Learning Noun and Adjective Meanings: A Connectionist Account

Michael Gasser
gasser@cs.indiana.edu
Computer Science and Linguistics Departments
Indiana University
Bloomington, IN 47405, USA

Linda B. Smith
smith4@ucs.indiana.edu
Psychology Department
Indiana University
Bloomington, IN 47405, USA

Abstract

Why do children learn nouns such as *cup* faster than dimensional adjectives such as *big*? Most explanations of this well-known phenomenon rely on prior knowledge in the child of the noun-adjective distinction or on the logical priority of nouns as the arguments of predicates. In this paper we examine an alternative account, one which seeks to explain the relative ease of nouns over adjectives in terms of the response of the learner to various properties of the semantic categories to be learned and of the word learning task itself. We isolate four such properties: the relative size and the relative compactness of the regions in representational space associated with the categories, the presence or absence of lexical dimensions in the linguistic context of a word (*what color is it?* vs. *what is it?*), and the number of words of a particular type to be learned. In a set of five experiments, we trained a simple connectionist categorization device to label input objects, in particular linguistic contexts, as nouns or adjectives. We show that, for the network, the first three of the above properties favor the more rapid learning of nouns, while the fourth favors the more rapid learning of adjectives. Our experiments demonstrate that the advantage for nouns over adjectives does not require prior knowledge of the distinction between nouns and adjectives and suggest that this distinction may instead emerge as the child learns to associate the different properties of noun and adjective categories with the different morphosyntactic contexts which elicit them.

1 The Phenomenon

Children learn to correctly label objects with dimensional adjectives such as *red* and *big* later than they learn to label objects with nouns such as *block* and *dog* (Blewitt, 1983; Carey, 1982; Ehri, 1976; Macnamara, 1982; Rescorla, 1980). From the point of view of category learning, this might be considered surprising for two reasons. First, children are faced with a much larger set of nouns than adjectives to learn. Second, the nouns would, at least from an adult perspective, seem to be organized in a much more complex fashion than the adjectives. Nouns are generally characterized by a such a wide range of features that many have argued that they almost defy definition (see, e.g., Murphy & Medin, 1985; Rosch, 1973; Wittgenstein, 1953). Dimensional adjectives, in contrast, refer to a restricted range of values on a single perceptual dimension.

Nonetheless, children learn the meanings of dimensional adjectives only slowly. The protracted course of learning dimensional adjectives has been well studied by students of child language. The evidence suggests that children have some ideas about dimensional word meanings before they understand the details of what it is that distinguishes the different words associated with a given dimension

(Blewitt, 1983; Carey, 1982; Smith & Sera, 1991). For example, in response to the question *what color is it?*, children may err by saying *green* for a red object, but they rarely err by saying *big* (Cruse, 1977).

Why are adjectives harder to learn than nouns? What does the fact that children learn which adjectives answer which questions before they learn to map specific adjectives to specific attributes mean for the difference between learning nouns and learning dimensional adjectives? We attempt to answer these question by asking what properties of the learning task lead to faster learning of "noun" vs. "adjective" categories in a simple connectionist network.

2 Accounts of the Phenomenon

One kind of possible explanations for the relative ease of learning nouns is based on the nature of the meanings of nouns and adjectives: nouns and adjectives are learned at different rates because of differences in the inherent difficulty of the their meanings. One account, proposed in somewhat different forms by Gentner (1978) and Maratsos (1988) (see also Markman 1989), goes as follows: All languages make a distinction between arguments, or objects, conveyed by nouns, and predicates, or relations, conveyed by verbs and adjectives. This distinction must be fundamental to the way people view the world. But nouns are in a sense prior to verbs and adjectives because while predicates presuppose arguments, the reverse is not true. Thus children learn nouns before they learn verbs and adjectives because the meanings of nouns are logically more basic. It is easier to figure out what *dog* means from examples like *the dog is big* than it is to figure out what *big* means from similar examples because, to figure out *dog*, you don't need to have already understood *big*, whereas, to figure out *big*, you do need to have already understood *dog*. Note that, on this view, the difference between nouns on the one hand and verbs and adjectives on the other is one of kind.

However, we question whether a difference in kind, in logical simplicity, is sufficient to cause a difference in the acquisition of nouns and verbs. The argument that the child learns nouns easily because they are logically simple requires that the child know in some way that one is logically simpler than the other. But how, upon hearing a specific unknown word uttered in the context of an object, can the child know that that word is the logically simpler noun or the logically more complex adjective? Both adjectives and nouns can be predicated of objects—we can say *it's big* or *it's a dog*. Given no other information, why should the child's first hypotheses in both cases not be the same? Consider input utterances in which an adjective or noun is the only content word: *it's snergelly, this is the lorax*. The very young child, if she can make sense out of these utterances at all, may view them as assertions, that is, as predicating some set of properties to a referent, or as references, that is, as pointing acts. But she would not have strong reasons to believe that *lorax* is a label for an object and *snergelly* is a label for a predicate or attribute. Given the evidence, a reasonable hypothesis might be that *lorax* is a label for objects with LORAX properties and *snergelly* a label for objects with SNERGELLY properties. In many respects, this hypothesis would not be far from wrong.[1] But notice that this form of hypothesis accords no special status to nouns over adjectives.

Although the input utterances to the child might commonly be of the sort *it's big* and *it's a dog*, the child will also hear utterances containing more than one content word. Some of these will include noun phrases containing an adjective and a noun: *a green thneed, the snergelly box*. But how is the child to know that adjectives such as *green* and *snergelly* in phrases like these have different functions than nouns such as *box* and *thneed*? A reasonable hypothesis for the initial learner to form would be that

[1] Note that this argument would not hold for two-place predicates; the presence of two arguments with *hit* would be strong evidence that *hit* is not simply a label for objects that have HIT properties. In fairness, it should also be noted that the argument for the priority of arguments over predicates has usually been made with respect to verbs, rather than adjectives.

the meaning of *the snergelly box* is an object to which both SNERGELLY and BOX properties are being attributed. Again such a hypothesis would not be terribly wrong, but again this hypothesis accords no differential status to nouns and adjectives.

If the child does not treat nouns and adjectives differently at first, where does this difference come from? What is the origin of the early advantage of nouns over adjectives? Eventually the child does sort out the different functions of nouns and adjectives. One kind of input that probably matters in the eventual distinction of nouns and adjectives is a third type of utterance, one such as *the lorax is snergelly*. If it is clear from the context that a set of properties is being associated with a given object, then the child might infer from such examples that it is *snergelly* that is doing the predicating and *lorax* that is doing the referring. The child may also be guided by morphosyntactic cues marking the distinction between adjectives and nouns, though these cues are minimal in some languages. Although there is some evidence that very young children can use morphosyntactic differences as small as these to differentiate intended meaning (Macnamara, 1982), there is more evidence to suggest that the use of such cues emerges slowly and errorfully (Soja, Carey, & Spelke, 1991; Smith, Jones, & Landau, 1992). Morphosyntax certainly does not provide foolproof cues at the start of learning.

To summarize, many of the input utterances containing adjectives or nouns are simply not helpful to the child in distinguishing adjectives and nouns as logically distinct kinds of terms. Even if the child were provided with the innate predisposition to expect some words to denote predicates and others arguments, the task of determining which words belonged to which category would not be a trivial one.

Accordingly, we propose a different kind of account as to why adjectives are harder to learn than nouns. We argue that a prior understanding of the logical differences between nouns and adjectives is not necessary to account for the difference in their learning rate. Instead, we propose that the noun advantage emerges even though the child's learning can be described as progressing through identically formed hypotheses for nouns and adjectives—hypotheses of the form: "*big* refers to objects with BIG properties" and "*dog* refers to objects with DOG properties". Our account is based on the nature of the categories represented by nouns and dimensional adjectives initially learned by children, but it does not make a qualitative distinction between these categories. The nouns learned first by children generally refer to concrete objects at the basic level: *dog*, *cup*, *chair*. The dimensional adjectives generally learned first are terms such as *big*, *little*, *red*, *green*, *dark*, and *light* that refer to perceptual properties. This fact of language means that noun categories and adjective categories differ markedly in their size, overlap, and number of relevant perceptual properties. We propose that it is these kinds of differences that make nouns easier to learn than adjectives, and we show that this proposal has merit by demonstrating that the noun advantage could emerge through such differences alone.

We can illustrate our arguments by considering how words refer to regions in the multidimensional space defined by the perceptible properties of objects. Within the space of possible objects that might be labeled for the child, adjectives and nouns differ with respect to the proportion of this space for which they are appropriate labels. Dimensional adjectives such as *little* and *dark* are applicable to a very large proportion of the space since for these words most other sensory dimensions are completely irrelevant. The bounds on the regions in representational space associated with these words are defined by constraints on a very small number of dimensions. That is, many different kinds of objects can be dark, and many different kinds of objects can be little. Nouns such as *dog* and *box*, on the other hand, apply to a very small proportion of the representational space since there are constraints on the possible values for many different sensory dimensions for the referents of these words. Put another way, dogs are—in comparison to all the things that can be little—very much alike. Thus, nouns and adjectives tend to differ in what we call **representational span**, in the size of the region in internal similarity space that is occupied by the objects labeled by the words. The difference is related to a difference in **representational compactness**, the degree to which a category covers a tighter region in representational space because it is defined in terms of relatively many dimensions. This difference in representational span and compactness does not depend on or require a categorical distinction between the meanings of nouns and adjectives because it is only a tendency. While noun meanings usually involve many dimensions, one, such as shape, may predominate over the others (Landau, Smith, & Jones, 1988). And for some adjectives, more than one sensory dimension may be relevant, and the range of relevant properties may be modulated by other properties (Clark, 1991). Figure 1 illustrates the character of this general difference between adjective and noun meanings for a representational space consisting of only three dimensions. For the adjectives only one of the three dimensions is relevant; for the nouns all three are.

