Barbara Starosta
NOTES ON THE CONCEPT OF INFORMATION

Originally published as ”Uwagi o pojęciu informacji,” Studia Semiotyczne 4 (1973), 95–107. Translated by Małgorzata Szubartowska.

1. INTRODUCTION

The term ”information” is currently one of the popular expressions that make natural language utterances seem modern.”Information” is replaced by expressions such as ”signal,” ”sign,” ”symptom,” ”message,” ”list,” ”advertisement,” ”text,” ”report,” ”meaning,” ”sense,” ”content”(of a sentences or a word) etc. In science, ”information” is used as a synonym for ”entropy,” ”vagueness,” ”indeterminacy,” ”unawareness,” ”ignorance” or (with a different interpretation) as a synonym of ”negative entropy,” ”definiteness,” ”determinacy,” ”order,” and ”knowledge.”

The term ”information,” used in the first half of the 20th century, especially in physics and telecommunications in the description of the problems associated with the technique of transmitting messages can be found in the dictionaries of most sciences, especially economics, biology, psychology, sociology, even linguistics and history. For several years now it has been present in the theoretical terminology of scientific methodologies, epistemology and semiotics.

The expression ”information” did not enter these scientific languages directly from the natural language, but was taken from the theory of digital computers and communication theory (the theory of telecommunications). This was, among others, due to systematic mathematization and computerization of science, especially humanities.

The question arises: what is this ”information” which the exact sciences pass on — with all that it entails — to other branches of knowledge? This question about definition can be expressed more accurately: ”What does the term 'information' mean?" If we assume that the meaning of an expression is determined by the procedures of use (so by the rules operating in a particular language) and by its uses, i.e. the specific context and extra-linguistic situations in which a word occurs, the above question can be formulated as follows: "what are the uses and the procedures of use of the expression 'information'?"

In general, we can say that the expression "information" is used in the majority of cases as either a concrete name (the name of an object or an individual) or the name of a FUNCTION. Since the significant part of research is related to the definition and the use of the concept of information understood as a function, this procedure of use will be the chief topic of this paper. But before we get to the main point, a few comments are required on the first way of using the term "information."

2. THE TERM "INFORMATION" AS A CONCRETE NAME

"Information" generally performs the role of a concrete name when it is used as the name of various kinds of strings of signs, in the broad sense of the latter word. These are usually strings of digits or letters. They may, but need not, form certain entities, which, in turn, are also referred to as information. The synonyms for thus understood information are, as listed at the beginning, words like "signal," "sign," "symptom," "message," "list," "advertisement," "data" (numeric or letter), "text," "report" and others. It is worth noting that, in such cases, it is not about the content or the meaning of the used signs, but only about their physical structure. In the case of inscriptions, what comes into play is the amount and the distribution of ink on paper. In the case of vocal signs — the intensity, the pitch and the tone color of sounds. This is how the word "information" is used e.g. in the theory of digital machinery and in numerous works concerning their use. In such works, the term "information" is usually synonymous to numeric or alphanumeric DATA. Currently, more and more often it is being replaced by the term "data" in order to avoid ambiguity. For example, the expression "mathematical machines are used to process information" is replaced by "mathematical machines are used for data processing."

In communication theory, the term "information" is in many cases used interchangeably with the term "message." For example, in the phrase "transfer of information is one of the most important tasks of modern technology," the term "information" can be replaced with the term "message" without changing the meaning of the whole sentence. "Information" as a "message" is the name for words or images (transmitted in telecommunication systems), various physical quantities characterizing objects that are remote or impossible to observe directly (which are transmitted by telemetry systems), so in all these instances it is used as a concrete name.

For a better illustration of this procedure of use, I will refer to a simple example taken from data preparation techniques for digital machines. The main task in this case is to represent data using a sequence consisting of only two characters: zero and one (0,1). In other words, a given set of characters, such as letters of the alphabet Z = $\{a, b, c, d, e, f, g, h\}$, has to be represented using elements of the set $B = \{0{,}1\}$, in other words, an eight-element set has to be represented by a two-element set. Depending on how big the set Z is, the zero-one sequences required to represent its elements may be short or long. Let us denote this "length," that is, the amount of ones and zeroes in the sequence, by the letter N. When $N = 1$, only two different letters, like b or d, can be represented by the elements of the set B:b as 0 and d as 1, or the other way round — b as 1 and d as 0.The following letters could only be denoted by 0 and 1, and so they would not be any different from b and d. When $N = 2$, you can already write them down with four characters using zeroes and ones. For example:

