The phoneme-grapheme relationship in Slovene

Emmerich Kelih

1 Introduction

In this contribution we discuss some problems concerning the synergetic control cycle of script theory proposed by Altmann (this volume, pp. 149ff.). The following aspects will be treated on the basis of Slovene: (a) the orthographic uncertainty of phonemes and a statistical comparison with other languages, (b) the distribution of grapheme size, (c) the mean grapheme length, (d) the mean graphemic load, and (e) letter utility and its positional weighting. The procedure adheres to the order proposed by Altmann.

Let us begin with some qualitative data. The Slovene phoneme system contains 29 phonemes. There are 8 vowels /i, e, ɛ, ə, a, o, ɔ, u/. The phonemes /eː/ and /ɛː/ are long but differ in their degree of openness: /eː/ in [peːti] "to sing" is pronounced long and narrow while /ɛː/ in [pɛːti] "the fifth" is long and open. The phonemes /oː/ and /ɔː/ differ in their degree of opening, too: long and narrow /oː/ in [moːra] "he/she must" and the long open /ɔː/ in [mɔːra] "nightmare" are examples. Hence the degree of opening is phonemically relevant (cf. Rehder 1998: 231f., Lencek 1982: 160ff. and on Slovene phonemics in general Toporišic 2000: 46ff.).[1]

While there is a general agreement concerning vowel inventory, the size of the consonant inventory is defined differently in the secondary literature. Lencek (1982: 169) defines 20 consonants in the inventory: /v, m, n, r, l, j, b, d, g, z, ž, p, t, k, s, š, c, č, f, h/ while Rehder (1998: 232) differentiates 22: /p, b, f, v, m, t, d, s, z, n, n', r, l, l', š, ž, c, č, j, k, g, h/. However, the palatalized /n'/ and /l'/ can be considered as positional variants.

The present investigation starts from 21 consonantal phonemes defined according to Toporišic (2000: 85): /p, b, f, v, m, t, d, s, z, n, r, l, š, ž, dž, c, č, j, k, g, h/. The phoneme /dž/ must be emphasized here because it usually

1. An alternative counting of Slovene vowels would distinguish the opposition stressed-unstressed and the quantities short-long respectively. In that case Slovene would have 19 vowels (cf. The Survey in Toporišic 2000: 71). In order to make comparison with other languages possible, we refrain from this maximal system here. We start from the 8 vowels defined above.

does not occur in standard presentations of Slovene. Nevertheless, it must be unequivocally identified as a Slovene phoneme. The respective examples are: [džem] "jam" vs. [čem] "to whom", [džez] "jazz" vs. [čez] "through" (cf. Toporišic 1978: 73).

The Slovene script system contains 25 letters: <a, b, c, č, d, e, f, g, h, i, j, k, l, m, n, o, p, r, s, š, t, u, v, z, ž>. The first conspicuous characteristic is the fact that phoneme /ə/ has no written correspondence. It is represented by <e> or not at all. The second characteristic is the use of diacritics as in <č, š, ž>, which were gradually introduced in Slovene since the forties of the 19th century with the Illyric Movement[2] and have become part of the written variant of Slovene up to now. One can find them in several other Slavic languages.

This elementary presentation of Slovene phoneme and script system is a sufficient basis to begin the investigation. In Table 12 (p. 71ff.) one finds each of the 29 Slovene phonemes and the respective graphemes. The attribution of a phoneme to a grapheme is based on the Table in *Slovenski Pravopis* (1990: 143ff.). One starts there, however, from sounds. But the attribution of sounds to letters is irrelevant from the point of view of the investigation. Hence, the inventory of 64 sounds is here "reduced" to 29 phonemes of Slovene in order to study adequately the phoneme-grapheme correspondence.

First we state that the 29 phonemes of Slovene are represented by 28 graphemes. We emphasize here that graphemes are the letters or letter combinations that are used to represent the phoneme system of language on the graphemic level. Hence grapheme is defined as a unit constituted from its relation to a phoneme.

