First words learning: A cortical model

Extract of paper that appeared in Cognitive Computation 2010, Volume 2, Number 3, Pages 217-229
(with Mazzone M. and Plebe A.)

Vivian De La Cruz - vdelacruz@unime.it
Dept. of Cognitive Sciences, Educational and Cultural Studies

Abstract

Many hypotheses have been proposed to explain the learning of first words with some emerging from the particular characteristics observed in child development. One is the peculiar trend in the speed with which words are learned, that have been referred to in the literature as "fast mapping". Using a neural network model trained in stages that parallel developmental ones and that simulate cortical processes of self-organization, we explored early stages of first word learning. This was done taking into account strictly visual and acoustic perceptions only. The results obtained show evidence of the emergence in the artificial maps used in the model, of cortical functions similar to those found in the biological correlates in the brain. Evidence of non-catastrophic fast mapping based on the quantity of objects and labels gradually learned by the model was also found. We interpreted these results as meaning that early stages of first word learning may be explained by strictly perceptual learning processes, coupled with cortical processes of self-organization and of fast mapping, without invoking specialized word learning mechanisms, at least not at these early stage in the word learning process.

1. Introduction

Humans come to recognize an infinite variety of natural and man-made objects in their lifetime and make use of sounds to identify and categorize them. How does this life-long learning process begin? By the time children are 8–10 months old their visual object categories are sufficiently stable and flexible to be used as the foundation for labeling and referencing actions. In the first year of life, acoustic perception becomes language specific, and children acquire the segmental inventory and many phonotactic regularities of their mother tongue. Increasing amounts of evidence point to the growing capacity of infants at this stage to reliably map arbitrary sounds onto meanings, and this mapping process is crucial to the acquisition of language. Many of the hypotheses proposed to explain the learning of first words invoke special processes other than those involving development driven by sensorial experience. The reasons behind this are many, with some emerging from the particular characteristics observed in child development. One is the peculiar trend in the speed with which words are learned that have been referred to in the literature as "vocabulary spurt" and "fast mapping". Vocabulary growth is initially very slow, but suddenly speeds up at around 18 months of age. Once at this stage, children grasp aspects of the meaning of a new word on the basis of only a few incidental exposures (Carey 1978, pp. 264–93; Dickinson 1984, pp. 359–73). Another is the tendency to extend object names on the basis of shape similarity, as opposed to other visual cues (Landau et al. 1988).

By design, our model avoids invoking any "special mechanism" to explain the early acquisition of first words. What it does is simulate brain processes during an early crucial stage of development, by taking into account strictly visual and acoustic perceptions only. What follows, briefly summarizes explanations that have invoked mechanisms other than strictly perceptual ones.

An idea that has been influential is that categorization involves essentially the reference to natural (and functional) "kinds" and that representing kinds implies the exercise of specific capacities other than simply registering perceptual regularities in the environment. This idea spread widely in psychology after the notion of "natural kind" was introduced in the philosophical literature by (Putnam, 1975; but also see Kripke, 1972). What they claimed is that when we name natural kinds (animals, chemical substances, etc.) we want our names to refer to the very essence of those kinds, irrespective of the contingent representations we have formed of them.

While in the philosophical literature, these positions were intended to set apart ontology (category individuation) and conceptualization, with conceptualizations performing solely a contingent role in the probabilistic assessment of category membership, but with no role in defining the essential structure of the category, in psychology a different conclusion has been drawn. The idea of natural kinds endowed with an essential structure has been taken to mean that categories are not just based on a simple evaluation of perceptual similarity, there must be a deeper common structure for people to treat a set of objects as a real category, and our conceptualizations have to be attuned to this structure, which is mainly conceived as a theoretical core based on causal relationships (Murphy et al 1985). Some scholars have proposed that in humans there must be an innate predisposition to detect some of these causal structures: this could explain how infants can initially form the basic domains of objects, narrowing the hypothesis space with respect to the patterns of features to be focused on first (Carey et al 1996). Other scholars have emphasized the importance of causal relations in later stages of child development for instance, the ability to detect hidden causal powers would systematically lead the process of category formation in 3-year-olds (Gopnik et al 2003). This approach has been generalized by also applying it to functional categories, and to artifacts, primarily. In this case, the deep structure of the category is not thought to be based on natural (causal) properties, rather it would depend on what objects are for, that is, how they can be used by us: it is the function of objects that seems to drive categorization at two-years of age and after (Kemler Nelson et al 2000).

