WHY IT IS HARD TO LABEL OUR CONCEPTS

Jesse Snedeker and Lila Gleitman

Harvard University and University of Pennsylvania

(To appear in Hall & Waxman (eds.), *Weaving a Lexicon*. Cambridge, MA: MIT Press.)

Jesse Snedeker
Harvard University
1136 WJH, 33 Kirkland St.
Cambridge MA 02138
Telephone: 617-495-3873 FAX: 617-
snedeker@wjh.harvard.edu

> But if the knowledge which we acquired before birth was lost by us at birth, and afterwards by the use of the senses we recovered that which we previously knew, will not that which we call learning be a process of recovering our knowledge, and may not this be rightly termed recollection?
>
> --Plato, *Phaedo (c.a.* 412BCE)

One way to think about infants before they acquire their native tongue is as second-language (or reincarnated) learners whose task is only to find out how priorly known concepts, at the level of both words and sentences, are mapped onto linguistic forms. Life for these Platonic infants would be much easier than that faced by Quine's linguist, trapped in an exotic land where not only pronunciations but, simultaneously, the modes of thought, may be different from one's own. We ask here which of these models is closer to the real case, concentrating attention on the task of acquiring a first lexicon.

The test bed for this question as raised in this chapter concerns the changing character of child vocabularies in the first three years of life. Specifically, the first vocabulary (the first 100 words or so), in various languages and under marvelously varying child-rearing conditions, massively over-represents the noun lexical class, compared to its frequency in the input corpus, and massively under-represents the verbs. A second widely observed property is that those verbs that do appear at the earliest stages of child speech and comprehension are largely limited to action terms such as *go, run,* and *throw* even though verbs like *think* and *look* are frequent in maternal speech. What accounts for such input-output disparities? Speaking very broadly, we can distinguish two general kinds of explanations for these systematic asynchronies in aspects of vocabulary development. The first of these emphasizes children's cognitive development. The alternative, which we will try to defend in this chapter, emphasizes linguistic development. The burden in maintaining this latter stance is to understand why, all the same, the abstract words are manifestly harder to acquire than the concrete ones.

Cognitive development and vocabulary learning

Plausibly enough, the changing character of vocabularies over developmental time is usually taken to reflect the changing character of the child's conceptual life. This kind of hypothesis comes in two main flavors, depending on whether the causal machinery for cognitive change is taken to be maturational or experiential. But in either case, the changes in the function over time are assigned to changes of mentality within the learner. Smiley and Huttenlocher (1995) present a particularly perspicuous version of this view:

> ...Even a very few uses may enable the child to learn words if a particular concept is accessible. Conversely, even highly frequent and salient words may not be learned if the child is not yet capable of forming the concepts they encode. These two cases, in which effects of input frequency and salience are weak, suggest that conceptual development exerts strong enabling or limiting effects, respectively, on which words are acquired. (p. 24).

The claim is that children fail to learn certain words despite the fact that they are often uttered in their presence just because the concepts they encode are beyond their grasp. Indeed, the word-learning facts are often used as more-or-less straightforward indices of concept attainment (e.g., Huttenlocher, Smiley, & Charney, 1983; Dromi, 1987; Gopnik & Meltzoff, 1997). Two cases in which cognitive development is posited as an explanatory mechanism are relevant to the findings we will present in this chapter.

Nouns before verbs

A first case that appears to lend itself particularly well to explanation in terms of conceptual change has to do with the temporal priority of *noun-learning* over *verb-learning*, a phenomenon that is apparently robust both to architectural distinctions between languages and to input differences that are correlated with linguistic distinctions (for a masterful review, Gentner & Boroditsky, 2001): Even though children hear both verbs and nouns from earliest infancy, their earliest vocabulary is overwhelmingly nominal (Goldin-Meadow, Seligman, & Gelman, 1976; Bates, Dale, & Thal, 1995), with only a very few true verbs[1]. Verbs are not seen in proportions similar to their proportion in the adult input until the child is about two and a half years old, or even older in many cases. It has been

---

1. In further detail, the finding is that nouns occupy a much larger proportion of the vocabulary than verbs than can be predicted from the input facts. When English speaking children's production vocabularies are between 20-50 words, 45% of these words are common nouns and only 3% are verbs (for Italian the corresponding percentages are 37% and 4%) Caselli, Bates, Casadio, Fenson, Fenson, Sanderl, & Weir, 1995). This is so even though the incidence of each word (that is, the token frequency) is higher for the verbs than for the nouns in the common set used by mothers (Sandhofer, Smith & Luo, 2000). These proportions both of types and of tokens are of course influenced by several cross-cutting factors having to do with the languages themselves and even with child rearing practices that differ cross-culturally (see Choi and Gopnik, 1995; Tardif, Naigles & Shatz, 1997). Still, early noun-dominance is visible cross-linguistically despite the countervailing influence of these additional variables.

conjectured that it is the typical object-labeling function of nouns vs. the relational functions of verbs that accounts for these learning facts: Necessarily the acquisition of concepts describing relations between objects would await acquisition of the object concepts themselves (O'Grady, 1987; Caselli, Casadio, & Bates, 1999).

The position, then, is that noun and verb learning occur *seriatim* as a straightforward reflex of the fact that object- and relational-type concepts become available *seriatim* to the infant mind. The littlest children just can't think about relations, or at least can't easily think about them at the level requisite to word learning, so they can't learn the words that express such concepts.

Concreteness of Verbs

A second linguistic-developmental finding provides further grist for the maturational account. The earliest verbs are not an unbiased sample of those that appear frequently in the input. Children all over the world say words like *throw* and *run* well in advance of *think* and *know.* The first verbs that children produce describe actions or movement events and encode the physical motion of the agent (Bloom, Lightblown, & Hood, 1975) and these verbs seem to be understood earlier or better as well (Huttenlocher, et al, 1983; Gentner, 1978). Mental state terms are not used as such until at least two and a half years of age (Shatz, Wellman, & Silber, 1983; Furrow, Moore, Davidge & Chiasson, 1992) and are not fully distinguished from one another in comprehension until around age 4 (see e.g., Johnson & Maratsos, 1977; Moore, Gilbert, & Sapp, 1995).[2] Their absence in early vocabularies is often taken as evidence that the child doesn't have control of the relevant concepts. As Gopnik and Meltzoff (1997) put this:

> ...the emergence of belief words like "know" and "think" during the fourth year of life, after "see" is well established. In this case...changes in the children's spontaneous extensions of these terms parallel changes in their predictions and explanations. The developing theory of mind is apparent both in semantic change and in conceptual change. (p. 121)

Summarizing, the position just sketched is that the word-learning function

---

2. There is some disagreement about the timing of situationally appropriate production of mental state terms, with some investigators seeing evidence of this development well before age three (Shatz et al., 1983; Bretherton & Beeghly, 1982). There are also substantial differences within the class of mental state predicates. In spontaneous production predicates encoding desire, perception, and affect begin appearing around 2;0 about a year before predicates for cognition (Bartsch & Wellman, 1995; Bretherton & Beeghly, 1982). But there is little disagreement that concrete action verbs are spoken and understood in advance of the mental state verbs.

follows from, and more or less directly reflects, the concept-acquisition function. At the limit, no one could dispute the relevance of conceptual organization to word learning. One cannot acquire productive use of a term that expresses a concept that one cannot entertain. Babies couldn't learn *casuistry* and *transmigration* until steeped in the appropriate theological traditions, and nomads for related reasons might have trouble with *picnic*.[3] But whether these limitations are imposed by the mentality of the learner rather than ("merely") by his state of world-knowledge is open to question.

Linguistic development and the vocabulary attainment sequence

A number of influential proposals relate word learning to biases and abilities that represent stages in the development of language rather than, or in addition to, development in the general mental capacities of the learner. Particularly, Markman and her colleagues (for a review, Markman, 1994) and Clark (1987) have experimentally documented biases that do not characterize learners generally, but which become operative in the specific context of word learning. Well-known examples are the whole object bias, the principle of mutual exclusivity (or "no-synonym" constraint), and the taxonomic constraint.

Constraints of this kind provide a principled account for the noun-dominance effect in early child language. Biases such as the whole-object constraint, which assist in the acquisition of nouns, could make it more difficult to learn other types of words (Kuczaj, 1990). If the child initially attempts to map every label to an object, verb learning will be delayed until enough evidence accumulates to override this default hypothesis. By positing a hierarchy of constraints which prioritize the child's hypotheses about word learning, it might be possible to extend this explanation to account for within-class sequences as well, including the order in which different types of verbs are acquired.

