DOCUMENT RESUME

ED 111 386 IR 002 440

AUTHOR Pearson, Charls
TITLE Goals for Long-Range Research in Information Science.
PUB DATE 1 May 75
NOTE 16p.; Paper presented at the National Information Retrieval Colloquium (12th, Philadelphia, Pennsylvania, May 1-2, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS Communication (Thought Transfer); *Information Science; *Information Theory; Models; Objectives; Research; *Semiotics
IDENTIFIERS Georgia Institute of Technology; Peirce (Charles Sanders)

ABSTRACT

In order to discuss the research goals of information science (IS), both its logical and its specific nature must be determined. Peircean logical analysis shows that IS may be classified in three parts: pure science, applied science, and technology. The deficiency in the present state of the art is in the pure science, or theoretical portion, of IS which must be developed in order to fully and properly apply the existing technology. The specific nature of the science underlying information technology is semiotics, the science of signs and sign processes, their structure, and their applications to the communication process and information transfer. Long-range research goals should therefore be formed in order to (1) develop a theory of the structure of all categories of signs, sign systems, and sign processes; (2) investigate the measurable properties of all sign components; (3) explore the basic regularities existing between the measurable properties; (4) develop theories which explain these regularities; (5) investigate the relationship between various information processes and semiotic processes; and (6) develop lab instrumentation with which to carry out the critical experiments in the above areas. (Author/LS)

GOALS FOR LONG-RANGE RESEARCH IN INFORMATION SCIENCE

Charls Pearson
School of Information and Computer Science
Georgia Institute of Technology

May 1, 1975
Atlanta, Georgia

An invited paper prepared for the 12th National Information Retrieval Colloquium; May 1-2, 1975; Drexel University, Philadelphia. Presented at the session on Long-Range Planning for Basic Research in Information Science.

Goals for Long-Range Research in Information Science

Charls Pearson

I. Introduction

I would like to limit my remarks today to the goals for research in information science, but such a narrow view is impossible given the current state of the field itself. Before we can even begin to analyze the goals for research in information science, long-range or otherwise, we must answer two questions that are fundamental to this consideration. The fact that these two questions are just now being asked by many serious investigators who call themselves 'information scientists' and that no definitive answer has yet been given that is generally accepted marks this field as a discipline in transition.

The two questions of interest both concern the nature of information science as a discipline and can both be summarized by the formula: What *is* Information Science?, or perhaps more epigrammatically,

(1) What IS I.S.?

However, despite the identity of their surface structure, these two questions are fundamentally distinct. For example, we might try to give one of the more popular answers to (1), such as (2), (3), and (4):

(2) I.S. IS SEMIOTICS.

(3) I.S. IS CYBERNETICS.

(4) I.S. IS GENERAL SYSTEMS THEORY.

But this would be to get the value cart before the logical horse. (2), (3), and (4) attempt to find specific values for some entity before we have determined the logical nature of that entity. If the logical nature of I.S. is a song then in place of the symbol 'SEMIOTICS' in (2) we ought to have the name of a song. To ease further discussion let us set up what might be called an 'epigram schemata' as in (5):

(5) I.S. IS________

or what amounts to the same thing

(6) I.S. = x

If I.S. were a number, then x would be replaced by a numeral —— the name of a number. It appears that 'SEMIOTICS', 'CYBERNETICS', AND 'GENERAL SYSTEMS THEORY' are specific values of three entirely different kinds of logical entities. Our first problem, then, is to determine the logical nature of the variable x and its range. Only then can we pick a particular value from this range to assign to the variable x. We can represent the prior problem by formula (7), and the latter by (8):

(7) What is the logical nature of I.S.?

(8) What is the specific nature of I.S.?

Once (7) and (8) have been answered, I think the goals for long-range research will be obvious to everyone. I have already remarked that no generally accepted definitive answers to these two questions yet exist. I would now like to give my opinions as to how they should be answered.

II. The logical Nature of Information Science

In order to answer (7) I should like to adopt the method for logical analysis of knowledge first developed by Peirce in the late 1800's [5]. This consists in trichotomizing all of cognitive knowledge according to whether it has the quality of 'firstness', 'secondness', or 'thirdness' and results in classifications which may be labeled 'science', 'engineering', and 'technology'. The labels 'pure science', 'applied science', and 'technology' are often used to mark the same distinction and thus serve as synonyms to the above terms, as do also 'basic science', 'practical science', and 'practics'.

