + 6A^2 + 5A + (6\mu - 6A - 6)\sum_{k=0}^{\infty} k\varepsilon_k + \sum_{k=0}^{\infty} k^3 (\varepsilon_k - \varepsilon_{k+1}).$$

The only difference as compared to Fucks's approach thus is the manner how the parameters ε_k are estimated. As opposed to Fucks, the Georgian authors suggest to estimate ε_k not with recourse to the central moments, but to the initial moments of the empirical distribution. Obviously, central moments and initial moments can be transformed into each other, i. e., both methods lead to identical parameter estimates. Still, the numerical procedure of estimating is different and shall be demonstrated in detail, here.

Gačečiladze/Cilosani (1971, 115) discussed two possibilities to estimate ε_k:

(a) by deriving the theoretical initial moments from the generating function of the Fucks-Gačečiladze distribution (44);
(b) by approximately equating the frequencies of the empirical and the theoretical distributions, using the fact that $\sum_{i=0}^{n} p_i \approx 1$.

Arguing that the second way is more convenient to be pursued, since the first includes a system of transcendental equations, Gačečiladze/Cilosani (1971, 116) favored the second option. Since, in our case, we consider only special cases of the generalized Fucks-Gačečiladze distribution, the system is reduced to less complex systems which today can easily solved by help of computer programs.

Gačečiladze/Cilosani (1971) did not show how the theoretical initial moments can be derived from the generating function. Therefore, it seems reasonable to recapitulate this step, following the same line of thinking already presented in context of the Fucks GPD (see above, Sec. 2.1.1). As a result, the initial moments of the Fucks-Gačečiladze distribution (40), are given as follows, with $\sum_{k=1}^{\infty} \varepsilon_k = A$,

(a) The first initial moment:

$$\mu'_1 = \mu_{(1)} = \left. \frac{\partial G(t)}{\partial t} \right|_{t=1} = \mu, \qquad (45)$$

(b) The second initial moment

$$\mu'_2 = \mu_{(1)} + \mu_{(2)} = \frac{\partial^2 G(t)}{\partial t^2}\Big|_{t=1} + \frac{\partial G(t)}{\partial t}\Big|_{t=1} = \frac{4}{3}\mu^2 - \frac{2}{3}\mu A + \mu - \frac{2}{3}A^2 - 2A + 2\sum_{k=1}^{\infty} k\varepsilon_k, \quad (46)$$

(c) The third initial moment

$$\mu'_3 = \mu_{(3)} + 3\mu_{(2)} + \mu_{(1)} = \frac{\partial^3 G(t)}{\partial t^3}\Big|_{t=1} + 3\frac{\partial^2 G(t)}{\partial t^2}\Big|_{t=1} + \frac{\partial G(t)}{\partial t}\Big|_{t=1}$$
$$= 2\mu^3 - 2\mu^2 A + 4\mu^2 - 2\mu A^2 + \mu - 8\mu A + 2A^3 + 4A^2 - A + (6\mu - 6A)\sum_{k=1}^{\infty} k\varepsilon_k + \sum_{k=1}^{\infty} k^3(\varepsilon_k - \varepsilon_{k+1}). \quad (47)$$

The empirical initial moments are defined as

$$m'_r = \frac{1}{N}\sum_i i^r \cdot f_i.$$

The initial moments are necessary for the establishment of the equation system, which, in turn, is needed for the estimation of the parameters ε_k of the distribution (40). Thus, using the relations between factorial and central moments (6), the second and third central moments of the Fucks-Gačečiladze distribution are given as

$$\mu_2 = \frac{1}{3}\mu^2 - \frac{2}{3}\mu A - \frac{2}{3}A^2 - 2A + 2\sum_{k=1}^{\infty} k\varepsilon_k + \mu$$

$$\mu_3 = \mu^2 + \mu - 2\mu A + 2A^3 + 4A^2 - A - 6A\sum_{k=1}^{\infty} k\varepsilon_k + \sum_{k=1}^{\infty} k^3(\varepsilon_k - \varepsilon_{k+1}). \quad (48)$$

Given these definitions, we can now, in the next step, direct our attention on the three-parameter special case of the Fucks-Gačečiladze distribution.

3.1. A three-parameter special case of the Fucks-Gačečiladze Distribution

In case of the Fucks-Gačečiladze three-parameter model, the three parameter μ, ε_2 and ε_3 have to be estimated. The estimation depends on the fact whether a class of 0-syllable words has to be taken into consideration, or not.