$$\begin{array}{ll} a — 0\ 0 & c — 1\ 0 \\ b — 0\ 1 & d — 1\ 1 \end{array}$$

When the length N= 3, 8 signs can be written using 2 characters:

$$\begin{array}{ll} a — 0\ 0\ 0 & e — 1\ 1\ 0 \\ b — 0\ 0\ 1 & f — 1\ 0\ 1 \\ c — 0\ 1\ 0 & g — 0\ 1\ 1 \\ d — 1\ 0\ 0 & h — 1\ 1\ 1 \end{array}$$

Generally, L number of elements in the set Z which can be represented using a sequence that has length N elements of the set B is:

(1)

$$L = 2^N$$

The assignment of elements of the set B to elements of the set Z is called CODING, or simply a CODE. Elements of the set Z are called INFORMATION. It is stated, for example, that with an 8-element sequence $N = 8$ of binary elements, i.e. 0 and 1, we can code 256 units of information ($2^8 = 256$). In such a case, units of information are counted "by piece" and the elements of the set Z are the designations of the name "information." e.g. for the set under discussion these are the letters of the alphabet *a, b, c, d, e, f, g, h.*

However, we can assume that it is not the elements of the set Z, but sequences N of elements of the set $\{0,1\}$ needed to encode the elements of the set Z that will be called units of information I_h. We can write this in the form of the following relation:

$$I_h = N$$

The amount of information is therefore calculated by using the formula $L = 2^N$ in the following way:

$$N \log_2 2 = \log_2 L$$

Since $\log_2 2 = 1$

(2)

$$I_h = N = \log_2 L$$

N equals one, when L equals two. This unit is called a bit (binary unit). N is sometimes called Hartley's information measure after the author of one of the first works on the communication theory entitled *Transmission of Information*, published in 1928. Hartley's "information" is the name used for the method of assigning such a set of elements where each element consists of N binary elements — zero and one — to the elements of the set Z. The term "information" is thus used here as the name of the function.

3. THE TERM "INFORMATION" AS THE NAME OF A FUNCTION

Linking the concept of information with the concept of function has paved the way for generalizing and refining the latter concept even further. The concept of function is relatively well defined by the concepts of set and membership to the set, as well as the concept of ordered pair, so by the basic concepts of set theory. Function is also referred to as a particular kind of relation, namely, an injective relation whose domain is the set X and whose codomain is contained in the set Y. In other words, a function assigns to each element of X a particular (exactly one) element of the set Y, or, it transforms (maps) the set X into the set Y. A formal way to write that is presented as follows:

$$f : X \rightarrow Y$$

We ought to distinguish FUNCTION f from its specific values $f(x)$. A function is often defined by its domain, that is, the set X, and the formula *W(x)* which represents the element *f(x)* for $x \in X$. We say then that the function is defined by the formula *W(x)*. For example, the function of coding was defined by (1) $L = 2^N$, while the function of Hartley's information was defined by the formula

$$I_h = \log_2 L$$

Linking information with the concept of function is a good attempt at explaining the term under discussion, as it is closely related to the way the term "information" is used in natural language. In many cases, "information" stands for certain relations between linguistic expressions or utterances and the extralinguistic reality, or between signs and their user. In other words, following the intuitive sense of the language, "information" means certain semiotic — that is, semantic, syntactic or pragmatic — relations. To paraphrase the saying that without information there is no control, it could also be claimed that without information there is no reasoning and definitely no inferring, since the information function seems to be strongly connected to logical thinking.

Also, in natural language, certain properties of the information function are defined. We say that a piece of information may be bigger or smaller, it may have a zero or a non-negative value. We often consider the information contained in a sentence like "a dog is a dog" or in any tautological statement to have a zero value. On the other hand, utterances which appear unexpectedly and are somehow new, different from the original ones are treated as highly informative. Pieces of information can be added or compared, for example, the information of two unrelated utterances is perceived as a sum of the two pieces of information contained in each of these individual utterances.