Gentner (1982) offers another explanation of the noun-dominance effect that is rooted in the process of language learning. She suggests that, while babies may be equally facile with object and action categorization, they may face a harder task in discovering the extensions of verbs than of nouns. According to her view, the key is in the universality, or lack of it, in the mapping relations between words and concepts from language to language. By and large, languages accomplish object-description in the same way, namely with nouns, but differ in the semantic components that they characteristically encode within the verb (Talmy, 1985). To the extent that this is so, one could envisage a learning procedure predisposed by nature for the object-to-nominal mapping. In contrast, acquisition of the first

3. Thanks to Dan Reisberg for this example.

verbs would have to await discovery of the language-particular conflation patterns.[4]

Note, however, that this proposal leaves us in the dark about several of the gross facts about early vocabulary growth. Choi and Bowerman's work (1991) on cross-cutting lexicalization patterns for motion events in English and Korean suggests that language-specific conflation patterns do not always pose a particularly difficult problem for the young language learner. Moreover, the conflation-pattern solution will not extend to the timing of "concrete" (or *action)* versus "abstract" (or *mental*) verbs since, as far as we know, there is no reason to predict that the cross-linguistic differences in conflation patterns should be more complex for more abstract verbs.

The present proposal

The proposal we evaluate in the present chapter also places the origin of vocabulary change in linguistic rather than conceptual development. The young child's conceptual repertoire may be rich and varied enough from the start to accommodate the commonly heard words, including *cat, throw*, and *think*. All the same, different kinds of information must be brought to bear for solving the mapping problem for these words, not all of them available from the outset. Specifically, infants must initially learn the meanings of words by inspecting the extralinguistic contexts (or situations) in which the word is used. After all, linguistic novices can glean little from the utterance in which a word occurs for the simple reason that they haven't yet learned what these other words mean. This initial word-to-world pairing procedure will suffice to identify a small set of concrete nominal terms. To learn that *cat* is the English-language word for the concept 'cat', the child need only note that cats are the objects most systematically present in scenes wherein the sound /kat/ is uttered (just as proposed by Augustine, 398; Locke, 1690; Pinker, 1984, and many other commentators).[5] However, this situational evidence, taken alone, may be

---

4. This story is complicated by the fact that even within a language the conflation patterns for individual words vary considerably. While many common English motion verbs encode manner (*walk* or *run*), the most common ones encode deixis instead (*go*, *come*), and a rare few encode path (*enter*, *climb*). A contact event can be described by a verb encoding manner of action (*pound*), instrument (*hammer*), or result (*hit*). Of course, this kind of componential analysis does not necessarily enter into the learning process in the first place (for discussion, see Carey, 1978; Fodor, 1998).

5. We presuppose here the lesson from Baldwin (1993) that all such generalizations are modulo the child's determination that, in using a novel word in a particular setting, the adult means to refer to that setting by that word use. For example a child playing with a new toy and hearing his mother say "Wow, look at the blitso!" will learn that "blitso" names that toy only if the mother

insufficient for the mapping of most of the vocabulary. For example, words like *think, idea*, *bachelor* and *pet* seem to lack obvious and stable observational correlates (and see Plato on *virtue*). Then so long as inspection of extralinguistic context is the only learning tool available, learners should be able to acquire only a special and very limited kind of words: those that are both straightforwardly available to perception and lexically "favored," perhaps in Markman's sense.

We will demonstrate that the vocabulary could become rich and diversified owing to a succession of bootstrapping operations, *grounded in the prior acquisition of concrete nominals*, that build the words in tandem with the clause-level syntax. In this new informational state, the mapping problem is solved by the child's joint consideration of the extralinguistic and linguistic-structural contingencies for word use (a learning procedure we will be calling *structure-to-world pairing).* Acquisition of the lexicon and acquisition of the grammar, are, in these regards, parts of a single underlying process. In earlier incarnations, the structure-to-world pairing approach was called "syntactic bootstrapping (Landau and Gleitman, 1985; Gleitman, 1990; Fisher, Gleitman, and Gleitman, 1991; Gleitman & Gleitman, 1997). However, under that title the position was misunderstood as refering exlusively to syntax-based cues to word learning. Then as now, however, our approach emphasizes the fact that learners beyond the very earliest rudiments of language discovery use multiple converging cues to acquire the vocabulary.[6]

An experimental analysis: The human simulation paradigm

The findings we review here are from experimentation in which adult subjects try to identify words from partial information about the contexts in which they occur (see also Snedeker, Brent, and Gleitman, 1999; Gillette, Gleitman, Gleitman, and Lederer, 1999; and Snedeker, 2000). Conceptually, these experiments are analogous to computer simulations in which a device, endowed with whatever ("innate") ideas and learning procedures its makers deem desirable to program into it, is exposed to data of the kind naturally received by the target learner it is simulating. The measure of success of the simulation is how faithfully it reproduces the learning function for that target using these authentic data.

Our test device is a population of undergraduates (hence *Human* Simulations). Their preprogramming includes, *inter alia,* knowledge of English. Their task is to

---

was attending to it while so saying. For the moment, we leave aside the usual skeptical objections ala Quine.

6. For analyses that emphasize the incremental nature of this procedure, see Fisher, Hall, Rakowitz and Gleitman (1994) and Gillette, Gleitman, Gleitman, and Lederer (1999).

identify words they already know when these are disguised as nonsense words, beeps, or tones. The data they receive are contextualized mother-to-child speech events, transcripts or videotapes of the situations in which a word was used. The form in which they receive information about these speech events is manipulated across conditions of the experiment.

These experiments serve two purposes. The first is to provide an estimate of the psychological potency of various cues to word meaning that are available in the real learning situation. What kinds of words can be identified by inspection of the contingencies for their use in the absence of all other cues, assuming a conceptually mature learner? The second is to estimate by reference to these outcomes something about the word- and structure-learning procedures used by children.

Restating, we attempt to reproduce in adults the learning function of the one- and two-year old child by appropriate changes in the information structure of the input. If successful, this exercise makes plausible that the order of events in child vocabulary acquisition (here, the developmental move from nominal categories to predicate categories, and from less to more abstract verbs) is assignable to information-structure developments rather than to cognitive developments in the learner, for we have removed such possible cognitive inadequacies from the equation. When our college-age subjects fail to learn, it is not because they have some mental deficit that disbars them from attaining words like *ball* or *get*. In fact, since the test items are the most common nouns and verbs uttered by mothers in conversation with their language-learning infants, we know that these words are already in our subjects' vocabularies. All they have to do is to recover what they previously knew. Our first experimental probe asked how efficiently they can do so by using the evidence of the senses.

Simulating word-to-world pairing

The stimuli for these experiments were generated by Gillette et al. (1999) who videotaped mothers interacting with their 18 to 24-month old children in an unstructured situation. The maternal speech was transcribed to find the 24 most frequent nouns and the 24 most frequent verbs that these mothers uttered during these taped sessions. To simulate a condition under which learners were presumed able only to identify recurrences of the same word (*qua* sound) in the speech stream and to match these with their extralinguistic contexts of use, they selected 6 video clips during which the mother was uttering each of these words. Each video clip started about 30 seconds before the mother uttered the word, and ended 10 seconds afterwards. A pilot study indicated that this was to enough to give the observer the gist of the mother-child interaction.

There were 6 clips for each "mystery word" these subjects were to identify. This simulates the fact that real learners aren't forced to acquire meanings from a single encounter with a word and its context. Rather, by examining the use of the word in a variety of contexts, the observer can attempt to parse out those properties of the world common to all these encounters. The 6 video-clips for the word were then spliced together with brief pauses between them. The subjects were told for each word whether it would be a noun or a verb, but *they did not hear what the mother was saying, because the audio had been removed*. Instead of the original sound-track a single audible beep occurred at the very instant during the depicted event when the mother had actually uttered the mystery word. Subjects wrote down their guess after viewing each of the 6 clips for a word; that is, as cross-situational evidence accumulated, then they were asked to make a final conjecture based on all the evidence.

Notice that this procedure cannot tell us whether a word can or cannot be acquired from extralinguistic contexts, even if word-learning mechanisms do not vary between adults and children and even if (as we suppose) the vocabulary acquisition procedure is largely a mapping problem. This is because, though the word learning function in children is clearly efficient and rapid (e.g., Carey, 1978; Caselli et al., 1995), perhaps, for all we know, 7 or 40 or 900 exposures to new words are ordinarily required. What we can consider realistically is the *relative* efficiency of noun versus verb learning under this information condition.

The results of this procedure were very dramatic. When the subjects were limited to the information of their senses, that is, when they had to acquire the mappings of sounds (here, beeps) onto meanings solely via inspection of the extralinguistic contingencies for their use, they were able to identify 45% of the nouns but only 15% of the verbs, a difference of great magnitude and statistical reliability. Like infants learning their first words, these adult subjects "acquired" a vocabulary dominated by nouns, and under-populated by verbs.

Further analyses are even more revealing of the dramatic noun/verb difference. Every noun target was correctly identified by at least one of the subjects who were exposed to it, whereas fully a third of the verbs were never correctly identified. In addition, the learning curve (correctness as a function of trial) is reliably steeper for nouns than for verbs. Finally, and most revealingly, subjects converge toward the correct answer for the nouns; the most frequent final guess was the correct target for over half of the nouns. In strong contrast, the most frequent final guess for the verbs was the correct target only a quarter of the time.

