SEMANTIC THEORIES OF INFORMATION

Gilda Maria Braga
CTR/IBICT

ABSTRACT

All mathematical theories of Information in the late fourties suggested rival concepts of information; furthermore the notion of "meaning" seemed to have been left out Shannon's analysis of the amount of Information in a signal disclaimed explicity any concern with its meaning and had been qualified as inadequate by the Semanticists. Bar—Hillel and Carnap, 1952, suggested two possible measures of the Information content of statements in an artificial language system. Schreider, 1965. states that in several situations the receiver's ability to understand the communication is the most Important characteristic of the process. Goffman expanded Shannon's theory in his General Theory of Communication, where there are 3 large phenomena to be considered — Information generation, transmission and use. If we consider the effect of Relevance on the receiver and explore further its implications me might be closer to a better and unified theory of information than we are with the introduction of meaning — a point still in discussion among Semanticists. (HB)

Descri tores:
Teoria da Informação; Semântica; Teoria da Comunicação; Significado; Relevância; teoria Semântica da Informação; Informeção.

In 1948 Claude Shannon, an engineer from the Bell Co., published "A Mathematical theory of communication"(9), which proposed for communication signals a measure based on the statistical improbability of the signal. Since the log. of improbability is additive for independent signals, this had attractive properties. Rapoport(8) points out that a 1953 bibliography of information theory had already some 800 entries. A wide variety of papers speculated (and still do) on how Shannon's theory related to the "structural" and "metrical" measures already existing.

Shannon was by no means the first one to have the idea of measuring information. In 1946 Dennis Gabor published the 'Theory of Communication", applying the Fourier transform theory to the frequency-time domain of communication, suggesting that a signal occupying an elementary area $(\Delta f . \Delta t = 1/2)$ could be regarded as a unit of information (logon). Even earlier than that (1935) the Statistician R. A. Fischer had proposed a measure of the "information" in a statistical sample which amounted, in the simplest case, to the reciprocal of the variance(5).

All those theories, in the late forties, suggested rival concepts of information; furthermore, the notion of "meaning" seemed to have been left out: Shannon's analysis of the amount of information in a signal disclaimed explicitly any concern with its meaning.

There are several papers attempting to define "information" (in Shannon's sense) in non-mathematical language(8). Roughly speaking, we may say that we gain information when we know something that we didn't know before. The classical example of coin tossing illustrates the simplest case of gaining information in a case where *2* choices (answers) are possible. The overall idea is the enabling of a selection from a set of possibilities or to narrow the range of possibilities about which we are ignorant. So the selective information-content of a message or of the result of a scientific experiment, for example, has to do with the number of independent choices between two possibilities which it enables us to make — the number of independent "yes" or "no" (or "head" or "tail") to which it is equivalent.

The number of choices is at a minimum when we arrange to choose between (two) equally likely possibilities.

Communication is a process — i.e., a sequence of events where information is transmitted from one object to another. In the communication process the sequence leads to the transmission of information.

Shannon's measure of the quantity or amount of information transmitted (communicated) has been qualified as inadequate by the Semanticists; in the paper where Shannon defined amount of information (essentially a log measure of the statistical

* Trabalho elaborada durante o Programa de Doutoramento na Case Western Reserve University, USA, 1976

* * Bolsista da CAPES, Proc. 601/75

unexpectedness) he insisted that the concept of meaning was outside the scope of his theory.

In 1952 Bar-Hillel and Carnap(1) suggested two possible measures of the information content of statements in an artificial language system.

Their language system deals with a finite number of *individual constants* which stand for individuals (things, events or positions) and a finite number of *primitive one-place predicates* which designate primitive properties of individuals. In *an atomic statement* — e.g., $P_1 a_1$ (the individual has the property is P_1) a primitive property is asserted to hold for an individual. For the derivation of quantitative measures, this theory considers not the structure of a given representation, but its power to imply possible statements (analogy with Shannon's "what one could say", instead of what one says)(9). Bar—Hillel and Carnap start considering the large class of statements formed by each possible atomic statement — or its negation — and their combination with the disjunction (B1 or B2 or not — C1). These are the weakest possible statements, called "content elements". The class of content—elements logically implied by a given statement is called its *content,* and this content is suggested as an explication of the information conveyed by it; A measure of this class (content—measure) is suggested as one possible explanation of the amount of information conveyed, in its semantic sense. This measure equals *0* if a statement is true, and equals / if it is false; it can be viewed as the inductive probability of the negation of the statement. One important point is that the measure *is not additive* for independent statements. So, the negative log of the inductive probability of the statement itself is proposed as a measure of information — this measure is additive and the only difference with Shannon's measure is the use of *inductive* probability (Shannon uses *statistical* probability).

