J. Child Lang. **33** (2006), 759–790. © 2006 Cambridge University Press
doi:10.1017/S0305000906007574 Printed in the United Kingdom

Language input and semantic categories: a relation between cognition and early word learning*

ARIELLE BOROVSKY AND JEFF ELMAN
University of California, San Diego

(*Received* 9 *November* 2004. *Revised* 4 *February* 2006)

ABSTRACT

Variations in the amount and nature of early language to which children are exposed have been linked to their subsequent ability (e.g. Huttenlocher, Haight, Bryk, Seltzer & Lyons, 1991; Hart & Risley, 1995). In three computational simulations, we explore how differences in linguistic experience can explain differences in word learning ability due to changes in the development of semantic category structure. More specifically, we manipulate the amount of language input, sentential complexity, and the frequency distribution of words within categories. In each of these simulations, improvements in category structure, are tightly correlated with subsequent improvements in word learning ability even when the nature of the input remains the same over time. These simulations suggest that variation in early language environments may result in differences in lexical proficiency by altering underlying cognitive abilities like categorization.

INTRODUCTION

The ability to group objects into categories based on some similarity of function, form or meaning is arguably one of our most important cognitive behaviours. While it may be a common-sense notion that we form categories based on our own direct experience of them, we also develop categories for things we may have never directly experienced like Roman emperors, subatomic particles, and vacuum-tube computers. A reasonable assumption is that language plays a key role in making this possible. But a similar issue arises in the case of language: we ultimately are able to learn

[*] This work was supported by NIMH grant (MH6051701A2) to Jeffrey Elman. We are grateful to Holger Keibel for his invaluable advice on measurement of category structure. This work is dedicated to Elizabeth Bates, whose input was crucial in influencing earlier versions of work, and without whom this project would have never begun. Address for correspondence: Arielle Borovsky, Department of Cognitive Science #0515, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA. e-mail: aborovsk@cogsci.ucsd.edu

words for things outside our direct experience. This poses a chicken-and-egg problem. Intriguingly, during early stages of language learning there is little evidence of category knowledge, and the rate of vocabulary acquisition is slow. At the point when children undergo a 'vocabulary spurt' – in which the pace of word learning increases rapidly – they also begin to display the ability to sort sets of objects into multiple categories (Gopnik & Meltzoff, 1987, 1993). This suggests that these two phenomena, i.e. the ability to learn new words and knowledge of categories, may be related in a synergistic fashion. In this paper, we use computational simulations to explore how language input influences development of category knowledge, and how category knowledge in turn influences subsequent lexical acquisition.

We begin with a brief review of the claims that have appeared in the literature regarding the nature of the relationship between category knowledge and lexical development. We then turn to a review of what is known about the effects of language input on vocabulary acquisition. Finally, we focus on a specific hypothesis – that even in the absence of experiential information, vocabulary acquisition can shape category knowledge, which once in place, then facilitates the rate of learning new words – and study the conditions under which these two phenomena interact.

The relationship between lexical development and category knowledge in children

How might the ability to categorize objects and the ability to learn new words be related? Several logical possibilities exist, ranging along a spectrum of being tightly interrelated to not related at all. The former position is suggested by Fodor (1975), who argues that linguistic and cognitive abilities are completely modular or domain specific and develop independently of each other. According to this account, semantic structure is not acquired due to language input. Rather, conceptual structure is innate, and words are learned that correspond to this innate 'mentalese.' Chomsky (1981) and Pinker (1991) have advanced similar – though not identical – positions.

But because there appears to be a confluence of rapid gains in cognitive and linguistic functioning around the middle of the second year, others have proposed that linguistic and other cognitive abilities such as categorization, develop in a more tightly coupled manner. Within this camp, theories tend to differ on degree of this relationship, ranging from the strongly Whorfian (Whorf, 1956) hypothesis that thought is impossible without the use of language, to more interactive versions where language and cognition are completely integrated and rely on each other during development. We take this to be the position of, for example, Gopnik & Meltzoff (1993).

The possibility that word-learning might specifically drive conceptual organization has been proposed by Bowerman and colleagues (Bowerman, 1996; Choi & Bowerman, 1991). Bowerman cites evidence from crosslinguistic studies that children will readily learn the spatial categorization scheme present in their language. For example, Korean focuses on the fit between objects, while English tends to emphasize the kind of containment or method of support. However, Choi & Bowerman (1991) find that children learning each language have no problem describing these spatial relations in terms appropriate to their language. So, Korean children will describe a peg in a hole, as tight or loose fitting, while English-speaking children will say whether the peg is 'in' or 'on' the hole.

Mandler (1996) has countered this claim by proposing that younger children might be able to carve-up space in ways that are not solely linguistically determined. She suggests that there are a number of different ways that the world can be categorized – all of which are initially available, but that language constrains these possibilities into a regularized convention that differs from language to language. Consistent with this claim is the finding that children at 0;9, 0;11 and 1;2 are all able to distinguish between tight/loose and in/out distinctions, regardless of linguistic environment (Choi, McDonough, Mandler & Bowerman, 1999).

The above debate delineates two possibilities regarding the language-categorization interaction. The former proposes that language drives the kind of categorical structure that is formed. The latter maintains that nonlinguistic categories are already formed from the infant's physical knowledge of the world, and that language 'slots' into these prior categories. By extension, one might assume that category knowledge facilitates vocabulary acquisition.

Lastly, Gopnik & Meltzoff (1993, 1997) have argued that cognitive and linguistic development is related to each other in a more tightly interactive way, and that both can influence each other equally. For instance, they propose that exposure to a particular word will lead to a drive to learn the underlying concept, but that this new knowledge will then interact with knowledge already in place, which might then lead to new word learning. Evidence for this idea is found from crosslinguistic studies in which the appearance of categorization abilities arise later in Korean speaking children than in English speaking children (Gopnik, Choi & Baumberger, 1996). In Korean, verbs are much more prevalent than nouns, but the opposite is true in English. Since English speaking children are able to sort objects into categories earlier, this suggests that the language's emphasis on nouns drives the organization of (at least nominal) concepts into categories earlier than in Korean, and that this improved categorization ability is what facilitates subsequent word learning proficiency.

Logically, whatever the relationship between vocabulary acquisition and categorization abilities, the role of the actual input presented to a child must itself play a critical role in the acquisition process. We turn now to a review of what is known about the effect that input has on vocabulary acquisition.

The role of input

A growing body of research has found that early language input is key to predicting levels of lexical proficiency. For example, a number of studies (Huttenlocher *et al.*, 1991; Hart & Risley, 1995) have found that children who have had more language input from their parents also know more words.

Additionally, some of the earliest observations about child directed speech (CDS) (e.g. Snow & Ferguson, 1977; Newport, 1977) have found that the language that children hear is simpler in both syntactic structure and the kinds of words used. For instance, one ambitious study of CDS to 12 children between the ages of 2;0 and 3;0 years old (Cameron-Faulkner, Lieven & Tomasello, 2003) found that 20% of all the utterances recorded during this period were sentence fragments, which usually were responses to a question, and that 32% of the utterances were questions. Complex sentences only accounted for 6% of all utterances heard by children at this age. These results were also compared with a similar, but smaller study by Wells (1981), who reported strikingly similar results.

These findings imply that simplifications in the structural complexity of CDS could aid word learning by reducing the processing demands placed on the child when a new word is encountered. However, in order to support this idea it is necessary to compare the kinds of input heard with the words actually learned by the child. To this end, Brent & Siskind (2001) examined the role of single word input in infants between 0;9 and 1;3 on the words that the child knows at 1;6 – i.e. at the beginning of the vocabulary spurt. They found that the number of times a child hears a word in isolation is a more reliable predictor of whether the word is known at 1;6 than the total number of times the word is heard. By examining the role of isolated words, the investigators were able to study how the words heard in the simplest kind of grammatical construction affect the learning of that word. When taken into account with the other CDS findings described above, this work indicates that word learning as whole might be easier when the sentence structures remain simple.

On the other hand, there are also recent findings that more structural complexity in CDS improves word learning at 2;0 (Hoff & Naigles, 2002). However, here syntactic complexity was not measured through detailed constructional analysis, but rather through the mean length of utterance (MLU) in maternal speech. Hoff & Naigles (2002) concluded that it is more syntactic complexity, not less, that is important in aiding word learning.

In addition to differences in the structure of the input, there is also evidence that the distribution of words in input may differ amongst children (Bates, Bretherton & Snyder, 1988; Broen, 1972). Weizman & Snow (2001) report that the usage of low frequency words varies between families, and that five year old children who encounter a higher proportion of these 'sophisticated words' from their environment also tend to have larger vocabularies. More recently, Pan, Rowe, Singer & Snow (2005) also find that variation in type and token frequency in maternal speech to children in the first three years of life also affects vocabulary growth. Moreover, they find that having a increased variation in maternal types was more significant than just overall amount of speech input alone.

