When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth*

Cynthia Fisher[a,**], D. Geoffrey Hall[b], Susan Rakowitz[c] and Lila Gleitman[d]

[a] Department of Psychology, University of Illinois, 603 E. Daniel Street, Champaign, IL 61820, USA

[b] Medical Research Council, Cognitive Development Unit, 4 Taviton Street, London WC1H 0BT, UK

[c] Department of Psychology, Fairfield University, Fairfield, CT 06430, USA

[d] Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104, USA

We ask how children solve the mapping problem for verb acquistion: how they pair concepts with their phonological realizations in their language. There is evidence that nouns but not verbs can be acquired by pairing each sound (e.g., 'elephant') with a concept inferred from the world circumstances in which that sound occurs. Verb meanings pose problems for this word-world mapping procedure, motivating a model of verb mapping mediated by attention to the syntactic structures in which verbs occur (Landau and Gleitman 1985, Gleitman 1990). We present an experiment examining the interaction between a conceptual influence (the bias to interpret observed situations as involving a causal agent) and syntactic influences, as these jointly contribute to children's conjectures about new verb meanings. Children were shown scenes ambiguous as to two interpretations (e.g., *giving* and *getting* or *chasing* and *fleeing*) and were asked to guess the meaning of novel verbs used to describe the scenes, presented in varying syntactic contexts. Both conceptual and syntactic constraints influenced children's responses, but syntactic information largely overwhelmed the conceptual bias. This finding, with collatoral evidence, supports a syntax-mediated procedure for verb acquisition.

* Thanks are due to Steven and Marcia Roth for funding that supported the experimental work, through a grant to Lila Gleitman. Preparation of this paper was also supported by NSF grant DBC 9113580 to C. Fisher. We wish to thank Renee Baillargeon, Judy DeLoache, Henry Gleitman, Jane Grimshaw, Aravind Joshi, Michael Kelly, Barbara Landau, Anne Lederer, Ellen Markman, Kevin Miller, Letty Naigles. Elissa Newport, Brian Ross, and Sandy Waxman for helpful comments on earlier drafts. Noam Chomsky is also thanked for continuing commentary on this line of work. Finally, we wish to thank the children, staff, and parents of the Saint Faith Nursery School in Philadelphia for their cooperation.

** *Corresponding author.*

SSDI 0024-3841(93)E0050-H

1. Introduction

Every standard text in psychology (or in education or linguistics for that matter) asserts that children aged 18 months to 6 years acquire 5 to 10 new words a day. How do they manage to do so?

We concentrate here on a single aspect of the word-learning problem: Granted that children can hypothesize some appropriate set of concepts, how do they decide which sound segment corresponds to each such concept? For instance, granted that they can entertain the concepts 'elephant' and 'give', how do they come to select the sound /elephant/ for elephants and /give/ for giving? This aspect of acquisition is called the *mapping* problem for the lexicon.[1]

Solution of the mapping problem has traditionally been assigned to a word-to-world pairing procedure in which the learner lines up the utterance of a word with the co-occurring extralinguistic contexts. Thus *elephant* comes to mean 'elephant' just because it is standardly uttered by caregivers in the presence of elephants.

Gillette and Gleitman (forthcoming) have begun to document just how well the word-to-world pairing procedure works in practice for simple nouns. In these manipulations, adult subjects watch a video, five or ten minutes in length, of mothers and their young children (MLU < 2) at play, but with the audio turned off. These lengthy situational segments allow the subjects to pick up whatever clues are available from the pragmatic concomitants of the speech event. They are told that, at the instant some particular noun is being uttered by the mother, they will hear a beep, their task being to guess what noun that was. For the nouns most frequently used in these mother/child interchanges, the subjects are almost at ceiling. Usually, even a single scene/beep pair is enough for the subject to identify the noun the mother was uttering.

These findings imply two things. The first concerns the input situation itself: Evidently, mothers of very young children usually say nouns just when the objects that these label are the focus of conversation and are being manipulated by the participants. This makes their recovery from context easy (see Bruner, 1975, and Slobin, 1975, for prior evidence of this here-and-now property of maternal speech to children). The second concerns 'natural' interpretations of situational information: The observer seems efficient at

[1] In our notation, /slashes/ represent sound, 'single quotes' the concept, "double quotes" the utterance, and *italics* the word as an abstract object.

guessing the level of specificity at which the speaker is making reference – elephants rather than animals or puppets – despite the fact that all of these interpretations fit the observed scenes equally well (*pace* Quine 1960; for evidence from child word learning, see Hall 1993; Hall and Waxman 1993).

So far so good: One can learn that the word for 'elephant' is /elephant/ (or /beep/) because it is said in the presence of elephants. However, when we turn to the acquisition of lexical categories other than the noun, this promising story appears to fall apart. Subjects cannot correctly guess which *verb* the mother is saying under the same circumstances – observation of the mother/child scenes without audio other than the beep. Though the subjects do choose as their guesses the most common maternal verbs of all (e.g., *come* and *put* as opposed to *arrive* and *situate*), they fail to select the one that the mother was actually uttering at the sound of the beep. Their success rate is between 0 and 7%, depending on details of the manipulation.

Why is the observed scene so decisive for nouns and so uninformative for verbs? One factor proposed by Gentner (1978, 1982) has to do with the concepts that these lexical classes standardly encode, namely the difference between object-reference concepts and relational concepts (see also Nelson 1974). The reference of many nouns can apparently be extracted by appeal to principles of object perception and pragmatic inference, but even the homeliest verb meanings express relations *among* such concepts. Which such relation the speaker has in mind to convey is rarely accessible from observation alone. Moreover, the nouns are frequently used in deictic-ostensive contexts to young learners: "This is a ball" (Ninio 1980, Bruner 1983), while verbs are much rarer in such contexts as "This is hopping".

