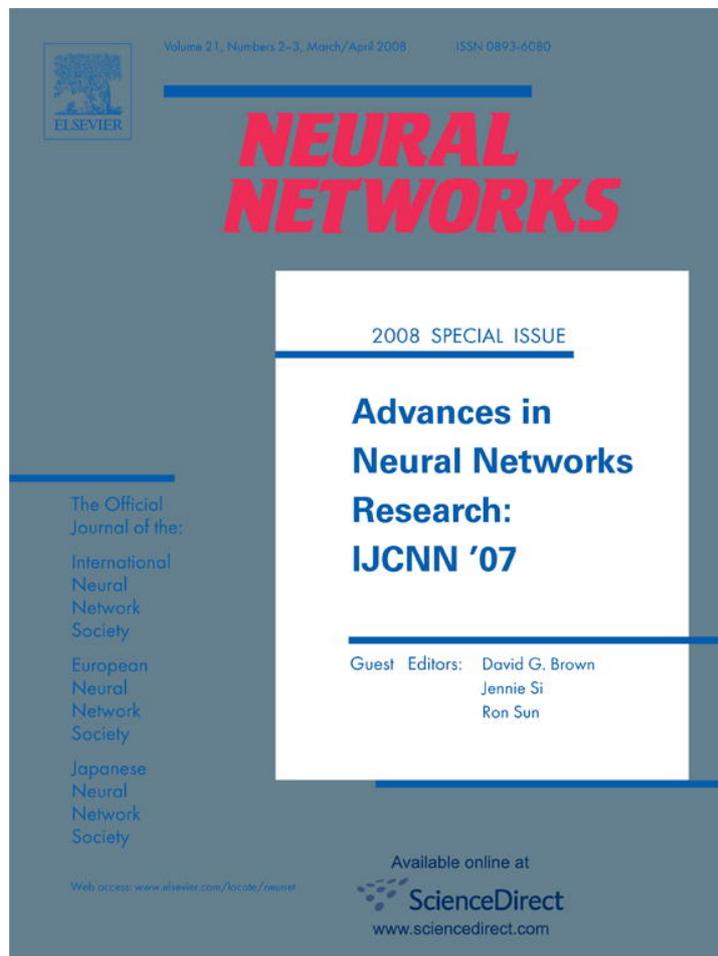


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



2008 Special Issue

# How language can help discrimination in the Neural Modelling Fields framework<sup>☆</sup>

José F. Fontanari<sup>a,\*</sup>, Leonid I. Perlovsky<sup>b,c</sup>

<sup>a</sup> Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560-970 São Carlos SP, Brazil

<sup>b</sup> Harvard University, 33 Oxford St, Rm 336, Cambridge MA 02138, United States

<sup>c</sup> Air Force Research Laboratory, 80 Scott Drive, Hanscom Air Force Base, MA, United States

Received 2 August 2007; received in revised form 30 October 2007; accepted 3 December 2007

## Abstract

The relationship between thought and language and, in particular, the issue of whether and how language influences thought is still a matter of fierce debate. Here we consider a discrimination task scenario to study language acquisition in which an agent receives linguistic input from an external teacher, in addition to sensory stimuli from the objects that exemplify the overlapping categories that make up the environment. Sensory and linguistic input signals are fused using the Neural Modelling Fields (NMF) categorization algorithm. We find that the agent with language is capable of differentiating object features that it could not distinguish without language. In this sense, the linguistic stimuli prompt the agent to redefine and refine the discrimination capacity of its sensory channels.

© 2007 Elsevier Ltd. All rights reserved.

**Keywords:** Acquisition of language; Clustering algorithms; Neural Modeling Fields

## 1. Introduction

In contrast to research on the origin and evolution of language, for which the lack of empirical evidences to support theory prompted the *Société Linguistique de Paris* to ban papers on the subject in 1866 (see, e.g. Aitchison (1996)), language acquisition has always been considered an area of legitimate scientific investigation. In fact, controlled experiments on language acquisition by children (Bloom, 2000) as well as by nonhuman primates (Fitch & Hauser, 2004; Savage-Rumbaugh et al., 1993) have set strict constraints to theories in this field. Curiously enough, however, most long standing controversies involving language are connected to language acquisition rather than to language evolution. For instance, the issue of whether acquiring a language is the result of an exclusive human innate language competence, i.e. language is unlearnable (Chomsky, 1972, 1980) or is the result of the use of purely inductive, statistical learning procedures (Bates & Elman, 1996; Saffran,

Aslin, & Newport, 1996; Seidenberg, MacDonald, & Saffran, 2002) is still a matter of fierce contention (see Deacon (1997, Chap. 4) for an overview).

Another issue on dispute is the so-called Sapir–Whorf or language determinism hypothesis which asserts that language determines thought, or in a weaker version, that language partially influences thought. The following quote by Ferdinand de Saussure sums up this idea: “Without language, thought is a vague, uncharted nebula. There are no pre-existing ideas, and nothing is distinct before the appearance of language” (de Saussure, 1966). This excerpt fits very well the notion that language is primarily a representational system rather than an ability evolved for the sole purpose of communication (Bickerton, 1990). Despite strong criticism from the Establishment (see, e.g., Pinker (1994, Chap. 3)), this powerful idea, which can be traced back to von Humboldt in the nineteenth century, resurges time after time. Its more recent upsurge is the claim that the difficulty of the Pirahã people to recount numbers higher than three may be related to the fact that their language contains three counting words only, namely, one, two and many (Gordon, 2004). (We mention also the possibility that the Pirahã may have no words for numerals at all, but words similar to ‘a few’, ‘some’, and ‘many’ which they use to describe quantity.) In support to the weak version of the Sapir–Whorf hypothesis, we mention

<sup>☆</sup> An abbreviated version of some portions of this article appeared in Fontanari and Perlovsky (2007) as part of the IJCNN 2007 Conference Proceedings, published under IEE copyright.