We are thus left with a number of unresolved questions. That the input matters, and that there is some relationship between language development and category knowledge, seems almost trivially self-evident. But the specific effect of the input that a child hears upon vocabulary acquisition, and what the precise relationship between the word learning and category knowledge, remains unclear. Although the answers to these questions will ultimately come from empirical investigations, it would seem that there is a useful role to be played by using computer simulations to explore the process of lexical acquisition. In this way, it is possible to develop more specific hypotheses about the effect of different types of input, and the relationship between lexical acquisition and category knowledge.

Computational studies have been used in the past to model lexical learning (i.e. Plunkett, Sinha, Moller & Strandsby, 1992; Li, Farkas & MacWhinney, 2004). These models have been able to replicate phenomena that have been observed in lexical development of children. Of relevance to this paper, Li, Farkas & MacWhinney (2004) find that organization of lexical categories (nouns, verbs, adjectives and closed class words) in self organizing networks improve as the networks learn more words. This suggests that modelling can reveal important links between category and lexical development.

There are a number ways in which computational simulations complement behavioural research. While it is difficult to devise a task that can directly measure internal representations in young children, and it would be unethical to alter language input to such an extent that it might disrupt normal processes of language acquisition, computational simulations can overcome these difficulties. This makes it possible to measure category structure both by artificially manipulating how well category members fit together, and more importantly, to evaluate the actual representations that form. That is, rather than inferring category knowledge from task behaviour (as is done with children), the network's category knowledge can be assessed directly through analysis of its internal representations. We use artificially generated language input in order to have precise control over the properties of input, as it difficult to find these kinds of neat and tidy kinds of variation in corpus data of CDS.

Another benefit of the computational methods employed in this paper is that it is possible to isolate the role that linguistic input alone may play in the development of conceptual structure, apart from other non-linguistic kinds of information that are undoubtedly used in lexical acquisition. In this way, our networks learn about words by their co-occurrence with other words. Previous computational simulations have demonstrated that such information can provide important information about category structure (Elman, 1990, 1998). In the following simulations, we probe how categories that are learned in this way may be related to rate of word learning.

As a start, we propose that associations made from language input may alter underlying category structure, and that it is this change in category structure that can be related to proficiency in word learning. Under this hypothesis, category development is an important factor in word learning, so it can be expected that factors in language input that affect lexical learning outcomes should also be reflected in the development of category structure similarly. This leads to two key predictions:

(1) Variation in language input that affects lexical acquisition also affects development of category coherence in a similar manner; and
(2) It is the development of categories themselves, and not solely language input, that facilitates lexical acquisition.

In the remainder of this paper we examine this hypothesis and its predictions in three computational simulations with connectionist networks. In the first study we examine the role of the amount of input in developing category structure and subsequent acquisition of new words. The second investigates the role of syntactic complexity on this relation. The third experiment examines how differences in the distribution of word frequency in affecting this relationship.

The modelling task

The purpose of the following simulations is to explore the ways in which language experience might affect cognitive development, and how such development might in turn impact word learning ability. To this end, the network's category knowledge is assessed directly through analysis of its internal representations. Before proceeding to describe the simulations, we provide a brief explanation of the neural network model that is used and explain how this type of model allows us to measure development of category structure.

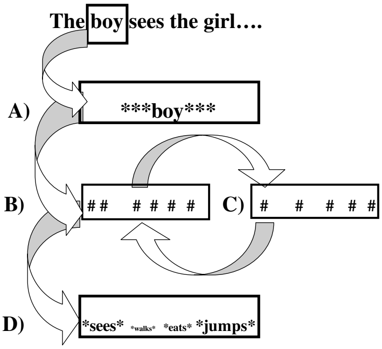


Fig. 1. Above, words from the sentence 'the boy sees the girl' are presented one word at a time to the network to the input layer (A). Next, the word is fed into the hidden layer (B), which also receives information about the immediately previous network states from the recurrent layer, (C). As training proceeds, the hidden layer will develop numerical internal representations that can be used to generalize to similar inputs. In the output layer (D) the network predicts the next possible words after 'boy'. Generally, the SRNs will learn to predict a range of words that are possible that correspond to their frequency with which they are associated. Here, the network is predicting both the actual next word 'sees' but also other possible words like 'jumps' and 'eats'.

Simple recurrent models

In this paper, we use the simple recurrent network architecture (SRN, Elman, 1990). Simulations were run using the TLEARN software (Plunkett & Elman, 1997). This type of network is especially useful for processing elements that are sequential in nature, such as words in a sentence. Figure 1 illustrates the architecture of an example network. In our simulations, the network receives individual words as input, one at a time. The network's task is to predict the next word in the sentence as its output. The output that is produced is a function of both the current word input to the network and the prior internal state of the network. stored in the context layer (Figure 1c). Importantly, this prior internal state is not a literal tape recording of preceding words, but is rather an abstract representation – that must be learned – of that sequence. The hidden layer that reflects these internal states (Figure 1b) is the part of the network that will be analyzed to provide evidence the network's knowledge of category structure, as described below.

Measurement of category structure

The use of language-like input allows for examination of the kinds of representations that might form when words are related by their occurrence in similar contexts to other words in the category. This involves active probing of the network to measure the coherence of representations that the

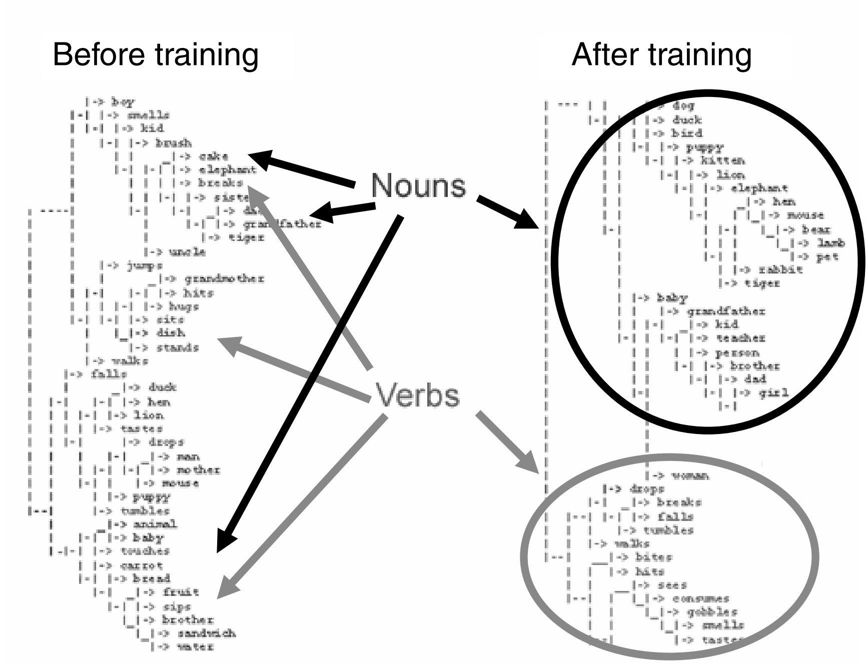


Fig. 2. Hidden Unit representation of vocabulary items in a young network before extensive training, after 20,000 sweeps and then at the end of training at 140,000 sweeps.

networks have developed from this kind of input. More specifically, measurement of category formation is accomplished through calculation of the network's internal (hidden unit) representations of words in a particular category (see Figure 1 for an example network).

Here, words that the network has learned to be similar (more precisely, in the sense that they share similar linguistic properties), or to belong to a similar category, will share hidden unit activations that are more similar than those that are not within a learned category. For example, it is possible to cluster these representations graphically (Figures 2 and 3) to visually reveal the similarity of hidden unit values for each word.

