Taxonomies and Part-Whole Hierarchies in the Acquisition of Word Meaning - A Connectionist Model

Georg Dorffner

Dept. of Medical Cybernetics and Artificial Intelligence, University of Vienna, and Austrian Research Institute for Artificial Intelligence[1]
Schottengasse 3, A-1010 Vienna, Austria
georg@ai.univie.ac.at

Abstract

The aim of this paper is to introduce a simple connectionist model for the acquisition of word meaning, and to demonstrate how this model can be enhanced based on empirical observations about language learning in children. The main sources are observations by Markman (1989, 1990) about constraints children place on word meaning, and Nelson (1988), as well as Benelli (1988), about the role of language in the acquisition of concept taxonomies. The model enhancements based on these observations, and those authors' conclusions, are mainly built on well-known neural mechanisms such as resonance, reset and recruitment, as first introduced in the adaptive resonance theory (ART) models by Grossberg (1976). This way the strength of connectionist models in plausibly modeling detailed aspects of natural language is underlined.

A Simple Model of Word Meaning

The connectionist model introduced in Dorffner (in press) is designed to demonstrate the abilities of a self-organizing system to acquire the meaning of simple words. Virtually no knowledge about the world is included a priori, mainly general pre-wired architecture. It concentrates on some important basic aspects while (necessarily) leaving out many details. The core ideas of the original model are the following.

- Words are primarily symbols in their referential sense (see, for instance, Dorffner 1992b). Learning the meaning of words in a first approach therefore means learning to identify symbols and their function.
- A basic task for every cognitive agent is to categorize environmental situations based on rich sensory stimuli and thus form concepts "about the world".
- Words as symbols in their simplest form refer to concepts. As a result, interpreting a symbol means building an internal link to one of the concepts in the above sense.
- Concepts are mental states clearly separable from other states. As such they are independent from each other, but can grow associative links so as to establish relations among each other. As a result, any conceptual structure such as hierarchies, usually attributed to conceptual schemata, is not reflected in the model architecture beforehand. It is viewed as being either "in the eye of the observer" of the model, or at best localizable after learning as associative traces between conceptual states.

In summary, meaning becomes defined with respect to the subjective experiences of the individual agent, leading to a constructivist core theory of word semantics. According to these ideas, the model consists of two sensory inputs, two components for concept formation based on categorization, and a set of layers for building the *referential links* (Fig. 1). For the sake of simplicity the two parts of the model (including the two different inputs) are kept separate—one being used for perceiving and clearly identifying the words (or so-called *external embodiments* of the symbols), the other for forming concepts about the perceived environment other than the elements of language. In the implementation, primitive acoustic input (stationary speech signals) was used for the former, simple visual input for the latter. In essence, both parts work the same way. Recognizing words is done by categorizing acoustic stimuli, the same way forming concepts is done by categorizing visual stimuli,

[1] *The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Science and Research*

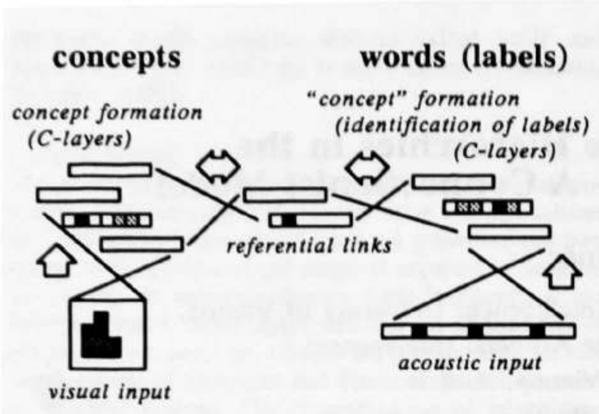

Fig. 1: The outline of the model

although for the latter internal activations can also have a major influence.

Concept Formation

The component for concept formation consists of a pool of so-called *C-layers*, which are connected to other C-layers and to the input via full associative feed-forward connections. Those connections are adapted with a "soft" competitive learning mechanism (Dorffner 1992a) that gradually compresses initially distributed patterns. Through learning it develops states we call *conceptual* or *identifiable states*. In the implementation they are basically defined as states where the winner of a C-layer is considerably larger than the average of the others. For this, the interactive activation rule (McClelland & Rumelhart, 1981) and negative weights between competing units, which are also adapted by the learning rule, are used.