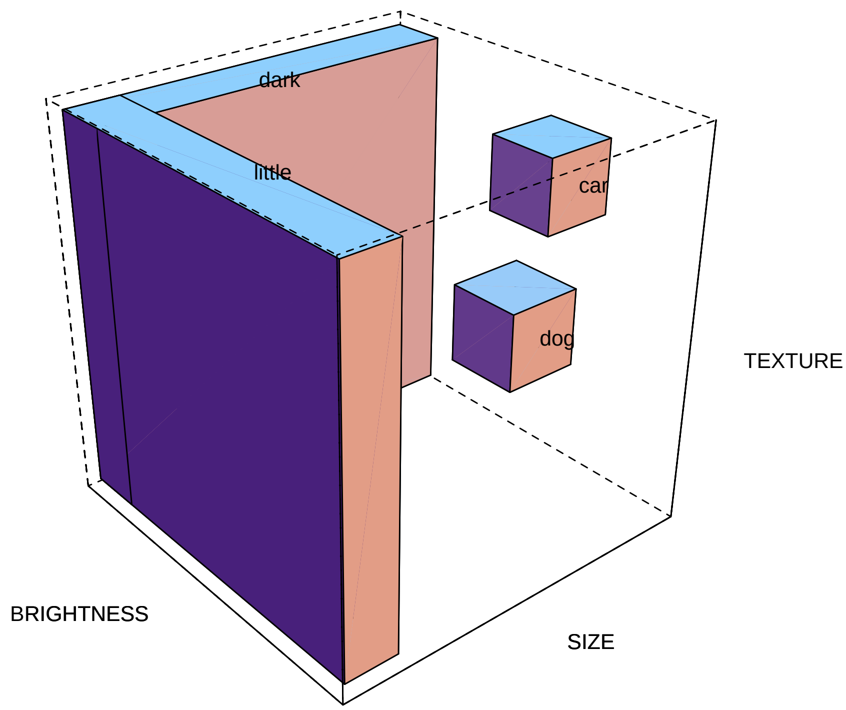

Figure 1: Adjectives and Nouns: Typical Span Difference

What might such differences in representational span and compactness have to do with ease of acquisition? We believe they may be one factor behind the advantage of nouns over adjectives. Given several specific instances of a nominal category, the bounds on the category are relatively clear because the bounded region is small. Given the same number of instances of objects labeled by a dimensional adjective, however, there would still be considerable uncertainty about what defines the category. This difference is illustrated in Figure 2. The white cubes represent known instances of a nominal category (*car* from Figure 1), and the black cubes represent known instances of an adjective category (*little* in Figure 1). For novel objects (for example, the shaded cubes in the figure), it is much easier to determine which nominal category they belong to than to determine which adjective category they belong to from these known instances. The specific purpose of the present study was to test the hypothesis that differences in representational span and compactness are sufficient for the emergence of a noun advantage in learning.

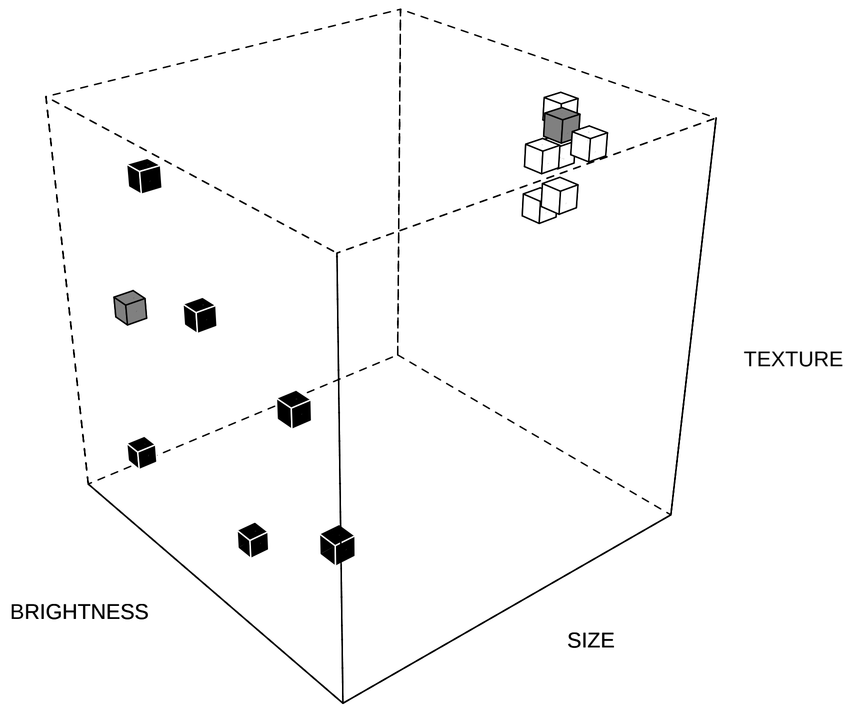

Figure 2: Learning Adjectives and Nouns

3 NIEC

We tested our hypothesis about representational span and compactness in the context of a connectionist model called the **Network for Implicit and Explicit Comparison** (NIEC) (Gasser & Smith, 1991; Smith, 1993). The NIEC architecture consists of two interconnected modules, one dedicated to categorization, which we refer to as **implicit comparison** and the other to **explicit comparison**. Explicit comparison involves two-argument predicates and is not addressed in this paper. Rather only the implicit comparison module of NIEC is of present concern. This consists of a very simple, general-purpose categorization device.

Like all connectionist models, this one consists of a network of interconnected processing units. The state of the network at any time is represented by the activations of the processing units. A unit's activation depends on the net input to it. Each (non-input) unit sums the inputs it receives along its input connections and computes an activation.

The connections joining units are weighted to reflect the degree of association between them. The connection weights are adjusted according to the learning rule and training regimen used by the network. NIEC makes use of **supervised** learning. That is, for each input, the system receives a **target** representing the correct response to the input. The learning rule, **backpropagation** in this case, compares the output of the network to the target and adjusts the network's connection weights in such a way that the network will come closer to the target the next time it receives a similar input (Rumelhart, Hinton, & Williams, 1986). The use of supervised learning is theoretically appropriate in the present case because children are explicitly taught words. Parents put objects before children, label them, ask children questions about objects, and correct their errors (Callanan, 1990; Mervis, 1987; Snow, 1977;

Wood, 1980).

Figure 3 shows the architecture of the categorization module of NIEC, the module that learns words. Each rectangle represents a layer of processing units and each thin arrow complete connectivity between two layers. The task of the network is to take visual objects and a linguistic context as inputs and to produce a noun or adjective as output. Inputs to the network are presented to two layers of processing units, one for the representation of the object itself and one for a linguistic context corresponding to a question the network is asked. Input objects consist of patterns of activation representing a visually presented object in terms of a set of sensory dimensions. For the simulations discussed in this paper, the inputs contain 4 dimensions. Each sensory dimension is represented by 12 units in the input layer of the network. Values along each dimension are coded in "thermometer" fashion. That is, the units within each dimension group correspond to positions along a scale. Each unit has a maximum activation of 1, so two of the possible values along one 12-unit dimension were [1, 1, 1, .3, 0, 0, 0, 0, 0, 0, 0, 0] and [1, 1, 1, 1, 1, 1, 1, .8, 0, 0, 0].

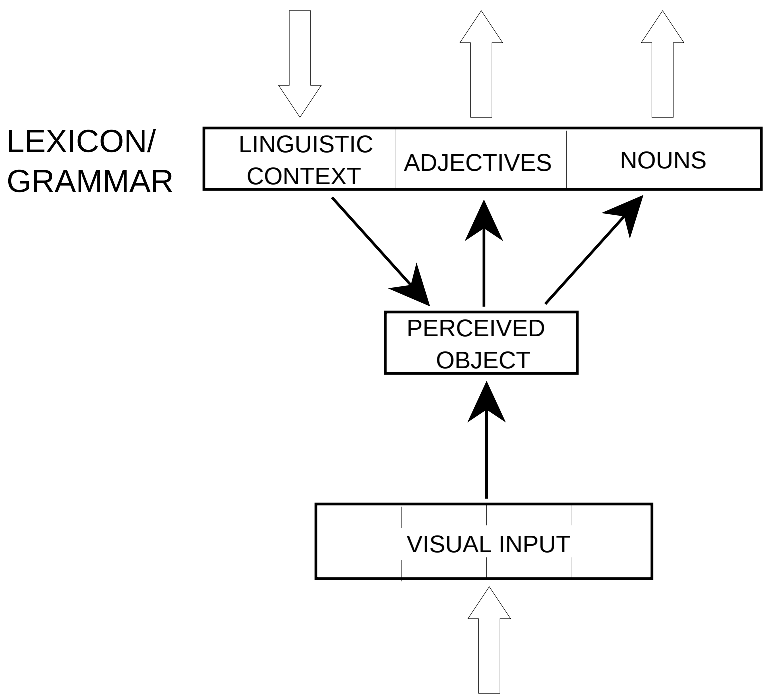

Figure 3: NIEC: Implicit Comparison Subnetwork

The linguistic context for each input consists of a question of the form *what size is it?* or *what is it?* A question such as *what size is it?* calls for an adjective response. The number of such questions that are possible depends on the number of **lexical dimensions** that are represented in the network. The question *what is it?* calls for a noun response. The linguistic context layer has one unit for this question and for each of the lexical dimensions along which adjectives are defined. It is important to note that, because the network is given no syntactic context, the noun context is indistinguishable from the adjective contexts at the start of training. That is, if we taught the network about nouns and three sets of dimesional adjectives, for example, color, size, and texture, as far as the network would be concerned, there would be only four kinds of inputs, noun, color, size, and texture, not two kinds, one for nouns and one for adjectives, with the adjective category subdivided into three classes. Our question is whether the network can learn such a noun-adjective partition in the course of learning individual nouns and adjectives which differ in representational span and other properties.