Figure 2 outlines how hidden unit representations change over training in this study. Before the network has adequately learned about category structure, words are clustered without any noticeable relation to each other. Yet, at the end of training it is clear that the network has learned not only the differences between nouns and verbs, but also subcategories between them (Figure 3). By comparing how close all members of a category are to each other, it is possible then to measure the 'global coherence'. Here, we follow Keibel, Elman, Lieven & Tomasello's (submitted) use of Average

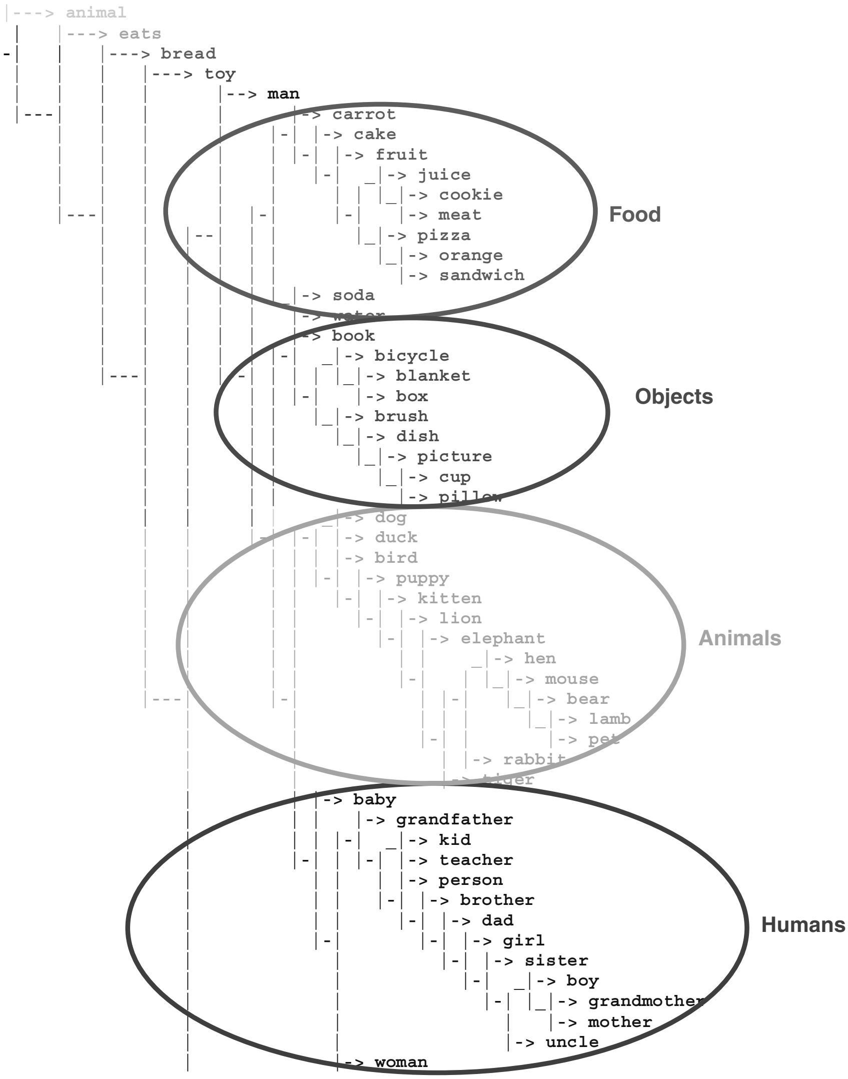


Fig. 3. Close-up of hidden unit cluster of noun items in older network.

Precision (Zavrel, 1996) for calculating global coherence. In this technique, the vector distance between each pair of words is calculated, and then for each word, other words are ranked on this distance measure. A coherence score is then calculated for each word based on how close other members of its category are ranked. The coherence score is then averaged across all members of a category to determine the overall coherence of the category. As Keibel *et al.* (submitted) points out, AVERAGE PRECISION (AP) is appropriate in situations where the number of members in different categories is not the same, which is the case in this study. AP values range on a scale between 0 and 1, with higher values signifying that members of a category have more similar hidden unit activation values. Essentially, categories that are well-formed should have higher AP than those that do not (see Appendix 1 for a detailed explanation of AP).

The use of computational simulations also allows for identical networks to be exposed to different training environments. In this case, we alter the exposure to both the quantity of input, but also qualities of its structure and frequency of words within a category, as outlined in the next three sections.

Effect of quantity

The amount of input has been shown to play a role in lexical acquisition, and the appearance of categorization abilities seems to be coincident with improvements in lexical acquisition. If these two are related, then simulations should show that networks that receive larger amounts of input also will have higher category coherence. At the same time, these networks should also be able to learn new words more quickly. This would lend support to the idea that improved linguistic learning is related to category development. This hypothesis can be further explored by manipulations of the input that more directly but subtly affect category development.

Effect of frequency

In order to fully examine how word learning might be influenced by category coherence, it is necessary to compare input conditions that are very similar, but in which one leads to higher coherence values than does the other. A recent proposal by Goldberg, Casenheiser & Sethuraman (2004) suggests one way this may be possible.

In that work, a corpus study revealed that there are five highly frequent verbs in CDS that correspond to five common constructional categories. Furthermore, an accompanying experimental study found that when new verb constructions are taught with a highly frequent exemplar, novel verbs with the same meaning construction are learned more easily for both adults and children (Casenheiser & Goldberg, 2005). This suggests that networks also might form a particular category such as ANIMAL more readily if they are exposed to input in which one animal word, such as *cat*, is more frequent than other animal words. Conversely, networks that are exposed to all words in a category with equal frequency should form less coherent categories (at least initially), and thus, when given a task to learn a new category member word, should learn less quickly than networks that have been induced to form a more coherent category through skewed word frequency exposure.

To summarize, two predictions are made. First, networks that are exposed to input where there is one very frequent word per category should form more coherent categories than those that have no frequency differences between words in the category. Second, networks that form categories with lower coherence, as measured by AP values, should learn new words in the low coherence category more slowly than networks with higher coherence values for a category.

Effect of syntax

Finally, as discussed earlier, there is some uncertainty in the literature about whether more or less syntactic complexity in CDS is better for early language learning. One drawback in studies of speech to toddlers is that data are often collected sporadically in lab visits, or, in cases where children are recorded many times and at home (Cameron-Faulkner *et al.*, 2003), the sheer amount of the data precludes a having large number of participants from a variety of backgrounds. Additionally, the success of these studies hinges upon that amount of variation that can actually be observed through recording sessions since it is not possible to systematically change the kinds of language input a child may hear on a large scale. On the other hand, it is possible to know in detail about the entirety of experience a neural network has with language, and to experimentally vary important properties of this input.

Of course, what counts as grammatical complexity is itself a complex question and one can imagine many ways in which utterances might be judged to be more or less complex. In this case, the most straightforward manipulation that lends itself to examination of how grammatical complexity of input may affect word learning is to present networks with input that contain only simple, transitive and intransitive sentences or networks that contain this simple input plus more complex ditransitive and matrix sentence constructions. If this additional grammatical complexity (as defined in this very specific manner) does indeed hinder vocabulary growth, then networks trained with the latter type of input should learn new words more slowly than networks that have been exposed to only more simple constructions. Second, this slower word learning should then be associated with lower AP values.

EXPERIMENT 1: EFFECT OF QUANTITY

METHOD

Input

The input for all simulations that are described in this paper was constructed using a language generator program (SLG; Rohde, 1999). In the Experiments 1 and 2, a vocabulary of 85 words (52 nouns and 33 verbs) was used and sentences were formed corresponding to two simple syntactic constructions containing only nouns and verbs: NV and NVN (intransitive and transitive, respectively).

Nouns were assigned to the following semantic categories: ANIMALS (15), HUMANS (15), FOOD (12), and OBJECTS (10). Nouns in these categories are commonly observed in the early vocabularies of toddlers (i.e. Nelson, 1973), and are included on checklists for vocabulary checklists at this age (Fenson Dale, Resnick, Bates, Thal & Pethick, 1994). Verbs belonged to the following categories: CHANGE OF STATE (5), COMMUNICATION (6), MOTION (6), EATING (5), PERCEPTION (6), and ACTION (5). These verb categories were chosen because they are typically included on checklists for vocabulary (Fenson *et al.*, 1994).

In the artificial grammar that was used, nouns and verbs were required to agree both semantically and syntactically, meaning that a sentence had to be grammatically correct and 'make sense'. Each word was coded in a localist fashion as an 85 element binary-valued vector with each bit representing a distinct word. Appendix 2 contains appropriate categorical semantic relations between sentences. Appendix 3 contains examples of sentences used in the study.

The amount of input was manipulated by altering the numbers of sentences to which the network is exposed. Training involved input ranging between 20 and 1000 sentences, depending on condition; there were five conditions with corpus sizes of 20, 50, 100, 200, 500 and 1000 sentences respectively. Table 1 contains information about the number and kinds of types and tokens for each corpus size.

Although the relationships between categories was fairly simple, the largest input condition (1000 sentences) still presented the network with only a small subset of all possible sentences possible in this grammar. It is estimated that in order to see every possible sentential combination the network would have to see nearly half a million sentences. By training with only a subset of all possible data, this allowed for the network to be exposed to both a range of types and tokens, but still not see every possible combination, such that category membership was not plainly 'given away'. Instead the network was forced to generalize from incomplete input in order to figure out which words belong to which categories.

