

Meaningful Agents: A Semiotic Approach

Antônio Gomes, Ricardo Gudwin, & João Queiroz
{asrgomes, gudwin, queiroz}@dca.fee.unicamp.br
DCA-FEEC-UNICAMP, Cx. Postal 6101, 13081-970, Campinas – SP – Brazil.

Abstract – *Meaning is a key problem in intelligent systems design. Most approaches in literature still adopt naïve definitions, which cause them many subtle problems such as the symbol grounding problem. In this work we propose a semiotic approach to this problem based on the triadic model of sign developed by C.S. Peirce. In order to observe the key aspects of this model of simulated semiosis (meaning processes) we developed an experimental protocol inspired on the work of Luc Steels (the talking heads experiment). We argue that, with the proposed theory, meaning emerges within each agent without any external interference (from other agents or the creature designer himself); they can exhibit semiosis, in a simulated form.*

1. INTRODUCTION

Meaning is a central problem in intelligent systems design. Most approaches in literature still adopt a strict symbolic definition of meaning, which usually results in deficiencies such as the lack of symbol grounding [7]. The objective of this research is to come up with a computational model of the emergence of meaning in autonomous agents based on the pragmatic notion of sign as defined by Charles Sanders Peirce.

Peirce's theory of sign provides a more comprehensive theoretical framework to deal with the problem of meaning than other purely symbolic approaches investigated so far. His notion of sign (the thing that signifies) does not suffer from these problems because it is firmly based on a triadic relation between Sign, Object (in the world) and Interpretant (the effect on the interpreter).

Our article is derived from previous work to describe semiosis (meaning processes) from a computational perspective [1] [2], in which we explained semiosis as a hierarchically organized process. In this context, meaning emerges from interactions between elements in different hierarchical levels. Also, a very general algorithm to simulate the emergence of meaning was sketched. However, at that time, we did not provide any experiment to discuss the results, and left many specific details open.

In order to observe the results of the proposed framework we derived an experimental protocol of the talking heads experiment [3], in which pairs of autonomous agents play a form of interactive language game.

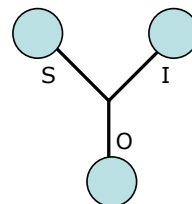


Figure 1 - Visual notation for a sign triad

In our case, the interactions result in the emergence of a simplified form of meaning. We argue that, with the proposed theory, meaning emerges within each agent without shared representations or external interference. These semiotic agents are able to ascribe meaning to their perceptions; they can exhibit semiosis, in a simulated form.

Organization of the Article

The article is structured as follows. Section 2 provides an overview of the underlying concepts of Peircean semiotics used in the model. Section 3 reviews our previous work on models of semiosis. Section 4 develops a model for simulation of Sign-Object. In section 5 we discuss the results of the experiment carried out.

2. OVERVIEW OF PEIRCEAN SEMIOTICS

The Peircean list of categories (Firstness, Secondness, Thirdness) is logically described as an exhaustive system of hierarchically organized classes of relations (monadic, dyadic, triadic) [9]¹. This system is the formal foundation of his architectonic philosophy and of his model of semiosis (meaning process). Peirce defined semiosis as an irreducible triadic relation [5]² between Sign-Object-Interpretant [4]³. According to Peirce, any description of semiosis involves a relation constituted by three irreducibly connected terms (S-O-I), which are its minimal constitutive elements [5]⁴ [6]⁵ (see Figure 1).

“A Sign is anything which is related to a Second thing, its Object, in respect to a Quality, in such a way as to bring a Third thing, its Interpretant, into relation to the same Object, and that in such a way as to bring a Fourth

¹ See page 14.

² See §171.

³ See vol. 1, §363; and vol. 8, §331.

⁴ See vol. 2, §242 and §274.

⁵ See vol. 318, §81

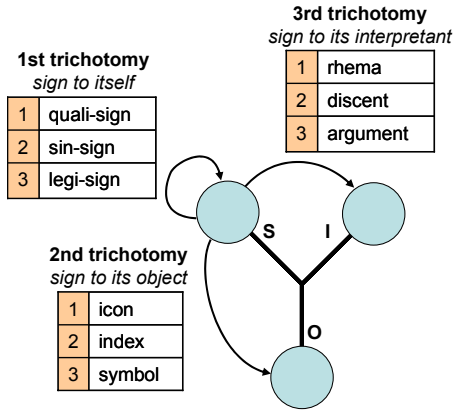


Figure 2 - General form of the sign and its main trichotomies

into relation to that Object in the same form, ad infinitum” [4]⁶.

