



Soviet-era science, translated into English

MATHEMATICS

S. M. OSOVETS

1962

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196201.56784>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

MATHEMATICS

S. M. OSOVETS

ON SOME STATISTICAL REGULARITIES IN THE WRITING OF LETTERS

(Presented by Academician L. A. Artsimovich, 5 V 1962)

Examination of the numerals and letters of the most widely used alphabets makes it possible to establish a definite regularity in the distribution of these signs according to their topological properties. The simplest example in this sense is provided by the numerals we use. Let us write out these numerals and, under each, indicate the corresponding topological genre.

$$\begin{array}{cccccccccc}
 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 1 & 1
 \end{array} \tag{1}$$

One may assert that the distribution of these signs by topological genre is described with sufficient accuracy by the relation:

$$n_k = n \left(1 - \frac{1}{e}\right) e^{-k}, \tag{2}$$

where n_k is the number of signs belonging to the given topological genre k ; n is the total number of signs. Here $1 - \frac{1}{e}$ is the normalizing factor, since

$$\sum_0^{\infty} e^{-k} = \frac{1}{1 - \frac{1}{e}}.$$

Let us write, in tabular form, the relative quantities obtained from relation (2), and the corresponding values in row (1).

	$k = 0$	$k = 1$	$k = 2$
Calculated	0.63	0.23	0.085
Corresponding to row (1)	0.6	0.3	0.1

It is seen that, to within an integer, these quantities coincide. Let us give an analogous table for the letters of the Russian, Latin, Gothic, and Greek

alphabets, for capital and lowercase letters (above the line are capitals, below the line lowercase):

	$k = 0$	$k = 1$	$k = 2$
Russian	0.595/0.57	0.31/0.34	0.063/0.063
Latin	0.72/0.64	0.24/0.33	0.042/0.042
Gothic	0.6/0.625	0.32/0.36	0.08/0.04
Greek	0.67/0.625	0.21/0.25	0.12/0.12

It follows from the table that the writing of letters in these languages also satisfies relation (2) with sufficient accuracy. For the Arabic languages, a predominance of genre 0 and of breaks is characteristic, since the basic writing is accompanied by strokes above the letters and below them. This, apparently, is less convenient both for memorization and for the process of writing the letters itself. It is characteristic that Roman numerals, whose writing was restricted to the zero genre, have practically gone out of use and are used mainly for designations, since they are inconvenient for memorization. The letters of the languages used in India, according to their writing, are divided into two groups: the devanagari script—these are the Marathi, Pahari, Sanskrit, Hindi, and other languages,

and a script going back to the ancient Indian syllabic Brahmi alphabet—these are the Kannada and Telugu languages. Although both of these scripts—Devanagari and Brahmi—have different roots and developed independently of the European alphabets, the writing of letters in these languages also satisfies relation (2). It is interesting to note that the letters in Kannada and Telugu contain a manner of writing corresponding to topological genre 3. This fits relation (2), since in these languages the number of letters reaches 55 and, consequently,

$$n \left(1 - \frac{1}{e}\right) e^{-3} > 1.$$

Let us give two letters of the Kannada and Telugu languages with $k = 3$:

[[two displayed letter glyphs]]

The existing manner of writing letters in the most widespread languages has undergone a certain evolution. Earlier letter forms differed from those used now and deviated substantially from distribution (2). Therefore there are grounds to assert that such a distribution is not accidental, but has some definite meaning—though not an entirely clear one. It is possible that such a distribution requires a minimal number of features for recognition.

What has been set out in the present note may prove useful, for example, in studies connected with the recognition of letters by automata.

Received
26 IV 1962

CITED LITERATURE

¹ R. S. Gilyarevsky, V. S. Grivnin, *Languages of the World*, 1957.

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.