LANGUAGE AND THE ACQUISITION OF IMPLICIT AND EXPLICIT KNOWLEDGE: A PILOT STUDY USING NEURAL NETWORKS

Alberto Greco, University of Genoa, Italy
Angelo Cangelosi, University of Plymouth, UK.

LANGUAGE AND THE ACQUISITION OF IMPLICIT AND EXPLICIT KNOWLEDGE: A PILOT STUDY USING NEURAL NETWORKS

Alberto Greco, University of Genoa, Italy
Angelo Cangelosi, University of Plymouth, UK.

Abstract

In experimental psychology there is wide evidence that language supports thinking. How this "support" works, however, is still not clear. One hypothesis is that categorization is easier when linguistic labels are available, because implicitly detected similarities and rules can be made explicit. We want to test this hypothesis using a neural-network simulation.

Language is not a common sensorial input, but acts as a "comment" on the world (Parisi, 1994). When linguistic labels are systematically coupled with objects, either of the two inputs can elicit one single response (e.g. articulating a name). In real situations labels can be names for the objects or may denote specific features or functions of them.

We constructed a neural-network simulation which learned to label a small set of stimuli in three input conditions (visual features, label, label + visual features), classifying them according to colour, category, object name. Network internal representations were analyzed by using cluster analysis in order to show the influence of linguistic cues in categorization. In the three input conditions a single object was represented very similarly but it had different representations in the label + features condition, depending on the label. These results support evidence on the mediating role of linguistic labels. Future development lines and model improvements are discussed.

In order to test how the "representational redescription" hypothesis (Clark & Karmiloff-Smith, 1993) can be implemented to allow the explicitation of previously acquired knowledge, we augmented the model requiring the network to use the already acquired knowledge to extract the explicit 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 and the new 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. Language, categorization, and representation in connectionist models

In this paper a pilot study is presented of a general project aimed at studying the role of linguistic labelling on categorization. Asking about the relationships between language and categorization is in a sense something new and in a sense something old. It is new if we consider that in the cognitivist paradigm it seemed not to make much sense to ask about the relationships between language and other cognitive processes, simply because mental activity in that paradigm consisted in representing the world by using symbols which worked similarly to words, and were manipulated according to rules similar to grammatical rules. Thus the question about the relationships between language and nonsymbolic processes was not an issue. On the other hand, however, this question is not so new, because it was posed - surely in a different form but perhaps substantially with a similar content - in classical psychology, namely in the chapter concerning the relationships between language and thought.

According to a classical psychological theory, dating back at least to Vygotsky, language substantially supports thinking. This idea had been endorsed also by the representationalist view, typical of the "human information processing" approach, but it has been accepted only in a particular version, inasmuch as it was believed that cognition is only possible if internal symbols are available for coding and processing information (hence the "language-of-thought" hypothesis, which concerns how thought is formally coded rather than how this coding affects the development of thought).

There is a wide classical set of psychological experiments, some confirming that verbal coding helps short-term memory storage, some saying that verbal information helps recall from long-term memory (these experiments are in every handbook of psychology: e.g. see Conrad, 1964; Bransford & Franks, 1971). None of these and others, however, show a necessary relationship between language and thought. At least, not the way it had been hypothesized in the so-called "linguistic relativism", that is according to the strong Whorfian hypothesis which says that language controls thought and perception. Yet we know that studies in categorical perception, on the contrary, have shown that we are able to categorize at least some attributes of the world (e.g. colours) without linguistic support (cf. Bornstein, 1987) and there is evidence that some mental operations are independent from language, as one can see for example from the difficulty in getting "thinking aloud" protocols during processes, like problem solving (Ericsson & Simon, 1993).

However, the weaker hypothesis, which says that language "influences" thought, has never been rejected. On the contrary, today there is a revival on this subject: for example, there is a recent area of research, about "implicit" or "tacit" knowledge (Reber, 1989; Seger, 1994), where the distance between what one knows and what one can tell is being explored, and here the role of language in categorization again is an important issue to be clarified.

In sum, from the psychological literature, it seems that there is agreement upon the fact that there is at least an influence of language on thought, though not deterministic.

What still remains to be clarified is the way language influences thought. One hypothesis is that linguistic labels make categorization easier, because they help in making explicit regularities and similarities that were previously only implicitly detected in the cognitive system. Following Werner (1963), a classical psychologist close to Gestalt and to the so-called "organismic" psychology, this path, from implicit states to explicit ones, may be called "microgenesis", that is the development of thought.