In sum, an input to the network has two components, a pattern consisting of 48 numbers representing an object in terms of values on four sensory dimensions and a pattern representing a question defining the network's task.

The hidden layer of the network, which we call the PERCEIVED OBJECT layer, compresses the input patterns into a smaller set of units, 22 in the experiments we report here. Thus at this level, the system no longer has direct access to the sensory input dimensions.

At the output layer, there is a single unit for each adjective and noun. A +1 activation on an output unit represents the network's labelling the input object with the corresponding word. A -1 activation represents the network's decision that the corresponding word is inappropriate for the input object, and a 0 activation represents an intermediate response, for example, one (e.g., *red*) that might be made if an object is described by the category (it is a red dog) but that is not an appropriate answer to the linguistic input question *what is it?*. During training, there is a target associated with each input pattern, representing an appropriate response to the input.

In ordinary backpropagation, each output unit receives a target on each trial. But this is an implausible procedure, as it means that all possible responses which are not appropriate are punished. Accordingly, we use an alternate training procedure. We provide targets only for the correct response and for all incorrect responses for which the network's output units are activated above a fixed threshold value. For example, say the network is presented a large, red object and asked "what color is it?". If none of its output units responds with an activation greater than the response threshold, then only the unit corresponding to the correct answer (*red*) receives a target, namely, a value of +1. There are no targets for the other output units, hence no adjustments are made to the weights on the connections into those units. If the correct (*red*) output unit is activated above the threshold, the procedure is the same. If, however, one of more incorrect output units (for example, those for *large* or *blue*) is activated above the response threshold, then these also receive a target. The target for the *large* unit is 0 (since the object is in fact large), while the target for the *blue* unit is -1 (since the object is not blue). So on such a trial the weights on the connections into these units are also altered, this time in such a way that, given a similar input later, they are less likely to be as highly activated.

In sum, the categorization network is a very simple device which is trained to take a pattern representing an object and one representing a question and to output an appropriate noun or adjective. The network has no knowledge pre-wired into it about the ways in which adjectives and nouns differ. Whatever it comes to know about nouns and adjectives it will derive from the inputs that it sees during training.

4 Experiments

The nouns that children learn first are basic-level nouns that refer to categories of objects that are highly similar across many dimensions. One class of adjectives that children learn early are dimensional terms that label all varieties of objects so long as they possess the critical property. In Experiment 1 we ask whether a network that has no prior knowledge of the differences of nouns and adjectives learns the nouns more rapidly than the adjectives. We specifically ask the network to learn to answer questions about a visually presented object: *what is it?*, *what size is it?*, *what color is it?*. Thus Experiment 1 instantiates the differences that exist between nouns and adjectives and asks whether this is enough for the learning rate differences to emerge.

The subsequent experiments tease apart potential components of these task differences. In Experiment 1 (and in the world), noun categories are small because they are organized by values on many dimensions and adjective categories are large because they are organized by values on just one dimension. If these differences matter for training rate, then we need to ask next whether it is representational span (size) that matters or number of relevant dimensions. We address this question in Experiment 2 by asking whether small "nouns" organized by restricted variation on many dimensions are learned faster than large "adjectives" organized by less restricted variation on many dimensions.

In Experiments 1 and 2 (and the world), the task of learning adjectives is also more complex than that of learning nouns in that adjectives require learning of *lexical* dimensions. When an object is green and the question is *what size is it?*, the network must learn that *green* is not an appropriate response. For nouns, there are no lexical dimensions to learn. Put another way, nouns are one kind of category, but adjectives are of multiple kinds, and the multiple kinds, together with dimension words such as *size*, *color*, and *shape* designate what must be learned. We investigate the role that having to learn the lexical dimensions plays in Experiment 3.

In Experiment 1 (and the world) many sensory dimensions are relevant to noun categories whereas just one (or few) is relevant for dimensional adjectives. In the present model, separate sensory dimensions are compressed at the PERCEIVED OBJECT layer so that the network has to learn to selectively attend in order to respond correctly to adjective contexts. This aspect of the network specifically models young children's well-documented difficulty in attending selectively to individual dimensions (e.g., Gibson, 1969; Smith, 1989). In Experiment 4, we ask whether learning to selectively attend rather than differences in representational span might be the key factor in the noun advantage.

Finally, in Experiment 5, we investigate the role of the number of noun and adjective categories. These do not vary in Experiment 1, but in the world children are exposed to more nouns than adjectives and this could be a factor in the nominal advantage.

In all the experiments, the network was trained simultaneously to learn the various categories of words. We were interested in the number of training instances required for the network to learn the different categories and in the character of errors made during learning. Specifically, are categories resembling nouns learned faster than categories resembling adjectives, and are errors made within rather than between the syntactic categories and dimensions?

4.1 Experiment 1: Nouns and Adjectives

This experiment instantiates what we believe may be the central task differences in learning nouns and adjectives in the world. The experiment asks whether small categories organized by values on many dimensions are learned faster than categories organized by a small range of values on one dimension. Since our concern was whether this difference was sufficient for a noun-adjective distinction to emerge, we did not include other differences between nouns and adjectives that might matter to children's learning—namely, the greater frequency of nouns relative to adjectives and (possibly) the training of nouns prior to adjectives. The network's task and training in this and the subsequent experiments is modelled after one real-world context in which children learn words: an object is visually presented (the visual input), the parent asks a question such as "what is it?" or "what size is it?" (the linguistic input), the child responds saying the word "dog" (the output), and the child is told the correct response ("that's not a dog; it's a cat").

4.1.1 Stimuli

The input to the network consisted of visual input plus linguistic context patterns. The visual inputs were all instances of the 36 categories that the network was trained on. Eighteen of these were nouns, and 18 adjectives. The 18 adjectives were organized into 3 lexical dimensions, each of which had 6 associated adjectives. For each training instance, the inputs were generated as follows. First, an output category was selected at random from the set of 36. Each category was defined in terms of a range of values along each input sensory dimension, and for each of these dimensions a possible value was picked at random for the selected category. This yielded a complete instance of the category. For example, the adjective *big* is defined as shown in Figure 4. In the visual input, black circles depict units which must be completely activated ($output = 1.0$) for an input to be an instance of a category, white circles depict units which cannot be activated for that category ($output = 0.0$), and circles with fuzzy boundaries indicate the range of variation possible within this category. Thus for *big*, there is variation possible for the outputs of two of the units within the first input dimension and for the outputs of all of the units within the other three input dimensions. A possible randomly selected instance of *big* is also shown in the figure. Again black circles show completely activated units and white circles units which are not activated at all. Fuzzy circles depict partially activated units ($0.0 < output < 1.0$). The noun *lorax* is defined as shown at the bottom of the figure.

The linguistic context input, also shown in the figure, consisted of the pattern representing a question that would be appropriate for the selected category, each question corresponding to a lexical dimension. For example, if the category was *big*, the input unit representing *what size it is?* was turned on (that is, its output was set to 1.0), and the other linguistic context units were turned off. If the category was *lorax*, the input unit representing *what is it?* was turned on, and the other linguistic context units were turned off.

For this experiment, each adjective was defined in terms of a range of 1/6 of the possible values along one of the input sensory dimensions and any value along the other three. No adjectival categories were defined for one of the four input dimensions. Each noun was defined in terms of a range of 1/6 of the possible values along each of the four input sensory dimensions. Thus each noun spanned $1/6 \times 1/6 \times 1/6 \times 1/6 = 0.00077$ of the representational space whereas each adjective spanned $1/6$ of the space.

4.1.2 Method

On each training trial, the network was presented with an input (visual plus linguistic context), generated as just described, and an appropriate target on the output (see Figure 4). The weights in the network were then adjusted according to the backpropagation algorithm. Following each presentation of 1000 input patterns the network was tested on 500 inputs generated in the same fashion as the training patterns. For each test input, it was determined whether the output unit with the highest activation was the appropriate word. Performance for each category of word was measured as the proportion of test trials for which this was true.

4.1.3 Results

Figure 5 shows the learning rates for adjectives and nouns in this experiment. The data shown are averages over 10 runs with different initial random weights on the network's connections. The nouns, which span much less of the representational space than the adjectives, are learned much faster. Performance on the nouns is close to perfect by the 4000th training trial. Performance on the adjectives continues to improve, but never reaches the level of the nouns (though it does for larger hidden-layer sizes).