Let $\varepsilon_0 = 1$, $\varepsilon_1 \neq 0$, $\varepsilon_2 \neq 0$ and $\varepsilon_k = 0$ for $k \geq 3$. Replacing ε_1 with α, and ε_2 with β, and denoting $a = \alpha + \beta$, it is possible to get the following 2×2 equations system, from equations (46) and (47):

$$\text{(a)}\ \mu'_2 = \frac{4}{3}\mu^2 - \frac{2}{3}\mu a + \mu - \frac{2}{3}a^2 + 2\beta$$

$$\text{(b)}\ \mu'_3 = 2\mu^3 - 2\mu^2 a + 4\mu^2 - 2\mu a^2 + \mu - 8\mu a + 2a^3 + 4a^2 + (6\mu - 6a)(a + \beta) + 6\beta.$$

Subsequent to the solution for α and β, we thus have the following distribution:

$$p_0 = e^{-\lambda}(1-\alpha)\,\varphi_0(0)$$
$$p_1 = e^{-\lambda}[(1-\alpha)\,\lambda\,\varphi_0(1) + (\alpha-\beta)\,\varphi_1(1)]$$
$$p_i = e^{-\lambda}\Big[(1-\alpha)\frac{\lambda^i}{i!}\varphi_0(i) + (\alpha-\beta)\frac{\lambda^{i-1}}{(i-1)!}\varphi_1(i) + \beta\frac{\lambda^{i-2}!}{(i-2)!}\varphi_2(i)\Big]. \quad i \geq 2 \quad (49)$$

Now, setting $\varepsilon_0 = \varepsilon_1 = 1$, $\varepsilon_k = 0$ for $k \geq 4$ and $\varepsilon_2 \neq 0$, $\varepsilon_3 \neq 0$, results in the 1-displaced three-parameter special case of the Fucks-Gačečiladze distribution. Furthermore replacing ε_2 with α, and ε_3 with β, and denoting $a = 1 + \alpha + \beta$, we obtain the following system of equations:

$$\text{(a)}\quad \mu'_2 = \frac{4}{3}\mu^2 - \frac{2}{3}\mu a + \mu - \frac{2}{3}a^2 + 2(\alpha + 2\beta)$$

$$\text{(b)}\quad \mu'_3 = 2\mu^3 - 2\mu^2 a + 4\mu^2 - 2\mu a^2 + \mu - 8\mu a + 2a^3 + 4a^2 + (6\mu - 6a)\cdot(a + \alpha + 2\beta) + 6\alpha + 18\beta.$$

After the solution for α and β, we thus have the following probabilities:

$$p_1 = e^{-\lambda}(1-\alpha)\,\varphi_1(1)$$
$$p_2 = e^{-\lambda}[(1-\alpha)\,\lambda\varphi_1(2) + (\alpha-\beta)\,\varphi_2(2)]$$
$$p_i = e^{-\lambda}\Big[(1-\alpha)\frac{\lambda^{i-1}}{(i-1)!}\varphi_1(i) + (\alpha-\beta)\frac{\lambda^{i-2}}{(i-2)!}\varphi_2(i) + \beta\frac{\lambda^{i-3}}{(i-3)!}\varphi_3(i)\Big],\ i \geq 3$$
$$\text{with } \lambda = \mu - 1 - \alpha - \beta. \quad (50)$$

Table 11a.7: Discrepancy coefficient C as a result of fitting 1-displaced Fucks-Gačečiladze (three-parameter) distribution to different languages

	English	German	Esperanto	Arabic	Greek
C (3-par.)	∅	∅	0.0007	0.0088	0.0014
$\hat{\varepsilon}_2$	–	–	0.4490	0.7251	−0.0731
$\hat{\varepsilon}_3$	–	–	0.1261	0.1986	−0.1508
	Japanese	Russian	Latin	Turkish	
C (3-par.)	∅	0.0028	0.0035	0.0087	
$\hat{\varepsilon}_2$	–	0.3821	0.6230	0.6870	
$\hat{\varepsilon}_3$	–	0.1885	0.3050	0.2606	

It is interesting to see, now, which results are obtained with regard to the Fucks data repeatedly analyzed above. Table 11a.7 represents the values of the discrepancy coefficient C as a result of the relevant re-analysis. As can be seen from Table 11a.7, an acceptable result is indeed obtained for Greek, too, what was not the case when fitting the three-parameter Fucks distribution (cf. Table 11a.5). In fact, The Fucks-Gačečiladze distribution provides very good fits in six of the nine samples ($C \leq 0.01$). Table 11a.7 also shows that the values for $\hat{\varepsilon}_2$ and $\hat{\varepsilon}_3$ are negative in case of the Greek sample. Again, this is due to the fact that we confine ourselves to Fucks's conditions (a)–(c) outlined above (cf. Sec. 2). The introduction of the additional condition, $0 < \varepsilon_k < 1$, $k = 2, 3$, results in another solution which is slightly worse, with $\hat{\varepsilon}_2 = 0.3013$, $\hat{\varepsilon}_3 = 0.1511$, and $C = 0.0144$. Still, there are no solutions for the English, German and Japanese data. The reason for this failure might be the fact that, for $\varphi_k \to 1$, the Fucks-Gačečiladze distribution (40) converges to the Fucks GPD, and under this condition provides identical results.