\* Corresponding author. Tel.: +55 16 33739849; fax: +55 16 33739877.

E-mail addresses: [fontanari@ifsc.usp.br](mailto:fontanari@ifsc.usp.br) (J.F. Fontanari), [Leonid.Perlovsky@hanscom.af.mil](mailto:Leonid.Perlovsky@hanscom.af.mil) (L.I. Perlovsky).

the possibility of grounding meanings – a typical cognitive task (Harnard, 1990) – using purely linguistic representations (Cangelosi, Greco, & Harnard, 2002). In general, however, language is viewed as a result of the high cognitive abilities of humans (Pinker, 1994), as well of their complex social organization (Fitch, 2004; Knight, 1998).

To address quantitatively the relationship between language and thought, in this paper we consider a simple artificial scenario in which an agent endowed with the Neural Modeling Fields (NMF) categorization system (Perlovsky, 2001) must discriminate and identify a variety of inputs that compose its environment. This differentiation task is a typical cognitive function which, in principle, can be performed without the aid of language using a variety of unsupervised learning algorithms (Blatt, Wiseman, & Domany, 1997; Fontanari, 2006; Fontanari & Perlovsky, 2006; Kohonen, 1995; Steels, 1996). However, if the input categories are highly overlapping, so that the separation between their means is smaller than the Cramér–Rao bound of each category (see, e.g. Perlovsky (2001, Chap. 9)), then the categorization mechanisms will fail to distinguish the input classes. We show that, in this case, learning from an external teacher symbolic or linguistic representations for members of the categories can be helpful to refine the discrimination ability of the agent. The NMF framework is designed to facilitate the integration (fusion) of inputs of very distinct nature, so this formalism suits very well to study the relationship between language and cognition (Perlovsky, 2004, 2006; Tikhonoff, Fontanari, Cangelosi, & Perlovsky, 2006).

In the next section, we describe the environment in which the agent lives as well as the task posed to it. In Section 3 we briefly review the NMF formalism within the context of the specific categorization problem addressed in this paper. In Section 4 we present the results of the simulation of the NMF dynamic in the case the agent does not receive a linguistic input, and in Section 5 we address the more interesting case where an external teacher names the category to which the samples belong. Finally, Section 6 summarizes the main conclusions. An abridged version of the present paper was published in Fontanari and Perlovsky (2007).

## 2. The discrimination task

The agent's environment contains a certain number of objects sampled from six overlapping categories. The objects are characterized by two features – feature A and feature B – which can be thought of as describing physical properties of the objects (e.g. texture and colour). For simplicity, these features are represented by real numbers drawn from Gaussian distributions. Fig. 1 displays the particular instance we will consider in this paper. The categories can be identified by assigning the integers 1, . . . , 6 in a one-to-one correspondence with the means of the Gaussians used to generate the samples. For example, the category whose samples are centered at coordinate (0.1, 0.1) is labelled 1, the one centered at (0.2, 0.2) is labelled 2, and so on. The environment is set so that the agent cannot categorize and identify all examples, because of the particularly high overlap between the features that describe the categories 3 and 4 (see

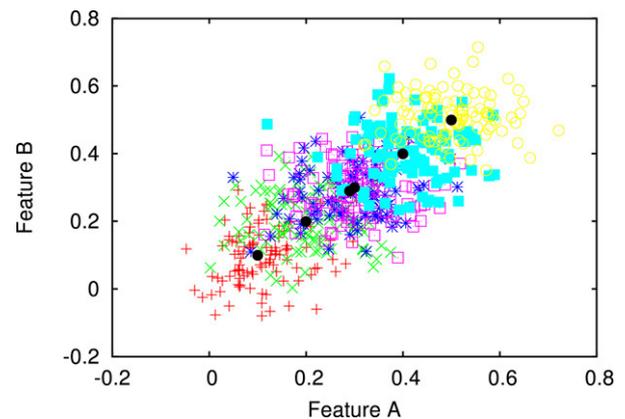


Fig. 1. The six sets of 100 examples, represented by the features A and B (e.g. texture and color), of six overlapping categories. The coordinates of each pixel are drawn from Gaussian distributions of means 0.1, 0.2, 0.29, 0.3, 0.4 and 0.5 (large black circles) and standard deviation 0.08. To each example of the first category we associate the label or word 1, which represents the linguistic signal produced by an external teacher. Labels 2–6 are similarly assigned to the other categories.

Fig. 1). Inspired by the “mushroom world” scenario (Cangelosi, 2001; Parisi & Cangelosi, 2002) we allow the agent to receive from the environment an additional sensory input: a heard linguistic signal. Here, we assume that this signal is produced by an external teacher who has perfect knowledge of the environment. Again, for simplicity, we represent these signals by the integer labels 1, . . . , 6. In practice, this amounts to assuming that the teacher utters the word ‘1’ whenever the agent is presented to an example of the category 1, and similarly for the other categories.