The model also exhibits the same tendency as do children for errors with adjectives to be within-dimension errors. Figure 6 shows, for 10 separate runs of the network, the proportion of those adjective

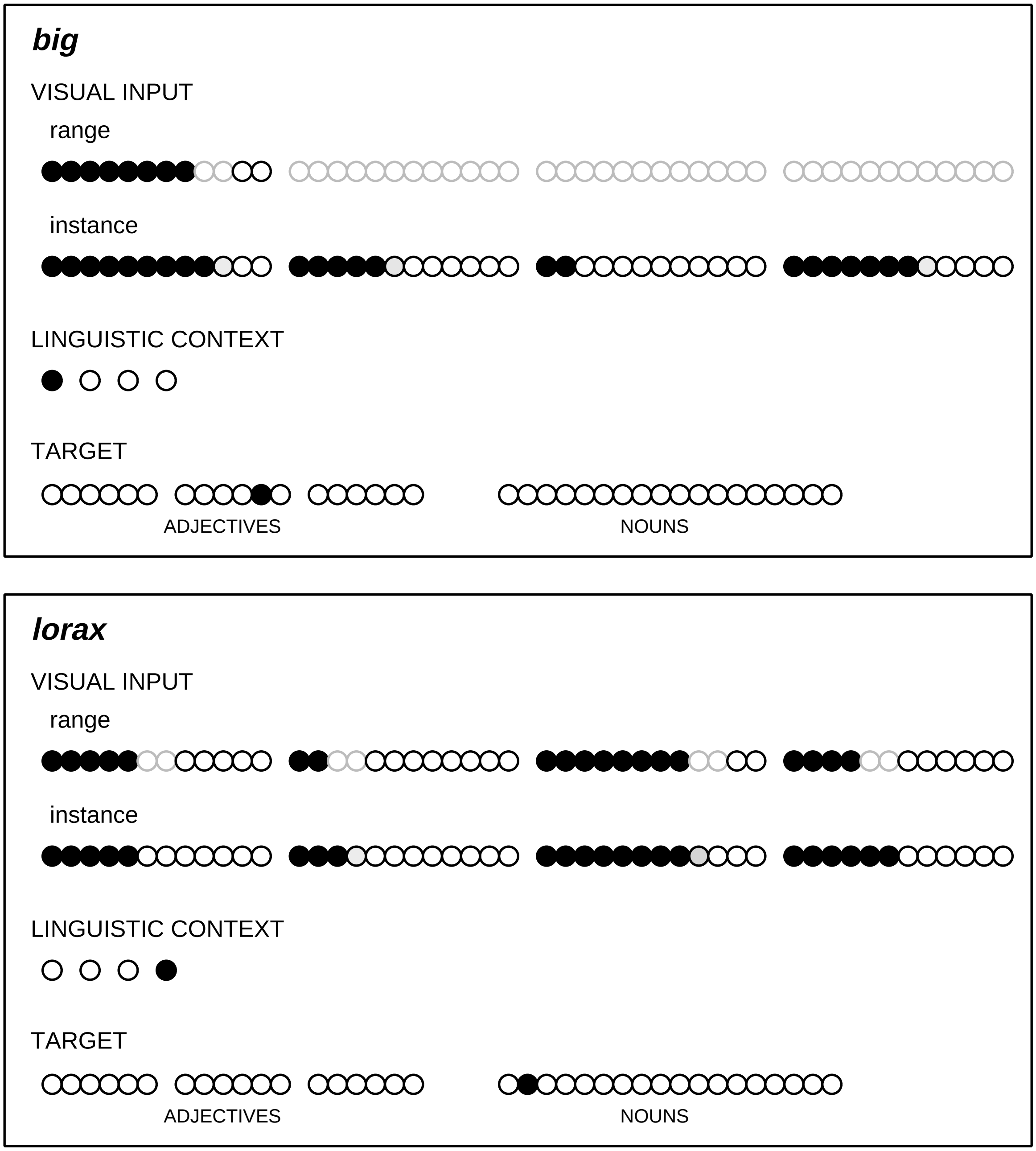

Figure 4: Example Input Patterns, Experiment 1

errors involving an adjective response for which the response was within the appropriate dimension. For example, given the linguistic context *what color is it?* and a *red* input, the probability that the network responds with *orange*, *yellow*, *green*, *blue*, or *purple*, is greater than the probability that it responds with *big*, *rough*, or some other non-color adjective. Noun responses to adjective contexts were not counted for the results shown in the figure. Relatively early in training, the proportion of within-dimension errors rises well above chance, where it remains. When noun responses are also taken into consideration, the results are similar: by the 4000th training pattern, the proportion of within-dimension errors has risen

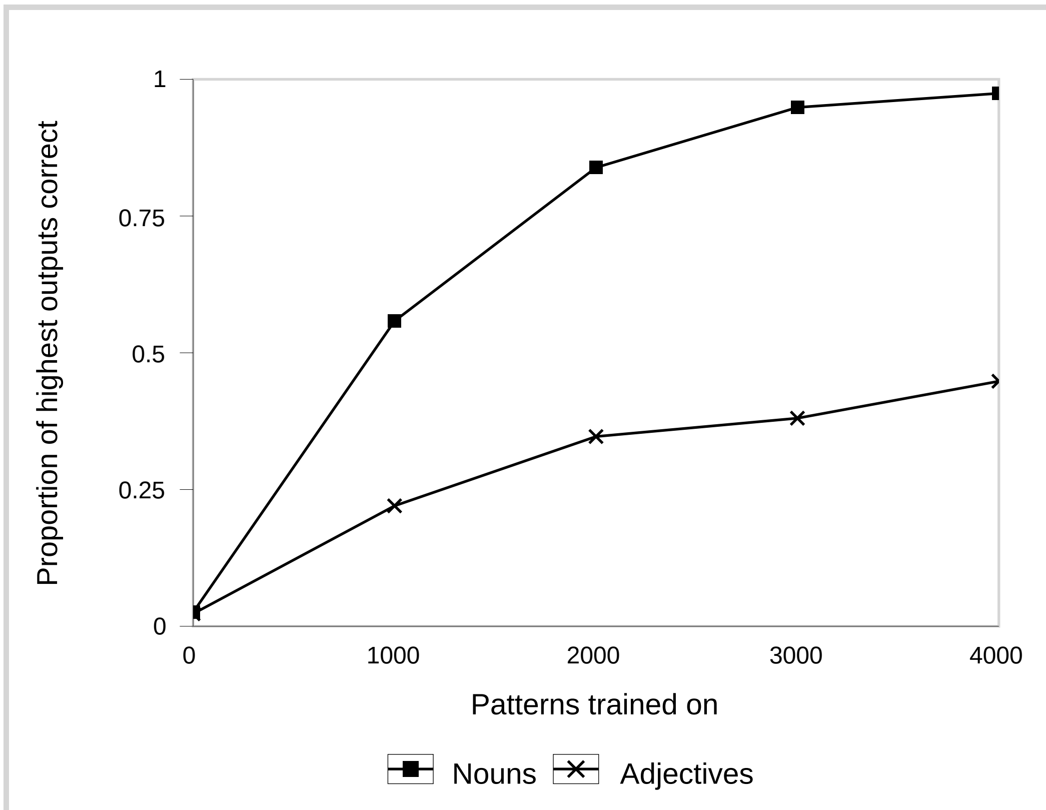

Figure 5: Learning Rates: Nouns and Adjectives (Experiment 1)

to 0.56 (chance: 0.14).

4.1.4 Discussion

In sum, the network, trained on adjectives and nouns presented with equal frequency, learns the nouns faster, as children do. Yet this difference is not due to any built-in preferences on the part of the network or to any pre-training representation of a difference between nouns and adjectives. It is due entirely to the nature of the categories which the network learns. The network also exhibits a structure pattern of errors on the adjective trials, making principally within-dimension errors. The network knows, in a sense, that *red*, *blue*, and *green* are words of the same dimensional kind before it knows which specific sensory inputs are red. This knowledge derives from early association of adjective outputs with the appropriate linguistic-context inputs. This experiment shows that differences in the learning rate of nouns and adjectives might emerge simply from the different category structures of nouns and adjectives.

In Experiment 1 we modelled noun categories that were very small because they were organized by values on many dimensions and adjective categories as big because there were organized by a narrow range of values on just one dimension. This description of nouns and adjectives fits the basic nouns and dimensional adjectives learned first by children. But the question arises whether the noun advantage demonstrated in Experiment 1 arises from representational span per se. Or is the nominal advantage related to the number of dimensions relevant to a category? That is, are small categories learned more

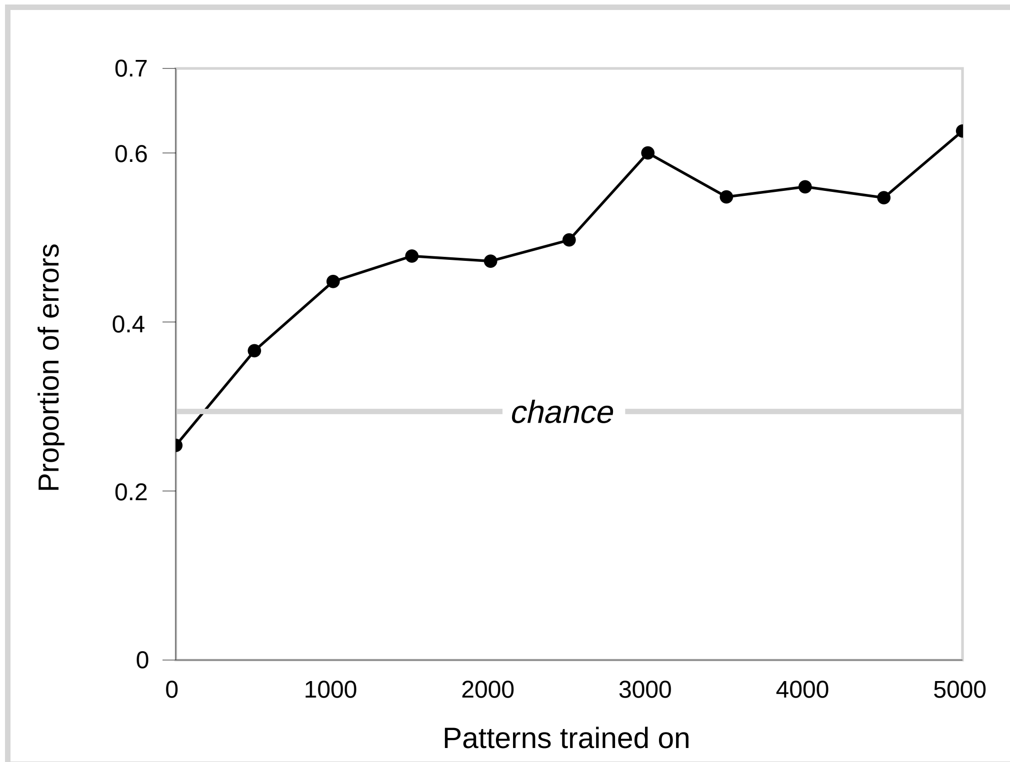

Figure 6: Proportion of Within-Dimension Adjective Errors (Experiment 1)

rapidly than large categories when both are organized by many dimensions? We answer this question in Experiment 2.

