G-36-611

final

SEMIOTIC FOUNDATIONS OF INFORMATION SCIENCE

CHARLS PEARSON

VLADIMIR SLAMECKA

JANUARY 1977

SCHOOL OF INFORMATION AND COMPUTER SCIENCE

GEORGIA INSTITUTE OF TECHNOLOGY

ATLANTA, GEORGIA 30332

GIT-ICS-77-01

SEMIOTIC FOUNDATIONS OF INFORMATION SCIENCE

Final Project Report

National Science Foundation Grant GN-40952

CHARLS PEARSON
VLADIMIR SLAMECKA

Principal Investigators

January 1977

School of Information and Computer Science
Georgia Institute of Technology
Atlanta, Georgia 30332

ABSTRACT

A new theory of sign structure is proposed which explains the syntactic, semantic, and pragmatic classification of signs due to C. Peirce. The theory comprises, in part, a language capable of relating studies of information processes across a range of disciplines, including communication science, psychology, computer science, and linguistics. The power and utility of the theory and the language are illustrated by explicating empirically such syntactic and semantic processes as perception, syntactic communication, and memory coding. The report indicates the relevance of the theory to selected applied problems of information engineering. The development and activities of the SemLab, a semiotic research laboratory dedicated to empirical investigations of information phenomena, are described. The effort reported is part of a basic research program in information science performed at the School of Information and Computer Science, Georgia Institute of Technology. This final report on National Science Foundation Grant GN-40952 covers work performed between January 1974 and December 1976.

ACKNOWLEDGEMENTS

Several members of the faculty and staff of the School of Information and Computer Science have participated in the work described in this report. We are pleased to acknowledge the contributions of Pranas Zunde, professor; Philip J. Siegmann, associate professor; Albert N. Badre, assistance professor; and of the following graduate research assistants: Dan Canyas, Francis Coyle, Jeff Flowers, Irene Gan, James D. Ray, Jr., and Keum Shin.

This material is based upon research supported by the National Science Foundation under Grant No. GN-4092. This support, and particularly the cooperation and assistance of Dr. Edward C. Weiss and his staff, are sincerely appreciated.

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CHARLS PEARSON
VLADIMIR SLAMECKA
Principal Investigators

Atlanta, Georgia
January 1977

TABLE OF CONTENTS

TABLE OF FIGURES

I. INTRODUCTION

The research reported here is the beginning of a long-term program of fundamental investigations in quantitative semiotics, a field which we believe lies at the foundation of information science *qua* science [51]. The program has the following major goals:

1. Development of a theory of the structure of signs, sign systems, and sign processes.
2. Investigations of the measurable properties of all sign components.
3. Investigations of the basic regularities existing between the measurable properties of all sign components.
4. Development of theories which explain these regularities.
5. Study of the relationship between various information and semiotic processes, such as perception, memory, recall, conception, communication, classification, recognition, decision, and others.

The principal motivation and goal of this research program is the study and elucidation of information processes. Within the framework of this program, the effort described in this report has had as its objective the further development of the crucial concept of sign structure. We have focused on an investigation of signs because in our experience all fundamental questions pertaining to information processes invariably boil down to the problem of understanding the nature and the structure of signs.

The report summarizes the results of research and related activities carried out during the period of January 1974 through December 1976. It is organized into three narrative sections. Chapter II proposes a new theory of sign structure. The theory, derived from a general model of sign structure, explicates the relationship between the structure of signs and their classification. Chapter III exemplifies the utility of the theory and the model of sign structure for interpreting a range of basic information processes, and it broaches the question of the utility of this research for technological problems. Chapter IV describes eclectically the facilities and activities of the Semiotics Laboratory of the School of Information and Computer Science (SemLab). The SemLab is a substantial as well as unique resource supporting education and research in this area.

* * *

It is perhaps appropriate that we explain, at the outset of this report, the relationship of this work to the field called information science.

In its currently popular trend, information science shows an almost complete preoccupation with technological problems and products. The fact that its applied research has been unable to attain many of the more important objectives enthusiastically predicted for it twenty years ago is attributed today largely to the absence in the information field of a core of basic, or scientific, results--such as were available, for instance, from physics for aeronautical engineering. This notwithstanding, a corollary of the technological preoccupation of present-day information science is a deep impatience with all efforts which do not immediately affect information technology.

It may behoove us at this point to consider a historical analogy. Today we credit physics with a major contribution to many of the triumphs of modern technology, such as the uses of atomic energy, or space travel and communications. Yet the basic scientific discoveries which underlie these technological accomplishments occurred largely during the 16th through the 18th centuries: the language revolution (Copernicus), the empirical revolution (Galileo), and the theory revolution (Newton). These efforts built the foundation for physics as a science, and they provided for its subsequent development culminating in the engineering accomplishments of our time.

If the reader will tolerate this analogy, then in its terms the current level of development of the science of information is somewhere at the level of physics of the 18th century. We view our own work as belonging to that level of development: the theory of sign structure proposed in Chapter II constitutes, in part, a new language devised for studying information phenomena; and the work described in Chapter III illustrates and demonstrates the utility of this language for explicating, empirically, a number of these phenomena. Our research thus aspires to be an early contribution toward the establishment of a science of information.

The study of signs and sign processes is, of course, not proprietary to any one field of institutionalized science. Indeed, our own research is related to ongoing work in an array of such fields: computer science, communication science, psychology, and linguistics. As the common denominator of these efforts is the study of information processes, the notion of an "information science" as the envelope for these studies is appropriate. Whether or not such a basic science will become institutionalized depends very much on the existence of one or more paradigms relating and unifying its efforts. We would like to believe that our work, incipient as it is, demonstrates such a portent. For this reason we view our research firmly as lying in information science.

II. A THEORY OF SIGN STRUCTURE

Traditionally, major advances in systematic science have been made by quantification and measurement. In information science, the need for better understanding of the concepts of information measures and measurement is well recognized. Our approach to the study of information measures and measurement is from the viewpoint of semiotics, the study of signs and sign processes. The role that signs play in information processes (that is, in semiotic interactions) is determined by the properties of the sign; in turn, sign properties are determined by the kind of sign and its structure. From this viewpoint, we regard an information measure as any observable property of the sign structure; and the measurement of information as the development of a measurement system for carrying out the observation of that property.

Our purpose in developing a theory of sign structure is to have a tool for explicating the nature of information measurement and its relationship to semiotic processes, and for classifying information measures according to their semiotic dimensionality and interrelationships. A theory of sign structure useful for these purposes has evolved gradually over the past three years. It is called the Universal Sign Structure Model.

A. Peirce's Taxonomy of Signs

Throughout our investigations we have had occasion to use several different taxonomies, or classification schemes, for signs. Of these only the classification by Charles Peirce [45] has proved to be satisfactory in every empirical setting for which a classification was wanted. We therefore ascribe the Peircean schmem an empirical reality, and would like our theory of sign structure to explain the applicability and usefulness of the Peircean scheme in terms of the structure of the sign.

Peirce defines the sign as a three-place relation:

> A sign, or *representamen*, is something which stands to somebody for something in some respect or capacity [45, 2.228].
>
> In consequence of every representamen being thus connected with three things, . . . the science of semiotics has three branches [45, 2.229].

Peirce called these three branches "pure grammar", "logic proper", and "pure rhetoric". Subsequently, Charles Morris called these the three 'dimensions' of semiotics and gave them their accepted names: syntactics, semantics, and pragmatics.

Peirce's taxonomy has three classification schemes, leading to nine categories of signs. Definitions 1-3 pertain to a syntactic classification; definitions 4-6 to a semantic classification; and definitions 7-9 to a pragmatic classification of signs.

Definition 1: A sign which exists as an abstract quality both in itself and in its relation to other signs is called a 'TONE'*.

Definition 2: A sign which exists as a general kind, both in itself and distinguishable from other signs is called a 'TYPE'.

Definition 3: A sign which exists as an actual, single, physically existing individual is called a 'TOKEN'.

Definition 4: A sign which is related to its object by an actual, single, existential, cause and effect relation is called an 'INDEX'.

Definition 5: A sign which is related to its object by a concrete similarity between the shape of the sign and its object is called an 'ICON'.

Definition 6: A sign which is related to its object by an arbitrary convention, agreement, or general law, is called a 'SYMBOL'.

Definition 7; A sign whose interpretant represents it as a sign of possible reference to its interpreter is called a 'RHEME'.

Definition 8: A sign whose interpretant represents it as a sign of fact or actual reference to its interpreter is called a 'PHEME'.

Definition 9: A sign whose interpretant represents it as a sign of reason to its interpreter is called a 'DOLEME'**.

Because of the rather opaque nature of several of these definitions it may be well to give some examples. An example of a tone in linguistics would be a nonterminal node of a phrase structure diagram, a context category, or a set of allowable (including obligatory) transformations on a sign (word, sentence, or discourse). An example of a tone in logic would be a functional combinator, i.e. a categorical analysis of a sign. An

*It must be remembered that Peirce employed a great number of different and differing nomenclatures. The one adopted here was used in [29].
**Peirce's actual term was 'deloam' from the Greek δελωμ.

example of a type in linguistics would be a terminal node of a phrase structure diagram or a lexical item (word, sentence, or discourse) at the morphological level, before the phonetic transformations have been applied. An example of a type in logic would be a well formed expression (term, formula, argument). An example of a type in statistical linguistics would be a general sign of which a particular occurrence token is a specific instance. Classical linguistics and classical logic do not concern themselves with the study of tokens. An example of a token in statistical linguistics would be the single, particular occurrence of some sign that actually occurs at a specific point in the computer scan of a machine readable text. An example of a token in psycholinguistics is one actual stimulus that is exposed in a teescope.