Obviously, those positions are controversial, and the overall debate is quite complex. Mandler (2004) denies the need for the innate core knowledge claimed by Carey and Spelke (1996), and assumes that the first trends in category formation in infants could emerge from environmental regularities and physiological constraints. This would mean that causal and functional properties could be abstracted away from the sensorial input we are exposed to. Nonetheless, she thinks that low-level sensitivity to environmental regularities is not all there is to category formation; one should also consider a different learning mechanism, consisting in attentive focusing on certain features. In other words, there would be a genuine difference between a purely perceptual learning process and a truly conceptual one, with only the latter delivering representations that are accessible to deliberate thought. Eimas and Quinn (1994), are critical of the distinction Mandler makes between the perceptual and conceptual categorization mechanisms. Nazzi and Bertoncini (2003), specifically address the difference in learning speed before and after the "vocabulary spurt", and propose that the transition corresponds to a shift from an associational to a referential lexical acquisition mechanism. Only the latter would allow the acquisition of genuine words as links between phonetically specified sound patterns and object categories. Booth and Waxman (2002) and Diesendruck and Bloom (2003), argue that the shape bias is in itself a mechanism for learning new object names, rooted in the understanding that shape is a reliable cue to the object kind. Gershkoff-Stowe and Smith (2004) and Smith (2005), challenge Booth and others' idea that the shape bias is a learning mechanism, they instead propose it as being a developmental product of children's acquisition of language that presents systematic co-variations between visual appearances and acoustic features.

Finally, we should mention what Tomasello has proposed (1999; 2003), according to which inductive generalization alone cannot explain the way humans form linguistic categories. Tomasello's issue is not how we categorize objects per se, but rather how words can be mapped onto objects (events, actions, etc.). This sort of mapping, however, can play a major role in subsequent category formation, since some categories could be formed as a consequence of word learning, rather than existing prior to it. What Tomasello proposes, is that this mapping takes place not by induction alone, but rather by induction plus a sensitivity to social cues (direction of eye gaze, etc.), which reveals to the child what the speaker is focusing on.

From our point of view, we do not deny that in building lexical categorization, humans must certainly rely on a great variety of skills that go beyond what is overt in visual and acoustic signals. The purpose of our model, however, is to try to assess to what extent sensorial experiences alone can account for early word learning. Similar models have been proposed in the past, as demonstration of what is "computationally possible" to achieve by perception alone (Regier, 2005; Rogers et al 2003). They have provided important evidence of how much categorization can develop on the basis of perception, including phenomena such as the vocabulary spurt, fast mapping, and the selection of categorical salient features. Still, one may suspect that those results have been achieved thanks to a combination of wholly artificial phonological or semantic features and powerful algorithms, which are far removed from the reality of human development. We are attempting to work one time step behind, by searching for what is "brain-possible" to achieve by perception alone. That is to say, we are trying to exclude from the model any algorithmic power whatsoever that might go beyond what a brain can do. On the other hand, the model attempts to reconstruct in a plausible way, the human cortical architecture responsible for the visual and acoustic paths of the process. We believe that in shifting our perspective on inner brain mechanisms, certain dichotomies used in the above summary of theoretical positions, might fade away. One is the distinction between what is called "associative" learning and something else, for example, referential, attentive, or based on causation, learning, by virtue of the well-known Hebbian principle (Hebb 1949) that is yet today acknowledged as describing most of the neural representation capabilities, which sees learning as triggered by the repeated temporal coincidence of different stimuli. As far as one class of stimuli becomes predictor of other stimuli on a regular basis, the connection between these classes naturally strengthens. Under this perspective, it is not surprising, that after the initial discovery that sound patterns might point to object categories, the same kind of association will gradually become easier, not to mention stronger.

We would add that recent studies have shown that the vocabulary spurt is not so much of a spurt at all, in that it instead demonstrates a gradual increase rate without an inflection point (Ganger et al 2004), and that fast mapping is a common feature of neural learning, not specific to object naming (Markson et al 1997). Similarly, for what concerns causation knowledge, there is a tradition both in philosophy and in psychology that has emphasized its perceptual basis rather than the idea that causes cannot be observed (Michotte 1963; Searle 1983). At the very least, it appears that there are perceptual cues, which are responsible for the fact that people come to perceive an event as causally determined (Scholl et al, 2000). Attentive phenomena are certainly at play during learning and are different from learning itself, even though established connections of stimuli facilitate later attentional selection of similar stimuli. Yet, Pruden et al. (2006), have demonstrated that while 10-month-old infants are sensitive to social cues, they cannot recruit them for word learning. In any case, attention in this context acts essentially as a preliminary selector of what object is referred to by a name in the visual field of the learner, after this selection, what happens is perceptual learning.

2. The model

Our artificial model of name and object learners, is a collection of neural networks organized as a simplified version of the visual and auditory paths in the cortex. In the computational structure of the model, there is an overall minimum of mathematical design, specific to the functions to be acquired. Most of the effort is in the inclusion of mechanisms of plasticity and in reproducing a correct hierarchy of cortical maps.

We simulate plasticity in the cortex, and in how cortical maps are organized as a result of developmental processes, using a self-organization mathematical framework. This approach has been the object of several proposals for artificial neural networks. The first implementation was proposed by von der Malsburg (1973), and Willshaw and von der Malsburg (1976), in models of the development of aspects of the visual system, based exclusively on the local interaction of neurons. The difficulties of the system of differential equations in this early formulation made it unsuitable for building cognitive models. Since then, much progress has been achieved, and today several models are available for the self-organizing development of cortical maps.