In sum, on the basis of accumulating cross-situational evidence, subjects identified nouns quite efficiently, and verbs very poorly. For many of the verbs the information provided by the scenes is not only inadequate, but actually misleading, leading subjects to settle on the same incorrect hypotheses.

Why should cross-situational observation have been so much more useful for noun identification than it was for verb identification? Gillette et al. (1999) showed that the noun dominance effect is itself an artifact of another distinction; namely one of imagability or concreteness which is closely correlated with the noun/verb distinction in the stimulus materials. Mothers frequently say words like *think* and *like* to their toddlers but they rarely say *thought* or *liking.* Both within and across lexical class, imagability ratings were highly correlated with number of subjects who identified a word. Nouns like *music* were less often identified than nouns like *drum;* and verbs like *want* were less often identified than verbs like *throw.* What these subjects' responses seem to be telling us is that, because they were limited to learning words from extralinguistic observation, they could only identify words whose real-world concomitants were stable and observable.[7]

This tautology ("only observables can be learned from observation") has to be taken very seriously in understanding the nature of infant word learning. Let us suppose that infants have adequate conceptual sophistication but are forced to deduce the meaning of terms solely by examining the world and attempting to decipher the intentions of those around them. That is, *suppose that the only problem in acquiring a mental lexicon is the mapping problem*, *with no concept acquisition required at all*. Then just because the verb repertoire in infant directed speech is systematically less concrete than the noun repertoire, as was shown, it should be less amenable to acquisition by word-to-world pairing.

As we next show, changing the information structure of the input removes -- in fact reverses--these effects of concreteness on the word-mapping function.

---

7. What can be "perceived," is among the most vexed of questions, which we do not pretend here to engage. There is considerable evidence that infants can, *pace* Hume, extract properties such as 'cause', 'animate agent,' and 'intentionality' from their observations of ongoing events (e.g., Woodward, 1998; and see Kellman and Arterberry, 1998 for an important general discussion and review). Moreover, young children do acquire terms like *look* and *see* despite the fact that their conversationally targeted objects are often occluded; even blind children rapidly become proficient with these terms (Landau and Gleitman, 1985). Yet we suppose no one will doubt that the conditions of applicability of the terms *think* and *know* are harder to extract from inspection of ongoing scenes than are *plane* and *elephant.*

The learning patterns for verbs

Smiley and Huttenlocher (1995) have shown strong timing differences in the appearance of verbs from different semantic domains in child vocabularies. For example, verbs that encode characteristic self-caused movements of an organism (e.g., *jumping*) are understood and produced earlier in life than verbs that encode change and require information about causes and goals (e.g., *opening*). And verbs that describe mental states and acts on average appear later still (Huttenlocher, et al., 1983; Shatz, et al., 1983; Gopnik and Meltzoff, 1997) despite the fact that verbs from all three classes are frequent in mother-to-child speech.

As we stated in introductory remarks, these timing effects have usually been assigned to the cognitive-representational limitations of very young children. But once again it may not be—or it may not be *only*—that these verbs are harder to acquire because the child lacks the conceptual wherewithal to represent their meanings. The same solution that applies to the acquisition differences across lexical classes (nouns before verbs) may apply within lexical class as well. It may be that the children just breaking into language lack the linguistic-learning tools for solving the mapping problem for abstract words. This seems plausible because, just as for the noun-verb learning function described earlier, the original learning machinery (word-to-world pairing) might yield adequate cues for learning which sound in English encodes *'kick'* and *'jump,'* but not for learning which sound encodes invisible mental acts like '*think.'* Indeed, the verbs with the highest identification scores in the Gillette study were *throw, come,* and *push,* and those with the lowest scores included all the mental-content verbs (*like, think, know, love*). Perhaps, just like our college sophomore subjects, infant learners have the concepts encoded in these abstract terms, but insufficient tools for discovering their encoding in English.

The experiment now reported attempted to adjudicate between these two kinds of explanation via an extension of the Human Simulation paradigm. Again adult subjects were asked to solve the mapping problem. But this time we varied the information available to them.[8] In one condition, as before, they had to identify words (that is, solve the mapping problem) on the basis of observed scenes. But

8. Some of these manipulations were also performed also in Lederer, Gleitman, and Gleitman, 1991, and reported in Gillette et al. (1999). However, in the current experiment, the procedural properties of the several information-condition manipulations were precisely matched, and all possible combinations of the information sources were tested. This allows us to make better-motivated comparisons of the potency of various information sources, and to directly test for interactions between information type and verb-subtype. In addition, the current experiment provides a more complete representation of the extralinguistic contexts of word use (longer color video clips, shot with a more mobile camera).

other subject groups were given alternative or additional information. These new information sources were attempts to simulate different ways the input might be represented at various points during language acquisition. For example, at earliest stages we assume learners are limited to word-to-world pairing. After all, as we just showed experimentally, that assumption is sufficient to explain why they learn many concrete nouns and few verbs. But later they ought to be able to use, as well, the statistically observable facts about words' frequent local neighbors (e.g., *the* near nouns but *-ed* near verbs, food words in construction with verbs like *eat*). These auxiliary sources of evidence, increasingly available as language experience grows, might broaden the class of identifiable verbs.

Eighty-four undergraduate students participated in this new experiment, 12 in each of experimental conditions, described below. The stimulus materials were taken from the same videotaped maternal speech as Gillette et al. (1999), but this time we focused only on the verbs, for these were the locus of difficulty in the scene-interpretation study. The 24 most frequent verbs in these maternal corpora were selected as the target items. As before, six instances of the use of each verb were culled from the corpora and presented to the subjects. It is important to keep in mind that while subject groups received different representations of this information (depending on which condition they were assigned to), each group was being asked to identify the *same* verbs, and based on the *same* 6 maternal utterances of that word.

Information conditions for verb learning

The 7 learning conditions were as follows:

Scenes Condition: This condition simply replicated the Gillette study. The extralinguistic contexts were presented on videotape. The original audio had been removed and a beep inserted at the point where the mother actually uttered the target word. Subjects attempted to identify the mystery word after each of the scenes and then offered his best guess in the light of all of them.

Nouns Condition: Here, subjects simply saw a written list of the nouns and pronouns that had been uttered in construction with each instance of the use of the verb. For example, the maternal utterance "Did you play with the elephant?" was presented to subjects in the Nouns Condition as: "elephant, you". These subjects did not see the scene in which this utterance occurred. This condition simulates a hypothetical stage of learning that assumes the prior learning of some frequent nouns, which can now -- even in the absence of knowledge of language-specific syntax -- support verb learning. For example, the child who hears food names repeatedly with some novel verb might make the canny guess that it meant something like *eat.* If the probabilities of this noun-verb association are stable enough, the learner could make use of such information even without knowing anything more refined about the structure of such sentences.

Frames Condition: Here, subjects were shown representations of the six maternal sentences as nonsense frames, much in the style of Lewis Carroll's "Jabberwocky," (1865). The closed-class words (e.g., determiners or prepositions) and morphemes were left intact but all the nouns, pronouns, adjectives, and verbs appeared as nonsense forms. For example, the sentence given above ("Did you play with the elephant?") was presented as: "Did er PILK with the ramermok?" This orderly distortion of the original sentences preserved systematic clues to the syntactic category of words while removing information about their identity. Because of the strong correlations between syntactic privileges of verbs and their semantic interpretations (see Levin, 1993 and Fisher, Gleitman, & Gleitman, 1991), it is plausible that learners once in possession of grammars could exploit this information to constrain their conjectures about verb semantics (for experimental evidence that young children do just this, Fisher, Hall, Rakowitz, & Gleitman, 1994; Naigles, Gleitman, & Gleitman, 1993; Naigles, 1996).

Scenes + Nouns Condition: Subjects were presented with both the video clip in which the verb was used and the list of nouns that it was used with.

Scenes + Frames Condition: Subjects in this condition were shown the video clip and the syntactic frame for each of the uses of the mystery verb.

Nouns + Frames Condition: This condition combined the two types of linguistic context, noun co-occurrence (*Nouns Condition)* and syntactic frame information (*Frames Condition),* simply by putting the nouns and pronouns words back into the nonsense frames. Note that this manipulation converts co-occurrence information to selectional information.

Full Information Condition: In this condition, we presented the videotaped scenes for each target, along with the sentences from the *Nouns + Frames Condition*.

Results

Information Sources for Verb Identification

Figure 1 presents the percent correct identification at each trial in each condition;

that is, the learning curves. Figure 2 illustrates the percentage correct on the final trial in each condition. As the figures suggest and as verified by statistical analysis, some of the information sources we tested were more potent than others. Higher potency is manifest as an increase in correct responses on all trials, a greater improvement across trials, and a higher proportion correct on the final trial.