Another important factor is that the verbs are not uttered even to young learners in a tight time-lock with the events (Tomasello and Kruger 1992, Lederer et al. 1991). Even when the events and verb utterances are relatively close in time, their seriation differs, a problem we have called *interleaving*. For instance, consider a scene in which the child is pushing a car, and then upon request from his mother carries it over to show to his grandmother, who beams. The serial order of events here is *push*, *go*, *show*, *beam*. But the mother actually says "Go show Granny what you're doing, she'll think you push the car so well". Little problem arises for these adult subjects (or, we presume, for children) in getting the gist of the conversation, but the gist is very far from explicit identification of the verbs. The problem posed for identification is that the number and order of verb utterance (*go*, *show*, *do*, *think*, *push*) do not line up with the event sequence. Notice, as well, that certain verbs commonly used by the caregivers are so general (*do*) or so abstract (*think*) as to be difficult to relate at all to what's actually going on (in this latter example, beaming).[2]

All these complexities bear on an extremely robust finding in the language learning literature: Verbs are very rare in the first spoken (or comprehended) 50 words of child vocabularies, rather most items are nouns with a scattering of social items ("bye-bye") and spatial prepositions (Goldin-Meadow et al. 1976, Nelson 1974, Bowerman 1976, Dromi 1987). This striking dominance of nouns (above their type frequency in maternal speech) persists until the third year of life (Gentner 1982).

We will contend here that, owing to the kinds of problems just sketched, verbs must be learned by a procedure that differs from the early noun-learning procedure that pairs isolated words (or beeps) to their real-world contingencies. According to our hypothesis, verb learners recruit evidence from the syntactic structure in which new verbs appear, and pair this structural evidence with the information present in the scene. Thus we postulate a sentence-to-world mapping procedure for verbs rather than the word-to-world procedure that is satisfactory for explaining first nouns (for earlier statements of this position, see Landau and Gleitman 1985, Fisher et al. 1991, Gleitman 1990). This would begin to explain the noun-before-verb developmental findings: It takes time to acquire structural knowledge, and nouns but not verbs can be acquired efficiently in the absence of such knowledge. Moreover, knowledge of the noun meanings is, as we shall see, a prerequisite to extracting the verb meanings.

We will present here an experiment that assesses children's use of situational and syntactic evidence for solving the mapping problem. But before doing so, we want to describe informally the ideas behind the syntax-sensitive learning procedure we have in mind. Fuller discussion is reserved until the experimental findings have been presented.

[2] Of course it is easy to think of nouns that are similarly 'abstract', such as *liberty*, so relative ease of learning via extralinguistic observation is not theoretically identifiable with the noun/verb distinction. But it is as a practical matter: Abstract verbs are common in usage to children (5 of the most frequent verbs in maternal use to children under two years refer to mental states and acts, *want*, *like*, *think*, *know*, and *see*) but all the most frequent nouns in our corpus refer to visible object classes or names, e.g., *Mommy*. The more important point is that subjects cannot reconstruct even the maternal verbs that refer to observable actions (*go*, *eat*, *catch*, etc.) by watching the scene in the presence of evidence (the beep) of just when they were uttered.

2. Syntactic supports for verb learning

As just discussed, observation of the pragmatics of scenes appears to underdetermine verb construals. The latitude for interpretation seems much too broad if we grant (as we must) formidable perceptual, conceptual, and pragmatic interpretive capacities to youngsters. Though it is clear (see Pinker, this volume) that there are severe constraints on what can be lexicalized as a verb, and on just how humans conceptualize a scene in view, within these limits the hypothesis space is still vast. It is often and truly said that a picture is worth a thousand words, but evidently that's just the problem for verb vocabulary acquisition.

One plausible way to distentangle verbs in ambiguous circumstances is to note which known nouns occurred in construction with the verb (Was the patient mentioned?) and the structural positions of these nouns (Which one was the subject of the sentence?). In principle this information can be useful because the surface-structural properties of sentences are well correlated with (in fact, are projections from) certain aspects of their semantics. To use such clues, the learning device must analyze the structure within which the novel verb is being heard. We have termed such a structure-dependent learning procedure *syntactic bootstrapping*.[3]

This approach is similar to prior proposals for solving the mapping problem in many regards. It posits that learners inspect ongoing events for clues to the verb meanings, armed with sophisticated perceptual, conceptual, and pragmatic knowledge.[4] In addition, and in accord with the known facts

[3] Pinker (this volume) has objected to this terminology, but what's in a name? We mean by it that the learner is presumed to converge on verb meaning by joint use of structural and situational evidence. *Semantic bootstrapping*, a term coined by Pinker (1984), ultimately concerns acquisition of a phrase-structure grammar. The first steps in this procedure extract word interpretations solely by observing the contingencies for their use – by word-to-world pairing. If this can be done for verbs (which we contend it cannot) as well as for nouns (which we contend it can), then the next step in Pinker's procedure is to assign the items to lexical categories based on the acquired meanings (e.g., a thing is likely to be labeled by a noun while an action is likely to be labeled by a verb, a generalization that holds cross-linguistically; Grimshaw 1981). So the terms *syntactic* and *semantic bootstrapping* are not really parallel. The question remains whether the background assumption of semantic bootstrapping can be satisfied, namely, whether the verb meanings can be acquired solely from extralinguistic evidence, prior to acquisition of the phrase structure.

[4] Pinker (1989, this volume) asserts that our position has limited the infant's conceptual repertoire to 'sensory properties' à la Locke. Quite the contrary. It is the latitude of the hypothesis space in the conceptually well-endowed infant that opens the door to multiple interpretations of single words and sentences (Chomsky 1957). Of course, if learners were as

about language learning, it posits that verb learning occurs in the presence of a priorly learned vocabulary of nominal items (Lenneberg 1967, Gentner 1978, 1982).

The innovation has to do with the way learners are posited to represent the *linguistic* input that is to be paired with the extralinguistic input: as a parse tree within which the novel verb occurs. The structured sentence representation can help in acquiring the novel verb just because it is revealing of the argument-taking aspects of that verb's interpretation. If phrase-structural knowledge of the exposure language facilitates verb learning, then the developmental priority of nouns begins to be understandable; and, so does the explosion of verb vocabulary acquisition simultaneous with the appearance of rudimentary sentences in speech (Lenneberg 1967).