Now I believe there is no dispute that information science concerns some part of cognitive knowledge and that it lies wholly or in part in one or more of the above classifications, but there is some disagreement as to which one, if indeed it does fall entirely within one of these classes.

The easy way to anwer this question would be to say that the Peircean classification motivates us to distinguish between information sciences. information engineering, and information technology and then information science is obviously information science and there is nothing more to be said about it. Now I happen to agree with these dinstinctions and the conclusion they lead to, but not by the above reasoning. I think there are some important, nontaxonomic, reasons for reaching this conclusion.

Since 1950, when the Eighty-first U.S. Congress authorized the National Science Foundation to "foster an interchange of scientific information among scientists in the United States and foreign countries," the Federal Government has been funding various science information centers and information science research organizations in an attempt to develop an information technology. Efforts were initially directed where it was thought that fast benefits could most easily be found —— towards technological applications.

During the past twenty years, we have witnessed the remarkable growth in the development and application of information engineering and technology. The engineering progress in this field has been rapid and dramatic in terms of developing equipment of high information processing capacity and high reliability. The technological progress has been equally swift with continued and sometimes drastic lowering of costs and continued increases in the kind and scope of applications. This technology has been (and is continuing to be) applied to information handling situations in all facets of our society. And there are strong pressures to further exploit this technology and to use it fruitfully in still new areas.

But the development of technology has often not proven to be the easy matter it was hoped for. There are many areas where the technology has been applied but where it is grossly under-used, or misunderstood. When it comes to the automatic handling of non-numerical information for the purposes of interrogation, search, and retrieval, we find a surprising mismatch between the high capacity of the technology and the rather incipient logical level at which it is employed for problems of information storage and retrieval. There were also strong expectations, with the emergence of new ideas in cybernetics and communication theory and with the development of the digital computer, that we would have a *science* of information retrieval from which effective systems would derive. But very little theoretical work on information retrieval has emerged.

From a conceptual viewpoint, we are still at a very primitive level with respect to a theoretical understanding of the full problems of information retrieval. For instance, technologists attempted to solve the information storage and retrieval problem without having first a well-developed measure of the value of information, a theory of information representation, or an understanding of human information processes. As a result, science information storage and retrieval systems have not met with the huge success originally anticipated

The problem would appear to be with the state of the discipline itself. It can now be seen that we were trying to develop a technology without having first a well-developed scientific foundation upon which to support it. There were no natural laws to apply in solving technological problems--no Newton's Laws for information to integrate or differentiate as the technological goals demanded.

The realization that a scientific foundation must be established to provide insight and support for technological developments is beginning to grow. The 1972 NATO Advanced Study Institute in Information Science reached substantially this conclusion. Its working group #9 reported

> . . . . . a discipline consists of interaction among three parts of a system: a science, applications, and education. . . no discipline can survive without all three components--each supporting the others. In information science, the weakest component is the science. Without it the applications and education will eventually not survive. Therefore, the problem is: what are we now doing, or could we do to strengthen the scientific component of information science regarded as such a three-component discipline? [2]

The working group on Information Technology reached a similar conclusion. They recommended:

> . . . society must be willing to risk a certain amount of resources or "venture capital" in efforts directed at furthering the base of fundamental knowledge with the hope that widening of this knowledge base will lead ultimately to benefits for society. [2]

It is now necessary to provide a theoretical foundation for information processes which would enable the transformation of information science from a vague, or nebulous, field into a "true" scientific discipline. Such a transformation is important for reasons beyond mere academic status. In the absence of any scientific framework, a great deal of the application of information concepts to human purposes is taking place on an adhoc basis. Little, if any, of a common nature exists to provide guidance in systems design and evaluation. If information engineering is ever to become a "true" engineering field, then information science must first become a "true" science by establishing a framework of theory, law, and experiment such as exists in physics or psychology.

This view, that information science *is*, or at least *should be*, information *science*, may be summarized by the epigram (9)

(9) I.S. IS I.S.

This may be conveniently called the $(IS)^3$, or "IS-cubed" viewpoint. Formula (9) is not as trivial nor as silly as it looks. It marks the fact that we have a technical language in which we distinguish science from engineering and technology, and an informal language in which we talk about the business of our daily lives and that we have identified 'information science' from our informal language with 'information science' in our technical language instead of with 'information engineering', or 'information technology'.