Figure 1 and Figure 2 about here

By comparing the three basic cue types, we can test the relative potency of each information source in isolation. By the final trial, identification of verbs is reliably higher in the *Frames Condition* (38% correct) than in the *Nouns* (17%) or *Scenes* (18%) *Condition*. Thus, we replicate Gillette et al's finding that scenes provide little guidance for mapping common verbs to their meanings, and confirm previous analyses of the input which showed that frames are a powerful information source for verb learning (Fisher et al., 1991; Geyer, 1998; Lederer, Gleitman, & Gleitman, 1995; Kako, 1998). The results of comparing Final-Trial correctness scores on all cues and cue combinations yields a statistically reliable partial ordering of the potency of these different informational conditions:

(1) *Scenes, Nouns < S+N, Frames < S+F, N+F < Full Information*

As we shall later discuss, the rising rate of identification in our adult subjects under these different conditions of stimulus-representation has the potential to support an information-based theory for why the child's verb learning is initially slow, improves somewhat in the older toddler who has amassed statistical knowledge of the co-occurrence of verbs with particular types of nouns, and then makes a dramatic move upward in the middle of the third year of life, when the phrase-structure of the exposure language is well established (see also Lenneberg, 1967; Caselli et al, 1995). Intriguingly, the same information-change solution that can account for the learning differences between the noun and verb lexical classes appears to hold once again when we look at the acquisition function within the verb class. Recall that the results for the verbs as a class yielded the ordering of cue-type potency given in (1). But this pattern of performance varied widely from verb to verb. This variation across items may throw light on the internal course of verb learning in young children too.

Different verbs require different information sources for identification

The maternally-frequent verbs studied in these experiments fall into three classes that have been identified and discussed in the prior literature. These are the relatively concrete verbs that describe specific actions in and on the observable world (*fall, stand, turn, play, wait, hammer, push, throw, pop*), the more abstract mental-content verbs (*know, like, see, say, think, love, look, want*), and a third set, of what have been called light verbs (*come, do, get, go, have, make, put).* Light verbs are those that are relatively "bleached" of specific semantic content, but which combine with nouns in English predicate phrases to form a contentful whole (as, "make a bet" or "go skiing"). One might argue with a few of our placements of the test verbs in this trichotomy but we have settled for following the practice of prior commentators, especially with regard to the identification of the light verbs (Clark, 1996; Ninio, 1996; Goldberg, 1999).

Figure 3 presents the percent identification scores for items in these three verb classes for each of the three basic information types. As inspection of the Figure reveals, while neither scenes nor nouns are very powerful cues to verb identity, they are more effective for the action verbs than for the other two classes.[9] The pattern for the *Frames Condition* is sharply different. First, this is the condition that is overall most potent as a learning cue for verbs. Second, it is the mental and light verbs whose learning is chiefly supported by the structural information. This is hardly surprising in light of the discussion thus far. Mental events and states are hard to observe and thus are unlikely to be acquired via a stand-alone machinery of extralinguistic observation. As for light verbs, the less they mean (or the more general their meanings) the less informative their situational concomitants will be.

Figure 3

The third generalization that can be drawn from Figure 3 is that frames are quite uninformative for identifying action verbs. This final result comes as something of a surprise. After all, while concrete action verbs might be readily identifiable in terms of their real-world contexts, there is no *a priori* reason to expect that they would not also be well-differentiated by their syntactic contexts. But they are not. Action verbs were more likely to appear in intransitive or simple transitive frames. Because these frames are associated with a broad range of meanings they provide little constraint on the types of verbs that can appear in

9. Assigning *come* and *go* to the action class (as seems intuitively correct to us) rather than to the light-verb class (following Clark, 1996 and Goldberg, 1999) magnifies the difference between the verb classes still further. With this reclassification, the identification of action verbs increases from 27% to 29% while the identification of light verbs in the Scenes Condition drops from 18% to 8%.

them. Mental verbs and light verbs were more likely to appear with prepositional phrases, particles, or sentence complements. These frames are more selective and thus provide more information about the verbs that occur in them. This "functional load" or "entropy" analysis of differential frame-informativeness was shown in the context of maternal-speech patterns by Lederer et al. (1995; see also Goldberg, 1995; and Kako, 1998, for experimental documentation with adult subjects).

The rationale for this distinction in the differential informativeness of frames across verb subtypes may be implicit in the needs of the learning device. Specifically, the information quality of the input is more efficient than generous. Syntactic differentiation of the verb set is provided in language design, but primarily where it is required to solve the mapping problem, i.e., for the abstract component of the vocabulary for which observation fails. Notice that it is now possible to understand the course of vocabulary learning within the verb class without reference to conceptual change: Novice learners who do not yet know the language-specific syntax of the exposure language, no matter their conceptual sophistication cannot solve the mapping problem for abstract verbs because the information relevant to identify these items resides almost solely in their distinctive syntactic privileges.

Bootstrapping into language knowledge

Though syntactic frames appear to be a powerful source of information for verb learning, they of course do not occur in nature outside of laboratory demonstrations (e.g., Epstein, 1961) and Lewis Carroll poems. By the time real children have the linguistic sophistication to identify the form that a syntactic structure takes in the exposure language, they have learned the meanings of many nouns from extralinguistic contexts (Gillette et al., 1999). This gain in syntactic knowledge clearly does not cause children to ignore the extralinguistic contexts of word use. Rather, it has the effect of enhancing the *linguistic* representations of events so that they are commensurate with the *extralinguistic* event representations that have been available from early infancy.

Several of the conditions of our simulation presented the adult learners with combinations of cues, which are by hypothesis simulations of successive states of the language learner. These were designed to support theorizing about two crucial aspects of the procedure by which novices turn into experts. The first concerns how successively more complex representations of the linguistic input could be built by the learner from the resources available in the prior learning-stage; that is, *how bootstrapping can be grounded.* The second pertains to the temporal succession from the concrete to the abstract within in verb learning.[10]

Here, we will concentrate attention on the three conditions (*Scenes*, *Scenes + Nouns*, *Full Information*) of the simulation which are intended to model successive representational states. If the simulations mirror the vocabulary-acquisition process, these comparisons allow us to predict at which stage the child is likely to have the representational wherewithal to learn different types of verbs. The performance of subjects in these conditions, again subdivided for the 3 verb types (action, light, mental), is graphically presented in Figure 4.

Figure 4

The first 100 words (word-to-world pairing). In previous studies we demonstrated that nouns submit most readily to the word-to-world pairing procedure simulated in the *Scenes* condition, mirroring the noun-dominant children acquiring their first vocabulary items (see Gillette et al., 1999; Snedeker et al., 1999). Verbs are harder for our subjects to identify from this single evidentiary source; moreover, only for the concrete action verbs is there any substantial success in this word-to-world simulation.

First word combinations (noun-to-verb co-occurrence). Adding *Noun* context information to the *Scene* information ("S+N" in Figure 4) raises the identification scores of our subjects in all three verb subcategories. It is easy to see why. For one thing, certain kinds of noun tend to cluster with certain kinds of verb. A noun like *telephone* can be a give-away to a small class of common verbs, such as *talk, listen,* and *call*. Moreover, the number of nouns in the maternal sentence (that is, within the prosodic contour of the clause) provides an additional clue to the verb meaning. This is because in the very short sentences used to infants (and therefore in the simulated input), the number of nouns is a soft indicator of the number of arguments (Fisher et al., 1994, Fisher, 1996; Gleitman, 1990; Naigles, 1990; Naigles et al., 1993). Thus *gorp* in a prosodically bounded sequence such as "... *John ... gorp...*" is more likely to mean '*sneeze*' than *'kick*', even if John is in the observed scene simultaneously sneezing and kicking a ball. And '*kick*' is a better guess than '*sneeze*' for either *"...John...gorp...ball..."* or "*...ball...gorp...John ..."* even if, because of insufficient language-specific syntactic knowledge, one cannot tell one of these last two from the other.

---

[10] We have discussed some of these issues elsewhere (see particularly Gleitman and Gleitman, in press; and Gillette et al., 1999).

We have conjectured a learning device that uses this primitive quasi-structural noun-to-verb information for purposes that go beyond acquiring the specifics of verb meaning: as a bootstrap into the language-specific clause level syntax. This is because the noun occurrences (e.g., *ball, John*) taken together with the scene information (say, John kicking a ball) allows tentative assignment of the nominals themselves to different argument positions. John is the agent, hence the subject, of the kicking event. As Figure 4 illustrates, this combination of information is more useful for identifying action verbs and light verbs than it is for identifying mental verbs[11]

Structure-to-world pairing: Learning in the presence of coordinated structural and semantic event representations. Looking now at the rightmost columns of Figure 4, we see the dramatic effect on verb identification of increasing representational resources: All three kinds of verb are more efficiently identified in the presence of full information (that is, when subjects saw the scenes and were shown the mother's sentences with their nouns occurring in their original structural positions) and the presence of syntactic-frame information has a stronger impact on the identification of light and mental verbs than it does on action verbs. The result is that in the *full-information condition* the identification of light and mental verbs is as efficient as the identification of action verbs

In this experimental condition, the adult subjects were highly efficient "learners": They correctly identified 78% the target verbs and all but 3 verbs were identified by more than half of them. This efficiency, accuracy, and scope of learning in the presence of full information fits well with the observed vocabulary-acquisition feats of child learners who give evidence of even rudimentary syntactic knowledge. Such children, somewhere between the third and fourth years of life are just the ones whose vocabulary contains both a complement of verbs roughly proportional to the percentage of verbs in the input (Caselli et al., 1995; Lenneberg, 1967), including a substantial number of mental verbs.