Defining "information" as a function did not yet give this notion the opportunity to be introduced to the arsenal of basic theoretical scientific concepts. What "gave it a green light," was the linking of the concept of information with the concept of PROBABILITY proposed by Shannon in *The Mathematical Theory of Communication* (1948), who devised the following formula:

(3)

$$I = — \sum_{i=1}^{n} p_i \log p_i$$

where $p_i$is the probability of the i-th event, to form an n-element set. This formula resulted from the adaptation of the so-called Boltzmann-Planck formula used in statistical physics to the communication theory. It was introduced by Planck to determine ENTROPY, that is, a state of disorder (chaos) of a thermodynamic system

(4)

$$S = k \ln P$$

where S is the entropy of the system, P is the thermodynamic probability and k is the Boltzmann constant:

$$k = 1.38 \cdot 10^{-16} \text{ erg/deg}$$

3.1. Information and entropy

The term ENTROPY, first introduced by Clausius in 1865, served initially to determine the state of disorder and chaos in the motion of particles at a temperature above absolute zero. A typical manifestation of such chaotic motion is the so-called Brownian motion, which is a random, zig-zag motion of small particles suspended in liquid or gas. The function of entropy, which ties this concept with the concept of probability, was determined by Planck in 1900 by the above-mentioned formula (4) $S = k \ln P$. In the case of a ”heat” interpretation, entropy is connected with the temperature of the system, it is always a positive quantity and it takes on zero value at the temperature of absolute zero, in which, theoretically, all motion of particles ceases.

In his work, Shannon proved that it is possible to extend the scope of the term ”entropy” and assume that it can be designated to the uncertainties of ANY system, e.g. a system of signs transmitted through channels of communication. Under this assumption, the entropy of a given system is reduced when some ORDERING factor is operating, such as, in some cases, the lowering of the temperature of the system, ordering operations of the human mind etc. The chaotic motion of a particle in a liquid or gas may be interpreted, for example, as the inability to predict the direction in which this particle moves. In the case of such movement, all directions are equally privileged and equally POSSIBLE. The entropy of such a system is at its maximum.

But the moment some ordering factor operates, it increases the probability of any particular DIRECTION. In an extreme case, this direction can be completely determined. At that point, the entropy of the system becomes zero. The factor which reduces the entropy of the system was referred to as INFORMATION by L. Szilard (1928) and E. Shannon (1948). With this interpretation, information equals the amount of entropy reduction. Thus, the amount of information can be determined on the basis of the formula used to calculate entropy:

$$I = H = — \sum_{i=1}^{n} p_i \log_2 p$$

In the works by Brillouin and his continuators, the term ”negentropy” was adopted in order to emphasize that, although information is equivalent to entropy, it is in fact its opposite. In some works, the term ”negentropy” is used synonymously with the term ”information.” But it is a purely conventional matter and the formula (3) remains unchanged. It seems that with a specific interpretation of the formula (3) there is no need for a new term. For example, if a state of disorder or uncertainty will be interpreted as the state of our ignorance about some source of messages, then entropy is the measure of this ignorance. Any message obtained from this source reduces the state of our ignorance and brings in information. When we collect all messages provided in this source, the state of our ignorance becomes zero — we gained information, or, to put it differently, we became fully informed. Generally, we can say that information is a factor which imposes ORDER on the chaos of ignorance.

With a different interpretation, if we introduce the concepts such as MEASUREMENT and OBSERVATION, the relationship between entropy and information is as follows: the uncertainty about the results of measurement is determined by entropy. Each result reduces entropy and reduces uncertainty. This reduction of uncertainty provides information. Uncertainty before measurement corresponds with certainty after measurement.

3.2. Axiomatizing the concept of information

Shannon’s formula (3) was the result of an intuitive application of mathematical formulas to the specific needs of the experiment. The first question, which Shannon already posed himself, was about the possibility of deriving this formula from more general assumptions. In other words, the question arose whether, or to what extent, the concept of information is ”derivable” from other mathematical concepts.

In any situation where you have to deal with a certain function, it is convenient to describe the essential characteristics of this function with a familiar conceptual apparatus, and to identify other properties only on the basis of these fundamental axioms. The description of these properties depends on the choice of language. The language should be rich enough to be able to express through it a relatively large set of attributes associated with a given concept. One of such languages, used by many authors, is the language of modern mathematics, especially algebra. In order to describe information different algebraic systems are used, such as ring, field, lattice, Boolean algebra or vector spaces.