III. The Specific Nature of Information Science

Just as we must guard against taking (9) too lightly, we must also not read more into it than is there. It tells us that information science is a science but it says nothing about what kind of science that science may be. This is the question that will be taken up in this section, and is represented by formula (8). It is this question that formulas (2), (3), and (4) are purported answers for. But once we have proposed an answer to (7) we are not free to consider (2), (3), and (4) equally. The variable in (6) now ranges over the domain of sciences while cybernetics and general systems theory are specific engineering and/or technological disciplines. Our problem then, is to answer (8) by picking a value for the variable x in (6) from the domain of sciences. Thus (2) is still a possible answer since semiotics is a science — a study of the regularities of semiotic nature for "the pure joy of learning and understanding and not for any ulterior practical purpose", [5]. But other possible answers exist for there are many sciences: physics, chemistry, geology, pyschology, sociology, and esthetics, for example.

Some of these answers can be ruled out as obviously unsuitable, others are less easy. For instance, I believe no one would dispute that information science is not a branch of chemistry even though some problems in chemistry require techniques similar to some techniques used in information science. On the other hand there are a great many similarities in both problems and techniques between information science and psychology. One very large group of sciences are classified as the physical sciences since physics serves as the paradigm for these sciences. But information technology is not an application of physics as is mechanical engineering or even astronautical engineering. Physics has a backlog of discoveries upon which these technologies can draw for their development. For information technology we not only do not have such a background of knowledge, we may not even know what the proper science is.

However, our preliminary studies in this area lead us to believe that the science underlying information technology is semiotics, the science of signs and sign processes.

Research in information science has been conducted on how accurately the symbols of communication can be transmitted. This is the province of communication theory. A theory of information transfer would focus on how accurately the transmitted symbols convey the desired meaning and how effectively the received meaning affects behavior in the desired way. Studies of signs are necessary in order to enable a formal explication of information and its general relation to measurable properties of signs and sign processes. This knowledge will facilitate the analysis of such functions as the generation, transmission, and storage of complex information.

One of the participants on this panel has expressed the opinion that Information Science concerns, among other things, the addressing of "notions" in cognitive memory. And since notions are signs which represent concepts, this involves determining the structure of signs and their addressing capability. In fact this is closely related to the memory coding problem which I will elaborate on later.

Some of the most essential elements of human information processing cluster around the mechanisms of remembering and recalling signs, and not just symbols either, but icons and indexes as well. Most of our knowledge of information storage and retrieval in the human is formulated in the language of neurophysiology, molecular biology, and psychology. Most of our knowledge of other information processes in the human is formulated in the languages of linguistics, logic, philosophy, psychology, and psychoanalysis. Semiotics is the mother science that unifies and communicates between these sciences just as physics does for the physical sciences. In my opinion, then, we should accept (2) as our answer to (8), giving (10):

(10) The science of information science is semiotics.

IV. Fundamental Questions in the Foundation of Information Science

In all of our studies —— in information measures; in human processing of information; in linguistic studies of natural language; in artificial communication systems —— the research eventually boils down to the question, "What is a sign, and what kind of structure does it have?" This question of structure appears to be the fundamental problem in the foundations of information science.

A sign is something that can stand for something else for some cognizing body. This is a trinary relation: a sign x stands for an object y to some cognizing body z. Whereas the fundamental relations of the physical sciences are all reducible to binary relations--a point mass x interacting with a point mass y; a field source x attracting a test particle y--the trinary relation of signification is *essentially* trinary: it is not reducible to any product of binary relations.

This is the minimum structure a sign must have, but particular classes of signs may have additional structure. The properties of these signs then depend on the kind of sign and the kind of additional structure they posses. It is these properties and the kinds of additional structure they possess that determine what role the signs play in information processes, that is, in semiotic interactions.

The School of Information Science at the Georgia Institute of Technology is studying the structure of various kinds of signs and attempting to determine the relationship between sign structure and information properties. The purpose of this research is to gain a fundamental understanding of the role that semiosis plays in information processes, [3].

The results of this research will begin establishing a firmer basis upon which to found new advances in the semiotic-related technologies, especially in information science. For instance, the creation, destruction, transmission, and storage of information require different sign structures. A knowledge of the internal structure of these signs will help facilitate analysis of such processes. As another example of the possible applications of the results of this research, Loveland's complexity measure (also called 'algorithmic information') is related to the shape of signs. A knowledge of internal semiotic structure will also help facilitate analysis of the creation, destruction, transmission, and storage of complexity.