Although a more realistic phonological model as well as more realistic descriptions of the category examples (through coloured pictures, for instance) could give more ‘credibility’ to our virtual scenario, these complicating factors are peripheral to our primary goal of studying the relationship between language and cognition. Therefore, to avoid diversion by unnecessary complications here we stick to the simplest possible situation in which the inputs are described by real or integer numbers. Explicitly, we model the inputs to the categorization system by the triples  $(O_{i1}, O_{i2}, O_{i3})$  where  $i = 1, \dots, N$  labels the distinct category examples (we set  $N = 600$  in this paper, see Fig. 1). The first two components represent the two physical features – feature A and feature B – of the examples. In particular, we choose  $O_{ie}$ ,  $e = 1, 2$  as independent random variables drawn from six Gaussian distributions of standard deviation 0.08 and means 0.1 for  $i = 1, \dots, 100$ , 0.2 for  $i = 101, \dots, 200$ , etc. The third component  $O_{i3}$  is the linguistic input which takes on the integer values 1, . . . , 6 depending on the category of example  $i$  (see Fig. 1). To study the case of learning without language we simply need to omit the third component of the input triple.

## 3. The neural modelling fields formalism

The basic idea behind Neural Modelling Fields is the association between lower-level signals (e.g. inputs) and higher-level concept models (internal representations) avoiding

the combinatorial complexity inherent to such a task. This is achieved by using measures of similarity between concept models and input signals together with a new type of logic, so-called dynamic logic. We refer the reader to Perlovsky (2001) for a complete presentation of NMF; here we particularize the discussion to the discrimination task presented in the previous section. Accordingly, we define  $M$  concept-models or modeling fields through the triples  $(S_{k1}, S_{k2}, S_{k3})$ ,  $k = 1, \dots, M$ , where  $S_{ke}$ ,  $e = 1, 2, 3$  are real variables which should capture the key features of the input signals. Of course, the nature of the modelling field variables is determined by the representation of the input signals. For example, if the category samples were collections of pixels (i.e., pictures) then we should use a similar representation for the modelling fields. The essential point here is that the number of concepts  $M$  should be much smaller than the number of input signals  $N$  so that the categorization system is forced to compress the information provided by the environment. In the NMF framework, this compression results in the creation of a few iconic representations (concepts) for the input data. Clearly, for the data depicted in Fig. 1 the simplest iconic representation is the average value of the features that characterize the examples of a given category, i.e. essentially the mean of the distribution used to produce the category samples.

Setting an appropriate value for the number of modelling fields  $M$  is crucial to the success of the NMF categorization system. In a previous contribution (Fontanari & Perlovsky, 2006), we have combined the NMF approach with the Akaike Information Criterion (Akaike, 1974) to design a categorization system that infers correctly the true number of categories in an environment similar to that exhibited in Fig. 1, but for the purposes of this paper, any choice of  $M \geq 6$  is satisfactory.

We begin the derivation of NMF algorithm, i.e. of the equations that govern the dynamics of the modelling fields, by defining a measure for the similarity between input  $i$  and concept  $k$ , namely,

$$l(i | k) = \prod_{e=1}^d [2\pi\sigma_{ke}^2]^{-1/2} \exp\left[-(O_{ie} - S_{ke})^2 / 2\sigma_{ke}^2\right] \quad (1)$$

where, at this stage, the fuzziness  $\sigma_{ke}^2$  are parameters given *a priori*. As mentioned before, the cases of learning with and without language can be considered simply by setting  $d = 3$  and  $d = 2$ , respectively. The goal is to find an assignment between concept models  $k$  and examples  $i$  such that the global similarity

$$L = \sum_{i=1}^N \ln \sum_{k=1}^M l(i | k) \quad (2)$$

is maximized. This can be easily achieved by changing  $S_{ke}$  such that  $dS_{ke}/dt = \partial L / \partial S_{ke}$  since then

$$\frac{dL}{dt} = \sum_{ke} \frac{\partial L}{\partial S_{ke}} \frac{dS_{ke}}{dt} = \sum_{ke} \left( \frac{\partial L}{\partial S_{ke}} \right)^2 \geq 0 \quad (3)$$

as required. The calculation of  $\partial L / \partial S_{ke}$  is straightforward and yields

$$\frac{\partial L}{\partial S_{ke}} = \sum_{i=1}^N \frac{1}{\sum_{k'=1}^M l(i | k')} \frac{\partial l(i | k)}{\partial S_{ke}}. \quad (4)$$

Using the identity  $\partial y / \partial x = y \partial \ln y / \partial x$  and defining the fuzzy associations as

$$f(k | i) = \frac{l(i | k)}{\sum_{k'=1}^M l(i | k')} \quad (5)$$

we can immediately write the equations for the modelling fields (Perlovsky, 2001)

$$\frac{dS_{ke}}{dt} = \sum_{i=1}^N f(k | i) \frac{\partial \ln l(i | k)}{\partial S_{ke}} \quad (6)$$

for  $k = 1, \dots, M$  and  $e = 1, \dots, d$ . The fuzzy associations (5) play a fundamental role in the interpretation of the NMF dynamics by giving a measure of the correspondence between sample  $i$  and model  $k$  relative to all other models  $k'$ . We note that although the features A and B of a category example are independent random variables, the components of a modelling field  $k$  are coupled dynamic variables. Actually, the term  $f(k | i)$  in Eq. (6) couples also components of different modelling fields.