As opposed to this, the fact that the Georgian authors base their estimations on the initial moments, rather than on the central moments, plays no role, because central moments and initial moments can be transformed into each other. Thus, the results can be expected to be identical in either case. This will be shown for our data.

The first three initial moments of Fucks distribution, which are necessary for the equation system to be established, are given as:

$$\mu'_1 = \mu$$

$$\mu'_2 = \mu^2 + \mu - \varepsilon'^2 - 2\varepsilon' + 2\sum_{k=1}^{\infty} k\varepsilon_k$$

$$\mu'_3 = \mu^3 + 3\mu^2 + \mu + 2\varepsilon'^3 + 3\varepsilon'^2 - \varepsilon' - - 3\mu\varepsilon'^2 - 6\mu\varepsilon' + \sum_{k=0}^{\infty} k^3 (\varepsilon_k - \varepsilon_{k+1}) + + 6(\mu - \varepsilon')\sum_{k=1}^{\infty} k\varepsilon_k. \tag{51}$$

Now, replacing ε_2 with α, and ε_3 with β, and denoting $a = 1 + \alpha + \beta$, we obtain the following system of equations:

(a) $\mu'_2 = \mu^2 + \mu - a^2 - 2a + 2(a + \alpha + 2\beta)$

(b) $\mu'_3 = \mu^3 + 3\mu^2 + \mu + 2a^3 + + 3(1-\mu)a^2 + 6\alpha + 18\beta - - 6\mu a + 6(\mu - a)(a + \alpha + 2\beta)$.

After the solution for α and β, we thus have the following probabilities:

$$p_1 = e^{-\lambda} \cdot (1 - \alpha)$$

$$p_2 = e^{-\lambda} \cdot [(1-\alpha) \cdot \lambda + (\alpha - \beta)]$$

$$p_i = e^{-\lambda}\left[(1-\alpha)\frac{\lambda^{i-1}}{(i-1)!} + + (\alpha - \beta)\frac{\lambda^{i-2}}{(i-2)!} + \beta\frac{\lambda^{i-3}}{(i-3)!}\right], \quad i \geq 3 \tag{52}$$

with $\lambda = \mu - 1 - \alpha - \beta$.

Table 11a.8 contains the results with parameter estimations based both on central and initial moments. As can easily be seen, the results are almost identical, as was to be expected.

In summary, one can thus state that neither the Fucks-Gačečiladze distribution nor the Fucks GPD, as one of its special cases (for $\varphi_k \to 1$), turn out to be adequate as a general standard model, capable to cover all nine data sets presented by Fucks.

Table 11a.8: Fucks' three-parameter model, with parameter estimation, based on moments

	Esperanto	Arabic	Russian	Latin	Turkish
m_2, m_3					
$\hat{\varepsilon}_2$	0.3933	0.5463	0.2083	0.5728	0.6164
$\hat{\varepsilon}_3$	0.0995	−0.1402	0.1686	0.2416	0.1452
C	0.00004	0.0021	0.0005	0.0003	0.0023
m'_2, m'_3					
$\hat{\varepsilon}_2$	0.3933	0.5464	0.2083	0.5728	0.6164
$\hat{\varepsilon}_3$	0.0994	−0.1402	0.1685	0.2415	0.1451
C	0.00004	0.0021	0.0005	0.0004	0.0023

The Fucks-Gačečiladze distribution seems to be a better model only for some special data compared to three-parameter Fucks distribution, since it provides good fits for Greek, too (cf. Table 11a.7). As to the estimation of the parameters, there are no differences as to the method of estimation (be it based on the initial or central moments).

4. The Fucks GPD: parameter estimation based on μ, μ_2, and first-class frequency (Bartkowiakowa/Gleichgewicht)

An alternative to estimate the two parameters ε_2 and ε_3 of the Fucks' three-parameter distribution was suggested by two Polish authors (Bartkowiakowa/Gleichgewicht 1964; 1965). Based on the Poisson distribution, given as

$$g_k = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, 2, \dots \qquad (53)$$

and referring to Fucks' GPD (2), the authors re-formulated the latter as

$$p_i = \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1})\, e^{-\lambda} \frac{\lambda^{i-k}}{(i-k)!} = \sum_{k=0}^{\infty} (\varepsilon_k - \varepsilon_{k+1}) \cdot g_{i-k}. \qquad (54)$$

Determining $\varepsilon_0 = \varepsilon_1 = 1$, and $\varepsilon_k = 0$ for $k > 3$, the two parameters $\varepsilon_2 = \varepsilon_3 \neq 0$ remain to be estimated on basis of the empirical distribution. Based on these assumptions, the following special cases are obtained for (54):

$$p_1 = (1 - \varepsilon_2) \cdot g_0$$

$$p_2 = (1 - \varepsilon_2) \cdot g_1 + (\varepsilon_2 - \varepsilon_3) \cdot g_0$$

$$p_i = (1 - \varepsilon_2) \cdot g_{i-1} + (\varepsilon_2 - \varepsilon_3) \cdot g_{i-2} + \varepsilon_3 \cdot g_{i-3},\ i \geq 3 \qquad (55)$$

with $\lambda = \mu - (1 + \varepsilon_2 + \varepsilon_3)$.