An algebraic system, or, in short, algebra, is an ordered pair (U, F), where U is any given set and F a finite system of operations. The basic terms used in the description of information are: the set of elementary events Q (otherwise called the space of elementary events), the set of subsets of Q (i.e. the class of sets called the set of events U) and elements of the class U, which are denoted by letters A, B, C etc.

The basic operations for the set $U$are: the union $\cup$ and the intersection $\cap$.

Algebra $(U, \cap\ , \cap)$ is a lattice if and only if the two operations, are defined as follows: For any two elements A and B of the set U:

(K 1)

$$A \cup B = B \cup A$$

(K 2)

$$A \cup (B \cup C) = (A \cup B) \cup C$$

(K 3)

$$A \cup A = A$$

(K 4)

$$A \cap B = B \cap A$$

(K 5)

$$A \cap (B \cap C) = (A \cap B) \cap C$$

(K 6)

$$A \cap A = A$$

(K 7)

$$A \cap (A \cup B) = A$$

(K 8)

$$A \cup (A \cap B) = A$$

A lattice is distributive if:

(K 9)

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

(K 10)

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

In the set U the relation of inclusion is a relation of partial order $\leqslant$

(K 11)

$$A \leqslant B \Leftrightarrow A \cup B = B$$
$$A \leqslant B \Leftrightarrow A \cap B = A$$

Thus, each of the operations $\cap$ and $\cup$, determines the relation of partial order.

A distributive lattice with two values 0 and 1, in which every element has a complement $\neg$ is called BOOLEAN ALGEBRA. We can write it

$$(U, \cup, \cap, \neg, 0, 1).$$

In this case, to the above-mentioned axioms we add the axioms which define the element 0, 1 and the unary relation $\neg$.

(K 12)

$$A \cup 0 = A$$

(K 13)

$$A \cap 1 = A$$

(K 14)

$$A \cap \neg A = 0$$

(K 15)

$$A \cup \neg A = 1$$

The elements 0 and 1 may be interpreted as follows: as either impossible and certain events or empty and the entire set Q.

Such a system of axioms is insufficient to define the field in which you can specify the information function. Class U has to be closed due to the operations of union, intersection and complement and it should also contain the entire set Q and the empty set $\emptyset$. Therefore, the axioms mentioned above should be followed up by:

(K 16)

$$\emptyset \in U$$

(K 17)

$$Q \in U$$

If $A, B \in U$ then

(K 18)

$$A \cup B \in U$$

(K 19)

$$A \cap B \in U$$

(K 20)

$$\neg A \in U$$

Thus defined class U is called the BOREL FIELD. The Borel field may be used to define the probability and the information functions. However, it requires two additional concepts: SET FUNCTION and MEASURE.

A SET FUNCTION μ is a function whose domain is a class of sets. Hence, it is a function defined not on the elements of the set Q, but on the elements of class U.

The set function defined on the elements of class U is called an ADDITIVE set function if for any two DISJOINT sets A, B in U , there is

(K 21)

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

(K 22)

$$\mu(\emptyset) = 0$$

A set function defined on the elements of class U is called COUNTABLY ADDITIVE if it satisfies condition K 22 and condition:

(K 23)

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu A_i$$

for every sequence of pairs of disjoint sets A_1, A_2 ... $\in U$.

MEASURE is a non-negative, countably additive set function with real numbers defined on the basis of the Borel field of events. We can write it

$$\mu : U \to R^+$$

where R^+ is the set of real, non-negative numbers. In other words, the function assigns positive real numbers and zero to the elements of the set U.

Measure μ is finite (that is, it takes on values which are only real, finite numbers) if and only if

$$\mu(Q) < \infty$$

In this case for every A U there is 0 (A) (Q). A measure is a probability measure if

$$\mu(Q) = 1$$

When μ is the measure of probability it is denoted by P. Hence, the probability function satisfies the following axioms:

(P 1)

$$0 \leqslant P(A) \leqslant 1$$

(P 2)

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

when $A_i \cap A_j = \emptyset$, i, j, $= 1, 2 \ldots\ldots$, $i \neq j$.