Classes of signs

Peirce developed several classifications of signs based on different trichotomies [5]⁷ which are aspects according to which semiosis can be observed, and can be translated into questions: (i) “What is the relation of the Sign with itself?”, 1st trichotomy; (ii) “What is the relation between the Sign and its Object?”, 2nd trichotomy; (iii) “What is the relation between the Sign and its Object for its Interpretant?”, 3rd trichotomy. In this article we are concerned with only two (first and second) of the three trichotomies explained in Figure 2 and Figure 3.

In the next section, we investigate these definitions and identify the major theoretical constraints required to simulate semiosis.

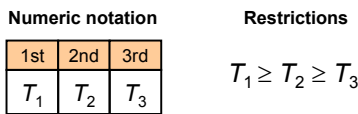


Figure 3 - Numerical notation for sign classes

3. SEMIOSIS: SIGN REPRESENTING ITS OBJECT

This section reviews the work on the theoretical framework extensively described in [2] and later in [1].

Levels of Semiosis

Semiosis is a dynamical process that happens *in time*. Triads are appended to the chain of triads according to the constraints of semiosis given in [1], that is:

$$\dots \rightarrow (S_{i-1} O_{i-1} I_{i-1}) \rightarrow (S_i O_i I_i) \rightarrow (S_{i+1} O_{i+1} I_{i+1}) \rightarrow \dots$$

This dynamical process happens at three levels: (i) *macro-semiotic level*: network of evolving chains of triads; (ii) *focal-semiotic level*: each chain of triads is simulated; and (iii) *micro-semiotic level*: relations of determination between triad elements (S-O-I) are simulated.

How a Sign Represents Its Object?

A Sign stands for its Object in three possible modes:

- (i) *intrinsic quality* of S (first term dependent) – such as the relation between a photograph of a cat and the cat itself. The Sign (photograph) shares an intrinsic quality (shape of the body, color, etc.) with the Object (cat), which means that the photograph, in a certain way, represents the real cat;
- (ii) S-O *relational quality* (first-second relation dependent) – such as the relation between smoke and fire. The Sign (smoke) shares an efficient causal relation with its object (fire), because we had a previous experience in which we perceived fire just after we perceived smoke. This means that smoke, in a sense, represents fire; and finally
- (iii) *imputed quality* by I to S-O (third term dependent) – such as the relation between the word “car” and its meaning, a typical car. The sign (word “car”) represents the object (car) by a convention or habit (here given by I).

Basically, this is strongly linked with the second trichotomy illustrated in Figure 2, which is one of the most important according to Peirce.

General Algorithm for Semiosis

In [2] we elaborated an algorithm to simulate the production of a triad. All steps are summarized below:

- 1) Choose a collection of potential signs $S' = \{s'_i\}$;
- 2) Choose one potential sign s' from this collection;
- 3) Propose a potential object o' and a potential interpretant i' , such that there exists a relation in one of the three possible modes. Then, we assume that o' determines s' relatively to i' .

Potential signs, objects and interpretants are things that may become effective if interpreted as such. In practical terms this depends on the overall result of the simulation. In the next section we will provide further details about the real realization of these steps in an artificial system.

⁶ See vol. 2, §92.

⁷ See vol. 2, § 289, §99, § 478, §91

4. SEMIOTIC INSPIRED ARCHITECTURE

In this section we extend the concepts reviewed in Section 3 in order to provide a real implementation. First, we develop a mathematical model for the modes of relation between Sign and Object (second trichotomy) and then propose a minimum agent architecture in which the kind of phenomena described so far can be observed. For the purpose of current work symbols (in Peirce's terms) will not be considered, and will be subject of investigation in future developments.

Notice that we don't intend to say that this computational approach is the only way to observe semiosis in artificial systems. As one can see, we left most complicating factors out and kept the main focus on the problem of emergence of meaning in its simplest form. What we do claim, indeed, is that this approach may provide many valuable insights on how artificial intelligence may escape some of its current limitations.