In the experiment that follows we examine an example of this problem and its proposed solution: There are many meaningfully distinct paired verbs that occur in virtually all and only the same real-world contexts, for example, *give* and *receive*, or *chase* and *flee*, *lead* and *follow*. When John gives a ball to Mary, Mary receives the ball from John. Movie directors make an art of distinguishing such notions visually. They can zoom in on the recipient's grateful mien, the giver out of focus or off the screen completely. Using the word *receive* rather than *give* is a linguistic way of making the same distinction. But only for a listener who understands their meanings. Without a zoom lens, how is a learner to acquire the distinction in the first place?[5]

If the learner considers the novel verb use within a syntactic structure, and requires an interpretation that is congruent *both with the scene and the*

Locke proposed, they would fail to acquire language for quite different reasons; namely, that almost no words refer to sensory properties or combinations of these (Fodor 1981, Armstrong et al. 1983).

[5] For all these perspective-changing verb pairs, distinguishing environmental conditions are not really nonexistent, but are very rare. For instance, it is reasonable to say *The people fled the city* but not so reasonable to say *The city chased the people* (example from Pinker, personal communication). So in principle one can flee without being chased. The question is whether these rare dissociating environments play a role in the child's differentiation of these paired verbs. We know that close to a third of verb uses to young children are in the absence of their referents – not about the here-and-now (Beckwith et al. 1989), as in "Granny is *coming* to visit next week", which occurs in the absence of visible coming (or visible Granny). This means that the learning device must be quite tolerant in evaluating scene-to-world conjectures. /Come/ must be mapped onto 'come' though it is often said when nothing is coming, and often *not* said when something *is* coming. No learning procedure willing to discount the large percentage of scene/usage mismatches for *come* could treat the vanishingly rare mismatches for *give/receive* or *chase/flee* as anything but noise.

structure, there is a solution to the mapping problem for these verbs. Consider a listener hearing one of these sentences:

(1) Look, biffing!
(2) The rabbit is biffing the ball to the elephant.
(3) The elephant is biffing the ball from the rabbit.

while watching a rabbit give and an elephant receive a ball. As we will show, if the listener has no access to the syntactic framework, as in (1), she will probably interpret /biff/ as related in meaning to English *give*. Hearing sentence (2) bolsters this choice. But a learner who inspects sentence (3) favors *receive*.

There are two clues to this choice in sentences (2) and (3). First is the *to/from* distinction, which indicated which entity is source and which is goal of the moving ball. Second is the placement of *rabbit* and *elephant* within the structure, for whatever entity showed up as the subject of the sentence has been selected, *in the utterance*, as the one that the sentence is 'about' – the entity of whom the act is predicated. The notional interpretation of /biff/ must be one that still fits the scene observed but casts it in a different light: If the subject of the predication was "rabbit", then the act was giving; if it was "elephant", then the act was receiving. In essence, for a listener sensitive to the full sentence, the interpretation of the observed scene will have been affected by the linguistic observation that accompanies it.[6]

The difficulty of the mapping problem is not restricted to the perspective-changing verbs that we have just discussed. Consider a learner observing a scene in which a rabbit pushes a duck, who falls; and hearing one of these three sentences:

[6] Presaging later discussion, note that a discovery procedure that implicates semantic deductions from surface structure must confront the fact that the relation between surface syntax and argument structure, even within a single language, is complex at best and can be misleading in some cases. Consider the case of *get*, a near relative of the two verbs (*give* and *receive*) just discussed. *Get* is subject to two interpretations. When we say "Emmanuel got a book from the library", the subject (Emmanuel) is certainly the causal agent in the book's moving out of the library. But when we say "Emmanuel got the flu from Aaron", Aaron was the intended causal agent, assuming that Emmanuel wanted no part of the flu. Thus surface position of the nominals does not uniformly reflect distinctions in their thematic roles. Moreover, *get* can appear in two-argument sentences such as "Emmanuel got the flu" in which its transfer-of-possession sense is masked (if intended) and may not be intended in the first place. We will return to these issues.

(4) Look, ziking!
(5) The rabbit is ziking the duck.
(6) The duck is ziking.

Hearing (4) should not support the selection of 'push'-like vs. 'fall'-like interpretations. But (5) must be 'push' and (6) must be 'fall'. This time it is the number of noun phrases (in separate argument positions) in the sentences that bears on the interpretation of the verb item; the intransitive sentence (6) simply will not support the causal property of *push*.

For this latter pair (as opposed to *give/receive*), cross-situational observation is available as an alternate route for acquiring the distinction between them, requiring no attention to syntax: Eventually there will be falling scenes without a pusher, allowing their disentanglement (Pinker 1984). All the same, in the real case learners may recruit syntactic cues to facilitate the choice.

3. Salient interpretations of the action

Though we have just conjectured that linguistic observation will affect listeners' interpretation of scenes, we also mentioned in passing that – syntax all aside – there are bound to be biases in how any event is most naturally represented. Thus in a ball-exchange by rabbit and elephant, its giving turns out to be more salient than its receipt. This differential salience of two interpretations of a scene is the second topic that will be taken up experimentally here.

Specifically, we will consider a distinction in plausibility that we term the 'agency bias'. Even infants appear to be inclined to interpret action scenes as involving a causal *agent* and an affected entity, or *patient* (Michotte 1963, Fritz and Suci 1981, Leslie 1982, Leslie and Keeble 1987, Mandler 1991). This is in preference to an interpretation which excludes reference to the agent. As an example, consider a scene in which a rabbit is feeding an elephant, who eats. The notions 'feed' and 'eat' appear to be equally likely interpretations of a novel verb then uttered. Yet as we will demonstrate, observers usually interpret the novel verb in such a scenario as expressing 'feed', (a causal act) rather than 'eat' (the noncausal option).

