Robert E. Taranto

A SEMIOTIC THEORY OF CODES

1.

This theory came to paper after a long period of research regarding the prerequisites for Artificial Intelligence involving signs, the connection between Artificial Intelligence and codes. This is based on the recent systematic and historic approach to codes in view of Artificial Intelligence and from the viewpoint of Theoretical Semiotics as defined by Charles S. Peirce (1839-1914) and developed to the point where it became an applied science, by Max Bense, Elisabeth Walther and many others at the University of Stuttgart in Germany.

Principally, each point of a semiotic space can be designated as a sign by means of a triadic connection between an object, the medium of representation and its interpretant. Not only that "anything" can be used as a medium for representation but also the designated object, i.e. the sign as an object of designation, can be used for representation.

The fact that all "things" in this world fall into one of the ten semiotic classes is scientifically important when communication is to be analyzed and hence will be of significance for understanding the nature of codes, information and, consequently, for further development of Linguistics and all the sciences involved in the development of Artificial Intelligence.

If some day a theory of the phenomena of consciousness could be established, then it will have been made possible by the pioneering development work of Max Bense realized at the University of Stuttgart.

From the fundamentals of Ch.S. Peirce' semiotics Max Bense postulated operational and foundational theoretical semiotics. Its theoretical and hypothetic evidence is supported by set theory, particularly by the axiomatic scheme of class formation by von Neumann-Bernays-Gödel.

It may be left open whether the postulates in theoretical semiotics deal with real objects or with idealizations, i.e. the rules of formation and deduction of a formal language and logic. There is every reason to believe that the codification of intuitive concepts and the representation of accepted principles will continue also in the future, and will bring new advances into territory still uncharted.

The author wishes to record here his indebtedness to Max Bense who, many years ago, deeply influenced his thinking on problems of epistemology, and whose friendly criticism was material in achieving whatever clarity was achieved.

2.

That which makes a language learnable is the fact that knowledge is transferable, i.e. communication of information can take place. The "vehicles" for this communication, i.e. of data transfer, are energetic or material mediums to which intrinsic "object" relationships are assigned with a particular character for interpretation. This triadic relationship of medium, object and interpretant is nothing else but a sign! The sign as defined by Ch.S. Peirce is the foundational element of each and every code, and consequently of every language and metalanguage.

The notion that each and every sign is actually a triadic representational scheme was introduced by Ch.S. Peirce, who defined the categorical, respectively fundamental aspects of the sign as being: a medium as firstness, the object in question as secondness and its associated interpretation as thirdness. In order to facilitate the sign functional and operational research, Max Bense introduced the primes of .1., .2. and .3. as functors for Peirce's categories. The complete graduating sequence for the representational scheme became an operational matrix, from Peirce's

M (quality, singularity, legality),
O (icon, index, symbol),
I (rhema, dicent, argument),

where the threefold, trichotomic interpretants are:

Rhema ==> "neither true or false" (Peirce) = open (Bense),
Dicent ==> "true or false" (Peirce) = closed (Bense), and
Argument ==> "always true" (Peirce) = complete (Bense),

to the operational, graduating function of Bense's "Small Matrix",

	.1.	.2.	.3.
.1.	1.1	1.2	1.3
.2.	2.1	2.2	2.3
.3.	3.1	3.2	3.3

The semiotic system of triadic sign relationships is the lowest limit beyond which no more reduction can take place in describing the world. The ontological-epistemological analysis approaches the limits of reductionism for the ontological understanding of the universe. The formal translation of representation into thematization, i.e. the dualization (Bense) of signs relates to its object language as a meta-language does. Therefore, the sign classes and their representation result in a dual system, where selective and ordaining stepping functors act according to the hierarchical founding scheme of .1. .2. .3.; in .1. all repertoires, in .2. all possible (object relational) structures, and in .3. the stati of the interpretant, i.e. the mind, bringing together the "key" to each and every code.

Semiotically, a code is a sign with intrinsic meaning attached to it, in theoretical semiotics each code is a hierarchically structured complex sign. Which will say it is a set of interrelated signs, sequenced or nested within each other, designed to convey an abstraction.