2. The inventory of 25 letters has been to a certain extent codified in the first Slovene orthography (*Pravopis*) in 1899 by Fran Levec (cf. Glozančev 1997: 85f.). This does not exclude the use of some "foreign" letters like <ć, đ, q, w, y, ...>. The publication of the first Slovene orthographic handbook initiated a discussion concerning not only the adequacy of letters for Slovene sound (this dispute has been performed in the so-called ABC-war in the twenties and thirties of the 19th century), but a series of special questions of Slovene orthography (cf. Dobrovoljc 2004: 113ff.). These concern the problem of capitalization, the writing of foreign words, that of personal names etc. The last orthography reform performed in 1990/2001 led to a lively scientific discussion (cf. Dobrovoljc 2004: 92ff.) and touched in general theses problems. However, this orthography reform did not cause any dramatic changes concerning the relation between phonemes and graphemes.

2 Economy of the script system

In order to scrutinize the phoneme-grapheme relationship in Slovene, we first compute the economy of the script system as the ratio of phoneme number to grapheme number (Best & Altmann 2005: 34). Denoting script economy by $SE = \#P/\#G$, for Slovene we obtain $SE(Sl) = 29/28 = 1.03$. This value can be considered the index of phoneme-grapheme correspondence.

For the Slovene script system we actually obtain a very high economy. A comparison with German $SE(G) = 0.57$ and Swedish $SE(Sw) = 0.63$ shows the economy of Slovene still more markedly. However, this index is global and does not express the phoneme-grapheme relationship in individual cases. In other words, one must take into account not only the fact that a phoneme can be represented by a single grapheme, but also the fact that graphemes are "polyfunctional" and can represent different phonemes.

3 Orthographic uncertainty of the phonemes

Here we examine the question how many graphemes are necessary to express a phoneme. It is reasonable to assume that the more different graphemes are "necessary" to represent a phoneme, the smaller the effectivity of the writing system. An extreme position is taken by English (cf. Fan & Altmann, this volume, pp. 25ff.). Using Best & Altmann's (2005: 32) definition the orthographic uncertainty of a phoneme can be given as "... the binary logarithm of the cardinal number of the representing grapheme set". The formula is

$$U_{/x/} = \log_2 n_x \,. \tag{1}$$

Here $U_{/x/}$ is the orthographic uncertainty of the phoneme /x/, $\log_2$ is the dyadic logarithm and n_x is the number of graphemes representing the phoneme /x/. A survey of representations and uncertainties for individual phonemes is presented in Table 1. Here for each phoneme the number of representing graphemes, the computed orthographic uncertainty and the number of phonemes with the given uncertainty are shown (n_x denoting the number of representing graphemes, U_x the uncertainty, and f_x the number of phonemes with uncertainty U_x).

It is conspicuous that in Slovene (cf. Table 1) the majority of phonemes (18 out of 29) can be represented by two or more different graphemes. This

Table 1: Unweighted orthographic uncertainty of Slovene phonemes

Phoneme	n_x	U_x	f_x
a, c, e, e, i, f, h, m, o, ɔ, r	1	0.00	11
b, d, g, k, j, l, n, p, t, z, ə, s	2	1.00	12
c, u, dž	3	1.58	3
š, v, ž	4	2.00	3

is an important hint at the fact that the sole interpretation of the phoneme-grapheme correspondence is not enough and must be completed by the unweighted orthographic uncertainty of the phoneme. The mean orthographic uncertainty of the whole system can be computed using data in Table 1. Let N be the number of all phonemes, f_x the number of phonemes represented by n_x graphemes (or graphemes with uncertainty U_x), then

$$\bar{U} = \frac{1}{N}\sum_x f_x U_x \,. \tag{2}$$

For Slovene we obtain $\bar{U} = [11(0)+12(1)+(3)1.58+3(2)]/29 = 0.7841$. Comparing the results with those presented in this volume, namely German ($\bar{U} = 0.965$), Swedish ($\bar{U} = 0.797$), Italian ($\bar{U} = 0.5641$) and Slovak ($\bar{U} = 0.7586$), one can state that the Slovene orthographic uncertainty is very similar to Swedish and Slovak. It is placed in the middle of the analysed languages, having the greatest "distance" from German and Italian.

In order not to rely on intuitive interpretation of the differences between the values, it is better to perform at least an asymptotic test answering the question about the significance of these differences in orthographic uncertainties. The criterion can be set up as

$$z = \frac{\bar{U}_1 - \bar{U}_2}{\sqrt{V(\bar{U}_1) + V(\bar{U}_1)}} \tag{3}$$

where $\bar{U}$ is the mean empirical uncertainty and $V(\bar{U})$ is its variance. For details see Bernhard & Altmann (this volume, pp. 13ff.). In Table 2 the individual values necessary for the computation are presented.