4.2 Experiment 2: Narrow and Wide Span

To test for the effect of representational span alone, we conducted a experiment in which the two types of categories differed only in representational span. Both "noun" and "adjective" categories were determined by ranges of variation on four sensory dimensions but the adjective categories allowed for a wider range of variation than the noun categories.

4.2.1 Stimuli and method

Stimuli for this experiment were generated exactly as in the last experiment. There were two types of categories, those which spanned relatively wide regions of the representational space and those which spanned relatively narrow regions. Each set contained 18 words. In the narrow set, each word was defined in terms of a range of 1/6 of the possible values along each sensory dimension. Thus each of these categories covered $1/6 \times 1/6 \times 1/6 \times 1/6 = 0.00077$ of the representational space. In the wide set, each word was defined in terms of a range of 1/3 of the possible values along each sensory dimension, a total of $1/3 \times 1/3 \times 1/3 \times 1/3 = 0.012$ of the representational space, that is, $1/16$ of the region occupied by each of the categories in the narrow set. Note that the spans of the two sets are closer than in the first experiment. There were no further dimensions distinguishing subsets of categories. For this experiment, there were only two possible linguistic contexts, one for which the wide-span words were appropriate responses, the other for which the narrow-span words were appropriate responses. Thus, assuming the narrow-span words were nouns and the wide-span words adjectives, the two linguistic contexts corresponded to the questions *what is it?* and *what's it like?*. For this experiment, we tested the network after every 500 training trials.

4.2.2 Results and discussion

Figure 7 shows the results of this experiment. Again the numbers shown are means over 10 separate runs of the network. As can be seen, the narrow-span words are again learned faster than the wide-span words. The difference is less than in Experiment 1 because the ratio of adjective-to-noun span is less: 16 to 1 in this experiment, 216 to 1 in Experiment 1. In this case the network eventually performs nearly perfectly on both sets.

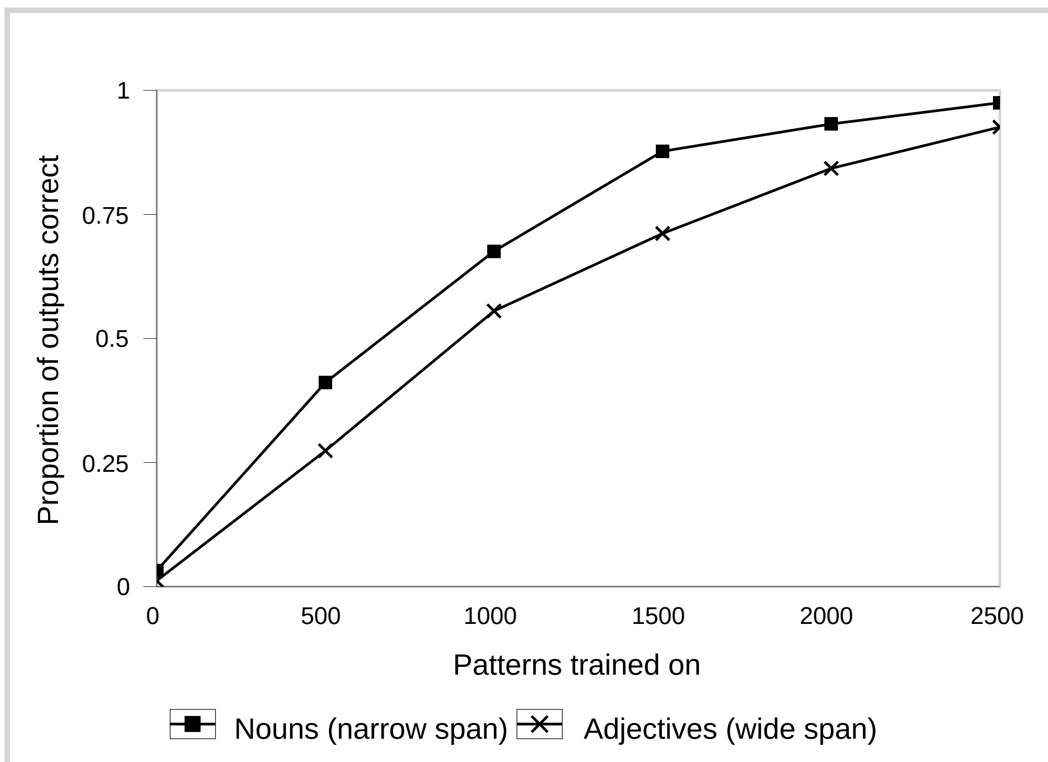

Figure 7: Learning Rates: Narrow and Wide Span (Experiment 2)

This experiment demonstrates that a difference in representational span by itself leads to a difference in learning rate. Words which are characterized by a relatively narrow span, those whose referents occupy a small portion of the representational space, are learned faster than those characterized by a wide span.

4.3 Experiment 3: Learning Lexical Dimensions

In Experiment 1, and in many of the labeling tasks faced by young children, adjectives are produced in response to questions about particular lexical dimensions such as *color* or *size*, whereas nouns are not broken down into subcategories in this way. This fact alone may be enough to make adjectives more difficult to learn. In Experiment 3, we examined the effects of the need to learn lexical dimensions for adjectives, but not for nouns. In this experiment, noun and adjective categories did not differ in span—each category, noun or adjective, was organized principally by variation along one sensory dimension.

4.3.1 Stimuli and method

As before, stimuli for this experiment were generated randomly, given the constraints which defined each of the categories. As in Experiment 1, adjectives were organized along lexical dimensions. In this case, there were four lexical dimensions, one each for the four input sensory dimensions. Again as in Experiment 1, the linguistic context input specified for adjective trials a question concerning one of the lexical dimensions for the input object (*what size is it?*), while the linguistic context for all noun trials was the same (*what is it?*).

Unlike in Experiment 1, however, the adjective and noun categories were identical in every other way; in fact, the same set of 16 categories was used for the 16 nouns as well as the 16 adjectives. For all categories a single sensory dimension was most relevant; that is, the range of variation possible along that dimension was considerably narrower than on the other three dimensions. For example, one of the "color" adjectives was defined in terms of a ranges spanning 2/3, 2/3, and 1/3 of the non-color sensory dimensions and 1/12 of the color dimension, and one of the noun categories was defined in exactly the same way. In this case, however, there was no overlap among the various adjectives or nouns (though there obviously was between adjectives and nouns since the categories were defined identically). All categories encompassed the same representational span, 1/81 of the total space, and 1/12 of the range along the most relevant dimension. Thus in this experiment, the only factor distinguishing "adjectives" from "nouns" was the necessity to learn the lexical dimensions and associate them with particular sensory dimensions in the case of adjectives.

4.3.2 Results

Figure 8 shows the results of this experiment. There is a modest advantage for the nouns. The need to learn lexical dimensions and associate them with individual sensory dimensions leads to only slightly worse performance. Thus learning lexical dimensions—subcategories of questions and responses in the case of dimensional adjectives—may play a small role in the nominal advantage.

4.4 Experiment 4: Selective Attention and Category Compactness

In addition to differences in representational span, nouns and adjectives tend to differ in terms of the number of sensory dimensions that enter into their definitions. Adjective categories may be consistent with the entire range of possible values along most dimensions and narrow ranges along a small number, or even just one, sensory dimension. Thus, for example, in learning to answer the question *what size is it?*, the child must learn to selectively attend—to emphasize information from one sensory dimension more than that from other sensory dimensions. Nominal categories, on the other hand, do not depend on single sensory dimensions, but are more "compact" in the sense that values along many dimensions matter equally. Such categories are compact in that the distance between all members of the category is minimal whereas for less compact adjective categories, the distance in the multi-dimensional space is great for at least some members. In Experiment 4, we investigated whether the number of relevant dimensions matters for the learning rate of categories. Are more compact categories more easily learned than less compact ones?