---

11. This difference across verb types is a presumptive consequence of two factors. First, the *Scenes* do not become more informative than they were before by addition of the nouns -- the mental acts and states are still hidden in the nontransparent brain, opaque to the device that scans the ongoing scene. Second, some of the nouns that co-occur with the mental verbs are in the embedded rather than the matrix clause and thus play no role in the selectional privileges of the matrix verbs. We should also note that the noun-verb comparison machinery we describe here is distinct from the *semantic bootstrapping proposal* of Grimshaw, 1981 and Pinker, 1984, which acquires nouns and verbs alike from observation of their situational concomitants, and as a prerequisite to the construction of phrase structure. In contrast, the procedure outlined here is necessarily and crucially sequential, with asyntactic noun learning forming the required scaffold for acquisition of specific verbs, and of the phrase structure.

The Logic of the Human Simulation Paradigm

We began this chapter with two puzzles about the course of early vocabulary growth: Why do nouns appear in much larger proportions in the earliest vocabularies compared to their proportions in the adult speech that infants hear? Why do children learn action verbs long before the equally common mental verbs? Our aim in the experimentation was to evaluate two approaches to this input-output disparity: changes in the learner's mentality versus changes in his language knowledge.

The classic method for distinguishing between internally and externally driven changes in organisms' attainments is by looking at the effects of isolation, produced either by nature (e.g., in the language-acquisition context, linguistically isolated deaf children and observationally deprived blind children; Feldman, Goldin-Meadow, & Gleitman, 1978; Singleton & Newport, 1994; Landau and Gleitman, 1985) or by experimental artifice (e.g., Herodotus, ca 410BCE; for a review Gleitman & Newport, 1995). The Human Simulation Paradigm is another way of testing performance under varying conditions of external information-availability. This paradigm has several critical properties that make it especially useful for our purpose: (1) It removes conceptual change as an explanation of the findings by testing adults only; (2) By the same token, it reduces the problem of vocabulary learning to mapping; (3) It therefore allows analysis of the effects of changing information on mapping success, both in rate and type of item that can be acquired; (4) It maintains control of the stimulus situation by testing the very same words in the very same sentences, differently represented across the conditions; (5) It maintains realism because the test items are the words mothers use most to novices, as they actually appeared within the mothers' own utterances.

Two important kinds of objection to the usefulness of this paradigm for theorizing have been raised in the literature. The first is a misunderstanding of what can or should be meant by "observation". To repeat a point we made earlier (see footnotes 5 and 8), nothing in what we say, or in the Human Simulation paradigm as instantiated here, implies that the observer, be he child or college sophomore, perceives the visible world only in terms of some low-level sensory-perceptual representation. We only ask: W*hatever* the set of interpretations that a sophisticated observer can take out of his transactions with the world, is it rich and restrictive enough to converge on the unique and specific interpretation implied by the actual lexical choice of the speaker? Can you tell from observation of, say, a book moving from Alice to John, what a speaker pertinently and saliently wants to say of it? For after all, Alice is *giving* the ball but also *handing* it and possibly *showing* it, the book itself is *moving*, and John is *getting*, *receiving*, *taking* and maybe *snatching* it; both of them are *looking* at it, and likely one or both, or the speaker herself, is thinking about the motives behind the transfer.

Indeed, it is because of the very richness of human perception and conception that the problem of language acquisition has become so difficult to explain in the modern era. Hume didn't have this problem. His babies all interpreted the scenes they saw in the same ways, those offered up by primitive sensory mechanisms. In contrast, modern babies and their simulated counterparts, while usually able to identify the referent event, are unable to assign a unique predicate-interpretation to that event, a finding that should have been predicted by anyone who ever read Chomsky's review of Skinner (1959).

A related objection that has been raised to the Human Simulation Paradigm is that its observational condition either denies or fails to explore the potency of the social cues that are available in the ongoing interaction between the mother and child (e.g., Pinker, 1994). But this is again a misunderstanding, this time of our subjects' situation. The primary cues that are available to the novice language learner are also available to the adults in these studies. During the 40 second video clip, they see the development of the ongoing interaction as well as the gestures, eye gaze, and facial expressions of both the mother and the child. And the subjects clearly use this information to restrict their hypotheses. They show a strong tendency to limit their responses to labels for the objects and events that the mother and child are focusing on. This sensitivity to social-pragmatic cues is evidenced most clearly by the target verb, *come*. In all 6 scenes, as it happens, each mother uses the verb to request that her child return to where she (the mother) is sitting. And in the 6 scenes the toddlers, true to their nature, continue to run away despite the mothers' pleas. But the lack of a direct correspondence between the depicted actions and the meaning of the word did not faze the subjects. In both studies, 75% of them identified the word from the video clips. The subjects' problem is not that social cues about the event in view are unavailable to them. It is that non-linguistic social cues are often ambiguous. Children can often use eye gaze, action sequences, and gestures to learn which event or object is being referred to (Baldwin, 1993, Tomasello, 1995) but these cues cannot tell them what perspective the speaker is taking on that entity. Clark has pointed out that adults often provide verbal information about the perspective that a word takes in the conversation in which they introduce it (1997). In these studies we threw out this information by erasing the audio. This was consistent with our goal of isolating the cues that are available to novice language learners.

Major findings

Summarizing the outcomes, both very young children and their simulated adult counterparts are highly "concrete" in their vocabulary learning under the very conditions that *must* obtain in the first moments of word learning, when the only inferential tool is discovery of recurrent aspects of the extralinguistic situations in which a particular word occurs. They exhibit noun dominance, plus concreteness in their verb learning. In contrast, three- and four-year olds, as well as the adult subjects when provided with coordinated structural and semantic information, seem to effortlessly acquire both abstract and concrete verbs. We conclude from the similarities between child and adult-experimental learning functions that changes in the character of early vocabularies may have less to do with the children becoming wiser over time than with their becoming experts in the semantically relevant syntax of their native tongue.

As for any experimental review, it is always possible to deny that the manipulations, limited as they are in fidelity to the real world of learners, truly model their target population. The simulations are mute about the extent to which verbs could be learned in the presence of more strenuous teaching techniques on the part of the adult community (for an important experimental demonstration and discussion, see Gropen, Pinker, Hollander, & Goldberg, 1991), or in the presence of a massively larger input database of situational evidence. Their only interpretable finding is that adult identification is sensitive to the type of information available, in a way plausibly related to the child acquisition sequence. This is an existence proof that vocabulary acquisition in the real case may reduce mainly to a mapping problem, and does not require invocation of conceptual change to explain.

Where did the phrase structure come from? Or why the children of the world oppose the Whorf-Sapir Hypothesis.

Explicit in the syntactic bootstrapping approach to word- and structure-acquisition (and, for that matter, in the semantic-bootstrapping approach of Grimshaw, 1990 and Pinker, 1984) are two background suppositions. The first is that infants, much like older humans, are disposed to interpret the world in terms of predicate-argument structures whose contents concern the relationships among entities, properties, states, and events. The second is that there is a mapping between these conceptual structures and the clause-level syntax of human languages which is universally the same and transparent in at least the following senses: (1) Events and states are described at the level of the clause; an important corollary of this universal is that, in every language, mental-content verbs license clause-like complements; (2) By and large, there is a one-to-one relation between participants in an act (or state) and noun-phrases in a clause: Unary relations (e.g., John falling) will be mapped onto intransitive structures, binary relations (e.g., John carrying a suitcase) will be mapped onto transitive structures, ternary relations (John giving a ball to Horace) onto ditransitive structures (i.e., the Theta Criterion and the Projection Principle, in some form, Chomsky, 1981)[12] (3) Conceptual dominance relations among actors in an event (the agent, the theme or patient, the recipient) map onto structural dominance in the sentence (the subject, the direct object, the indirect object); a thematic hierarchy. *These simplest syntax-semantics relations are unlearned, "natural," and therefore can serve as part of the presuppositional armamentarium that children bring into the language learning situation; indeed they establish the framework within which language learning is possible at all* (see Jackendoff, 1985 for detailed explication of relations between conceptual and linguistic structure, and their implications for learning).

We have invoked such unlearned principles repeatedly in the present discussion: Learners can choose among the many plausible interpretations of an observed event *because* they can inspect necessary semantic implications of an observed syntactic structure. Of course, beyond these first principles, there are many syntactic-semantic correspondences that differ across languages. To pick a few at random: Many languages have locative verbs where English has locative prepositions; languages differ in surface placement of subjects; not all languages have lexical causatives, and so on. So there is plenty of detailed work for a novice to carry out to achieve mature knowledge of the native tongue.