In 1965 Yu A. Schreider(10) made another round of attack — from the Semantic point of view — against the lack of "meaning" in Shannon's theory.

Schreider says that "where the amount of information is expressed by quantities of the entropy type, this strict concept is not sufficient for the description of all situations and does not cover all the properties of information with which ones has to deal". He states that in several situations the ability of the receiver to understand the communication is "the most important characteristic of the process" — and exemplifies with machine translation. He then proceeds to explain that in Shannon's theory, "before even considering the information contained in a statement about a given event, it is necessary to consider the a *priori* probability of the event concerned", and adds that this can be impossible in some cases.

If a set of events $A_1, A_2, \ldots, A_n$ are possible a *priori* with probabilities $P_1\ p_2, \ldots, p_n$, the receiver has a kind of guide that characterizes his "external world": it is a *priori* knowledge. Upon receipt of statements the description of the external world will change to:

$$A_1, A_2, \ldots A_n$$
$$p'_1, p'_2, \ldots p'_n$$

This change of the receiver is characterized by the "entropy" type quality in Shannon's theory.

Schreider says that it is important that the *quantity of information received is measured by the degree of change* of the guide to the external world. From there on he approaches the meaning characteristics of information, starting with some definitions:

a) let's assume the existence of a thesaurus *9,* which is a guide in which our knowledge is recorded; this 0 is equivalent to a list of events and their probabilities;

b) each statement T changes the state of 0 i.e. is equivalent of a transformation in it;

c) *the amount of information I (0,T) is the degree of change* of the thesaurus 0 under the action of a given statement *T,* which implies an operator *At,* for each 7 entering *6;*

d) a thesaurus changing into itself means / *(d,T)* = 0;

e) two statements (texts) *T* and T1 are synonymous (i.e., carry the same information) if they correspond to the same transformation operator of the thesaurus: $A_t = A_{t1}$

f) two texts can have identical amounts of information $I_{(T1\ 0)} = I\ (T_2 0)$ but the information may be different. In this case the texts are not synonymous;

g) if we consider the rules for constructing the operator A_t as something outside *8,* then the quantity / *(T,0)* depends on these rules; otherwise / *(T, 0)* is determined by *T* and by the thesaurus (in the wide meaning of the word);

h) if we consider separately a thesaurus, in the narrow sense, and don't include rules for A_t then the quantity / *(T,0)* can also increase;

i) a primitive thesaurus cannot comprehend a text *T* (gains no information); a developed one understands *T* and obtain maximum information from it; a satured thesaurus cannot gain information because it knows already everything (But is this situation possible, in the "real world"?).

From these preliminary definitions Schreider builds a formal description of his model, considering 3 sets for the construction of the thesaurus:

P {a,b,. . . } predicates, with 2 relationships: consequence (⊃) and applicability (⊢)
M {x,y,. . . } objects
C {α,β, . . . } events
The relationships of *P* have the properties of:
1. a ⊃ a
2. If a ⊃ b and b ⊃ c then a ⊃ c
3. If a ⊃ b then b ⊢ a
4. If a ⊢ b and b ⊢ c, then a ⊢ c

From 3 and 4:
5. If a ⊃ b and c ⊢ b then c ⊢ a
Pairs from sets *P* and *M* can be formed as:
ax (object x possesses property a)
¬x(object x does not possess property a)

After introducing the notion of "atomic proposition" (front the ternary function R (a,x,K) where a & P, x & M and K is an integer and $(ax_1, x_2 \ldots x_i)$ is admissible if and only if (∀ K) R(a,xk,K) is true) Schreider introduces a list of "quantors", which aid the determination of events.

Basically, there are 2 quantors:

d — generation (rise, introduction)
Ex.: (dx) (ax) "there arises x which has property a"
L — description

Ex.: a(lx) (bx) "that x which has the property b has the property a". An extension of this latter is in the quantors for set *description,* which acts only on objects belonging to M: (a(Ex) (bx) "those x which have the property b have the property a") and *general description,* which acts on all objects which may later belong to M: (a{8 x) (bx) "all those x, which have property b, have the property a). The proposition a (9 x) (bx) is equivalent to the relationship b D a.