Peirce analyzed the logical relationships of what is now called the 'external' structure of signs (or the relationship between signs and a non-sign entities, such as objects of signs and sign interpreters) and thereby arrived at what is regarded by the majority of workers in the field as the most useful classification of signs for most technical purposes. Certainly in our own investigation of sign phenomena, Peirce's classification scheme has proven to yield the most useful results.

Altho Peirce developed several successive classification schemes, each involving the external structure of signs, and each involving successively more refined analysis, the best-known scheme and the one that has to date proven the most useful is also the least complex, involving a three-way analysis into ten sign categories. (One other scheme of his, involving a ten-way analysis into sixty-six sign categories, has received some attention in the literature [6] but has not been extensively developed. The categories of this latter scheme do not conflict with the simpler scheme, being only refinements of its categories).

Signs may be classified according to their three modes of existence into tones, tokens, and types; according to their three modes of reference into icons, indexes, and symbols; and according to their three modes of interpretation into rhemes, phemes, and dolemes. This could yield $3^3 = 27$ sign categories, except that certain combinations, such as indexical dolemes, are impossible, so that only ten possible combinations result.

Table 1 defines the nine sign types. The ten categories that this classification scheme results in are shown in Figure 1.

As examples of the utility of this scheme for information science we may point to Brillouin's distinction between absolute and distributed information. [1, p265f], which corresponds to Peirce's distinctions between types and tokens. The referential classification provides a useful scheme for categorizing computers. Digital computers are symbol processors; simulators are index processors; and analog computers are icon processors. In logic, terms are examples of rhemes; propositions are examples of phemes; and completed arguments are examples of dolemes.

Table 1. External Structure of Signs: Definitions

Tone:	A tone is a sign which has a potential mode of existence embodied in a sheer quality. Example: any quality insofar as it is a sign.
Token:	A token is a sign whose existence is a single actual instance. Example: any actual existent thing or event which is a sign.
Type:	A type is a sign whose existence is in the abstract via a general law. Example: all conventions are types.
Icon:	An icon is a sign which represents its referent via a similar quality. Example: a paint chip.
Index:	An index is a sign which represents its referent via a single causal connection. Example: smoke as a sign of fire.
Symbol:	A symbol is a sign which represents its referent via a general convention. Example: natural language signs.
Rheme:	A rheme is a sign whose interpretant is determined by a qualitative possibility. Example: words or logical terms, etc.
Pheme:	A pheme is a sign whose interpretant is determined by an actual existence. Example: sentences or logical propositions, etc.
Doleme:	A doleme is a sign whose interpretant is determined by a general law. Example: extended discourse or logical argument, etc.

We have found some use for the more refined ten-way analysis of external structure, but a more critical problem appears to involve an entirely different structure than the one Peirce devoted his attention to. Since the problem we face seems to involve the basic structure of the sign itself, as opposed to the relationships between the sign and its object or the sign and its interpretant (the result of interpreting the sign within a cognizing body), we have called the structure we are investigating 'internal structure' to distinguish it from Peirce's 'external structure'. The internal structure of signs appears to be related to the problem of how signs bear meaning and what meaning is; the measurable properties of signs and their relation to information measures; the manner in which signs can be combined to build additional structure and form messages, etc. Each of these problems has a critical bearing on one or more branches of information science and technology.

Our research has concentrated on an investigation of internal structure, its form and properties, its relation to external structure, its relations to information measures of all kinds, and its relation to the nature of the meaning of signs [4]. We have been able to develop a Universal Structure Model which explains the structure of each kind of sign and the relationships between the various kinds of signs. See figure 2.

Our analysis of various information measures that have been proposed in the literature show that every measure that we have analyzed so far has been a measurable property of one of the sign components shown on this model. Therefore, one task of future research is to make a systematic determination of the measurable properties of all the sign components and the explore the relationships among them.