Training

For each of the input conditions listed above, 10 simple recurrent networks (SRNs, see Figure 1 for an example) with 50 hidden units were trained on a next-word prediction task. In this task, the network is presented with successive words in a sentence, one at a time, and is trained to predict the

TABLE 1. *Number of word types in each condition*

No. of sentences	Condition Even	 Uneven	 Simple	 Complex
1000	85	85	85	95
Verbs	33	33	33	43
Animals	15	15	15	15
Humans	15	15	15	15
Food	12	12	12	12
Objects	10	10	10	10
500	85	—	—	—
Verbs	33	—	—	—
Animals	15	—	—	—
Humans	15	—	—	—
Food	12	—	—	—
Objects	10	—	—	—
200	79	—	—	—
Verbs	32	—	—	—
Animals	14	—	—	—
Humans	15	—	—	—
Food	10	—	—	—
Objects	8	—	—	—
100	62	—	—	—
Verbs	25	—	—	—
Animals	11	—	—	—
Humans	12	—	—	—
Food	7	—	—	—
Objects	7	—	—	—
50	44	—	—	—
Verbs	19	—	—	—
Animals	11	—	—	—
Humans	6	—	—	—
Food	4	—	—	—
Objects	4	—	—	—
20	33	—	—	—
Verbs	15	—	—	—
Animals	5	—	—	—
Humans	6	—	—	—
Food	4	—	—	—
Objects	3	—	—	—

next word. Because the task is non-deterministic, the network's optimal strategy should be to learn the implicit classes of words that are appropriate in each context. Note that these categories are defined solely by privilege of occurrence (i.e. the input vectors themselves contain no information regarding category membership).

The learning rate was 0·01, no momentum was used, and training was carried out to 140,000 sweeps (one sweep corresponds to presentation of one word; pilot studies had determined that was sufficient training to result in asymptotic performance on the task). Thus, each network saw an equal number of words, but was exposed to each corpus a differing number of times, depending on the size of the corpus. This training scheme was meant to simulate a number of children who are the same age, but have heard different amounts of language input. Therefore, all networks had the benefit of being trained for an equal duration. The only difference was the range of sentences seen.

Analysis

At intervals of 20,000 sweeps, each network was probed to assess its ability to learn new words and also the category structure of its internal representations (i.e. hidden unit activations in response to seeing each word) as measured by AP scores.

New word learning was measured by exposing the network to novel sentences containing words the network has not previously seen. First, five sentences containing five instances of one new noun were exposed to the network for 50 sweeps, with the learning state of the network being captured every five sweeps. Over 50 sweeps, this translates to the new word being exposed between 13–14 times. Then, at the five sweep intervals, the network was tested for its ability to predict the new word in five previously unseen sentences, and the node activations of the new word was recorded. This was done for four new nouns – one for each noun category. This was meant to be analogous to examining the ability of a child to name a new word in a cued context after hearing the word a certain number of times.

At the same time, the AP value was also determined for individual noun categories, to be able to compare improvement in AP with word learning across corpus sizes.

RESULTS AND DISCUSSION

Figure 4 shows the average output unit prediction over training averaged over five new nouns that were taught to the networks after 140,000 sweeps. Higher node activation indicates better performance on the prediction task. Figure 5 shows the AP values of the network over time. From Fig. 4, we see the trend that larger corpus sizes provided an earlier advantage in new word learning, with new word being predicted in appropriate contexts earlier in training than occurs in networks that have been exposed to smaller corpus sizes. Consistent with our hypothesis, there were significant differences as a function of size of training corpus $F(5, 234) = 42{\cdot}4$, $p < 0{\cdot}0001$. *Post hoc* tests using Tukey's HSD (Table 2) revealed that word learning between 20 and 50 sentences were comparable, but smaller than all other training input

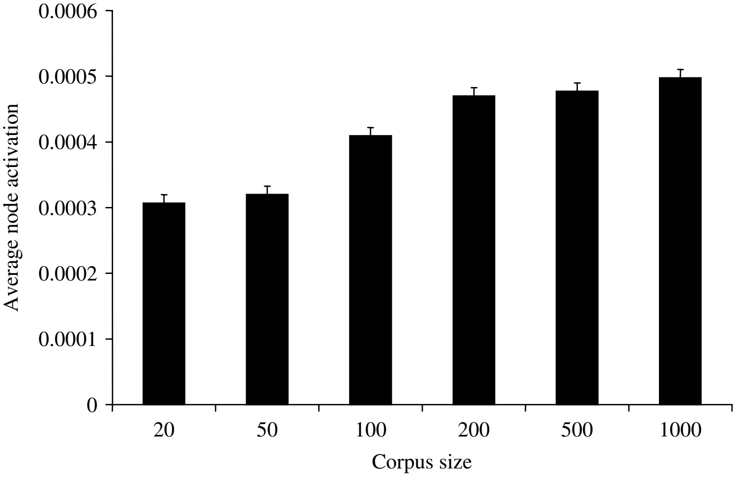


Fig. 4. Word learning across size of input. Average node activation is plotted across corpus size for the prediction of the new word in a novel context.

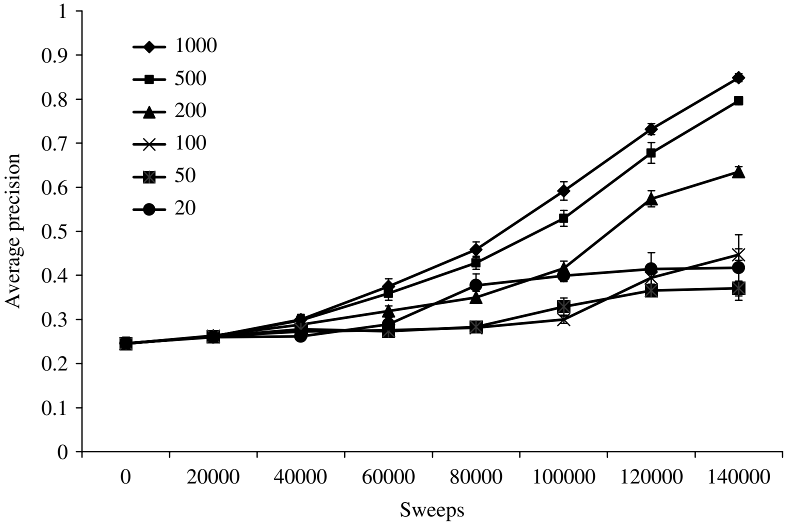


Fig. 5. Change in average precision values over training for each corpus size.

sizes. Additionally, word learning between 200, 500, and 1000 sentences did not differ. However, the word learning with the 100 sentence corpus was larger than 20 and 50 sentences, but smaller than 200, 500 and 1000 sentence corpora.

TABLE 2. *Word learning means and standard deviations for different training corpus sizes*

Corpus size in sentences	Mean (S.D.)
20	0·307$_a$ (0·000070)
50	0·320$_a$ (0·000042)
100	0·409$_b$ (0·000073)
200	0·470$_c$ (0·000084)
500	0·477$_c$ (0·000099)
1000	0·498$_c$ (0·000100)

Note: Subscripts indicate *post hoc* comparisons using Tukey's HSD. Means that do not share a common subscript are significantly different at $p<0{\cdot}05$.

TABLE 3. *Average precision means and standard deviations for different training corpus sizes*

Corpus size in sentences	Mean (S.D.)
20	0·418$_a$ (0·074)
50	0·371$_b$ (0·017)
100	0·447$_a$ (0·0133)
200	0·636$_c$ (0·011)
500	0·796$_d$ (0·008)
1000	0·849$_e$ (0·009)

Note: Subscripts indicate *post hoc* comparisons using Tukey's HSD. Means that do not share a common subscript are significantly different at $p<0{\cdot}05$.

Next, AP values at the end of training were compared across size. It was predicted initially that higher AP values (which reflect greater category coherence) would also be associated with larger training corpora. Figure 5 shows the change in AP by corpus size over training. There were significant differences at the endpoint in training of AP value as a function of size of training corpus, $F(5,54)=404{\cdot}21$, $p<0{\cdot}0001$. *Post hoc* tests using Tukey's HSD (Table 3) revealed that AP differences were higher between larger corpus sizes, except between the 100 and 20 sentence corpus, where there was no difference. This supports the hypothesis that like word learning, higher AP values are attained by the network over time with networks that have the benefit of larger corpora.

In order to better ascertain the relationship between word learning and category coherence, a simple regression was conducted with the AP values from the final point in training for each corpus size across the word learning node activation values. Figure 6 shows that higher AP scores were very highly correlated with better rates of word learning across corpus size, $R^2=0{\cdot}68$, $F(1,58)=123{\cdot}27$, $p<0{\cdot}0001$.

