

## The use of continuity in modelling semantic phenomena

Bernard Victorri

### Introduction

Why should we use continuous models in semantics ? At first glance, this question seems simple : we have to use continuous models if and only if semantic phenomena are continuous. But this last statement is wrong for at least two reasons. First, continuity or discreteness are not properties of phenomena, they are characterizations of theories upon phenomena. Second, as stressed by D. Kayser in this volume, one can use discrete models to represent continuous concepts, and the other way round. So, our first question must be split into two questions : (1) What kinds of linguistic theories concerning semantic phenomena need concepts related to continuity ? (2) What kinds of mathematical and computer tools can deal with these concepts ?

### 1. Linguistic issues

#### 1.1. Categorisation

At every level of linguistic description, we are confronted with the problem of classification. The scenario is always the same : in order to reach some degree of generality in the description of any phenomenon, one must define classes of linguistic expressions or relations and state rules in terms of these classes. But how to decide if one given expression or relation belongs to one given class ? In most cases, a single criterion proves to be insufficient since linguistic data show a great variability. So linguists tend to use sets of criteria to define classes. As a consequence, some graduality is obtained : expressions satisfying the whole set of criteria can be said typical elements of the corresponding class, whereas other expressions can be viewed as more peripheral, further from the center of the class as they satisfy a smaller number of criteria.

As we said, this situation prevails in every domain of linguistics. Even in syntax, we can find some sort of graduality and typicality in the definition of syntactical categories, functional relations, classes of transformations, and so on (cf. P. Le Goffic, in this volume). But the domain where these notions most obviously apply is semantics. Examples are numerous : semantic lexical features (like 'animate' versus 'inanimate'), types of process ('activity', 'accomplishment', 'achievement',...), aspect and modal classifications. In each case, it is relatively easy to exhibit typical examples showing all the features which can characterize

a class. But it is also easy to find examples on the border of two classes, where we need to refine the classification, to distinguish between different characterizations, and to introduce new factors that can take into account these differences. As the number of factors grows, and their interrelations become more complex, an alternative to the classical combinatorial representation comes out : the construction of a space representation, whose dimensions can combine the effects of different factors, each one with a specific strength. The advantage is to preserve the relation of proximity between elements (from the same class as well as from different classes) within the frame of a tractable low-dimensional space, by means of a distance on this space. Every element can be assigned a place in this representation, and the notions of center and borders of classes are fully accounted for by their geometrical counterparts. Obviously, some details will be lost from the original complexity, but with a judicious choice of the dimensions, such a compromise can be the best solution to represent the main tendencies in an efficient way, without completely eliminating any relevant factor.

It is important to note that the choice of a space representation by no means implies that the phenomena are considered as continuous. Only graduality and typicality are assumed : in other words, here continuity is nothing but an efficient tool to deal with a multiplicity of interrelating discrete factors.

## **1.2. Compositionality**

One of the main issues for any semantic theory consists in explaining how the meaning of a whole (phrase, sentence, ...) can be computed from the meaning of its components. The classical approach to this problem is compositionality. The starting point is the syntactical structure : at each node of the syntactic tree, meaning is computed bottom-up by applying rules that give the meaning of the current node as a function of the meaning of its directly dependant nodes. Two main difficulties arise in this approach. First, local rules cannot be sufficient : very often, an element, far away in the syntactical tree, exerts a definite influence over the meaning of a given part of the sentence. The second point concerns the pervasive phenomenon of polysemy. It is well known that many words, and specially the most frequent ones, are highly polysemous. Their precise meanings depend upon the rest of the sentence and have to be computed during the process, so they cannot be taken as a basis for a bottom-up computation.