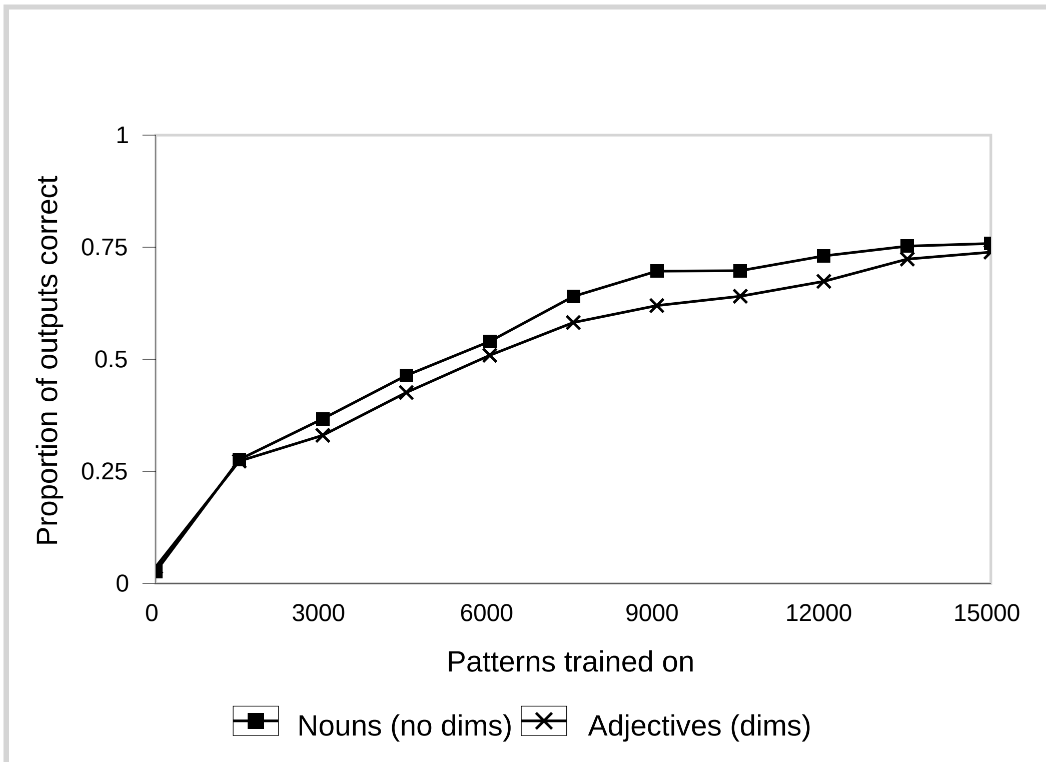

Figure 8: Learning Rates: Presence/Absence of Lexical Dimensions (Experiment 3)

4.4.1 Stimuli and method

The stimuli were again generated as before. For this experiment, as in Experiment 2, there were only two types of linguistic inputs, *what is it?* and *what's it like?*. Thus the network was not required to learn lexical dimensions. The 16 "adjective" categories were identical to the categories used in Experiment 3. They were defined in terms of ranges of 2/3, 2/3, 1/3, and 1/12 of the possible values along the four input dimensions. That is, one dimension, the one for which the range was 1/12 of the whole, was much more relevant than the other three in defining the category. Each of the four dimensions played this role for four of the adjectives. Each of the "nouns", on the other hand, was defined in terms of a range of 1/3 of the possible values along each input dimension. Thus one sensory dimension was more important for determining adjective than noun categories. However, the nouns and adjectives were identical in every other way: (1) they encompassed the same span (1/81 of the space); (2) their linguistic contexts were equally complex, neither requiring the learning of lexical dimensions; (3) there were the same number of adjectives and nouns to learn.

4.4.2 Results

Figure 9 shows the results of Experiment 4. The categories that were organized by an equally restricted range of variation on all four sensory dimensions were learned more rapidly than categories in which the range of variation on some dimensions was wide and on others narrow. Evenly compact categories are more rapidly learned than elongated ones. Again, this is a difference—like span—which favors the kinds of nouns children learn early over adjectives.

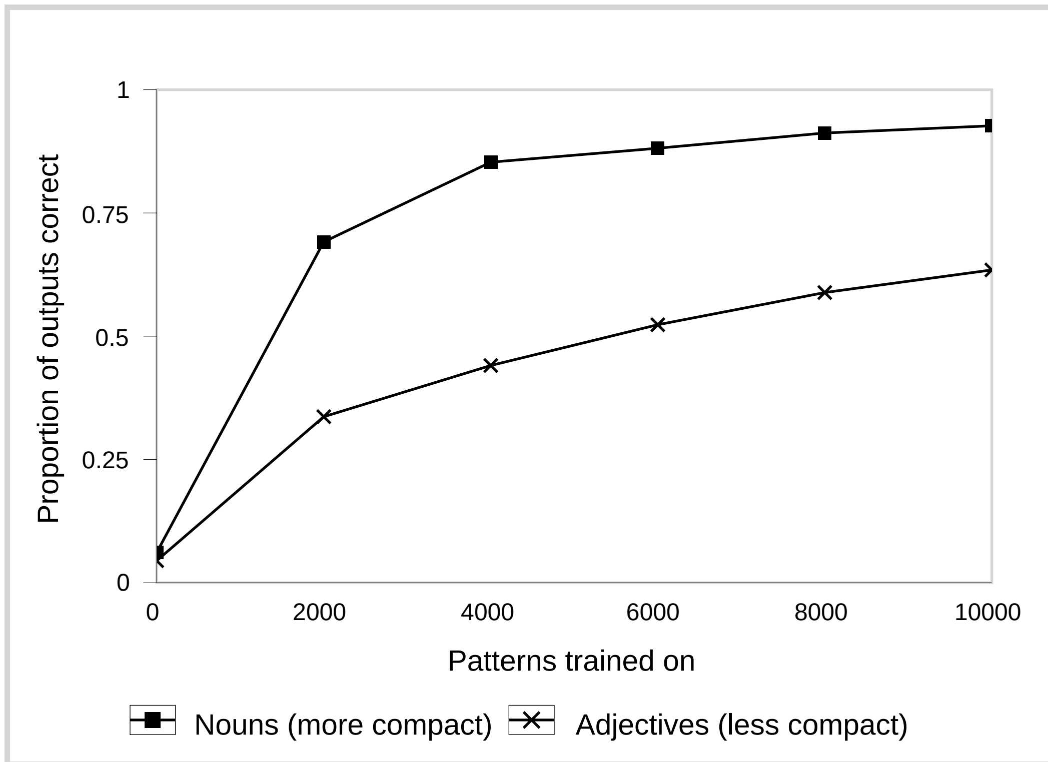

Figure 9: Learning Rates: Category Compactness (Experiment 4)

4.5 Experiment 5: Number of Categories

In Experiments 1 through 4, we asked the network to learn equal numbers of noun and adjective categories. This learning task contrasts sharply with children's task in learning nouns and adjectives. Children are faced with many more nouns than adjectives. Even if individual categories occurred with equal frequency, the child would be exposed to nouns more often. Does this fact make nouns as a class easier to learn? We have shown in this paper that a noun advantage can emerge without this difference. But is the frequency difference that exists in language sufficient for a noun advantage? In fact, it is not immediately clear whether this factor would lead to an advantage: the child has more experience with many nouns but has fewer distinctions to learn and less potential for confusion with adjectives. We investigated this factor in Experiment 5.

4.5.1 Stimuli and method

Stimuli were generated as above. As in Experiments 2 and 4, there were only two types of categories. Each category was defined in terms of ranges of 1/3, 1/3, 1/2, and 2/3 of the possible values along the input dimensions. Thus noun and adjective categories were characterized by identical representational span and compactness. However, there were 9 adjectives and 27 nouns. During training, instances of individual categories, together with the appropriate linguistic context (either *what is it?* or *what is it like?*), were presented to the network with equal frequency. Thus nouns occurred on the average 3 times as often as adjectives.

4.5.2 Results

The results of Experiment 5 are shown in Figure 10. In this case, performance on the adjectives is much higher than on the nouns. Thus the greater frequency of nouns over adjectives does not appear to be a contributory factor in the nominal advantage but instead may push in the opposite direction. The relatively large number of nouns leads to more confusions among the nominal categories than are found with adjectives. Not only the number of within-noun confusions, but the proportion of noun errors that are within the nouns, is higher than for adjectives (see Figure 11[2]). This difference outweighs the obvious advantage which accrues from the more frequent presentations of nouns. For comparison, the same data are shown for Experiments 2 and 4 in Figure 12. Here we see that the proportion of within-noun and within-adjective errors is closer, indicating that this difference is not sensitive to category span or compactness.

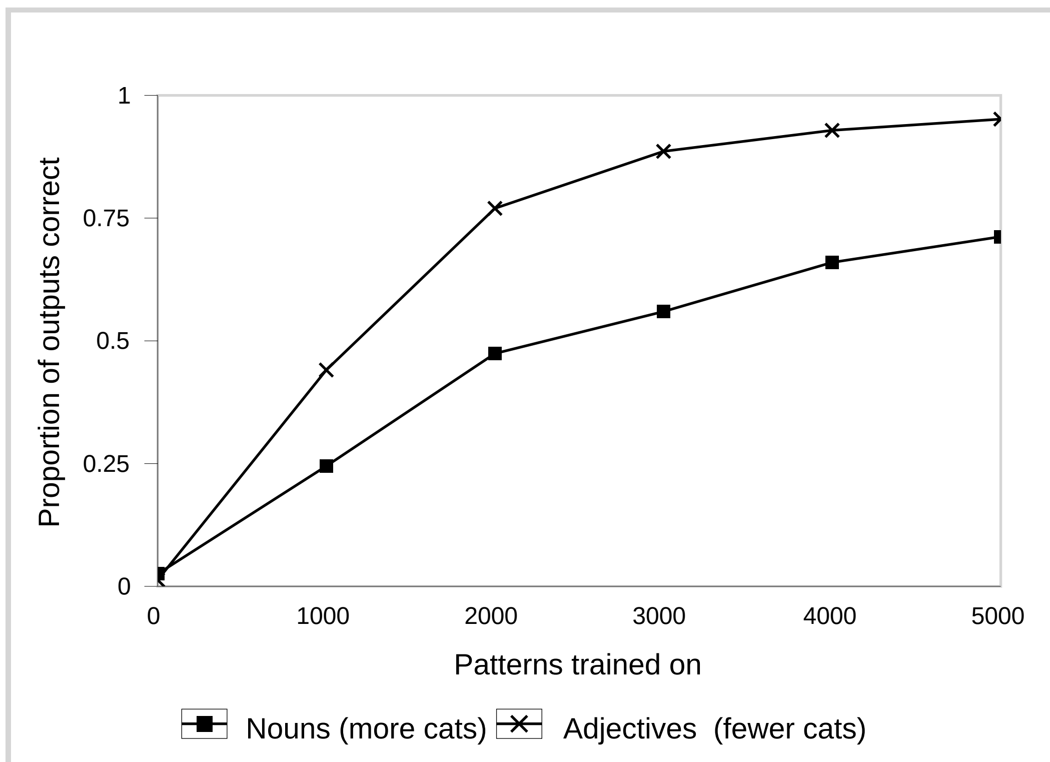

Figure 10: Learning Rates: Number of Categories (Experiment 5)

5 Discussion

Why do children learn nouns faster than adjectives? Most previous accounts explain this phenomenon in terms of knowledge the child brings to the language learning task—assumptions or biases about the meanings of words (e.g., Markman, 1989). In this paper we have shown that the differences in

[2] In this figure the results for the first 1000 training patterns are not shown because early in training there is a disproportionate likelihood that a single noun will be the network's response to all inputs. This makes the potential for errors within the nouns artificially high during this phase.