As to the estimation of ε_2 and ε_3, the authors did not set up an equation system on the basis of the second and third central moments (μ_2 and μ_3), as did Fucks, thus arriving at a cubic equation. Rather, they first defined the portion of one-syllable words (p_1), and then modelled the whole distribution on that proportion. Thus, by way of a logarithmic transformation of $p_1 = (1 - \varepsilon_2) \cdot g_0$ in formula (55), one obtains the following sequence of transformations:

$$\ln \frac{p_1}{(1 - \varepsilon_2)} = \ln g_0$$

$$\ln \frac{p_1}{(1 - \varepsilon_2)} = -\lambda$$

$$\ln \frac{p_1}{(1 - \varepsilon_2)} = -[\mu - (1 + \varepsilon_2 + \varepsilon_3)].$$

Referring to the empirical distribution, a first equation for an equation system to be solved (see below) can thus be gained from the first probability p_1 of the empirical distribution:

$$\ln \frac{\hat{p}_1}{(1 - \hat{\varepsilon}_2)} = -[\bar{x} - (1 + \hat{\varepsilon}_2 + \hat{\varepsilon}_3)]. \qquad (56)$$

The second equation for this system is then gained from the variance of the empirical distribution. Thus one gets

$$\mu_2 = \mu - (1 + \varepsilon_2 + \varepsilon_3)^2 + 2 \cdot (\varepsilon_2 + 2 \cdot \varepsilon_3)$$

resulting in the second equation for the equation system to be established:

$$m_2 = \bar{x} - (1 + \hat{\varepsilon}_2 + \hat{\varepsilon}_3)^2 + 2 \cdot (\hat{\varepsilon}_2 + 2\hat{\varepsilon}_3). \qquad (57)$$

With the two equations (56) and (57), we thus have the following system of equations, adequate to arrive at a solution for ε_2 and ε_3:

$$\text{(a)} \quad \ln \frac{\hat{p}_1}{(1-\hat{\varepsilon}_2)} = -[\bar{x} - (1 + \hat{\varepsilon}_2 + \hat{\varepsilon}_3)]$$

$$\text{(b)} \quad m_2 - \bar{x} = -(1 + \hat{\varepsilon}_2 + \hat{\varepsilon}_3)^2 + + 2(\hat{\varepsilon}_2 + 2\hat{\varepsilon}_3)$$

Bartkowiakowa/Gleichgewicht (1964) did not theoretically discuss the corresponding estimation procedure of ε_2 and ε_3 in detail. Rather, they preferred to present the results of empirical studies, based on selected Polish literary texts, which served as a test of their approach.

In addition to the difference as to the estimation of ε_2 and ε_3, the two Polish authors argued in favor of a statistical test in order to evaluate the goodness of fit of the theoretical distribution, comparing it with the empirical distribution on the basis of the χ^2-test; in doing so, they suggest to repeat the process of estimation as long as a minimal value for the χ^2 function is obtained (the so-called χ^2 minimization method). Of course, this method is much more expensive as compared to the other approach, based on the relative frequency $\hat{p}_1$ and the two moments $\bar{x}$ and m_2.

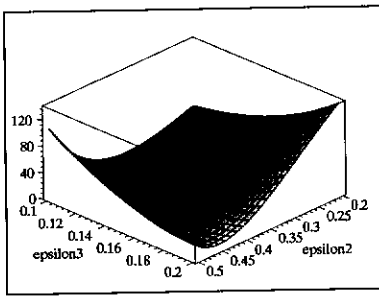
The analyses included nine Polish literary texts, or specific segments thereof (e. g., only selected passages, only the dialogical sequences of a given text, etc.). The results of the analysis indeed proved their approach to be successful. As can be seen from Table 11a.9, the discrepancy coefficient is $C < 0.01$ in all cases; furthermore, in six of the nine samples, the result is clearly better as compared to Fucks' original estimation.

For the sake of comparison, Table 11a.9 contains both the dispersion quotient d and the difference $M = \bar{x} - m_2$ between the mean of the empirical distribution and its variance for each of the samples, in addition to the results for the (1-displaced) Poisson and the (1-displaced) Dacey-Poisson distributions, which were calculated in a re-analysis of the raw data provided by the Polish authors. A closer look at these data shows that the Polish text samples are relative homogeneous: for all texts, the dispersion quotient is in the interval $0.88 \le d \le 1.04$, and $0.95 \le M \le 1.09$. This may explain, why the theoretical model turns out to be adequate for all samples. Therefore, the Polish authors' approach will also be tested with Fucks' linguistic data repeatedly analyzed above.