Using Slovene as the basis of comparison we obtain the z-values as shown in Table 3. We can state that at the 95% confidence level there is *no significant difference* between Slovene and other languages.

We can state that concerning the degree of orthographic vagueness the languages examined show similar behaviour. However, only further research can

Table 2: Mean uncertainty and its variance

Language	$\bar{U}$	$V(\bar{U})$
German	0.9650	0.012602
Swedish	0.7970	0.022763
Italian	0.5641	0.012189
Slovak	0.7586	0.012030
Slovene	0.7841	0.014057

show whether this fact can be explained by the time of script introduction or other cultural and linguistic factors. In any case, we could show that Slovene, having a small number of phonemes, can be represented by a small number of graphemes, which can partly be used in several functions.

Table 3: z-values of the comparison between Slovene and other languages

	German	Swedish	Italian	Slovak
Slovene	1.13	0.07	1.39	0.16

4 The distribution of graphemic representation

In connection with the orthographic uncertainty, we can ask whether the number of graphemes per phoneme can be captured by a theoretical model. Evidently, there is no one-to-one representation, hence the distribution is not deterministic, as can be seen in Table 1. There are phonemes which can be represented maximally by four graphemes. Using Bernhard & Altmann's (this volume, pp. 13ff.) and Mačutek's (this volume, pp. 75ff.) approach, we try to fit the Shenton-Skees geometric distribution to the data. As can be seen, there is no monotonic decrease; we have to do with a slightly bell-shaped distribution. In Slovene and German the maximum is in $x = 2$ (cf. Table 4).

The empirical data can be successfully captured by the Shenton-Skees-geometric distribution. Using the chi-square goodness-of-fit test we get $P = 0.30$ for Slovene, which is a good corroboration of the model. In Table 5 one finds the computed values for all analysed languages.

Table 4: Distribution of representation size of phonemes

x	Italian	German	Swedish	Slovak	Slovene
1	34	10	16	19	11
2	14	18	10	13	12
3	8	7	6	10	3
4	1	3	1	1	3
5	2	0	2	0	
6		1	1	1	

5 Grapheme size

The graphemes of a language may consist of one letter or of combinations of several letters. Besides digraphs, trigraphs etc., there is still the possibility of adding diacritical marks to letters in order to enhance the effectivity of a script system. In the Slovene script system there is a special mark above some letters called "haček": <č,š,ž>.

Adhering to the analysis in Best & Altmann (2005), Bernhard & Altmann (this volume, pp. 13ff.) and Nemcová & Altmann (this volume, pp. 79ff.), we shall determine the size of graphemes in two ways. We consider the number of Latin letters

1. in which additional marks are not considered as separate components;
2. in which additional marks are considered as separate components.

Table 5: Results of fitting

x	Italian	German	Swedish	Slovak	Slovene
1	33.31	9.99	15.79	19.36	12.20
2	14.92	18.00	9.99	13.40	11.33
3	6.37	7.54	5.35	6.58	4.91
4	2.64	2.47	2.64	2.83	2.56
5	1.76	0.73	1.24	1.14	
6		0.27	1.00	0.69	
p	0.6488	0.7768	0.6152	0.6813	0.7495
a	0.2398	2.3323	0.4588	0.7569	1.4216
DF	2	1	2	2	1
X^2	1.55	0.12	1.36	3.36	1.07
P	0.46	0.73	0.51	0.19	0.30

In Tables 6a and 6b one finds the respective grapheme size, the concerned units and their number.

Table 6a: Size of Slovene graphemes: Method 1

Size	Grapheme	Number
1	a, b, c, č, d, e, f, g, h, i, j, k, l, m, n, o, p, r, s, š, t, u, v, z, ž	25
2	dž, lj, nj	3

In order to get a compact characterization of the whole system, we compute in the first step the mean grapheme length in Slovene. Using Table 6a (without diacritical marks) we obtain 1.11.

Table 6b: Size of Slovene graphemes: Method 2

Size	Grapheme	Number
1	a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, r, s, t, u, v, z	22
2	č, š, ž, lj, nj	5
3	dž	1

Using method (2), which considers special symbols as grapheme components, we obtain the mean length as 1.25. In order to interpret intuitively the extent of these values we compare them with Slovak, Italian, German and Swedish mean grapheme sizes. The individual values (see Table 7) show in particular that the two Slavic languages have a low mean grapheme size.