As an example of how this model can be used to determine research strategies we may look at the problem of cognitive memory. The term 'cognitive' marks a distinction between the objective, and knowable, external (cognitive) world and the subjective, and experiencable, internal (emotive) world. Since the cognitive world represents the domain of objects of our sign processes, cognition is inherently related to the semantic dimension of semiosis. Cognitive psychologists such as Kintsch, Broadbent, and Garner, have postulated three separate cognitive memories (not necessarily three separate physiological functions or areas of the brain), which they call 'sensory' (span time of a few seconds); 'short-term memory', or 'STM' (span time of a few minutes); and 'long-term memory', or 'LTM' (span time of a few hours). As information scientists, we are interested in how these memory functions are coded. From an analysis of the pertinent psychological experiments it appears that the sensory memory is accessible primarily by indexical coding, STM by iconic coding, and

LTM by symbolic coding. However, no critical experiment has ever been run to determine the degree of this relationship. Is it complete and totally exclusive, or is it just a statistical tendency? Our semiotics lab is attempting a critical experiment in this area. At UCLA-Berkeley, Wallace Chafe is doing much the same kind of thing by investigating the ways that various kinds of knowledge are stored internally. Chafe is using linguistic and sociological techniques; we are using semiotic and psychological techniques.

V. Goals for Basic Research in Information Science

We may briefly summarize this discussion by listing some of the goals that have already been mentioned. Because of the identification that was made in section III, we may identify goals for long-range research in information science with goals for basic researc' in semiotics.

1. Develop a theory of the structure of all categories of signs, sign systems, and sign processes.
2. Investigate the measurable properties of all sign components.
3. Explore the basic regularities existing between the measurable properties.
4. Develop theories which explain these regularities.
5. Investigate the relationship between various information processes and semiotic processes.
    a) perception
    b) memory
    c) recall
    d) conception
    e) communication
    f) classification
    g) recognition
    h) decision
    etc.
6. Develop lab instrumentation with which to carry out the critical experiments in the above areas.

This last is of special importance and is crucial to each of the above goals. Because the phenomena under investigation is semiotic and not physical, this requires the invention of entirely new concepts in scientific instruments. It will therefore go hand in hand with goals #2 and 3.

Fig. 1. Peirce's Ten Fundamental Sign Types

Mode of Existence of the Sign

Category	1st	2nd	3rd				
Name	Tone	Token	Type				
Existence of the sign is	Potential	Actual	Abstract	Existence of also required	**Name**	**Category**	**Relation between sign & object is**
	1.	2.	5.	Ø	Icon	1st	Similar Quality
		3. 4.	6. 7.	Object	Index	2nd	Actual Direct Relation
			8. 9. 10.	Interpretant	Symbol	3rd	General Law