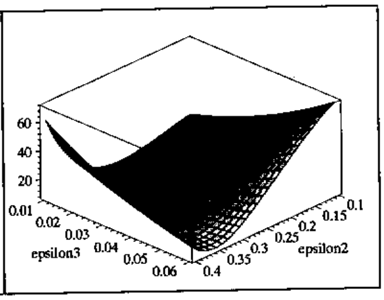
Before this additional re-analysis, it seems worthwhile testing the performance of the χ^2 minimization method suggested by Bartkowiakowa/Gleichgewicht (1964). The question is if there is an additional improve-

Table 11a.9: Fucks' three-parameter model, with parameter estimation (Polish data)

	1	2	3	4	5
$\bar{x}$	1.81	1.82	1.96	1.93	2.07
m_2	0.76	0.73	0.87	0.94	1.07
d	0.93	0.88	0.91	1.00	0.99
M	1.05	1.09	1.09	0.99	1.00
C values:					
Poisson	0.00420	0.00540	0.00370	0.00170	0.00520
Dacey-Poisson	0.00250	0.00060	0.00200	∅	0.00531
m_2, m_3	0.00240	0.00017	0.00226	0.00125	0.00085
$\hat{p}_1, m_{22}$	0.00197	0.00043	0.00260	0.00194	0.00032
	6	7	8	9	
$\bar{x}$	2.12	2.05	2.18	2.16	
m_2	1.10	0.98	1.21	1.21	
d	0.98	0.94	1.03	1.04	
M	1.02	1.07	0.97	0.95	
C values:					
Poisson	0.00810	0.00220	0.01360	0.00940	
Dacey-Poisson	0.00862	0.00145	∅	∅	
m_2, m_3	0.00084	0.00120	0.00344	0.00383	
$\hat{p}_1, m_2$	0.00030	0.00077	0.00216	0.00271	



(a) "Anielka (without dialogues)"



(b) " Żywot"

Fig. 11a.1: χ^2 function for two Polish texts

ment of the results trying to minimize χ^2 subsequent to the described estimation of the parameters ε_2 and ε_3.

By way of an example, let us test the productivity of this additional procedure, analyzing two Polish texts by B. Prus and M. Rej, taking the relevant data from Bartkowiakowa/Gleichgewicht's article.

Estimating the parameters ε_2 and ε_3 for "Anielka" (without dialogues) by B. Prus, and minimizing the χ^2 function, the Polish authors obtained the following estimates: $\hat{\varepsilon}_2 = 0.390$, $\hat{\varepsilon}_3 = 0.145$ and $\chi^2 = 5.041$. It seems that we are concerned here with an error: taking the estimated values for ε_2 and ε_3 suggested by Polish authors, and calculating once again the expected theoretical distribution, it turns out that there is a mistake in the Polish calculation, since for these ε values, a much better value of $\chi^2 = 3.534$ is obtained.

Our further re-analysis of the Polish data, which is based on more exact computer methods as compared to the methods available at the times of the Polish study, includes two steps:

(a) In a first step, we look for a numerical solution of the equation system (56) and (57), thus obtaining estimators for ε_2 and ε_3.
(b) In a second step, we plot the χ^2 function, since we are interested to know whether the estimators obtained in the first step differ from the minimum values of the χ^2 function, or not.

As to the numerical solution, solving the system of equations (56) and (57), we obtain a value of $\chi^2 = 1.5696$ with $\hat{\varepsilon}_2 = 0.4378$ and $\hat{\varepsilon}_3 = 0.1784$. As minimal value of χ^2 one obtains $\chi^2 \approx 1$. Figure 1(a) exhibits the plot of the χ^2 function for the above-mentioned text.

A closer look at it shows that the values obtained for ε_2 and ε_3 by way of the numerical re-analysis, are near to the values where the minimum of χ^2 is attained.

The same observation can be made with regard to M. Rej's text "Żywot". As a result of our re-analysis, a value of $\chi^2 = 7.880$ is obtained (with $\hat{\varepsilon}_2 = 0.3564$ and $\hat{\varepsilon}_3 = 0.0599$). As compared to this, the Polish authors arrived at $\chi^2 = 7.934$ (for $\hat{\varepsilon}_2 = 0.345$ and $\hat{\varepsilon}_3 = 0.054$). Our re-analysis shows that the minimum is smaller than $\chi^2 = 7.88$. Thus, again the ordinary estimation method yields good results. Figure 1(b) illustrates the relevant plot of the χ^2 function.

Summarizingly, one can say that, at least as far as these two exemplary texts are concerned, the best results of estimation are already reached by solving equation system (56) and (57), without any remarkable improvement by way of the χ^2 minimization.