It should be noted that concept formation based on, for instance, visual input would have to incorporate the great complexities of any natural visual system if it wanted to remain plausible in all respects as compared to humans. For the aspects this model is focused on such complexities would go far beyond the available computational resources. Therefore it is assumed that the patterns in the input are the result of preprocessing mechanisms, such as feature detectors or other transformations. Categorization as discovering invariants in stimuli classes starts after this preprocessing has happened. It is assumed that appropriate preprocessing can always transform real sensory stimuli into patterns showing similarity structures that can be processed by neural networks of the introduced kind.

Activation spreads to and within a pool of C-layers in two phases. First, any activation pattern originating outside the layer pool can activate patterns in the C-layers, plus the C-layers can cross-activate each other, all depending on the weight matrices shaped through learning. Secondly, competition sets in, both within layers and *between* layers. By introducing such an interconnected pool of C-layers the model is able to develop more than one conceptual state given one sensory stimulus. Examples would be concepts on different taxonomic levels. Furthermore, conceptual states develop either cleanly bottom-up (based on sensory input alone), or influenced by any other conceptual state to any degree. Competition in the second phase of a cycle ensures that always one (or at best, a few) conceptual state remains active. The underlying assumption is that a cognitive agent at one point in time can only concentrate on one conceptual perspective. For instance, a dog is either seen as a *poodle*, a *pet*, or an *animal*, but hardly as all three at the same time. It is, however, unclear as to which mechanism should select the most appropriate concept to win in competition—one drawback that has lead to the extensions described below.

An additonal part of concept formation based on sensory input is an automatic *focus of attention*—in this case a window whose size and location are adaptive. It can automatically be centered on any part of the original input in order to cut out (or enhance) this part. This window is then used as the input to the C-layers. By changing the size, the system can focus on parts of an object, building conceptual states based on those. Again, it is rather unclear when and how a shift of window position and size should be triggered.

Referential Links

For the implementation of referential links between identifiable states, a layer of specialized units—called *SY-layer*—with a winner-take-all (WTA) characteristics was introduced. This layer learns to identify co-occuring concepts and link them via a link unit. Learning of symbolic reference can be divided into the following phases. In the *fuzzy phase* no link unit has learned to respond to an identifiable state (clear concept). In the *identification phase* two identifiable states occur at the same time in different parts of the model. The winning link unit initiates WTA and the weights are strengthened. If this happens often enough, in the *recognition phase* one identifiable state is sufficient to let a link unit win and associate the corresponding concept. A special value of that unit—called the "symbol status"—must be above threshold to distinguish this case from the fuzzy phase. The links are built by a special "one-to-one rule" that lets weights grow for co-occuring identifiable states on both sides of the SY-layer. In other words, labels (words) are only linked to concepts that consistently occur at the same time, neither less nor more often.

The important properties of referential links implemented this way are as follows. First, not just any stimulus can be linked to a concept, it has to be clearly identified first. This means that referential links are built relatively late, i.e. after a considerable number of concepts have been learned. This corresponds to observations on early child language (e.g. Aitchison 1987), where consistent names for object classes are learned well after sounds are recognized and reproduced. Secondly, through WTA it is ensured that similarity structures on one side of the model do not map onto similarity structures on the other one. This corresponds to the arbitrariness of symbols by which their form is not related to their meaning (see, e.g., Lyons 1976 – ch.2, Dorffner 1992b, for more details).

As SY-layers are connected to C-layers they too can influence concept formation. For this, however, two modi operandi of referential links have to be distinguished. The modus described above presupposes equal treatment of the two model parts. Put differently, words (labels) are identified at the same time concepts are, whereafter the two get linked. This is plausible for the acquisition of the very first words, when neither the words, nor their potential referential power are known. Later, however, words are identified as referring to something even if no concept is activated at the same time. Words, in order to influence concept formation, should thus be permitted to grow referential links from the label side only ("directed link"). Therefore, the model can be switched to a mode where novel link units can be activated by identifiable states on the acoustic part. Examples for concepts influenced by language this way are superordinates such as 'furniture' or 'vehicle.' This, too, leads us to the model extensions described in the next section.

