A cross-situational algorithm for learning a lexicon using Neural Modeling Fields

José F. Fontanari, Vadim Tikhanoff, Angelo Cangelosi and Leonid I. Perlovsky

Abstract— **Cross-situational learning is based on the idea that a learner can determine the meaning of a word by finding something in common across all observed uses of that word. Although cross-situational learning is usually modeled through stochastic guessing games in which the input data vary erratically with time (or rounds of the game), here we investigate the possibility of applying the deterministic Neural Modeling Fields (NMF) categorization mechanism to infer the correct object-word mapping. Two different representations of the input data were considered. The first is termed object-word representation because it takes as inputs all possible object-word pairs and weighs them by their frequencies of occurrence in the stochastic guessing game. A re-interpretation of the problem within the perspective of learning with noise indicates that the cross-situational scenario produces a too low signal-to-noise ratio, explaining thus the failure of NMF to infer the correct object-word mapping. The second representation, termed context-word, takes as inputs all the objects that are in the pupil's visual field (context) when a word is uttered by the teacher. In this case we show that use of two levels of hierarchy of NMF allows the inference of the correct object-word mapping.**

I. Introduction

The origin of human language is truly secret and marvelous, being probably the most fundamental unsolved problem of science and, perhaps, of the humanities as well. In fact, we often hear the claim that language is the very feature that distinguishes our species from the other animals. Sometimes language is even set above thought, the hallmark of human rationality, as this quote by Ferdinand de Saussure illustrates so vehemently "Without language, thought is a vague, uncharted nebula. There are no pre-existing ideas, and nothing is distinct before the appearance of language" [1]. Although nowadays few researchers subscribe to such a radical viewpoint, the more established stance has equally controversial claims such as the position that language evolved from animal cognition, *not* from animal communication [2], to the despair of an entire generation of ethologists who sought for evidences of a continuous line of development between animal communication systems and language [3]. The current predominant view, at least among linguists, is that language and thought are distinct abilities of the mind [4] and a quick comparison between the cognitive and linguistic abilities of apes and parrots appears to be a convincing argument for many researchers.

In this contribution we carry further the rather ambitious research program of integrating language and cognition within the Neural Modeling Fields Framework (NMF) [5]-[8]. This is a task of enormous breadth that encompasses many unsolved (and, perhaps, unsolvable) problems such as object perception [9]-[11], symbol grounding which addresses the question of how physical signs can be given meaning [12]-[14], and the emergence of a common lexicon in a population of interacting agents [15]-[19]. As pointed out in [13], language is not an isolated capability of the individual and cannot be fully comprehended if one ignores its intrinsic relationships with the cognitive and social abilities.

Here we assume that the first two problems, namely, object perception and symbol grounding, were solved so we address the issue of how agents can communicate with each other using the meanings they have constructed. Clearly, to achieve that the agents must be able to use some sort of signals, and so they must be endowed with the ability to create signals as well as to express and infer the meanings of those signals. In other words, it is necessary to assume the existence of a mapping between concepts or internal representations of the outside objects and signals – the meaning-signal mapping. Explaining how this mapping emerges from the agents' interactions is an open problem [20]-[24].

There are a few radically distinct approaches to studying the emergence of communication codes (or meaning-signal mappings). On the one hand, there is the approach based on a direct analogy with biological evolution [15],[16],[19] that makes use of the explicit assumption that the communication codes are transmitted from parents to children (vertical transmission in the population genetics jargon) and that possessing an efficient communication code confers a fitness advantage to the individual. On the other hand, there is the cultural-based approach, of which the so-called Iterated Learning Framework (ILF) is the most important representative [25],[26]. In the ILF there are (typically) only two agents involved – the teacher and an initially *tabula rasa* pupil. After learning, the pupil replaces the teacher and

The research at São Carlos was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Project No. 04/06156-3.

J. F. Fontanari is with Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP Brazil (phone: +55-16-33739849; fax: +55-16-33739877; e-mail: fontanari@ifsc.usp.br).

V. Tikhanoff and A. Cangelosi are with Adaptive Behaviour & Cognition Research Group, University of Plymouth, Plymouth, UK (e-mail: vadim.tikhanoff@plymouth.ac.uk; a.cangelosi@plymouth.ac.uk)

L. I. Perlovsky is with Harvard University, Cambridge MA and The Air Force Research Laboratory, SN, Hanscom Air Force Base, MA (e-mail: Leonid.Perlovsky@hanscom.af.mil).

a fresh pupil is introduced in the learning scenario. This procedure is repeated until the communication code becomes fixed.

The cross-situational lexicon acquisition scenario we consider here differs from the biological and cultural approaches in the sense that the concept of evolution (variations passed from generation to generation) plays no role. In fact, a cross-situational algorithm to learn a communication code (lexicon) differs from the previous approaches in the sense that it involves the repeated interaction between the same two individuals. In that sense, these algorithms are also known as guessing or naming games [27]. Cross-situational learning is based on a very intuitive idea (which may actually bear no relevance at all on the way children learn a vocabulary [28]) that one way that a learner can determine the meaning of a word is to find something in common across all observed uses of that word [29]. Although, the general notion of cross-situational learning has been proposed by many authors (see, e.g., [30], [31]) the translation of those rather abstract theories into a mathematical or computational model is still subject of intense research in the literature [18],[20]-[24],[29].