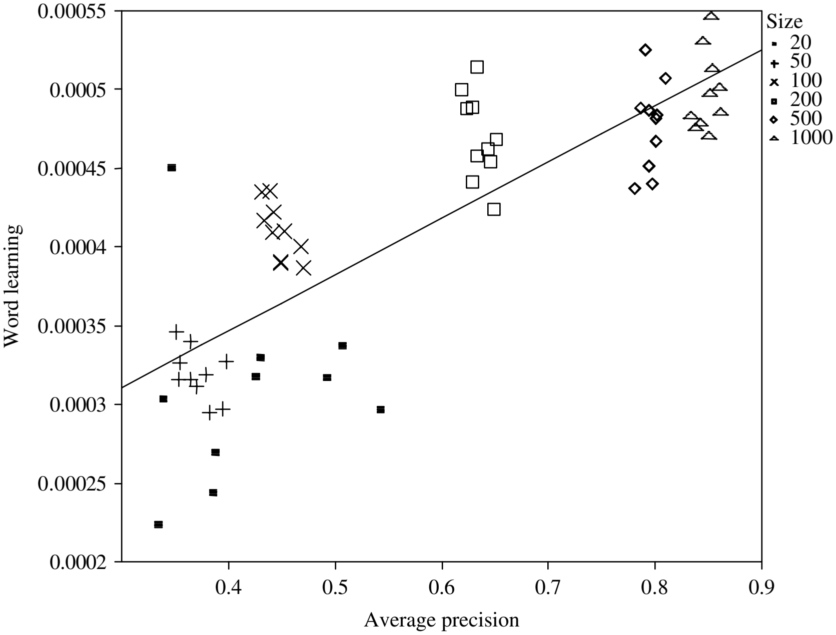


Fig. 6. Average node activation across average precision value for each corpus size.

In sum, these results are consistent with previous findings in the acquisition literature that increased exposure to language results in better lexical acquisition skills. Here, networks that were exposed to input that was both more varied, and had more tokens reaped the benefits of being better able to learn new nouns that they were subsequently exposed to. Additionally, this simulation also supports the hypothesis that better noun learning is indeed associated with increased coherence of noun categories. This is consistent with both the notion that categorization abilities that appear around the time lexical learning improves could index better category structure, as well as the idea that it is possible to influence cognitive development (in the specific sense of inducing category structure) from linguistic input alone.

But can subtler differences in input also affect rate of word learning and category structure? It has long been known that the vocabulary to which children are exposed is skewed, in the sense that there are dramatic differences in the frequency of different lexical items (Bates, Bretherton & Snyder, 1988; Broen, 1972). More recent work suggests that not only do these differences vary across families and children, but that the frequency with which different members of a category occur may play a role in learning and generalization. Bybee (1995) has noted such effects in the domain of morphological generalization, and Goldberg and colleagues present experimental findings that suggest that the presence of a high frequency category exemplar during learning can facilitate the learning of categories (Casenhiser & Goldberg, 2005; Goldberg, Casenhiser & Sethuraman, 2004). In the next study, we test this possibility by keeping the type frequency (number of different words) constant, but varying the token frequency of individual words.

EXPERIMENT 2: EFFECT OF FREQUENCY

METHODS

Input

The training input was constructed with the same characteristics from Experiment 1, with the same number of sentences, semantic and syntactic relations and vocabulary. The major difference was that the frequency of the word tokens was altered. There were two conditions: Even frequency and Uneven frequency. In the Even frequency condition, all members in a category were equally frequent, while in the Uneven frequency condition, one member of a category was much more frequent than the other members, while all other members shared the same low frequency.

Training

Each word frequency condition was presented to 10 SRNs with 50 hidden units and trained on a next-word prediction task. Learning rate was 0·01, no momentum was used, and training was carried out to 140,000 sweeps. Thus, the network saw an equal number of word types and each corpus the same number of times, but was exposed to different frequencies of the same words.

Analysis. Analysis proceeded in identical fashion as in Experiment 1, with AP values and new word learning being measured at regular intervals.

RESULTS AND DISCUSSION

Figure 7 shows the AP values over training for the networks trained in each condition. There were no differences between AP values before and at 60,000 sweeps, nor at 140,000 sweeps. However, there were significant differences in AP values at 80,000 sweeps, $F(1, 144) = 920{\cdot}77$, $p < 0{\cdot}0001$, at 100,000 sweeps, $F(1, 144) = 926{\cdot}57$, $p < 0{\cdot}0001$, and at 120,000 sweeps, $F(1, 144) = 229{\cdot}93$, $p < 0{\cdot}0001$. These results suggest that there is an advantage for the network that sees words presented with uneven frequency, such that it enjoys an early boost that levels off, while the other network catches up. In other words, the token frequency manipulation did aid in initial category formation as predicted. (Note that, because of the limited

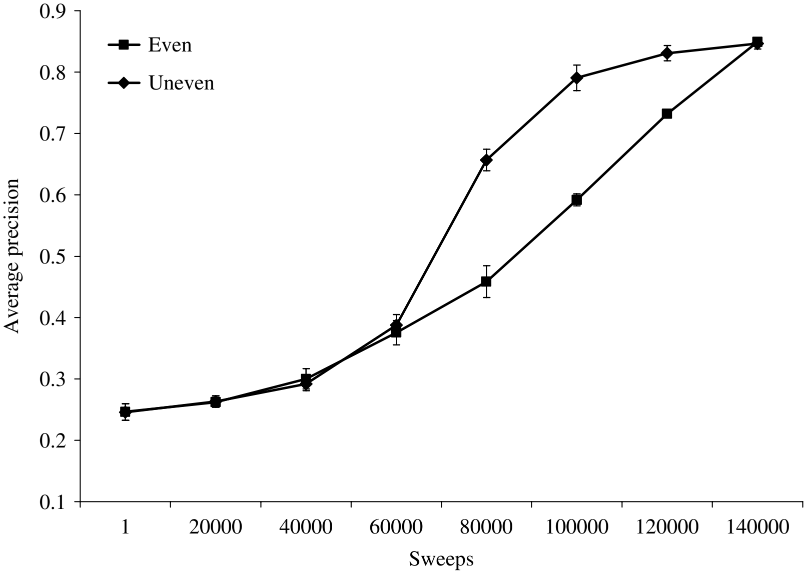


Fig. 7. Average precision values over training and frequency condition.

size of the vocabulary, there is a ceiling effect in performance such that all learning regimes converge on the same level of performance. A similar phenomenon is observed with growth curves from the CDI: learning appears to level off for all children. But this is because the CDI tests a fixed (and relatively small) set of words. Eventually, like our networks, all children learn the words in this limited set. Therefore, such growth curves are most informative in the middle regions, above the floor and below the ceiling of performance.)

Next, the network was probed for new word learning at 80,000 and 100,000 sweeps, where the largest difference in AP values was found. The results of this learning at 80,000, 100,000 and 140,000 sweeps can be seen in Figure 8. When measuring the rate of new word learning by taking the average node activations across the 50 sweeps of training, new noun learning for the uneven condition shows an advantage at 80,000 sweeps, $F(1, 312) = 34{\cdot}07$, $p < 0{\cdot}0001$, but not at 100,000 sweeps, $F(1, 312) = 0{\cdot}03$, *ns*. The result for 80,000 sweeps but not 100,000 falls in line with predictions that uneven token frequency in each word category should provide some sort of benefit in word learning that is tied to improvements in categorical structure. In order to examine word learning when there are no differences in AP values, Figure 8 also plots how well new nouns are being learned at the end of training at 140,000 sweeps, where no differences in AP values were observed. Examining this portion of the graph, the networks trained

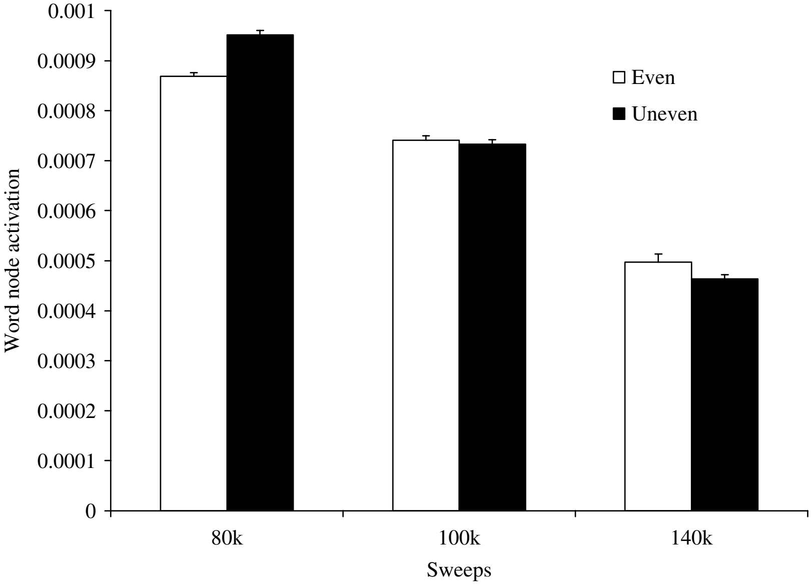


Fig. 8. Rate of word learning at 80k, 100k and 140k sweeps in both frequency conditions as measured by average word node activations over training.

with an evenly distributed word frequency have a significant learning advantage, $F(1, 312) = 5{\cdot}51$, $p < 0{\cdot}01$. Although there are no differences in AP value at this point, we do observe a difference in word learning in the Even condition. This seems to reverse the trend seen at 80,000 sweeps in training where the Uneven condition held the advantage in both word learning rates and AP values. It appears that as differences in AP values get smaller between the two frequency conditions, word learning in the Even condition improves.