By hypothesis, this bias operates as well for the choice between *give* and *receive* even though both these verbs occur in sentences that mention the agent, i.e., in sentence (2) the rabbit is agent of giving and in sentence

(3) the elephant is agent of receiving. But there may be a preference between them all the same. In most *give/receive* scenarios, the *giver* seems more volitional and thus is the plausible candidate for the agent role – the cause, first-mover, or instigator (Dowty 1991).

Our experiment will examine the joint effects of preferences in event representation (the agency bias) and syntactic deduction. Sometimes these factors work together to reveal the verb's meaning. For example, causal agent is linked to subject position in the sentence in all known languages (Clark and Begun 1971, Grimshaw 1981, Bates and MacWhinney 1982, Pinker 1984, Givón 1986, Schlesinger 1988, Dowty 1991). Thus in sentence (5) the listener's event bias will mesh with syntactic deduction. In contrast, sentence (6) pits the two evidentiary sources against each other, for the preferred causal interpretation ('push') is in this case incompatible with the intransitive syntax.[7]

4. Experiment

We have proposed that perspective-changing verb pairs like *give/receive* and *chase/flee* (which are legion in the verb lexicon) pose a special problem for an observational word-mapping scheme and hence offer a useful testing ground for determining whether learners might be sensitive to other kinds of evidence. Specifically, we asked whether young children show sensitivity to syntactic structure in disentangling the senses of such pairs of verbs, as well as other pairs which might be learned via cross-situational observation. To find out, we taught 3- and 4-year-olds novel (nonsense) verbs by using them to describe action scenes. These scenes depicted single events which could be interpreted in two complementary ways.

The manipulations of interest concerned the linguistic context in which the novel verb was presented. A child whose attention is directed to a

[7] For adults, *feed* can occur intransitively, e.g., *The cattle are feeding*. But as predicted, it then is synonymous with *eat*. The wary reader will have noticed, as well, that *eat* often occurs transitively, as in *The cattle eat the fodder*, but does not mean that the cattle cause the fodder to eat. In short, *eat* can drop its object while *feed* can omit its causative subject. Then the two verbs share their licensed syntactic environments, as both can be both transitive and intransitive. It is the positioning of nominals in the structures, as mapped against the scene in view, that reveals the difference in their argument structures (see Levin and Rappaport, this volume, for a discussion of these verb types). The mapping problem can no more be solved by attention to syntax alone than by attention to observation alone. It is the joint operation of the two evidentiary sources that does the work.

relevant scene should be more likely to interpret a novel verb then heard as notionally resembling *give* if it is presented in sentence (2) than if it is presented in sentence (3). Symmetrically, taking a novel verb to resemble *get* or *receive* should be more likely upon hearing sentence (3) than upon hearing sentence (2).

To test these predictions, we needed some way of assessing how our subjects interpreted novel verbs. The method used was a straightforward one: We simply asked the children what they thought the words meant. This method has two main advantages. The first is its very straightforwardness. When children can provide a paraphrase of a novel word, it is unnecessary to attempt to infer from some less obviously relevant aspect of their behavior what they consider the novel word to mean. The second advantage is more central to our question: The essence of the problem we address in this work is that there are pairs of verbs that will be difficult or impossible to differentiate from observational evidence alone. By our own argument there will be no event we could show subjects that would isolate *giving* from *receiving* or *chasing* from *fleeing*. Thus we could not assess what children learned about the novel verbs by, for example, asking them to choose among pictured events (as in Brown 1957). The paraphrase method allowed us to examine just those cases that we have argued pose the knottiest problem for verb mapping.

This method has one disadvantage as well, in that two-year-olds learning first verbs cannot be induced to provide glosses or paraphrases for made-up words, so our youngest subjects are three-year-olds. As we will show, these children are capable of answering the question "What does biffing mean?" in revealing ways. No doubt can arise as to the relevance of three- and four-year-olds to the question posed, for the bulk of the basic verb vocabulary is acquired by these age groups.

4.1. Method

Video-taped scenes were shown to preschoolers and to adult controls. Each scene was described by the experimenter with a sentence that contained a nonsense verb. The subjects' task was to paraphrase the verb.

4.1.1. Subjects

The child subjects were twenty-four 3-year-old children (mean age 3;8, range 3;1–4;0), and thirty 4-year-old children (mean age 4;8, range 4;3–5;0). Nine children (five 3-year-olds and four 4-year-olds) were replaced in the design for failure to respond (see the Procedure section below). Eighteen adults were included to provide a baseline measure of competent performance in this task. A third of the subjects in each age group were randomly assigned to each of three introducing context conditions (see *Procedure*, below).

4.1.2. Stimuli

Six brief motion scenes with puppet actors were video-taped. The scenes were designed to be naturally describable with two English verbs that differed in their semantic and syntactic properties. One of the sentence contexts that could accompany each scene was arbitrarily called the 'X' context, and the other was called the 'Y' context. Descriptions of the scenes and these sentence contexts are shown in table 1.

Table 1
Scenes/sentence pairs

Scene	Sentences
1: A rabbit is feeding an elephant with a spoon.	X: The elephant is ---ing. (eat) Y: The bunny is ---ing the elephant. (feed)
2: A rabbit comes up and pushes a monkey off a box.	X: The bunny is ---ing the monkey. (push) Y: The monkey is ---ing. (fall)
3: A rabbit runs across the screen, followed by a skunk.	X: The bunny is ---ing the skunk. (flee) Y: The skunk is ---ing the bunny. (chase)
4: A monkey is riding piggy-back on a rabbit.	X: The monkey is ---ing the bunny. (ride) Y: The bunny is ---ing the monkey. (carry)
5: An elephant hands a ball to a rabbit.	X: The elephant is ---ing the ball to the bunny. (give) Y: The bunny is ---ing the ball from the elephant. (take)
6: A rabbit puts a blanket over a monkey.	X: The bunny is ---ing the blanket onto the monkey. (put) Y: The bunny is ---ing the monkey with the blanket. (cover)