In our model, we use a combination of two different approaches, one more suitable for replicating maps physically existing in the cortex, and another for abstracting higher-level functions that are not carried out in a well-defined area in the brain, but most likely distributed over many areas in ways that still remain unclear.

For the first purpose, we chose a mathematical abstraction of cortical maps, which is to a degree, faithful enough in reproducing a biological learning mechanism, through the combination of Hebb's principle and neural homeostasis, yet simple enough to allow the building of high level models: the LISSOM (*Laterally Interconnected Synergetically Self-Organizing Map*) architecture (Bednar 2002; Sirosh et al 1997), a two-dimensional arrangement of neurons, with intracortical excitatory and inhibitory connections.

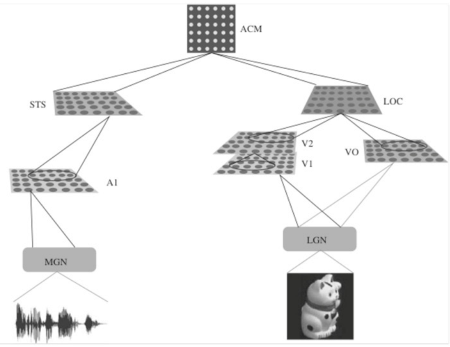
The other type of artificial network included in the model is the Kohonen Map or SOM (*Self-Organizing Map*) that achieves self-organization through a simple winner-takes-all mechanism (Kohonen 1995). The winner-takes-all mechanism is a significant departure from the behavior of biological cortical circuits. It works as a mathematical substitution for the effect of lateral neural connections, but only assuming fixed connections and uniform neighborhoods. However, it is an efficient non-supervised representational device, useful at a level of abstraction higher then single cortical maps. This network is used in the model as an abstract conceptual map.

2.1 Model components

An outline of the modules that make up the model is shown in Fig. 1. There are two main paths, one for the visual process and another for the auditory channel. Both paths include thalamic modules, which are still partially driven by development. However, since the detailed shape of their functions is not relevant in the scope of this study, all subcortical processes have been hardwired according to what is known regarding their functions.

In the visual path, LGN is implemented with simple on-center and off-center receptive fields. More precisely, there are three pairs of sheets in the LGN maps: one connected

Fig. 1 Overall scheme of the model. The auditory path, on the left, is composed by MGN (*Medial Geniculated Nucleus*), A1 (*Primary Auditory Area*), STS (*Superior Temporal Sulcus*). The ventral visual path, on the right, is composed by LGN (*Lateral Geniculated Nucleus*), V1 (*Primary Visual Area*), V2 (*Secondary Visual Area*), VO (*Ventral Occipital Area*), and LOC (*Lateral Occipital Complex*). At the top, the abstract map ACM (*Abstract Conceptual Map*



to the achromatic intensity image plane, and the other two connected to the medium and long wavelength planes. In the color channels, the internal excitatory portion of the receptive field is connected to the channel of one color, and the surrounding inhibitory part to the opposite color. The cortical process proceeds along two different streams: the achromatic component is connected to the primary visual map V1 followed by V2, the two spectral components are processed by map VO, the color center, also called hV4 or V8. The two streams rejoin in the cortical map LOC, the area recently suggested as being the first involved in object recognition in humans (Kanwisher 2003). Details of the visual path are in (Plebe et al. 2007).

The hardwired extracortical MGN component is just a placeholder for the spectrogram representation of the sound pressure waves, which have been extracted with tools of the Festival software (Black et al. 1997). It is justified by evidence of the spectrotemporal processes performed by the cochlear-thalamic circuits. The thalamic afferents are collected by a LISSOM module, acting as the auditory primary cortex. The next map in the auditory path of the model is STS, because the superior temporal sulcus is believed to be the main brain area responsive to vocal sounds. The two higher cortical maps in the visual and the auditory paths, LOC and STS, carry the best representation coded by the models on object visual features and word features. These two representations are associated in an abstract type map, called ACM (Abstract Categorical Map). This component is implemented using the SOM architecture described earlier, known to provide non-linear bi-dimensional ordering of input vectors by unsupervised mechanisms. It is the only component of the model that cannot be conceptually referred to as a precise cortical area. It is an abstraction of processes that actually involve several brain areas in a complex way, and as such departs computationally from realistic cortical architecture.

2.2 Experiments and Results

2.2.1 Phases and development

The training of the model takes place in separate stages utilizing different sets of training samples. The different cortical components that comprise the model, are exposed to a variety of stimuli sequentially. Though highly simplified, this process parallels some of the natural phases of a child's development. There are three main phases: prenatal, pre-linguistic and linguistic, as shown in Fig 2.