These difficulties are tackled by the classical approach. For instance, the so-called mechanism of 'recategorization' enables a node value to be changed in accordance with some operation on an upper node in the tree. But here again, the combinatorial complexity of these mechanisms grow very quickly and one can be sceptical about their capacity to handle the whole set of semantic interrelations in non-trivial sentences (for a discussion about the limits of recategorization, see Fuchs et al, 1991, pp 157-162). The alternative is to consider a sentence as a 'Gestalt' where relations between whole and parts are fully bi-directional. In this view,

each component of the sentence interacts with each of the others, in no precise predefined order. What is important is the relative strength of each interaction which acts as a constraint upon the potential of meanings carried by each polysemous element. Construction of meaning can be seen as the result of a dynamical global process in which a stable solution is obtained when the maximum of constraints are satisfied (Victorri, 1992).

Related to this issue is the question of 'degrees of analysability' pointed out by R. Langacker in this volume. The compositional approach imposes a dichotomic choice : a phrase must be fully analysable (computable from its components) or fully idiomatic (considered as a new genuine element). In the gestaltist approach, these configurations are only two extreme cases of a more general situation where meaning is partly due to each component and partly an irreducible quality acquired by their interaction.

The need of continuity in gestaltist approaches is obvious. Claiming that construction of meaning is a dynamical process implies to define a space where this process can take place, where stable states can be defined, and so on. We must be aware that precise quantification is not necessary : the point of interest is the qualitative behavior of the process. A continuous space is the natural frame in which qualitative properties of dynamical systems can be handled, and nothing more.

### **1.3. Representation of meaning**

Another big issue for invoking continuity comes from representation of meaning. The main trend is to use the apparatus of logics, in one way or another, for this purpose. But a significant number of linguists challenge the prevailing views by advocating the use of topological representation. This is a very attractive idea : many lexical and almost all grammatical units can be associated to small graphic configurations that outline the kernel of their semantic value whereas logical representations tend to split the different precise meanings in as many different representations. Moreover, topological concepts and their perceptive counterparts seem to be more efficient than logical tools to explain the functional properties of these units in a cognitive perspective. The works of A. Culioli (1991), R. Langacker (1987), G. Lakoff (1987) or L. Talmy (1988) are representative of the diversity and creativity one can find in this area.

In these theories, continuity and other related topological, geometrical and dynamical notions are used as a conceptual framework. The abstract properties of, say, open intervals, boundaries, attractors, ... are the basic elements of the representation. Here quantification is completely irrelevant. As a matter of fact, one can say that the deep reason to use these concepts is their capacity to group in a single class situations which differ in their quantitative and domain-reference aspects.

## **2. Mathematical definitions**

To discuss the role continuity can play in semantic models, we first have to define this term. Most often, continuity is seen as a property of variables for which there is always an intermediate value between any two given values. Another frequent formulation uses a notion of distance : around any point in a continuous space one can always find other points as close as one wants. But from a mathematical point of view, at least three different notions related to continuity can be defined, and none of them corresponds precisely to these intuitive definitions.

### **2.1. Continuity versus discontinuity**

The opposition continuity/discontinuity applies to functions. It was first defined for numeric functions. Technically speaking, a real function  $f$  of one real variable is continuous in a given point  $x$  if for any open interval  $J$  comprising  $f(x)$  one can find an open interval  $I$  comprising  $x$  such as  $f(I)$  is included in  $J$ . In other words, there is no sudden "jump" in the value of the function when the variable passes through  $x$ . In terms of distance, one can say that a continuous function preserves 'closeness' : as the variable get closer to  $x$ , the value of the function gets closer to  $f(x)$ . This definition can be easily extended to real functions of several variables, and more generally for any function from one multi-dimensional real space to another. When using such a function in a model, points where the function is discontinuous are most often the most interesting ones, because they correspond to situations where the phenomenon under study changes its behaviour in an observable way.

### **2.2. Discreteness**

The definition of continuity is by no means limited to functions of real variables. This property can be defined for functions from any set to any another, as soon as these sets have been provided with a topological structure. To provide a set with a topology, one must define a family of subsets, called open subsets, verifying a small number of rather simple axioms : the entire set and the empty set must belong to the family, the intersection of a finite number of members of the family must belong to the family, as well as the union of any (possibly infinite) number of members. Any set can be provided with several more or less interesting topological structures. One of them is the discrete topology, for which the family of open subsets is constituted by all the subsets of the set. As a matter of fact, this topology is not very helpful because every function defined on a discrete space is continuous !