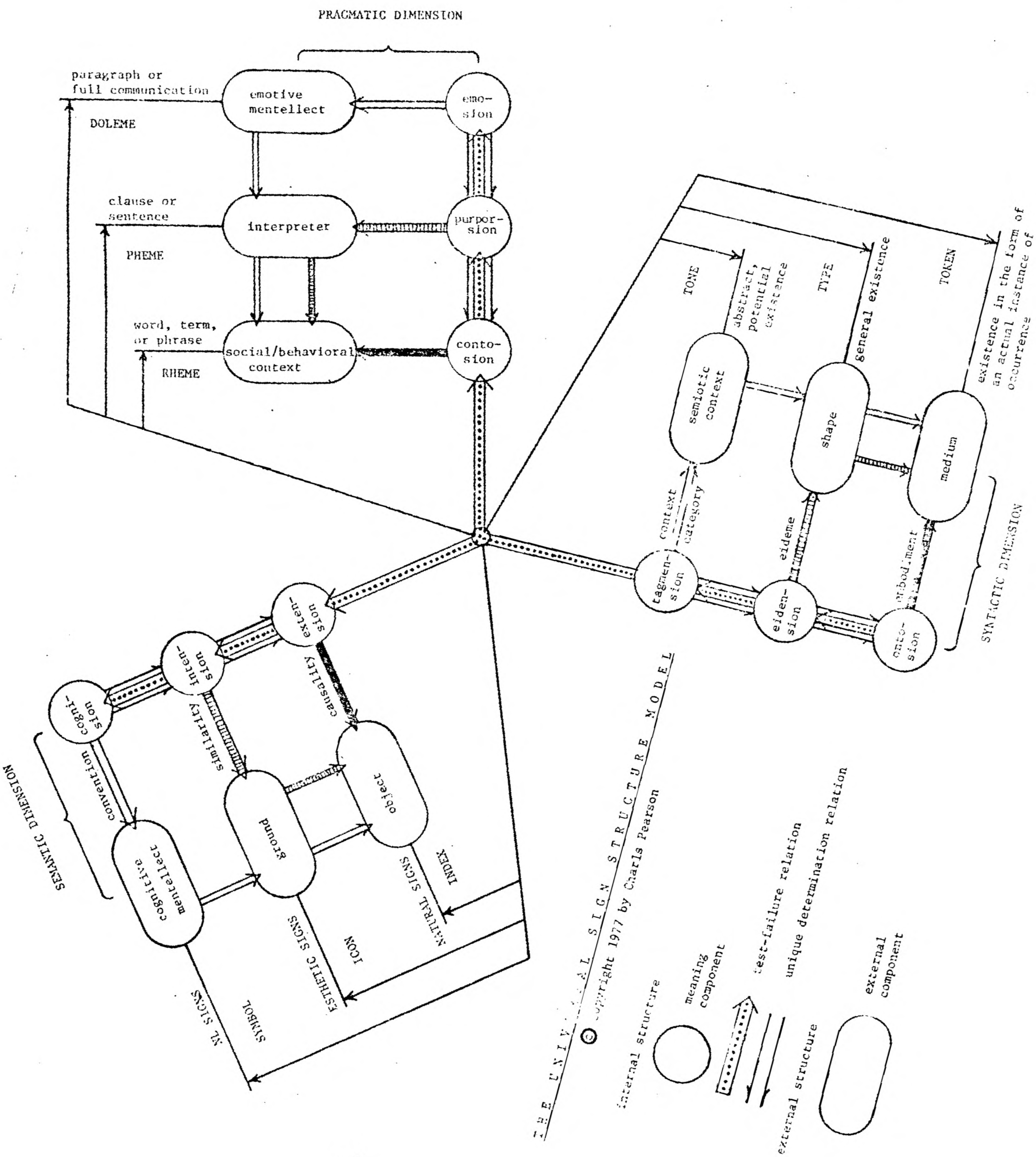
An example of an index in cognitive psychology is Bruner's 'enactive response'. An example from ordinary life would be a pillar of smoke in a dry forest taken by a ranger as a sign for fire, or a knock on a closed door taken by someone on the inside as a sign that someone or something was present on the outside. An example of an icon from cognitive psychology is Bruner's 'ikon'. An example from ordinary life is a paint chip that denotes paint in a can, of the same color as the chip, or a rhythmically repeated note in a melody that holds the music together by the similarities that it establishes. An example of a symbol from cognitive psychology is Bruner's 'symbol'. Natural language signs are all symbolic, including those called 'indexical' and those called 'onomatopoetic'.

An example of a rheme in logic would be a term; an example of a rheme from natural language would be a word or a phrase. An example of a pheme from logic is a statement; from natural language a clause or sentence. An example of a doleme from logic is an argument; from natural language a paragraph or a complete communication.

B. A Universal Sign Structure Model

The proposed theory of sign structure is embodied in the Universal Sign Structure Model shown in Figure 1. In order to show how this model explains the Peircean taxonomy, we first state the following three principles of the theory.

Fig. 1. The Universal Sign Structure Model



The Representation Principle: A sign must consist of a trinary relation, and it must represent. A sign, therefore, consists of three parts: A syntactic structure, a semantic structure, and a pragmatic structure.

The Principle of Internal/External Balance: The internal and the external structure of a sign must be balanced, consisting of exactly one internal component for each external component and vice versa. The internal components are called components of meaning.

The Principle of Additional Structure: Whenever a sign has more than the minimum structure, the additional structure is built up from the center out (as per Figure 1), and for each dimension independently.

Example. From Figure 1 we isolate the minimum structure (Figure 2) which we shall later find is the structure of the indexical rhematic tone. If we want to add to it one layer of semantic structure, we derive (according to the Principle of Additional Structure) the structure of the iconic rhematic tone (Figure 3).

Using the universal sign structure diagram of Figure 1 and these three principles we can now explain the Peircean Taxonomy of signs by means of nine representation theorems. ('Representation' is used here in its mathematical rather than its semiotic sense.) Certain rules of interpretation or translation between the theoretical vocabulary and the observational (or less theoretical) vocabulary will become apparent as we proceed with the proofs of these theorems. The rules of interpretation are obvious, and they form an integral part of the theory. The nine representation theorems and their proofs are as follows.

Theorem 1: A sign is a tone iff it has exactly one level of syntactic structure. It therefore has one component of syntactic meaning (tagmension) and one external syntactic component (the semiotic context).

Proof: By the Representation Principle and the Principle of Additional Structure any sign must have at least one level of syntactic structure and this must be the innermost, or tagmatic, level. According to the Universal Sign Structure Model (Figure 1), the outermost syntactic level consists of the embodiment of a sign in a physical medium. But if a sign had an embodiment in a physical medium it would exist as an actual, single, physically existing individual and could not exist merely as an abstract

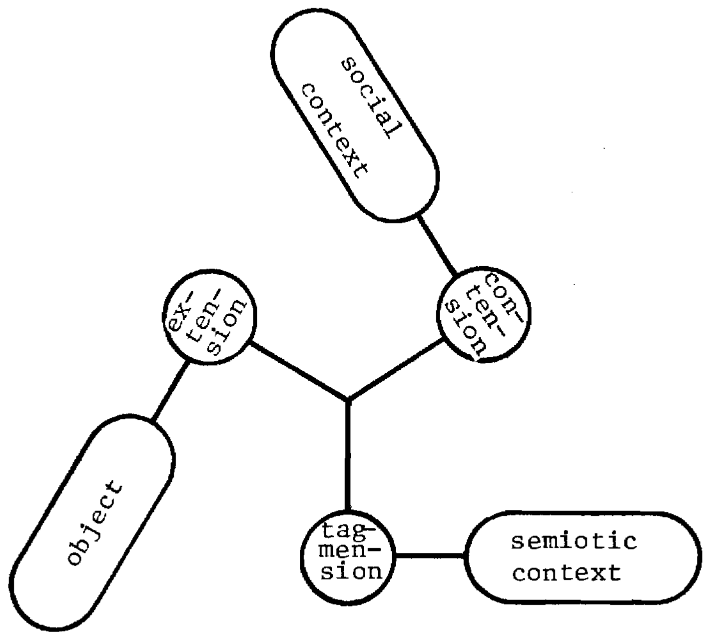


Fig. 2. The Minimum Semiotic Structure

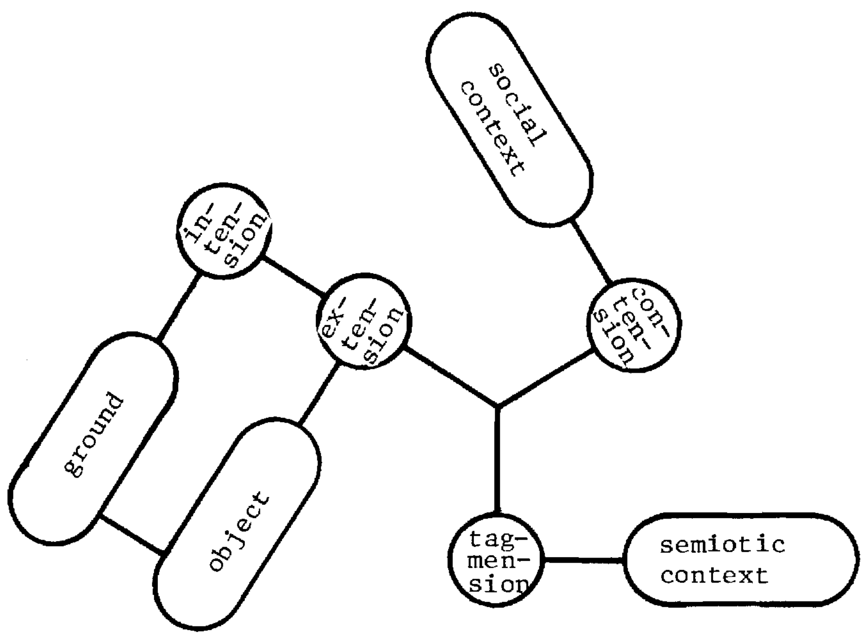


Fig. 3. A Sign With the Minimum Additional
Semantic Structure

quality. It would be a token, not a tone; therefore a tone cannot have an ontotic level of syntactic structure. Also from Figure 1, the second (or middle) syntactic level consists of the distinguishability of a sign by a shape. But, if a sign had a distinctive, distinguishable shape, it would exist as a concrete general, serving as an archtype for all tokens of the same type and could not exist etc. It would be a type, not a tone. Therefore, a tone cannot have an eidontic level of syntactic structure.

Thus a tone has exactly one level of syntactic structure, which is the tagmatic structure. By the Principle of Internal/External Balance, this structure will consist of both an internal component and an external component. From Figure 1 we see that the internal component is tagmension, the meaning component abstracted from the semiotic context, and the external component is the semiotic context itself. *Q E D*

Thus the structure for a tone as given by our theory is shown in Figure 4.

Theorem 2: A sign is a type iff it has exactly two levels of syntactic structure. It therefore has two components of syntactic meaning (tagmension and eidension) and two external syntactic components (the semiotic context and the shape of the sign).

Proof: As in Theorem 1, any sign must have the tagmatic level of structure. However, from Definition 2 we see that an abstract existence as given by the tagmatic structure is not sufficient for a type which must have a concrete general existence and must be distinguishable from other sign types. To enable distinguishability the type must have a shape, which then determines its existence as a general. A type must therefore have the second or eidontic layer of structure in addition to the tagmatic level. On the other hand, if a type also had the third (ontotic) level it would exist as an actual, single, physically existing individual (as argued in Theorem 1) and could not be a general as required by Definition 2. Therefore a type cannot have an ontotic level of syntactic structure.

Thus a type has exactly two levels of syntactic structure, which are the tagmatic and the eidontic structure. By the Principle of Internal/External Balance, this structure consists of two internal components and two external components. From Figure 1 we see that these internal components

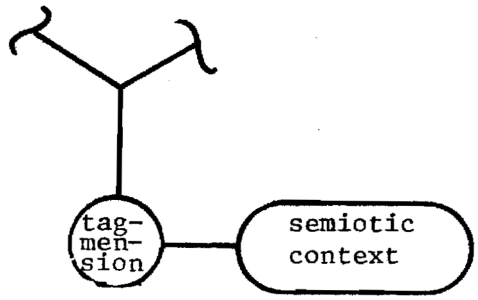


Fig. 4. Tone Structure

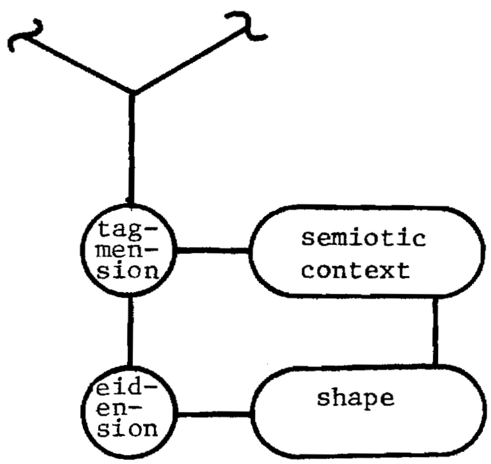


Fig. 5. Type Structure

are tagmension and eidension (the meaning component abstracted from the semiotic shape), and the external components are the semiotic context and the shape itself. *Q E D*

Figure 5 shows the structure for a type, as given by the Sign Structure theory.