Several authors have taken a more radical position than the one defended herein: Responding to the fact that there are indubitably differences in the correspondence rules of various languages, they try to bite the bullet and claim that all such correspondences are induced via detailed environmental exposure.

12. "By and large" hedges for certain architectural commitments of languages that complicate this picture, e.g., the requirement for a surface subject in English such that zero-argument predicates have one too many associated noun-phrases ("It rains"); symmetrically, the broad licensing of denominal verbs means that some will have too few, e.g., *butter*; in addition, each language exhibits various unsystematic quirks in these mappings. As we have discussed elsewhere, implicit arguments (and the massive argument-dropping of various languages, e.g., Mandarin Chinese) are easily encompassed within our theoretical approach, just because we hold that the learner is attentive to the full *range* of structures associated with a verb, and their semantic implications. Thus one can say "I gave at the office," "I gave my life," and "I gave Jay a good idea," all these privileges consistent with and informative of the argument-level semantics of *give.* Such verbs as *go, fall, happen*, and even *rain* are licensed for the first of these three structures but crucially never could appear ditransitively because they are incompatible with the notion of transfer.

The results of our simulation bear on the plausibility of this claim.

Generalizing syntactic-semantic correspondences from meaning correspondences: Most of these authors propose that children derive the syntactic-semantic correspondences from generalizations based on a large and diverse set of verbs whose properties the child priorly acquired one by one, by observation (e.g., Bates & MacWhinney, 1987; Bowerman, 1990). This view is perhaps most clearly represented in Tomasello's Verb Island Hypothesis (1992). Tomasello claims that each verb the child learns is initially an isolated island of meaning; the child does not recognize that verbs belong to a single class and therefore does not productively use morphology, passives, or syntactic alternations. At about 2;6 children are supposed to first notice the similarities amongst verbs and begin to build a syntactic class. Of course, if the children have no verb class, they have no subclasses based on refined syntactic-semantic correspondences either. For example, according to this view the correspondence between word order and thematic role assignment in caused motion events isn't learned until around 3;0 (Akhtar & Tomasello, 1997), by which time the child has learned hundreds of verbs. Previous spontaneous production of correct word order is attributed to the use of word-specific patterns.

The prerequisites of such generalization hypotheses are clearest in the Allen connectionist model which essentially instantiates this process. The model "learns about verbs and their argument structures from naturalistic input" (Allen, 1997 as reported in Seidenberg, 1997; p. 1602). It is trained with a set of sentences, each paired with an interpretation. From this information the model learns a set of syntactic-semantic correspondences, as evidenced by its ability to correctly interpret novel verbs in previously encountered constructions.

The Allen model performs its learning feats only because it is given the meaning of each verb and each sentence in the training set; presumptively, in the real case that meaning must be derived by inspecting co-occurring events in the world. Our experimental child surrogates were unable to "learn about verbs and their argument structures..." by comparing the situational contingencies for their occurrence; that is, by intuiting their interpretations. The subjects failed to come up with correct interpretations 85 percent of the time (Figure 1). And when we look at their preferences at the final exposure, the choice made by the largest number of subjects is a false one for three-quarters of the items (see Snedeker, 2000).

The systematicity of the errors is particularly important. Modeling work, such as Siskind (1996), shows that even a very scantily clad learning device can acquire something like an interpreted phrase-structure system by distributional analysis over highly errorful data. However, this is on the assumption, to the best of our understanding, that the errors in that data are not highly organized, i.e., that they slow learning rather than causing it to converge on false solutions. What we see in the adult observational learners is hypothesis generation that is marching in two wrong directions: toward solutions that are either overly general (with accumulating evidence, many verbs come to mean "look" or "play") or overly concrete; generalizations that are false of the input.

Our point is that all proposals requiring that verbs be learned by a procedure that gleans the intended event-structure of heard sentences from extralinguistic observation alone fails to answer to the real chronology of lexical acquisition, or to its simulation in adult laboratory subjects. These proposals founder just because their rock-bottom presupposition cannot be met: The events accompanying verb use do not "naturally" and without other aids create a search-space of interpretations that is at once both narrow enough to acquire verb meanings efficiently and broad enough to encompass their typology (abstract as well as concrete items). If the syntactic-semantic correspondences could be anything at all, that is, if learning had to be open-minded induction as in the more blatant forms of the Whorf-Sapir hypothesis, language knowledge would be impossible to attain.

The light verb hypothesis: Recently, several researchers have suggested another possibility in the continuing search for the learning story that makes the fewest advance commitments about how language expresses reality. They propose that light verbs play a critical role, and enable the acquisition of argument structure (Ninio, 1996; Clark, 1996; Goldberg,1999). This proposal goes something like this: Novice learners do not know about the linkages between syntactic frames (or constructions) and event semantics. They learn their first few verbs without the benefit of syntax, just as in the Allen model, by noting the extralinguistic contexts in which these verbs occur (making use, perhaps, of the nouns that occur in the utterance). Among these early-acquired verbs are the light verbs. They are learned quickly because they are so frequent. The child notices that each of these words*, in its use as a light verb,* has a fairly minimal meaning and appears most frequently in a single syntactic frame. The child associates the meaning of the verb with that frame and begins to attribute that meaning to other verbs that appear in that frame (adding for each verb the other bits of meaning suggested by the unique extralinguistic contexts in which it appears). Notice then, that this approach potentially does away with the need to posit not only a god-given ability in the infant to interpret the world but a just-as-god-given disposition to map that world only in certain ways onto linguistic structures.

But again this story critically depends on the child being able to learn verbs from its extra-linguistic and nominal contexts, prior to syntax. Particularly -- and this alone will beg the question that the hypothesis was designed to resolve -- such a learner must recognize just when this verb is being used in its *light verb* context. How is that to be done? Consider, e.g., the frequent item *bring.* How is a presyntactic infant to disentangle its standard interpretations, its “light” interpretations, and its “idiomatic” interpretations (as in, e.g., *John brought up his suitcase, his dinner, his baby, his objections; brought down his suitcase, the government; the house).* Moreover, as the Human Simulation demonstrated, the commonly used verbs that can be assigned to the light-verb class are not the sort most easily acquired by observation or by attention to their nominal contexts, even by sophisticated adults (Figures 3 and 4). Close inspection of our results suggests that the only light verbs for which there was some successful identification in the *Scenes Condition* were *come*, *go*, and *put* (whose status as light verbs is least clear), and only in cases where these could be interpreted in their standard -- not light -- interpretation, i.e., when somebody in the scene was saliently going or putting away objects.

In any case, though learned young, light verbs appear to be a feature of early word combinations rather than single word speech (see the examples provided by Clark, 1996; Goldberg, 1999). Work on early comprehension and word learning indicates that children at this level of development also have the ability to use syntactic context as a cue for interpreting an ambiguous or novel verb (Naigles, 1990; Naigles, Hirsh-Pasek, & Golinkoff, 1996; Fisher, 1999). We suspect that the acquisition of light verbs may be one of the first products of structure guided word learning, rather than the engine that makes it possible.

Final thoughts

We began this chapter by asking about the extent to which vocabulary learning reduces largely to mapping: Plato’s problem purged of the error of reincarnation. Is the child’s task merely to recover the conventional pronunciations for preexisting concepts? Usually, the vocabulary acquisition task is taken to involve more than this, to be a problem of concept acquisition as well as a problem of mapping. This position, as we saw, is rendered plausible by noticing that concrete object terms and concrete verbs are learned earlier in life, on average, than abstract terms. The experimental work presented herein was designed to show that a closer analysis of evolving mapping strategies adequately explains these sequential aspects of word learning, without invoking conceptual change at all.

What makes the mapping problem a hard one, we have argued, is that observation offers up myriad salient and relevant representations for any single scene. The enormous breadth of the available hypothesis space for interpreting the world therefore poses real obstacles for word learning. Not only are there many possibilities for just how a heard word matches up with a single observed scene, but the next use of this same word often occurs in circumstances quite different from the first. Thus any learning device limited to inspection of external conditions for word learning faces a many-to-many sorting task of unknown complexity. The more sophisticated the learner in analyzing the social, pragmatic, conceptual, and perceptual concomitants of the scene in view, the larger the problem-space becomes. As we tried to show, the best and most efficient solution is for the learner to construct a more sophisticated representation of the linguistic stimulus, one that exposes its semantically informative syntactic structure. Structure-to-world matching procedures are highly efficient, as we saw, to rein in interpretation of the novel word.

More generally, the work reported here fits into an emerging picture of language acquisition that explains the child's semantic and syntactic accomplishments in terms of layers of information that become available in sequence as a consequence of solving prior parts of the learning problem. In their current form, these bootstrapping approaches leave open almost all questions about the particulars of the lexical and grammatical representations that are finally attained. They are incompatible only with proposed learning procedures for language that assert that vocabulary attainment is prior to and independent of grammar attainment.