Whenever a distance is provided on a space, a corresponding 'natural' topology can be derived for which continuous functions preserve closeness. But this topology can be the discrete one, as is the case, for instance, for the set of integers with the standard distance. In

fact, no interesting topology can be defined on such a set, where the distance between any two points is greater than a constant value.

### **2.3. Continuum**

The third definition to be introduced is the notion of a continuum, sometimes called a continuous space. A continuum is a non-discrete space, but the converse is not true. For instance, the set of rational numbers with the standard distance is neither a discrete space nor a continuum. Though one can find another rational number as close as one wants to any given one, there are nevertheless "holes" in this set. In a sense, filling these holes is equivalent to the axiomatic construction of the set of real numbers, which is a continuum. Many simple geometrical properties expected from continuous spaces depend crucially on topological properties of real numbers. To give only one example, given any closed curve in the plane, one expects that a line joining a point inside the curve to a point outside must intersect the curve : such a property would not be true if one considers only points with rational coordinates in the plane.

Therefore, the whole complexity of real numbers is required to build an adequate framework for most mathematical models using continuity in a geometrical sense. In particular, dynamical systems are defined on so-called differential manifolds, which are generalizations of curves and surfaces, and the topological properties of these manifolds play a central role in the theory.

## **3. Modelling considerations**

### **3.1. Qualitative modelling**

Continuity is generally associated with quantitative modelling. We have just seen why one needs real numbers to get the advantages of continuum topology properties. But it does not mean that the model must be quantitative. Once a continuum framework has been built up, one can use it to represent a phenomenon, and very often only its qualitative features are of interest. For instance, to catch the notion of graduality, the best solution is to adopt a continuous representation where graduality can be differentiated from sudden jumps by means of continuity and discontinuities in functions, even if one knows that not every point in the space of parameters corresponds to an observable value. To have a discontinuity in one point of the space is a qualitative property for a function, and we do not need to specify the exact position of the point nor the exact value of the jump at this point to characterize the class of functions presenting this outstanding feature. On the contrary, if one adopts a discrete representation, the only way to distinguish this feature from graduality is some kind of threshold : one must define

"small" jumps and "big" ones, since there is anyway a jump when passing from one point to the next one.

As shown by this last example, in a qualitative model, the focus is not on one particular numeric function but on a class of functions exhibiting a common behavior. Nevertheless, a qualitative model can be predictive. Many qualitative relationships between data can be tested to ascertain its validity. For instance, a model may imply that data must respect a given order relative to the importance of a gradual phenomenon, that some "jump" in its behavior must be observed during the combined variations of a set of parameters, and so on. Such predictions are as useful as quantitative ones to validate or invalidate a model.

### **3.2. Dynamical systems**

As shown by our earlier discussion about linguistic issues, dynamical systems theory looks likely to play a central part in linguistic continuous models. It can be used to represent meaning of units, and also to model units interactions in a sentence. In both cases, qualitative modelling is needed. Actually, dynamical systems theory lends itself remarkably well to qualitative modelling. One can define classes of equivalent systems, characterized by a similar behavior, including in a same class systems on spaces of different dimensionality. This last point is very important in linguistics, where the same unit is used for a great variety of domain-reference spaces. One important example is the notion of bifurcation, discussed in this volume by R. Thom, and used as a central concept in Culioli's works on determination.