In order to examine more closely how differences in AP values between the two conditions relate to differences in word learning rates, normalized differences of AP scores and word learning between networks with the same initial random weight setting (this is analogous to using the same human subject) at 80,000, 100,000 and 140,000 sweeps are plotted in Figure 9. Simple linear regression reveals a highly significant relationship between differences in AP and differences in word learning between the even and uneven condition $R^2 = 0{\cdot}85$ $F(1, 28) = 703{\cdot}58$, $p < 0{\cdot}0001$. This analysis suggests that even though we find differences between the two word frequency conditions in rates of new word learning at 140,000 sweeps when there are not differences in AP scores, and that we find no differences at 100,000 sweeps, even though there are differences in AP scores, there is still

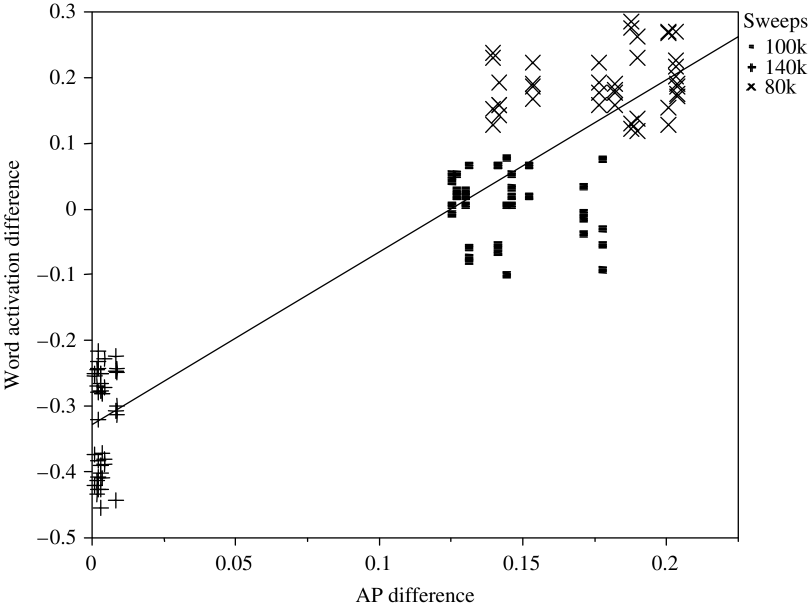


Fig. 9. Differences between even and uneven new word node activations plotted across difference in average precision values.

a very predictable relationship between changes in AP values and new word learning. This indicates that the overall relationship between the difference in learning rate and coherence in each condition, even though raw values might fluctuate. Here, our regression model reveals a very highly significant and positive relationship between differences in AP and subsequent differences in new word learning, such that when conditions improve new word learning ability in a particular network, we can also expect improved AP scores.

Finally, we turn to a third way in which input conditions might differ and ask what the effect of grammatical complexity might be on new word learning. The literature on this point reports mixed findings, perhaps because different measures of grammatical complexity are used in different studies. Recognizing that there are many dimensions along which such complexity might be defined, we begin with a very straightforward manipulation. We will define complexity in terms of the number of different arguments that are involved in a construction. The grammatically simple condition will involve only transitive and intransitive constructions (NVN and NV); the grammatically complex condition will additionally include ditransitive and sentential complement constructions (NVNN, NVNV, NVNVN).

EXPERIMENT 3: EFFECT OF GRAMMATICAL COMPLEXITY

METHOD

Input

Two 1000 sentence corpora were constructed, one for the Simple condition and one for the Complex condition. The Simple corpus was constructed with the same characteristics as the 1000 sentence corpus in Experiment 1. The Complex corpus was constructed with 10 additional verbs that belonged to two new verb categories: PSYCH (5) and TRANSFER (5). Semantic and syntactic relations between these two verb categories are included in Appendix 2. Examples of sentences containing verbs in these categories are also included in Appendix 3. No new nouns were added. Thus, the makeup of tokens in the complex grammar contained 95 total words (52 nouns and 43 verbs; the network architecture was adjusted to reflect the larger input and output vectors).

The new verb categories allowed for additional syntactic complexity by allowing for three more complex constructions to be added to the already present NV and NVN constructions. TRANSFER verbs allowed for ditransitive constructions of the form: NVNN. PSYCH verbs allowed for NVNV or NVNVN) constructions, with PSYCH verbs only occurring in the first verb position in these sentences.

Training

Each of the two corpora were presented to 10 SRNs with 50 hidden units and trained on a next-word prediction task. Learning rate was 0·01, no momentum was used, and training was carried out to 140,000 sweeps. Thus, the network saw an equal number of word tokens. Because the average length of sentences was different between each condition, each corpus was not seen the same number of times.

Analysis

Analysis proceeded in identical fashion as in Experiments 1 and 2, with AP values and new word learning being measured at regular intervals.

RESULTS AND DISCUSSION

Figure 10 shows the AP values over training in networks with the syntactically simple and complex input. The graph shows that the simple input has significantly higher AP values at 80,000 sweeps, $F(1, 18) = 195{\cdot}02$, $p < 0{\cdot}0001$, 100,000 sweeps, $F(1, 18) = 674{\cdot}05$, $p < 0{\cdot}0001$, 120,000 sweeps, $F(1, 18) = 628{\cdot}42$, $p < 0{\cdot}0001$, and 140,000 sweeps, $F(1, 18) = 637{\cdot}84$,

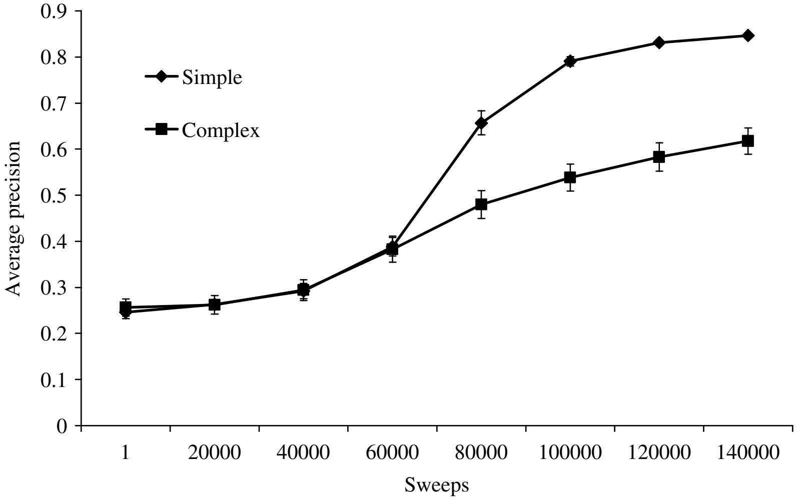


Fig. 10. Average precision values across training in simple and complex input conditions.

$p<0{\cdot}0001$. AP is equivalent between each syntactic condition earlier. These findings suggest that simpler grammatical constructions do indeed aid in early categorical formation, because simpler syntax in this manipulation displayed higher AP values after training.

Figure 11 relates these findings to new word learning. This graph shows how well the networks trained with each kind of input learned new words after 140,000 sweeps. Here, the networks trained with simpler input indicate stronger activation to predict new nouns. The rate of new learning was then assessed by taking the average activation value of the word node to predict the new noun from the initial value. We find that there is also a significant difference between new word learning with simple syntax learning showing higher rates of word learning than more complex syntax, $F(1,78)=130{\cdot}41$, $p<0{\cdot}0001$. These results are consistent both with the hypothesis that simpler grammar should improve lexical acquisition, and that higher category coherence also predicts better word learning.

GENERAL DISCUSSION

These three experiments tested two predictions that follow from the hypothesis that categorization is used as a tool in lexical acquisition.

(1) Variation in language input that affects lexical acquisition also affects development of category coherence in a similar manner; and
(2) It is the development of categories themselves, and not solely language input, that facilitates lexical acquisition.