The rest of this paper is organized as follows. In the next section, we describe the cross-situational learning scenario and introduce the necessary adjustments to apply the NMF framework. In section III we briefly review the NMF formalism within the context of the specific problem addressed in this paper. In section IV we present and discuss the results of the simulations of the NMF dynamics. Finally, section V summarizes the main conclusions

II. The cross-situational learning scenario

A minimal model of cross-situational lexicon learning involves two agents who take turns in the roles of “speaker” and “hearer” [20]-[24]. (Henceforth we will make no distinction between objects and meanings). The agent who plays the role of the “hearer” at a certain round of the game has access to a set of objects - the context – which is a random subset of the entire set of objects that make up the agents’ world. The hearer also has access to a single word emitted by the other agent who plays the role of the “speaker” in that round of the game. The hearer’s task is to guess to which object in the context the word refers to. In addition, we assume that there is a fixed number N of objects and that the agents have access to a fixed number H of words (the lexicon size).

The NMF formulation capable of both creating meanings and evolving an organized structure (an object-word mapping) from scratch is given in [32]. In this paper, as a first step, we use NMF in its form described in [33]. This formulation) maximizes the similarity between input data and parametric models used to represent (and usually compress) that data. This is sufficient to demonstrate an ability for a cross-situational learning, however this framework is not appropriate for the typical guessing game settings [20]-[24]. Therefore a standard cross-situational setting has to be slightly modified to apply the NMF parametric model formulation [33]. We assume then that one of the agents – the teacher – has complete domain of the language which is described by a one-to-one object-word mapping ($N = H$). The *tabula rasa* pupil must then infer that mapping from the examples provided by the teacher. Let us again emphasize that the problem addressed here is that of acquiring or learning an existing Saussurean communication code [15], it is assumed that such a code already exists and is manifested in the examples provided by the teacher. In the terminology of cross-situational learning of the beginning of this section, the teacher always plays the speaker role whereas the pupil takes up the hearer role. Figure 1 illustrates the modified cross-situational lexicon acquisition scenario.

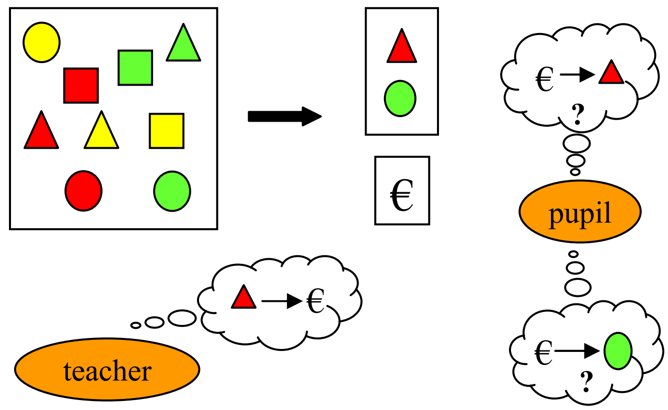

Fig. 1. Scheme of the cross-situational lexicon acquisition scenario. The pupil has access to a context comprised of two objects selected at random from the full set of N objects and to a word (represented by € in the figure) uttered by the teacher. The pupil’s task is to infer the word-object mapping from the co-occurrence of objects in the context and the accompanying words.

The second change in the standard cross-situational scenario is that the deterministic nature of the NMF algorithm requires that all input data (e.g., all possible examples of object-word pairs) be presented at once to the pupil. This sort of batch-mode learning setting differs considerably from the on-line learning of the typical guessing game scenario. For the sake of concreteness, let us label objects and words by the integers $1,\ldots,N$ and $1,\ldots,H$, respectively, and assume without lack of generality that the correct one-to-one object-word mapping is such that object 1 corresponds to word 1, object 2 to word 2, and so on (recall that $N = H$). In what follows we will introduce two different representations of the input data which, of course, lead to distinct implementations of the NMF algorithm.

In the first representation, which we term object-word input representation, the input data to the pupil consists of the N^2 object-word pairs (1,1), (1,2), (1,3),…, (N,N) that are weighed according to their frequencies of occurrence in the guessing game scenario. These frequencies (or weights) can be obtained by noting that a round of the guessing game consists of the presentation of one of the following $N(N-1)$ situations: $\{(1,1),(2,1)\}$, $\{(1,2),(2,2)\}$, $\{(1,1),(3,1)\}$, …, $\{(N,N-1),(N-1,N-1)\}$. Here the first brace represents the situation where the context comprises objects 1 and 2 and the teacher utters the word that corresponds to object 1; the second brace represent the same context but now the accompanying word corresponds to object 2, etc. We have separated the input data by braces which represent different learning events for the sake of clarity only: because of the batch-mode of learning, the pupil has no access to that kind of information. Clearly, the pair (1,1) appears $N-1$ times, whereas the pair $(1,n), n \neq 1$ appears just once. A similar result holds for the other objects, so we can conclude that the weight of a correct object-word association is $N-1$ whereas the weight of any other object-word association is 1. The pupil attempts to model the input data through object-word mapping (S_{1k}, S_{2k}), with $k=1,\ldots,M$, where the components $S_{ek}, e=1,2$ - the so-called modeling fields – are real variables given by the NMF equations described in the next section. The question is whether the pupil can recover the correct object-word mapping (1,1) (2,2) … (N,N) having access only to the information available in the cross-situational learning scenario summarized in Fig. 1, or more precisely, having access to the relative weights of each object-word example. There is a considerable loss of information involved in the procedure of replacing the actual presentations of contexts plus words by the weights of the object-word examples, since one can easily imagine different situations which result in the same weights.