Moreover, dynamical systems can easily deal with the full range of linguistic phenomena known as ambiguity, indetermination and vagueness. In our work on polysemy (Fuchs et Victorri, 1988 ; Victorri et Fuchs, 1992), we built a mathematical model in which the precise meaning of a polysemous unit in any given sentence is represented by a dynamical system on a semantic space. The dynamics is parametrized by the other units present in the sentence. Each stable state (i. e. point attractor) of the dynamics correspond to a possible semantic value of the polysemous unit, so that the number of attractors and the form of the basins of attractors characterizes the meaning of the unit in the given sentence. Thus the presence of two (or more) attractors is related to the existence of an actual ambiguity. A large shallow basin of attractor represents an indetermination whereas a deep narrow one represent a specific precise meaning. These different cases can be classified and the semantic behavior of the unit can be defined in terms of the relation between these classes of dynamics and the parameters depending on the other units present in the sentence which are responsible for the form of the dynamics. For instance, one can observe how small modifications of the sentence induce qualitative changes in the dynamics, such as the appearance or disappearance of an ambiguity as an element of the sentence is replaced by another.

### **3.3. General framework**

If we try to outline the general framework emerging from the preceding considerations, we can bring out a few principles which constitute a common basis to continuous modelling in semantics.

Two representations can be associated to each linguistic element. The first one is a representation of its kernel of meaning, sometimes called its 'iconic' representation, which specifies the constant contribution of this unit to the meaning of any sentence comprising it. The second one is what we called here its 'semantic space', whose dimensions reflect the degrees of freedom corresponding to the variable precise meanings this unit can convey in different sentences. The interactions between units in a given sentence are then twofold. On the one hand, kernel representations interact, bringing out the full representation of the meaning of the whole sentence. On the other hand, each unit receives from the others a set of constraints which defines its behavior in its semantic space. In both cases dynamical systems theory seems to be the appropriate tool to compute these interactions. It is at least the right tool to model these interactions as a gestaltist process.

As it stands, this framework is not actually an effective semantic model. Our claim is that it defines a research program in which most of linguistic studies using continuity as a main ingredient can be included.

## **4. Computer tools**

### **4.1. Continuity on a digital computer**

If we turn now towards computer implementation, representing continuity on a machine is all but a simple problem. From a rigorous mathematical point of view, continuity cannot be reached with a digital computer : even the so-called 'real' variables take their values, whatever the precision, on a discrete finite set of numbers. So, the best we can do is to approximate continuous functions by discrete gradual operations and the subtle distinctions we made at the mathematical level no longer apply in this context. Nevertheless, in many domains, and specially in physics, computer simulations are used to study continuous mathematical models, and machine precision is sufficient to obtain reliable results. So it cannot be argued that a digital computer is not suitable for representing continuity.

But the problem is elsewhere. In domains like physics, the focus is on quantitative simulations, and computers are used for what they do the best : numeric computations. In our case, we are most interested in qualitative behavior and choosing numeric values is most of the time an irrelevant burden. As we have shown, quantification is the opposite of what is needed

in the mathematical apparatus related to continuity. If computer implementation of a qualitative continuous mathematical model imposes arbitrary numeric coding, it will be devoid of interest.

## **4.2. Connectionism**

Connectionist networks seem to provide an elegant solution to this problem. They are essentially numeric, of course, since the relations between units are defined by a 'real' number, i.e. the weight of their link. But these weights are not to be arbitrarily chosen by the designer of the system. They are adjusted as the result of a learning process. In concrete terms, what is needed is the encoding of a sample of input data and of expected corresponding output results. Then the learning algorithm automatically computes weight values so that the system gives a correct response when presented with data close to the learned sample.

The most simple example is given by the so-called 'feed-forward' networks. They are constituted by an ordered set of layers of units, each unit of one layer being connected to all units of the next layer. The first and the last layers are respectively the input and output layers. Mathematically speaking, these two layers implement two spaces, and the learning process is equivalent to computing the most regular function from the input space to the output space satisfying the constraints given by the learning sample. To use such a system, one has to design the two basic spaces by specifying their dimensions and the rules of encoding data and results onto these spaces. Strictly speaking, encoding operations are also numeric, but here the situation is different : each unit must correspond to a linguistic criterion which is part of the model, and the choice of coding values which can be limited to two or three values is not arbitrary from a linguistic point of view. Connectionist approach is often opposed to symbolic approach, but in fact, there is a symbolic aspect in any connectionist network, precisely because input and output units are inevitably given a symbolic meaning, even in so-called 'distributed representations', otherwise the system could not be of any use for modelling. The non-symbolic aspect of connectionist networks is limited to what happens strictly inside the network, in the correspondance between input and output, where the learning process takes place.