Acknowledgments

We thank Jane Gillette and Anne Lederer for their advice and their stimuli. We also gratefully acknowledge Tova Rosman, who collected the data for verb-learning experiment. This work was supported by National Institute for Health Grant 1-R01-HD3750707-01 and a National Science Foundation Center Grant to the University of Pennsylvania Institute for Research in Cognitive Science.

References

Akhtar, N. & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology*, **33**, 952-965.

Augustine (398) *Confessions*, commentary by J. J. O'Donnell, New York : Oxford University Press, 1992.

Baldwin, D.A (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, **29**, 832-843.

Bartsch, K. & Wellman, H.(1995) *Children talk about the mind.* New York, NY: Oxford University Press.

Bates, E. & MacWhinney (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157-193). Hillsdale, N.J.: Lawrence Erlbaum Associates

Bates, E., Dale, P.S., and Thal, D. (1995). Individual Differences and Their Implications for Theories of Language Development. In Fletcher, P. & MacWhinney, B. (Eds.*), The handbook of child language* (pp. 96-151), Oxford, England: Blackwell Publishers.

Bloom, L., Lightblown, P., & Hood, L.(1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, **40**, 1-97.

Bowerman, M. (1990). Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics*, **28**, 1253-1289.

Bretherton, I. & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *Developmental Psychology*, **18(6)**, 906-921.

Carey, S. (1978). The child as a word learner. In M. Halle, J. Bresnan & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264-293), Cambridge MA, MIT Press.

Carroll, L. (1865). *Alice's adventures in wonderland*. New York: Bantam Books.

Caselli, M.C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian, *Journal of Child Language*, **26**, 69-111.

Caselli, M.C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development*, **10**, 159-199.

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. & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language*, **22**, 497-529.

Chomsky, N. (1959). Verbal Behavior, By B.F. Skinner, *Language*, **35**, 26-58.

Chomsky, N. (1981). *Lectures on the theory of government and binding*. Foris: Dordrectht.

Clark, E.V. (1987). The principle of contrast: a constraint in language acquisition. In B. MacWhinney (Ed.) *Mechanisms of language acquisition* (pp. 1-33), Hillsdale NJ: Erlbaum.

Clark, E.V. (1996). Early verbs, event types, and inflections. In Johnson, C.E., Gilbert, J., et al. (Eds.), *Children's language, Vol. 9*, (pp. 61-73). Mahwah, NJ: Lawrence Erlbaum Associates.

Clark, E.V. (1997) Conceptual Perspective and Lexical Choice in Acquisition. *Cognition*, **64**, 1-37.

Dromi, E. (1987). *Early lexical development*, Cambridge: Cambridge University Press.

Epstein, W. (1961). The influence of syntactical structure on learning. *American Journal of Psychology*, **74**, 80-85.

Feldman, H., Goldin-Meadow, S., & Gleitman, L. (1978). Beyond Herodotus: The creation of language by linguistically deprived deaf children. In A. Lock (Ed.), *Action, symbol, and gesture*. New York: Academic Press.

Fisher, C. (1996) Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, **31**, 41-81.

Fisher, C. (1999) Transitivity and verb arguments in acquisition, Paper presented at the Thirtieth Annual Child Language Research Forum, Stanford University, April, 1999.

Fisher, C., Gleitman, H., Gleitman, L.R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, **23**, 331-392.

Fisher, C., Hall, G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, **92**, 333-375.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. New York, NY, USA: Clarendon Press.

Furrow, D., Moore, C., Davidge, J., & Chiasson, (1992). Mental terms in mothers' and children's speech: Similarities and relationships. *Journal of Child Language*. **19**, 617-631.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S.A. Kuczaj, II (Ed.), *Language development: Vol 2. language, thought, and culture* (pp. 31-53). Hillsdale, NJ: Erlbaum.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development*. England: Cambridge University Press.

Geyer, H. (1998). Subcategorization as a predictor of verb meaning: Evidence from Hebrew. Manuscript, University of Pennsylvania.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulation of vocabulary learning. *Cognition*, **73**, 135-176.

Gleitman, L. (1990). The structural sources of word meaning. *Language Acquisition*, **1**, 3-55.

Gleitman, L.R & Newport, E.L.(1995) The invention of language by children: Environmental and biological influences on the acquisition of language. In L. Gleitman & M. Liberman et al. (Eds.), *An invitation to cognitive science: Vol. 1 Language* (pp. 1-24), Cambridge, MA, USA: MIT Press

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*, Chicago: University of Chicago Press.

Goldberg, A. (1999). The Emergence of the Semantics of Argument Structure Constructions, In B. MacWhinney (Ed.), *The emergence of language* (pp. 197-212), Mahwah, NJ: Lawrence Erlbaum.

Goldin-Meadow, S., Seligman, M.E. & Gelman, R. (1976) Language in the two-year-old. *Cognition*, **4**, 189-202.

Gopnik, A. & Meltzoff, A.N. (1997) *Words, thoughts, and theories*. Cambridge, MA: MIT Press.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991). Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition*, **41**, 153-195.

Herodotus (410BCE). *The histories*, translated by W. Blanco, W. Blanco & J.T. Roberts (Eds.), New York: Norton, 1992.

Huttenlocher, J., Smiley, P., Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review*, **90**, 72-93.

Jackendoff, R. (1985). *Semantics and cognition*, Cambridge, MA: MIT Press.

Johnson, C.N. & Maratsos, M.P. (1977). Early comprehension of mental verbs: think and know. *Child Development*, **48**, 1743-1747.

Kako, E. (1998) The Event Semantics of Syntactic Structures, Dissertation, University of Pennsylvania.

Kellman, P.J. & Arterberry, M.E. (1998). *The cradle of knowledge: development of perception in infancy*, Cambridge, Mass.: MIT Press.

Kelly, M.H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review,* **99**, 349-364.

Kuczaj, S.A. (1990). Constraining constraint theories. *Cognitive Development*, **5**, 341-344.

Landau, B. & Gleitman, L.R. (1985). *Language and experience: Evidence from the blind child.* Cambridge, MA: Harvard University Press.

Lederer, A., Gleitman, H., & Gleitman, L. (1991). The informativeness of cross-situational and cross-sentential evidence for learning the meaning of verbs. Paper presented at the Boston University Conference on Language Development. Boston.

Lederer, A., Gleitman, H., & Gleitman, L. (1995) Verbs of a feather flock together: Semantic information in the structure of maternal speech. In M. Tomasello & W.E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 277-297), Hillsdale, NJ: Lawrence Erlbaum Associates.

Lenneberg, E.H. (1967). *Biological foundations of language*. New York: Wiley.
Levin, B. (1993). *English verb classes and alternations*. Chicago: University of Chicago Press.

Locke, J. (1690). An Essay Concerning Human Understanding (pp. 259-298). Cleveland: Meridian Books, 1964.

Markman, E.M. (1994). Constraints on Word Meaning in Early Language Acquisition. In L. Gleitman & B. Landau [Eds.], *The acquisition of the lexicon* (pp. 199-227), Cambridge, MA: MIT Press.

Moore, C., Gilbert, C., & Sapp, F. (1995). Children's comprehension of the distinction between want and need. *Journal of Child Language*, **22**, 687-701.

Naigles, L. (1990) Children use syntax to learn verb meanings. *Journal of Child Language*, **17**, 357-374.

Naigles, L.R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, **58**, 221-251.

Naigles, L.R., Gleitman, H., & Gleitman, L.R. (1993). Children acquire words meaning components from syntactic evidence. In E. Dromi (Ed.), *Language and cognition: a developmental perspective* (pp. 87-102). Norwood, NJ: Ablex.

Naigles, L.R., Hirsh-Pasek, K., & Golinkoff, R.M. (1996), Young children’s ability to use syntactic frames to derive meaning, In K. Hirsh-Pasek & R.M. Golinkoff*, The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press.

Ninio, A. (1996) Pathbreaking verbs in syntactic development. Paper presented at the 7th International Congress for the Study of Child Language, Istanbul, Turkey, July, 1996.

O'Grady, W. (1987). *Principles of grammar & learning.* Chicago, IL: University of Chicago Press.

Pinker, S. (1984). *Language learnability and language development*, Cambridge, MA, USA: Harvard University Press.

Plato (412 BCE). *Phaedo*, D. Gallop (Translator), New York: Clarendon Press, 1990.

Sandhoffer, C., Smith, L., & Luo, J. (2000). Counting nouns and verbs in the input: Differential frequencies, different kinds of learning? *Journal of Child Language*, **27**, 561-585

Seidenberg, M.S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, **275**, 1599-1603.

Shatz, M., Wellman, H.M. & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, **14**, 301-321.

Singleton, J. & Newport, E. (1994). When learners surpass their models: The acquisition of American Sign Language from impoverished input. Manuscript, University of Rochester.

Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to meaning mappings. *Cognition*, **61**, 39-91.

Smiley, P. & Huttenlocher, J. (1995). Conceptual development and the child's early words for events, objects, and persons. In M. Tomasello & W.E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 21-62), Hillsdale, NJ: Lawrence Erlbaum Associates.

