Rapporto Tecnico
Technical Report

n. 96-03

Language, categorization, and representation:
a pilot study using neural networks

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

Language, categorization, and representation: a pilot study using neural networks *

Alberto Greco
Angelo Cangelosi

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

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 NN simulation.

Language is not a common sensorial input, but acts as a "comment" on the world (Parisi, 1994). When linguistic labels are sistematically 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 denotate specific features or functions of them.

We constructed a NN which learned labeling a small set of stimuli in three input conditions (visual features, label, label + visual features), classifying them according to color, category, object name. Network internal representations were analyzed 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.

Introduction

In this paper a part 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 using symbols which worked so similarly to words, and were manipulated according to rules so similar to grammatical rules. Thus the question about the relationships between language and nonsymbolic processes was some sort of nonsense. 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

* Paper presented at the ESSCS 13th Annual Workshop (5-8 Sept. 1995), Oxford, England.

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 has 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 STM storage, some saying that verbal information helps recall from LTM (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. colors) 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 using a neural network (NN) simulation. But the first step toward this achievement is to explore how is it possible to simulate, using NN, the linguistic influence on

categorization (though not-deterministic, as psychological literature has shown). The work being presented here tries to take this first step, that is to see how the *implicit* emerges; the next step will be to study how it becomes explicit.

Connectionist research on categorizing and naming

The connectionist approach has been already 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 the converse (e.g. see early pattern completion models implemented using interactive activation, in Rumelhart & McClelland, 1986).

For example, one straightforward kind of model is exemplified by Nosofsky et al. (1992), known as 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 given in input 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 sistematically 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 it does not happen the same thing 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. Among others it is worth mentioning two articles that appeared in *Cognitive Science*, one by Miikkulainen & Dyer (1991) and one by Schyns (1991). The former authors suggest a very simple solution concerning 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 occurence 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 appears too simple.

We can assume that when object's visual features *and* labels (arbitrarily coupled) are input, at some low level different representations are created for visual features and for the label. In Piagetian terms perhaps we 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 could not be 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 this fact, but it seems needless to say: 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 himself) that having labels can make *easier* category construction or retrieving. 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 mother does with her child when she points out at 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", 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 denotate 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.

Simulation

Our simulation tries to reproduce such a situation. Our idea is to construct a NN which has to learn to output a label being trained 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 fact that the network internal representation of a physical object can be affected by the presence of a linguistic stimulus (a label) and that this label can direct the feature extraction process for the categorization task.

The different conditions are set by presenting the network sometimes **objects+labels** (what we can call an *ostension* situation), other times **objects only** 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 color *or* its category; sometimes the child sees only objects, 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 or what it has to say (that is, the object's name, or color or category). Learning occurs because it is corrected each time is wrong. Considering that during the training the labels may refer to different aspects of the input (name, color, category), the network does not simply learn to read but at the same time it learns to categorize.

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

Figure 1 - Stimuli set
(adapted from Plaut & Shallice, 1993)

Visual features

	Component Shape			Size	Color
	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
box-thin	0 0 0 1
cylinder-short	0 0 1 0
cylinder-long	0 1 0 0
cylinder-hollow	1 0 0 0
hole	0 0 1 1
taper-to-point	0 1 1 0
top	1 1 0 0
liquid	0 1 1 1

SIZE	CODE
less 3 inches	0 1
3 to 6 inches	1 0
1 to 2 feet	1 1

COLOR	CODE
red	0 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

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 color of the object. 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 colors (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 color 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).

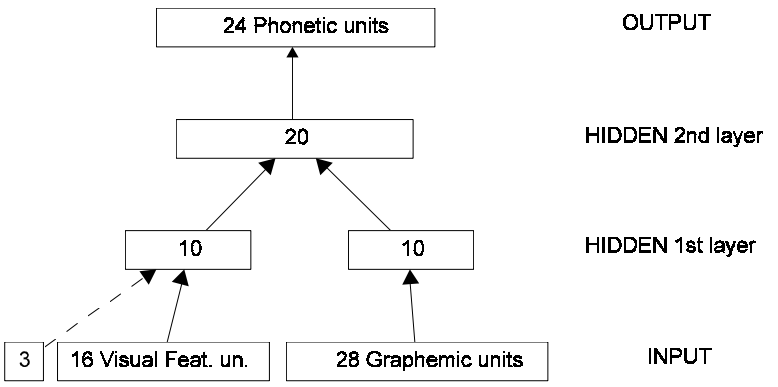
Thus the total number of input conditions is 72 (4 objects X 2 colors X 3 subtasks X 3 input conditions).

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 group of units, each processing the visual features of the object or the label. The second layer of hidden units receive input from both the units groups of the lower hidden layer. The use of such a network structure has been suggested by Parisi and colleagues in one of their works (Parisi, Pagliarini & Floreano, 1994) and it is required considering that the task implies an arbitrary coordination of schemata.

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

The network was trained, using the back-propagation algorithm, for 1 thousand (1000) epochs, and it learned the task after 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 stimuli 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 the network uses when the physical features of the object are presented.

A cluster analysis was also separately made for the two units 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.