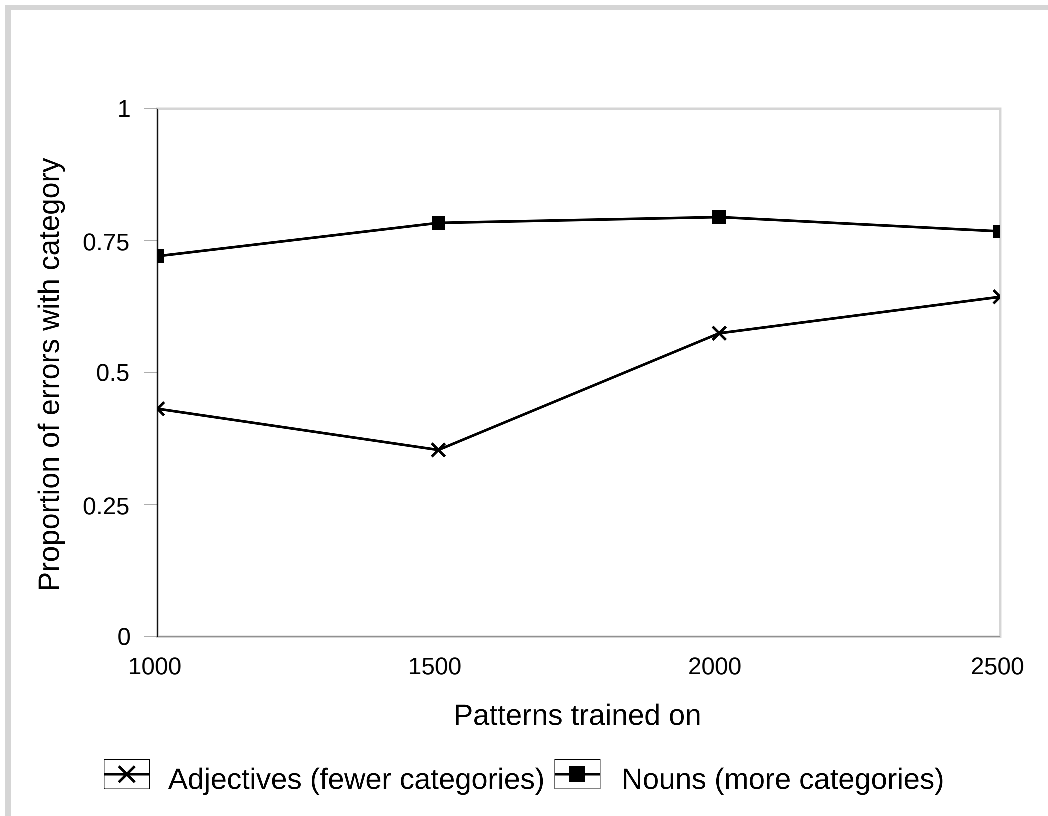

Figure 11: Proportion of Within-Noun and Within-Adjective Errors, Experiment 5

learning rate between nouns and adjectives emerge in a simple network, unendowed with special-purpose predispositions. That is, the differences emerge from the operation of general processes that learn "small" categories faster than "large" ones and that learn categories that do not require selective attention faster than those that do. In the present model, learning nouns is favored over adjectives because these two consequences of the general processing mechanism favor the sorts of *meanings* that nouns have over those that adjectives have.

The experiments reported in this paper have been primarily concerned with differences in the meanings of nouns and adjectives. Adjective categories tend to have greater representational span than nominal categories, and we showed in Experiment 2 that a connectionist network learns categories with smaller span more easily. Adjective categories are also more diffuse than nominal categories: fewer sensory dimensions are relevant for their definition; they cut large swaths thorough the representational space rather than the compact chunks that characterize nouns. In Experiment 4, we showed that a connectionist network learns compact categories more easily than diffuse ones. These two differences between the meanings of nouns and adjectives—span and compactness—appear to be the principal cause of the noun advantage in our model.

We also found a small noun advantage in Experiment 3 that derived from the added task in the adjective case of learning dimensional subcategories. In the case of learning nouns, the linguistic context of the question *what is it?* is associated with a large class of outputs. In the case of learning dimensional adjectives, the specific linguistic contexts of the questions *what color is it?*, *what size is*

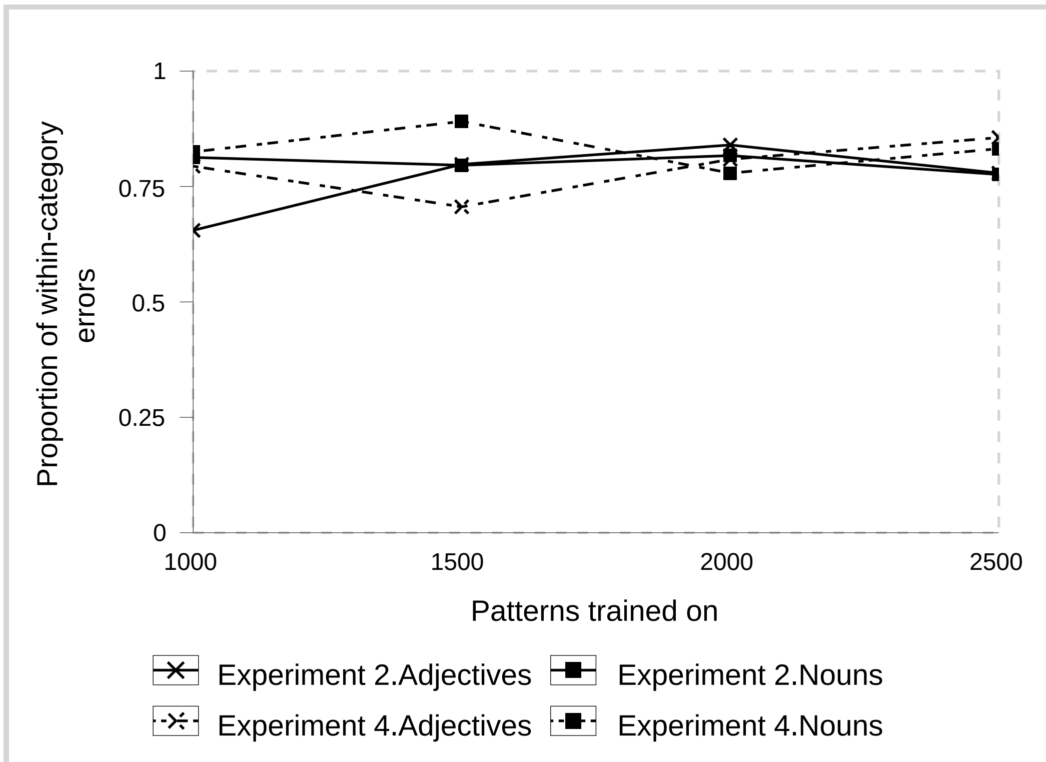

Figure 12: Proportion of Within-Noun and Within-Adjective Errors, Experiments 2 and 4

it?, *what shape is it?*, and *what texture is it?* are each associated with a small number of outputs (in the present cases with four words each). Experiment 3 showed that this difference in the learning task favored nouns slightly over adjectives. What is perhaps surprising in Figure 8 is the fact that the need to learn the three lexical dimensions in the adjective case does not cause the network more difficulty than it does. As shown in Figure 6, the network begins to sort out lexical dimensions rather quickly; most of its errors with adjectives are within-dimension errors. In other words, the network tends to learn to group adjectives with others from the same dimension early in learning. The network has, in a sense, learned to link certain possible responses (*red*, *blue*, *green*) to certain linguistic inputs (*what color is it?*) earlier than it learns to link a specific response (*red*) to a specific sensory input in a specific linguistic context. This aspect of the network's learning looks remarkably like that of children.

We also found one aspect of noun-adjective differences in their respective learning tasks that did not favor nouns—the greater number of noun categories relative to adjective categories. Given two sets of categories to learn, each associated with a particular linguistic context, our general category learner learns the set which is fewer in number faster than the set which is larger in number. Although this factor may work against nouns in the real learning of children, it may be overwhelmed by the the other factors working in favor of nouns. This idea points to a further conclusion to be drawn from our results: the complex developmental course of real children's word learning and the robust facts about that learning such as the noun advantage may have at their root multiple causes. They may be the product of a complex consortium of the properties of the learning task as they interact with the properties of the learner.