Table 7: Mean grapheme sizes according to two methods

Language	Mean size (Method 1)	Mean size (Method 2)
German	1.68	1.78
Swedish	1.61	1.67
Italian	1.65	1.70
Slovak	1.16	–
Slovene	1.11	1.25

This is not surprising because the use of di- and trigraphs is not a central property of these script systems. The small grapheme size in Slovene can partly be explained historically: in the thirties and forties of the 19th century the first orthography leaned against the principle "write as you speak" which

means a 1:1 relation of phoneme/sound to grapheme. This principle asserted itself to a great extent.

A more thorough interpretation of grapheme size and of the position of Slovene will be possible after more languages have been analysed. Compared with other languages, Slovene has the lowest mean grapheme size, which is an indication of the "simple" construction of the Slovene graphemic system.

6 The graphemic load of letters

Having analysed the grapheme size, we examine the graphemic load of individual letters. We consider the presence of a letter in different graphemes. From another point of view we measure the graphemic "polytexty" of Latin letters in Slovene. This property is usually scrutinized with words but it has relevance in graphemics, too. In this analysis letters are taken into account without diacritical marks, i.e. in their Latin form. Hence the inventory consists of 22 letters, see Table 8. In this way comparability with other languages is made possible.

Table 8: Participation of Latin letters in Slovene graphemes

Component in x graphemes	Latin letter	Number of letters
1	a, b, c, e, f, g, h, i, k, m, o, p, r, s, t, u, v	17
2	d, l, n	3
3	j, z	2

Using the numbers in Table 8, we can compute the mean graphemic load of letters, resulting in 1.32, i.e. on the average, a letter occurs in "slightly more" than one grapheme. This is, of course, a very low value explained by the small number of digraphs: <lj, nj, dž>. A comparison with other languages is shown in Table 9.

7 Letter utility

The participation of a letter in a grapheme can be measured not only directly as presence or absence but also in a weighted way. Best & Altmann (2005: 37) take into account also the position of a letter in the grapheme. It is supposed

Table 9: Graphemic load in five languages

Language	Graphemic load
German	3.96
Swedish	3.36
Italian	3.92
Slovak	2.23
Slovene	1.27

that the later a letter occurs in the grapheme, the smaller its relevance. Though this need not be always the case, in the majority of cases it is true. Hence the position of a letter serves as the weight of its utility. We just have to consider all graphemes and for each letter its position. The results for Slovene are shown in Table 10.

Table 10: Letter utility

Weight	Letter x	Number of letters $f(x)$
1	a, b, c, e, f, g, h, i, k, m, o, p, r, s, t, u, v	17
2	d, l, n	3
3	–	–
4	z	1
5	j	1

The mean positional weight for Slovene is $PW(Sl) = 1.5$. Compared with other languages this value is very low. A comparison is shown in Table 11.

Table 11: Positional weight in five languages

Language	Positional weight
German	6.12
Swedish	5.41
Italian	6.48
Slovak	2.50
Slovene	1.45
English	47.35

Lastly we can compute the distance from the ideal state (cf. Nemcová & Altmann, this volume, pp. 79ff.). The ideal state is achieved if all letters are monographs. For Slovene we get the ideal state as $I = 22$, and the real

state as the vector of frequencies from Table 10 $W(Sl) = [17,3,2]$ yielding $D(Sl) = 6.16$.

Consequently, there is a very small distance between the real and the ideal state in Slovene. Slovak ($D = 17.78$) and Italian ($D = 21.6$) are more distant. Hence one can state that Slovene has an almost optimal graphemic system.

8 Conclusion

In this investigation we ascertain a number of characteristics of the Slovene graphemic system. A central property is a relatively small inventory of phonemes (world wide it is ca. 32) represented also by a small number of graphemes. While the phoneme-grapheme index indicates an "over-economy"; the computation of the orthographic uncertainty shows that Slovene graphemes can represent different phonemes. A statistical comparison of this property with that of other languages has shown that there is no significant difference between Slovene and German, Italian, Swedish and Slovak respectively.

It could be shown that the Slovene grapheme system has quite optimal characteristics. This concerns grapheme size, letter load, letter utility and positional weighting. Deeper insights can be drawn by further analyses, especially those of Slavic languages. A further step would be the embedding of the characteristics of the Slovene grapheme system in the synergetic control cycle of script theory (cf. Altmann, this volume, pp. 149ff.).