The second representation of the input data, which we term context-word input representation, is a more direct one, in which, for example, the situation where the context comprises objects 1 and 2 and the teacher utters the word that corresponds to object 1 is represented by a triplet (1,2,1). The first two components represent the object labels, whereas the third component, the word label. In this case the pupil models the input data using a context-word mapping (S_{1k}, S_{2k}, S_{3k}), $k=1,\ldots,M$. In addition, because these components have a totally distinct nature we allow the NMF algorithm to give more importance to the linguistic component of the input when determining the assigning between input and models.

III. The Neural Modeling Fields equations

Perlovsky's Neural Modeling Fields (NMF) algorithm consists basically of an iterative, self-consistent, deterministic process designed to maximize the similarity between models and incoming signals [33]. In this sense it shares some elements with the Hopfield-Tank neural network [34] and the mean-field annealing [35]. In fact, these two deterministic heuristics have been extensively used to search for optimal or quasi-optimal solutions of a variety of optimization problems, whereas NMF searches for the maximum of a global similarity function. The main feature that sets NMF apart from these heuristics, as well as from many other neural networks, is the presence of the so-called fuzzy association variables, which can be roughly thought of as weights of a neural network. These variables give a measure of the probability of association between input data and concept-models although, as already pointed out, NMF is a deterministic algorithm.

We refer the reader to [33] for the NMF formalism presentation; here we modify this general framework for the object-word representation of the input data produced by the guessing game introduced in the previous section. The equations for the context-word representation can be immediately derived from the equations presented here, as indicated in the end of this section.

As mentioned before, each example (input signal) is described by the pair of integer variables (O_{1i}, O_{2i}) with $i=1,\ldots,N^2$. Let us assume that there are M concept-models described by the pairs (S_{1k}, S_{2k}) with $k=1,\cdots,M$ that should "model" (i.e., create internal representation for) the original examples; hence the denomination "modeling fields" to the mathematical quantities S_{ek}. We define the following partial similarity measure between example i and model k

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

where, at this stage, the fuzziness σ_{ek} are parameters given *a priori* (see [7] for another application of NMF to multi-component inputs). Of course, in the object-word representation we must set $d=2$. The goal is to find an assignment between models and examples such that the logarithmic global similarity

$$L = \sum_i a_i \ln\left(\sum_k l(i \mid k)\right) \quad (2)$$

is maximized. Here the parameters a_i are new elements in the theory and give the weight (relative importance) of input i in the calculation of the global similarity. The maximization of L can be achieved using the NMF mechanism of model formation which is obtained through the direct maximization of (2) with respect to S_{ek}. The aim here is to derive a dynamical equation for the modeling fields S_{ek} such that $dL/dt \geq 0$ for all time t. This condition can easily be met by choosing $dS_{ek}/dt = \partial L/\partial S_{ek}$ since then

$$dL/dt = \sum_{e,k} \left(\partial L/\partial S_{ek}\right)\left(dS_{ek}/dt\right) = \sum_{e,k} \left(\partial L/\partial S_{ek}\right)^2 \geq 0 \quad (3)$$

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

$$\frac{\partial L}{\partial S_{ek}} = \sum_i a_i \frac{1}{\sum_{k'} l(i \mid k')} \frac{\partial l(i \mid k)}{\partial S_{ek}} \quad (4)$$

and leads to the following dynamics for the modeling fields

$$dS_{ek}/dt = \sum_i a_i f(k \mid i)\left[\partial \log l(i \mid k)/\partial S_{ek}\right], \tag{5}$$

for $e = 1,\ldots,d$ and $k = 1,\cdots,M$ and where we have used the identity $\partial y/\partial x = y\,\partial \log y/\partial x$. The fuzzy association variables $f(k \mid i)$ defined by

$$f(k \mid i) = l(i \mid k)\Big/\sum_{k'} l(i \mid k') \tag{6}$$

play a fundamental role in the interpretation of the NMF dynamics by giving a measure of the correspondence between object *i* and model *k* relative to all other models k'. We note that the term $f(k \mid i)$ in Eq. (5) couples not only S_{1k} and S_{2k}, which is critical for producing a sensible object-word mapping, but also the components of different modeling fields.