Figure 3 - Error during training

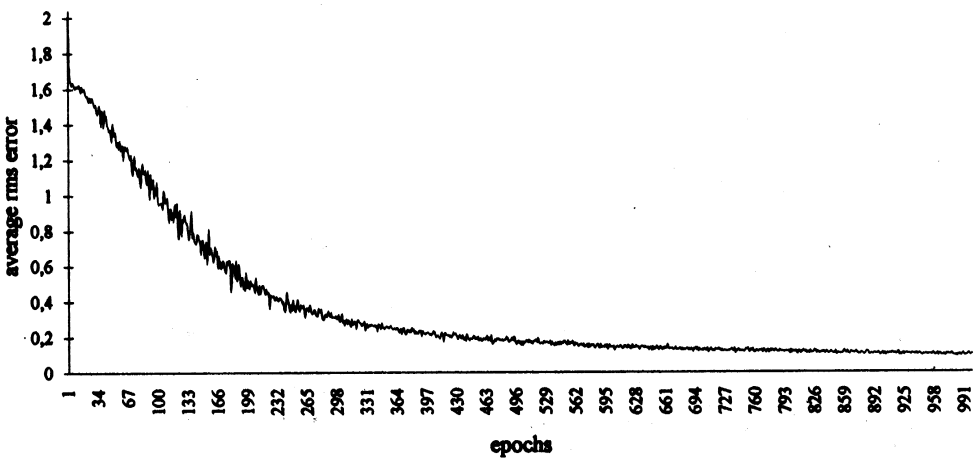
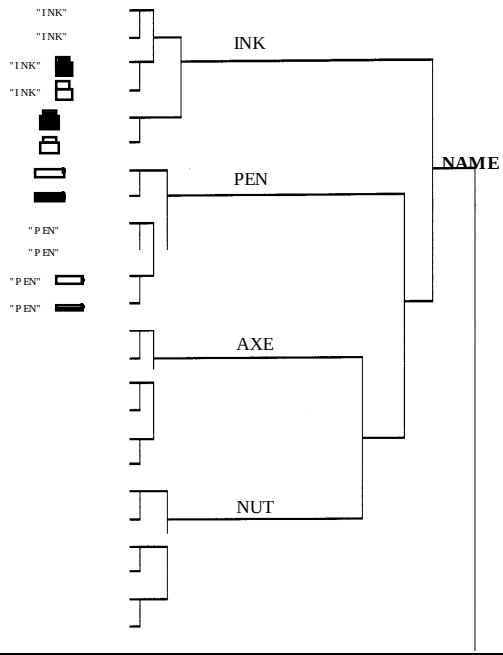


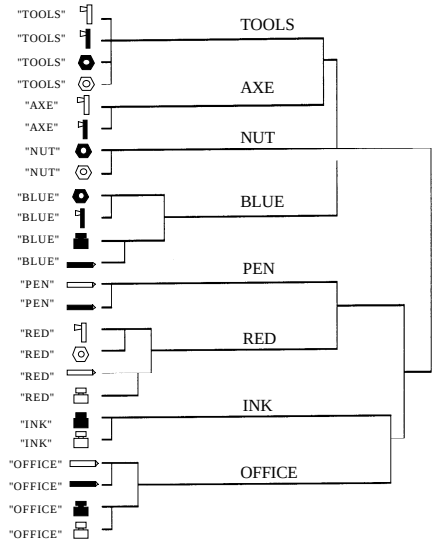
Figure 4 - Hierarchical cluster analysis (zoomed on the *name* subtask)



3) We also wanted to study how a *single* object is represented. Does the 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.

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.

Figure 5 - Hierarchical cluster analysis in the *label+feature* condition



Discussion

These 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.

The results here reported 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 allow the system itself to be able to extract information that presently we read using cluster analysis. This can be done by adding a new group of units which codes the features; the network should be trained with a non-supervised algorithm (e.g. with competitive learning) so that only the winner feature wins.

- 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 must think that some unit gives no contribution, and that there may be redundancy in our network.

These and other enhancements of this model are possible. But the most important enhancement is beyond this specific model and concerns the second step of our project. Its aim will be to find 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 that, differently from others, in obtaining the "implicit" representation did not work independently from language. The second part is to see how the implicit becomes explicit, available for other tasks. We certainly need more powerful techniques for the analysis of the network internal representation, but we need also more powerful *systems*.

For this second part of the research, we can profit from some very good proposals expressed by Clark and Karmiloff-Smith (1993). We find that their idea of "representational redescription" is appealing for exploration. We have tried to construct a model where language can direct categorization. It was the first step in order to investigate *how* this happens.

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.

Clark E.V., Karmiloff-Smith A. (1993) The cognizer's innards: a psychological and philosophical perspective on the development of thought. *Mind & Language*, 8, 487-519.

Conrad R. (1964) Acoustic confusions in immediate memory. *British Journ. Psychol.*, 55, 75-84.

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

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

Nosofsky R.M., Kruschke J.K., McKinley S.C. (1992) Combining exemplar-based category representations and connectionist learning rules. *Journal 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 (*Towards a neural model in language learning: imitation and arbitrary coordination of schemata*) in Laudanna A., Burani C., Il lessico: processi e rappresentazioni, NIS Nuova Italia Scientifica, Firenze.

Plaut D.C., Shallice T. (1993) Perseverative and semantic influences on visual object naming errors in optic aphasia: a connectionist account. *Journal of Cognitive Neuroscience*, 5, 1, 89-117.

Reber A.S. (1989) Implicit learning and tacit knowledge. *Journal 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.

Werner H. (1963) *Symbol formation*. Wiley, New York.