Icons

An icon is a sign in a certain way similar (or analogous) to the object it represents. This similarity means that a sign has the same properties of its object. This notion is strongly linked with the incapacity of the interpreter to distinguish S from O with regard to an aspect of S. For instance, consider a mouse that interprets a photo of a cat as an icon of a cat. It is clear that this photo has only some of the many properties presented by the cat itself, but is obvious that there are many features not conveyed by the photograph such as smell, sound, etc. This leads us to conclude that iconicity may be associated to a partial view that S expresses about the object.

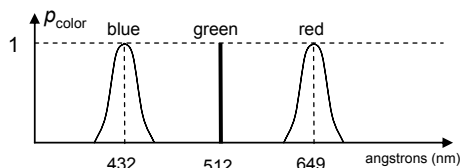


Figure 4 – Sample properties

In order to provide a formal definition for iconic signs, we introduce the concept of *fuzzy entity*, which are “things” with properties that represent entities' qualities. Some examples (see Figure 4) of properties are color, weight, texture, structure, etc. Also, consider a set S that contains all universes of discourse, in the form $S = \{S_1, \dots, S_p\}$. Then, an *entity property* is defined as a fuzzy set v_i defined in some $S_j \in S$. Thus, an entity is described by a tuple v :

$$v = (v_1, v_2, \dots, v_n)$$

where

$$v_i = S_{\alpha(i)} \rightarrow [0,1], \text{ or, alternatively}$$

$$v_i \in 2^{S_{\alpha(i)}}, \text{ where } \alpha \text{ is a property descriptor}$$

$$\alpha: \{1, \dots, n\} \rightarrow \{1, \dots, p\}$$

Likewise, an *aspect* is defined as a partial view of S . This allows us, for example, to focus on some types of properties, such as color and position, and ignore others, such as weight and temperature. Formally, an aspect is defined as a tuple with the indices of the universes S_i that form the aspect:

$$\gamma = (\gamma_1, \dots, \gamma_q)$$

where

$$\gamma_i \in \{1, \dots, p\} \text{ and } q \in \{1, \dots, p\}$$

$$\forall i, j \quad (\gamma_i = \gamma_j) \Leftrightarrow i = j$$

It then follows from this definition that an aspect represents a sub-space of the space given by all elements in S . This space is expressed as:

$$\mathbf{V}^\gamma = 2^{S_{\gamma_1}} \times 2^{S_{\gamma_2}} \times \dots \times 2^{S_{\gamma_q}}$$

Now, assuming that both⁸ v and w have all S_i indicated by γ , we are able to provide a very general definition of iconicity, in the form:

$$PIC(v, w)^\gamma: \mathbf{V}^\gamma \times \mathbf{V}^\gamma \rightarrow [0, 1]$$

PIC should be implemented as a similarity measure that is applied to the projection of v and w onto the space restricted by the aspect γ ⁹. This means that v has an iconic relation with w , considering that $S \equiv v$ and $O \equiv w$. Technically speaking, the potential of iconicity varies from a bottom limit 0 in which v and w are absolutely distinct to an upper limit 1 in which v and w cannot be distinguished at all.

Indices

An index is a sign that represents its object by means of a cause-effect relation. That is, there must be something, like a past event, that relates one to the other. This relation represents a causal connection, space-time correlation, or physical connection. In this context, an *event* is anything that can be perceived by an interpreter as a specific thing (a singular, a particular or *token*). Notice that we consider only causality that may be identified in space/time.

As just explained, an index, as such, must be an indication that two things are causally related. To do so, we propose a formal model to characterize this kind of relation and then a method by which they can be synthesized in computers.

⁸ It is implicit that v and w must have some type of quality in common to be compared.

⁹ That is, both tuples compared in some sub-space they may share with each other.