For the first two scenes in table 1 (*feed/eat* and *push/fall*), the syntax of the two sentences differs in the *number* of noun phrases, i.e., transitive *feed* expresses the causal relationship while intransitive *eat* does not. For the next two scenes (*chase/flee* and *carry/ride*), the number of noun phrases is equal but the *order* of the nouns encodes two perspectives on the event and, consequently, who is the agent.[8]

[8] Notice that *ride/carry* and *chase/flee* don't differ syntactically, i.e., in their subcategorization frames; both appear in simple transitive sentences. Moreover, both members of these pairs have

In the last two scenes (*give/receive* and *put/cover*), the sentences for the two standard choices differ in the *order* of NP arguments in the sentence as well as in the *preposition* used to mark the indirect object (*to* vs. *from* and *onto* vs. *with*). These last two cases can be subdivided into one pair relevant to the choice of agent (*give/receive*), thus subject (elephant vs. rabbit), and one pair relevant to the choice of goal and located object (*put/cover*), thus direct object (blanket vs. monkey). Thus the stimuli overall can provide some indication of the kinds of syntactic-semantic linkages that young children can recruit for verb mapping.

Each sentence structure was randomly paired with one of six nonsense syllables for each subject (*zike, blick, pilk, dack, moke, nade*). All sentences were presented with the verb in the progressive form (*blicking*) to maximize intelligibility and pragmatic felicity as descriptions of ongoing actions.

4.1.3. Procedure

Two experimenters tested each child individually; one showed the video-tapes and uttered the stimulus sentences, while the other recorded the subjects' responses. The sessions were also audio-taped, to allow later checking of the accuracy of the recording experimenter.

A puppet (Mac) was introduced to the child who was then told "Mac doesn't speak English very well, so sometimes he uses puppet words. Can you help us figure out what the puppet words mean?". Assent received, the child was then given a practice trial in which Mac said "Look! The elephant is zorping!" as an experimenter made a hand-held elephant puppet laugh. The child was then asked, "What does zorping mean?" and prompted by asking "What is the elephant doing?" This latter prompt was used only in the practice trial, to help the child understand the task. For the adult subjects, Mac the puppet was omitted for obvious reasons. They were simply informed that their task would be to guess the meanings of nonsense words.

On each test trial, the subject first heard the stimulus sentence and then saw the video-taped scene. The scene was repeated for up to one minute as the subject watched. The experimenter repeated the stimulus sentence at least

---

the same thematic-role assignment to syntactic position: the subject is agent, the object is theme (or patient). What we mean by 'attention to the syntax' in solving the mapping problem for the members of these pairs is that the observer's conjecture about whether the verb means 'chase' or 'flee' is consequent on noticing which observed entity receives the subject-agent slot. As we have emphasized earlier, this is one reason why a priorly acquired nominal vocabulary is prerequisite to acquiring verb meanings.

once while the videotape played. The subject was asked "What does gorping mean?". Additional prompts were "What's happening?" or "What's going on?". Subjects were encouraged to guess. Subjects who said nothing during the practice trial and during the first two experimental trials were dropped from the study. Scenes were presented in two orders, chosen to allow all subjects to begin with less difficult items (as revealed during pilot testing). The first order was *feed/eat, give/receive, chase/flee, cover/put, carry/ride, push/fall.* The second order was its reverse.

The nonsense verbs were presented to the subjects in one of three linguistic contexts:

Neutral syntax ('No sentence') trials. A third of the subjects in each age group heard the nonsense words in the syntactically uninformative context "Look! Ziking!". These subjects were cued (by the *-ing* suffix, Brown 1957) that the novel word was a verb but received no information about its specific syntactic behavior. Even if young children use syntactic evidence in real-life verb learning, this experimental condition withheld such evidence, allowing us to assess any biases subjects might show in interpreting the scenes and providing a baseline against which to compare their performance when given syntactic information.

Sentence trials. The remaining two-thirds of the subjects saw the scenes accompanied by one of two introducing sentential contexts (X or Y in table 1, for each scene). Each subject heard only one of the two sentences for a single scene. One group of subjects heard the sentences designated 'X' in table 1, and the other group heard those designated 'Y'. We reiterate that the assignment of sentences to the categories X and Y was entirely arbitrary. Thus, e.g., transitive *feed* was assigned to the Y category while transitive *push* was assigned to the X category. This arbitrary assignment assured that each subject group heard some sentences of both kinds.

4.1.4. Coding and scoring

Subjects' paraphrases were sorted into three categories:

(1) *Response X*: Responses that fit sentence X (shown in table 1), both syntactically and in describing the scene, were coded as X responses. For example, for scene 1, *eat*, or a phrasal equivalent ("He's drinking soup"), is congruent with the scene and with the construal implied by the structure of sentence X. Notice that a subject might be exposed to a Y sentence (*The bunny is gorping the elephant*) but give the X response all the same. If so, the syntax of the input sentence has failed to influence that response.

(2) *Response Y*: Responses that fit sentence Y (table 1), both syntactically and in describing the scene, were coded as Y responses. For example, for scene 1, *feed*, or a phrasal equivalent ("He's giving him medicine") would be congruent both with the scene and with the (causative) construal implied by sentence Y of this pair.
(3) *Other (O)*: Failures to respond and responses that fit neither sentence X nor Y were coded as Other. These were almost always relevant to the scene in some way (e.g., the response "They're playing" for the *give/receive* scene), but incongruent with the ditransitive syntax. Also included in this category were responses that fit *both* X and Y contexts, and therefore could not demonstrate sensitivity to the sentence structure.