Problems with the Simple Approach

In this model, assigning word meaning becomes the problem of developing an appropriate referential link between the identifiable state corresponding to an external symbol embodiment (i.e. word or label) and the concept the symbol should refer to. Not surprisingly, this is not as trivial as described above. Words do not simply name categories of objects in a unique way. Among others, complexities arise as

- there are words on different taxonomic levels, such as 'poodle,' 'dog,' 'pet,' or 'animal.' This means that for one given object a large set of words would be applicable, depending on context and the intent of the reference.
- there are words for parts of objects. When the object is presented, its parts are too, which leads to the problem as to which of these the word should refer to.
- there are words for overall properties of an object, such as size or color. The same problems arise as with parts of an object.

The one-to-one learning rule briefly described above already captures some of the complexities in that it builds consistent links only for consistent pairs of co-occuring conceptual states. Words on a higher taxonomic level (say 'animal') than a given concept (say *dog*) will not develop such a link, as it co-occurs with other concepts as well. Words on a lower lever (say 'poodle'), on the other hand, co-occur with only a few instances of *dog*. However, such a mechanism is not sufficient. Complexities like the ones described cannot simply be explained by using mere statistical correlations. As a result, extensions to the simple approach need to be made. In the spirit of the introduced model such extensions should preferably not require too complex an architectural implementation, while the model's behavior should remain at least psychologically plausible.

Constraints Children Place on Word Meaning

One source for the model extensions we have chosen are proposals put forward by Markman (1989, 1990). Markman has observed that children must face very similar problems during the learning of word meanings as the ones identified above for the model. Time is too short and the possible combinations of concepts and words are too large to let statistical correlations decide alone upon how to construct the mappings between the two. She suggests that there must be some constraints children implicitly apply when learning words, and derives the following, empirically supported, assumptions.

(a) the **whole object assumption**:
When faced with one or more objects, each consisting of several parts, children obviously assume that "a novel label is likely to refer to the whole object and not to its parts, substance or other properties" (Markman 1990, p.59).

(b) the **taxonomic assumption**:
"This assumption states that labels refer to objects of the same kind rather than to objects that are thematically related" (Markman 1990, p.59). In other words, children obviously assume that the concept a label refers to is based on categories built on similarities and not on thematic relations. For instance, 'dog' refers to a class of objects that look and behave similarly, and not to a class of objects in a certain thematic context (e.g. objects being petted by the mother).

(c) the **mutual exclusivity assumption**:
This assumption states that usually only one label can be attached to each concept. Thus it "helps children override the whole object assumption, thereby enabling them to acquire terms other than object labels" (Markman 1990, p.66). In experiments it could be shown that children faced with an object—for which they do not have a word yet—and a label tend to take this label for the whole (assumption (a)). Children who know a word already tend to take the new word as standing for a salient part of the object (e.g. a receiver of a telephone – see Markman 1990), due to assumption (c).

These three assumptions together, among other principles, permit children to efficiently learn the meanings of words (or better, nouns). As it turns out, all three assumptions can be nicely transferred to the connectionist model for word meaning introduced earlier.

Assumption (a) can be introduced by assuming that the model starts with a large focus of attention.[2] In other words, each label is first attached to concepts corresponding to whole objects rather than parts. Assumption (b) has already implicitly been built into the model by letting concepts be based on similarities.

According to assumption (c), novel identifiable states corresponding to labels should not be linked to previously labeled concepts. This is reminiscent of principles realized in *adaptive resonance theory (ART)* models (Grossberg 1976). There, in a kind of competitive learning, the goal is to prevent overgeneralization of categories to patterns that just happen to activate the same winner "by accident." It is achieved by letting each category learn a prototype and comparing this prototype with the current pattern. If the two patterns are similar enough ("resonance") the category is maintained. If they are not ("mismatch") the winner is reset and a new unit gets a chance of becoming active.

The same mechanism can be used to implement the mutual exclusivity assumption. After training, a conceptual state can activate the corresponding link unit representing its label. Now consider the case that in such a situation a novel link is about to be built, i.e. a novel link unit is active together with the same concept. This situation can be detected by letting the concept activate its learned label (link unit) and comparing the two in a kind of resonance or mismatch. In fact, comparison simply comes down to checking whether the two link units are identical. If they are (resonance), learning can continue. If they are not (mismatch) then the concept is supressed (reset), and another concept can be linked, provided the same cycle now leads to resonance.