Theorem 3: A sign is a token iff it has all three levels of syntactic structure. It therefore has three components of syntactic meaning (tagmension, eidension, and ontosion) and three external syntactic components (the semiotic context, the shape of the sign, and the medium in which it is embodied).

Proof: In order to have the actual, single, physical existence as an individual required by Definition 3, a token must be embodied in some physical medium. Figure 1 gives this as the third level of syntactic structure; and by the Principle of Additional Structure the token must therefore have all three levels of syntactic structure. By the Principle of Internal/External Balance, this structure consists of three internal components and three external components. It follows that these internal components are tagmension, eidension, and ontosion (the meaning component abstracted from the semiotic medium), and the external components are the semiotic context, the shape, and the medium itself. *Q E D*

The proof of Theorem 3 follows very simply from the calculus of the theory, but because of this terseness it leaves something to be desired of our understanding. This can be supplied by motivating the need for all three levels of syntactic structure in the token. In order to exist as a token, a sign must be embodied in some medium, but it cannot be so embodied in the medium without supplying a shape to the medium as well. It is this shape that is used to detect and distinguish the existence as as instance of this type rather than some other type. This determines the presence of the eidontic structure in the token. The necessity of the tagmatic structure is more subtle. We can very well imagine a message seeming to consist of a single sign, for instance a cross standing beside a rural road. But such messages do not actually consist of a single sign and never appear in isolation from some semiotic system which determines their contextual relations. (The cross appears beside the road where we have come to expect such signs, contrasted with situations along interstate highways in which such signs don't occur. These are contextual relations; they form part of the semiotic context for interpreting the cross as a sign, and determine the presence of the tagmatic structure in the token.) Thus the result of Theorem 3 is well motivated both intuitively and empirically.

Figure 6 gives the structure for a token determined by Theorem 3.

Theorem 4: A sign is an index iff it has exactly one level of semantic structure. It therefore has one component of semantic meaning (extension) and one external semantic component (the object of the sign).

Proof: By the Representation Principle and the Principle of Additional Structure any sign must have at least one level of semantic structure and this must be the innermost, or deictic, level. From the universal sign structure diagram (Figure 1), the outermost semantic level consists of the arbitrary, but conventionalized, cognitive interpretation of the sign. By Definition 4, however, an index determines the object by a single cause-and-effect relation actually existing between the sign and the object. Therefore we are not free to form arbitrary conventions as to how we shall interpret an index, and hence an index cannot have a noetic level of semantic structure. Also by Figure 1, the second, or middle, semantic level consists of an interpretation of a sign as determining its object via a similarity between properties in the object and properties in the shape of the sign. Since a cause need not bear any sensible similarity to its effect and vice versa, we are not free to interpret indexes via similarities, and hence an index cannot have a hypotic level of semantic structure.

Thus an index has exactly one level of semantic structure, which is the deictic structure. By the Principle of Internal/External Balance, this structure will consist of both an internal and an external component. From the universal structure diagram we see that the internal component is extension (the meaning component abstracted from the object), while the external component is the object itself. *Q E D*

Theorem 4 gives us the structure for indexes shown in Figure 7.

Theorem 5: A sign is an icon iff it has exactly two levels of semantic structure. It therefore has two components of semantic meaning (extension and intension), and two external semantic components (the object of the sign and its ground).

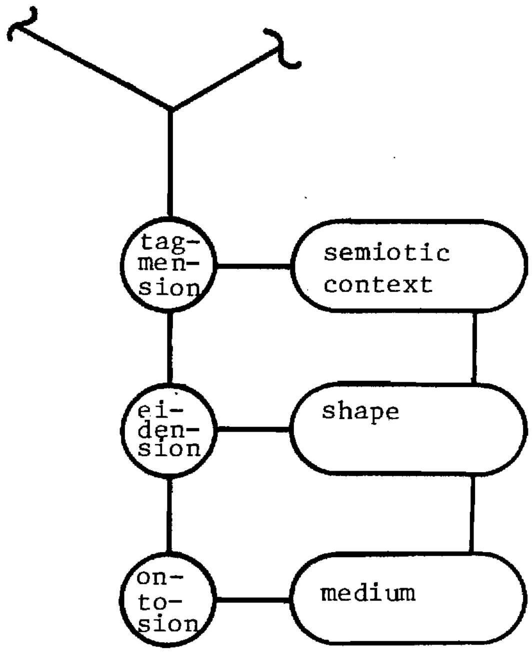


Fig. 6. Token Structure

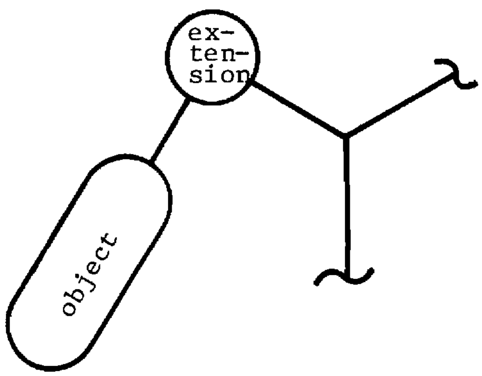


Fig. 7. Indexical Structure

Proof: As in Theorem 4, an icon must have a deictic level of structure which by the Principle of Internal/External Balance consists of the extension and object of the sign. But by Definition 5, the icon's object is determined by a similarity. Since no similarity is encompased by a deictic relation (cause-and-effect, pointing, or otherwise), no determination would be made if the deictic level were the only semantic structure present in the icon. Therefore icons possess an additional level of semantic structure. By the Principle of Additional Structure this must include at least the hypotic level. This can also be justified intuitively since in order to represent its object by a likeness, there must be a set of properties in the object by which this likeness is determined. These properties constitute the ground of the sign which is the hypotic component of external structure. By the Principle of Internal/External Balance, there must also be an internal hypotic component, which is the intension of the icon.

As in the proof of Theorem 4, we are not free to form arbitrary conventions as to how we shall interpret an icon. We must use those properties for judging a similarity which is actually present in the shape of the sign. Therefore an icon cannot have a noetic level of semantic structure.

Thus an icon has exactly two levels of semantic structure, which are the deictic and hypotic levels. By the Principle of Internal/External Balance each level of this structure will consist of both an internal component and an external component. From the universal sign structure diagram (Figure 1) we see that the internal components are the extension and intension (the meaning component abstracted from the ground of the sign) and the external components are the object and the groud itself. *Q E D*

Figure 8 gives us the structure for icons determined from Theorem 5.

Theorem 6: A sign is a symbol iff it has all three levels of semantic structure. It therefore has three components of semantic meaning (extension, intension, and cognesion), and three external semantic components (the object, the ground, and the cognitive mentellect of the sign).

Proof: In order to determine its object according to an arbitrary convention, agreement, or general law, as required for a symbol by Definition 6, a symbol must be interpreted via a cognitive mentellect

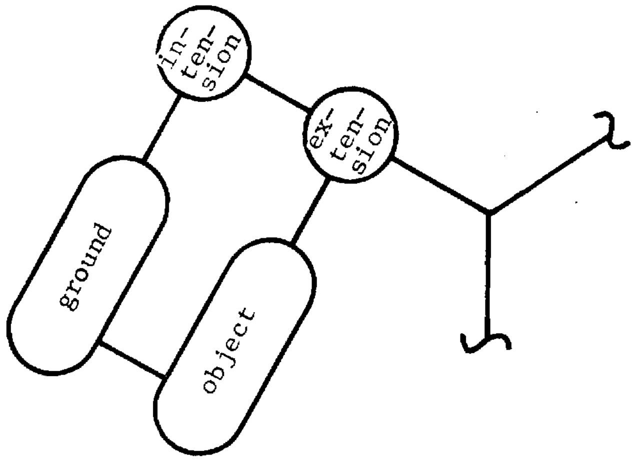


Fig. 8. Iconic Structure

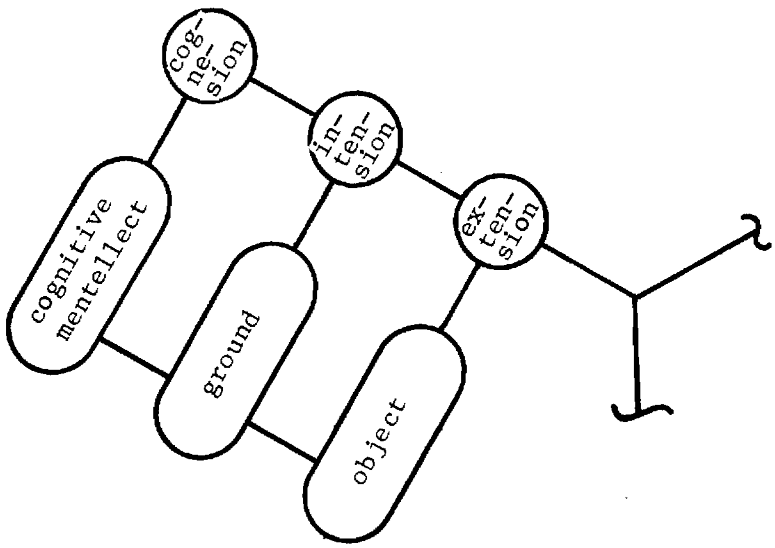


Fig. 9. Symbolic Structure

(which the universal sign structure diagram (Figure 1) gives as the third level of semantic structure). By the Principle of Additional Structure the symbol must therefore have all three levels of semantic structure. By the Principle of Internal/External Balance, this structure consists of three internal components and three external components. It follows from the universal sign structure diagram (Figure 1) that these internal components are extension, intension, and cognision (the meaning component abstracted from the cognitive mentellect), and the external components are the object, the ground, and the cognitive mentellect itself (Figure 9). *Q E D*

Theorem 7: A sign is a rheme iff it has exactly one level of pragmatic structure. It therefore has one component of pragmatic meaning (contension) and one external pragmatic component (the social/behavioral context of the sign).

Proof: By the Representation Principle and the Principle of Additional Structure any sign must have at least one level of pragmatic structure and this must be the innermost, or contotic, level. From the universal sign structure diagram (Figure 1) we see that the second level of pragmatic structure sets up an actual relation of fact between the sign and the interpreter, and therefore represents the kind of relation defined for a pheme (and not for a rheme which must express to the interpreter only a possible reference). Therefore, a rheme cannot have a second level of pragmatic structure at the purportic level, and (by the Principle of Additional Structure) it also cannot have a second level of pragmatic structure at the emotic level. Thus a rheme has exactly one level of pragmatic structure, which is the contotic structure. By the Principle of Internal/External Balance, this structure will consist of both an internal component and an external component. From the universal sign structure diagram (Figure 1) we see that the internal component is contension (the meaning component abstracted from the social/behavioral context), and the external component is the social/behavioral context itself. *Q E D*

Thus, Theorem 7 gives us the structure for rhemes shown in Figure 10.