The final aim of our project is to test this hypothesis by using a neural-network (NN) simulation. But the first step toward this achievement is to explore how is it possible to simulate, by using NN, the linguistic influence on categorization (though non-deterministic, as psychological literature has shown). The work being presented here tries to take this first step. Initially we are interested to know how the implicit linguistic-dependent knwoledge emerges; the next step will be to study how this knowledge can become explicit and available for further cognitive tasks.

The connectionist approach has already been used to simulate the role of language in categorization. The NN categorization capabilities have been well established since longtime, and we know that networks can efficiently extract features from input, recovering its categorical structure. Also to establish stimulus-response-like associations between labels and contents is not difficult: given the name, features can be retrieved and converse (e.g. see early pattern completion models implemented by using interactive activation, in Rumelhart & McClelland, 1986).

For example, one straightforward kind of model is exemplified by Nosofsky et al. (1992), known as the ALCOVE system, where categories are learned by means of a repeated association of exemplars with their name. The task is to decide how much the stimulus belongs to a category (by computing its similarity with previous exemplars) firing the output nodes that represent the appropriate category.

The main problem with connectionist research, however, is that it still gives little help in clarifying some problems that exist with the psychological relationships between categorizing and naming. We believe that this happens because, in fact, connectionist research perhaps has neglected that language is not a common sensorial input. Language is not like other objects in perception, but it has something special, because it acts as a "comment" on the world (Parisi, 1994).

The typical tasks where language is used as a comment on the world are of this sort: first, linguistic labels are systematically coupled with objects; then at a subsequent presentation only one of the two inputs, all alone, can elicit some specific response (for example articulating a name). The important thing to be considered in this situation, in our opinion, is the fact that to perceive an object and an object + a comment on it (in the simplest case, perceiving it with a linguistic label) are different cases. In the second case the same thing does not happen as before plus a second thing, but it is just a new thing. From the representational point of view, the question is if the object+label situation does elicit an old representation plus something more, or rather a fresh new representation.

This problem appears in some recent connectionist work, e.g. Miikkulainen & Dyer (1991) and Schyns (1991). They suggest a very simple solution on how language and categories could interact. As usual in this kind of networks, categorical features are extracted from the sensory properties of input, and they are coded in the network's hidden units: we can say this corresponds to the representation of types. But what about token labels? The proposed solution is that each time a new occurrence or token is encountered, the representation corresponding to its type is cloned and the individual identifier (that is, the name) is simply "attached" to it, constructing a sort of twofold representation. However, in our opinion this approach is too simple.

We can assume that when visual features and labels of objects (arbitrarily coupled) are input, at some low level different representations are created for visual features and for the label. In Piagetian terms we perhaps can speak of two different schemata. But those two representations (or schemata) are not just superimposed, rather they must be coordinated, because - as we have said - the coupling is arbitrary (that is, there is no rule for predicting which labels are coupled with objects). So the idea of putting together in a single representation categorical features and unique tags for single instances, as Miikkulainen and Dyer do, looks unnatural because it ignores this need for coordination, wich is not superimposition.

The Schyns' solution, in turn, is to have separate networks for categorizing and for naming. This is not a bad idea in itself; the problem however is that naming is not a completely independent task because, as we have seen, it often takes place simultaneously with categorizing. We shall give more examples later.

In a sense it is true that categorizing and naming are independent functions. Schyns reminds us of this fact, but the very fact that networks exist which can categorize without using labels is further evidence of it. Categorizing and naming are independent but related functions, however. Related because it is generally admitted (also by Schyns) that having labels can make category construction or retrieving easier. But why this happens is unclear. Then the real problem is how those two functions work and are related.

In our opinion, to work out this problem, it is necessary to consider that:
1. the process of naming in real life often starts "by ostension" (that is label+object is presented, as a parent does with her child when she points to an apple and says something, perhaps "apple", perhaps "red" or "good" or "it's to eat", etc.), depending on the context;
2. categorizing when also labels are given is "special", and is not the same case as when no labels are given. As we have seen, to perceive an object or an object + a label are different cases.

We speak of labels and not of names because, as the example shows, in real situations labels can be names for the objects (apple) but they also may denote specific features or functions of them. For example features (red) or a function (something to eat).