Table 12: The phoneme-grapheme relationship in Slovene
P = Phoneme, **G** = Grapheme, <cl.> = cluster

P	G	Examples	Remarks
/a/	<a>	sam, brat, riba, dna	
/b/	<b>	biba, objem, oblati, ob očetu/materi	
	<p>	snop daj	In front of voiced stops
/c/	<c>	Cene, stric, Micka	
/č/	<č>	peč, človek, pečka	
	<dž>	bridžka, bridž prinesi, bridž, bridž igra, bridž je	In front of voiceless stops, pause, vowel, in front of the sonorant of the next word
	<c>	stric šepa	In front of voiceless stops
/dž/	<dž>	Džungla	
	<c>	odločba, proč daj	In front of voiced stops
	<c>	stric Džon	In front of voiced stops
/d/	<d>	dedje, odgnati	
	<t>	svatba, svat bo	In front of voiced stops
/e/	<e>	kmet, ribe, pet	
/ɛː/	<e>	teta, seja	Long, open (ê)
/ə/	<e>	pes, temen	Semivowel
	<ø>^{<cl.>}	rdeč, Vltava, Mboja, Nkrumah	In neighbourhood of consonants (in front of r, l, m, n, v)
/f/	<f>	fant, Škofljica, Štefka	
/g/	<g>	glagol, Bogdan	
	<k>	kdaj k domu, vsak dan	In front of voiced stops
/h/	<h>	hahljati se, mehko, meh	
/i/	<i>	pila, sit, mami	
/j/	<j>	jajce	
	<i>	celuloid, bo imela	In quoted words; at word beginning, if the previous word ends with a vowel
/k/	<k>	krik, kura, pakt	
	<g>	bogca, rog, rog odnesi, rog nesi	In front of voiceless stops, pause, vowel, in front of the sonorant of the next word

(continued on next page)

Table 12 (continued from previous page)

P	G	Examples	Remarks
/l/	<l>	polje, ohol	
	<lj>	polj, poljski	Not in front of a vowel
/m/	<m>	imam	
/n/	<n>	Nina, njena	
	<nj>	sanj, sanjski	Not in front of a vowel
/o/	<o>	nos, gol, bratov	
/ɔ:/	<o>	noga, voda	Long, open (ô)
/p/	<p>	pipa, pnem, snop	
	<b>	Robca, rob strehe, Rob, Rob ima, Rob je	In front of voiceless stops, pause, vowel (except of prepositions), in front of the sonorant of the next word (except of prepositions)
/r/	<r>	trava, morda, dar, rdeč	
/s/	<s>	sila, sneti, stati, trs	
	<z>	nizka, niz, niz ima, niz je	In front of voiceless stops, pause, vowel, in front of the sonorant of the next word
/š/	<š>	šumiš, smešna, šteti	
	<ž>	ježka, jež, jež odhaja, jež je	In front of voiceless stops, pause, vowel, in front of the sonorant of the next word
	<s>	sčasoma	In front of voiceless sibilant
	<z>	iz časov	In front of voiceless sibilant
/t/	<t>	tat, tknalo, tkati	
	<d>	gladka, glad, glad ubija, glad je	In front of voiceless stops, pause, vowel, in front of the sonorant of the next word
/u/	<u>	kruhu, kruh	
	<e, cl.><v>	bukev, barv	In front of stressless semivowels, in consonant clusters
	<i,e, cl.><l>	nesel, nosil, videl, trl	With unstressed semivowels, with consonant clusters,
/v/	<v>	siva, vera	labiodental
		siv, sivka	A diphthong, when it alters to [v] (siva)
		v očeh	As preposition in front of vowels

(continued on next page)

Table 12 (continued from previous page)

P	G	Examples	Remarks
		vsak, predvsem	Voiceless bilabial, no diphthong
	<u>	Dachaua	In foreign words
		Nauk	In rare Slovene words
		bo ušla, bo uran	In word combinations
		ubiti	As prefix
	<f>	Afgan, grof gre	In front of voiced stops
	<l>	bral, bralca, kosil, topel, bel	In alternation with [l] (brala)
		trl	At word end
/z/	<z>	zet, zrelost, zdaj, cez oceta, mater	
	<s>	glasba, glas bobni	In front of voiced stops
/ž/	<ž>	žaga	
	<š>	izvršba	In front of voiced stops
	<z>	iz džungle	In front of voiced sibilant
	<s>	kos džungle	In front of voiced sibilant

References

Altmann, Gabriel
2007 "Towards a theory of script". This volume, pp. 149–164.