With this in mind, it will be interesting to see in how far the estimation procedure suggested by Bartkowiakowa/Gleichgewicht (1964) is able to improve the results for the nine different languages analyzed by Fucks (cf. Table 11a.1). Table 11a.10 represents the results of the corresponding re-analysis.

Table 11a.10 compares the results obtained by two different ways of parameter estimation: the original procedure as suggested by Fucks and the modification suggested by the Polish authors. The comparison is done only for data sets, appropriate for the application of Fucks' three-parameter distribution model. As a result, we can conclude that the procedure to estimate the two parameters ε_2 and ε_3, as suggested by

Table 11a.10: Fucks' three-parameter model, with parameter estimation

	Esperanto	Arabic	Russian	Latin	Turkish
m_2, m_3					
$\hat{\varepsilon}_2$	0.3933	0.5463	0.2083	0.5728	0.6164
$\hat{\varepsilon}_3$	0.0995	−0.1402	0.1686	0.2416	0.1452
C	0.00004	0.0021	0.0005	0.0003	0.0023
$\hat{p}_1, m_2$					
$\hat{\varepsilon}_2$	0.3893	0.7148	0.2098	0.5744	0.6034
$\hat{\varepsilon}_3$	0.0957	0.1599	0.1695	0.2490	0.1090
C	0.00001	0.0042	0.0005	0.0003	0.0018

Bartkowiakowa/Gleichgewicht (1964), results in better estimates, at least for the data analyzed.

A possible interpretation is the better aptness of the modified approach which might have its foundation in the fact that this estimation procedure is particularly adequate when the frequency in the first class is relatively large; thus, more weight is given to this large frequency value of the first class, as compared to the the third moment which is more affected by the frequencies of the higher classes.

5. Summary

In the history of word length studies, an important step was made by German physicist Wilhelm Fucks: his theoretical model, often termed as "the Fucks model", turned out to be the most important model discussed from the 1950s until the late 1970s. From that time on until today, Fucks is broadly credited for having promoted the 1-displaced Poisson distribution to be an adequate model for word length frequencies, at least for all those languages, which form their words from syllables.

At a closer mathematical inspection of Fucks' works, it turns out, however, that the 1-displaced Poisson distribution is only one special case of a broader generalization of the Poisson distribution suggested by Fucks; this model, a specific sum of weighted Poisson probabilities, is termed the Fucks Generalized Poisson Distribution (GPD) throughout this paper.

In addition to presenting this model, in detail, different methods to estimate its parameters are discussed, which have been suggested not only by Fucks himself, but by researchers following him. Also an even more far-reaching generalization of the Fucks GPD, as developed by Georgian scholars in the early 1960s, is discussed, on the background of which Fucks' GPD, in turn, appears to be a special case of one further generalization, namely, the so called Fucks-Gačečiladze distribution.

By way of a number of re-analyses, not only the Fucks GPD, but also some of its special cases, as well as the above-mentioned Georgian generalization are fitted to data from various languages, presented by Fucks himself (cf. Fucks 1956a). As a result of these analyses, a number of conclusions can be drawn. Ultimately, these conclusions are related to the summarizing Tables 11a.11 and 11a.12: whereas Table 11a.11 contains four relevant characteristics of the nine language samples, Table 11a.12 summarizes the results of fitting the discussed distributions.

Table 11a.11: Mean word length, variance, d value and difference M for nine languages

	English	German	Esperanto	Arabic	Greek
$\bar{x}$	1.4064	1.6333	1.8971	2.1032	2.1106
m_2	0.5645	0.7442	0.8532	0.6579	1.3526
d	1.3890	1.1751	0.9511	0.5964	1.2179
M	0.8420	0.8891	1.0438	1.4453	0.7580

	Japanese	Russian	Latin	Turkish
$\bar{x}$	2.1325	2.2268	2.3894	2.4588
m_2	1.3952	1.4220	1.2093	1.1692
d	1.2319	1.1591	0.8704	0.8015
M	0.7374	0.8048	1.1800	1.2896

1. 1. The 1-displaced Poisson distribution, as one-parameter special case of the Fucks GPD, cannot be accepted as a general standard model for word length frequency distributions: specifically, it can be an adequate model only, as long as the dispersion quotient $d = m_2/(\bar{x}-1) \approx 1$ in an empirical sample.