What about internal representations? It seems reasonable to hypothesize that, at higher levels, they are affected by both inputs and thus the internal representation of a unique object must be different depending on the particular label that occurs with it (this would mean that language can mediate perception). On the other hand, the same old problem is that there must be something common in representations for similar objects, and in representations for similar features, otherwise categorization or concepts could not occur at all. But if our hypothesis holds, the common parts in this composite representation are not clear-cut separable parts.

2.   Simulation 1: The emergence of Implicit Knowledge

Our simulation tries to reproduce a situation that concepts or objects have features in common.The idea is to construct a NN which has to learn to output a label in different input conditions, and then to analyze its internal representations in order to show the influence of linguistic cues in categorization. This simulation is a first pilot study, where a small set of stimuli and a simple neural architecture have been used. We hope to use the results of this initial study for designing a more complex and more comprehensive model.

The aim of our model is to simulate the influence of a linguistic stimulus (label) on the internal representation of a physical object, and on the feature-extraction process for the categorization task. The different conditions are set by presenting the network sometimes with objects+labels (what we can call an ostension situation), at other times with objects or labels only. The modeled situation is similar to the one previously described, where a child learns to read different labels while seeing objects or pictures. These labels can show the name of the object or its colour or its category; sometimes the child sees only objects, at other times only labels.

When the model sees some label (alone or with the object), its task is to read this label, when it sees only the object's picture it receives an extra signal (a context flag) that indicates where attention must be directed to, or what it has to say (that is, the object's name, or colour or category). Learning occurs because it is corrected each time it is wrong. Given that during the training the labels may refer to different aspects of the input (name, colour, category), the network does not simply learn to read but at the same time it learns to categorize.

The stimulus set (Figure 1) includes four different objects: axe, nut, pen, ink. For each object, the visual features and the linguistic labels are coded according to the representation used in Plaut & Shallice (1993). For linguistic labels a localistic representation of the graphemes of the object name is used; for visual features the representation is distributed.

The task is to output the label corresponding to one of three subtask requests: (i) name of the object, (ii) name of its functional category, and (iii) name of the colour of the object.

Visual features

		Component Shape		Size	Colour
	Main	*Second*	*Third*		
AXE	box-thin	taper-to-point	cylinder-long	1 to 2 feet	red (or blue)
NUT	cylinder-short	hole		less 3 inches	red (or blue)
PEN	cylinder-long	taper-to-point	top	3 to 6 inches	red (or blue)
INK	cylinder-hollow	top	liquid	less 3 inches	red (or blue)

Binary coding

SHAPE	CODE	SIZE	CODE
box-thin	0 0 0 1	less 3 inches	0 1
cylinder-short	0 0 1 0	3 to 6 inches	1 0
cylinder-long	0 1 0 0	1 to 2 feet	1 1
cylinder-hollow	1 0 0 0		
hole	0 0 1 1	COLOUR	CODE
taper-to-point	0 1 1 0		
top	1 1 0 0	red	0 1
liquid	0 1 1 1	blue	1 0

Example

red AXE	box-thin	taper-to-point	cylinder-long	1 to 2 feet	red
	0 0 0 1	0 1 1 0	0 1 0 0	1 1	1 0

Figure 1. Stimulus set (adapted from Plaut & Shallice, 1993)

This according to the label being read or according to the contextual flag. As the output, a phonetic representation of the name is used.

The four objects belong to two different functional categories (pen and ink to OFFICE, and axe and nut to TOOLS. Each object is presented to the network in two different colours (blue and red) so we can say there are 8 objects. The number of each object exposition to the network in one epoch is three times (for the name, category, and colour subtask request). Moreover, each object is presented to the network in three different input conditions:

- presentation of only the object visual features (F condition) with an extra input for one of the three subtask requests;
- visual presentation together with the linguistic label referring to one subtask request (F+L),
- label-only presentation (L).

Thus the total number of input conditions is 72 (4 objects × 2 colours × 3 subtasks × 3 input conditions).

The neural architecture (Figure 2) consists of a four-layer feed-forward network. There are 47 input units, 28 for the label, 16 for the visual features and 3 for the flag used for specifying the subtask request only in the F (features only) condition. In output there are 24 localistic phonetic units. The first hidden layer consists of two separate groups of units, each processing the visual features of the object or the label. The second layer of hidden units receives input from both the units groups of the lower hidden layer. The use of such a network structure has been suggested by Parisi e.a. (Parisi, Pagliarini & Floreano, 1994) to allow an arbitrary coordination of schemata.

Five different simulations were run, starting with new random weights, each time producing very similar results. The data here shown come from only one of these simulations, they are not weighted data.