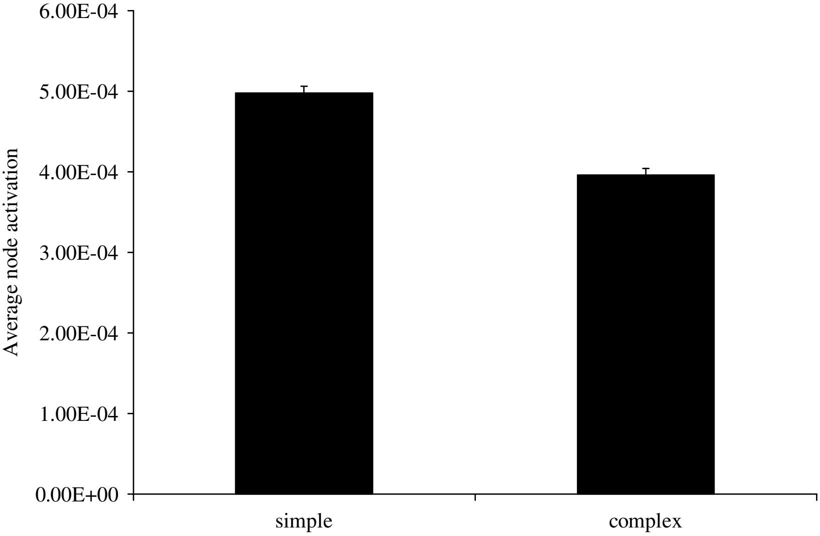


Fig. 11. Rate of word learning across complexity of input.

The results of these three experiments support these predictions. All three show direct relationships between input condition and corresponding improvements in lexical acquisition and category coherence.

A very clear illustration of this relationship is demonstrated in Figure 6, from Experiment 1, where increasing amounts of input positively influenced both category coherence and word learning. This supports the idea that there is a relationship between category development and speed of word learning. However, it is still not clear from these results if this relationship is of the nature as described in our second prediction, or is driven by language input driving both factors independently, but in coincidentally similar ways.

A key finding that addresses this issue comes from Experiment 2, where the frequency of the words in each category was manipulated. Here, while the nature of the input remained constant – in terms of the richness of the vocabulary – we see a direct relationship between improvements in category and lexical development. This suggests that lexical acquisition is affected by changes that the input has on category structure.

Word learning mechanisms

Much of the work that has been devoted to explaining word learning focuses on the dramatic changes in rate of lexical acquisition that occur during the middle of the second year of life. Two basic types of theories have been proposed. One class of theories hypothesizes that language specific constraints and principles appear during the vocabulary spurt and have the effect of improving the ability of children to acquire new words. These accounts have tended to emphasize the language-specific nature of the constraints involved in word learning (Woodward & Markman, 1998; Markman, 1989), although it is also possible that these domain-specific word learning constraints may arise from more domain-general processes, as in the 'lexical principles' model (Golinkoff, Mervis & Hirsh-Pasek, 1994).

An alternative to the constraints and principles view is one that words are learned through domain-general processes that make note of statistical associations between words and other properties that accompany their referents. These theories are entirely compatible with a view that other cognitive processes like categorization may aid in knowledge generalization from similar words in the same category. The simulations we report here illustrate how such a mechanism might work. The simulations demonstrate that a single learning mechanism that does not change over the course of development can account for a number of ways in which input differences seem able to influence word learning ability. The effects are mediated by category knowledge; interestingly, the development of these categories can be manipulated not only by quantitative variations in the input, but also by differences in grammatical complexity and frequency distributions across the input vocabulary.

These results are thus consistent with Gopnik & Meltzoff's (1993) account of the use of categorization as a tool in learning language as a 'complex bi-directional interaction'. We find a direct inter-relationship between improvement in category scores and word learning, when starting from clean conceptual slate. By training neural networks, we were able to examine how linguistic input might serve to carve out categorical space in the absence of any other kind of perceptual input, and how this categorical space in turn improves proficiency in lexical acquisition. In this way, we find support for a bidirectional influence of language and categorical development. Through semantic information that is encoded solely in language, it is possible to find relationships between improvements in linguistic ability and category development in the absence of perceptual input.

Overall, we have found that a single domain-general mechanism can account for a number of patterns in child word learning. Initially, our networks show very low category coherence, as it learns about individual items. This pattern is also observed in children, where analysis of early vocabulary shows that children tend to learn about basic level words before superordinate or subordinate items (Mervis, 1983). Category coherence improves because the neural networks eventually learn to categorize word items that are more similar to each other. More input aids in this process, by allowing the network to have a larger variety of experience and examples in which it may use to more appropriately classify items into groups.

Simpler syntax is useful in allowing the network to process simpler relations that are easier to understand and more easily grouped. Also, having a highly frequent exemplar in each noun category was useful to allow the networks to understand one example very well in a variety of contexts, thereby allowing other members of the same category to be organized more easily into a group. Word learning was improved in cases of better developed category coherence, because the networks were able to generalize from its knowledge of other members in the category to new words seen in similar contexts.

Role of input

The role of language input is highly emphasized in this paper. However, it is important to acknowledge several ways in which the experience of the networks differs from that of children. First, in these simulations we are limited to using a simplified artificial language that does not represent the full richness and complexity of natural language. Undoubtedly there are many other factors in CDS that may affect the outcome of this study. For instance, earlier disagreement about the role of structural complexity in language might actually be related to a connection in increasing complexity of CDS with age. Hoff & Naigles (2002) found that increased syntactic complexity in CDS was a better predictor of vocabulary size in 2;0 year olds. On the other hand, Brent & Siskind's (2001) study suggest that simpler complexity is beneficial was achieved with children at 0;9 to 1;3. It could be that increased complexity is beneficial for older children, like those in Hoff & Naigles (2002). A similar sort of result is reported by Elman (1993), where networks trained on input complexity that was incrementally increased in complexity demonstrated better learning than starting off with complex input initially.

Second, the input used in this study completely ignores the role that other social and perceptual cues may play in language learning. We have no doubt that input in other forms and modalities is crucial in both language and cognitive development. In fact, there is evidence that around the time of the vocabulary spurt, children seem to be making important developments in their social abilities as well. Thus, we emphasize that we do not believe that information from verbal input is the sole determinant of conceptual development.

Nonetheless, it is a striking result that the linguistic input alone – absent the experiential input available to real children – is such a rich source of information about category structure. Furthermore, we suspect that the additional information available to children will be learned through a similar kind of association mechanism that organizes experience based upon similarity. Essentially, input comes in many forms, but the underlying principles for learning is the same for all of them.

CONCLUSIONS

It is clear from these experiments that there are potentially many factors in CDS that may influence both a child's ability to learn words and her categorical knowledge. Here it is implied that disadvantages that arise from deficient input will persist due to influences that remain from the cognitive implications of this deficit. However, by identifying at least one underlying cognitive factor that is affected by change in language input, it is possible that interventions may be designed for children in normally impoverished linguistic environments. For instance, Hart & Risley (1995) mention somewhat pessimistically that differences in the amount of language input that are related to socioeconomic status are so large that it would require a continuous intervention of 40 hours per week to make up for 'lost input'. Perhaps focused training on category development may boost word learning ability in these children that could at least partially make up for deficiencies in language experience by aiding them to make the most efficient use of language that they do hear. This may also apply to cases where children may have difficulty learning language, or where there may be a delay of input due to conditions like deafness. Indeed studies by Smith, Jones, Landau, Gershkoff-Stowe & Samuelson (2002) have shown that training children to attend to shape rather than texture of objects boosts the number of object words known several months later. It is still left to be determined if this kind of approach could also apply to other types of categories, and other aspects of input. These are all important questions that merit further study.

In conclusion, while there are a number of limitations to this study, both in its simplification of language input and in the exclusion of other forms of perceptual input, the results we have found still are able to account for several patterns in language acquisition in children. Indeed, our results do mimic trends that we have already seen in the developmental literature. For instance, we replicated findings by Huttenlocher and colleagues (1991) and Hart & Risley (1995), that less exposure to linguistic input seems to put children at a disadvantage in language knowledge and learning. If anything, this type of replication suggests that the kinds of phenomena we have measured in our simulations, should also relate to learning in children. Most importantly, overall, this work provides a clearer understanding of a possible mechanism by which the development of category knowledge and word learning may be related.

REFERENCES

Bates, E., Bretherton, I. & Snyder, L. (1988). *From first words to grammar: individual differences and dissociable mechanisms*. New York: Cambridge University Press.

Bowerman, M. (1996). Learning how to structure space for language. In P. Bloom, M. A. Peterson, L. Nadel & M. F Garrett (eds), *Language and space*. Cambridge, MA: MIT Press.

Brent, M & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition* **81**, B33–B44.

Broen, P. A. (1972). The verbal environment of the English-learning child. ASHA Monographs, 17, Washington, D.C.: American Speech and Hearing Association.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* **10**, 425–55.

Cameron-Faulkner, T., Lieven, E. & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science* **27**, 843–73.

Casenhiser, D. & Goldberg, A. E. (2005). Fast mapping of a phrasal form and meaning. *Developmental Science* **8**, 500–8.