One of most interesting classes of connectionist networks is the family of 'recurrent' networks. As opposed to feed-forward networks, they allow bi-directional links between units and so they are direct implementations of dynamical systems which we argued to be of cardinal importance in semantic continuous models. With these systems, one can capture the notions of attractors, bifurcations, and so on. As an example, we used a recurrent architecture to implement our model of polysemy, and it enabled us to differentiate phenomena of ambiguity, indetermination, ... by the form of the basins of attractors of the dynamics created inside the

recurrent network associated to the polysemous unit in different sentences (Victorri et al, 1989 ; Gosselin et al, 1990).

### **4.3. Current limits of connectionism**

So connectionism is already an essential tool to implement continuous models. But it has a drawback that prevents it from playing in continuous modelling the same role as classical artificial intelligence tools play in discrete modelling. This flaw is related to an important notion developed in artificial intelligence : the notion of control. A connectionist network remains a "black box" which does not allow much reasoning about its functioning. Our experience with polysemy modelling showed us how frustrating it was to work with such a system. Even when it gave us roughly satisfactory results, we could not use it to answer the most important questions which motivated the implementation of the model : what were the decisive factors that explained the good performances ? In which direction might we modify the system to improve it ? How could we characterize the class of systems giving acceptable results ?

To answer these questions, one must "open the black box", i.e. study the relation between the performances of the network and its internal configuration. Theoretical work is implied to classify networks in terms of qualitative behavior. This direction of research constitutes a major challenge for connectionism. This work has already started, as shown for instance by the important theoretical results obtained by D. Amit on one family of recurrent systems, the Hopfield networks. The usefulness of continuous model implementations crucially depends upon progress in this area.

## **Conclusion**

Continuous models look likely to play a more and more important role in current research in semantics. In this paper, we tried to delimit to what extent mathematical and computer tools are adapted to this task. Clearly, many efforts rare still to be made in this field to become an actual alternative to discrete formalisation. But an original theoretical framework already emerges : it can bring new light to many well-known linguistic issues, which cannot be taken into account by discrete modelling.

## REFERENCES

- Culioli A. (1990) : *Pour une linguistique de l'énonciation. Opérations et représentations*. tome 1, Paris, Ophrys.
- Fuchs C. et Victorri B. (1988) : Vers un traitement automatique de la polysémie grammaticale, Paris, *T.A. Informations*, 29.
- Fuchs C., Gosselin L. et Victorri B. (1991) : Polysémie, glissements de sens et calcul des types de procès, Strasbourg, *Travaux de linguistique et de philologie*, XXIX:1, pp137-169.
- Gosselin L., Konfé A. et Raysz J.P. (1990) : Un analyseur sémantique automatique d'adverbes aspectuels du français, *Actes du 4<sup>e</sup> colloque de l'A.R.C.*, Paris.
- Lakoff G. (1987) : *Women, fire and dangerous things : what categories reveal about the mind*, University of Chicago Press.
- Langacker R.W. (1987) : *Foundations of cognitive grammar*, Stanford University Press.
- Talmy L. (1988) : Force dynamics in language and cognition, *Cognitive Science*, 9:1.
- Victorri B. et Fuchs C. (1992) : Construction de l'espace sémantique associé à un marqueur grammatical polysémique, *Linguistica Investigationes*, 16:1.
- Victorri B. (1992) : Un modèle opératoire de la construction dynamique de la signification, in *La théorie d'Antoine Culioli, Ouvertures et incidences*, Paris, Ophrys, pp 185-201.
- Victorri B., Raysz, J. P., Konfé, A. (1989), Un modèle connexionniste de la polysémie, *Actes de Neuro Nimes 89*, EC2.