The network was trained, by the back-propagation algorithm, for 1000 epochs, and it learned the task after a few hundreds epochs, reaching a very low error level (Figure 3). After the training, a test was made to check the error for each of the 72 conditions, and it was always lower than 1 percent.Then we made a study of the network internal representation in order to show which kind of "semantic" representation each object activates in the different input conditions. To achieve this, a cluster analysis of the units activation values for the objects in all the input conditions was made for the second hidden layer.
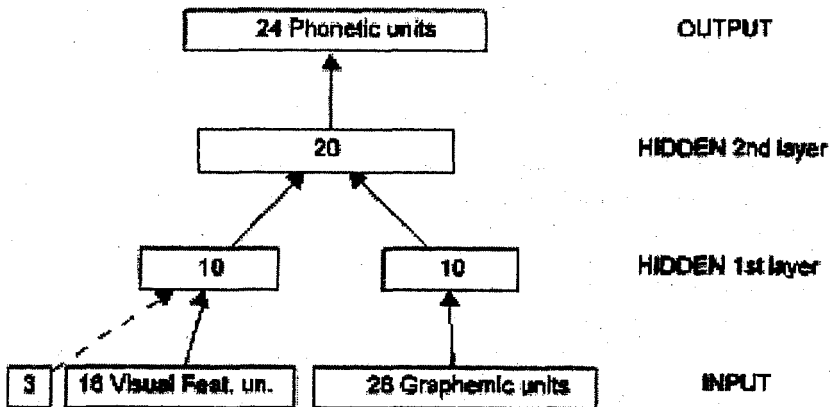

Figure 2. Neural architecture

Results

1. The general dendrogram (partially shown in Figure 4, only for the name subtask) suggests that an internal representation emerges, reflecting the explicit semantic structure that was used in the construction of the training stimulus set.

2. We can ask how similar the network internal representations are in the different input conditions. The question is whether the three input conditions elicite different internal representations or only one. If we look at the similarities between representations (considering the distance at which clusters are formed) in the dendrogram shown in Figure 4, we see that there is no difference for the same object in the three input conditions. The label+feature condition (e.g. "INK"+ink) and, noteworthy, the label-only presentation ("INK"), that is only linguistic, activate a "semantic" representation very similar to that used by the network when the physical features of the object are presented. A cluster analysis was also separately made for the two unit groups of the first hidden layer. The dendrograms show that for the units group of the visual features, the network builds a different activation pattern for the four objects, while each object activates a similar pattern in all the different input conditions. The group of units for the label input activates different representations for all 8 labels.

3. We also wanted to study how a single object is represented. Does language have a role in differentiating the semantic representation of a single object? We did a separate cluster analysis considering only the 24 input conditions where label+object features (L+F) occurred. Also in this case we analyzed the activation values of the second hidden-layer units.
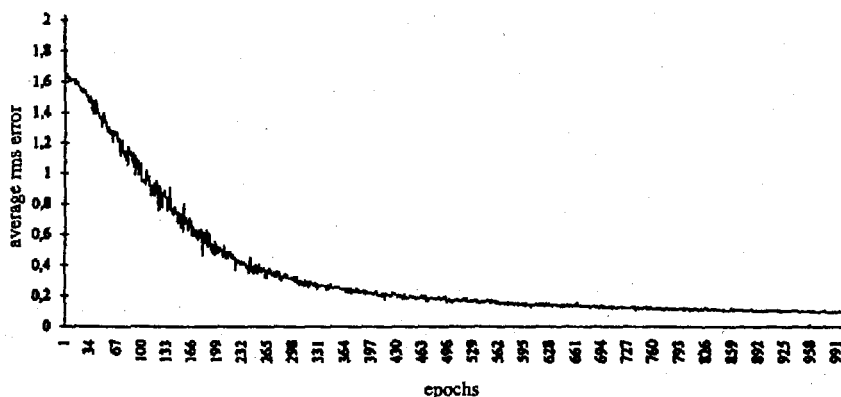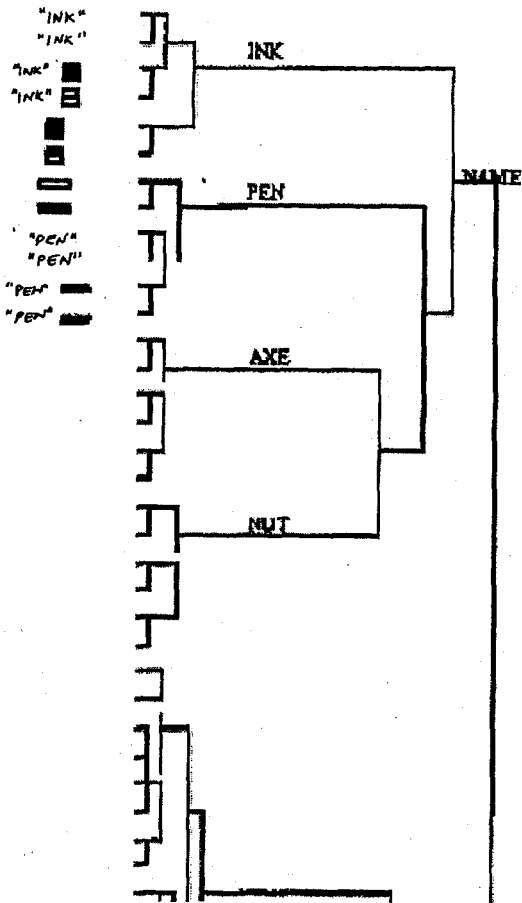