If no other conceptual state can be activated given the current situation, mismatch can further trigger a shift in focus. Remember that it was stated earlier that principles have to be found to automatically set the size and center of the focus of attention. The presence of a word, mismatch and the absence of another concpetual state can now be introduced as one mechanism to trigger a shift. It can either be a positional one onto another object, or one in size onto one of the salient parts of the object. Although this has only partially been implemented (mainly the positional shift), the technical realization of the shift itself appears straightforward. If the shift is onto one of the object's part, the exact same behavior can be realized in the model that was roughly observed with children—namely that the model first tends to link words with whole object concepts, and later, only if a label has already been attached, with parts of objects.

The Origins of Taxonomies

With the inclusion of the three assumptions it is still not clear how the model should deal with sub- and superordinate concepts and their labels. They are not a matter of focus. If concepts on several taxonomic levels were activated when

[2] In this discussion we assume that objects do not overlap, thus bypassing the problem of how to focus on one object in a scene.

presenting sensory input, it would still be a problem as to which link should be built, aside from what can be handled by the one-to-one rule (see above). We need to look at further observations.

It is generally acknowledged that in human categorization there exists something called a 'basic level' (Rosch 1978, et al.), where concepts are formed most naturally and easily. For sensory-based concepts this level can be identified as the one where categories are found mainly due to natural similarities. Objects are named on the basic level (e.g. 'dog' or 'cat') much more frequently than on any other taxonomic level (e.g. 'poodle' or 'animal') and children learn such words much earlier than others (see, e.g., Aitchison 1987). Further empirical studies by Nelson (1988) and Benelli (1988) suggest that for learning concepts that are not on the basic level, such as superordinates, language itself plays a major role. They even go as far as suggesting that "superordinate terms are defined *in the language* and not in the world" (Nelson, 1988, p.4, emphasis by author).

Nelson (1988) further distinguishes two cases. First, she introduces superordinates as so-called "slot filler" categories. With this she means categories of objects (which might be in different basic-level categories) that are seen as belonging together through a specific context, such as food one usually eats for breakfast. Secondly, Nelson describes real taxonomic categories which she claims are defined through language and can only be learned linguistically.

This observations can, at least partially, also be reflected in the introduced model. The conceptualization mechanism based on similarities very nicely shows basic-level behavior, in that some categories are learned naturally without further reinforcement or influence ("bottom-up"). For superordinate categories at least two cases can be distinguished.

(a) They are categories of patterns that have too little similarities to be naturally thrown into one class.
(b) They are categories of patterns whose "similarities" are mainly defined through context, that is through invariant activations in other model components.

Case (b) corresponds to the slot-filler case, case (a) to real taxonomic concepts. In both cases further influence is necessary in order to cause a conceptual state to develop. This further influence, according to Nelson and Benelli, is language, that is, the labeling of those categories by words. In the model this can be implemented by conceptual states that are induced by a referential link. In other words, novel link units that cannot be mapped to any existing concept are permitted to recruit uncommitted units in a C-layer. This too is similar to the ART-model where units can be recruited to stand for novel classes. This recruitment can be supported by existing similarities, however small they may be. A learning rule—which is a slight variation of the soft competitive learning described above—can adapt weights so as to grow strong links between the invariants that are there and the concept unit. Thus after some training in many cases the concept can be activated bottom-up without first presenting a label.

A Simulation Run

The results of a simple simulation run of the major model components should make clear the basic functions of the extended model. The eight patterns in Fig. 2, chosen rather arbitrarily, were used as visual input. Eight different acoustic signals (call them 'a' to 'h') were used as labels ("words"). The model started in modus one and began to categorize both types of inputs. For this all eight visual, as well as the eight acoustic patterns were presented, each around 50 times. The visual patterns were presented in random order. The acoustic patterns were presented so as to correspond to one of the visual patterns' labels. For instance, label 'a' was used together with visual pattern 1 or 2, label 'e' for patterns 1 through 4, and so on (among the possibilities the choice was again random).