Choi, S. & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition* **41**, 83–121.

Choi, S., McDonough, L., Bowerman, M. & Mandler, J. (1999). Early sensitivity to language-specific spatial terms in English and Korean. *Cognitive Developmen*, **14**, 241–68.

Chomsky, Noam A. (1981). *Lectures on Government and Binding*. Dordrecht, Holland: Foris Publications.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science* **14**, 179–211.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition* **48**(1), 71–99.

Elman, J. (1998). Generalization, simple recurrent networks and the emergence of structure. In M. A. Gernsbacher & S. Derry (eds), *Proceedings of the 20th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum, 1998.

Fenson, L., Dale, P. S., Resnick, J. S., Bates, E., Thal, D. J. & Pethick, S. J. (1994). Variability in early communicative development. *Monograph of the Society for Research in Child Development* **59**, 174–9 (serial no. 242).

Fodor, J. L. (1975). *The language of thought*. Harvard University Press.

Goldberg, A. E., Casenhiser, D. & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics* **14**, 289–316.

Golinkoff, R. M., Mervis, C. & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework. *Journal of Child Language* **21**, 125–55.

Gopnik, A., Choi, S. & Baumberger, T. (1996). Crosslinguistic differences in semantic and cognitive development. *Cognitive Development* **11**(2) 197–227.

Gopnik, A. & Meltzoff, A. N. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child development* **58**, 1523–31.

Gopnik, A. & Meltzoff, A. N. (1993). Words and thoughts in infancy: the specificity hypothesis and categorization and naming. In C. Rovee-Collier & L. Lipsitt (eds), *Advances in infancy research*. New Jersey: Ablex.

Hart, B. & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H Brookes Publishing CO, 1995.

Hoff, E. & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development* **73**(2), 418–33.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early vocabulary growth: relation to language input andgender. *Developmental Psychology* **27**(2),236–48.

Keibel, H., Elman, J. L., Lieven, E. & Tomasello, M. (submitted). From words to categories: distributional regularities in German child-directed speech.

Li, P., Farkas, I. & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks* **17**, 1345–62.

Mandler, J. M. (1996). Preverbal representation and language. In P. Bloom, M. Peterson, L. Nadel & M. Garrett (eds), *Language and space*. Cambridge, MA: MIT Press.

Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

Mervis, C. (1983). Acquisition of a lexicon. *Contemporary educational psychology* **8**, 210–36.

Nelson, K. (1973). *Structure and Strategy in Learning to Talk*. No. 149 in Monographs of the Society for Research in Child Development. University of Chicago Press: Chicago.

Newport, E. L. (1977). Motherese: the speech of mothers to young children. In N. J. Castellan, D. B. Pisoni & G. R. Potts (eds), *Cognitive theory* (vol. 2). Hillsdale, NJ: Erlbaum.

Pan, B. A., Rowe, M. L., Singer, J. D. & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development* **76**(4), 763–82.

Pinker, S. (1991). Rules of Language. *Science* **253**, 530–5.

Plunket, K. & Elman, J. L. (1997). *A handbook for connectionist simulations.* Cambridge, MA: MIT Press.

Plunkett, K., Sinha, C., Moller, M. F. & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science* **4**, 293–312.

Rohde, D. L. T. (1999). The simple language generator: encoding complex languages with simple grammars. Technical Report, CMU-CS-99-123. Carnegie Mellon, Department of Computer Science, Pittsburgh, PA.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L. & Samuelson, S. (2002). Early noun learning provides on-the-job training for attention. *Psychological Science* **13**, 13–19.

Snow, C. & Ferguson, C. (eds) (1977). *Talking to children: language input and acquisition.* Cambridge: CUP.

Weizman, Z. O. & Snow, C. E. (2001). Lexical input as related to children's vocabulary acquisition: effects of sophisticated exposure and support for meaning. *Developmental Psychology* **37**(2), 263–79.

Wells, C. G. (1981). *Learning through interaction: the study of language development.* Cambridge: CUP.

Whorf, B. L. (1956). *Language, Thought and Reality* (ed. J. B. Carroll). Cambridge, MA: MIT Press.

Woodward, A. L. & Markman, E. M. (1998). Early word learning. In D. Kuhn & R. S. Siegler (eds), *handbook of child psychology: Vol. 2. Cognition, perception and language* (pp. 371–420). New York: John Wiley & Sons.

Zavrel, J. (1996). *Lexical space: Learning and using continuous linguistic representations.* Unpublished doctoral dissertation, Utrecht University, Utrecht, Netherlands.

APPENDIX 1

AVERAGE PRECISION

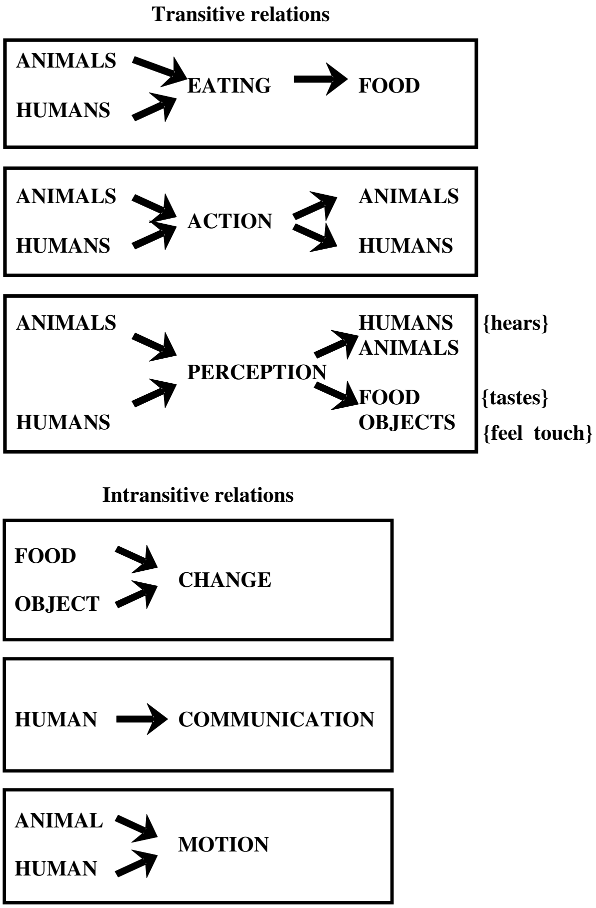
Average precision is calculated in several steps. First, the Euclidean distance between each word's hidden unit vector and every other word is calculated. This is to find the pair-wise distance between each word and every other in representation space. The values are then ranked, so that for each word the other words that have the most similar hidden unit representation are ranked closer than those that are not. Then, the average precision of each word is calculated with the following formula:

$$P(w) = \frac{1}{|C_w|} \sum_{i \in C_w} \frac{|n_{wi}(C_w)|}{|n_{wi}(all)|}$$

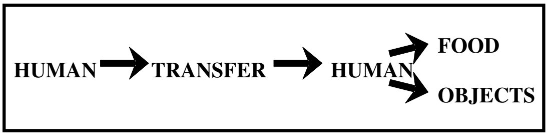
Where $P(w)$ stands for the average precision of one word, C_w is the number of words in the category, n_{wi} is the rank number of a particular word, and $n_{wi}(C)$ is the number of words in the target words category that have appeared before the particular rank. Simply, this algorithm calculates the proportion of words that belong to a target word's category at each rank, and then divides this proportion by the number of words in the category. In a best case scenario, where all the words in a category are the closest ranked members, this would yield a value of one. Theoretically, these values can approach zero, where all within category words are infinitely far away from the target word. However, in practice, this result is difficult to achieve, so normally, just by chance, average precision values will hover around 0·2 without any real structure.

APPENDIX 2

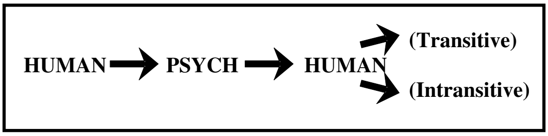
SEMANTIC RELATIONS BETWEEN CATEGORIES



Ditransitive relations



Matrix relations



APPENDIX 3

EXAMPLES OF SENTENCES USED IN THE STUDY

Transitive sentences

EATING:

kid gobbles pizza.
bird drinks water.

PERCEPTION:

rabbit sees bread.
grandmother touches book.

ACTION:

animal bites puppy.
teacher hugs lamb.

Intransitive sentences

CHANGE:

cake falls.
box breaks.

COMMUNICATION:

boy talks.
kid laughs.

MOTION:

tiger moves.
mother jumps.

Ditransitive sentences

TRANSFER

mother buys brother cup.
brother offers grandfather box.

Matrix sentences

PSYCH

uncle wants grandmother sees duck.
grandfather convinces brother sits.