By construction the dynamics (5) always converges to a (possibly local) maximum of the similarity *L* for fixed σ_{ek} [33]. A salient feature of the NMF is a match between parameter uncertainty and fuzziness of similarity. By properly decreasing the value of the fuzziness σ_{ek} a unique assignment between inputs and models is attained. In fact, for fixed $\sigma_{ek} > 0$ we obtain the fuzzy logic limit, whereas for $\sigma_{ek} = 0$ we obtain the usual crispy, Aristotelian logic limit. The basic idea of NMF is to reduce the fuzziness during the time evolution of the modeling fields and so, because of this interpolation, the algorithm is also referred as Dynamic Logic. Of course, this procedure is similar to the cooling schedule of simulated annealing [36], although, as pointed out before, a more appropriate comparison is with mean-field annealing [35] since both algorithms are deterministic.

In what follows we decrease the fuzziness on the fly, i.e., simultaneously with the change of the modeling fields, according to the following prescription [37]

$$\sigma_{ek}^2(t) = \sigma_{ae}^2 \exp(-\alpha_e t) + \sigma_{be}^2 \tag{7}$$

with $\alpha_e = 2500$. As a guideline for setting the values of the parameters σ_{ae} and σ_{be}, we note that σ_{ae} must be chosen large enough such that, at the beginning, all examples are described by all fields, whereas the baseline resolution σ_{be} must be small enough such that, at the end, a given modeling field will describe a single category. However, σ_{be} should not be set to a too small value to avoid numerical instabilities in the calculation of the partial similarities (1). Finally, $1/\alpha_e$ yields the relevant time-scale of the modeling fields dynamics. The results of the next section will illustrate the main ideas involved in the numerical evaluation of equations (5) and (7).

To obtain the equations that describe the NMF algorithm for the context-word representation we simply have to define the input as triplets (O_{1i}, O_{2i}, O_{3i}) with $i = 1,\ldots,2N(N-1)$ and set $a_i = 1$ for all *i*. The modeling fields must be defined accordingly, (S_{1i}, S_{2i}, S_{3i}). So, keeping these remarks in mind we can leave the basic equations (1), (5) and (7) unchanged by setting $d = 3$.

In what follows we will set $M = N = H$ so Eq. (5) stands for a set of $2N$ (object-word representation) or 3N (context-word representation) nonlinear coupled equations, which are solved with Euler's method using the step-size $h = 10^{-5}$. In a previous contribution [14], we have combined the NMF approach with the Akaike Information Criterion [38] to design a system that infers automatically the optimum number of models. Since this paper reports a preliminary attempt to put together two well-established mathematical frameworks, namely, cross-situational learning and NMF, we opted to make the task as easy as possible for NMF, hence the choice of $M = N$, which should bias the categorization system to find the correct object-word mapping.

IV. SIMULATION RESULTS

A. Object-word input representation

To apply the NMF formalism described above to the cross-situational learning scenario of Fig. 1 we have to set the weights a_i appropriately. In particular, if the input *i* represents a correct object-word association (n,n) then we set $a_i = N-1$, otherwise we set $a_i = 1$. In addition, in the simulations reported here we set $\sigma_{ae} = 1$ and $\sigma_{be} = 0.1$ for $e = 1,2$.

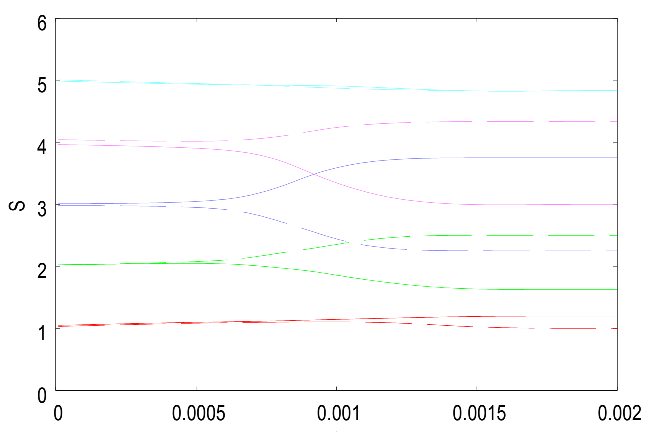

Fig. 2. Time evolution of the two components of the modeling fields in the case of $N = 5$ objects. The weights a_i are chosen so as to mimic the co-occurrence frequency of objects and words in the cross situational learning scenario of Fig. 1. The second component, associated to the label of the word, is represented by a solid line and the first component, associated to the object, is represented by a dashed line. The $M = 5$ different modeling fields are represented by different colors. The initial condition is a small perturbation of the correct object-word mapping.

In Fig. 2 we show the results of a typical run in which the initial condition (i.e., the values of the modeling fields at $t = 0$) are set very close to the correct object-word mapping $(1,1),\ldots,(5,5)$. We have verified that in the case the initial condition is identical to the correct mapping then the two components S_{1k} and S_{2k} remain equal to each other ($S_{1k} = S_{2k}$) at all times, although they drift slightly from their correct value ($S_{1k} = S_{2k} \approx k$), as seen in the evolution of the modeling field associated to the input $(5,5)$ in Fig. 2. The important point illustrated in this figure is that the correct object-word mapping is not an attracting fixed point of the dynamics and, more importantly, the two components of a same model split up in 4 out of 5 cases producing a object-word mapping quite different from the correct one.