Figure 3. Error during training

Figure 4. Hierarchical cluster analysis (zoomed on the name subtask)

The dendrogram in Figure 5 shows that the input corresponding to the visual features of a single object activates three internal representations which are different according to the three categorization subtasks. For example, the blue pen input - that is, the visual features of the blue pen - when presented with the label "blue" activates the "blueness" units, the "pen-ness" units when the label is "pen" etc. This clearly happens because of the presence of the linguistic label together with features. Remark that the network does not always look at the language; in fact, the same network is able to properly categorize items from only visual information.

The results reported here come from a very simple and limited model. However, these pilot simulations encourage the design of a more complex model for the study of the linguistic input role in categorization. Some future development lines for improving this model could be:

- to use more realistic representations of the label codes (e.g. by using phonetic inputs, or a more linguistic-like input);
- to use also a more complex stimuli set, and different categorization and naming tasks;
- to adopt the lesion method to analyze the role of hidden units. We have already tried a preliminary lesion study of the single units in the second hidden layer and it seems to show that some units have a different role in some of the three input conditions. But this must be done more carefully. In fact, since we obtained similar results with a reduced network, we think that some unit gives no contribution, and that there may be redundancy in our network.
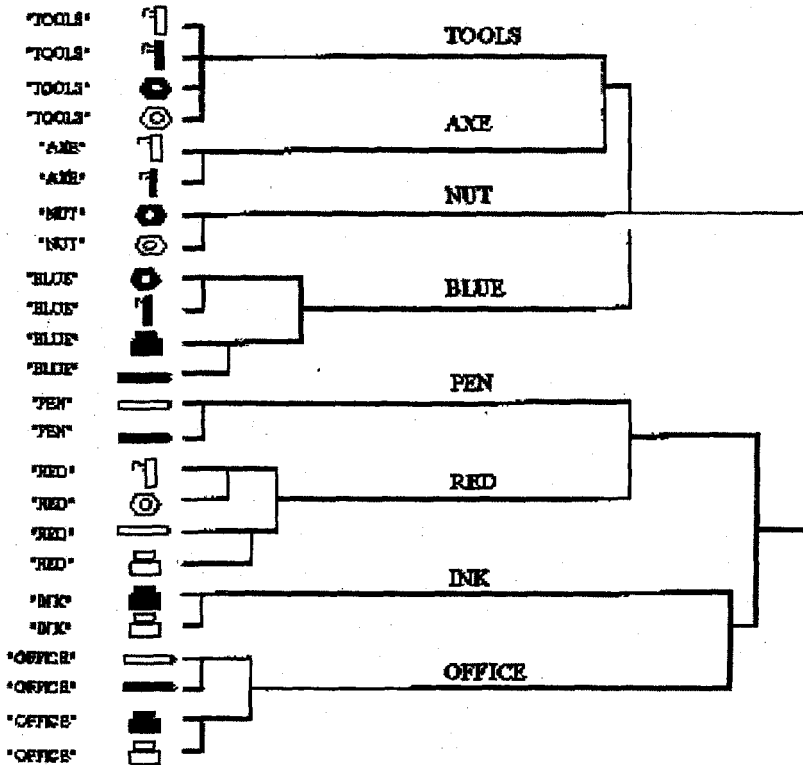
Figure 5 - Hierarchical cluster analysis in the label+feature condition

3. The representational-redescription hypothesis

The previous results support the assumption that this model is able to simulate the way linguistic information is used as a relevant property of the world that we perceive, and it is consistent with the general idea that the way we semantically organize the different objects into categories is affected by language.