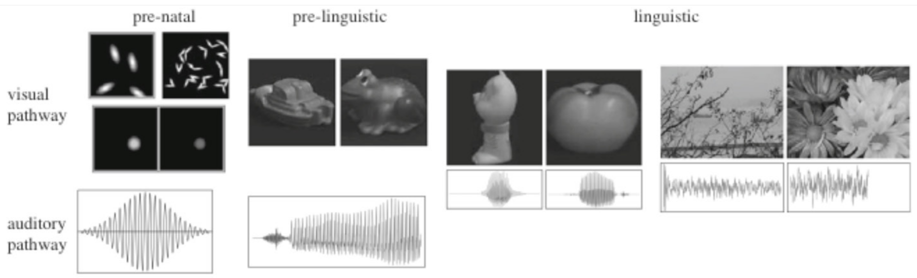


Fig. 2 Stimuli used in the experiments. In the pre-natal stage at the *top* elongated blobs input to V1, couple of blobs for V2, hue *circular blobs* for VO, wave trains for A1. In the pre-linguistic stage, real images for the visual path, and common word waves for the auditory path. In the linguistic phase, combination of view of real objects and waves of their category names, and random landscapes combined with music waves

During the pre-natal phase, initially only V1, VO and A1 maps are allowed to modify their synaptic weights, with the modification of V2 taking place shortly after, within the same phase. The stimuli presented to V1 and VO are synthetic random blobs that simulate pre-natal waves of spontaneous activity, known to play an essential role in the early development of the visual system, A1 is stimulated with simple random Gabor-shaped trains of sound waves. The visual stimuli to V2 comprises pairs of elongated blobs with a coinciding end point in order to enhance the experience of patterns that are slightly more complex than lines, such as corners.

The pre-linguistic phase, corresponding to that just after eye opening, involve the LOC and STS maps. For the visual pathway, natural images are now the primary stimuli. In order to include the primary and most realistic difficulty in recognition, which is the identification of an object under different views, the COIL-100 collection has been used (Nayar et al 1995) where for each of the 100 objects, 72 different views are available. Only 8 views per object were used during the learning phase of the model. All 72 views, however, were later used in the testing phases.

All the auditory maps are exposed to the 7,200 most common English words (from http://www.bckelk.uklinux. net/menu.html) with lengths between 3 and 10 characters. All words are converted from text to waves using the Festival Software, with cepstral order 64 and a unified time window of 2.3 s.

At the end of this first phase of development, each map in the model has evolved its own function. Orientation selectivity is the main organization in the primary visual cortex, where the responsiveness of neurons to oriented segments is arranged over repeated patterns of gradually changing orientations, broken by few discontinuities. This sort of arrangement emerges in the model's V1, as already demonstrated in (Plebe et al 2006) and (Sirosh et al 1997). Angle selectivity, or the responsiveness of neurons to stimuli containing angled lines, emerges in the model's V2 map, as is found in the biological correlate (details are in Plebe 2007). The model's VO map develops a simple form of color constancy. Color constancy is the tendency of the color of a surface to appear more constant than it is in reality. This property is helpful in object recognition, and develops sometimes between two and four months of age. One of the main functions shown by the LOC layer in the model is visual invariance, the property neurons have of responding to peculiar object features despite changes in the object's appearance due to different points of view. Invariance is an important characteristic for an object recognition area to have, and is found in the human LOC.

Tonotopic mapping is a known feature of the primary auditory cortex that represents the dimensions of frequency and time sequence in a sound pattern, and in the model it is achieved in A1. The spectrotemporal mapping obtained in STS is a population coding of features, in frequency and time domains, representative of the sound patterns heard during the development phase.

In the linguistic phase of development the ACM is involved as well, and the visual and auditory paths are exposed contemporaneously to stimuli occurring at the same time. Two kinds of events are simulated: the intentional ostensive naming of an object, and the casual association of sound patterns with natural scenes. For the first kind of event, the 100 objects of the COIL-100 collection are grouped manually into 38 categories, and the corresponding names converted to waves. Each name is replicated in four utterances, using the en1 "Roger" male voice, and the us1 female American speaker, both duplicated at standard and slower speeds, using the 1.3 value of the Duration Stretch parameter in the Festival software. For the meaningless coincidences of visual scenes and sounds, images from the McGill Flowers and Landscape collections (http://www.tabby.vision.mcgill.ca/) and music sounds (Wagner, Der Fliegende Hollander) are used.

In this phase, 500 clones of the model are individualized through the exposure to different sets of stimuli. From the 100 COIL objects, 500 different subsets are randomly extracted, grouped according to 5 different sizes. At the end of this phase, there will be 5 groups of models that can be considered as representing 5 different stages of de-

Table 2 The five stages of development, in terms of objects known by the model, and known words corresponding to object categories

Stage	# Objects	# Words
Stage I	30	20.6
Stage II	50	28.3
Stage III	60	31.1
Stage IV	70	33.5
Stage V	80	35.5

velopment, each with a progressive vocabulary of known words, composed of 100 exemplar models (see Table 2). In should be noted that each stage has a fixed number of known objects, but because categories of objects in the COIL collection have an uneven number of exemplars, and being that the set of stimuli is selected randomly, the number of known words in a single stage of development varies slightly between individual models. The figures shown in Table 2 are the averages.