For fixed fuzziness  $\sigma_{ke}^2$ , the dynamics (6) is guaranteed by construction to converge to a maxima of  $L$ . However, the fixed point towards which the dynamics evolves is very sensitive to the choice of the (fixed) value of  $\sigma_{ke}^2$  (Fontanari & Perlovsky, 2005). To get around this difficulty we could maximize the global similarity  $L$  with respect to  $\sigma_{ke}^2$  as well (Perlovsky, 2001). This procedure works very well in the case the categories are characterized by large standard deviations, such as those illustrated in Fig. 1, but it is somewhat problematic in the case of narrow categories and, even more so, in the case each category is represented by a single example (Fontanari & Perlovsky, 2005). In fact, the difficulty is that  $\sigma_{ke}^2$  can become too small, leading to a numerical instability in the evaluation of the partial similarities Eq. (1). An alternative scheme for changing the values of the fuzziness on the fly, i.e. simultaneously with the change of the modelling fields, is the following prescription (Fontanari & Perlovsky, 2005, 2006)

$$\sigma_{ke}^2(t) = a_{ke}^2 + b_{ke} \exp(-\alpha_e t) \quad (7)$$

where  $\alpha_e$ ,  $a_{ke}$ , and  $b_{ke}$  are time-independent control parameters. As a guideline for setting the values of these parameters, we note that  $b_{ke}$  must be chosen large enough such that, at the beginning, all category examples can be described by all modelling fields, whereas the baseline resolution  $a_{ke}$  must be small enough such that, at the end, a given modelling field will describe a single category. However, as pointed out before  $a_{ke}$  should not be set to a too small value to avoid numerical instabilities in the calculation of the partial similarities defined

by Eq. (1). The results of the next section will illustrate the main ideas involved in the numerical evaluation of Eqs. (6) and (7).

A word is in order about the connection between the NMF framework and neural networks, before we embark on the numerical analysis of the modelling field dynamics. A NMF neural architecture was described in Perlovsky (2001), which combines architecture with models of objects or category examples. Essentially, input neurons or bottom-up signals encode the feature values of the category examples  $O_{ie}$ , and top-down or priming signal fields to these neurons are generated by the modelling fields  $S_{ke}$ . Interaction between bottom-up and top-down signals is determined by the neural weights  $f(k|i)$  that associate signals and models. As described before, these weights are functions of the fields  $S_{ke}$ , which in turn are dynamically adjusted so as to maximize the overall similarity between category examples and models. This formulation sets NMF apart from many other neural networks. There is, on the other hand, a certain formal similarity between the NMF approach and the Hopfield–Tank neural network approach to tackle optimization problems (Hopfield & Tank, 1985). This becomes apparent when one recognizes that the nature of perceptual problems dealt with here is similar to that of other optimization problems. In fact, in both systems it is the time evolution of analogue neurons that drives the neural configuration to a maximum of the cost function (the global similarity  $L$  in our case). Moreover, the quality of the solutions found by the neural network is greatly improved by annealing the analogue gain parameter (Vandenbout & Miller, 1989) in a similar manner as the slow decrease of the fuzziness according to Eq. (7) leads ultimately to perfect categorization. In addition, the competition between different concept models to match the category examples is reminiscent of the dynamics of unsupervised learning algorithms and, in particular, of self-organizing maps (Kohonen, 1995).

#### 4. Agent without language

The choice of the particular environment depicted in Fig. 1 is motivated by the inability of the agent to distinguish between all six categories into which the 600 examples are organized. The difficulty, of course, is to distinguish between the two sets of examples centred at the coordinates (0.29, 0.29) and (0.3, 0.3), which are labelled 3 and 4, respectively, by the external teacher. Fig. 2 illustrates the time dependence of the first components of the modelling fields  $S_{k1}$ ,  $k = 1, \dots, 6$  obtained by solving the set of ordinary differential equation (6) using the Euler method with the step size  $h = 10^{-6}$ . Since the second components  $S_{k2}$  are essentially equivalent to the first ones, they exhibit a very similar time dependence and so will not be considered here. The third component is fixed to zero, since in this case the agent has no access to the extra linguistic input produced by the external teacher. The parameters that control the slow decrease of the fuzziness (7) are  $a_{ke} = 0.1$ ,  $b_{ke} = 1.5$  and  $\alpha_e = 2 \times 10^{-4}$  for  $k = 1, \dots, 6$  and  $e = 1, 2$ . This parameter setting implies that the fuzziness is independent of the model and component labels, i.e.  $\sigma_{ke}^2 = \sigma^2$ . The initial values of the components of the modeling fields are chosen randomly from the uniform

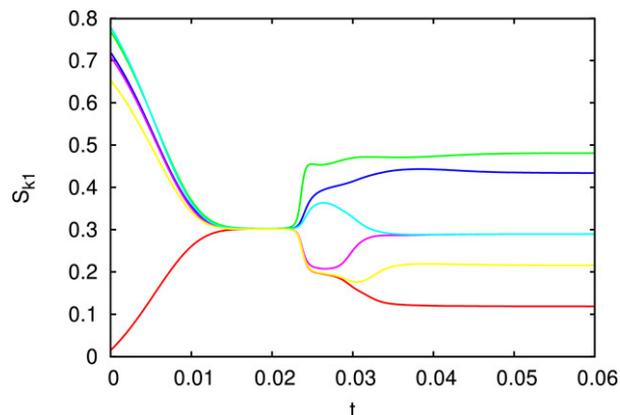


Fig. 2. Evolution of the component  $e = 1$  (feature A) of the six modelling fields (from top to bottom)  $k = 2, 3, 5, 1, 6$  and  $4$  when the agent without language perceives the environment composed of 600 examples that belong to six categories as illustrated in Fig. 1. Note that the agent is unable to distinguish between the two categories labeled 3 and 4 by the external teacher.