Bernhard, Gerald; Altmann, Gabriel
2007 "The phoneme-grapheme relation in Italian". This volume, pp. 13–23.

Best, Karl-Heinz; Altmann, Gabriel
2005 "Some properties of graphemic systems." In: *Glottometrics*, 9; 29–39.

Dobrovoljc, Helena
2004 *Pravopisje na Slovenskem.* Ljubljana: Založba ZRC.

Fan, Fengxiang; Altmann, Gabriel
2007 "Graphemic representation of English phonemes". This volume, pp. 25–59.

Gložančev, Alenka
1997 "Kratek pregled slovenskego pravopisja od konca devetnajstega stoletja do danes." In: *Jezikoslovni zapiski*, 3; 85–104.

Lencek, Rado L.
1982 *The structure and history of the Slovene language.* Columbus, Ohio: Slavia.

Mačutek, Ján
2007 "On the distribution of graphemic representations". This volume, pp. 75–78.

Nemcová, Emília; Altmann, Gabriel
2007 "The phoneme-grapheme relation in Slovak". This volume, pp. 79–87.
Pravopis, slovenski
1990 *Slovenski Pravopis. 1. Pravila.* Ljubljana: Državna Založba Slovenije.
Rehder, Peter
1998 "Das Slovenische". In: Rehder, Peter (Ed.), *Einführung in die slavischen Sprachen (Mit einer Einführung in die Balkanphilologie).* Darmstadt: Wissenschaftliche Buchgesellschaft, 230–245.
Toporišič, Jože
1978 *Glasovna in naglasna podoba slovenskega jezika.* Maribor: Založba Obzorja.
2000 *Slovenska Slovnica.* Maribor: Založba Obzorja.

On the distribution of graphemic representations

Ján Mačutek

In several articles of the present volume it has been shown that the graphemic representations of phonemes abide by the Shenton-Skees geometric distribution, which is, as a matter of fact, a modified (1-displaced) geometric distribution. It can be obtained from the geometric distribution by the Gram-Charlier expansion (cf. Shenton & Skees 1970; Berg 1985; Johnson, Kotz & Kemp 1992: 442 ff.). Moreover, a probability distribution can be constructed from a parent distribution P^* by modifying P^* in the following way,

$$P_x = P_x^* \left(1 + a(x - \mu*)\right), \tag{1}$$

where μ^* is the first moment of P^* about zero (mean) and the parameter a fulfils certain conditions. In general, $a \leq \frac{1}{\mu^* - n}$, where n is such an integer that $\mu^* > n$ and $\mu^* \leq n + 1$. In our case

$$P_x^* = p(1-p)^{x-1}, \quad x = 1,2,3,\ldots \tag{2}$$

with $0 < p < 1$. The expectation (mean) of (2) is $\mu* = 1/p$, hence we obtain from (1) and (2) that

$$P_x = p(1-p)^{x-1}\left(1 + a(x - \frac{1}{p})\right), \quad x = 1,2,3,\ldots \tag{3}$$

where $0 < p \leq 1$, $0 \leq a \leq 1/(1-p) - 1$ (if $p = 1$ then $a = 0$) representing the Shenton-Skees geometric distribution (cf. Wimmer & Altmann 1999: 593).

If the number of empirical frequency classes is greater than 3, the distribution can be fitted using the available software (which tests the fit by means of the chi-square criterion), but this is not always the case, as can be seen in the articles in this volume. Hence some elementary estimators must be derived. Mohanty & Altmann (this volume) proposed the estimators computed from the frequency classes; we show two other estimators.

Since the probability generating function of the Shenton-Skees geometric distribution is (writing $q = 1 - p$)

$$G(t) = \frac{t(p-a)}{1-qt} + \frac{apt}{(1-qt)^2}, \tag{4}$$

Quantitative Linguistics 63

Editors

Reinhard Köhler
Gabriel Altmann
Peter Grzybek

Mouton de Gruyter
Berlin · New York

Analyses of Script

Properties of Characters and Writing Systems

Edited by

Gabriel Altmann
Fan Fengxiang

Mouton de Gruyter
Berlin · New York