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

The second step of our project is then aimed at finding some method that allows what we can call a "microgenetic" analysis of the network processes, that is to show how language works in helping to make explicit similarities and regularities that are automatically detected by the categorization system.

Before trying such a model, we needed a categorization system to include language. Secondly the transition from implicit to explicit had to become clear.

In a system like the one 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, as 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 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 in different degrees on a scale 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 this distinction explicit, 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 by 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 the least relevant for performance. They propose not to replace the original network 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, in wich 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.

4. Simulation 2: The emergence of explicit knowledge

We have tried to implement a different solution in our model. We augmented our model by 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 stimulus 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 are integrated (Figure 6). 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.
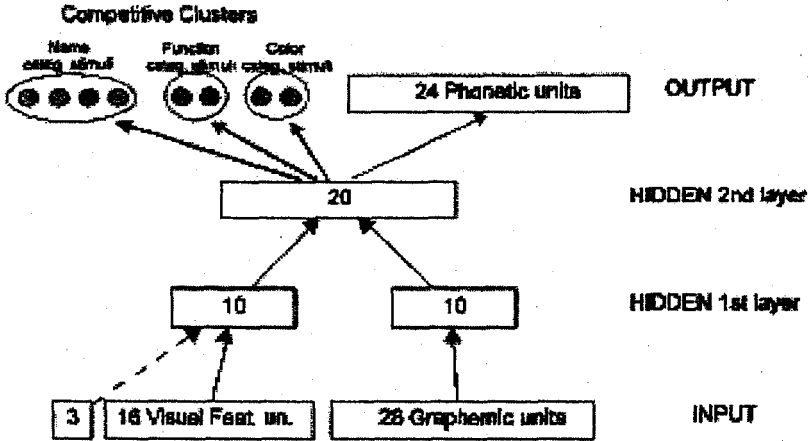

Figure 6 (see text)

Results

At the end of the competitive learning, a testing phase was performed. We checked that the data were grouped according to the semantic structure built in the stimulus 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. Independently 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 should the representational redescription output 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 number of 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 asking 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 consider 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 should a hybrid system 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.

There is 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 alternate. 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 continue). We must think that auto-oriented and hetero-oriented processes continuously interact, if we want to think of a truly self-organizing system.

References

Bornstein M.H. (1987) Perceptual categories in vision and audition, In Harnad S., *Categorical perception*, Cambridge Univ. Press, Cambridge, Mass.

Bransford J.D., Franks J.J (1971) The abstraction of linguistic ideas, *Cognitive Psychology* 2, 331-350.

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.

Conrad R. (1964) Acoustic confusions in immediate memory, *British J. of Psychology* 55, 75-84.

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

Ericsson K.A., Simon H.A. (1993) *Protocol analysis: Verbal reports as data*, Cambridge, Mass: MIT Press.

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

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

Karmiloff-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.

Miikkulainen R., Dyer M.G. (1991) Natural language processing with modular PDP networks and distributed lexicon, *Cognitive Science* 15, 343-399.

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

Nosofsky R.M., Kruschke J.K., McKinley S.C. (1992) Combining exemplar-based category representations and connectionist learning rules, *J. of Experimental Psychology: Learn.,Mem.,and Cogn.* 18, 2, 211-233.

Parisi D., Pagliarini L., Floreano D. (1994) Verso un modello neurale dell'apprendimento del linguaggio: imitazione e coordinamento arbitrario di schemi, In Laudanna A., Burani C., *Il lessico: processi e rappresentazioni*, NIS Nuova Italia Scientifica, Firenze.

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.

Plaut D.C., Shallice T. (1993) Perseverative and semantic influences on visual object naming errors in optic aphasia: a connectionist account, *J. of Cognitive Neuroscience* 5, 89-117.

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

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

Schyns P.G. (1991) A modular neural network model of concept acquisition, *Cognitive Science* 15, 461-508.

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.

Addresses author:

Alberto Greco
Dip. di Scienze Antropologiche (DISA)
Laboratorio di Psicologia
Via Balbi, 4
16126 Genova, Italy
e-mail: greco@disa.unige.it

Angelo Cangelosi
School of Computing
University of Plymouth
Drake Circus
Plymouth PL4 8AA, UK

e-mail: angelo@soc.plym.ac.uk