This basic statement could be illustrated with an example like the following sequence:

What is the meaning of the first dot? The second? Well, for somebody uninitiated it is just a sequence of dots and dashes, but others will recognize in it something "in Morse code", the sequence that stands for SOS, still others could even know that SOS stands for "Save Our Souls". How much information can be imbedded in the eight times

interrupted line? Semiotically, the qualitative elements or qualisign (1.1) can be combined to various sequences of these dots and dashes, each sequence being a special "Gestalt" or sinsign (1.2), and all possible sequences together form the Morse alphabet representing letters, numbers etc. of a language.

Codes like signs are "vehicles" for communication, i.e. for information transfer, therefore in addition to the pure theoretical aspects they have also a communicative one which is of primary interest when a theory of codes is approached.

An analysis of the firstness, i.e. of the Medium, in the above-coded message in Morse code, shows plainly that in addition to the repertoire of (visible) dots and dashes there are also the repertoires of letters and words (for which the letters SOS stand for) involved. The meaning of the message is of no interest to the analysis at this point because it is contextually and procedurally dependent from the context in which it appears.

Secondness is in this case the signal itself, i.e. the interruption of the continuing line. Let us not forget that a signal is defined as the interruption of a continual of whatever energetic or material nature it be. Therefore the signal, i.e. the fact that there are interruptions (and not the formal dot-dash sequences) are themselves the Object. So, there is the medium of, say "ink on paper" trace out of which a signal is generated. (The signal becomes a sign only if it has an interpretant, of course).

Thirdness is the Interpretant, which can be "open", "closed" or "complete" (Bense) or, with other words, the interpretant is the addition to the dyadic relationship Medium - Object combinational interpretation, which can be "true", "false" or "don't-care" and without which no sign can exist; (a sign communicates only when perceived).

In the foregoing example of the coded message of "Save Our Souls", the metasemiotical description of realities has the trichotomic triad:

1.1 1.2 1.3 = Medium thematized Medium
2.1 2.2 1.3 = Object thematized Medium a)
3.1 3.2 1.3 = Interpretant thematized Medium,

The next substrate in the hierarchical structure of a code, i.e. the next level in which the "message" of the code is represented, is the metasemiotic description of the realities involved with the trichotomic triad:

3.1 1.2 1.3 = Medium thematized Interpretant
2.1 2.2 1.3 = Object thematized Interpretant b)
3.1 3.2 1.3 = Interpretant thematized Medium

Analyses of the communicative aspects of codes show that more than one sign class is involved in the coding and decoding functions of even the simplest code as shown above.

Out of the (monadic) medium a signal is generated whose (dyadic) function becomes an operative sign, then and only then, when the interpretant (triadic) has been integrated.

(2.1 2.2 1.3 2.1 2.2 2.3) = (3.1 3.2 1.3 - 2.1 2.2 2.3 - 3.1 1.2 1.3)

this can be expressed also symbolically as:

(O(M) > O(O) ⇒ (I(M) ⟶ O(O) ⟶ M(I);

3.1 3.2 1.3 = Interpretant thematized Medium,
2.1 2.2 2.3 = Object thematized Object, and c)
3.1 1.2 1.3 = Medium thematized Interpretant.

The thematized (dualized) small matrixes of a, b, and c are:

$$\mathcal{Z}_a = \begin{matrix} 3.1\ 2.1\ 1.1 \\ 3.1\ 2.2\ 1.2 \\ 3.1\ 2.3\ 1.3 \end{matrix} \qquad \mathcal{Z}_b = \begin{matrix} 3.1\ 2.1\ 1.3 \\ 3.1\ 2.2\ 1.2 \\ 3.1\ 2.3\ 1.3 \end{matrix} \qquad \mathcal{Z}_c = \begin{matrix} 3.1\ 2.3\ 1.3 \\ 3.2\ 2.2\ 1.2 \\ 3.1\ 2.1\ 1.3 \end{matrix}$$

The foregoing dualization (Bense) results in a formal translation of representation into thematization of the realities involved

(semiotically) when codes communicate information. The first three levels of any code as far as the communicative (semiotic) aspects of codes are concerned are:

1. Perception of (Morse) code sign as in a),
2. Recognition of the letters as in b),
3. Understanding of the message as in c).

Codes are operators that have syntactical (formal) as well as semantical (contextual) internal functions caused by functors which are basic (procedural) signs. Logic operation with codes add another pragmatic dimension to that which can be external(metatheoretical) or internal, a function of quantization and modalization.