distribution in  $(0, 1)$ , which leads to an arbitrary assignment between models and categories. In particular, category 1 is (iconically) represented by model  $k = 4$ , category 2 by model  $k = 6$ , category 5 by model  $k = 3$ , category 6 by model  $k = 2$ , whereas categories 3 and 4 are clumped by models  $k = 1$  and  $k = 5$  (see Fig. 2).

The point of Fig. 2 is to stress the rather expected failure of most categorization methods to distinguish highly overlapping categories. In fact, the agent was able to identify four of the six overlapping categories displayed in Fig. 1 and, in addition, it succeeded in discriminating categories 3 and 4 from the remaining ones. Inspection of this figure reveals some interesting features of the categorization procedure. After an initial transient stage, the algorithm reaches a regime where all examples are classified into a single category, which corresponds to the homogeneous solution  $S_{ke} = S_e = \sum_i^N O_{ie}/N$  for all  $k$ . Then the input space is divided into two sectors, depending whether the example features are greater or smaller than  $S_e$ . This pattern is reminiscent of that displayed by the Super Paramagnetic Clustering algorithm in which the clusters (categories) are formed following a similar hierarchy as the temperature of a system of Potts spins is slowly lowered (Blatt et al., 1997).

In Fig. 3 we illustrate the time dependence of the global similarity density,  $L/N$ , as well as of the fuzziness  $\sigma^2$  for the same run shown in Fig. 2. The relevant feature here is that the modelling field dynamics (6) increases the global similarity despite the fact that the fuzziness  $\sigma_{ke}^2 = \sigma^2$  decreases steadily with increasing time. We note that the time dependence of  $\sigma^2$  exhibited in Fig. 3, analogously to the arbitrary cooling schedule of Simulated Annealing (Kirkpatrick, Gelatt, & Vecchi, 1983), is not affected by the dynamics of the modelling fields.

Although the stationary values of the modelling fields are an important outcome of the NMF framework, since they are interpreted as the iconic, internal representations created by the agent to make sense of its environment, the fuzzy association variables  $f(k|i)$  are the most relevant quantities as far as

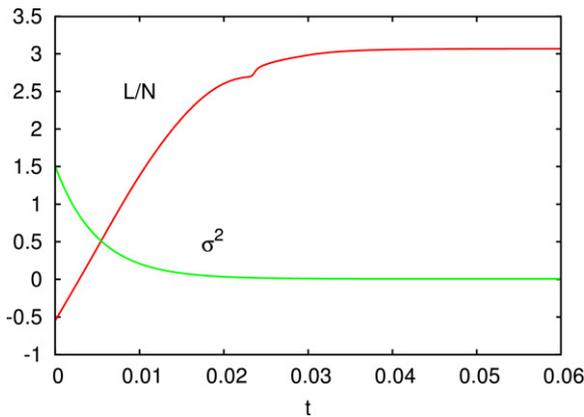


Fig. 3. Time dependence of the global similarity per example  $L/N$  and of the model independent fuzziness  $\sigma^2$  for the same run shown in Fig. 2. Note that  $L$  is an increasing function of time  $t$ , regardless of the fact that  $\sigma^2$  is not fixed.

the categorization problem is concerned. In fact,  $f(k | i)$  yields the degree of confidence with which model  $k$  describes a particular sample  $i$ . Fig. 4 illustrates the results for the cleanest case of models  $k = 4$  and  $k = 2$ , which are associated to the more definite categories 1 ( $i = 1, \dots, 100$ ) and 6 ( $i = 501, \dots, 600$ ). These results show that the ill-definiteness of the categorization problem illustrated in Fig. 1 is reflected on the fuzzy association variables rather than on the modelling field variables. In fact, samples of the neighbour categories 2 ( $i = 101, \dots, 200$ ) and 5 ( $i = 401, \dots, 500$ ), which have large deviations from the means, are “wrongly” associated to models  $k = 4$  and  $k = 2$ . To quantify this effect, we determine the model  $k$  for which  $f(k | i)$  is maximum for each sample  $i$ , and then compare this prediction with the correspondence between models and categories displayed in Fig. 2. For example, a perfect (but unrealizable, unless the agent receives additional non-fuzzy information regarding the class membership of each sample; see Section 5) model should predict that  $k = 4$  maximizes  $f(k | i)$  for  $i = 1, \dots, 100$ . In our framework we find that 83% of the samples of category 1 were correctly assigned to model  $k = 4$ , 56% of category 2 to model  $k = 6$ , 57% of category 3 to model  $k = 5$ , 64% of category 4 to model  $k = 1$ , 59% of category 5 to model  $k = 3$ , and 81% of category 6 to model  $k = 2$ . Random guessing would yield a success rate of 17% in any case.