It is plausible to assume that an observer that perceives an event at a certain instant of time keeps¹⁰ a memory of this event for a certain duration $\Delta t > 0$. Being vivid within the interpreter, this event can function as a building block for indexical signs. The idea is that it can be related to other events by possible causal relations. To do so we introduce the notion of *potential of evocation* (*PEV*), which is a measure of the possibility to relate a given event to any other within the interpreter. This value is expressed as a mapping from a universe of discourse **T** (time) to the unit interval, and can be treated as fuzzy set.

$$PEV: \mathbf{T} \rightarrow [0, 1]$$

In practical terms, as one can see in Figure 5, it is reasonable to assume that a potential of evocation assumes its maximum value when the event is presented to the interpreter and then tends to decrease monotonically to zero from this moment on (if this is was last evocation). However, there's no other specific requirement on the shape of this fuzzy set. It may assume any form.

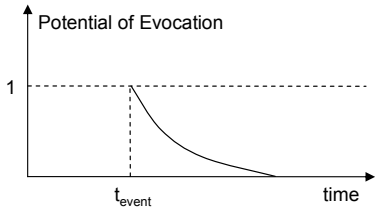


Figure 5 - Typical potential of evocation

An indexical sign, as defined earlier in this section, is a sign that is related to its object as a result of a causal relation. So, the problem boils down to the identification of causal relations between possible signs and possible objects. We propose that distinct events can be seen as causally related as long as the similarity between their *PEV*'s is not zero. That is, the more similar their *PEV*'s, the higher will be the strength of their indexical relation (defined here as the *potential of indexicality*, or simply *PIN*). In fuzzy set theory there are many ways to evaluate the similarity of two fuzzy sets. Typical choices are possibility and necessity measures [10]. In mathematical terms we have:

$$PIN(v, w) = \text{similarity}(PEV_v, PEV_w) \in [0, 1]$$

Consider for example the case described in Figure 6 in which we can see three curves: PEV_{button} , PEV_{food} , and $POSSIBILITY(PEV_{button}, PEV_{food})$. As one can see, the interpreter has experienced two events (at t_{button} and t_{food} respectively). The similarity between the two initial curves is an indication that these two event may be correlated in time.

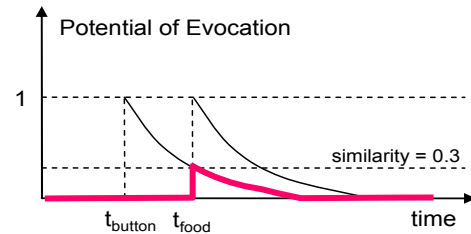


Figure 6 – Sample case of potential indexicality

Preliminary Agent Architecture ⁽⁶⁾

In order to simulate the models of iconicity and indexicality developed earlier in this section, we must have some sort of cognitive apparatus. Much work has been already done in this field, and complexity may grow very fast. For this work we sketched a minimum agent architecture, which is slightly based on the work of Steels in his famous Talking Heads experiment [3]. Our architecture has

- a *perception* module which converts external signals captured in the environment, performs segmentation and produce perceptions (modeled as fuzzy entities);
- a *relation engine* module which implements a version of the meta-algorithm cited in Section 3 (see [1] and [2]), producing new interpretations from perceptions and current accumulated experience;
- an *effector* module that affects the world as a result of some interpretation; and finally
- a *pragmatic* module which carries out the scripts required for the experiment.

In this example we are not interested in syntactic relations in internal structures, because we assume that this is not required for the emergence o meaning.

5. SIMULATED SEMIOSIS: THE FIRST EXPERIMENTAL APPROACH

In this section we describe a simple experiment in which we apply the models introduced in [1] and [2], and further improved in this Section 4.

Inspiration: The Talking Heads Experiment

The underlying formulation of the problem is derived from the Talking Heads Experiment, which is extensively described in [3], and deals with the emergence of meaning from interactions between pairs of robots (in the original case physical robots).

The environment is based on the GEOMS world, which is a simple kind of white board that have simple objects (geometric forms such as triangles, rectangles and circles).

¹⁰ In some sort of short term memory.

Elements may vary in size, position and grayscale. Agents perceive the world through sensory channels, which encapsulate perceptions of things in the world.

The original game (the guessing game) is played by a speaker and a hearer. The speaker chooses an object as the topic (his focus of attention) and gives a verbal hint about the selected topic. Then, the speaker guess which of the geometric forms represents the hint. The game succeeds if the hearer identifies the correct topic by “pointing to it”¹¹. The game proceeds with agents changing their roles during many rounds.

Sign Relations as a Graph

Traditionally, the sign is depicted as a tripod. However, it is of paramount importance to be able to express in a clear way the internal details of the sign. The notation shown in Figure 7 allows the reader to represent the categories of the first two trichotomies, which are the focus of the example.

Each node of the graph is a fuzzy entity, and are primarily created by the sensory module. Each one has properties from the sensory channels. The idea is that these entities form a graph of representations that is in fact the experience of the agent. Another point is that each node has an associated *PEV* that tends to decrease if the entity is not evoked (that is, used in some new representation).

Our Experimental Protocol

In this example, in order to keep things as simple as possible and yet with sufficient theoretical constraints to validate the results, we restricted the formulation of problem. At the present time we are not interested in the emergence of language *per se*, but in the emergence of relations that exhibit at least some sort of indexical competence. The apparent simplicity of the question (at least for a naïve reader) eclipses many subtle arguments typically ignored in similar experiments, including [3] itself. We do argue that the emergence of such phenomena is not obvious.

Initially, assume that: (i) the experiment start with a “virgin” artificial agent labeled AG_A and an human experimenter labeled AG_H ; (ii) AG_A can perceive objects and verbalizations by two groups of sensory channels (let’s say: HPOS/VPOS/GRAYSCALE, and VERBAL). Also, the white board starts empty. The steps are as follows (see Figure 8 for a graph view).

- *Instant 1* – AG_A is constantly reacting to visual and sound stimuli. As the board is still empty no perception is performed.
- *Instant 2* – AG_H put 2 objects on the white board. The sensory module of AG_A perceive 2 things. These



Figure 7 - general graph notation of the 1st and 2nd trichotomies

“things” are fuzzy entities in the form $v = (hpos, vpos, grayscale)$, labeled e_a and e_b . In terms of the first trichotomy, these elements are of the category of singulars (because every perception is unique), which means that they are potentially sin-signs. Additionally, the *PEV* of e_a and e_b become PEV_0 ¹².

- *Instant 3* – AG_H point to the figure in the board represented internally by e_a , and, approximately at the same time, generates an audible verbalization “yada”. The sensory module of AG_A generates three perceptions given by e'_a , e'_b , and e_c (this one represents the verbalization). Assuming that the world has not changed, the agent will notice the complete similarity ($PIC \approx 1$) between e_a and e'_a , and e_b and e'_b . This will cause these new entities to be considered the same (referring to the same “real” object) and then merged into e_a and e_b . Additionally, $PEV(e_a)$ and $PEV(e_b)$ are kept at a certain level as they were evoked. However, $PEV(e_a)$ stabilizes at a higher level than the others because it is the focus of attention (being pointed). So, the following is true: $PEV(e_a) > PEV(e_b)$. As before, e_c is an existent, and then a sin-sign. At this moment, $PEV(e_a)$ and $PEV(e_c)$ are high enough to result in a possible indexical relation between e_a and e_c . In another words, these two entities are now part of an indexical sin-sign (or 22)(a singular, e_c , represents by means of a causal connection its object, e_a). This is the first interesting result.
- *Instant 4* – Consider the previous step all over again. As a result, we will have another indexical sin-sign that associates a specific verbalization to a specific object. If this continues, more and more indexical sin-signs will be added. This may trigger the emergence of an indexical legi-sign, a sign that is in some sense a law and represents an object by causal connection. That is, this legi-sign is the generalization of many specific (and similar signs) related to the same object in the same way. This is the element “E” in Figure 8.
- *Instant 5* – AG_A perceives e_a , e_b , and e_f (another verbalization “yada”). In this case there’s no focus of attention. Now, the indexical sin-sign that relates e_c to e_a acts as law and brings e_f to the same relation with e_a ,

¹¹ Actually, real robots must have some way to detect the location in the white board to where the other robot is looking at.

¹² This is the initial value of *PEV* when the fuzzy entity is first evoked (created).

such that the object given by e_a is evoked again (in numerical terms, $32 \rightarrow 22$).

- *Instant 6* – The converse of the last discussion is true. That is, if e_a is pointed to AG_A , then e_c may be evoked as well (AG_A may speak “yada”).

What we have shown with these arguments is that AG_A , given all required constraints, is able to construct internal representations in terms of iconic and indexical relations without any direct interference of AG_H , except from interactions in the world.

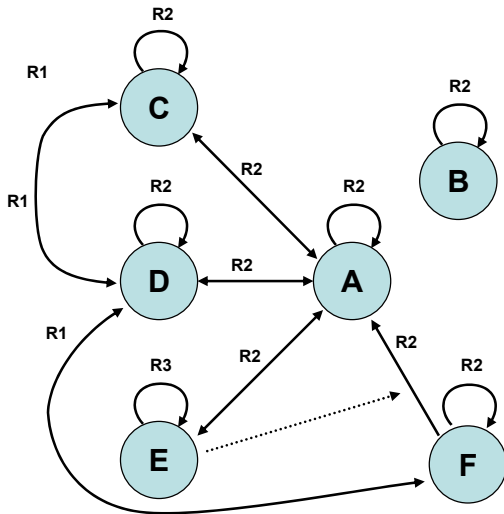


Figure 8 - Graph of relations for the discussed example

6. CONCLUSIONS

In this article we proposed a mathematical model for the simulation of relations between sign and object, as developed by C.S Peirce. We have also shown and discussed in detail a simple experimental protocol to simulated semiosis by means of the emergence of sign-object relations.

Future Work

A computer implementation is required to observe the autonomous emergence of meaning. In doing so, many computational problems arise, such as the combinatorial explosion in the number of options presented to the algorithm (combination of fuzzy entities, aspects, modes of relation, etc.). For the last but not the least, it is required a proper model for symbolic competences, which is a crucial requirement for certain abilities [6].

Acknowledgments

Antônio Gomes and Ricardo Gudwin are grateful to CAPES/Brazil. João Queiroz thanks to FAPESP for the grant 02/09763-2.

REFERENCES

[1] Antônio Gomes, Charbel Niño El-Hani, Ricardo Gudwin, & João Queiroz (2004), Towards Emergence of Meaning Processes in Computers from Peircean Semiotics, *Proceedings of European Computing and Philosophy Conference 2004* (forthcoming), Pavia, Italy, Collegio Ghislieri, June 3-5, 2004.

[2] A. Gomes, R. Gudwin, J. Queiroz (2003) - On a Computational Model of the Peircean Semiosis - *Proceedings of the IEEE International Conference on Integration of Knowledge Intensive Multi-Agent Systems - KIMAS'03*, 30 September - 04 October, 2003, Cambridge, MA, USA, pp. 703-708. ISBN 0-7803-7958-6.

[3] Steels, L. (1999). *The Talking Heads Experiment: Volume I. Words and Meanings*. Pre-Edition. VUB Artificial Intelligence Laboratory, LABORATORIUM, Antwerpen.

[4] PEIRCE, C.S. 1931–1935. *The collected papers of Charles Sanders Peirce*. Electronic edition reproducing Vols. I–VI [C. Hartshorne & P. Weiss (Eds.), Cambridge: Harvard University Press, 1931–1935], Vols. VII–VIII [A. W. Burks (Ed.), same publisher, 1958]. Charlottesville: Intelix Corporation. (Here referred as CP, followed by volume and paragraph number.)

[5] Peirce, Charles S. (EP1, EP2). *The Essential Peirce. Selected Philosophical Writings*. Vol. 1 (1867–1893) (ed. by Nathan Houser and Christian Kloesel). Vol. 2 (1893–1913) (ed. by the Peirce Edition Project). Bloomington and Indianapolis: Indiana University Press, 1992 and 1998.

[6] Deacon, T. (1997). *Symbolic Species: The Co-evolution of Language and the Brain*. New York: Norton.

[7] Harnad, S. (1990). *The symbol grounding problem*. *Physica D* 42, 335-346.

[8] Farias, P. & Queiroz, J. (2003) On diagrams for Peirce’s 10, 28, and 66 classes of signs. *Semiotica – Journal of the manuscripts and letters from C.S Peirce*: MS and L according to the catalogue]

[9] HOUSER, N. (1997). *Introduction: Peirce as a logician*, In: *Studies* 147 (1/4): 165-184.

[10] W. Pedrycz, F. Gomide. *An Introduction to Fuzzy Sets*. The MIT Press, 1998;