In this context, it is important to note that there are other differences between nouns and adjectives that may contribute (or modulate) the noun advantage that we have not examined yet. In the present work, we concentrated on aspects of the individual categories, but there may be differences between noun and adjective categories that have to do with the relations between categories. For example, nouns and adjectives might differ in the tendency of their meanings to overlap. Adjective categories typically overlap with categories on other dimensions (red things can be big or small) but noun categories at the concrete basic level typically do not (a dog cannot be a house or table). Nouns and adjectives also differ in the proportion of the representational space they cover as a class. Adjectives from a single dimension when considered together tend to cover all the representational space. That is, every point in the space can be described as *big*, *little*, or *medium-sized*. But nouns seem not to cover the entire space. Rather, there are combinations of values on the various dimensions that have no name. Put another way, nouns leave large gaps in the representational space and the fact of these gaps may make learning nouns somehow easier. We intend to investigate these factors, as well as others, in future work.

The approach we have taken in this work is to specify a general cateory learner, with no predispositions, no prior knowledge of the meanings of nouns and adjectives, and ask what factors in the learning task itself (rather than within the individual) may cause the general character of development to look the way it does, in the present case to cause nouns to be more rapidly learned than adjectives. The value of this modelling approach to word learning is precisely that it enables us to theoretically tease apart and examine the multiple factors alone and together that may contribute to causing learning to be the way it is. This approach is reminiscent of Plunkett & Marchman's (1991) investigation of the effects of type and token frequency on the learning by a network of morphological rules. We believe that *the least* this approach will contribute to the understanding of word learning, even if the basic aspects of the model and its fundamental assumptions about the learner are wrong, is a finer grained understanding of the learning task itself and how the the properties of the task may constrain the developmental trajectory. Of course, our principal theoretical goal in the present study is larger than this: the current aim is to show how the noun advantage can emerge from a general category learner and from the nature of the categories to be learned. In this larger theoretical context, there are two further points to be discussed: (1) the nature of the general category learner that we propose relative to other category learning models and (2) the possible emergence with learning of a distinction between nouns and adjectives of the kind assumed to be there by some theorists at the start of learning. We consider each of these issues in turn.

The network that we used to study the learning of nouns and adjectives is a general learnig device with few special properties to distinguish it. This is its strength for the task that we have set for ourselves, namely, how much of the noun advantage could come from the nature of learning task itself. Nonetheless, our model of category learning and some of our results different importantly from some other current models of category learning. One property that appears to distinguish category learning models is whether attentional learning is organized dimensionally. In models such as ours, Gluck & Bower's (1987), and Anderson's (1991) attention is not dimensionally organized. Thus, although we have dimensions in the sensory input, the representations of values on those dimensions are distributed and compressed at the hidden layer. Kruschke's (1992) model of category learning, on the other hand, has learning proceed by the altering of attention weights on dimensions. One difference between these two kinds of models is the relative ease with which they learn categories which are wholistically (and disjunctively) organized along many dimensions versus categories which are well-organized by one or two dimensions. Kruschke has argued from the classic results of Shepard, Hovland & Jenkins (1961) that people learn about dimensions and learn about categories organized by values on a single dimension far more rapidly than they learn about categories organized by values on many dimensions. This statement of the data to be explained—single dimension categories are easier than multiple dimension categories—is the exact opposite of natural category learning in children.

Kruschke's, Gluck & Bower's, and others' models in the literature on adult cognition, like the classic Shepard et al. (1961) paper, concentrate on adult's learning of a single partition of a small number of objects (16 to 50) into just two categories. Each individual object must be placed in one and only category and one and only one solution need be found. Children, in learning natural categories have a much harder task. They must learn to place individual objects in multiple categories; they must learn, for example, to call a dog "dog", "red", "big", and "furry". They must possess a general category learner that can learn multiple categories at the same time. We have proposed here such a multiple category learner and have asked it to learn a harder task than any contempory model of category learning that we know about. And our general category learner, like children, learns categories organized by many dimensions and spanning a small proportion of the representational space faster than it learns categories organized by values on a single dimension and spanning a large proportion of the representational space. Our model of category learning on these grounds alone seems a more plausible model of category development.

Our simple general category learning device knows nothing about the distinctions of nouns and adjectives at the start of learning. Of course, whether they know it at the start or not, children eventually do understand the deeper significances of nouns and adjectives—that nouns name things and that adjectives name the properties of those same things Markman (1989). Our model provides some insight as to how children might arrive at this distinction rather than start from it. The results of Experiment 5, in particular, suggest that the network had learned two distinctive syntactic categories—one for nouns and one for adjectives. The evidence for that distinction is this: given the linguistic context associated with nouns (*what is it?*), the network gave noun answers—either the correct answer or an incorrect noun. Given the linguistic context associated with adjectives (*what's it like?*), the network gave adjective answers—either the correct answer or another adjective. As shown in Figure 12, in Expiments 2 and 4 there were also many more errors within adjectives and within nouns than between the two categories. The fact that confusions were mainly between nouns or between adjectives shows that the system has begun to organize itself internally in terms of the distinction between adjectives and nouns; the syntactic differences have emerged. Thus although our models began learning without the tendency to treat to nouns and adjectives differently, it developed the disposition to treat them differently. Although our model began learning without any knowledge of the differences between nouns and adjectives, it learned that nouns and adjectives are different kinds of words.

6 Conclusions

We have shown that a general category learning device can, without prior knowledge of the distinction, learn nouns faster than dimensional adjectives. This demonstration, of course, does not show that children are general category learners nor that they approach word learning without some prior knowledge of the distinction of nouns and adjectives. Whatever the nature of children's knowledge at the start of learning, however, our results show that some basic properties of the task of learning nouns and learning adjectives may contribute to the noun advantage. They also suggest that prior ontological knowledge of nouns versus adjectives and the logical priority of nouns as the arguments of predicates, are not necessary to explain the noun advantage. Indeed, such knowledge could emerge in development and in part through the operations of a general category learner such as our network. An important next step in the pursuit of this idea is to test fine-grained predictions of the model against the category learning of children. Discovering, as we have done in these five experiments, the specific aspects of noun and adjective structure and of the child's learning task which lead to the noun advantage is the first step toward empirical tests of the model.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Blewitt, P. (1983). *Dog* vs. *collie*: vocabulary in speech to young children. *Developmental Psychology*, *19*, 602–606.

Callanan, M. A. (1990). Parents' description of objects: potential data for children's inferences about category principles. *Cognitive Development*, *5*, 101–122.

Carey, S. (1982). Semantic development: the state of the art. In Wanner, G. & Gleitman, L. R. (Eds.), *Language Acquisition: The State of the Art*. Cambridge University Press, Cambridge, MA.

Clark, H. H. (1991). Words, the world, and their possibilities. In Lockhead, G. R. & Pomerantz, J. R. (Eds.), *The Perception of Structure*, pp. 263–278. American Psychological Association, Washington, DC.

Cruse, D. A. (1977). A note on the learning of colour names. *Journal of Child Language*, *3*, 364–384.

Ehri, L. (1976). Comprehension and production of adjectives and seriation. *Journal of Child Language*, *3*, 369–384.

Gasser, M. & Smith, L. B. (1991). The development of the notion of sameness: a connectionist model. *Annual Conference of the Cognitive Science Society*, *13*, 719–723.

Gentner, D. (1978). On relational meaning: the acquisition of verb meaning. *Child Development*, *48*, 988–998.

Gibson, E. J. (1969). *Principles of Perceptual Learning and Development*. Appleton-Century-Crofts, New York.

Gluck, M. & Bower, G. (1987). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*.

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.

Macnamara, J. (1982). *Names for Things: A Study of Human Learning*. MIT Press, Cambridge, MA.

Maratsos, M. (1988). Crosslinguistic analysis, universals, and language acquisition. In Kessel, F. (Ed.), *The Development of Language and Language Researchers: Essays in Honor of Roger Brown*, pp. 121–152. Lawrence Erlbaum, Hillsdale, NJ.

Markman, E. M. (1989). *Categorization and Naming in Children: Problems of Induction*. MIT Press, Cambridge, MA.

Mervis, C. B. (1987). Child-basic object categories and lexical development. In Neisser, U. (Ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge University Press, Cambridge.

Murphy, G. L. & Medin, D. L. (1985). Role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.

Plunkett, K. & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition*, *38*, 1–60.

Rescorla, L. A. (1980). Overextension in early language. *Journal of Child Language*, *7*, 321–335.

Rosch, E. (1973). Natural categories. *Cognitive Psychology*, *7*, 573–605.

Rumelhart, D. E., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In Rumelhart, D. E. & McClelland, J. L. (Eds.), *Parallel Distributed Processing, Volume 1*, pp. 318–364. MIT Press, Cambridge, MA.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*, (13, Whole No. 517).

Smith, L. B. & Sera, M. (1991). A developmental analysis of the polar structure of dimensions. *Cognitive Psychology*.

Smith, L. B., Jones, S., & Landau, B. (1992). Count nouns, adjectives, and perceptual properties in children's novel word interpretations. *Developmental Psychology*.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, *96*, 125–144.

Smith, L. B. (1993). The concept of same. In Reese, H. W. (Ed.), *Advances in Child Development and Behavior, Vol. 24*. Academic Press, New York.

Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language*, *4*, 1–22.

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meanings: object terms and substance terms. *Cognition*, *38*, 179–211.

Wittgenstein, L. (1953). *Philosophical Investigations*. Macmillan, New York.

Wood, D. J. (1980). Teaching the young child: some relationships between social interaction, language, and thought. In Olson, D. R. (Ed.), *The Social Foundations of Language and Thought*. Norton, New York.