2. The first special case of the Fucks GPD, the two-parameter (1-displaced) Dacey-Poisson distribution, is an adequate theoretical model only for a specific type of empirical distributions, too: this model is likely to be an adequate theoretical model for empirical samples with $d < 1$. Therefore, of the nine languages tested, only four (Esperanto, Arabic, Latin, and Turkish) meeting this condition can be successfully modelled (cf. Table 11a.12).
3. The three-parameter special case of Fucks' GPD provides clearly better results as compared to those of the one-parameter and two-parameter special cases; and can be an appropriate model also for empirical distributions in which $d > 1$. However, valid estimators exist only if both $M = \bar{x} - m_2 \leq 0.75$ and parameter $a = \hat{\varepsilon}_2 + \hat{\varepsilon}_3$ fulfill some additional conditions. Thus, the first restriction rules out the Japanese sample, and the second restriction is violated for the English, German and Greek data. A very good fit is obtained for those four texts with $d < 1$, and for the Russian sample with $d \approx 1.16$.
   The three-parameter Fucks model thus too, is adequate only for particular types of empirical distributions, and it cannot serve as an overall model for language, even if restricted to syllabic languages.
4. The generalization of the three-parameter GPD by Gačečiladze et al. additionally delivers valid estimations and appropriate models for the Greek ($d = 1.1218$) sample, but it too, is not able to fit the English, German and Japanese data.
5. The three-parameter GPD, combined with the alternative estimation method based on $\bar{x}$, m_2 and the first-class frequency $\hat{p}_1$ provides a slightly better fit than the method using the first three moments $\bar{x}$, m_2 and m_3 and is also comparable with the much more expensive χ^2 minimization method. The results obtained for the Fucks-Gačečiladze distribution were slightly better as compared to those for the three-parameter Fucks GPD.

Figure 11a.2 illustrates the results of fitting of the discussed estimation methods for the two- and three-parameter modifications.

Generally speaking, from a contemporary point of view, there are a number of theoretical and practical drawbacks of the Fucks GPD, including its modifications and generalizations:

(a) There is no *a priori* information how many components of the ε-spectrum are necessary for a linguistic application of the Fucks GPD; furthermore, all available suggestions as to a linguistic interpretation are nothing but heterogeneous ad hoc assumptions;
(b) the support of a word length frequency distribution cannot be infinite, notwithstanding the fact that this circumstance tends to be ignored in practical applications;

Therefore, our present research on word length frequencies focuses not only on an empirical test of the Fucks GPD, including its modifications and generalizations discussed in this chapter (cf. Grzybek/Stadlober 2005a), but also on further generalizations and modifications of the Poisson distribution (cf. Grzybek/Stadlober 2005b). One of these lines is along the three-parameter GPD distribution, hereby dropping the unnecessary condition $\varepsilon_{k+1} \leq \varepsilon_k$ as stated by

Table 11a.12: Comparison of fitting nine languages, based on discrepancy coefficient C (acceptable best fits for for each language are in bold face)

	Discrepancy coefficient values (C) for				
	Poisson	*Dacey-Poisson*	(m_2, m_3)	$(\hat{p}_1, m_2)$	*Fucks-Gačečiladze*
English	0.0903	∅	∅	∅	∅
German	**0.0186**	∅	∅	∅	∅
Esperanto	0.0023	0.0019	**0.0000**	**0.0000**	0.0007
Arabic	0.1071	0.0077	0.0021	**0.0042**	0.0088
Greek	0.0328	∅	∅	∅	**0.0014**
Japanese	0.0380	∅	∅	∅	∅
Russian	0.0208	∅	**0.0005**	**0.0005**	0.0028
Latin	0.0181	0.0149	**0.0003**	**0.0003**	0.0035
Turkish	0.0231	0.0021	0.0023	**0.0018**	0.0087

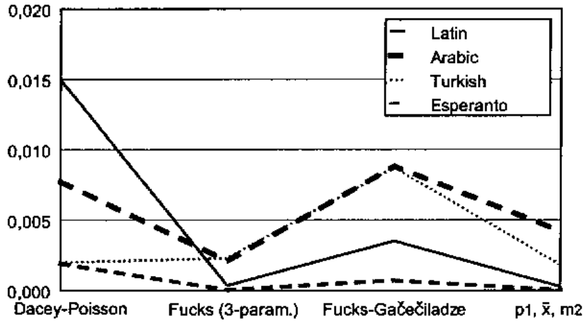


Fig. 11a.2: Comparison of fitting results (C) for the two- and three-parameter models

Fucks. Another line of research in this context concentrates on the generalization of the Poisson distribution as discussed by Consul (1989), and on the Hyperpoisson distribution (cf. Wimmer/Altmann 1999: 281 f.).

In any case, we concentrate on concrete individual texts from various languages, rather than on text mixtures (corpora) allegedly describing the abstract norm of a given language. First results show that these generalizations may lead to valid estimations and appropriate fits of a large class of word length distributions with empirical dispersion quotients $0.5 \leq d \leq 1.5$.

6. Literature (a selection)

Bartkowiakowa, Anna/Gleichgewicht, Boleslaw (1964), Zastosowanie dwuparametrowych rozkładów Fucksa do opisu długości sylabicznej wyrazów w różnych utworach prozaicznych autorów polskich. In: *Zastosowania matematyki* 7, 345–352.