Snedeker, J. (2000). Cross-situational observation and the semantic bootstrapping hypothesis. In E. Clark (ed.), *Proceedings of the Thirtieth Annual Child Language Research Forum.* Stanford, CA: Center for the Study of Language and Information.

Snedeker, J, Gleitman, L & Brent, M. (1999), The Successes and Failures of Word-to-World Mapping. In A. Greenhill, M. Hughs, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the Twenty-third Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Talmy, L. (1985). Semantics and syntax of motion. In J. Kimball (Ed.), *Syntax* *and semantics* (Vol. 4, pp. 181-238). New York: Academic Press.

Tardif, T., Shatz, M. & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, **24**, 535-565.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development.* Cambridge: Cambridge University Press.

Tomasello, M. (1995). Pragmatic contexts for early verb learning, In M. Tomasello & W.E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 115-146), Hillsdale, NJ: Lawrence Erlbaum Associates.

Woodward, A.L. (1998). Infants selectively encode the goal object of an actor's reach, *Cognition*, **69**, 1-34.

Figure 1: Performance across trials for each information condition (verb targets).

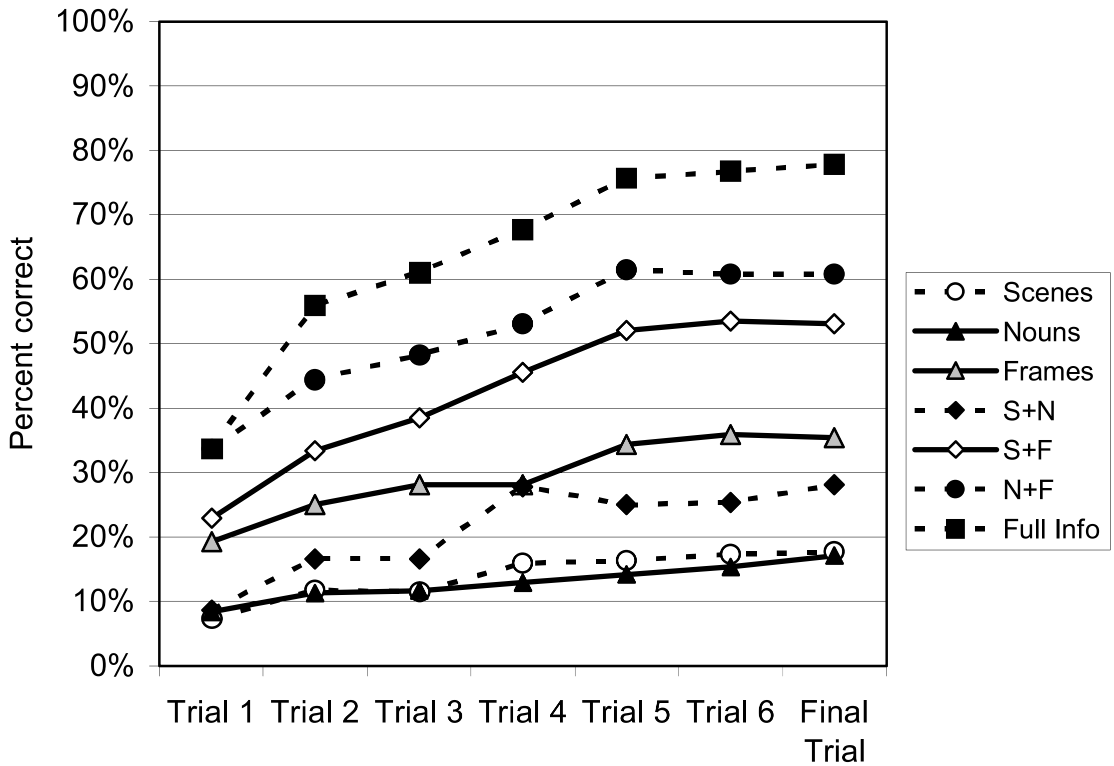

Figure 2: Performance on the final trial for all conditions.

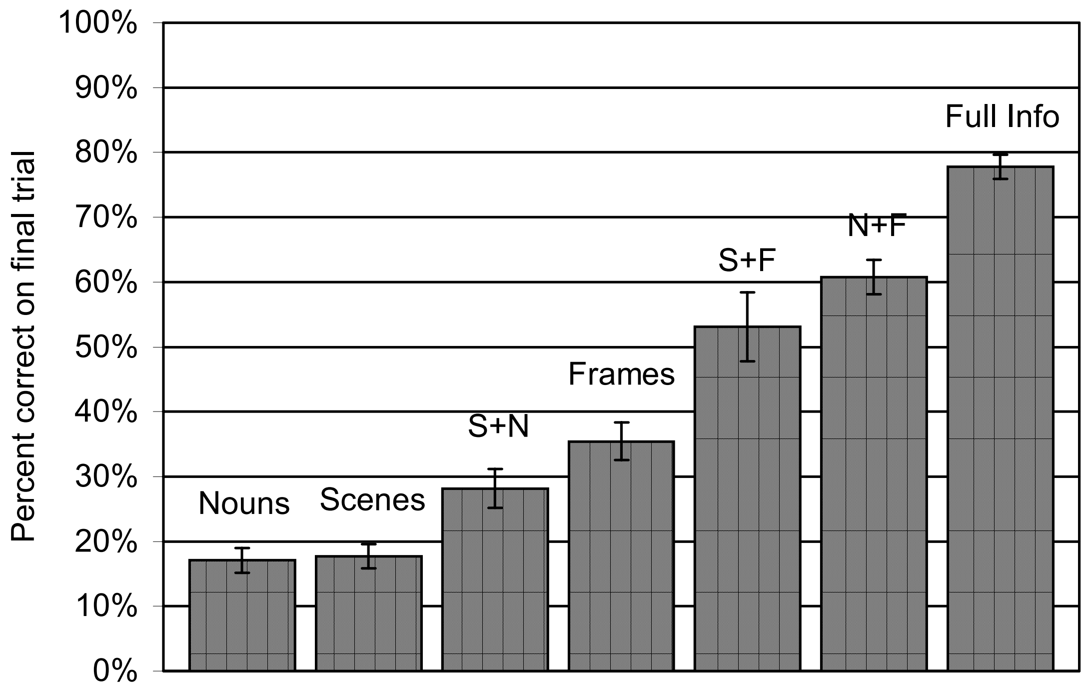

Figure 3: Different types of verbs require different information sources.

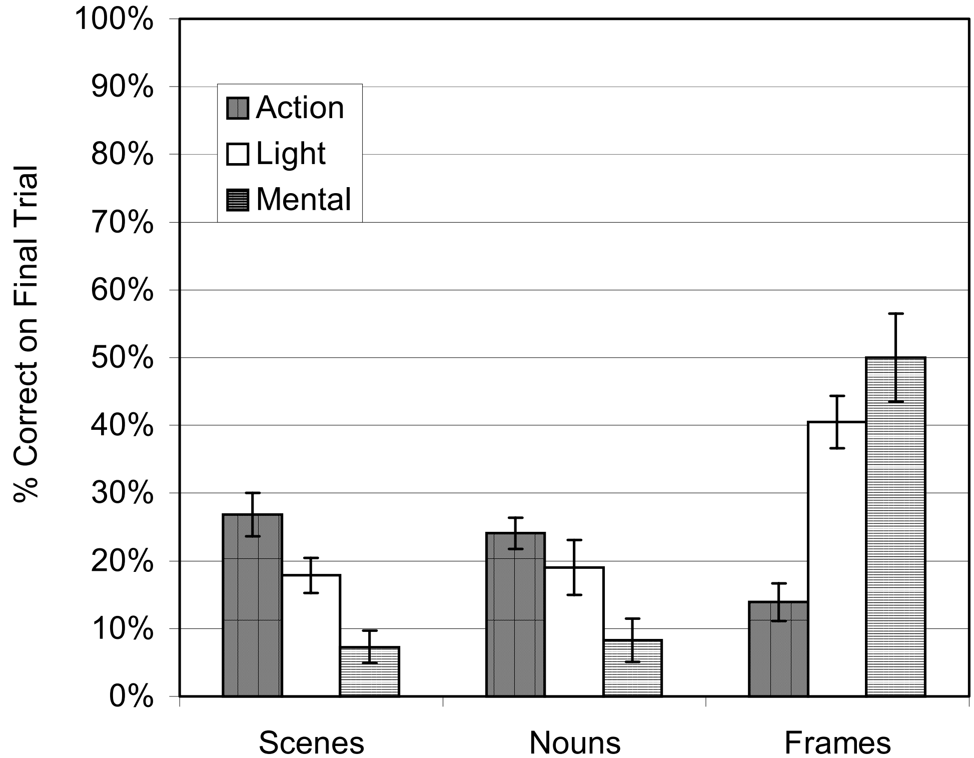

Figure 4: Performance for each verb class improves with increased representational resources.