Based on the relations and quantors, an *event* (C [a, ft...}) is defined — and there are several rules imposed on *C,* based on relations between objects and predicates.

Schreider defines his thesaurus *6 as* the "aggregate of the sets P, M, and C with the relationships and given in P".

There are several operations acting on the thesaurus, denoting the changes performed. Schreider calls them "canonical forms". Basically, these forms analyse the a *priori* state of the thesaurus, the introduction of an expression containing a quantor and the transformation of the thesaurus under elementary operators.

The quickest way to understand what is meant is to look at one of the examples given by Schreider:

Let's consider the text "Today a new house was built". Also let's assume these predicates had been given beforehand: a_t (house, to be a house) a_2 (new) a_3 (to be built) a_4 (today). The canonical expression is:

1. (dx) $(a_1 x)$
2. a_2 (Lx) $a_1 x$
3. a_3 (lx) $a_2 x = \alpha$
4. a_4 α_1

The statement started with elements of set P; sets M and C were filled in the process of text analysis.

One important feature is that if we cannot distinguish any canonical expression in T, then this text cannot transform the thesaurus. It would be not difficult to think about jtexts that are not identifiable as a canonical expression. A concept such as "destruction" — can it be considered as a negation of the quantor *d,* "construction"?

Several rules have to be considered for the actual construction of an algorithm for text analysis. Schreider avoids the issue saying that this problem is closely connected with the problems of mechanical translation.

One important aspect is pointed out: in this theory the two thesauri must have the same order of complexity if they are to understand each other — man-machine communication poses a challenge to this issue.

Schreider points out that his theory is only a crude model of the process of comprehension, and that "attempts to describe fully the semantic portrait of the words of the given language, independently of a concrete thesaurus, lead, it seems, to immensely complicated and badly formulated problems".

The Semanticists have been too eager to introduce the concept of meaning in Shannon's theory — but there are several kinds of meaning: the receiver's, the sender's and what may be called the "conventional meaning".

Schreider's theory attempts to measure the degree of change in our guide (thesaurus), which represents our knowledge about the external world. But he accomplishes little and his theory is too limited by all sets of rules imposed upon the transformations, operators and our lack of knowledge of message transfer in machines. I wonder whether this latter paint can bring all the help he expects. Machine translation seems still a dream for everyday purposes. Also his statement that no transformation of the thesaurus equals no information seems rather strict - a message does have some meaning, even if the receiver already knows what is being said: a message does not lose meaning because, it is repeated.

An important point is the concept of meaning in itself: what does it mean, "meaningless" or "meaningful"? May be the Semanticists have been too hasty in trying to introduce in information theory a concept that is not clearly defined nor limited. We are still far from knowing what are the processes that determine a question and bring into being a "meaningful answer".

Certainly Shannon's theory is a limited view of information, but before throwing it out as "meaningless" or "incomplete", let's have a closer look at its limitations and possibilities.

The theory is limited by being under the constraints of a finite scheme and probabilities. But in real situation, are our choices unlimited? And probabilities are far more reliable than it might seem at a first glance; even for hard-believers, a process of coin tossing can demonstrate how probability laws are close to "real life".

W. Goffman expanded Shannon's theory in his "General Theory of Communication"(4,3); what follows is a brief resume of some of its aspects.

In the information-communication theory there are 3 large phenomena to be considered:

— Information generation
— Information transmission
— Information use (still a big question mark)

What about information generation? If we take 3 basic concepts like data, information and knowledge, can we consider them as synonymous? are they different concepts? Certainly it is very hard to tell the differences in a complex level — and these 3 concepts are present at the very beginning of a "Science" process — *Science* is, under various aspects, a process of information generation.

Let's consider an arbitrary source of information S (objects, for example)

$S \; \{ S_1, S_2, \ldots, S_n \}$

Let's associate with S another set $\bar{S}$ (more chaotic)

$\bar{S} \; \{ \bar{s}_1 \; k_1, \bar{s}_2 \; k_2, \ldots, \bar{S}_{nkn} \}$

How can we relate these two sets? Obviously the first relationship is to look for common attributes in the two sets, whatever these attributes may be.