Bartkowiakowa, A./Gleichgewicht, B. (1965), O rozkładach długości sylabicznej wyrazów w różnych tekstach. In: *Poetyka i matematyka.* (ed M. R. Mayenowa). Warszwawa: Państwowy instytut wydawniczy, 164–173.

Bokučava, N. V./Gačečiladze, T. G. (1965), Ob odnom metode izučenija statističeskoj struktury pečatnoj informacii. In: *Trudy Tbilisskogo gosudarstvennogo universiteta*, t. 103, 174–180.

Bokučava, N. V./Gačečiladze, T. G./Nikoladze, K. Ja.; Cilosani, T. P. (1965), Zamečanie k matematičeskoj modeli dlja slogoobrazovanija v gruzinskom jazyke. In: *Trudy Tbilisskogo gosudarstvennogo universiteta*, t. 103, 169–172.

Cercvadze, G. N./Čikoidze, G. B./Gačečiladze, T. G. (1959), Primenenie matematičeskoj teorii slovoobrazovanija k gruzinskomu jazyku. In: *Soobšcenija akademii nauk Gruzinskoj SSR, t. 22/6,* 705–710.

Cercvadze, G. N./Čikoidze, G. B./Gačečiladze, T. G. (1962), see: Zerzwadse et al. (1962)

Consul, Prem C. (1989), *Generalized Poisson Distributions.* New York/Basel: Dekker.

Fucks, Wilhelm (1955a), Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen. *Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen.* Köln/Opladen: Westdeutscher Verlag.

Fucks, Wilhelm (1955b), Theorie der Wortbildung. In: *Mathematisch-Physikalische Semesterberichte zur Pflege des Zusammenhangs von Schule und Universität* 4, 195–212.

Fucks, Wilhelm (1955c), Eine statistische Verteilung mit Vorbelegung. Anwendung auf mathematische Sprachanalyse. In: *Die Naturwissenschaften* 42_1, 10.

Fucks, Wilhelm (1956a), Die mathematischen Gesetze der Bildung von Sprachelementen aus ihren Bestandteilen. *Nachrichtentechnische Fachberichte, [= Beiheft zu Nachrichtentechnische Fachzeitschrift]* 3, 7–21.

Fucks, Wilhelm (1956b), Mathematische Analyse von Werken der Sprache und der Musik. In: *Physikalische Blätter* 16, 452–459 & 545.

Fucks, Wilhelm (1956c), Statistische Verteilungen mit gebundenen Anteilen. In: *Zeitschrift für Physik* 145, 520–533.

Fucks, Wilhelm (1956d), Mathematical theory of word formation. In: *Information theory.* (ed. C. Cherry). London: Butterworth, 154–170.

Fucks, Wilhelm (1957), Matematičeskaja teorija slovoobrazovanija. In: *Teorija peredači soobščenij.* Moskva: Izdatel'stvo inostrannoj literatury, 221–247.

Gačečiladze, T. G./Cercvadze, G. N./Čikoidze, G. B. (1961), Ob ε-strukture raspredelenija probelov. In: *Trudy instituta élektroniki, avtomatiki i telemechaniki* 2, 3–15.

Gačečiladze, T. G./Cilosani, T. P. (1971), Ob odnom metode izučenija statističeskoj struktury teksta. In: *Statistika reči i avtomatičeskij analiz teksta.* Leningrad, Nauka: 113–133.

Grotjahn, Rüdiger (1982), Ein statistisches Modell für die Verteilung der Wortlänge. In: *Zeitschrift für Sprachwissenschaft* 1, 44–75.

Grzybek, Peter (2005), History and Methodology of Word Length Studies – The State of the Art. In: *Contributions to the Science of Language* (ed. P. Grzybek). [in print]

Grzybek, Peter/Stadlober, Ernst (2005a), The Performance of Fucks' Generalized Poisson Distribution in the Study of Word Length Frequencies. [in prep.]

Grzybek, Peter/Stadlober, Ernst (2005b), The Performance of Generalized Poisson Models in Word Length Frequency Studies. [in prep.]

Piotrovskij, Rajmond G./Bektaev, Kaldybay B./Piotrovskaja, Anna A. (1977), *Matematičeskaja lingvistika.* Leningrad: Nauka. [German trans.: Piotrowski, R. G.; Bektaev, K. B.; Piotrowskaja, A. A.: *Mathematische Linguistik.* Bochum: Brockmeyer, 1985. [= Quantitative Linguistics; 27]

Wimmer, Gejza/Altmann, Gabriel (1999), *Thesaurus of univariate discrete probability distributions.* Essen: Stamm.

Zerzwadse, G./Tschikoidse, G./Gatschetschiladse, Th. (1962), Die Anwendung der mathematischen Theorie der Wortbildung auf die georgische Sprache. In: *Grundlagenstudien aus Kybernetik und Geisteswissenschaft* 4, 110–118.

Gordana Antić, Peter Grzybek, Ernst Stadlober, Graz (Austria)