### 5. Agent with language

To study the situation in which the agent receives an additional (linguistic) input associated to each category sample, we just have to turn on the third component  $S_{ke}$  of the modelling fields. As already mentioned in Section 2, this setting is motivated by the “mushroom world” scenario (Cangelosi, 2001; Parisi & Cangelosi, 2002) where there is an external teacher who utters the word  $O_{i3}$  whenever the agent observes the physical input ( $O_{i1}, O_{i2}$ ). Of course, the teacher utters the same word for all examples belonging to a same category and so  $O_{i3}$  takes on only six different values (i.e. there are only six different words). To stress the fact that the nature of these signals (i.e. the words) is completely distinct from that

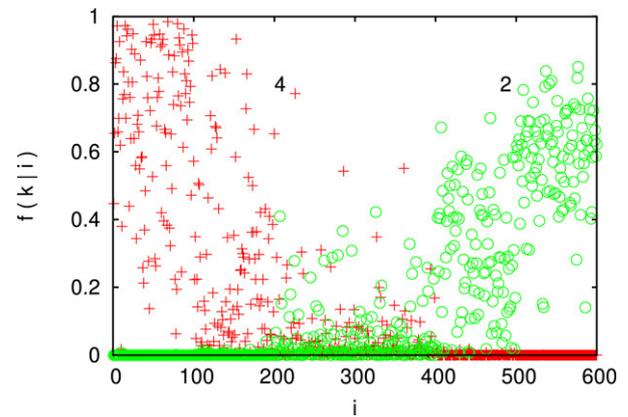


Fig. 4. Stationary values of the fuzzy association variables  $f(k | i)$  for  $k = 2$  ( $\circ$ ) and  $k = 4$  ( $+$ ) for each category sample  $i = 1, \dots, 600$ . According to Fig. 2, model  $k = 4$  should describe the examples from  $i = 1$  to 100 (category 1), whereas model  $k = 2$  the examples from  $i = 501$  to 600 (category 6).

of the physical inputs, we assign the integer values  $1, \dots, 6$  to the linguistic component of the input. These integers are the category labels presented in the caption of Fig. 1.

In the computational experiments reported in this section, we keep the same setting of control parameters that determine the evolution of the modelling fields associated to the physical inputs, including their initial values, used in the previous section. The control parameters for the linguistic component are setting as follows:  $a_{k3} = 0.1, b_{k3} = 2.5$  for all  $k$ , and  $\alpha_3 = 5 \times 10^{-5}$ . The reason for the larger value of  $b_{k3}$ , as compared with the values of  $b_{k1}$  and  $b_{k2}$ , is that the separation between the target words  $O_{k3}$  are much greater than the distances between the means of the Gaussians used to generate the category examples in Fig. 1. We recall that the successful convergence of the NMF scheme requires that one starts with large fuzziness values to guarantee that at the outset all models have a nonzero similarity with all input data. In addition, the small magnitude of  $\alpha_3$  as compared with  $\alpha_1$  and  $\alpha_2$  emphasizes the need for different cooling schedules for assimilation of inputs of distinct nature. We set the initial values of the components  $S_{k3}$  to zero so that at the beginning the linguistic components will have little effect on the dynamics of the other physical components.

Figs. 5 and 6 illustrate the time evolution of the modeling field component  $S_{k3}$  associated to the linguistic input, and of the component  $S_{k1}$  associated to the physical feature A, respectively. As before, the component associated to feature B exhibits a similar time dependence to that shown in Fig. 6. The first point to be noted is that, despite the initial values the modelling fields being the same as in the case the linguistic input is blocked, the resulting assignment between the model labels  $k$  and the categories is different in the experiments summarized in Figs. 2 and 6. This assignment is very sensitive to fluctuations – the splitting of the homogeneous solution shown in Fig. 2 is reminiscent of the spontaneous symmetry breaking of critical phenomena in statistical mechanics (Patrias, 1972) – and so it is expected that the presence of the extra, third component would alter that assignment.

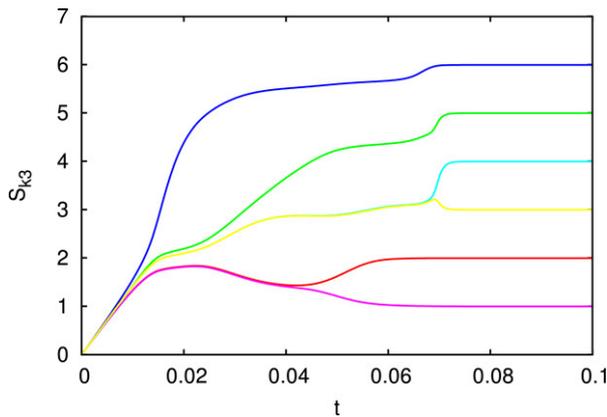


Fig. 5. Evolution of the linguistic component  $S_{k3}$  of the modelling fields (top to bottom)  $k = 3, 2, 5, 6, 4$  and  $1$ .

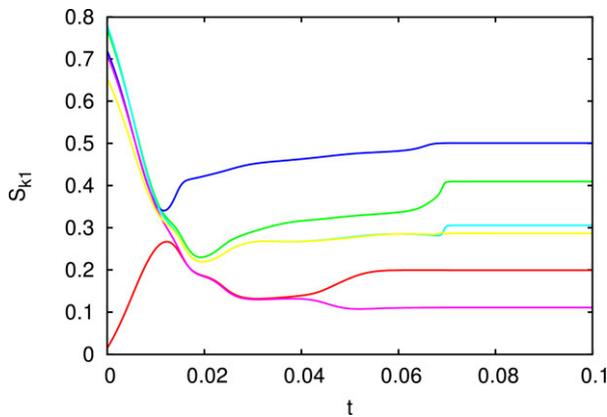


Fig. 6. Evolution of the component  $S_{k1}$  (feature A) of the modelling fields (top to bottom)  $k = 3, 2, 5, 6, 4$  and  $1$  for the agent with language. Because of the linguistic input, the agent was able to differentiate the physical features associated to categories 3 and 4.

Despite the complete lack of overlap between the distinct words used to label the categories, the matching between the linguistic inputs and the associate component of the modelling fields is diffcultated due to the influence of the two physical components, as shown in Figs. 5 and 6. However, as soon as the word models are differentiated, which happens at  $t \approx 0.068$ , categories 3 and 4 are disentangled, as illustrated in the figures. This is a remarkable finding: the extra information carried by the linguistic component allowed the agent to create distinct iconic representations for the category examples. In other words, the knowledge that categories 3 and 4 are distinct, which was obtained thanks to the linguistic input, allowed the agent to redefine and refine its expectations about features A and B. This result is a direct consequence of the particular manner that NMF handles linguistic and physical inputs. A phrase by Kerényi (1996) neatly summarizes our finding: “The interdependence of thought and speech makes it clear that languages are not so much a means of expressing truth that has already been established, but are a means of discovering truth that was previously unknown.”

We note that the extra linguistic input associated to each category sample makes the categorization problem trivial, since now there is a unique assignment of each sample to the category

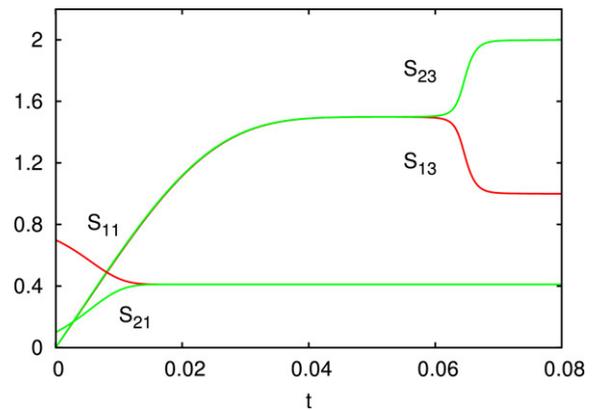


Fig. 7. Evolution of the modelling fields associated with feature A,  $S_{11}$  and  $S_{21}$ , and to the linguistic input,  $S_{13}$  and  $S_{23}$ , in the case the teacher assign different names to the same category.

labels, which is given by the linguistic input. In fact, the fuzzy association variables  $f(k|i)$  at the stationary regime takes on the values 1 or 0 depending on whether the sample  $i$  belongs or not to the category represented by the concept label  $k$  (see caption of Fig. 5 for the assignment of the labels  $k$  to the categories). For example,  $f(3|i) = 1$  for  $i = 501, \dots, 600$  and 0 otherwise. Of course, the point of introducing the linguistic input was to verify its influence on the ability of the agent to discriminate the other physical inputs, and so to address the Sapir–Whorf hypothesis within the restricted domain of the NMF unsupervised learning procedure.

To conclude this analysis, now we address the question of whether the agent would create fictitious distinctions between physically identical objects given that the external teacher assigns different names to them. In the experiment we devise to address this issue, the agent is exposed to samples of two categories described by two physical features  $O_{i,e}, e = 1, 2$ , and a linguistic signal  $O_{i,3}$ , as before. Samples  $i = 1, \dots, 100$  correspond to category 1, whereas samples  $i = 101, \dots, 200$  to category 2. The key ingredient of this experiment is that both categories are characterized by the same physical features, i.e.  $O_{j,e} = O_{j+100,e}, e = 1, 2$ , but different names,  $O_{j,3} = 1$  and  $O_{j+100,3} = 2$  with  $j = 1, \dots, 100$ . The physical inputs  $O_{j,1}$  and  $O_{j,2}$  are sampled from a Gaussian of mean 0.4 and standard deviation 0.08. The parameters that control the fuzziness reduction schedule are the same as those used in the previous experiment. The results are summarized in Fig. 7: although the agent eventually learns the category names (or category labels) it does not create distinct representations for the physical features. This result is reassuring because it shows that the refinement of the discrimination capability of the physical sensors is not a general consequence of the fact that the teacher provides names to the categories.

## 6. Conclusion

We have reported a computational experiment in which the addition of language, or more precisely of a linguistic signal, affects the manner that an agent processes its other sensory inputs. Remarkably, the agent with language is capable

of differentiating sensory inputs that it could not distinguish without language. The crucial role played by the linguistic signal in our experiment contrasts with a milder claim that language enhances performance only if the agent has already evolved an ability to respond appropriately to the visually perceived objects without language (Parisi & Cangelosi, 2002).

What distinguishes linguistic signals (e.g. word sounds) from other stimuli is the fact that the agent experiences the sounds in concomitance with nonlinguistic experience. Of course, it is difficult to distinguish a linguistic input from a nonlinguistic one, since ultimately linguistic signals are perceived through physical sensors (e.g. the ears in case of speech) as well. In fact, in case of humans this distinction is probably made by some innate mechanism which allows the hearer to recognize the intention of the speaker to transmit some information — essentially the agents must be endowed with a Theory of Mind (Bloom, 2000). In our scenario we avoid these complications, and define the linguistic input as a persistent signal associated to members of a same category. This naive view that learning category names is achieved through naming members of a category is certainly not correct — naming a single instance of a category is sufficient for children (and even apes) to realize that the name applies to the category, rather than to the observed example. However, since such high-level innate cognitive abilities cannot be taken for granted in the study of language acquisition by robots or virtual agents, the “learning by examples” approach adopted in this paper seems to be the only viable alternative. Finally, we note that the fact the linguistic signal must be expressed through some physical media is nicely accounted by the NMF approach in which all input are processed (and fused) in equal terms.

## Acknowledgments

This work was supported in part by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA9550-06-1-0202, and in part by CNPq and FAPESP, Project No. 04/06156-3. The US Government is authorized to reproduce and distribute reprints for governmental purpose notwithstanding any copyright notation thereon.

## References

- Aitchison, J. (1996). *The seeds of speech: Language origin and evolution*. Cambridge, UK: Cambridge University Press.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Bates, E., & Elman, J. (1996). Learning rediscovered. *Science*, *274*, 1849–1850.
- Bickerton, D. (1990). *Language and species*. Chicago: University of Chicago Press.
- Blatt, M., Wiseman, S., & Domany, E. (1997). Data clustering using a model granular magnet. *Neural Computation*, *9*, 1805–1842.
- Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.
- Cangelosi, A. (2001). Evolution of communication and language using signals, symbols, and words. *IEEE Transactions on Evolutionary Computation*, *5*, 93–101.
- Cangelosi, A., Greco, A., & Harnard, S. (2002). Symbol grounding and the symbolic theft hypothesis. In A. Cangelosi, & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 191–210). London: Springer-Verlag.
- Chomsky, N. (1972). *Language and mind*. New York: Harcourt Brace Jovanovich.
- Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
- Deacon, T. W. (1997). *The symbolic species*. New York: W.W. Norton & Company.
- de Saussure, F. (1966). *Course in general linguistics* (W. Baskin, Trans.). New York: McGraw-Hill Book Company.
- Fitch, W. T. (2004). Kin selection and mother tongues: A neglected component in language evolution. In K. Oller, & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach* (pp. 275–296). Cambridge, MA: MIT Press.
- Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*, 377–380.
- Fontanari, J. F. (2006). Statistical analysis of discrimination games. *European Physical Journal B*, *54*, 127–130.
- Fontanari, J. F., & Perlovsky, L. I. (2005). Meaning creation and modeling field theory. In *Proceedings of the IEEE conference on integration of knowledge intensive multi-agent systems* (pp. 405–410).
- Fontanari, J. F., & Perlovsky, L. I. (2006). Categorization and symbol grounding in a complex environment. In *Proceedings of the IEEE international joint conference on neural networks* (pp. 10039–10045).
- Fontanari, J. F., & Perlovsky, L. I. (2007). Language acquisition and category discrimination in the Modeling Field Theory framework. In *Proceedings of the IEEE international joint conference on neural networks* (pp. 1180–1185).
- Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*, 496–499.
- Harnard, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.
- Hopfield, J. J., & Tank, D. W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics*, *52*, 141–152.
- Kerényi, C. (1996). *Dionysos: Archetypal image of indestructible life* (R. Manheim, Trans.). Princeton, NJ: Princeton University Press.
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, *220*, 671–680.
- Knight, C. (1998). Ritual/speech coevolution: A solution to the problem of deception. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 68–91). Cambridge, UK: Cambridge University Press.
- Kohonen, T. (1995). *Springer series in information sciences: Vol. 30. Self-organizing maps*. Berlin: Springer.
- Parisi, D., & Cangelosi, A. (2002). A unified simulation scenario for language development, evolution and historical change. In A. Cangelosi, & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 255–275). London: Springer-Verlag.
- Pathria, R. K. (1972). *Statistical mechanics*. Oxford: Pergamon Press.
- Perlovsky, L. I. (2001). *Neural networks and intellect: Using model-based concepts*. Oxford: Oxford University Press.
- Perlovsky, L. I. (2004). Integrating language and cognition. *IEEE Connections*, *2*, 8–13.
- Perlovsky, L. I. (2006). Integrated emotions, cognition, and language. In *Proceedings of the IEEE international joint conference on neural networks* (pp. 1570–1575).
- Pinker, S. (1994). *The language instinct*. London: Penguin Press.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- Savage-Rumbaugh, E. S., Murphy, J., Sevoik, R. A., Brakke, K. E., Williams, S. L., & Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs on Social Research into Child Development*, *58*, 1–221.
- Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science*, *298*, 553–554.
- Steels, L. (1996). Perceptually grounded meaning creation. In M. Tokoro (Ed.), *Proceedings second international conference on multi-agent systems* (pp. 338–344). Menlo Park: AAAI Press.
- Tikhonoff, V., Fontanari, J. F., Cangelosi, A., & Perlovsky, L. I. (2006). Language and cognition integration through modeling field theory: Category formation for symbol grounding. *Lecture Notes in Computer Science*, *4131*, 376–385.
- Vandenbout, D. E., & Miller, T. K. (1989). Improving the performance of the Hopfield–Tank neural network through normalization and annealing. *Biological Cybernetics*, *62*, 129–139.