Mode of Representation

Mode of Interpretation

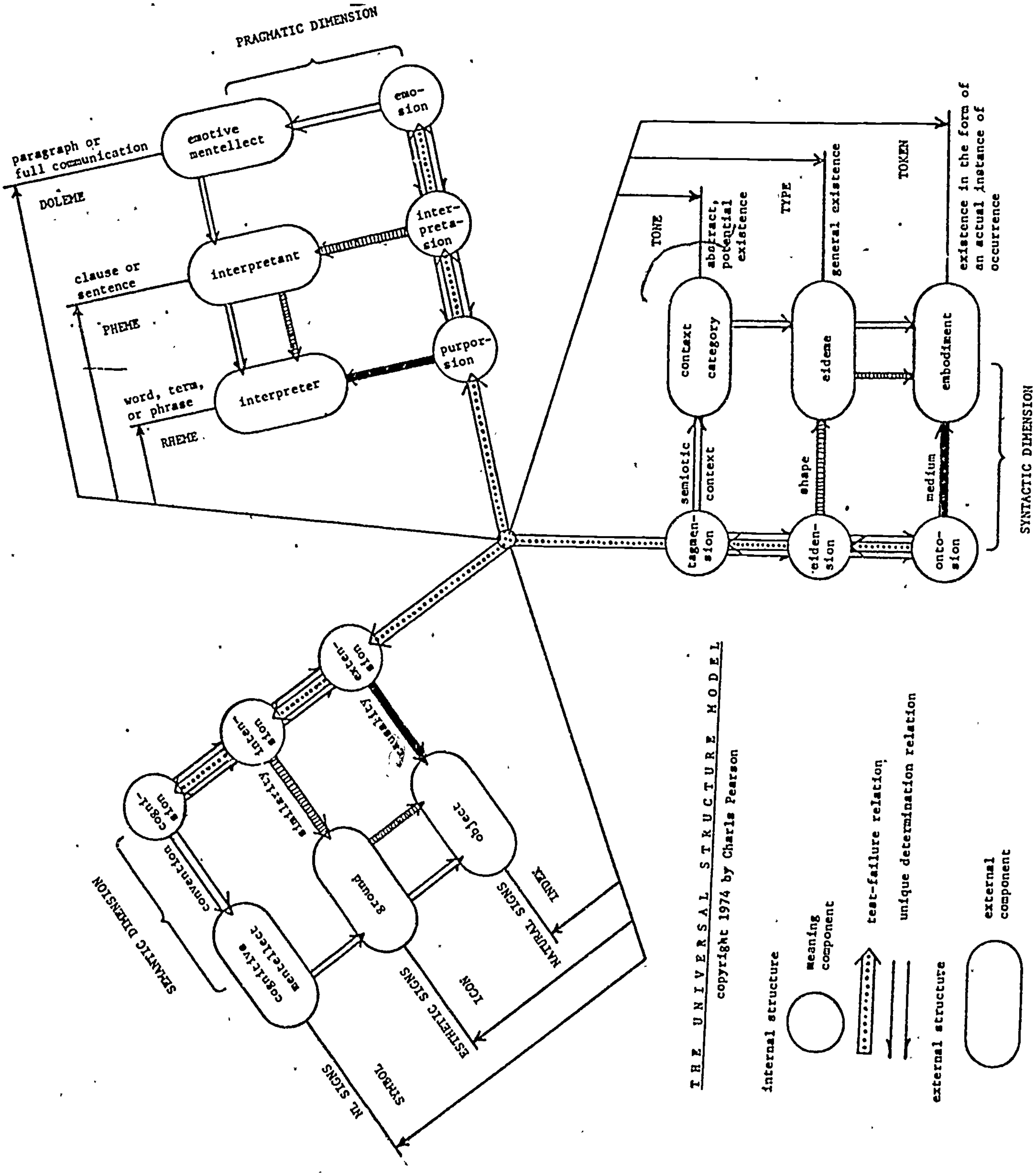
Category	Determination of Interpretant	Name
1st	Qualitative Possibility	Term
2nd	Actual Existence	Proposition
3rd	General Law	Argument

Terminology	
Original →	Final
Qualisign	Tone
Sinsign	Token
Legisign	Type
Likeness	Icon
Sign	Index
Symbol	Symbol
Rheme	Term
Dicisign	Proposition
Argument	Argument
Representamen	Sign

References:

Peirce
Buchler
Goudge
Burks
Alsyon
Morris

Fig. 2



PRAGMATIC DIMENSION
emo-sion
emotive mentellect
paragraph or full communication
DOLEME
inter-preta-sion
interpretant
clause or sentence
PHEME
purpor-sion
interpreter
word, term, or phrase
RHEME
TONE
abstract, potential existence
TYPE
general existence
TOKEN
existence in the form of an actual instance of occurrence
context category
eideme
embodiment
semiotic context
shape
medium
tagmen-sion
eiden-sion
onto-sion
SYNTACTIC DIMENSION
exten-sion
inten-sion
cogni-sion
causality
similarity
convention
object
ground
cognitive mentellect
SEMANTIC DIMENSION
INDEX
NATURAL SIGNS
ICON
ESTHETIC SIGNS
SYMBOL
NL SIGNS
THE UNIVERSAL STRUCTURE MODEL
copyright 1974 by Charls Pearson
internal structure
meaning component
test-failure relation
unique determination relation
external structure
external component

VI. References

1. Brillouin, L. Science and Information Theory. Academic Press, Second Edition, 1962.

2. Debons, A., ed. Challenges to the Development of a Science of Information: Proceedings of the 1972 NATO Advanced Study Institute in Information Science. New York, N.Y., Marcel Dekker, Inc., 1973.

3. Pearson, C.R. "Semiotic Research at Georgia Tech-1974". Introductory remarks at Technical Session 6, 1974 Annual Meeting of American Society of Information Science.

4. Pearson, C.R.; and Slamecka, V. "Semiotic Foundations of Information Science." Progress Report No. 1; March 1975; School of Information and Computer Science; Georgia Institue of Technology.

5. Peirce, C.S. The Collected Papers of Charles Sanders Peirce. Harvard UP, Vol. I-VIII, 1932-1956.

6. Weiss, P. and Burks, A.W. "Peirce's Sixty-Six Signs", Journal of Philosophy, 42 (1945), pp 383-388.