Codes that cannot be selected have to be introduced, i.e. have to be ordained. Which will say that in order to be able to convey information I shall have to select a given code out of an available repertoire or I shall "invent" one within a given system (in agreement with the receiver). The semiotic sign class of all codes is the same as the one of all languages, hence the generative (and creative in the sense of grammatics) possibilities are limitless and, as in languages, codes too have syntaxes and morphologies. The general formula for languages and codes is,

$$Scl\ (Cs) = Scl\ (Ls);$$

Information theory is to-day no longer accepted as the key to understanding speech perception as it was only a few years ago. The theory according to which the occurrence of each sound in a word and each word in an utterance are statistically determined by the preceding sounds and words, was based on Shannon's work on information transmission. However, since Noam Chomsky's system of transformational grammar was substantially modified in 1965-70, resulting in a system of generative semantics, the role of the phonological components was to "interpret" the strings of words generated by the syntactic components and not just statistical elements. Regarding the triadic nature of these phonological components (Medium, Object and Interpretant) and their "interpretation" within chains, analyses of the communicative aspects of complex signs can shed some light.

The grammatical foundation of any language is an immensely complicated structure of interrelated systems governed by syntactical, morphological and phonological rules. Each of these systems (rules are procedural knowledge imbedded in code) has its own elements encoded in a different code.

All essential functors, when communication is effected by means of written language, are letters and words (in addition to the above mentioned codes) where each letter is an encoded sound. Morphs and phones or phonemes, are speech sounds when the spoken language is used and are, naturally, also encoded signs because (on the foundation level) signs are the only means available for communication.

Code, in the naive sense of the word, is actually a hierarchical structure of individual codes, each code being a multitude of signs. In a language, the hierarchical structures of codes can be seen as being "packed" in "envelopes"[1] of codes. Each code having its intrinsic, intensional, (own) "alphabet" to carry information about its particulars, i.e., so and so many elements, so and so many combinations possible, permutations and so on.

The basic constituents of a language are codes enclosed in "envelopes" which are nothing else but again encoded information of the various (inherent) combinational modes in which a code should function. This information is again encoded in "envelopes" generated by the relationship between syntax and semantics, which is a nonlinear function of the intentional invariance of signs, and consequently is the basic component for generation of expressions.

In addition to the encoding levels mentioned above (the envelopes within envelopes of letters/phonemes and words/lexicals) there exist also morphemic, lexemic and phonemic "envelopes" or strati which pertain to the deep structure of sentences (see Chomsky).

Postulate 1.

Codes are signs. Each and every sign is a triadic-trichotomic relational scheme, therefore every code is a triadic-trichotomic sign-relationship between a Medium, an Object and an Interpretant.

Correlate to 1.1.

The repertoire from which the Medium (M) is chosen belongs always to the realm of the Object's axis of reference[2] and is thetically introduced by the thinking mind.

Postulate 2.

At the base of each hierarchical code-structure for any one given system, there is at least one code which has a Medium (M) chosen from a physical repertoire which is detectable by the senses, i.e. such as acoustic, visual phenomenons, signals and so on.

Correlate to 2.1.

No communication is possible without having at least one code within a certain code structure, such as a language, with a substantive repertoire for the Medium located in the physical world.

Correlate to 2.2.

Every code conveys information in two forms, namely contextual (elements) and syntactical (valid combinations of elements). The number of combinational sub-sets for a given code are a factorial function of the sum of elements for the given set and are called the 'potential' of the given code.

Every code can be a carrier of information, the knowledge content has aspects discussed above and can be seen as "encapsulated" knowledge which takes always the form of

Immediate <== Deklarative (Translate function),
and
Procedural ==> Epiphenomenon (Syntax function),
knowledge.

Intelligence depends crucially on the ability to create high-level descriptions of complex arrays by means of decoding and synthesizing the procedural information embedded in particular code.