All subsets used in the linguistic phase lack a category of objects used for the experiments as well as a small number of other objects used as exemplars in the triadic trials. The car category has been chosen because it is composed by a sufficient number of different samples (seven), with a variety of different shapes and colors.

A theoretical difficulty in differentiating levels of maturity within artificial networks is to be emphasized here. In general, there is a relationship between the performance of an artificial network in representing information, and its number of free parameters. Taking for the sake of clarity, only the final ACM map, it is well known that the number of different categories that can be ordered in a SOM map scales with the size of the map itself. For this reason, in an earlier preliminary version of this model (Plebe et al 2007), the size of the abstract SOM map was increased during the stages of development. This solution has two drawbacks: first it is not biologically realistic, as no similar growth process takes place in the cortex during development. Second, it introduces a subjective parameter that affects the results, due to the arbitrariness in the function relating SOM size with the knowledge level of the system. In this model, instead, the size of all the maps is fixed for all stages of development. This was possible, by keeping the overall size of the stimuli set constant during the linguistic phase, at all stages of development. At early stages, the model is exposed more often to random coincidences of scenes and sounds in the form of music, and less to purposeful object labeling, gradually the ratio is balanced, and in later stages the meaningful linguistic input predominates.

2.2.2 Fast Mapping-Trials

In the first phase of these experiments, each model is trained to learn an unknown target object, with its name. This object is #23, visible in Fig. 3, and is then presented to the model under 3 different views. The duration of the presentation of this stimulus is very short (40 training epochs only), corresponding to a typical new name training session in child experiments, that when the learning phase is very short and successful, corresponds to the fast mapping phenomena. Fast mapping is usually difficult to reproduce in artificial neural models, due to the fact that they typically suffer the drawback of "catastrophic inference": the deterioration of the knowledge acquired before the fast-mapping event. As will be discussed later, this does not occur in our model.

During the testing phase of the experiments, two objects, both new to the model, are presented. One object is a car, different from target #23 used in the training trial, the other is an unfamiliar object of a different category. Both objects are presented together with the verbal label "car". We assume that the whole procedure is reasonably similar to experimental protocols on fast mapping, in which a name for a new category is introduced in the training

Fig. 3 The standard used in the learning tests, object #23 of the COIL-100 collection, here reproduced in graylevel, seen by the model in its original red color

stage ("this is a DAX"), and then a small set of objects is presented in order to verify if, and how, the name has been mapped ("give me a DAX") (Regier 2005; Smith L.B. 2001).

3 Results and discussion

During preliminary screening, it was evident that for most of the COIL-100 objects, the task was easy, and the correct object was always chosen; therefore, the tests

Table 3 Results of all trials. Numbers are fraction of correct responses when presenting the object of the same category (column) and a object of different category (row). All trials are replicated for the 5 stages of development

			 #6	 #8	 #15	 #69	 #76	 #91
		stage I	0.25	0.23	0.58	0.25	0.96	0.69
		stage II	0.18	0.17	0.58	0.19	0.93	0.67
#38		stage III	0.14	0.14	0.63	0.18	0.94	0.71
		stage IV	0.26	0.24	0.76	0.27	0.96	0.81
		stage V	0.25	0.25	0.90	0.26	0.92	0.91
		stage I	0.78	0.87	1.00	0.87	1.00	1.00
		stage II	0.84	0.92	1.00	0.93	1.00	1.00
#46		stage III	0.80	0.91	1.00	0.91	1.00	1.00
		stage IV	0.92	0.95	1.00	0.95	1.00	1.00
		stage V	0.94	0.96	1.00	0.98	1.00	1.00
		stage I	0.60	0.63	0.88	0.64	0.98	0.90
		stage II	0.63	0.65	0.92	0.67	0.96	0.93
#47		stage III	0.63	0.70	0.93	0.69	0.97	0.96
		stage IV	0.80	0.83	0.96	0.81	1.00	0.97
		stage V	0.84	0.84	0.98	0.85	1.00	0.99
		stage I	0.35	0.39	0.79	0.35	0.96	0.87
		stage II	0.41	0.41	0.82	0.34	0.97	0.88
#100		stage III	0.40	0.45	0.85	0.37	0.98	0.90
		stage IV	0.54	0.55	0.91	0.41	1.00	0.94
		stage V	0.67	0.67	0.91	0.43	1.00	0.95

were performed on four non-car objects only that mostly confounded the models, compared with six exemplars of car, in all possible combinations.

Table 3 lists the detailed results of all trials, computed as the fraction of correct choice, over all 100 models in each stage of development.