On the visual side, three basic level categories were developed, grouping patterns 1&2, 3&4, and 5&6 (the fourth grouping did not happen to be discovered in this particular run). All eight acoustic patterns lead to different identifiable states. As the first four labels were used to name the four expected basic level concepts (1&2,

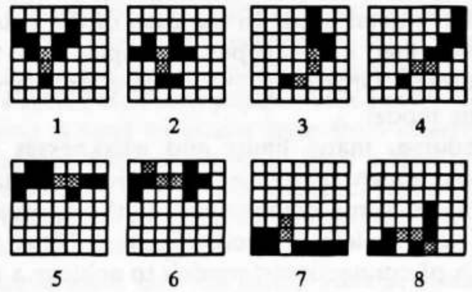

Fig. 2: Patterns used in the simulation

3&4, 5&6, and 7&8), three of them (the first three) were indeed learned perfectly. The one-to-one learning rule prevented any of the other labels to be associated with these concepts.

Then the model was switched to the modus of directed links. Now, whenever the new label 'e' was used to name the patterns 1 through 4 (superordinate concept), the novel link unit was compared with the perviously trained one, which lead to mismatch and reset of the conceptual state. No other concept could be activated. Shift of focus was not considered in this run. Therefore, after mismatch a new unit in one of the C-layers was recruited. Weight adaptation lead to the strengthening of the little similarities among patterns 1 through 4 (such as unit 3 in row 6), so that the new concept could also be activated bottom-up after some time. The model had thus learned a superordinate concept induced by naming. The previously learned basic level labels remained unchanged.

In another run, shift in focus of attention was considered, as well. Before recruiting a new unit, mismatch first triggered a shift. So labels for the most salient parts of each pattern (e.g. 'f' for the upper part in patterns 1 and 2) were learned, but only if a label for the whole object had been attached before. Otherwise the label was associated with the basic level concept, roughly mirroring Markman's observation on child language.

Conclusion

In this paper we have presented a psychologically plausible model for many aspects in the acquisition of word meaning. Many properties of the model can directly be compared to rough observations about early language learning in children; the model shows basic level conceptualization, builds links between labels and concepts after both have been identified clearly, and mirrors assumptions by Markman (1989, 1990) and Nelson (1988). It has been demonstrated that for achieving this mostly well-known neural mechanisms have been used. In the case of the extensions described in this paper, principles from the ART-model (Grossberg 1976) were transferred into this model.

Of course, many limits and weaknesses remain. Although many oversimplifications have been made, some complexities of the approach could nevertheless be demonstrated. Thus the strength of connectionist models to achieve a detailed model behavior has been underlined. Therefore the model is seen not only as a model with its psychological value, but also as a fruitful step toward alternative natural language understanding systems. Only the level of words—and there mainly nouns, and some adjectives—has been captured. It is believed that this is the primary level to be thoroughly understood if one wants to approach the phenomenon of natural language.

References

Aitchison J. 1987. *Words in the Mind*, an Introduction to the Mental Lexicon, Oxford: Basil Blackwell.

Benelli B. 1988. On the Linguistic Origin of Superordinate Categorization, *Hum.Dev.*, 31:20–27.

Dorffner G. 1992a. "Winner-take-more"—A mechanism for soft competitive learning, Technical Report 92–12, Austrian Research Institute for Artificial Intelligence.

Dorffner G. 1992b. On redefining symbols and reuniting connectionism with cognitively plausible symbol manipulation, Technical Report 92–13, Austrian Res.Inst.f.Artificial Intelligence.

Dorffner G. in press. A Step Toward Sub–Symbolic Language Models without Linguistic Representations, in Reilly R., Sharkey N. eds., *Connectionist Approaches to Natural Language Processing*, Hove: Lawrence Erlbaum, forthcoming.

Grossberg S. 1976. Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors, *Biological Cybernetics* 21:145–159.

Lyons J. 1977. *Semantics, Vol. I & II*, Cambridge: Cambridge University Press.

Markman E.M. 1990. Constraints Children Place on Word Meanings, *Cognitive Science* 14: 57–77.

Markman E.M. 1987: How children constrain the possible meanings of words, in Neisser U. ed., *Concepts and Conceptual Development*, Cambridge: Cambridge University Press.

McClelland J.L., Rumelhart D.E. 1981. An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings, *Psych.Review* 88: 375–407.

Nelson K. 1988. Where Do Taxonomic Categories Come from?, *Hum.Dev.* 31: 3–10.

Rosch E. 1978. Principles of Categorization, in E.Rosch, B.B.Lloyd eds., *Cognition and Categorization*, Hillsdale: Lawrence Erlbaum.