4.2. Results

To assess the reliability of the coding, all of the responses from the 3-year-old subjects were recoded by an independent coder who was blind to introducing context. The two coders agreed with each other in 94% of coding decisions. Residual disagreements were resolved through discussion. The subjects' responses are summarized in table 2. This table shows the proportion of subjects in each age group who produced each type of response for each scene. We now discuss these findings under several rubrics:

4.2.1. Breadth of the hypothesis space

The scenarios were designed to be quite simple, containing few distracting properties. Even so, as table 2 shows, child subjects gave a response that had to be relegated to the 'Other' category in a substantial proportion of the trials (29% across all conditions for 3-year-olds, 23% for 4-year-olds). This finding constitutes yet one more demonstration of the many–many relations between scene observation and verb interpretation. In several cases coded 'Other', the subjects explicitly mentioned both of the standard choices (e.g., *giving* and *getting*). These responses were quite common in the 'No sentence' condition (36% of adults' and 20% of children's total responses) but extremely rare in the sentence contexts (3% of adults' and 5% of children's responses). When a sentence context was provided, both children ($t(52) = 3.64$, $p < 0.001$) and adults ($t(16) = 5.89$, $p < 0.001$) were much less likely to propose both readings for the novel verb. This is a first indication that the syntactic contexts rein in the many interpretations made available by scene inspection, focusing subjects' attention on specific aspects of our scenarios.

Table 2
Proportion each response for each scene, by age and introducing context

Scene	'No sentence'			Sentence X			Sentence Y		
	X	Y	Other	X	Y	Other	X	Y	Other
Three-year-olds (n = 24):									
eat/feed	**0.25**	**0.38**	0.38	**0.50**	**0.13**	0.38	**0.25**	**0.75**	0.00
push/fall	**0.38**	**0.13**	0.50	**0.88**	**0.13**	0.00	**0.25**	**0.75**	0.00
flee/chase	**0.00**	**0.63**	0.38	**0.38**	**0.13**	0.50	**0.00**	**0.88**	0.13
ride/carry	**0.25**	**0.75**	0.00	**0.63**	**0.25**	0.13	**0.25**	**0.50**	0.25
give/take	**0.50**	**0.25**	0.25	**0.50**	**0.00**	0.50	**0.25**	**0.25**	0.50
put/cover	**0.25**	**0.13**	0.63	**0.38**	**0.25**	0.38	**0.38**	**0.25**	0.38
Mean	**0.27**	**0.38**	0.35	**0.54**	**0.15**	0.31	**0.23**	**0.56**	0.21
Four-year-olds (n = 30):									
eat/feed	**0.10**	**0.70**	0.20	**0.60**	**0.20**	0.20	**0.10**	**0.80**	0.10
push/fall	**0.70**	**0.00**	0.30	**1.00**	**0.00**	0.00	**0.50**	**0.40**	0.10
flee/chase	**0.00**	**0.70**	0.30	**0.40**	**0.10**	0.50	**0.00**	**0.90**	0.10
ride/carry	**0.00**	**0.80**	0.20	**0.40**	**0.20**	0.40	**0.20**	**0.70**	0.10
give/take	**0.40**	**0.10**	0.50	**0.50**	**0.00**	0.50	**0.30**	**0.60**	0.10
put/cover	**0.20**	**0.20**	0.60	**0.40**	**0.20**	0.40	**0.20**	**0.70**	0.10
Mean	**0.23**	**0.42**	0.35	**0.55**	**0.12**	0.33	**0.22**	**0.68**	0.10
Adults (n = 18):									
eat/feed	**0.00**	**0.17**	0.83	**1.00**	**0.00**	0.00	**0.00**	**1.00**	0.00
push/fall	**0.50**	**0.00**	0.50	**1.00**	**0.00**	0.00	**0.00**	**1.00**	0.00
flee/chase	**0.00**	**0.83**	0.17	**1.00**	**0.00**	0.00	**0.00**	**1.00**	0.00
ride/carry	**0.17**	**0.33**	0.50	**0.83**	**0.17**	0.00	**0.00**	**1.00**	0.00
give/take	**0.50**	**0.00**	0.50	**1.00**	**0.00**	0.00	**0.00**	**1.00**	0.00
put/cover	**0.00**	**0.83**	0.17	**0.50**	**0.17**	0.33	**0.00**	**1.00**	0.00
Mean	**0.19**	**0.36**	0.44	**0.89**	**0.06**	0.06	**0.00**	**1.00**	0.00

4.2.2. Narrowing the hypothesis space by attention to the syntax

Each of the standard responses was more frequent in the matching sentence context than in either of the other two introducing contexts. For example, subjects gave more 'give'-like responses to *give* syntax than to *receive* syntax or to the neutral condition. That is, X responses were more likely in the context of sentence X than sentence Y (children: $t(34) = 5.04$, adults: $t(10) = 14.56$, $p < 0.001$) or the 'No sentence' context (children: $t(34) = 4.19$, adults: $t(10) = 5.81$, $p < 0.001$). Similarly, Y responses were more frequent in the context of sentence Y than sentence X (children: $t(34) = 8.83$, adults: $t(10) = 16.14$, $p < 0.001$) or 'No sentence' (children: $t(34) = 5.16$, adults: $t(10) = 32.73$, p

< 0,001).[9] The effect of syntax on the interpretation of these nonsense words emerges in the same way for each scene with one exception (the 3-year-old subjects paraphrased a nonsense verb describing *put/cover* consistent with *put* syntax more often than *cover* syntax in all three introducing contexts). Overall, the result is that syntactic context had a powerful effect on subjects' construal of the nonsense verb.

The children's behavior is epitomized most poignantly in several contrasting responses to the scenes in the presence of syntactic context. Consider the *carry/ride* scene. Hearing the sentence that treats the carrier (rabbit) as subject, a child responds "He's holding him on his back". But hearing the sentence that treats the rider (monkey) as subject, another child responds "He's sitting on him's [sic] back". Clearly these children watch the scene and interpret what they see to derive the verb meaning: They induce the meaning from inspection of its real-world accompaniments. Perhaps unlike the adults, they do not seem to be aware of a linguistic puzzle. All the same, the sentence heard exerts a strong – albeit implicit – influence on just what they think the scene depicts: whether it is 'about' the one who holds/carries or the one who sits/rides. This is the sense in which sentence-to-world pairing can sharply limit the search-space for verb identification.