Theorem 8: A sign is a pheme iff it has exactly two levels of pragmatic structure. It therefore has two components of pragmatic meaning (contension and purposion) and two external pragmatic components (the social/behavioral context of the sign, and its interpreter).

Proof: As in Theorem 7, a pheme must have a contotic level of structure which by the Principle of Internal/External Balance consists of the contension and social/behavioral context of the sign. But by Definition 8, a pheme must express an actual relation of fact between the sign and interpreter. This cannot be done by the contotic structure which express only a possible relation of reference between the sign and interpreter. Therefore phemes possess an additional level of pragmatic structure. By the Principle of Additional Structure this must include at least the purportic level. This can also be justified intuitively: we saw from the universal sign structure diagram that the second level of pragmatic structure does set up an actual relation of fact between the sign and the interpreter. The interpreter, in fact, is the external component of the purportic level of pragmatic structure. By the Principle of Internal/External Balance then, there must also be an internal purportic component which is the purporsion of the pheme. If an emotive mentellect were added to the pragmatic structure of the pheme, its interpretant would express a relation of reason between the sign and the interpreter, or the pheme would be expressed as a sign of reason to the interpreter, not as an actual relation of fact between the sign and interpreter as is required by Definition 8. Therefore a pheme cannot have an emotic level of pragmatic structure.

Thus a pheme has exactly two levels of pragmatic structure, which are the contotic and purportic levels. By the Principle of Internal/External Balance each level of this structure will consist of both an internal component and an external component. From the universal sign structure diagram (Figure 1) we see that the internal components are the contension and purporsion (the meaning component abstracted from the interpreter of the sign), and the external components are the social/behavioral context and the interpreter itself. *Q E D*

Theorem 8 gives us the structure for phemes shown in Figure 11.

Theorem 9: A sign is a doleme iff it has all three levels of pragmatic structure. It therefore has three internal pragmatic components (contension, purporsion, and emosion), and three external

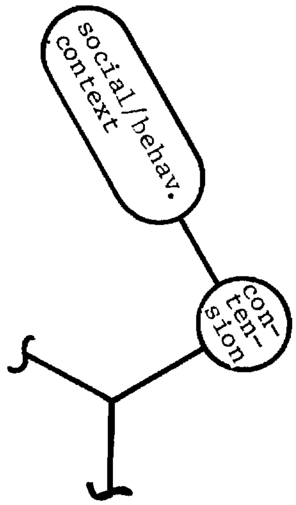


Fig. 10. Rhematic structure

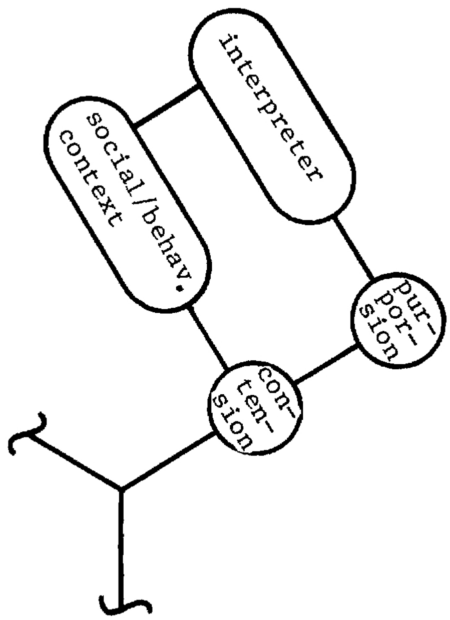


Fig. 11. Phematic Structure

pragmatic components (the social/behavioral context, the interpreter, and the emotive mentellect of the sign).

Proof: In order for a sign's interpretant to represent it as a sign of reason to its interpreter, as required for a doleme by Definition 9, a doleme must be expressed by an emotive mentellect, which the universal sign structure diagram (Figure 1) gives this as the third level of pragmatic structure; and by the Principle of Additional Structure the doleme must therefore have all three levels of pragmatic structure. By the Principle of Internal/External Balance, this structure consists of three internal components and three external components. It follows from the universal sign structure diagram that these internal components are contension, purporsion, and emosion (the meaning component abstracted from the emotive mentellect), and the external components are the social/behavioral context, the interpreter, and the emotive mentellect itself. *Q E D*

Theorem 9 yields the structure for dolemes shown in Figure 12.

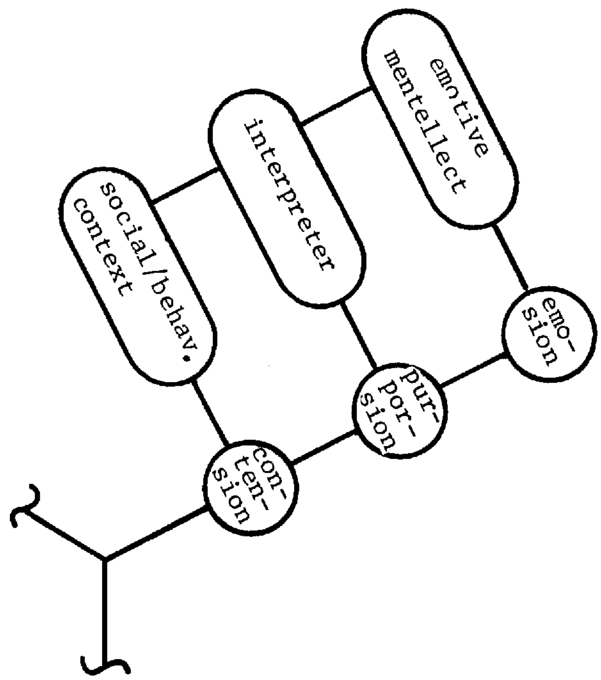


Fig. 12. Dolemic Structure

C. Summary

Theorems 1-9 explain the three trichotomies proposed by Peirce--the syntactic classification (Theorems 1-3), the semantic classification (Theorems 4-6), and the pragmatic classification of signs (Theorems 7-9). The proposed sign structure model is universal in the sense that it displays the structure of all categories of signs.

The theory presented in this section is the outgrowth of the dissertation research of one of the investigators into the structure of the symbolic rheme [29]. In this work the meaning of the sign is identified with its internal structure. Separate reports are in preparation surveying the various senses of the work 'meaning' found in 20th century literature [30], and explicating the distinction between internal and external sign structure and sign components. [31].

The proposed theory is a relatively elementary beginning pertaining to relational phenomena. In the future, information science should develop more refined theories of sign structure, particulary ones capable of predicting quantitative phenomena. To do so, information science research must focus heavily on the fundamental questions of sign structure and sign processing, both from the experimental and the theoretical side. In our opinion, significant progress along these directions may establish information science as a new paradigm for an alternate group of sciences.

III. INVESTIGATIONS INTO SIGN STRUCTURE

This section describes a number of investigations into the structure of signs and information processes, using the language and concepts developed and embodied in the theory outlined above. There is no particular reason underlying our choice of these "projects". In selecting these investigations our motivation has been to test and demonstrate the utility of the language and the theory across as broad a range of information processes as possible; at the same time the choice of projects has been affected by the backgrounds and interests of the investigators and their slight propensity toward theoretical questions, rather than toward technological problems of temporary significance. Potential utility of the research in applied research in information technology is broached at the end of this chapter.

One investigation described (into the nature of definition), concerns all three dimensions of semiotic processes--syntactic, semantic, and pragmatic. The remaining studies fall into one each of two categories: syntactic and semantic. So far; we have not pursued studies to advance our understanding of pragmatic structure, although we believe the Universal Sign Structure Model to be very useful and promising in this respect.

A. The Nature of Definition

Definition may be regarded as one of the more important information processes. We believe that our theory of sign structure permits us to systematize all previously proposed concepts of definition.

Many terms associated with definition have appeared in the literature, but apparently there has been no suggestion that these may be related to the various components of meaning in any systematic manner. Thus Robinson lists and analyzes eighteen kinds of definition found in good writers, [46] without attempting to systematize or interrelate them. Plato, Pascal, Locke, Whitehead and Russell, and Wittgenstein all appear intent on explicating certain concepts of definition without interrelating them. We do see efforts at a systematic account of definition in Leibniz and Peirce. In Leibniz clear and distinct definition leads to clear and distinct ideas, while Peirce introduces a third mode of definition that leads to a higher mode of understanding; these three modes of definition were already understood in Scholastic ages, however, by such semioticians as Duns-Scotus.

The menetic analysis of definition, first alluded to by Pearson [29], proceeds from the approach to definition propounded by J.S. Mill:

> A definition is a proposition declaratory of the meaning which it bears in common acceptance, or that which the speaker or writer, for the particular proposes of his discourse, intends to annex to it. [19].

Our theory of sign structure, which identifies meaning with the internal structure of signs and postulates nine meaning components, permits us to modify Mill's concept by introducing the concepts of 'elementary' and 'complete' definition, as follows:

> An ELEMENTARY DEFINITION is one which states one component of the meaning of a term.
>
> A COMPLETE DEFINITION is one which defines all nine components of meaning of a term, and hence incorporates nine elementary definitions.

The Universal Sign Structure Model predicts nine different kinds of elementary definition. The following are some of the kinds of elementary definition and their equivalents identified in the literature.

> Definition of the shape of a sign (eidontic definition) is called in the literature "definition by abbreviation". It is most often used in mathematics, as when it is declared that a certain newly introduced symbol . . . is to mean the same as a certain other combination of symbols of which the meaning is already know. [54, 2nd Ed., p11].
>
> Definition of the semiotic context of a sign (tagmatic definition) is called "contextual definition" or "definition in context". (Russell's definition of the meaning of a definite descriptive phrase is an example of a definition in context; as such it captures only the tagmension of this meaning.)
>
> Definition of the object of a sign (deictic definition) is called "ostensive definition" [15], "extensive definition" [46], and "denotative definition" [18].
>
> Definition of the ground of a sign (hypotic definition) is called "attributive definition" [46] and "connotative definition" [18].

Definition of the social and behavioral context of a sign (contotic definition) is called "the-method-of-rule-giving" in the sense of voluntary human rules as used by Robinson [46, p129f].

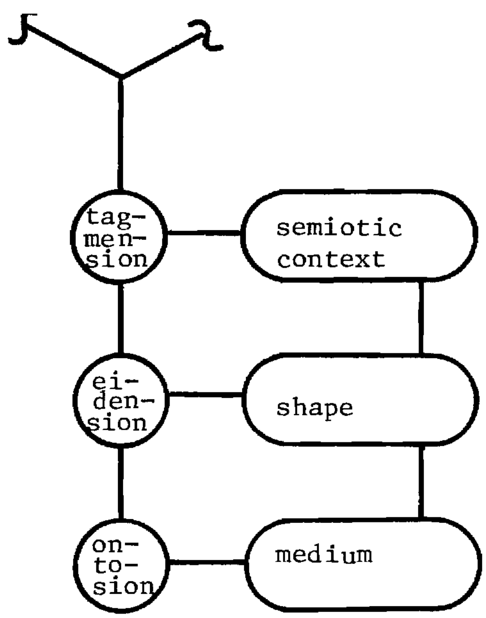
Furthermore, Bridgeman's concept of 'operational definition' comes close to being pure ergotic definition (definition of the purporsion of a term). Robinson give the label 'persuasive definition' to what appears to be emotic definition (the definition of the emosion of a term).

B. Syntactic Communication

The Universal Sign Structure Model predicts three levels of syntactic structure: ontotic, eidontic, and tagmatic. In the syntactics of natural language these levels may be identified with phonetics, morphophonemics, and tagmatics, respectively, although this identification has not been explicated as yet. Instead, early efforts have concentrated on using this prediction to explicate the statistical theory of syntactical communication. The Universal Sign Structure Model appears to offer the most natural explication for this theory.

In communication we use actually existing, embodied signs to carry out actual instances of communication. Communication thus requires the use of sign tokens; this syntactic structure is then our only concern in syntactic communication theory. Therefore according to our Theorem 3, the structure of communication is represented by the following diagram in Figure 3.

Fig. 13.
Structure of Communication



In the standard theory of syntactic communication as introduced by Shannon [48], however, we are not interested in the meaning of the messages communicated; hence, ignoring the internal portion of the above diagram and rotating the external portion, we obtain Figure 14.

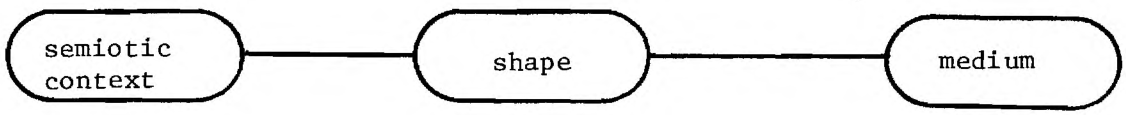


Fig. 14. External Syntactic Structure Rotated

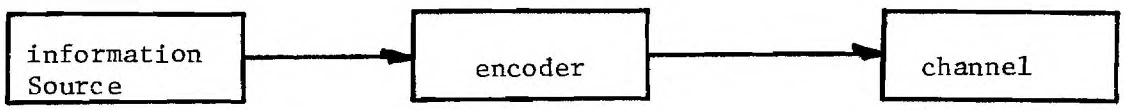


Fig. 15. The Communication Interpretation

We must now interpret this model in the communication setting. In generating or initiating communication we start with the semiotic context, since this is the first, or innermost, level (from the Principle of Additional Structure). Therefore, we first generate the semiotic context of a sign for communication; next, we add a shape to the sign and its context; and finally, we embody the sign in some physical medium so that the communication can actually be carried out. From these we derive Fig. 15; the communication component which generates the context of a sign has been called an 'information source' [1]; the component which adds a shape to a sign and its context is called an 'encoder'; the physical medium embodying the sign is called the 'communication channel'. Taking into account the fact that communication includes both a sender and a receiver we derive the traditional communication model (Figure 16). As usually presented, this diagram includes noise, a physical property of every real physical medium.

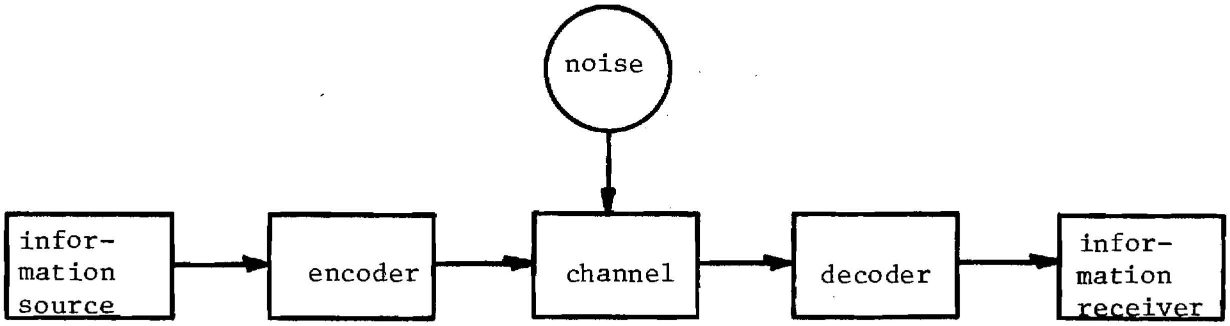


Fig. 16. The Communication Model

In most textbooks the "communication model" is usually presented unmotivated. We are able to derive the communication model rationally from the fact that in the theory of syntactic communication we are interested only in the external syntactic structure of tokens. From our viewpoint such theories of communication as presently exist are seen to be theories of communication physics, not general semiotic theories of communication. We suspect that further advances in communication science will require further development of more general semiotic theories.

The semiotic properties associated with tone, type and token phenomena may be used to understand the communication processes associated with each component. We have incorporated this approach into class notes for a senior level course on communication processes [26]; it makes these processes quite easy to explain. A textbook on this subject is in preparation [44].

C. Perception

According to at least one major school of philosophy, the object of perception is signs. Stated more precisely, signs are the vehicle of perception, and the denotata of signs are the objects of perception.

Perception as a semiotic, or information, process is similar to communication, with two important exceptions. First, we are only interested in receiving signs, not in generating or sending them. Second, we are interested in both the internal and the external structure of signs. In order to be received, signs must actually exist; hence, in perception we are interested in tokens. Furthermore, we are only interested in the syntactic structure. Thus our understanding of the syntactic structure of signs, and particularly our Theorem 3, should be useful for developing psychological theories of perception.

1. A Semiotic Model of Perception

Figure 17 shows a semiotic model of perception, an adaptation of the Universal Sign Structure Model. The major new concept introduced is that of 'menetic valves', which appears useful for integrating many of the isolated findings and theories of particular perceptual phenomena.

From Figure 17 we note that perceptual variables are divided into three categories: ontotic (variables associated with the physical medium), eidontic (variables associated with the semiotic shape of the sign), and tagmatic (variables associated with the tagmatic context). Ontotic variables are equivalent to Stevens' 'psychophysical variables' [52] and Garner's 'energic variables' [11]; eidontic variables are synonymous with Garner's 'structural' or 'informational' variables [11]; and tagmatic variables are equivalent to Jenkins' 'contextual variables' [14].

Our threefold catagorization of variables is motivated by the syntactic structure of the sign. It explains such experimental pheonomena as Garner's observation that at the informational level of perception all variables fall into two categories (i.e., ontotic and eidontic) having distinct and unique properties; and that there are two kinds of relationship (called by Garner 'state' and 'process') between perceptual variables and the sign processor. These and other applications of the theory of sign structure to perception are discussed below.

2. The Neural-Quantum Model and the Békésy-Stevens Valve

The Stevens' psychophysical power law [52] is a major improvement over the logarithmic law of Weber & Fechner in that it allows us to relate in a consistent way--namely, through a power relationship--several psychophysical variables, and to determine their relationship to each other and to the interpreting organism. It also enabled Stevens to upgrade the measurement of psychophysical variables from an interval measurement to a ratio measurement. As a result Stevens has gained enough insight into the processing of ontotic variables to ask new kinds of penetrating questions, and to refine experimental techniques to study how the interpreter processes ontotic variables. The results of this work are incorporated into a theory called the "neural-quantum model" [52], developed jointly by Békésy and Stevens.

In the language of the semiotic model, the stimulus variables are first processed by a neural-quantum (or Békésy-Stevens) valve which selects the ontotic (or psychophysical) variables, bypassing the eidontic (or structural) and tagmatic (or contextual) variables, and determines the organismic response strength corresponding to the physical intensity of these variables. The ontotic processing center also determines whether the present stimulus is a sign or merely a physical body, and it activates or deactivates the eidontic valve accordingly. This explains the metric relations existing between various psychophysical variables, the power relationship between physical itensity and psychological response to psychological variables, and the ability of the interpreter to selectively process the ontotic variables to the exclusion of variables of any other category (but not vice versa).

3. The Levels of Processing Model and the Day-Wood Valve

Attempts to explore the relation between the ontotic and eidontic levels of perception have been begun in a vague, unsystematic way (e.g., by Razran, Rommetveit, Jenkins, Skinner, Day and Wood, Garner, Posner and Mitchell). Garner and his colleagues [11] have carried out perhaps the most complete and systematic investigation of how the information processor interprets eidontic variables; but their results are open to interpretation until the more fundamental question of the relation between the levels is clarified.

To-date the most conclusive results on this question appear to be those of Day and Wood. What happens when an eidontic and ontotic property of the stimulus is processed in the same act of interpretation? Day and Wood [7] asked an equivalent question and found asymmetric interference: in an experiment involving six classification tasks, an ontotic variable produces interference when it is irrelevant to a judgment task requiring differentation of an eidontic variable. The opposite interference does not occur however.

These results are as expected for redundant coding. There is improvement in performance when the S needs it, as with correlated variables. There is, however, avoidance of interference only when the S is judging the ontotic variable; the S cannot avoid interference when judging the eidontic variable. Day and Wood suggest that such a situation makes sense if we consider that the ontotic variable is processed in a distinct processing center at a lower level than the eidontic variable. Thus an ontotic variable is processed before any eidontic variables, and it can be discriminated without interference from irrelevant variation in the eidontic property. If an eidontic property is the relevant variable, then any ontotic variation will interfere with discrimination of the eidontic variable, because that variation must be processed before the eidontic property can be perceived. Garner [11, p137] offers a partial explanation of this. A stimulus may exist with only ontotic properties (no eidontic properties); that is, an interpreter may process ontotic variables without processing any eidontic variables. A stimulus with eidontic properties cannot exist, however, without ontotic properties; i.e., an interpreter cannot process eidontic variables without first processing the ontotic variables.

By reference to Figure 17 we note that the diode symbol —|◀— represents the quantative aspects of the valve, and the mixerbox symbol, $\otimes$, represents the logical relations involved. The Day-Wood hypothesis of different processing levels is represented in the model by the distinct components and processing levels of the diagram. The necessity to process the ontotic variable before the eidontic, and the possibility of processing the ontotic variable without processing the eidontic are represented by feeding an output from the ontotic processing center into the control box of the eidontic, or Day-Wood, valve. With no input to the control box of the Day-Wood valve, the Day-Wood valve cannot ouput any eidontic variable into the eidontic processing center.

The semiotic model of perception also explains many of Garner's distinctions. According to Garner, the first important process in stimulus learning involves learning *what* exists; the second, discriminating

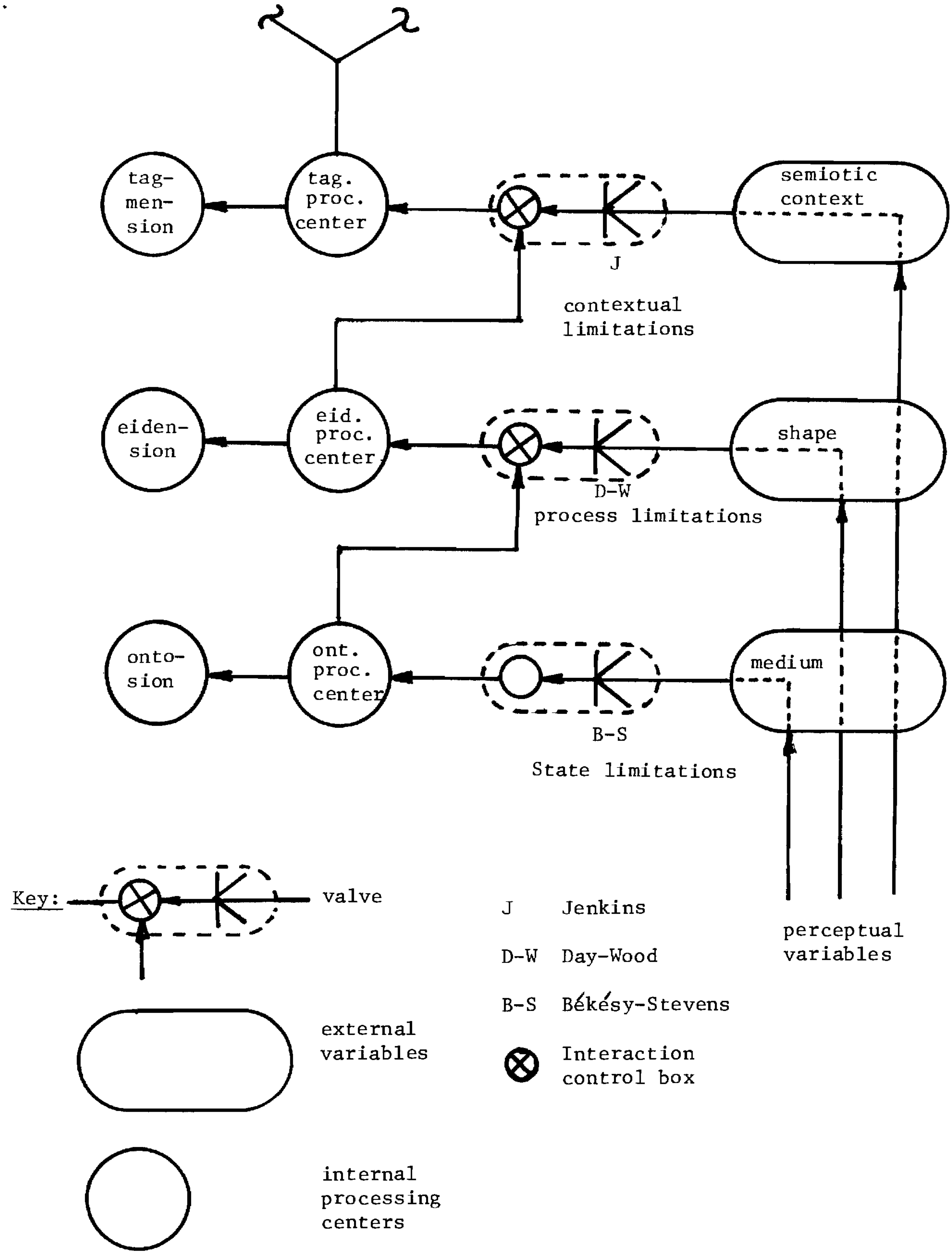


Fig. 17. A Semiotic Model of Perception

between the items *that* exist [11]. This distinction corresponds to perception at the ontotic and the eidontic levels. Furthermore, if letters are used as the levels and variables of multidimensional stimuli such that they form "real" words rather than nonsense words, the effects of eidontic and ontotic structure are considerably attenuated, but not eliminated. This indicates that the stimuluses are perceived as words and they form a higher level of interpretation, perhaps at the tagmatic level as suggested by the semiotic model of perception.

Garner finds it necessary to distinguish between the informational properties of a stimulus and its "energic" properties. This distinction differentiates alternative ways in which stimulus redundancy can affect discrimination performance. Energic properties provide activation of the sense organ and correspond to our ontotic properties. He uses the label 'informational' or 'structural' equivocally, sometimes referring to the shape properties of signs (the eidontic properties) and sometimes to semiotic properties in general. Garner also uses a distinction concerning factors that can limit perception. He calls these the 'state axis' and the 'process axis'. In our model, limitations concerning ontotic properties correspond to the state axis while limitations concerning eidontic properties correspond to the process axis.

4. An Interpretation of Jenkins' Phenomena: the Jenkins Valve

Jenkins [14] has also observed a "massive interference" effect in several experiments. However, in this case, both the ontotic and eidontic aspects of the stimulus were fully processed. As Jenkins himself observed, the effect seems to involve an asymmetry in the contextual relations of the sign.

We therefore postulate that this effect can be explained by a valve similar to the Békésy-Stevens and the Day-Wood valves, and propose to call it the 'Jenkins valve'. Based on Garner's analysis of the Day and Wood experiment we suggest that the Jenkins valve can best be isolated and observed by concentrating the S's attention on two processing levels simultaneously, perhaps by having S measure word spelling errors (eidontic property) while monitoring part of speech usage of the same word (tagmatic property) *vs.* measuring category errors while monitoring corrections of spelling.

5. Proposed Measurement of the Operating Characteristics of the Day-Wood Valve

The Békésy-Stevens neural-quantum model may be interpreted as a theory of the operating characteristics of the Békésy-Stevens valve. Compared to our knowledge of this valve, our understanding of the Day-Wood valve is scanty: we know little else than its control box is a summatable on-off device. It is therefore of interest to be able to measure the operating characteristics and to derive a theory of the Day-Wood valve.

We believe this can now be accomplished, using the concept of eidontic deviance and the instrumentation discussed in section D1 below. We have observed at one point of the eidontic deviance scale an effect which universally and radically changes the way artificial words are interpreted: on one side of this point S's process only length, while on the other they process both length and pattern. This point then appears to be the initial operating point for the Day-Wood valve--a phenomenon we are anxious to use for measuring that valve's operating characteristics. Other effects which we anticipate will enable us to explore the operating characteristics of this valve are the Terwilliger effect [53], the Miller-Selfridge effect [21], and also various effects involving eidontic deviance and visual acuity, and novelty.

D. Syntactic Shape

Our research into the nature of syntactic theory has concentrated on the eidontic level, with the semiotic concept of shape being of primary interest. This section reports on our work in this area.

1. Eidontic Deviance

The deviation of the shape of a natural language sign from the hypothetical norm, or average shape, of a sign in a given natural language is of considerable interest to information science, for both theoretical and applied reasons. To measure such a deviation we have developed an instrument called the 'eidontic deviometer' or, in short, 'eidometer' [42]. Measurements on artificial word forms using this instrument are both reliable and precise.

In previous work, Miller, Bruner and Postman showed that the interpretation of signs is affected by their shape [20]. We expect the eidometer to enable a precise measurement of this phenomena, and hence to lead to a better understanding of the interpretation process. Thus far, we have redesigned the Miller-Bruner-Postman experiment using an elementary tachistoscope (teescope) with the stimuluses measured by the eidometer, and have performed successfully an exploratory trial (the number of interpretation errors as measured on the teescope is directly and linearly proportional to the eidontic deviance as measured on the eidometer). This experiment will be refined and carried out with a sufficient number of subjects to enable satisfactory tolerances to be placed on error bounds. This research should lead to a direct measurement of the redundancy curve for natural language, a measurement which has not been made before (although Shannon [49] determined upper and lower bounds for this curve mathematically).

The eidometer permits the redesign of many other classical experiments involving the measurement of word shape, as well as the design of new experiments investigating various aspects of semiotic shape. A file of nearly 100 preliminary experiment designs employing the eidometer has been compiled. One such experiment, the proposed measurement of the operating characteristics of the Day-Wood valve, was described in the previous section.

2. Polygram Frequencies

Tables of polygram frequencies are useful for the generation of artificial word forms and the study of redundancy in natural language. Since tables for American English which are publicly available are at least half a century old and suspect (having most likely come from counts of military documents), and since access to later and more general counts which exist requires a security clearance and "need-to-know", we have prepared a table of polygram frequencies from a count of 5.5. million letters in the Brown Corpus of standard American [28].

During the analysis of this count data we discovered a rank-frequency regularity among the letters. However, unlike the rank-frequency law of Zipf and Estoup for words, which is log-log in nature, the regularity for letters is log-linear in nature. We analyzed all available data for other alphabets and phonemic systems, and found this relationship to hold in every case. A preliminary literature search shows no previous mention of this regularity; it is hoped that after additional analyses (still to be performed) we may be able to report a discovery of a universal relation for the shape elements of a system of discrete signs.

3. Tagmatic Deviance

Although several proposals for measuring tagmatic deviance (deviation in contextual constraints) have been suggested, none are straightforward. A way must be found to obtain measurement of tagmatic deviance by interlinking the measurement of eidontic deviance at both the word level and the sentence level. Several schemes for doing this are presently under evaluation at our School.

4. Algorithmic Information

A historical and tutorial paper has been prepared on the measurement of semiotic shape, to appear in International Journal of Computers and Information Science [43].

In many kinds of signs, shape is primarily concerned with length and pattern. In 1965 Kolomogorov proposed a measure of shape which is mainly a measure of the pattern [17]; called 'algorithmic information' or 'complexity', it pertains to the length of the shortest algorithm that will produce a given sign as its output.

Patterns, however, can be described verbally, whether for the purposes of internal coding or of long-term memory and reproduction. In 1963 Glanzer and Clark, using signs composed of linear arrays of black and white elements, showed that accuracy of reproduction of patterns was correlated with the length of description of the patterns [12]. In this case the correlations were based on average rather than minimum lengths, and length was measured as the number of words in a natural language (American) description rather than the number of steps in an algorithm. Using various outline shapes, Glanzer and Clark further showed that the length of the description was correlated with judged complexity of the shapes [13]; in general, longer descriptions go with greater difficulty of learning and with greater judged complexity.

Conceptually, the Kolmogorov and the Glanzer-Clark measures are the same. Kolmogorov's measure is a formal or mathematical model of Glanzer-Clark's empirical measure.

E. Semantic Structures

Another area of original investigation which has just begun concerns the semantic structure of signs. Although the Universal Sign Structure Model stems from research into natural language, this same structure should, if it has any corellation with reality at all, show up also in other disciplines which study sign processes--disciplines such as philosophy and psychology.

1. Analysis

A preliminary argument has been developed which shows the usefulness of the universal structure model for unraveling philosophical problems. G. E. Moore, an early twentieth century British philosopher, developed a paradox which has come to be called Moore's paradox of analysis and may be stated as follows: if the analysis of the meaning of a word has the same meaning, it is trivial; but if it has a different meaning, then it is wrong. Moore knew well that philosophers very often make correct and non-trivial analyses, but he was never able to develop a theory of analysis which overcame his own paradox. While other philosophers have tried with varying amounts of success, the problem has never been solved completely. The most popular approach is to say that the problem lies in the formulation of the paradox, which assumes that meaning is either a single or wholistic kind of thing which is either completely the same or else altogether different. Frege [10] and Carnap [6] assumed that the meaning of signs has two components, but their assumptions were for entirely different purposes. Carnap was able to delineate the character of scientific analysis fairly well with his 'extension' and 'intension', but he was never able to handle philosophic analysis. Moore himself said he thought philosophic analysis required something like determining the same objects by the same properties but understanding or cognizing this determination in a different way.

From our sign structure model (Figure 1), we note that cognision uniquely determines intension, which in turn uniquely determines extension, while a difference in extension ensures that two terms will have a difference in intension, which in turn ensures a difference in cognision. We may therefore state the solution of Moore's paradox as follows: Scientific analysis requires an identical extension with a difference in intension, while philosophic analysis requires an identical intension with a difference in cognision.

2. Memory Coding

Another area we have begun to explore concerns cognitive representation. Kintsch has reported three aspects of cognitive memory which he calls 'sensory', short term', and 'long term' [16]. Bruner has reported several modes of representation, or coding, including 'enactive', 'ikonic', and 'symbolic' [4]. He has studied the sequence in which these capabilities develop in children and the rate at which signs can be processed using the various modes of representation. It would appear as if there were just one form of coding associated with each aspect of cognitive memory; however, this is not clear because of confounding effects on the experiments.

An experimental program is being designed to critically isolate each memory aspect and the mode of representation that is associated with it. The first experiment, to isolate and determine the characteristics of iconic coding, uses an interference effect suggested by Siegmann [50]; in experimental trials the interference effect is well-marked and can be detected easily [36]. Additional experiments are planned, including ones using children to verify Bernbach's [2] results.

The advantage of achieving an answer to this question is to allow quantitative measurements of psychology to be used in future investigations of semantic structure. For instance, memory span times, processing rates, and age of development are all quantitative measurements, and all run in the sequence: index, icon, and symbol.

3. Semantic Linkage Strength

The memory coding experiments described above lead in a natural way to the development of measures for semantic linkage strength.

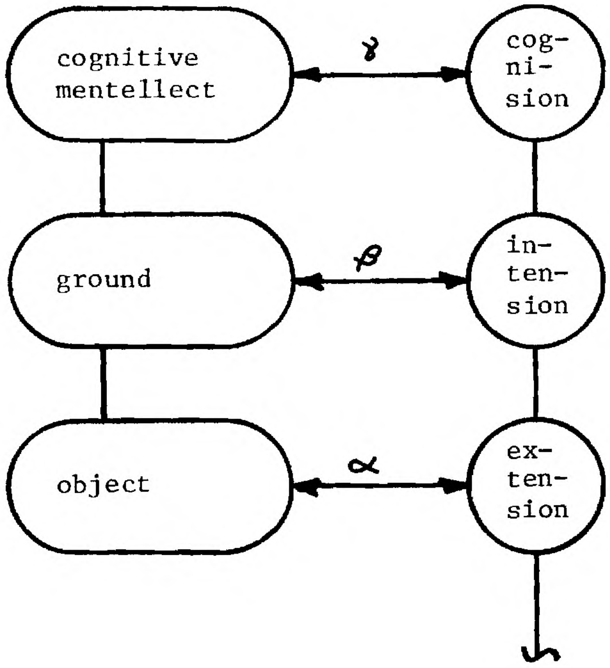


Fig. 18. Semantic Field Structure

According to the Universal Sign Structure Model some signs have the semantic field structure shown in Figure 18. One hypothesis that is being investigated (semantic field structure hypothesis) requires all signs to have all six semantic components and all three linkages: the indexical link α; the iconic link β; and the symbolic link γ.

The new concept of α, β, and γ 'linkages' requires empirical establishment. In our thinking, α, β, and γ become empirical measures of semantic linkage strength. They are information measures in the sense of Zunde and Pearson [55] and their practicality must be established by semiotic reinterpretation.

As an example of how one might go about developing measures of linkage strength, consider the iconic linkage strength β. From the paradigm of the Pearson-Siegmann experiment described above we have measures of what may be called iconic interference. From the Bernbach experiment we have motivation for interpreting this as a measure of iconic linkage strength.

By generalizing both experiments we may hope to find that the ratio of the short-term memory component to the long-term associative effect varies with the ratio of the iconic interference effect to what may be called 'non-interference'. This may be used both to develop an interval scale for measuring β and for establishing a semiotic reinterpretation for β as an information measure.

F. Applied Research Potential

We have shown initial evidence of the power of the proposed theory of sign structure to explicate sign phenomena, the utility of which must be further demonstrated. In addressing the question of the utility of this research to practical issues in information processing we may afford to be mildly speculative. In the body of this report we have alluded to our current and planned efforts at the applied plane (e.g., research into aphasia and related brain disorders); we also see our work relevant to a number of applied problems in information technology. The following illustrations may be given:

- Now that programming syntactics has reached an initial maturity of development, interest in programming theory has begun to turn to programming semantics. Early studies into programming semantics have concentrated on a single-level semantics. Our investigations into the structure of signs would suggest, however, that in order to achieve the full power of symbolic communication of which digital computers and their compilers are capable, a three-level semantics is required. A full understanding of symbol structure will be required to develop such theories.

- There exist almost no theories of programming pragmatics, and few studies of the subject have ever been made. Our studies of sign structure suggest that the pragmatic dimension is independent of the semantic dimension (a major departure from the Peirce-Morris theories), and that it may be at least as important to programming theory as the semantic dimension. In order to study programming pragmatics and develop appropriate theories, an understanding of the pragmatic structure of signs is required.

- Even a cursory look at the notations of control functions in programming systems shows beyond doubt the confusion facing computer programmers and computer users. The choice of control functions and their notations in individual programming systems has had no basis in theoretical principles or in the empiricism of human engineering. Our theory of sign structure appears to provide a useful framework and a tool for the empirical, science-based development of program control functions and their notations.

- Many issues in the vast problem area of human interaction with computer-based information systems concern the coding of symbols, indexes and icons. Most coding studies to-date have dealt with the coding of symbols only. Furthermore, there are two types of coding involved: 1) the creation, change, and interpretation of the shape of signs; and 2) the storage, linkage, and retrieval of signs into, in, and from memory. The coding theories of Shannon and Wiener address the former, while studies by Bruner, Broadbent, Kintch and others address only the latter. So far, there has been little reference to the common relationships involved between these two types of coding and studies have made either little or naive use of understanding of sign structure. Since it is plausible to argue that the man/machine interface problem concerns in part the relationship of the two types of coding, it would appear that our theory is a potential tool for this virgin area of applied research. This is so because this theory encompasses a language and a power to interrelate the semantic and syntactic structure of indexes, icons and symbols.

It is easy to expand on this list of applied problems which appear relevant to our theoretical work. There is no question in our minds that the practical, the applied and technological results will be forthcoming at the appropriate time when the understanding of the natural semiotic phenomena has been achieved. The search for such understanding is in the realm of a basic science of information.

IV. THE SEMIOTICS LAB

Early in the project the School of Information and Computer Science moved to establish a Semiotics Lab, in support of both research and instruction in several related courses. The initial development was reported in [37]. A lab manual has been published by the School and is available through the Georgia Tech Bookstore [23]. The instrumentation recently added to the SemLab includes several eidontic deviometers, a teescope, and a timer for memory coding experiments.

A. Computer Software

A major, resource of the SemLab is its bank of computer software for semiotic research. Documented programs developed by the SemLab are announced in Semiotic Scene, a publication of the Semiotic Society of America, and in Foundations, a newsletter of the ASIS Special Interest Group in Foundations of Information Science. Among the programs available at cost upon request are:

WORDGEN, A Markov artificial word generator, generates words of Markov order 0, 1, 2, and 3, using tables of relative polygram frequencies for monograms, diagrams, and trigrams.

WORDGN3, another Markov artificial word generator, generates artificial words of *all* finite Markov orders using machine readable natural language text.

TTKANAL, an instrument for measuring types and tokens in a sample of running text. As an additional feature, it also performs a rank-frequency analysis, a type-token analysis, and computes Yule's K coefficient for the sample.

SIMIMAB uses data collected from word sorting experiments to build a similarity matrix and generates linkages to word clustering routines which use the similarity matrix to compile word lists. The latter are used in list sorting experiments and in building eidometers.

B. Machine Readable Data Bases

Another substantial resource of the SemLab, developed over the period of this NSF grant, is its collection of machine readable texts for purposes of semiotic research. During the last year, this data base has been increased by a corpus of technical writing in pharmacology; three German corpuses (including one complete novel); a corpus of aphasic text; a copy of chimpanzee generated language text; and other materials. A typescript corpus of natural language text, generated while its authors were under controlled dosage rates of various drugs, has been obtained, and plans are underway to convert it into machine readable form.

C. Collaborative Projects

Because of the unique facilities and activities of the SemLab, a number of collaborative activities with other institutions have developed during the program period. Illustrative of these are the following:

1) We assisted the University of South Florida to design and conduct an experiment to collect samples of aphasic text under controlled conditions, along with control samples of normal text under the same conditions, to allow a comparative analysis of rank-frequency and type-token relationships. The experiments are carried out at the Camp Challenge facility in Florida; the results are analyzed in the SemLab. Preliminary results show no differences in the rank-frequency and type-token results between aphasics and normals [35]. Negotiations are underway with the Florida Easter Seal Society (which operates the Camp Challenge facility) and the Speech Department of the University of Florida (which coordinates research at Camp Challenge) to conduct a new set of experiments designed to capture a larger sample of text from the aphasics and to obtain a more precise classification of aphasic disability.

2) Negotiations have been opened with the Georgia Institute of Mental Health and, independently, the Virginia Institute for the Living to conduct similar experiments on schizophrenics and manic-depressives.

3) Lists of 4, 5, 6, and 7-letter words of high frequency appearing in the Brown corpus were made available, along with the polygram frequency lists compiled for these words, for experiments in information processing conducted at Georgia State University.

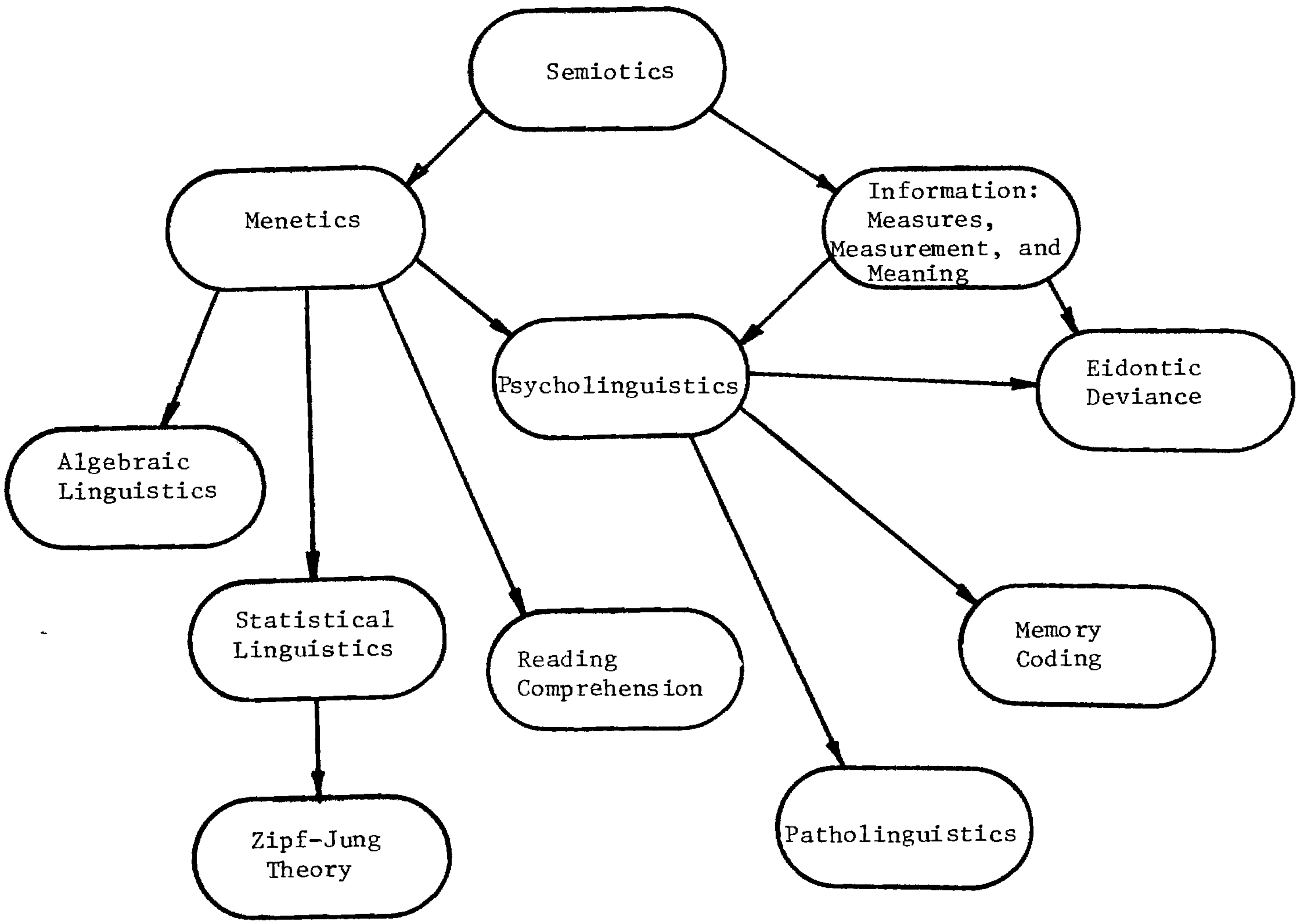
4) The SemLab conducted a rank-frequency and a type-token analysis on a running sample of chimpanzee-generated language text for Project "LANA" at Yerkes Primate Observatory.

D. Semiotics Bibliographies

Several bibliographies on different topics relating semiotics and information science have been prepared. The different bibliographies are related as shown in Figure 19.

Although prepared primarily for use by project personnel, the bibliographies are of general interest, and were made avaiable to Professor Umberto Eco, executive director of the International Association for Semiotic Studies. They will appear in a future issue of VS, the Italian journal of semiotics. Individual bibliographies have been made available, upon request, to scholars throughout the world.

Fig. 19 Bibliographic Subset Lattice for Semiotics



E. Organizational Activities

The principal investigators have been very active in the formation of the Semiotic Society of America in 1975/6. Partially in recognition of the Semiotic research at Georgia Tech, the SSA chose Atlanta as the site of its first national conference, hosted by the School of Information and Computer Science. Conference proceedings, edited by one of the project directors, are schedule to appear [33]. Also as a result of our efforts the SSA developed a special interest group structure emphasizing experimental, mathematical, and theoretical semiotics.

In conjunction with this conference the School of Information and Computer Science conducted a one-day Workshop in Experimental Semiotics on September 23, 1976. The Workship was attended by 40 scholars from the U.S. and abroad, and was enthusiastically received [34]. We plan to conduct a second workship in experimental semiotics in early 1978.

V. REFERENCES

1. Ash, R. Information Theory. Interscience, 1965.

2. Bernbach, H. A. "The Effect of Labels On Short-Term Memory for Colors With Nursery School Children". Psychon. Sci., 7(1967), p149-150.

3. Bridgeman, P. W. The Logic of Modern Physics. Macmillan, 1928.

4. Bruner, J. S. "On Cognitive Growth, I & II". In Bruner-1, p1-67, 1966.

5. Bruner, J. S.; et.al: (Bruner-1). Studies in Cognitive Growth. Wiley, 1966.

6. Carnap, R. Meaning & Necessity. University of Chicago Press, 1958.

7. Day, R. S.; and Wood, C. C. "Interactions Between Linguistic and Nonlinguistic Processing". Journal of the Acoustical Society of America., 51(1972), p79.

8. Flowers, J. "Progress in Semiotics Software at Georgia Tech's SemLab". The Semiotic Scene, (newsletter of the Semiotic Society of America), 1(1976).

9. Flowers, J. "Software Development Progress at Georgia Tech's SemLab". To appear in ASIS-SIG/FIS Foundations, 1977.

10. Frege, G. "Über Sinn und Bedeutung". Zeitschr. f. Philos. u. Philos. Kritik, (1892), p100. Reprinted as "On Sense and Reference" in The Philosophical Writings of Gottlob Frege. Tr. by M. Black.

11. Garner, W. R. The Processing of Information and Structure. Wiley, 1974.

12. Glanzer, M.; and Clark, W. H. "Accuracy of Perceptual Recall: An Analysis of Organization". Journal Verbal Learn. and Verbal Behavior, 1(1963), p289-299.

13. Glanzer, M.; and Clark, W. H. "The Verbal-loop Hypothesis: Conventional Figures". American Journal Psychol. 77(1964), p621-626.

14. Jenkins, J. J. "Remember That Old Theory of Memory? Well, Forget It". American Psychologist, 1974, p785-795.

15. Johnson, W. E. Logic. Cambridge, 1921.

16. Kintsch, W. Learning, Memory and Conceptual Processes. Wiley, 1970.

17. Kolmogorov, A. N. "Three Approaches to the Quantitative Definition of Information." Problems of Information Transmission. 1(1965), p1-7.

18. Levi, A. W.; and Frye, A. M. Rational Belief. Harcourt, Brace and World, 1941.

19. Mill, J. S. A System of Logic. 1843.

20. Miller, G. A., Bruner, J. S., and Postman, L. "Familiarity of Letter Sequences and Tachistoscopic Identification". Jour. Gen. Psychol., 50(1954), p129-139.

21. Miller, G. A.; and Selfridge, J. A. "Verbal Context and the Recall of Meaningful Material". Amer. Jour. Psych., 63(1950), p176-185.

22. Pearson, C. R. "Peirce's Theory of Signs and the Universal Structure Model". A seminar presented to the University of Pittsburgh Interdisciplinary Program in Information Science, Pittsburgh, Pennsylvania. October 21, 1975.

23. Pearson, C. R. "Lab Manual for Semiotics". Atlanta, Georgia, School of Information and Computer Science, Georgia Institute of Technology, 1975. Itek.

24. Pearson, C. R. "Goals for Long-Range Research in Information Science". Paper presented at the 12th National Information Retrieval Colloquium; Drexel University, Philadelphia, May 1-2, 1975.

25. Pearson, C. R. "Information Science and the Universal Structure Model". To appear in ASIS-SIG/FIS Foundations, 1976.

26. Pearson, C. R. Communication Processes (Course Notes). Atlanta, Georgia, Georgia Institute of Technology (School of ICS), 1976. (Mimeographed).

27. Pearson, C. R. "Review of Information and Structure by Garner, W.P." To appear in Computing Reviews. 1977.

28. Pearson, C. R. "A Table of Polygram Frequencies". (In preparation.)

29. Pearson, C. R. Towards an Empirical Foundation of Meaning. Atlanta, Georgia, Georgia Institute of Technology, 1977. (Ph.D. thesis)

30. Pearson, C. R. "The Senses of 'Meaning'". (In preparation.)

31. Pearson, C. R. "The Components of Signs". (In preparation.)

32. Pearson, C. R. "Quantitative Investigations Into the Type-Token Relation". To appear in Semiotica, 1977.

33. Pearson, C. R. (Ed.) Proceedings of the 1976 Annual Conference of the Semiotic Society of America. Atlanta, 1977.

34. Pearson, C. R. (Ed.) Proceedings of the 1976 Workshop on Experimental Semiotics. Atlanta, School of Information and Computer Science, Georgia Institute of Technology, 1977. (In preparation.)

35. Pearson, C. R.; and Goodman, C. "The Type-Token Relationship in Aphasics". (In preparation.)

36. Pearson, C. R.; Siegmann, P. J.; and Shin, K. J. "An Experimental Investigation of Iconic Coding". (In preparation.)

37. Pearson, C. R.; and Slamecka, V. "Semiotic Foundations of Information Science". In Research 1973/74: Annual Progress Report, School of Information and Computer Science, Georgia Institute of Technology; Atlanta, Georgia; p3-7.

38. Pearson, C. R.; and Slamecka, V. "Semiotic Foundations of Information Science". Internal Research Memorandum, March 1975, School of Information and Computer Science, Georgia Institute of Technology, Atlanta, Georgia.

39. Pearson, C. R.; and Slamecka, V. "Semiotic Foundations of Information Science". Progress Report No. 2. Internal Research Memorandum, March 1976, School of Information and Computer Science, Georgia Institute of Technology; Atlanta, Georgia.

40. Pearson, C. R.; and Slamecka, V. "Review of Experimental Work in Semiotics". To appear in Progress in the Communication Sciences, 1(1977).

41. Pearson, C. R.,; Slamecka, V.; Zunde, P.; Flowers, J.; Betts, R.; Shin, K. J.; and Canyas, D. "Semiotic Foundations of Information Science". In Research 1974/75: Annual Progress Report, School of Information and Computer Science, Georgia Institute of Technology; Atlanta, Georgia; p3-10.

42. Pearson, C. R.; and Smith, L.; and Ray, J., Jr. "The Eidontic Deviometer: An Instrument for Measuring Shape Deviance in Natural Language." (In preparation.)

43. Pearson, C. R.; and Zunde, P. "The Kolmogorov Potential as a Measure of Algorithmic Information". To appear in International Journal of Computers and Information Science.

44. Pearson, C. R.; and Zunde, P. A Semiotic Approach to the Statistical Theory of Syntactic Communication. (Textbook in preparation.)

45. Peirce, C. S. Collected Papers of Charles Sanders Peirce.* 8 vols. (Ed. by C. Hartshorne and P. Weiss, vols. 1-6; A. Burks, vols. 7-8). Harvard U. P., 1931-58.

46. Robinson, R. Definition (Corrected Ed., 1965), Oxford, 1954.

47. Russell, B.; and Whitehead, A. N. Principia Mathematica (2nd Ed., 1925) Cambridge U P, 1910.

*References to the Collected Papers will be given in the standard form where 2.228 refers to vol. 2, paragraph 228.

48. Shannon, C. E. "A Mathematical Theory of Communication". Bell Syst. Tech. Jour., 27(1948), p379-423, p623-656.

49. Shannon, C. E. "Prediction and Entropy of Printed English". Bell Syst. Tech. Jour., 30(1951), p50-64.

50. Siegmann, P. J., personal communication, 1975.

51. Slamecka, V.; and Pearson, C. R. "Information Science". In A. Ralston, (Ed.) Encyclopedia of Computer Science. Petrocelli Books, 1975.

52. Stevens, S. S. Psychophysics. Wiley, 1975.

53. Terwilliger, R. F. Meaning and Mind. Oxford U. P., 1968.

54. Whitehead, A. N.; and Russell, B. Principia Mathematica. 3 vols. Cambridge U. P., 1910-13.

55. Zunde, P.; and Pearson, C. R. Information Measurement. (Textbook in preparation.)