Rapporto Tecnico
Technical Report

n. 96-07

A representational redescription method using competitive learning

Alberto Greco, Angelo Cangelosi

Università degli Studi di Genova
Dipartimento di Scienze Antropologiche
Sezione e Laboratorio di Psicologia

Pubblicazione del
Dipartimento di Scienze Antropologiche (DISA)
Università di Genova
V. Balbi, 4 - 16126 Genova

DISA - University of Genoa, Italy

A representational redescription method using competitive learning *

Alberto Greco
Angelo Cangelosi

greco@igecuniv.csita.unige.it
angelo@caio.irmkant.rm.cnr.it

Abstract

In a previous simulation a network learned labeling a small set of stimuli in three input conditions (visual features, label, label + visual features), classifying them according to colour, function, object name. Results suggested that in different input conditions the network internal representation reflects the explicit semantic structure of stimuli. Evidence on the mediating role of linguistic label was suggested by cluster analysis: in the three input conditions a single object is represented very similarly but it has different representations in the label + features condition, depending on the label.

One limit of this kind of models is that transferring acquired knowledge to other tasks would require network retraining. A second shortcoming is that this model allows only one level of knowledge representation. Empirical evidence shows that knowledge must be represented at different levels, ranging from implicit to full explicit symbolism. A similar view has been suggested by the "representational redescription" hypothesis (Clark & Karmiloff-Smith, 1993).

In order to test how to meet these requirements, we augmented our model requiring the network to use the already acquired knowledge to extract the semantic structure of the stimulus set, for each of the three subtasks. The hidden unit layer was connected to a new module with three clusters of output units. Each output cluster had to make explicit the structure in each of the three categorization subtasks. To train this new module, all the other connection weights were frozen except the connection from the hidden layer to the new output units. These connection weights were trained with the competitive learning algorithm.

The results show that the network is able to exploit previously acquired knowledge and to make explicit the stimuli semantic structure using the hidden (implicit) representation. This structure corresponds to that obtained by cluster analysis in the previous research. This method can be considered a first step in testing the representational redescription hypothesis. It will require further exploration and testing of more complete models. Some related issues are discussed, such as the controversial need of hybrid models.

1 Introduction

The present research is the sequel of a general project that is aimed at studying how language affects categorization. The general statement that language influences thought seems to meet a wide agreement in the psychological literature. Such a statement, however, is too vague, since both terms, language and thought, may have different aspects. The *linguistic*

* Paper presented at the ESSCS 14th Annual Workshop (17-20 April 1996), St.Maximin-La S.te Baume, France.

influence, for example, can be understood either as the effect of having a lexicon, that is a system of labels systematically coupled with concepts or of symbols that stand for concepts, or having something more complex, such as a syntax made of some rules for composing symbols. *Thought* is also a complex concept: it may refer to categorizing or using concepts, to reasoning or making inferences, to problem solving, etc...

The research presented here is focused on a much more simple phenomenon than a general language-upon-thought influence, that is the influence of learning a lexicon (i.e., labels for objects) along with learning to discriminate features and creating categories for objects. A central question is: what happens when not only perceptual features but also linguistic labels are available? It is worth to remark that *linguistic* labels are not simple labels, like a car numberplate, because they are *systematic* and *arbitrary***.** Linguistic labels are a special input, because they act as a "comment" on what is perceived, a comment which is systematically coupled with features that are to be detected and categorized, but which is arbitrary because there is no *a priori* rule for predicting which labels are coupled with perceptual aspects, if not experience. What a child has to learn in such a situation, in fact, is what Piaget (1936, 1937) called a coordination of schemata. There are two representations, probably of a different kind, which at the start are independent, and that are not superimposed but must be somehow coordinated. Perhaps the best way to study this coordination is to consider categorizing and naming as two functions which must be put in relation. And the first issue here concerns whether we can conceive such functions as independent.

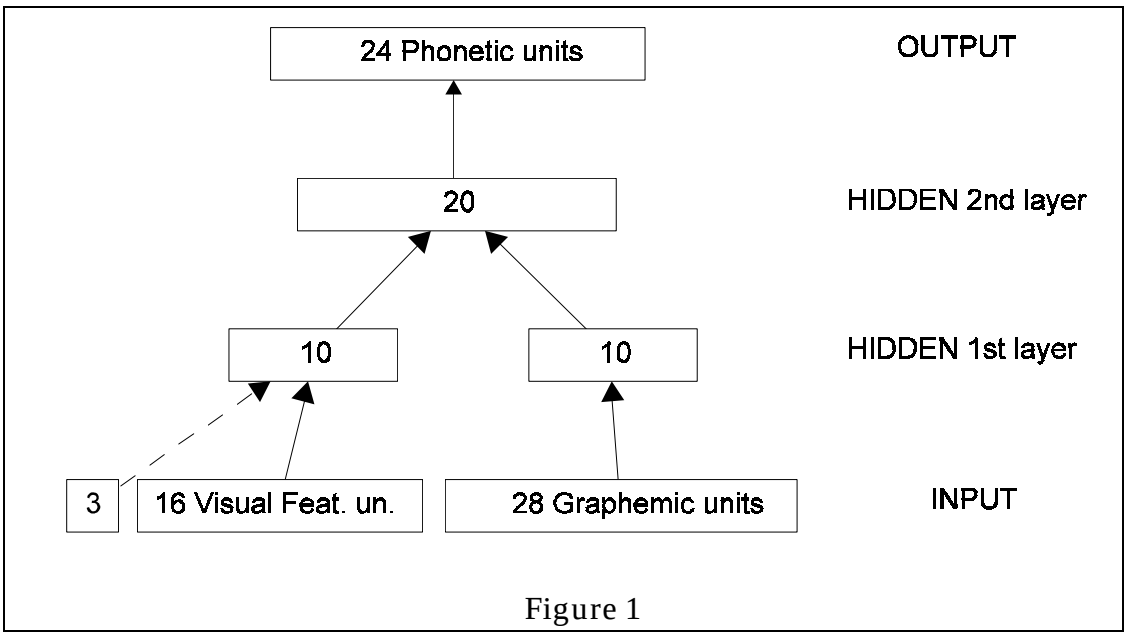

Figure 1

In a previous research (Greco & Cangelosi, 1996) we did a simulation that supported the idea that categorizing and naming are not independent. In this research a neural network learned a naming task. The situation was similar to that of a child who learns to read different labels while seeing objects or pictures. These labels can show the name of the object, or of its colour, or of its category. Sometimes labels may be presented along with objects (which we call an *obstension* situation), other times *labels only* or *objects only* may be input. The network was trained in three input modes corresponding to three conditions:

1) **object+label**, when the input to the network consisted of the object visual features plus the object name;
2) **label** condition when the only input consisted in object label;
3) **object** condition with the input presentation of the object visual features.

In the object-only condition, the network received an extra contextual flag indicating which aspect attention had to be focused on and which name (object, colour, category) had to be output. The object *input* consisted of a code representing simple visual features. A phonetic representation of the name was used as *output*. The architecture consisted of a four layer feed-forward neural network (see figure 1), trained using the supervised learning algorithm error back-propagation (Rumelhart & McClelland, 1986). The input layer had two groups of units, the feature units and the graphemic units, plus three input units that acted as the context flag in the feature-only input condition. The first hidden layer had two separate groups of 10 hidden units. The second hidden layer had 20 units, while the output layer had 24 phonetic units. After learning, we studied the network internal representation, using cluster analysis, in order to see whether the activation values of the hidden layer units (we focused on the second hidden layer) organized themselves in patterns reflecting the semantic structure of stimuli.

The two results of main interest are:
(1) a single object was represented very similarly across the three different input conditions. This means that stable categories were created;
(2) considering the ostension condition label+features, a single object had different representations, depending on the label.

From this analysis we can argue that the network did not simply learn to read but also to categorize items, and that this categorization was affected by the label presence or absence. When there are *names* along with other perceptual features, this can affect the whole object representation. Categorizing and naming are not independent functions.

This result, however, shows only an interaction between names and categories in producing a composite implicit representation, which is not the only way language can influence categorization. One additional hypothesis is that language helps categorization because helps it to become more explicit.

In other words, in a first time - when categorization occurs - regularities and similarities are implicitly detected in the input. The role of language would be manifest in a second time, when linguistic labels make categorization easier because they contribute in making *explicit* those regularities that previously were detected only implicitly.

2 The *representational redescription* hypothesis

In a system like the one we previosuly described, there is no way to make explicit the internal network representational structure. As a consequence, the acquired knowledge is not available to the system for further use in other tasks, for example in problem solving. In order to exploit this acquired knowledge, at least a network retraining would be necessary (a special retraining for each new task), if not the design of a fresh new structure. A second limit of this kind of model is clear if we consider some results of empirical research on implicit learning and in developmental psychology which suggest that knowledge must be represented at different levels of explicitedness.

If we briefly consider *implicit learning* experiments (Reber, 1989; Seger 1994; Lewicki et al., 1992), these were usually carried out with tasks like artificial grammar learning, probability learning, covariation learning. Such experiments show that subjects can acquire knowledge about structure or about rules, which they exhibit only in their behaviour, as it is not available to consciousness for a direct report. For example, subjects can make judgments on new stimuli or they can be more accurate in new tasks, exploiting knowledge abstracted from previous presentations of related stimuli. Even in such cases, subjects show a clear ability to transfer implicit knowledge to different tasks where the surface structure is changed but the deep structure remains the same. Current connectionist models are not able to do so, because their representations are linked to particular stimuli, not to abstract rules. Implicit knowledge probably is not all-or-none but it is represented at different levels of implicitness / explicitness.

In developmental psychology, an example is given by Clark & Karmiloff-Smith (1993, p.497) of French children who are able to mark the distinction between two different uses of the French word ‘un’, which means either the indefinite article ‘a’, or the numeral ‘one’. There is a level in development when French children make explicit this distinction, which previously was only implicit, saying ‘**un** petit four’ (**a** cake) or ‘**un de** petit four’ (**one** cake). At a next level, and in the adult age, the partitive ‘de’ is only used as an emphatic

These examples reveal the now usual distinction between nonsymbolic knowledge, which implicitly influences behaviour, and symbols that can be found only at explicit levels. In psychological but also in connectionist

literature there is a discussion on how to reconcile those two kinds of knowledge. One hypothesis by Karmiloff-Smith (Karmiloff-Smith, 1992; Clark & Karmiloff-Smith, 1993) states that symbolic or explicit knowledge is extracted from the composite or implicit representation using a process of redescription, which has been called *representational redescription* (RR). Even if such terminology may be misleading and simply "re-representation" could be a better expression, this hypothesis seems very attractive. The basic idea is that implicit, already acquired, knowledge would be explicitly available to other parts of the cognitive system by means of a process of recoding it into a new format.. Different levels of redescription are envisaged. At the first level, termed "Implicit" (I), the system can use some procedural knowledge as a whole but cannot access its parts. The "Explicit-1" (E1) level is a first redescription level of representations in a simpler, more flexible and general-purpose format. At this level, parts of procedures are available to the system but not to consciousness - which is only possible at the "Explicit-2" (E2) level - and to verbal report, possible at the E3 level. It is also hypothesised that in this process knowledge is "reduced" and some original detail is lost, but new representations are redundant in the system, they do not replace old formats. The appealing aspect of this hypothesis is that the implicit-explicit dichotomy is overcome and multiple levels of explicitness are envisaged, corresponding to multiple levels of redescription. In our opinion, this can give new insights on old hypotheses about language-thought relationships, such as *differentiation* (Werner & Kaplan, 1963) and *microgenesis* (Werner, 1957; Draguns, 1983).

Clark & Karmiloff-Smith proposed also that the connectionist approach is a suitable method to investigate RR. They suggested to use a mechanism proposed by Mozer & Smolensky (1989) for "skeletonization" of successful trained-up networks, by identifying and deleting hidden units which result least relevant for performance. Differently from those authors, the original network would not be replaced by the "pruned" one, which would be a duplicate designed for use in new tasks. As an alternative implementation, they suggest to augment current connectionist models by adding some mechanism that allows knowledge re-representation, like in the Finch & Chater (1991) model where cluster analysis is used as an explicit, symbolic description of a trained network internal representation.

Other authors (Shultz, 1994; Brook, 1995) proposed the cascade-correlation algorithm as a good connectionist model to test the RR hypothesis. These autors suggest a direct relation between the learning phases envisaged in the cascade-correlation models and levels of knowledge re-representation, such as the first error-driven phase would correspond to the implicit (I) level, the correlation-driven phase to the intermediate E1 level, and the second error-driven phase to the more explicit representations E2 and E3. However, as also Karmiloff-Smith (1994) claimed, such a model overestimates the role of error reduction, a process which makes sense in mastery learning, whereas RR can occur independently of behavioural mastery.

3 A competitive learning architecture for RR

We have tried to implement a different solution in our model. We augmented our model setting up a new task (which can be considered as a side-task) for our, previously trained, naming network. The new task is to make explicit the stimuli categories according to the main naming task being performed. To make explicit here means to activate a local symbolic output corresponding to the category presently being named.

We used the *competitive learning* algorithm (Rumelhart & McClelland, 1986, p.151) to test this model. The competitive algorithm is an unsupervised learning technique for feature extraction. The network is trained to autonomously select and activate only one unit in the cluster of output units. It is usually called the *winner-takes-all* method. The selected output unit represents the common feature of the group of stimuli that activated it.

By using this method we expect that the new competitive learning module is able to extract the data structure starting from the network hidden representation. Since, as we have said, the naming task and the categorization task are related, then a common internal representation should be usable by different networks which perform those tasks.

The hidden unit activations of the previously trained network is the input to the output units in the competitive module. We can imagine this module as including one cluster of units for each category. In each cluster only one unit is activated, i.e. wins, according to the feature being represented. After the new training, this module, to act as a representational redescription method, should be able to extract the input semantic structure firing the appropriate units. We sorted out the 72 stimuli in 3 groups of 24 data, according to the three main categorization tasks (object naming, colour, category or function). For each of these stimulus groups, we separately trained the corresponding cluster in the competitive learning module.

Thus, the overall network architecture consisted of the previous naming network, plus a cluster of competitive output units connected to the second layer of hidden units, the units where visual feature and label information is integrated (figure 2).
The output cluster had four output units for the naming data, two units for the function subtask data, and two units for the colour subtask.

During the competitive training, the connection weights of the naming network were frozen. The connection weights from the second layer of hidden units to the cluster of competitive output units were changed according to the learning algorithm. For each data set, the simulation run reached a stable state after about 100 (one hundred) epochs.

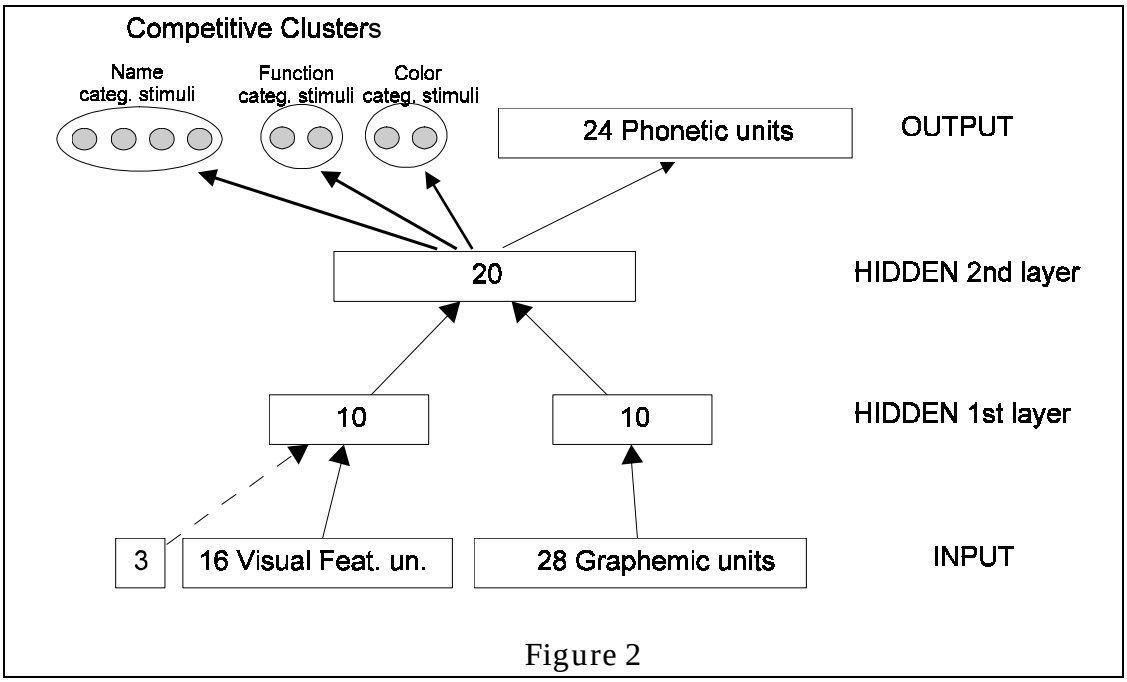

Figure 2

4 Results

At the end of the competitive learning, a testing phase was accomplished. We checked out that the data were grouped according to the semantic structure built in the stimuli set. 21 out of the 24 objects in the *naming* task were correctly classified in 4 groups (i.e. pen, ink, nut, axe) with a result of 87.5% correct / expected classification. 23 out of the 24 (96%) of the stimuli in the *colour* condition properly activated two units, one for the red objects and one for the green objects. 20 out of the 24 stimuli in the *function* condition were grouped in the two sets of tools and office objects (83% correct classification).

To avoid a possible bias in extracting the correct number of stimulus classes, due to the preset number of units in the output cluster, we rerun the three simulations using larger output clusters. We used 6 units for the four-classes subtask (object name) and 4 units for the two-classes subtasks (colour and function). The results were the same. Indipendently of the number of output units, the classification reflected the object semantic structure, and the extra output units were not used by the network.

5 Discussion

As we expected, the two tasks, labeling and classification, are very related. Then it is easy to extract the relevant information from the hidden representation of the stimuli. Beyond this result, we think that some important issues are to be considered and discussed. The first issue regards the question: What exactly the representational redescription *output* should be? In our model, it is a *local output* with clear-cut units for single concepts. It is clear, however, that it is not plausible to imagine a plethora of different units for the infinite concepts that can be implicitly thought and then made explicit. Then our output units are a first approximation, useful to test the competitive algorithm in this model, but output units clearly should be replaced, perhaps with symbolic language tokens. This, in turn, leads to ask other questions (what kind of tokens and what kind of language? something like the external natural language or an intermediate "language of thought"?). In any case, it seems reasonable to hypothesize that output units should be composable and reusable.

A second issue is that the model tries to make it clear that redescription depends on the *context* (or on what its use is). In our system this is modeled by the fact that the task currently being performed served as a prompt for redescription. A corollary is that we can have different redescriptions of the very same hidden state, depending on the context. This is another reason why we need composable and reusable output units. Of course, if even the same internal state can be redescribed in different ways, a local output is still less conceivable.

Following this line of reasoning, it is natural to think that *context* can influence the very process of redescription and differentiation of implicit into the explicit. But what is *context*? If it is something *mental*, it should be represented along with the main concept. In this case it is legitimate to ask *how* is it represented. For example we can ask whether contextual knowledge, in turn, is implicit or explicit. One possibility is that context is a part of the overall implicit knowledge. Then it could act as an agent which takes part in the process of redescription. For example its task could be to select what has to be redescribed.

But if we considered implicit knowledge (and its contextual part) as coded in a language of thought, then consciousness would not be oriented by nonsymbolic processes but by internal language, as Vygotsky (1962), Luria (1961), and other psychologists, who studied the role of self-direction, had envisaged. A related question: Is there an endogenous pressure to re-code or is it suggested by contextual knowledge? There is a series of fascinating hypotheses that could be tested by means of new models.

A last question concerns the very nature of these models. Even if we are to accept the need for a process of redescription like the one described by Clark and Karmiloff-Smith (1993), as we have seen, there is a discussion about how

to implement it. Is a full connectionist system enough, or a hybrid system should be devised? We think that this is a pseudo-problem. In fact, the important thing is to determine whether symbols are necessary or not. We think that there is no doubt that composable elements are needed, that act *like* words. How they are implemented is less important.

No doubt that only explicit symbols become available for introspective awareness. This is like meta-knowledge, which is now so in fashion in cognitive and educational psychology. Meta-knowledge could be a process of redescription which enables to put order (for example to introduce time constraints) in low-level knowledge, according to the context.

This is the right task for a competitive procedure. By competitive procedure here we mean not only the competitive lerning algorithm, but general control mechanisms where active representations are selected by means of inhibition of alternative ones. Of course one can question whether a competitive mode applies *after* supervised training or whether competition is present from the start. Our opinion is that supervised and competitive algorithms should be both used in the same simulation, at different stages. We think of a model like a sandwich where unsupervised and supervised modules are alternated. For example, the basis for any supervised learning is discrimination, which can be learnt by competitive procedures. But after learning new competitive processes can take place, like we have tried to show in our simulation (and since redescriptions can be corrected on the basis of some external or internal parameter, the process can go on further). We must think that auto-oriented and hetero-oriented processes continuously interact, if we want to think of a truly self-organizing system.

References

Brook J. (1995) Cascade correlation as a model of representational redescription. In Howell A.J., Wood J.A. (eds.) *The eighth white house papers: graduate research in the cognitive & computing sciences at Sussex*. Brighton, UK: University of Sussex, School of Cognitive & Computing Sciences. Research paper CSRP 390.

Clark A. & Karmiloff-Smith A. (1993) The cognizer's innards: a psychological and philosophical perspective on the development of thought. *Mind & Language*, 8, 487-568.

Draguns J.G. (1983) Why microgenesis? An enquire on the motivational sources of going beyond the information given. *Archiv für Psychologie*, 135, 5-16.

Finch S. & Chater N. (1991) A hybrid approach to the automatic learning of linguistic categories. *AISB Quarterly*, 78, 16-24.

Greco A. & Cangelosi A. (1996) Language, categorization, and representation: a pilot study using neural networks. *Technical Report* n. 96-03, Dip. di Scienze Antropologiche (Dept. of Anthropological Sciences), Univ. of Genoa.

Karmiloff-Smith A. (1992) *Beyond modularity: a developmental perspective on cognitive science*. Cambridge, Mass.: Bradford (MIT Press).

Kamiloff-Smith A. (1994). Transforming a partially structured brain into a creative mind. *Behavioral and Brain Sciences*, 17, 732-745.

Lewicki P., Hill T., Czyzewska M. (1992) Nonconscious acquisition of information. *American Psychologist*, 47, 6, 796-801.

Luria A. (1961) *The role of speech in the regulation of normal and abnormal behavior*. New York: Liveright.

Mozer M.C. & Smolensky P. (1989) Using relevance to reduce network size automatically. *Connection Science*, 1, 3-16.

Piaget J. (1936) *La naissance de l'intelligence*, Neuchâtel-Paris, Delachaux et Niestlé.

Piaget J. (1937) *La construction du réel chez l'enfant*, Neuchâtel-Paris, Delachaux et Niestlé.

Reber A.S. (1989) Implicit learning and tacit knowledge. *Journal of ExperimentalPsychology: General*, 118, 219-235.

Rumelhart D.E. & McClelland J.L. (eds.) (1986) *Parallel distributed processing.* Vol 1: Foundations. MIT Press, Cambridge, Mass.

Seger C.A. (1994) Implicit learning. *Psychological Bulletin*, 115, 163-196.

Shultz T.R. (1994) The challenge of representational redescription. *Behavioral and Brain Sciences*, 17, 728-729.

Vygotsky L.S. (1962) *Thought and language.* New York: Wiley & Sons.

Werner H. (1957) The concept of development from a comparative and organismic point of view. In Harris D.B. (ed.) *The concept of development*. Minneapolis: University of Minnesota Press.

Werner H. & Kaplan B. (1963) *Symbol formation*. New York: Wiley & Sons.