4.2.3. Semantic biases in the interpretation of verbs

In addition to constituting a baseline for comparison to the sentence trials in our task, the 'No sentence' condition provides a chance to look for semantic biases in verb interpretation. After all, in observing the *chase/flee* scene while hearing "Look! Ziking!", subjects are not really warranted in preferring one of these interpretations over the other. The scene fits them both. Our initial hypothesis was that subjects would therefore choose *chase* or *flee* ("run away") more or less at random in this condition, hence our original arbitrary division of the sentences (and their responses) into the 'X' and 'Y' categories. But our subjects were anything but open-minded in their guesses, as table 2 shows. All age groups had a preferred response in the 'No sentence' condition for 5 of the 6 scenes. They tended to

[9] Preliminary analyses indicated that there was no effect of order (which of the stimuli the subjects saw first) on the probability of sentence-congruent responses, so the two order groups were combined in this and all further analyses. The effect of syntax on verb paraphrases was shown in a series of planned 1-tailed *t*-tests. In each case the dependent variable was the proportion (arcsine transformed) of each response (X or Y), examined across sentence contexts. Separate analyses for the 3- and 4-year-olds yielded the same results in each case; to simplify presentation of the results, the two groups of children are pooled in the analyses presented here.

describe scene 1 as one of *feeding* rather than *eating*, scene 2 as *pushing* rather than *falling*, scene 3 as *chasing*, not *fleeing* or *running away*, scene 4 as *carrying* rather than *riding*, and scene 5 as *giving* rather than *taking* or *receiving*. The adults showed these biases even more regularly than the children.

In each of these cases, subjects evidently selected a 'more causal' or agentive participant in the scene, and took the verb to code the actions of that participant, who thus became sentential subject. Scene 6 (*put/cover*) leaves the agent choice unaffected and, interestingly enough, it is in this case only that children show no preference of choice in the 'No sentence' condition. Clearly, this unanticipated agency factor was strongly affecting our subjects' responses.

To study this semantic bias, we now recoded the X and Y sentences according to whether they matched or mismatched this agency bias. For *feed/eat* and *push/fall*, no difficulty in doing so arises: Only the *feed* and *push* sentences mention the causal agent so these interpretations should be favored if there is a bias to conceive scenes as an agent acting on a thing affected. For *chase/flee* and *give/receive*, this agent–act–patient interpretation applies to both members of the pairs. For these, we relied on Dowty's (1991) 'proto-agent' classification scheme in which several factors are postulated to lead to the choice of plausible agent: animacy (which does not distinguish for our stimuli), activity, and instigator of the action. Thus necessarily it is chasing that precedes and causes fleeing, and giving that precedes and causes receiving.

For the case of *carry/ride*, no such principled distinction in the verb meanings themselves exists for choosing the plausible agent. However, in our particular depiction of this act, the carrier (the rabbit) was actively running across the screen with an inert monkey sitting on his back. Apparently this distinction of relative activity vs. passiveness led subjects to choose the rabbit as instigator, thus plausible agent. (Had our scene instead shown, say, a child riding a mechanical bull, doubtless this choice would have been reversed for this verb pair.)

Table 3 reorganizes the findings according to this distinction between Agentive (A) versus NonAgentive and 'less plausible agent' (NA) responses.[10] The table shows the proportion of subjects in each age group and intro-

[10] Because the agency effect was unanticipated, the experiment was not balanced such that each subject would hear an equal number of A and NA sentences. So that this difference will not contaminate the statistical assessments here, all comparisons between the A and NA contexts are within-subjects. All other comparisons are between subjects, as before.

ducing context who produced A (e.g., *feed*, *carry*) and NA (e.g., *eat*, *ride*) responses, as defined above. These values are shown only for the five relevant scenarios (that is, excluding the *put/cover* scene which cast the same participant as subject in both context sentences). Across all presentation conditions and all age groups, A responses (187) outnumber NA responses (91) two to one. There is an agency bias.

Table 3
Proportion[a] agentive (A) and non-agentive (NA) responses for each biased scene, by age and introducing context

Scene		No sentence		Agentive		Non-agentive	
A	NA	A	NA	A	NA	A	NA
Three-year-olds (n = 24):							
feed	eat	0.38	0.25	0.75	0.25	0.13	0.50
push	fall	0.38	0.13	0.88	0.13	0.25	0.75
chase	flee	0.63	0.00	0.88	0.00	0.13	0.38
carry	ride	0.75	0.25	0.50	0.25	0.25	0.63
give	take	0.50	0.25	0.50	0.00	0.25	0.25
Means:		0.53	0.18	0.70	0.13	0.20	0.50
Four-year-olds (n = 30):							
feed	eat	0.70	0.10	0.80	0.10	0.20	0.60
push	fall	0.70	0.00	1	0.00	0.50	0.40
chase	flee	0.70	0.00	0.90	0.00	0.10	0.40
carry	ride	0.80	0.00	0.70	0.20	0.20	0.40
give	take	0.40	0.10	0.50	0.00	0.30	0.60
Means:		0.66	0.04	0.78	0.06	0.26	0.48
Adults (n = 18):							
feed	eat	0.17	0.00	1.00	0.00	0.00	1.00
push	fall	0.50	0.00	1.00	0.00	0.00	1.00
chase	flee	0.83	0.00	1.00	0.00	0.00	1.00
carry	ride	0.33	0.17	1.00	0.00	0.17	0.83
give	take	0.50	0.00	1.00	0.00	0.00	1.00
Means:		0.47	0.03	1.00	0.00	0.03	0.97

[a] Since 'Other' responses are left out of this table, proportions within each scene/context cell may not sum to 1.

Not surprisingly, this effect of the natural interpretation of the scenes is most powerful where it is not contaminated by mismatching syntactic information. Thus the preference for agentive responses appears most strongly in the 'No sentence' condition in table 3. This effect was assessed in post-hoc (Bonferroni adjusted) matched-pairs *t*-tests, which revealed that A choices were more frequent than NA ones in the 'No sentence' context for both children ($t(17) = 6.18$, $p < 0.001$) and adults ($t(5) = 5.93$, $p < 0.01$).