A good example perhaps is the following, quite simple cryptographic message of World-War II vintage,

TBWF PVS UQWMV

The message, like in our first example with the Morse code is "Save Our Souls". Like in all ciphers, also here, substituting of signs takes place. The rule for substituting of signs is called code, but code is also the "missing" signs, i.e. that what is to be substituted. In the example, the first rule is substitution of each letter with the follow-on in the alphabet. The next rule is that with each repetition of a letter in the text, the next follow-on is selected from the alphabet and so on. The key to each succeeding level of decoding the above message is carried by the message itself, i.e. it is embedded in the rules. The "key" to decoding the letter sequence requires enough samples, time and ingenious iterative algorithms, the letter sequence just has "to be found out", it is there:

S ⟶ T	O ⟶ P	S ⟹ U
A ⟶ B	U ⟶ V	O ⟶ Q
V ⟶ W	R ⟶ S	U ⟹ W
E ⟶ F	--------	L ⟶ M
--------	--------	S ⟶ V

Note that the letter 'S' is translated differently the first, second and third time. Some features of a code, like the paradox of Epimenides[3] cannot be specified in advance, thus there must exist rules for inventing new rules as the processing of information goes on. Without doubt, rules that change themselves directly and indirectly are the foundational prerequisite for intelligence. Thus, it is important to distinguish between processing information (words, sentences) in one system, supervising the 'rules' of information perception in a second, "look-ahead" in a third and so on, all in a hierarchical structure of processes taking place almost simultaneously. Decoding the manyfold layers of (input) information is aided by the perception of isomorphisms between structures which create meaning and allow interpretation, though gaps in some codes occur often enough.

Linguistics has become a highly technical subject and does indeed incorporate such major divisions as phonetics, grammar and semantics. Codes are just another form for whatever the names of the "vehicles" for communication (in linguistics) are given, and therefore belong to the subject.

The extent of the interdependence of language and thought, i.e. the so-called linguistic relativity, has once and for all separated language and thought; nevertheless, sooner or later linguists will have to deal with the foundational, elementary "particles" of language without neglecting the main function of it, namely communication and intention. Codes have to be theoretically defined as categorical, functional and operational entities, and that can be done only if theoretical semiotics with its triadic trichotomic sign-scheme is involved.

Codes are based on signs; already the rationalists in the 17h century knew that speaking is expressing thoughts by signs invented by the human thinking mind. Is there anyone who doubts this fact?

NOTES

1 The term "envelope" is commonly used in radio engineering and defines a modulated carrier wave form in which a number of different frequencies are bound together under one envelope, each frequency containing its own "intelligence" or communication channel.

2 If during a communication the subject-matter is, say, graphics, the M's repertoire will be one of graphics, if the subject-matter were mathematics, the repertoire from which the Medium would be picked will be that of mathematical entities, and if the Object is something scratched on the black-board, then the Medium is from a repertoire of chalks, red, white, green and so on.

3 Epimenides paradox does not allow the proper perception of a sentence until another one puts the first at doubt, like:

The following sentence is false.
The preceding sentence is true.

Therefore, the truth of the first sentence cannot be defined 'apriori'.

SEMIOSIS 39 40

Internationale Zeitschrift
für Semiotik und Ästhetik
10. Jahrgang, Heft 3/4, 1985

INHALT

Georg Nees:	*Kasus und Zeichen*	5
Robert E. Taranto:	*A semiotic theory of codes*	14
Olga Schulisch:	*Farbe und Kommunikation*	24
Günther Sigle:	*Eine semiotische Bemerkung zur axiomatischen und konstruktivistischen Grundlegung der Arithmetik*	33
Georg Galland:	*Zur Semiose des Begriffs*	36
Elisabeth Walther:	*Semiotik der natürlichen Sprache*	46
Barbara Wichelhaus:	*Märchentext - Märchenbild Eine semiotische Untersuchung*	62
Pietro Emanuele:	*La teória dei contrassegni (Merkmale) in Lambert e in Kant*	72
Werner Steffen:	*Die Beeinflussung des Kunstwerks als Zeichen durch den vom Künstler beigegebenen Titel*	78
Matthias Götz:	*Gegen "Phantasie". Eine Attacke*	91
Gerhard Wiesenfarth:	*Dynamik im Prozeß des Umgestaltens*	111
Wendelin Niedlich:	*Buchhändler N. notiert den Stand von Max Bense*	120

Nachtrag zu "Das Phänomen der Orthogonalität" von Gotthard Günther 124

Nachrichten 125

Inhalt von Jahrgang 9/10, 1984/1985 126

Wir danken Herrn Direktor Hans Lense, Laupheim, für seine großzügige finanzielle Unterstützung.