The first thing to be noted is the good overall performance of the models, that is, their good capacity to generalize on the basis of little exposure to a single exemplar of the new category. The rapid acquisition shown by the models could be considered as cases of both "fast categorization" and of "fast mapping". In fact, the models did not have a previous category for cars; they rapidly formed the category through few exposures to the image and the word. Subsequently the word simply behaves as a feature that co-varies coherently with other perceptual features of the object, without requiring any specialized mechani-

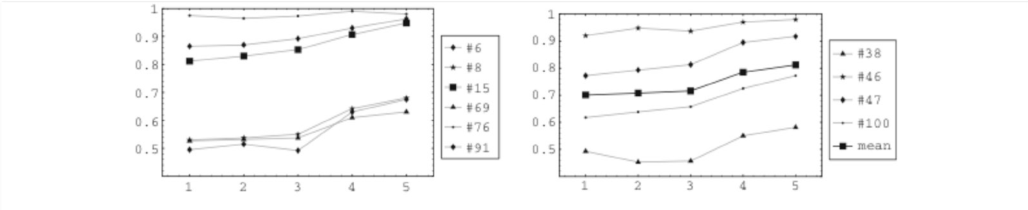


Fig. 5 Success in fast mapping at different stages of development. In the left plot curves of success for each object of the same category of the standard, in the *right* plot the curves for each object of category other than the standard, and the mean success curve

sm for word learning, this is in line with the hypotheses proposed in (Smith L.B. 2001) and the model proposed in (Regier 2005). Getting into the details of the single tests, all four objects that sometimes confused the model are, at least partially, red. However, color is not the discriminating feature, since there are 13 more red objects in the COIL-100 collection that were never chosen as the standard by the model. Object #38, the object that induced the most errors, is of a very different category (boat), but its contour is quite similar to the target. Object #100 is of a close category (truck), and of a similar color, but its shape is quite different. There is a structural similarity in the components (wheels), and the position of components with respect to the main body. All of the few confounding combinations improve in subsequent stages, except in the case of item #38 (the red boat), when compared with cars #6, #8, and #69. The confusion with object #38 improves in subsequent stages when compared with cars #15 and #91. The same improvement over subsequent stages is found for all the other non-car objects of the trials, including the truck #100. The larger difficulties in discriminating the newly learned car from item #38, suggest that at the outset the models base categorization on shape similarity. Sporadic choices of #100 (the red truck) during the first age groups may indicate an initial tendency to categorize by taking into account object components, as in (Biederman 1987).

Table 4 summarizes the accuracy of the models in categorizing all known objects and compares the performances before and after the trials with exposure to the standard object #23. Figures here, are mean and standard deviations over all 100 individual models for each stage. The accuracies are evaluated for object identification by visual appearance only, or by combination of vision and naming, and for categorization. This table confirms the good performance of the models that are able to learn objects and categories, at every stage of development. The small decrease in identification performance in subsequent stages is a normal consequence of the increase of the number of different objects to be identified. The size of the ACM SOM map allows a maximum coding of 100 different items and that number is reached during later stages of development. The main achievement here is the limited effect the new fast mapping has on the knowledge already acquired by the models. Noticeably, the ability of identifying objects when named, becomes significantly better than without naming after the fast mapping, and this effect is enhanced at later stages. It seems that as long as the language support improves, identification relies more and more on language itself.

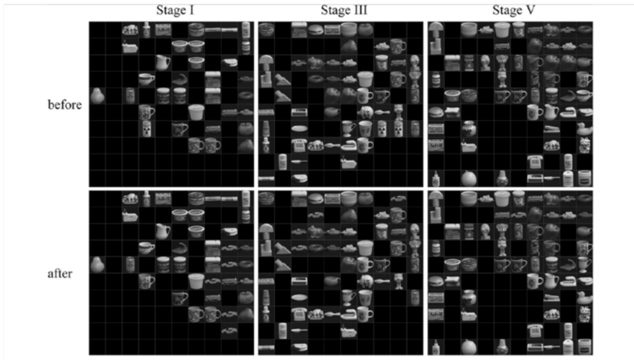


Fig. 4 Samples of the transformation in the Abstract Categorical Map by effect of the fast mapping, for 3 different stages of development. In each couple of maps, the *left* one is before fast mapping, and the *right* one is the same model modified after fast mapping

In Figure 4, we show examples of how the ACM categorial map is affected by the fast mapping. The pictorial representation of the maps is generally obtained by overlapping the image of the object that is mostly represented by a map neuron. In the first example, at stage I, the map appears sparse, since the model knows only 30 objects. After the learning phase of the trial, several neurons are recruited for representing the new object #23. Its representation is spread over 5 neurons, probably due to the variety of appearance under all viewpoints, compared to the few views experienced by the model. We should remember that in the fast-mapping trials only 3 views of the object are presented, while for the evaluation, the full set of 72 views are used. The placing of the object is reasonable, close to a red truck and the green boat that despite the color, as discussed earlier, shows a shape-bias effect. It can be seen that the other categorization in the map remains almost unchanged, with the only exception being the can below a cup, in the bottom-left quarter of the map. It is clearly a marginal side-effect of the rearrangement of the neural vector codes, since this object mapping is quite far from the car representing neurons in the map. The example at stage III, once again shows that neurons have been adapted for the new car category, without forgetting previously learned categories. The positioning, privileges again, closeness to boats and trucks, in this case, one green (surrounded by two #23 cars) and one red (top right).