4.2.4. The interaction of syntax and semantics

So far we have shown (table 2) that structure affects Ss' guesses when it is made available, in the X and Y (sentential) conditions of presentation; and that the agency bias (for which responses were recoded as A and NA) affects Ss' guesses heavily in the absence of syntactic information (the 'No sentence' condition of table 3). The question remains how these factors interact when (as in the real world of maternal speech) the new word is presented in a full sentence context.

To find out, we now look more closely at the finding (table 2) that, while the effects of syntactic context were very strong and reliable, they were nowhere near categorical for the child subjects (though they were nearly so for the adults). The reason is that while the syntactic introducing circumstances supported the agency bias in some cases (as when the subject heard "The rabbit is biffing the elephant" while watching a *feed/eat* scene), the syntactic and semantic cues pulled in opposing directions in other cases (as when the subject heard "The elephant is biffing" while watching this same scene). To repeat, subjects tend to view this scene as a feeding scene (the semantic-interpretive influence), but to encode the intransitive sentence structure as favoring *eat* (the syntactic influence).

The joint action of these two variables can be seen in table 4, which repeats the data of table 3, but summarizing across the five relevant scenarios. Thus table 4 shows the proportion of subjects in each group and introducing context who produced A and NA responses. Column 1 of this table shows the bias toward agentive responses which we have just documented. Column 2 shows a slight enhancement of this preference (slight, because the preference in this direction was already strong) when it is supported by the syntactic evidence; that is, in agentive syntactic contexts. Column 3 shows that the agency bias is heavily mitigated by mismatching syntactic evidence: When the syntax demands the NA interpretation, the A response is actually dispreferred, no longer the modal response for any age group. That is, we see the same effect of sentence context on verb interpretation when the normally disfavored non-agentive responses are examined separately: These responses are more frequent in the non-agentive than the agentive sentence context (children: $t(35) = 7.13$, adults: $t(11) = 29.63$, $p < 0.001$). Both children and adults made use of syntactic evidence to override what seems to be a strong semantic or observational preference in verb interpretation.

Table 4
Proportion[a] agentive (A) and non-agentive (NA) responses, by age and introducing context

Age group	Introducing context					
	No sentence		Agentive		Non-agentive	
	A	NA	A	NA	A	NA
Threes	0.53	0.18	0.70	0.13	0.20	0.50
Fours	0.66	0.04	0.78	0.06	0.26	0.48
Adults	0.47	0.03	1.00	0.00	0.03	0.97
Overall	0.57	0.08	0.81	0.07	0.18	0.61

[a] Since 'Other' responses are left out of this table, proportions within each scene/context cell may not sum to 1.

4.2.5. *The effect of age*

Inspection of table 4 also shows that there is an age effect on these response patterns, with adult responses almost categorical and child responses probabilistic with respect to the variables under investigation. The effect of age on the likelihood of producing frame-congruent responses in the two sentence contexts was shown in an ANOVA with age group as a between-subjects effect, and sentence context (A versus NA) as a within-subjects factor. The main effect of age was significant ($F(2,45) = 17.36$, $p < 0.001$): adults were more likely than children to produce a response that fit the frame, but the two groups of children did not differ significantly from each other. There was also a significant main effect of sentence context ($F(1,45) = 5.33$, $p < 0.05$): Frame-congruent responses were more likely in the A than the NA context for all age groups, the interaction of situation and syntax that we discussed earlier. However, the effects of age and sentence context did not interact ($F(2,45) < 1$). Thus the effect of age is a simple one – adult performance was more stable with regard to both the syntactic and salience variables, but all groups took syntactic information into account when it conflicted with the agency bias as well as when it did not.

4.2.6. Summary of the findings

The findings support the claim that a scene observed is insufficient for fixing the meaning of a new verb, for subjects' construal was often one that mismatched the interpretations we had in mind. But in these stripped-down puppet scenarios, the kinds of interpretation we had expected were the most frequent ones for each age group and for each scenario (table 2). A bias in event representation influenced the subjects' choice of the plausible 'doer' vs. 'done-to' in these scenes, independent of the syntactic introducing contexts (tables 3 and 4). But the most powerful influence on the choice of response in this regard was the structure of the sentence heard, for the sentences that pitted syntax against semantics reversed what had been subjects' modal preference in the neutral contexts (tables 2 and 4). As we hypothesized, the syntactic structure led to a principled choice of interpretation among those that were compatible with the scene in view.

4.2.7. An alternative interpretation of the findings

Before discussing a theory of verb mapping that comports with these findings, we digress to consider a persistent difficulty in interpreting experimental studies of word-learning. It is hard to rule out the hypothesis that subjects are relying solely on preexisting lexical knowledge in such tasks (Pinker this volume). That is, children could take the nonsense words to be synonyms for existing words, and respond with *eat* when appropriate in our task because they already know both the meaning and the syntax of that particular lexical item, rather than by interpreting the novel term in a syntax-sensitive learning procedure. The standard remedy for this problem is to teach young children new words for novel objects or actions (e.g., Markman 1989, Gropen et al. 1991). Many studies of syntactic bootstrapping use this strategy as well (Naigles 1990, Hirsh-Pasek et al. 1988).

In the current experiment, even though it allowed for the possibility of glossing the nonsense term with a known verb, our child subjects rejected this option in a substantial percentage of cases. Thirty-five percent of the 3-year-olds' and 48% of the 4-year-olds' responses were phrasal descriptions rather than single words (e.g., "licking it off the spoon" instead of "eating", "giving him medicine" rather than "feeding", "trying to smell him" rather than "chasing"). This is not to say that these children didn't know the words *feed*, *eat*, and *chase*. Rather, they seem to have taken the novel verbs to have specific meanings that were not equivalent to known verbs. This outcome is in line with much current work on the uses of lexical contrast in the acquisition of word meaning. