A Set-Theoretic Predicate for Semantics in Natural and Formal Languages

Adonai S. Sant'Anna Otávio Bueno
Newton C. A. da Costa

Department of Mathematics, Federal University of Paraná, Curitiba, PR, 81531-990, Brazil

Department of Philosophy, University of Miami, Coral Gables, FL 33124-4670, USA

Department of Philosophy, Federal University of Santa Catarina, Florianópolis, SC, 88040-900 Brazil

Abstract

We present an axiomatic framework for semantics that can be applied to natural and formal languages. Our main goal is to suggest a very simple mathematical model that describes fundamental cognitive aspects of the human brain and that can still be applied to artificial intelligence. One of our main results is a theorem that allows us to infer syntactical properties of a language out of its corresponding semantics. The role of pragmatics in semantics in our mathematical framework is also discussed.

Key words: Semantics, syntax, language.

1 Introduction

How did you learn English so that you are able to read this paper? Did you start with syntax?

In this paper we develop a specific formal approach to the semantic counterpart of natural and formal languages. By 'semantics' we understand a systematic correspondence between strings of a given language L and elements from a domain Δ that has no element in common with L. The intuition behind this is that such a correspondence allows us to endow a string with a specific meaning or class of meanings. We recognize that this understanding of semantics is not precise. Nevertheless it is somehow in accordance with the broad linguistic view that states: "semantics is the study of meaning in language" (Hurtford et al., 2007). In this sense, *language* and *meaning* work in two different domains, but somehow related through semantics.

The use of mathematical models for dealing with natural languages is not new. It is well known that generative grammars are quite useful for the study of syntax of natural languages (Chomsky, 1956; Chomsky, 1959; Dougherty, 1994), computer science (Davis et al., 1994), and even music (Steedman, 1989). Nevertheless, if we are interested in capturing formal aspects of natural languages, we cannot ignore either semantics or pragmatics. There are some interesting approaches to the semantic aspects of natural languages (Chomsky, 1972; Heim & Kratzer, 1998; Thomason, 1974). Heim and Kratzer's book (1998), in particular, is based on Gottlob Frege's vision about semantics of natural language, which is strongly supported by the usual set-theoretic approach to functions. At the end of this paper we briefly compare our proposal to others.

Here we introduce an axiomatic system for semantics, without any immediate concern with the pragmatic aspects of natural languages. We are not suggesting that semantics can be fully understood without considering pragmatics. After all, consider the following sentence: "Do you know what time it is?" This can be interpreted either as a request for information or as a request for someone to leave. The meaning of a such a sentence depends on a social context characterized by the pragmatic dimension of natural languages. Nevertheless, at the end of this paper we sketch a possible way to accommodate, at least in part, the pragmatic dimension of natural languages, which is an endeavor we intend to pursuit in detail in the future. First we need to establish the basics for our semantic approach, independently of any pragmatics. One of our main goals is the future development of a computer software that simulates social interactions among communicating individuals without any human assistance.

There have been reported experimental evidences that some neurons in the temporal cortex of the human brain are highly selective for object categories such as faces and hands (Desimone et al., 1984) and even specific people (Quiroga et al., 2005). Such results inspired the creation of a software capable to build high-level, class-specific feature detectors from unlabeled images (Le et al., 2012). The main result in (Le et al., 2012) was that such a software was able to identify (from a database of images) cat faces and human bodies without the need for any programmer to label which images were supposed to contain faces or bodies. In other words, machines are potentially able to create object categories, just as the human brain does. According to the terminology of the authors in (Le et al., 2012), their software "learns the concept of faces".

We propose the use of such object categories as an ingredient in an axiomatic framework for semantics in natural languages. This may be an important step toward the conception of machines that behave like human beings, from the cognitive point of view. In other words, it is not enough that a machine learns the concepts of faces or human bodies. It is necessary that the machine be able to communicate what it learned, without any human assistance.

2 Our Proposal

Let A be a finite set, called *alphabet*. The set of all *words*, or strings of elements of A, is denoted by A^*. A *language* L on A is any subset of A^*. We can understand the juxtaposition of words as a binary operation on A^*. In this sense, A^* endowed with such a binary operation is a free semigroup, where the empty word $\emptyset$ is the neutral element. For the sake of abbreviation, we say that L is a language if there is an alphabet A such that L is a language on A. Now we can define a *meaningful language*, where we consider that $\emptyset \in L$.

Definition 1 $\Lambda = \langle L, \Delta, s \rangle$ *is a* meaningful language *iff*

ML1 *L is a language.*

ML2 *$\Delta \neq \emptyset$ and $\Delta \cap L = \emptyset$.*

ML3 *$s : L \to \wp(\Delta)$ is a function, where $\wp(\Delta)$ is the power set of Δ.*

ML4 *If α, β, γ, and $\alpha\gamma\beta$ are strings of L, then $s(\alpha\gamma\beta) \subseteq s(\gamma)$.*

ML5 $s(\emptyset) = \Delta$.

Particularly, L may be a language generated by a phrase-structure grammar (Suppes, 2002) or even a set of strings of elements of a vocabulary in the sense presented by (Mendelson, 1997). In either case we must admit that the empty string $\emptyset$ belongs to L. The set Δ is the *domain of application* of the meaningful language Λ. Intuitively speaking, it corresponds to the set of categories [in the sense of (Le et al., 2012; Desimone et al., 1984; Quiroga et al., 2005)] of specific objects that endow meaning to a string of L. Within the metamathematical framework, we can interpret the elements of Δ as "Barack Obama's car", "Hilary Clinton's car", "the reader's favorite shirt", "the lion of MGM", "Isaac Newton's first toy", and so on. The intuitive interpretation of Δ depends on the intended application of our framework. Such a procedure is quite usual, e.g., in theoretical physics. In the axiomatic system of classical particle mechanics discussed by Patrick Suppes (2002) we may interpret particles as specific planets, bicycles or even apples. Nevertheless, according to axiom **ML2** there is no element in common between Δ and L. From now on we refer to the elements of Δ as *meanings*. The function s is the *semantic function*. Any string of L corresponds to one and only one subset of Δ. For example, if the language L is English, then we may have $s(car) = \{$Barack Obama's car, Hilary Clinton's car, the car that almost hit John Doe last month, ...$\}$. If $m \in s(\alpha)$ we say that m is a *possible meaning* for α, and that $s(\alpha)$ is the set of all possible meanings for α. Axiom **ML4** is the heart of our proposal. It simply states that longer strings (obtained by adding strings either to the left or to the right) of a language L never correspond to wider sets of meanings. For example, the string *car* may correspond to a lot of meanings, as we illustrated above. But the longer string *blue car* may correspond to a set that has all the elements of $s(car)$ except "the car that almost hit John Doe last month". That would happen if the car that almost hit John Doe last month was not blue, but red. Since axiom **ML4** does not demand that $s(\alpha\gamma\beta)$ be a proper subset of $s(\gamma)$, then we allow the possibility that $s(\alpha\gamma\beta) = s(\gamma)$. In this case we say that $\alpha\gamma\beta$ and γ correspond to *redundant meanings*. Finally, the last axiom states that the empty string corresponds to all possible meanings of Δ. Roughly speaking it says that silence may mean anything.

We need to say something more about Δ. We speculate that all concepts used in natural languages correspond to object categories in an extended version of the usual sense in neuroscience (Quiroga et al., 2005) [and applied to computer science (Le et al., 2012)]. So, according to axiom **ML3**,

$$s : L \to \Im(s)$$

is a surjective function, where $\Im(s)$ is the image of function s. We call the set $\Im(s)$ the class of all object categories of Λ.

Theorem 1 *If α, β, γ, $\alpha\gamma$, and $\gamma\beta$ are strings of a language L in a meaningful language $\Lambda = \langle L, \Delta, s\rangle$, then*

1. *$s(\alpha\gamma) \subseteq s(\gamma)$, and*
2. *$s(\gamma\beta) \subseteq s(\gamma)$.*

Proof: Item (1) is proved from axiom **ML4**, when $\beta = \emptyset$. Item (2) is proved from axiom **ML4** when $\alpha = \emptyset$.

Before we state the next theorem, let us consider an example of a meaningful language.

Example 1 *Let L be the language of all possible strings with letters A, B, and C, and $\Delta = \{1, 2, 3\}$. Then $\langle L, \Delta, s\rangle$ is a meaningful language if $s(A) = \{1, 3\}$, $s(B) = \{2, 3\}$, $s(C) = s(\emptyset) = \{1, 2, 3\}$, $s(ACB) = s(AC) = s(CB) = \{3\}$ and for the remaining strings α we have $s(\alpha) = \emptyset$. The reader can easily verify that all axioms* **ML1**$\sim$**ML5** *are satisfied. In this example, $\Im(s) = \{\{1, 2, 3\}, \{1, 3\}, \{2, 3\}, \{3\}, \emptyset\}$. This illustrates the fact that object categories are not equivalence classes with respect to any binary relationship defined between elements of Δ.*

Theorem 2 *Let $\Lambda = \langle L, \Delta, s\rangle$ be a meaningful language, where α, β, and γ are strings from L. If there is a meaning (an element of Δ) m such that $m \in s(\alpha\gamma\beta)$ but $m \notin s(\alpha\beta)$, then $m \in s(\alpha) \cap s(\beta) \cap s(\gamma)$ but $m \notin s(\alpha\beta\gamma)$ and $m \notin s(\gamma\alpha\beta)$.*

Proof: According to theorem 1 $s(\alpha\gamma\beta) \subseteq s(\alpha\gamma) \subseteq s(\alpha)$. Analogously, $s(\alpha\gamma\beta) \subseteq s(\gamma\beta) \subseteq s(\beta)$. And, $s(\alpha\gamma\beta) \subseteq s(\alpha\gamma) \subseteq s(\gamma)$. So, if $m \in s(\alpha\gamma\beta)$, then $m \in s(\alpha) \cap s(\beta) \cap s(\gamma)$. Using again theorem 1, we have that $s(\alpha\beta\gamma) \subseteq s(\alpha\beta)$ and $s(\gamma\alpha\beta) \subseteq s(\alpha\beta)$. So, if $m \notin s(\alpha\beta)$, then $m \notin s(\alpha\beta\gamma)$ and $m \notin s(\gamma\alpha\beta)$.

So, while theorem 1 states that longer strings, either to the left or to the right, narrow sets of possible meanings, this last theorem considers the case that the center-embedding of strings within strings does not necessarily narrow sets of possible meanings. Such a phenomenon is exactly what happens in example 1. The meaning 3 belongs to $s(ACB)$ but it does not belong to $s(AB)$ (since $s(AB)$ is empty). In other words, $s(ACB)$ is not a subset of $s(AB)$, despite the fact that ACB is longer than AB by the juxtaposition of C *between* A and B. Although our approach is not limited to languages generated by regular grammars (Suppes, 2002) (generative grammars which do not allow center-embedding of strings within strings), this last theorem states that the dynamics of center-embedding of strings within strings is nonlinear (in an intuitive sense) with respect to the semantic function s. For example, in the case of English, the possible meaning m of the string "John kicks the ball" is not necessarily a possible meaning for "John the ball". And if that is the case, m is not a possible meaning for "John the ball kicks" or "kicks John the ball", although m belongs to the classes of possible meanings of the strings "John", "kicks", and "the ball". Obviously we are considering that $\alpha =$ "John", $\gamma =$ "kicks", and $\beta =$ "the ball". Within this context, it seems reasonable to interpret theorem 2 as a way to (at least) partially infer some aspects of L's syntax. Furthermore, theorem 2 has another advantage. Even if function s is partially defined (we only know the images of some strings), other images may be inferred from what we already know. But, of course, this works only for the case in which some (previously unknown) images are the empty set.

Note that if a meaning m does not belong to $s(\alpha\beta\gamma)$, then m does not belong to $s(\delta\alpha\beta\gamma\varepsilon)$, which is the set of all possible meanings of a longer string than $\alpha\beta\gamma$ (by the juxtaposition of strings δ and ε to the left and to the right, respectively). One of the consequences of this is that a meaningless string (a string α such that $s(\alpha) = \emptyset$) cannot transform into a meaningful string (a string β such that $s(\beta) \neq \emptyset$) by the juxtaposition of new strings either to the left or to the right.

The next theorem refers to the case in which longer strings obtained from the center-embedding of strings within strings are associated to narrower sets of possible meanings.

Theorem 3 *Let $\Lambda = \langle L, \Delta, s\rangle$ be a meaningful language, where α, β, and γ are strings from L. If $s(\alpha\gamma\beta) \subseteq s(\alpha\beta)$, then $s(\alpha\gamma\beta) \subseteq s(\alpha\beta) \cap s(\gamma)$.*

Proof: Straightforward from the hypothesis and axiom **ML4**.

In other words, if the center-embedding of a string γ within $\alpha\beta$ narrows possible meanings for the new string $\alpha\gamma\beta$, then all possible meanings of the new string are common for both γ and $\alpha\beta$.

Before we discuss in more detail the syntax (grammars) in meaningful languages, it is worth making some complementary remarks about our semantic approach. In our mathematical framework, we can rigorously define some concepts usually discussed in the literature about semantics in natural languages. For example, a string α is *meaningless* iff $s(\alpha) = \emptyset$. A string α is *meaningful* iff $s(\alpha) \neq \emptyset$. Two strings α and β are *synonyms* iff there are strings γ and δ such that $s(\gamma\alpha\delta) = s(\gamma\beta\delta)$. A meaningful string γ is *ambiguous* iff there are meaningful strings α and β such that $s(\alpha) \subseteq s(\gamma)$, $s(\beta) \subseteq s(\gamma)$ and $s(\alpha) \cap s(\beta) = \emptyset$. So, auto-antonyms (e.g., in English, the words *fast*, *sanction*, and *stay*) are particular cases of ambiguous strings.

Antonyms are a bit more tricky. First, our axiomatic framework $\langle L, \Delta, s\rangle$ does not say anything about logical connectives, such as conjunction, disjunction, material implication or negation in the language L. Second, antonyms frequently depend on a given context. For example, the string "close" may be an antonym for "open" if we are comparing the strings "close the door" and "open the door". Nevertheless the same strings "open" and "close" are not antonyms if we are comparing the strings "you are close to me" and "you are open to me". We can admit that some languages do have special strings corresponding to logical connectives. We can then define antonyms by using such logical connectives. But besides the fact this maneuver would restrict too much the conception of antonym for those languages with such special strings, there still is the problem of context that was just noted. Moreover, we should not forget that logical connectives cannot be translated into natural languages as easily as it may seem. For example, a double negation in a natural language may be interpreted as an emphasis for a negation. So, the most we can say (for now) is: one necessary condition for strings α and β to be antonyms is that there are strings γ and δ such that $s(\gamma\alpha\delta) \cap s(\gamma\beta\delta) = \emptyset$. We do not dare (for now) to establish a sufficient condition for the relationship "α and β are antonyms".

Concerning logical connectives, it is worth remarking that in certain formal languages there are no logical connectives, even in cases of formal languages that capture significant aspects of modern mathematics (Tarski & Givant, 1987).

We can discuss the vagueness of strings of a given meaningful language $\Lambda = \langle L, \Delta, s\rangle$ as well. Consider the set $\Im(s)$ of all images of the semantic function s. Each element of $\Im(s)$ is an object category. Nevertheless, there are some concepts, quite usual in natural languages, that are well known to be vague: "young", "bald", "tall", and so on. One usual way to deal with vague concepts is through fuzzy set theory (Lee, 2005). So, for each string α we may define a fuzzy set $f_\alpha : s(\alpha) \to [0,1]$, in which $[0,1]$ is the set of real numbers between 0 and 1. So, if $m \in s(\alpha)$, then $f_\alpha(m)$ states how fuzzy or how vague is the meaning m in the object category $s(\alpha)$. If f_α is a characteristic function (in the usual set-theoretic sense), then α is not vague. Otherwise, α is said to be vague.

For example, the string $\alpha =$"bald" can be associated to an object category $s(\alpha)$, which can have many elements, including those corresponding to people. Two possible meanings m_1 and m_2 in $s(\alpha)$ can correspond to, respectively, "the actor Patrick Stewart" and "the President Barack Obama". In this case, we can have $f_\alpha(m_1) = 0.9$ and $f_\alpha(m_2) = 0.1$ in order to express the idea that it is much more plausible to associate the string "bald" to Patrick Stewart than to Barack Obama.

We do not intend to suggest from this brief discussion that the best way to deal with vagueness is through fuzzy set theory. We're just pointing out that such an approach is compatible with our axiomatic system.

3 Grammars

Let V be a finite, nonempty set, V^* be the set of all finite sequences of elements of V, and $V^+ = V^* - \{\emptyset\}$. It is well known (Suppes, 2002) that:

Definition 2 $G = \langle V, N, P, S\rangle$ *is a* phrase-structure grammar *iff*

PSG1 *V, N, and P are nonempty finite sets.*

PFG2 $N \subseteq V$.

PSG3 $P \subseteq V^+ \times V^*$.

PSG4 $S \in N$.

N is called the set of *nonterminal* symbols of V, and $V_T = V - N$ is the set of *terminal symbols* or words of V. P is the set of *productions*. If $(\alpha, \beta) \in P$ we denote this by $\alpha \to \beta$. If γ and δ belong to V^* then $\gamma\alpha\delta \Rightarrow_G \gamma\beta\delta$ iff

$\alpha \to \beta$. Finally, if γ_1 and γ_m belong to V^*, then $\gamma_1 \Rightarrow_G^* \gamma_m$ iff $\exists \gamma_2 \ldots \gamma_{m-1}$ such that $\gamma_1 \Rightarrow_G \gamma_2$, ..., $\gamma_{m-1} \Rightarrow_G \gamma_m$.

So, the *language generated by a phrase-structure grammar G* is:

$$L(G) = \{\alpha / \alpha \in V_T^* \wedge S \Rightarrow_G^* \alpha\}.$$

where V_T^* is called the set of lexical strings.

Definition 3 *Let A be an alphabet, A^* be the set of all strings of elements of A, and $\Lambda = \langle L, \Delta, s \rangle$ be a meaningful language. Then*

1. *If L is a proper subset of A^* ($L \subset A^*$) such that $\emptyset \in L$, and for all $\alpha \in L$ we have $s(\alpha) \neq \emptyset$, then $\Lambda = \langle L, \Delta, s \rangle$ is a* meaningful language with a nontrivial grammar*.*

2. *If for all $\alpha \in A^*$ we have $s(\alpha) \neq \emptyset$, then $\Lambda = \langle A^*, \Delta, s \rangle$ is a* meaningful language with a trivial grammar*.*

Meaningful languages with either trivial grammars or nontrivial grammars are called g-meaningful languages (g stands for grammar). It is easy to verify that any meaningful language $\Lambda = \langle A^*, \Delta, s \rangle$ with a trivial grammar can be associated to an extended version of phrase-structure grammar $G = \langle V, N, P, S \rangle$, if A is a finite, nonempty set. If $N = \{S\}$, $P = \{(S, \alpha) / \alpha \in A^*\}$, and $V_T^* = A^*$, then G generates A^*, if we generalize phrase structure grammars in order to allow that P is infinite. Naturally we are not suggesting that this is the only solution to the proposed problem. After all, it is an unsolvable problem (Hopcroft & Ullman, 1969) whether or not a given phrase-structure grammar generates the set of all strings that belong to V_T^*, which, in this case, is equal to A^*.

Nevertheless it is highly difficult to solve the more interesting problem of associating meaningful languages with nontrivial grammars to phrase-structure grammars. If L is a language in a g-meaningful language $\Lambda = \langle L, \Delta, s \rangle$, then there is no effective procedure (Hopcroft & Ullman, 1969) to ensure that a specific phrase-structure grammar $G = \langle V, N, P, S \rangle$ generates L, since (according to axiom **ML1**) L is an arbitrary language. In the case discussed in example 1 the relationship between our mathematical framework and phrase-structure grammars is no mystery at all, as we can see in the next example.

Example 2 *The ordered triple*

$$\langle \{\emptyset, A, B, C, ACB, AC, CB\}, \{1, 2, 3\}, s \rangle$$

is a meaningful language with nontrivial grammar if s is the same function given in example 1, restricted to the new set $\{\emptyset, A, B, C, ACB, AC, CB\}$. So, one corresponding phrase-structure grammar $G = \langle V, N, P, S \rangle$ is given by

$$V = \{S, A, B, C, ACB, AC, CB\},$$

and

$$P = \{(S, \emptyset), (S, A), (S, B), (S, C), (S, ACB), (S, AC), (S, CB)\}.$$

In this case, $L(G) = \{\emptyset, A, B, C, ACB, AC, CB\}$.

For a more exciting example, with an infinite language L, consider the following:

Example 3 *Let $\Lambda = \langle L, \Delta, s \rangle$ be a meaningful language in which:*

$$L = \{\emptyset, A, B, AB, ABB, ABBB, ABBBB, ...\},$$

Δ is the set ω of all natural numbers, and s is given by

$$s(\alpha) = \begin{cases} \Delta - \{1, ..., \text{length of } \alpha\} & \text{if } \alpha \neq \emptyset \text{ and } \alpha \neq B \\ \Delta & \text{if } \alpha = \emptyset \text{ or } \alpha = B \end{cases}$$

So, e.g., $s(ABB) = \{0, 4, 5, 6, 7, ...\}$. It is easy to verify that Λ is a meaningful language with a nontrivial grammar (definition 3, where the alphabet is given by the set $\{A, B\}$).

One natural phrase-structure grammar $G = \langle V, N, P, S \rangle$ that generates L is given by

$$V = \{S, X, Y, A, B\},$$

$$N = \{S, X, Y\},$$

$$P = \{(S, \emptyset), (S, A), (S, B), (S, XY), (X, A), (Y, B), (X, XY)\}.$$

If $\{A,B\}^*$ *is the set of all possible strings of elements of* $\{A,B\}$*, then* $\Lambda' = \langle \{A,B\}^*, \Delta, s' \rangle$ *is a meaningful language if*

$$s'(\alpha) = \begin{cases} s(\alpha) & \textit{if } \alpha \in L \\ \emptyset & \textit{if } \alpha \in \{A,B\}^* - L \end{cases}$$

in which L *is the language in* Λ*. However,* Λ' *is not a meaningful language with a nontrivial grammar. Note that the function* s *in* Λ *is a restriction of the function* s' *in* Λ' *to the domain* $L \subset \{A,B\}^*$*. Moreover, this particular semantic function* s *can be recursively defined as follows:*

First we establish that:

1. $\emptyset \in L$*,* $A \in L$*, and* $B \in L$*.*
2. *If* $\alpha \in L$ *and* $\alpha \neq \emptyset$ *and* $\alpha \neq B$*, then* $\alpha B \in L$*.*

Next, we define that

1. $s(\emptyset) = s(B) = \Delta$ *and* $s(A) = \Delta - \{1\}$*.*
2. *If* $\alpha \in L$ *and* $\alpha \neq \emptyset$ *and* $\alpha \neq B$*, then* $s(\alpha B) = s(\alpha) - \{\textit{length of } \alpha B\}$*.*

So, from an intuitive point of view, what do we mean by the expression "meaningful language with a nontrivial grammar"? In this case, we mean that, for any string α *from* Λ' *(a natural extension of* Λ*)* $s'(\alpha) \neq \emptyset$ *iff* $\alpha \in L(G)$*, in which* G *is the phrase-structure grammar that generates the language* L *of* Λ*.*

However, in the general case, the expression "meaningful language with a nontrivial grammar" means (loosely speaking) a meaningful language $\Lambda = \langle L, \Delta, s \rangle$ in which the set of all meaningful strings $\alpha \in L$ (those α such that $s(\alpha) \neq \emptyset$) can be associated to some sort of syntax, as suggested by theorem 2. This syntax basically states that strings cannot be arbitrarily juxtaposed in order to form longer strings and still have some meaning. Language L from example 3 has clearly two categories of strings: $\{A\}$ and $\{B\}$. String A should always be the first one in any longer string, while string B never is. At most, B is meaningful when it is alone.

So, while a phrase-structure grammar G, by itself, does not allow to infer any information at all concerning the meaning of lexical strings of a language L generated by G, our semantic approach to any arbitrary language L allows to derive some information concerning L's grammar. This result is related to the fact that children start learning their natural language (according to the region where they mostly live) from the semantic and pragmatic points of view. Syntax is the last studied subject in order to understand the basics of any natural language. Even formal languages (in mathematics) are learned this way. No working mathematician starts his/her studies from the point of view of formal logic. The usual starting point is the study of calculus, linear algebra, differential equations, geometry and other subjects that have a strong connection with physical problems of the real world.

4 Pragmatics

In (Horn & Ward, 2006) pragmatics is referred to as the study of "context-dependent aspects of meaning which are systematically abstracted away from in the construction of content or logical form". This notion of pragmatics seems to be vaguer than the first concept of semantics we discussed in the beginning of section 1. But we are inspired by this notion in the following way. The meaning of strings of a given language depend on social interactions among those individuals who share either elements of this language or elements of the way such individuals perceive their surrounding and themselves. So, in this paper, we take pragmatics to be the study of this dependence of the meaning of strings on social interactions. We interpret the context-dependent aspects of meaning pointed out by (Horn & Ward, 2006) as social interactions, and we discuss this point in more detail below.

We consider that meaningful languages correspond to some cognitive aspects of human beings and even machines (properly programmed). What are those cognitive aspects? If a meaningful language $\Lambda = \langle L, \Delta, s\rangle$ is associated to an individual, then L is the language known by such an individual and Δ corresponds to the way that the individual perceives the environment and oneself. In a future paper we intend to characterize pragmatics through interactions among meaningful languages. If $\Lambda_1 = \langle L_1, \Delta_1, s_1\rangle$ and $\Lambda_2 = \langle L_2, \Delta_2, s_2\rangle$ are two meaningful languages, we can (in principle) define mathematical relations between Λ_1 and Λ_2, in the sense of formally supporting changes (the definition of new sets) in L_1, Δ_1, s_1, L_2, Δ_2, and s_2. These mathematical relations among two or more meaningful languages characterize a social context.

Let $\Lambda = \langle L, \Delta, s\rangle$ be a meaningful language. Consider, for example, the string “Barack Obama’s car”, supposing that L is English. From the point of view of our set-theoretic approach, there are at least two possible ways to define the function s:

1. s(Barack Obama’s car) can be identified with a set of still images of cars (or a specific category of cars); or
2. s(Barack Obama’s car) can be identified with the empty set.

Of course, there are other possibilities, but let’s stick to these two, for the sake of argument.

The first case can be intuitively interpreted as Λ is “imagining” that any car from s(Barack Obama’s car) can be Barack Obama’s car. The second case can be interpreted as Λ “does not even ‘know’ what a car is or who is Barack Obama”. Now, if we admit interactions among meaningful languages, Λ may change the function s as follows:

s(Barack Obama’s car) $= \{$A given category of Lincoln cars$\}$.

In regard to the first case, the new function s can be interpreted as the situation in which Λ “learned” that Barack Obama owns a Lincoln. With respect to the second case, Λ gives its first step toward the comprehension of what a car is.

However, there is a severe shortcoming in our vision of semantics if we are interested in applying it to natural and formal languages. Our mathematical framework is too general. And excessive generality can be easily understood as a synonym for uselessness. Our main hope for the actual usefulness of our mathematical model lurks exactly in pragmatics.

Consider L as English. Then $s(blue)$ can correspond to *everything* that is blue and *everything* that is not blue, as well; even if we restrict ourselves to the notion that “*blue*” refers only to a color. For practical purposes, each $s(\alpha)$ is a very large set, assuming that we are interested in applications of meaningful languages to familiar natural languages. That is the main reason why we did not define logical connectives in our approach.

We believe that such a limitation in our proposal is due to the fact that natural languages are not just languages, in the usual sense described in section 2. Natural languages are always endowed with some sort of *rationale*, that is, some kind of logic (very broadly understood as certain inferential procedures). And the main issue is that this rationale is not detached from the natural language itself. Language and inference processes (rationale) are entangled in one single “structure” usually referred to as natural language.

In order to address this problem, we intend to define a *thought function* t, which, intuitively speaking, corresponds to what an individual really means by uttering a string α in a conversation.

Since $\wp(\Delta)$ is a σ-algebra (in the usual sense of probability theory), then we can define a probability function $p : \wp(\Delta) \to [0, 1]$, which we call the *propensity* function. Such a function p may be restricted to domains $\wp(s(\alpha))$ for each string α from language L. It is worth noting that for each $\alpha \in L$ we have that $\wp(s(\alpha)) \subseteq \wp(\Delta)$. The *thought function* t is a function $t : L \to \wp(\Delta)$ such that:

$$t(\alpha) = \{M \in s(\alpha) \,/\, \text{for all } m \in s(\alpha) \text{ we have } p(\{M\}) \geq p(\{m\})\}.$$

Applied to a given string α, the thought function t is simply the collection of all meanings with maximum value for the propensity function. Such a propensity function applied to $\wp(s(\alpha))$ tells the probability that α means m, for each and every m that belongs to $s(\alpha)$.

In order to address (some aspects of) pragmatics, we note that a *communicating individual* I can be represented by an ordered pair $\langle \Lambda, t \rangle$, in which Λ is a meaningful language and t is a corresponding thought function.

If $t(\alpha)$ has a single element M, we say that the individual I is thinking of M when I either utters α or listens to the string α. If $t(\alpha)$ has two or more elements, I is thinking that α can equally well correspond to either of them. If $t(\alpha)$ is empty, we say that I does not understand α at all. Social interactions will change not only the elements of a meaningful language, but the thought function as well.

We believe that any rationale involved in a natural language should be characterized by the thought function, and not by the language L as an ingredient of a meaningful language Λ. Thus, the context-dependent aspects of meaning mentioned in (Horn & Ward, 2006) are defined by the thought function t. For example, if a communicating individual utters the word "blue", many people will expect the individual to be referring to something that is blue, rather than something that is non-blue. In other words, it is more probable that this communicating individual means something that *is* blue. This expected value for the meaning of "blue" is directly associated with pragmatics.

5 Discussion

In a recent paper (Le et al., 2012), as noted above, researchers from Google X Lab and Stanford University have reported an experiment in which a computer network was able to identify cat faces in a database originated from ten million YouTube stills. The most remarkable aspect of this experiment is that the computer network was not programmed to look for cat faces. Such machines have been described as having *learned* to identify this pattern of images. From the point of view of our approach, this means that the Google X Lab network of computers just created a small domain of application Δ. So, our hope is that our proposal can be helpful for the future conception of machines that are able to effectively "understand" strings from natural languages. If someone utters the string "cat" (from English), then a computer (in principle) may be able to associate such a string to a class of images that intuitively correspond to cats. Naturally any definition of a domain Δ, for practical purposes, demands much more than the classification of object categories based only on still images. Some "meanings" are very difficult to translate into images, such as the concepts of "tomorrow", "there", "dream", and "God", among other examples. Other concepts are usually derived from non-visual stimuli, such as the notions of "tasty", "sweet", "smelly", and so on. And other concepts, in turn, have no sensorial appeal whatsoever, such as "sets", "functions", and "topological spaces". Clearly, there is a lot of work to do until programmers are able to implement any notion of semantics in a machine, at least from the point of view presented in this paper, in order to create a machine that can effectively dialogue with a human being as we do.

Concerning human beings only, we consider that each person at each instant of time can be associated to a particular meaningful language $\Lambda = \langle L, \Delta, s \rangle$. The learning process of any individual is just any change in either L, Δ or s, as well in the thought function, as described in section 4. In a future paper, we intend to explore this possibility in detail.

In section 3, we made a brief comparison between our semantic approach and generative grammars, at least from the point of view of phrase-structure grammars, which allow to define a whole hierarchy of grammars well known in the literature. A few additional remarks are in order. First, as noted above, usually people learn languages invoking semantic and pragmatic considerations. Syntax is (quite) usually the last step in any language learning process. We may want to take the same path when we aim to understand formal aspects of natural languages. It is unclear why we should start these formal studies from the syntactic point of view. Second, our axiomatic system for semantics is partially based on the notion of *function* (the semantic function s), a strategic set-theoretic concept that plays a fundamental role not only in mathematics but in a huge variety of applications as well. For a detailed discussion of the role of functions in pure and applied mathematics, see (Sant'Anna & Bueno, 2014). However, generative grammars, such as phrase-structure grammars, are strongly based on a set of productions, which is a relation, not a function. This fact certainly compromises the applicability of generative grammars. Despite this fact, the semantic function properties demanded by axiom **ML4** also suggest a relation, since the behavior of s is settled by the $\subseteq$ relation between sets of possible meanings. We hope that the appropriate use of further functions in a detailed formal picture of pragmatics will make our proposal more easily adjusted for practical purposes. Third, the vocabulary of a generative grammar includes several non-lexical strings, which could be understood as elements from a metalanguage associated to the generated language through the axioms of phrase-structure grammars. In our approach, there is (in a sense) a metalinguistic ingredient as well, namely, the domain of application Δ. In this context, we understand the semantic function s as a systematic mean to promote the translation between strings from a given a language and the way we feel and understand the world and ourselves.

Finally, our axiomatic picture is quite different from other formal approaches to semantics in natural languages. The mathematical model discussed in (Heim & Kratzer, 1998), e.g., is truth-conditional, in the sense that the authors understand the meaning of a string as something directly related to its truth-conditions. Meaningful languages, however, have (in principle) no concern with any notion of truth. We simply state that a communicating individual attributes some meaning to a string, when presented to it. Our thought function, introduced in section 4, emphasizes that there can be one (or a few) most probable meaning(s) chosen by the communicating individual given some string. And we believe that the choice of a specific meaning is related to the social context, as described in section 4. If the social context is somehow related to a particular notion of truth (such as the correspondence between the language and the world), then the communicating individual will have its thought function affected by the relevant social context. However, since the present paper on meaningful languages focuses on semantics without immediate concern for pragmatics, we prefer to leave the proper examination of the notion of social context for a future paper.

Some semantic approaches were specifically developed for generative grammars (Chomsky, 1972; Schiffer, forthcoming). But, as we pointed out in section 3, our semantic approach allows us to infer some information about the syntax of the language under consideration. Moreover, our mathematical model is perfectly compatible with generative grammars, as we noted in section 3. Stephen Schiffer (in a forthcoming paper) states the generative grammar hypothesis as: "The ability of a speaker of a natural language L to understand sentences of L requires her to have tacit knowledge of a generative grammar of L, that being a finitely specifiable theory of L that generates one or more syntactic structures for each sentence of L and interprets those structures both phonologically and semantically". This is not the place to discuss whether the tacit knowledge of a generative grammar is needed nor not. But we do wonder whether this tacit knowledge is better understood as an evolutionary consequence of mankind's history of social interactions. If this is correct, then social interactions can be understood as the proper starting point for the development of human languages.

There is also a relation between our proposal and a much earlier attempt at the use of formal methods in natural language semantics introduced by Richard Montague (Thomason, 1974). According to Montague, it is possible to comprehend the syntax and semantics of both natural and artificial languages (studied by logicians) within a single mathematical theory. The main obvious difference between our axiomatic method and Montague's theory, however, is the fact that the latter depends on higher-order predicate logic and the lambda calculus. We propose that a simple and quite usual set-theoretic framework is good enough for dealing with this important problem. After all, if someone wanted to use a formal framework as a starting point for computer simulations of communicating individuals, we do not believe that Montague's approach would provide the proper setting. A simpler mathematical model would be recommended. And we hope that our approach is up to the task.

6 Conclusions

Our main results may be summarized as follows:

- Our semantic approach is applicable to any language in a very broad sense, including natural and formal languages. In the case of natural languages, our mathematical model is applicable as long as all (or, at least, most) of the concepts used in natural languages correspond to object categories, in an extended formulation of the usual sense invoked in neuroscience (Quiroga et al., 2005). In the case of formal languages, our approach is applicable as long as machines in the future are able to create object categories in an extended formulation of the sense presented by artificial intelligence researchers in (Le et al., 2012).

- Phrase-structure grammars do not allow, by themselves, to infer semantic aspects of generated languages. But our semantic approach to arbitrary languages allows one to infer some basic syntactic properties of such languages.

- Our semantic approach allows one to define rigorously concepts that are usually referred to in semantics, such as 'synonym' and 'ambiguity'. Nevertheless, it does not seem to allow to define what are 'antonyms'. Moreover, our approach is compatible with the use of fuzzy set theory for dealing with the notion of vagueness.

- Our axiomatic system seems to allow for future studies of the influence of pragmatics in the meaning of strings of any language as long as we are able to separate purely linguistic elements of natural languages from their logical aspects.

References

Chomsky, N.: 1956, "Three models for the description of language" *IRE Transactions on Information Theory* **2** 113-124.

Chomsky, N.: 1959, "On certain formal properties of grammars" *Information and Control* **2** 137-167.

Chomsky, N.: 1972, Studies on Semantics in Generative Grammar, Mouton de Gruyter.

Davis, M. E., Sigal, R., Weyuker, E. J.: 1994, Computability, Complexity, and Languages, Boston Academic Press.

Desimone, R., Albright, T., Gross, C., Bruce, C.: 1984, “Stimulus-selective properties of inferior temporal neurons in the macaque” *The Journal of Neuroscience* **4** 2051-2062.

Dougherty, R. C.: 1994, Natural Language Computing, Psychology Press.

Heim, I., Kratzer, A.: 1998, Semantics in Generative Grammar, Wiley-Blackwell.

Hopcroft, J. E., Ullman, J. D.: 1969, Formal Languages and Their Relation to Automata, Addison-Wesley, Reading.

Horn, L. R., Ward, W. (eds.): 2006, The Handbook of Pragmatics, Blackwell, Oxford.

Hurford, J. R., Heasley, B., Smith, M. B.: 2007, Semantics: A Course Book, Cambridge University Press, Cambridge.

Le, Q. V., Ranzato, M., Monga, R., Devin, M., Chen, K., Corrado, G. S., Dean, J., Ng, A. Y.: 2012, “Building high-level features using large scale unsupervised learning”, http://arxiv.org/pdf/1112.6209.pdf.

Lee, K. H.: 2005, First Course on Fuzzy Theory and Applications, Springer Verlag, Berlin.

Mendelson, E.: 1997, Introduction to Mathematical Logic, Chapman & Hall, London.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C, Fried, E.: 2005, “Invariant visual representation by single neurons in the human brain” *Nature* **435** 1102-1107.

Sant’Anna, A. S., Bueno, O.: 2014, “Sets and functions in theoretical physics”, Erkenntnis **79** 257-281.

Schiffer, S.: “Meaning and formal semantics in generative grammar”, forthcoming in Erkenntnis.

Steedman, M. J.: 1989, “A generative grammar for jazz chord sequences”, Music Perception **2** 5277.

Suppes, P.: 2002, Representation and Invariance of Scientific Structures, CSLI, Stanford.

Tarski, A., Givant, S.: 1987, A Formalization of Set Theory Without Variables, AMS, Providence.

Thomason, R.: 1974, Formal Philosophy: Selected Papers by Richard Montague, New Haven.