Algebra $(U,\ P)$ is called probability space. Only in this space is the information function defined. For every A, B *in* U

(K 24)

$$I(A \cap B) = I(A) + I(B) \Leftrightarrow \text{A is independent of B}$$

(K 25)

$$P(A) \geqslant P(B) \Leftrightarrow I(A) \leqslant I(B)$$

(K 26)

$$I(A) = 0 \Leftrightarrow P(A) = 1$$

A real function which satisfies the conditions K 24 — K 26 is the function composition —log$P$. For the properties of function log x present as follows:

1.

log x is defined on the subaxis (0, $+\infty$)

2.

$$\log 1 = 0$$

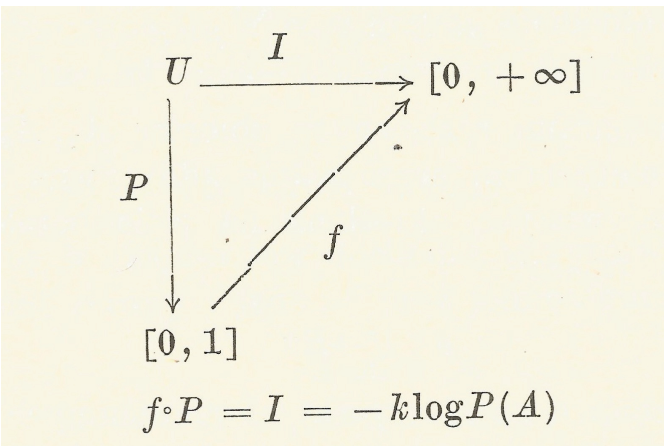
3.

the function log is strictly increasing when log > 0

4.

$$\log(x{\cdot}y) = \log x + \log y$$

Function I can be represented by the following scheme:



where k is any positive, real constant which takes on value 1 with a conventional choice of unit. f: [0,1] $\rightarrow$ [0, $+\infty$]

When we are dealing with a system G of mutually exclusive elements, it is possible to define the average information of this system by assuming that function I is a random variable which takes on values from [0, $+\infty$] with probability $P(A)$, $A \in G$. In that case, the expected value of $I(G)$ is:

$$I(G) = \sum_{A \in G} I(A)P(A) = - \sum_{A \in G} P(A) log P(A)$$

Thus, what we get is Shannon's formula. Schematically,

$$\begin{tikzcd} G \arrow[r, "I(G)"] \arrow[d, "\langle P, P..P \rangle"'] & {[0, +\infty]} \\ {[0,1] \times [0,1] \times [0,1] \times \ldots \times [0,1]} \arrow[ur, "I"'] & \end{tikzcd}$$

$$I \circ \langle P, P \ldots P \rangle = I(G).$$

Obviously, the axioms provided here are thus far the only ones. Numerous efforts have been made hitherto to make the conditions imposed on the domain (in which the information function is defined) less strict. For example, the distribution axiom K 9

(K 9)

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

can be replaced with the following expression when $C \leqslant A$:

(K 9)

$$A \cap (B \cup C) = (A \cap B) \cup C)$$

The lattice which satisfies the axiom K 9 is called a MODULAR lattice. Complemented modular lattices are not Boolean algebras. Yet, we can still determine their probability.

The measure of probability and information defined on a modular lattice is used e.g. in quantum mechanics.

Also, it is not necessary to define the concept of information on the basis of the concept of probability. The concept of information may be regarded as primary to the concept of probability. The first attempt in that direction was made by two Polish scholars R. Ingarden and K. Urbanik (1962), later by de Foriet and Forte (1967), Forte and Pintacuda (1968), and recently by Z. Domotor (1970).

However, these works still do not challenge the close relationship between the two concepts. For example, defining the information function in the Boolean domain of events (Ingarden and Urbanik) leads to univariate probability distribution.

4. CONCLUSION

The concept of information, used initially in communication theory, and then taken over along with the Shannon model by other branches of science, is closely linked with the concept of probability. Almost all works on the axioms of that concept use probability theory and measure theory. The mathematical apparatus used for this purpose is far richer than what was presented in Part 3, which was only meant to demonstrate the main idea behind this kind of axiomatization.

In contemporary works related to communication theory, the concept of information usually appears in the context of discussions about the transmission of messages over a noisy channel.

The function defined by Shannon's formula keeps, in this case, the name "entropy" as opposed to "information," which is used to determine the information transmitted over the channel (transmissional information). We denote the INPUT sequence (that is, random variables which describe the elementary input signals) by $\{S\}$ and the OUTPUT sequence by $\{Y\}$. The amount of information about the output sequence $\{Y\}$ provided by the input sequence $\{S\}$ equals the difference between the entropy of the output sequence and the average, conditional entropy of the output sequence where the input sequence is determined.

$$I(\{Y\} : \{S\}) = H(\{Y\}) — H(\{Y\} \mid \{S\})$$

If the input sequence is not known, the indeterminacy of the output sequence is $H(\{Y\})$. The indeterminacy decreases as the knowledge about the input sequence grows.

The properties of thus defined information function are as follows:

(1)

$$I(\{Y\} : \{S\}) \geqslant 0$$

(2)

$$I(\{Y\} : \{S\}) = I\ (\{S\} : \{Y\})$$

Therefore, in this discussion, the direction in which messages are transmitted is not important.

Another key concept related to information is channel capacity. It is the maximum amount of information which can be transmitted over a given channel:

$$C = \max_S I(\{Y\} : \{S\}) = \max_S H(\{S\}) — \mathrm{H}(\{S\} \mid \{Y\})$$

When the channel is noiseless $H(\{S\} \mid \{Y\}) = 0$

$$C = \max_S H(\{S\})$$

In the latter case, the amount of information transmitted over the channel equals the entropy of the input sequence. In other words, the produced output sequence eliminates the indeterminacy of the input sequence.

No matter if ”information” stands for entropy of the difference of entropies, it still remains the name of a function, which, in many cases, can be calculated. Like the terms ”weight,” ”pressure” and ”temperature,” it falls under the category of names designating classes of equivalence relations: ”equally informative” (just as ”equally heavy,” ”equally warm,” ”equally cold” etc.).

The need for finding measures of information measures and refining the concept of information has to do with communication difficulties of the modern world. These include the problem of the so-called information boom, i.e. the continuously increasing number of publications of all sorts. The question arises as to what novelties these publications introduce.

The issues raised more and more often concern the languages used in publications (especially in scientific publications) and the appearance of new expressions which are not always a response to new situations or the discovery of previously unknown relations. Gradually, it is becoming more difficult to transmit knowledge, which may consequently result in its rate of growth slowing down.

The concept of information is closely related to these issues and is often used to describe various types of analyses and possible solutions to these problems. A closer analysis of the concept of information makes it easier to use it properly.

The main task is to establish the connections between the concept of information and the concept of meaning or content — that is, the sense of linguistic expressions. These topics are explored by numerous researchers committed to analyzing the concept of information, generally referred to as "semantic information."

Bibliography

1. Białynicki-Birula, Andrzej (1971) *Algebra*. Warszawa: PWN.

2. Brillouin, Léon (1969) *Nauka a teoria informacji*. Warszawa: PWN.

3. Domotor, Zoltan (1970) "Qualitative Information and Entropy Structures." In *Information and Inference*, Jaakko Hintikka and Patrick Suppes (eds). Dordrecht: D. Reidel Publishing Company.

4. Fisz, Marek (1969) *Rachunek prawdopodobieństwa i statystyka matematyczna*. Warszawa: PWN.

5. Hartley, Ralph (1928) "Transmission of Information." *Bell System Technical Journal* 7: 535—563.

6. Ingarden, Roman (1963) "A Simplified Axiomatic Definition of Information." *Bull. de l'Acad. Pol. des Sciences, Sêrie des math., astr. et phys.* 9[4]: 209—212.

7. Ingarden, Roman (1964) "Teoria informacji i termodynamika maserów." In *Postępy Fizyki* 15[3]: 287.

8. Ingarden, Roman and Kazimierz Urbanik (1962) "Information Without Probability." *Colloquium Mathematicum* 9 131—150.

9. Kobozev, Nikolay I. (1971) *Issledovaniye w oblasti termodynamiki protsessov informatsii i myslenia*, Moskva: Izdatielstvo Moskowskogo Uniwersiteta.

10. Maurin, Krzysztof (1971) *Analiza*, part I. Warszawa: PWN.

11. Nowakowski, Jerzy and Władysław Sobczak (1971) *Teoria informacji*. Warszawa: Wydawnictwa Naukowo-Techniczne.

12. Shannon, Claude E. and Warren Weaver (1949) *The Mathematical Theory of Communication*. Urbana: The University of Illinois Press.

13. Traczyk, Tadeusz (1970) *Wstęp do teorii algebry Boola*. Warszawa: PWN.

14. Ursul, Arkadiy D. (1971) *Informacja*. Moskva: Nauka.

15. Watanabe, Satoshi (1969) *Knowing and Guessing*. New York: John Wiley and Son.