An interesting aspect that emerges from the experiments is that the ability to learn a new object by its label seems to improve with the quantity of known objects and labels. This trend can be appreciated in the graphs of Fig. 5, where success in making the correct choice is plotted against the stages of development. With respect to the full matrix of data given in Table 3, in the plot on the left, data have been averaged by columns (non-car objects), while in the right plot, data have been averaged by rows (car objects). A statistical analysis has been performed with two-ways ANOVA, grouping the same data according to stage of development, and either by non-car objects or by car objects. As reported in Table 5, there is a clear effect of the stages of development.

The direction of the trend has been further tested by positive linear correlation, results are reported in Table 6. In this analysis, the independent variable is treated ordinally, using the number of known objects, or known words. For both cases, data are grouped as before, either by non-car objects, or car objects. In all combinations, the correlation is strong and confirms the improvement of fast mapping abilities with the simulated linguistic development of the models.

Table 5 Two-ways ANOVA analysis of the significance of stage levels on the success of the trials

Groups	$f(\cdot,\cdot)$	p
Car objects	$f(4,20) = 20.0$	<0.001
Non-car objects	$f(4,12) = 16.4$	<0.001

4 Conclusions

We believe the model discussed here could be a useful tool in the study of language acquisition, both for the general principles applied and for the specific results obtained.

In general, the experiments described exemplify how a simulation model can be used in order to assess developmental issues within a biologically constrained framework. What distinguishes our model is the effort to replicate with a high level of fidelity the contribution of some cortical areas to the emergence of early words. Surely, our model omits a number of factors, which are certainly involved in word learning. Within the limits of the factors we focused on, however, the model reaches a considerable degree of biological realism. In particular, the model grows from a detailed, biologically well-founded simulation of the ventral pathway of the visual cortex. The auditory pathway, has been built, although in a more speculative way, in accordance with what is currently known about it, based on recent data coming from the neurobiological literature. Similarly, the training phases of the model have been designed so as to preserve a high level of accordance with developmental data. In this sense, our model is quite different from others that previously attempted to simulate some of the typical phenomena of early word learning. The emergence of fast mapping and fast categorization are embedded here, in a full developmental time course, showing in considerable detail how word learning evolves through stages of individual experience. We believe that this concept can be fruitful in investigating several aspects of early language acquisition. We have applied this approach in several other models, which explored the role of working memory in language evolution, for example (Plebe et al 2008), or which studied the acquisition of adjectives (Mazzone et al 2008).

As far as the results of the model here presented are concerned, we do not claim that they disprove any of the positions reviewed in the introduction that posit special non-perceptual mechanisms at the basis of language learning. On the contrary, we agree that the acquisition of full-blown language relies on a combination of several complex mechanisms, with some emerging from experience alone. Our results suggest, in fact, that certain phenomena often taken to support the idea of "special" mechanisms at work, can also be observed in the absence of anything other than perception. It is the case of the increase of the ability of learning new words, as long as the known vocabulary increases, or the bias for shape as cue for identifying named objects. In light of these results, we suspect that the limitations usually held against so-called associative learning might fade when the term is not used in its psychological sense, but instead used to refer to neural associative learning, at the synaptic level.

References

Bednar J.A. (2002) Learning to see: genetic and environmental influences on visual development. Ph.D. thesis, University of Texas at Austin, tech Report AI-TR-02-294.

Biederman I. (1987) Recognition-by-components: a theory of human image understanding. *Psychol Rev.*, 94, pp. 115–47.

Black A.W., Taylor P.A. (1997) The festival speech synthesis system: System documentation. Tech Rep HCRC/TR-83, Human Communication Research Centre, University of Edinburgh, Edinburgh

Booth A.E., Waxman S.R. (2002) Word learning is smart: evidence that conceptual information affects preschoolers' extension of novel words. *Cognition*, 84, pp. B11–B22.

Carey S. (1978) *The child as word learner*. In Halle M, Bresnan J, Miller G. (eds), "Linguistic theory and psychological reality", pp. 264–93, Cambridge, MIT Press.

Carey S., Spelke E. (1996) Science and core knowledge, J *Philos Sci*, 63, pp. 515–33.

Dickinson D.K. (1984) First impressions: children's knowledge of words gained from a single exposure. *Appl Psycholinguist*, 5, pp. 359–73.

Diesendruck G., Bloom P. (2003) How specific is the shape bias? *Child Dev*., 74, pp.168–78.

Eimas P.D., Quinn P.C. (1994), Studies on the formation of perceptually based basic-level categories in young infants. *Child Dev*. 3, pp. 903–17.

Ganger J., Brent M.R. (2004). Reexamining the vocabulary spurt. *Dev Psychol*., 40, pp. 621–32.

Gershkoff-Stowe L., Smith L.B. (2004) Shape and the first hundred nouns. *Child Dev*., 75, pp.1098–114.