To understand the mapping produced by the NMF mechanism we have to look at the fuzzy association variables $f(k \mid i)$ with $k = 1,\cdots,M$ and $i = 1,\ldots,N^2$. We find that after convergence $f(k \mid i)$ takes on the values 1 and 0 only. So we can say that the input or example *i* (actually an object-word pair) is represented by model *k* if $f(k \mid i) = 1$. For instance, for the run illustrated in Fig. 2 we found that model *k* = 1 describes the examples (1,1) (1,2); model *k* = 2, the examples (2,1) (2,2) (3,1) (3,2) (4,1); model *k* = 3, the examples (1,3) (1,4) (1,5) (2,3) (2,4) (2,5) (3,3) (3,4) (3,5); model *k* = 4, the examples (4,2) (4,3) (4,4) (5,1) (5,2) (5,3); and finally model *k* = 5, the examples (4,5) (5,4) (5,5). Note that each example is described by a single model but a given model can represent several different examples. Thus, for example, the NMF mechanism interprets both examples (1,1) (1,2) as instances of the same situation which is labeled $S_{11} = 1.0$ and is associated to the word $S_{12} = 1.2$ (see red curves in Fig. 2).

The criterion the NMF mechanism uses to lump together examples in a same class (model) is the similarity between the integers that label the objects and the words. In principle, those labels were intended to be used solely to identify the objects and words (i.e., as labels), and not to produce a metric in the object and word spaces (see, however, [39],[40] for a model where the metric in those spaces is exactly the difference between the labeling integers). If we recall that we are not making any distinction between the concepts of object and meaning (we have skipped the object-meaning mapping by identifying meaning and object so the usual meaning-word mapping turned into an object-word mapping) then we can realize that this difficulty is a reflex of a much deeper philosophical problem. In fact, the very notion of the similarity of meanings as compared to the identity of meanings is a highly controversial issue in cognitive science (see, e.g., [41]-[43]). However, within a connectionist perspective in which meanings are neural activation patterns, the concept of meaning similarity follows naturally and the NMF theory systematically exploits this similarity [33]. More precisely, a necessary condition for applying the NMF mechanism to a categorization problem is the definition of a metric in the input as well as in model spaces so that the similarity measure, Eq. (1), becomes a well-defined physical quantity.

Therefore by using the similarity measure (1) we were actually assuming that the relevant distance (metric) in both object and word space is the result of the usual subtraction between the integers that label the objects and the words [39],[40]. Once this metric is taken into account, the mapping produced by the NMF mechanism makes perfect sense.

Alternatively, we may interpret the input set as a realization of the problem of learning the mapping $(1,1),(2,2),\ldots,(N,N)$ from noisy examples. In this sense, the examples (1,2) (1,3) (1,4) and (1,5) may be viewed as noisy versions of the signal (1,1). Accordingly, we interpret the weights a_i as the strength of example *i*. The problem is that using the frequency of occurrence of the examples in the cross-situational scenario we find that the signal to noise ratio equals 1, which explains the instability of the correct object-word mapping observed in Fig. 2.

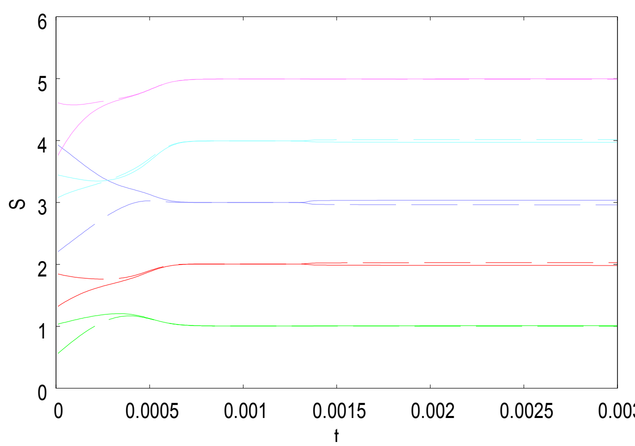

Fig. 3. Time evolution of the two components of the modeling fields in the case of $N = 5$ objects. The weights a_i are chosen so as to model the learning with noise situation. We set $a_i = 100$ for the N examples of the form (n,n) and $a_i = 1$ for the other $N(N-1)$ examples. The conventions for line color and style are the same as in Fig. 2. The initial condition for each component was chosen randomly in the range $[0, N]$.

In fact, if we set the strengths of the correct object-word associations to a much larger value such that the signal to noise ratio moves up to 25, we get a very different scenario (see Fig. 3) where now the modeling fields reproduce the correct object-word mapping. As before, to understand the mapping produced by the NMF mechanism we need to inspect the fuzzy associations $f(k \mid i)$. For the run shown in Fig. 3 we find that model *k* = 1 describes the examples (1,1) (1,2); model *k* = 2 the examples (1,3) (2,1) (2,2) (3,1) (3,2) (4,1); model *k* = 3 the examples (1,4) (1,5) (2,3) (2,4) (2,5) (3,3) (3,4) (4,2) (5,1); model *k* = 4 the examples (3,5) (4,3) (4,4) (5,2) (5,3) (5,4); and model *k* = 5 the examples (4,5) (5,5). The assignment of the model labels is arbitrary so we have chosen the convention to label *k* the model that describes the examples (*k*,*k*). Again, the similarity between the integers that represent the objects and words is the key to understand the assignment of the 25 distinct examples to the 5 categories. Unfortunately, we cannot devise a cross-situational learning scenario that yields the weights values a_i such as those necessary to produce the results of Fig. 3, unless we allow the possibility that, say, object 1 and word 1 are presented repeatedly to the pupil in a completely unambiguous situation. Of course, this possibility trivializes the lexicon acquisition scenario.

B. Context-word input representation

Since our attempt to model the inherently stochastic cross-situational learning scenario by using the frequency of occurrence of the object-word examples did not produce the intended results (see Fig.2) due essentially to the low signal to noise ratio, now we try a different representation of the input data, the context-word mapping mentioned before. In this case the $2N(N-1)=40$ different situations: (1,2,1) (2,1,1) (1,3,1) (3,1,1) (1,4,1) (4,1,1)… (4,5,5) (5,4,5) are presented just once to the pupil (i.e., $a_i = 1$ for all *i*). The first two components of these triplets label the objects that comprise the context, and the third component labels the word associated to the object corresponding to the first entry. In addition, to bias the NMF system so it gives more importance to the linguistic entry we set the fuzziness parameters in Eq. (7) to $\sigma_{ae} = 1$, $\sigma_{be} = 1/10$ for $e = 1,2$ (as before) and $\sigma_{a3} = 1/4$, $\sigma_{b3} = 1/40$ for the linguistic entry. This choice means that a small discrepancy regarding the third component of the input and model triplets produces an enormous effect on the partial similarity, see Eq. (1). In this way we can assign different weights to the components of the inputs.

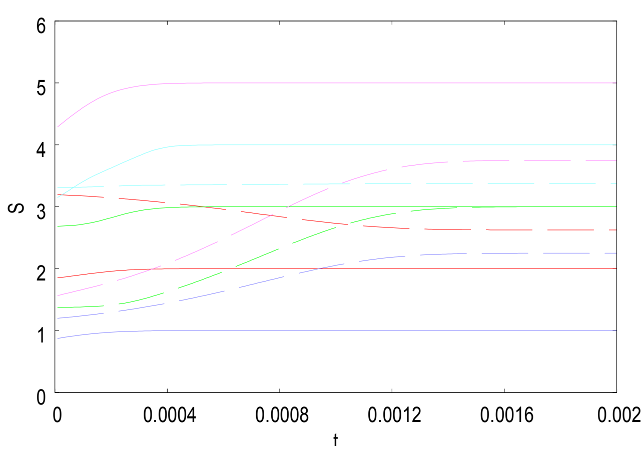

Fig. 4. Time evolution of the first (dashed line) and third (solid lines) entries of the modeling fields in the context-word input representation. The second entry (not shown in the figure) becomes identical to the first entry for t > 0.0012. The initial condition for each component was chosen randomly in the range $[0, N]$.

The results of a typical run are summarized in Fig. 4. As expected the entry associated to the linguistic feature of the input quickly reaches the correct value, whereas the entries associated to the other two features which are basically irrelevant to determine the assignment of model *k* to input *i*, cluster around the value of the center of mass of the input labels, namely, $(1+\ldots+5)/5 = 3$. It is interesting, however, that the dashed lines do not collapse into the single line given by $S_{ek} = 3 \;\; e = 1,2$ and all *k*. This is the same effect reported in [44] in which once the inputs are classified into different categories (because of the linguistic component) small differences in the other entries are enhanced, producing a better discrimination of the non-linguistic features. Of course, since there is really no ambiguity in the triplet representation of the inputs, the word-biased NMF mechanism lumps into a single model (say *k* = 1) all triplets whose third component takes on the value 1, into model *k* = 2 the triplets whose third component is 2, and so on.

In order to extract the object-word mapping given the context-word output of the NMF we must figure out how to extract the common object label from the subset of inputs (1,2,1) (2,1,1) (1,3,1) (3,1,1) (1,4,1) (4,1,1) (1,5,1) (5,1,1) associated to model *k* = 1. (Of course, an entirely similar procedure holds for the other models as well.) This can be easily achieved by introducing a single modeling field *S* the task of which is to extract the common label that appears in the first two entries of the 5 triplets listed above. In fact, introducing the similarity

$$l(n) = \exp\left[-\left(S - O_{1n}\right)^2 \left(S - O_{2n}\right)^2\right] \qquad (8)$$

with $n = 1,\ldots,5$ and minimizing the global similarity $L = \sum_n \ln[l(n)]$ with respect to S yields the following NMF dynamic equation

$$\frac{dS}{dt} = -4\sum_n \left(S - O_{1n}\right)\left(S - O_{2n}\right)\left(S - \frac{O_{1n} + O_{2n}}{2}\right). \qquad (9)$$

Recalling that for each n we have either $O_{1n} = 1$ or $O_{2n} = 1$ the stationary solution of Eq. (9), i.e. $dS/dt = 0$, is clearly $S = 1$, as desired. In that way we can recover the correct object-word mapping by implementing two NMF hierarchies.

V. Conclusion

There are some inherent difficulties to apply the NMF formalism to the cross-situational learning scenario. First, that scenario is based on labels (integers or symbols) that represent different objects and words, and so there is no notion of distance or similarity between different objects or different words. In fact, as pointed out by Pinker [4], “babies should not, and apparently do not, expect *cattle* to mean something similar to *battle*, or *singing* to be like *stinging*, or *coats* to resemble *goats*". (This point is reminiscent of Fodor's notion of meaning identity in contradistinction to meaning similarity [41].) The NMF mechanism, on the other hand, builds heavily on the notion of similarity between concept-model and input data. Hence, in a problem where the input space has no metric the use of NMF can be somewhat problematic. Second, the input data (context plus accompanying word; see Fig. 1) changes erratically in time whereas the NMF equations (see Eq. (5) for example) are applied to the case where the inputs are time-independent, otherwise the very notion of maximizing the global similarity would be meaningless.

In this paper we have tried to circumvent those obstacles by modifying some aspects of the cross-situational learning scenario while maintaining the NMF mathematical formulation intact. In particular, we assume that the numerical differences between the objects and words labels are in fact *bona fide* measures of similarities, which have already been used in the study of compositional and structured meaning-word mappings [39], [40]. As for the random dependence of the input data on time, we choose to average over infinitely many turns of the guessing game and keep only the relative frequencies of occurrence of the object-word pairs. These frequencies are then used to weigh the importance of the object-word pairs in the NMF categorization mechanism (see section III). The loss of temporal correlations involved in this averaging procedure, however, resulted in a low signal-to-noise ratio which prevented the NMF algorithm to find and stabilize the correct object-word mapping. A more successful approach was based on a divide-and-conquer strategy in which we first use the NMF mechanism to classify the different situations (context plus accompanying word) and then use a second NMF system to extract the common object label in the contexts that belong to the same concept-model of the first NMF hierarchy.

A more promising approach to apply NMF to cross-situational learning, however, may be to modify the NMF formalism so as to handle more naturally situations where the notion of similarity is extraneous and the input time dependence is stochastic. Interestingly, in the learning algorithms for cross-situational learning the measure of similarity is a self-organized quantity, similarly to $f(k \mid i)$, that emerges from the repeated interactions between the agents (see [21]-[24]). Work in this line is in progress.

Acknowledgment

J. F. F. thanks the hospitality of the Adaptive Behaviour & Cognition Research Group, University of Plymouth, where this research was initiated. The visit was supported by euCognition.org travel grant NA-097-6

References

[1] F. de Saussure, *Course in General Linguistics*. Translated by Wade Baskin. New York: McGraw-Hill Book Company, 1966.

[2] I. Ulbaek, "The origin of language and cognition," in *Approaches to the Evolution of Language*, J. R. Hurford, M. Studdert-Kennedy and C. Knight, Eds. Cambridge: Cambridge University Press, 1998.

[3] M. D. Hauser, *The Evolution of Communication*. Cambridge: MIT Press, 1997.

[4] S. Pinker, *The Language Instinct*. London: The Penguin Press, 1994.

[5] L. I. Perlovsky, "Integrating language and cognition," *IEEE Connections*, vol. 2, pp. 8-13, 2004.

[6] L. I. Perlovsky, "Integrated Emotions, Cognition, and Language," *Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN06*, pp. 1570 –1575, 2006

[7] V. Tikhanoff, J. F. Fontanari, A. Cangelosi and L I. Perlovsky, "Language and Cognition Integration Through Modeling Field Theory: Category Formation for Symbol Grounding," *Lecture Notes in Computer Science*, vol. 4131, pp. 376-385, 2006.

[8] A. Cangelosi, V. Tikhanoff, J. F. Fontanari and E. Hourdakis, "Integrating Language and Cognition: A Cognitive Robotics Approach," *IEEE Computational Intelligence Magazine*, vol. 2, pp. 65-70, 2007.

[9] D. Marr, *Vision*. San Francisco: Freeman, 1982

[10] P. J. Kellman and E. S. Spelke, "Perception of partly occluded objects in infancy," *Cognitive Psychology*, vol. 15, pp. 483-524, 1983.

[11] J. F. Fontanari and L. I. Perlovsky, "Object perception in the Neural Modeling Fields framework," *Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN08*, pp. 1775-1780, 2008.

[12] S. Harnard, "The symbol grounding problem," *Physica D* , vol. 42, pp. 335-346, 1990.

[13] A. Cangelosi, A. Greco, and S. Harnard, "Symbol grounding and the symbolic theft hypothesis," In: A. Cangelosi, D. Parisi (Eds.), *Simulating the Evolution of Language*, London: Springer-Verlag, pp. 191-210, 2002.

[14] J. F. Fontanari and L.I. Perlovsky, "Categorization and symbol grounding in a complex environment," *Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN06*, pp. 10039 –10045, 2006.

[15] J. R. Hurford, "Biological evolution of the Saussurean sign as a component of the language acquisition device," *Lingua*, vol. 77, pp. 187-222, 1989.

[16] B. J. MacLennan, "Synthetic ethology: an approach to the study of communication," in: C. G. Langton, C. Taylor, J. Doyne Farmer, S. Rasmussen, Eds. *Artificial Life II*, SFI Studies in the Sciences of Complexity, vol. X . Addison-Wesley, Redwood City, pp. 631-658, 1991.

[17] A. Cangelosi, "Evolution of Communication and Language using Signals, Symbols and Words," *IEEE Transactions on Evolutionary Computation*, vol. 5, pp. 93-101, 2001.

[18] L. Steels, "Evolving grounded communication for robots," Trends in Cognitive Science, vol. 7, pp. 308-312, 2003.

[19] J. F. Fontanari, and L. I. Perlovsky, "Evolution of communication in a community of simple-minded agents," *Proceedings International Conference on Integration of Knowledge Intensive Multi-Agent Systems*. IEEE Press, Boston, pp. 285-290, 2005.

[20] L. Steels and F. Kaplan, "Spontaneous lexicon change*," Proceedings of COLING-ACL*. Morgan Kaufmann, San Francisco, pp. 1243-1250, 1998.

[21] A.D.M. Smith, "Semantic generalization and the inference of meaning," *Lecture Notes in Artificial Intelligence*, vol. 2801, pp. 499-506, 2003.

[22] T. Lenaerts, B. Jansen, K. Tuyls and B. de Vylder, "The evolutionary language game: An orthogonal approach," J. Theor. Biol. vol. 235, pp. 566-582, 2005.

[23] J. de Beule, B. de Vylder, and T. Belpaeme, "A cross-situational learning algorithm for damping homonymy in the guessing game," In: L.M. Rocha, M. Bedau, D. Floreano, R. Goldstone, A. Vespignani, L. Yaeger, Eds. *Proceedings of the Xth Conference on Artificial Life*, The MIT Press, Cambridge, 2006.

[24] J.F. Fontanari and L.I. Perlovsky, "Meaning creation and communication in a community of agents*," Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN06*, pp. 2892 –2897, 2006.

[25] K. Smith, S. Kirby and H. Brighton, "Iterated Learning: a framework for the emergence of language," *Artificial Life*, vol. 9, pp. 371-386, 2003.

[26] H. Brighton, K. Smith and S. Kirby, "Language as an evolutionary system," *Phys. Life Rev*. vol. 2, pp. 177-226, 2005.

[27] V. Loreto and L. Steels, "Emergence of language," *Nature Physics*, vol. 3, pp. 758-760, 2007.

[28] P. Bloom, *How children learn the meaning of words*. Cambridge: MIT Press, 2000.

[29] J. M. Siskind, "A computational study of cross-situational techniques for learning word-to-meaning mappings," *Cognition*, vol. 61, pp. 39-91, 1996.

[30] S. Pinker, *Learnability and cognition*. Cambridge, MA: MIT Press, 1989.

[31] C. Fisher, G. Hall, S. Rakowitz and L. Gleitman, "When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth," *Lingua*, vol. 92, pp. 333-375, 1994.

[32] L. I. Perlovsky, "Language and Cognition," Neural Networks, 4(3), accepted for publication.

[33] L. I. Perlovsky, *Neural Networks and Intellect: Using Model-Based Concepts*, Oxford: Oxford University Press, 2001.

[34] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimization problems," *Biological Cybernetics*, vol. 52, pp. 141-152, 1985.

[35] G. Bilbro, R. Mann, and T. Miller, "Optimization by mean field annealing," in: D. S. Touretzky , Ed. *Advances in Neural Information Processing Systems I*. Morgan Kaufmann Publishers, San Mateo, CA, pp. 91-98, 1989

[36] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, pp. 671-680, 1983.

[37] J. F. Fontanari and L. I. Perlovsky, "Meaning Creation and Modeling Field Theory," *Proceedings of the IEEE Conference on Integration of Knowledge Intensive Multi-Agent Systems KIMAS 05*, pp. 405 –410, 2005

[38] H. Akaike, "A new look at the statistical model identification", *IEEE Transactions on Automatic Control*, vol. 19, pp. 716-723, 1974.

[39] J. F. Fontanari and L. I. Perlovsky, "Evolving compositionality in Evolutionary language Games," *IEEE Transactions on Evolutionary Computation*, vol. 11, pp. 758-769, 2007.

[40] J. F. Fontanari and L. I. Perlovsky, "A game theoretical approach to the evolution of structured communication codes," *Theory in Biosciences*, vol. 127, pp. 205–214, 2008.

[41] J. Fodor and E. Lepore, "All at Sea in Semantic Space: Churchland on Meaning Similarity," The Journal of Philosophy, vol. 96, pp. 381-403, 1999

[42] P. M. Churchland, "Conceptual Similarity across Sensory and Neural Diversity: The Fodor/Lepore Challenge Answered," The Journal of Philosophy, vol. 95, pp. 5-32, 1998

[43] B. Abbott, "Fodor and Lepore on Meaning Similarity and Compositionality," The Journal of Philosophy, vol. 97, pp. 454-455, 2000.

[44] J. F. Fontanari and L. I. Perlovsky, "How language can help discrimination in the Neural Modelling Fields framework," *Neural Networks*, vol. 21, pp. 250–256, 2008.