If we consider $m(s_i \wedge s_j)$ the number of attributes common to s_i and s_j and $m(s_j)$ the number of attributes associated with s_j, we can define a relation *r* as

$$r_{ij} = \frac{m(s_i \wedge s_j)}{m(Sj)}$$

This relation is a ratio of redundancy (common attributes) and can also be established between an observer S_0 and an arbitrary sequence ($s_1, s_2, \ldots, S_n$), showing a relation between the observer and the elements in terms of the observed attributes. By the way its a relation of Relevance, which can be established according to certain conditions imposed upon a lower bound.

Without entering into a discussion of *Relevance,* we can infer some of its implications, which don't require the complications of a semantic theory built upon canonical forms: — the main point in Schreider's theory is the degree of change introduced in the receiver's thesaurus upon receipt of some statement; since nothing is said about the primitive state of the thesaurus, can't we consider an analogy among an observer S_0 and an arbitrary set of events S and *measure* the common elements between S_0 and S? A set of events can be partitioned in disjoint classes; we can establish the relevant measure or let it be an arbitrary number between 0 and 1. The lowest bound B, zero, would mean no partition at all, i.e. no means of identifying common elements between S_0 and S - as the bound B would get close to 1 we would obtain more identifiable elements — the classes would decompose themselves in a finer way.

If we consider the effect of this measure r on the receiver and explored further its implications we might be closer to a better and unified theory of information than we are with the introduction of meaning — a point still in discussion among Semanticists.

And what is Relevance, in a broader sense? Isn't it "meaningful information" for the receiver?

REFERENCES

1 - BAR-HILLEL, Y & CARNAP, R. -Semantic Information.Sr/'f. *J. Philos. Science,* 9:12-27, 1954.

2 - BELKIN, N.J. - Some Soviet concepts of information of Information Science. *JASIS,* 26 56-63, 1975.

3 - GOFFMAN, W. - General Communication Theory. Notes taken in class, Fall 1975.

4 - GOFFMAN, W. - A General theory of communicatk communication. In: Saracevic, T.*Introduction to Information Science.* New York, Bowker, 1970. p. 726-747.

5 - McKAY, D. - *Information, mechanism and meaning.* Cambridge, MIT Press, 1969.

6 — MALONEY.. J. Semantic Information. *American Documentation,* 13:276-287, 1962.

7 — RAISBECK, G. - *Information theory:* an introduction for scientists and engineers. Cambridge, MIT, 1963.

8 - RAPOPORT, A. - The promise and pitfalls of Information theory. In: Saracevic, T. — *Introduction to Information Science.* New York, Bowker, 1970- p. 13-17.

9 - RAPOPORT, A. - What is information. In: Saracevic, T. — *Introduction to Information Science.* New York, Bowker, 1970. p. 5-12.

10 - SHANNON, C.E. - A Mathematical theory of communication. *Bell Syst. Tech. J.,* 27:379-423, 1948.

11 - SCHREIDER, Yu. - On the semantic characteristics of information. In: Saracevic, T. — *Introduction to Information Science,* New York, Bowker, 1970. p. 24-32.

12 - SHERA, J.S. - Of Librarianship, documentation and information science. In: Elias, A. ed. *Key Papers in Information Science.* Washington, ASIS, 1971. p. 4-11.

RESUMO

Ao final dos anos 40, todas as teorias matemáticas da informação sugerem conceitos antagônicos de informação, além de parecerem deixar de lado a noção de significado. A análise de Shannon sobre a quantidade de informação contida em um sinal declina explicitamente de qualquer interesse pelo significado, sendo qualificada pelos semanticistas de inadequada. Bar-Hillel & Carnap, 1952, sugerem duas possíveis medidas de conteúdo de informação nos signos em um sistema em linguagem artificial. Schreider, 1965, declara que, em diferentes situações, a habilidade do receptor em entender a comunicação é a característica mais importante do processo. Goffman, na sua Teoria Geral da Comunicação, expande a teoria de Shannon, onde existem três grandes fenômenos a serem considerados - geração, transmissão e uso da informação. Se considerarmos o efeito da medida Relevância sobre o receptor e analisarmos suas implicações estaremos talvez mais próximos de uma Teoria Unificada da Informação, do que estamos com a introdução do significado — ponto de discussão entre os semanticistas. (HB)