Gopnik A., Nazzi T. (2003) *Words, kinds, and causal powers: a theory theory perspective on early naming and categorization*. In Rakison D.H. (eds) "Early category and concept development", pp. 303–29, Oxford, Oxford University Press.

Hebb D.O. (1949) *The organization of behavior*. New York, Wiley.

Kanwisher N. (2003) The ventral visual object pathway in humans: evidence from fMRI. In

Chalupa L., Werner J. (eds.) *The visual neurosciences*. Cambridge, MIT Press.

Kemler Nelson D.G, Frankenfield A., Morris C., Blair E. (2000) Young children's use of functional information to categorize artifacts: three factors that matter. *Cognition*, 77, pp.133–68.

Kohonen T. (1995) *Self-organizing maps*. Berlin: Springer;.
Kripke S.A. (1972) Naming and necessity. In Davidson D, Harman G.H (eds), "Semantics of natural language", pp. 253–355. Dordrecht, Reidel Publishing Company.
Landau B., Smith L.B., Jones S. (1988) The importance of shape in early lexical learning. *Cogn Dev.* 3, pp 299–321.
Markson L., Bloom P. (1997) Evidence against a dedicated system for word learning in children. *Nature*. 385, pp. 813–15.
Mazzone M., De La Cruz V.M., Plebe A. (2008) Una simulazione biologicamente orientata dell'apprendimento di aggettivi: tra categorizzazione e concettualizzazione. *Sist. Intelligenti*, 20, pp. 293–305.
Michotte A. (1963;1946) La Perception de la causalite. Editions de *l'Institut supérieur de Philosophie*, Louvain, english translation by TR Miles, E Miles, Methuen, London.
Murphy G.L, Medin D.L. (1985) The role of theories in conceptual coherence. *Psychol Rev*, 92, pp. 289–316.
Nayar S., Murase H. (1995) Visual learning and recognition of 3-d object by appearance. *Int J Comput Vis.*, 14, pp. 5–24.
Nazzi T., Bertoncini J. (2003) Before and after the vocabulary spurt: two modes of word acquisition? *Dev Sci*. 6, pp. 136–42.
Plebe A. (2007) A model of angle selectivity development in visual area V2. *Neurocomputing*, 70, pp. 2060–6.
Plebe A., Domenella R.G. (2006) Early development of visual recognition. *BioSystems*, 86, pp. 63–74.
Plebe A., Domenella R.G. (2007) Object recognition by artificial cortical maps. *Neural Networks*, 20, pp. 763–80.
Plebe A., De La Cruz V.M., Mazzone M. (2007) *Synthetic learners of objects and names*. In Demiris Y., Scassellati B., Mareschal D. (eds.), "Proceedings of the 6th International Conference on Development and Learning", IEEE, pp. 300–5.
Plebe A., De La Cruz V.M., Mazzone M. (2008) *Modeling language emergence by way of working memory*. In Smith A.D.M., Smith K., Ferrer I., Cancho R. (eds.) "The Evolution of Language: Proceedings of the 7th International Conference (EVOLANG7)". Singapore, World Scientific Press, pp. 480–1.
Pruden S.M., Hirsh-Pasek K., Golinkoff R.M., Hennon E.A. (2006) The birth of words: ten-month-olds learn words through perceptual salience. *Child Dev*., 77, pp. 266–80.
Putnam H. (1975) *The meaning of "meaning"*. In Putnam H. (ed), "Mind, language and reality", vol 2. Cambridge, MIT Press.
Regier T. (2005) The emergence of words: attentional learning in form and meaning. *Cogn Sci*., 29, pp. 819–65.
Rogers T.T., McClelland J.L. (2006) Semantic Cognition—a parallel distributed processing approach. Cambridge, MIT Press.
Scholl B.J., Tremoulet P.D. (2000) Perceptual causality and animacy. *Trends Cogn Sci*., 4, pp. 299–309.
Searle J.R. (1983) *Intentionality: an essay in the philosophy of mind*. Cambridge, Cambridge University Press.
Sirosh J., Miikkulainen R. (1997) Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Comput*., 9, pp. 577–94.
Smith L.B. (2001) *How domain-general processes may create domain-specific biases*. In Bowerman M., Levinson S. (eds) "Language acquisition and conceptual development". Cambridge, Cambridge University Press.
Smith L.B. (2005) *Shape: A developmental product*. In Carlson L., VanderZee E. (eds) "Functional features in language and space". Oxford, Oxford University Press.
Tomasello M. (1999) *The cultural origins of human cognition*. Cambridge, Harvard University Press.
Tomasello M. (2003) *Constructing a language: a usage-based theory of language acquisition*. Cambridge, Harvard University Press.
von der Malsburg C. (1973) Self-organization of orientation sensitive cells in the striate cortex. *Kibernetic*, 14, pp. 85–100.
Willshaw DJ, von der Malsburg C. (1976) How patterned neural connections can be set up by self-organization. *Proceedings Royal Society of London*, B194, pp. 431–45.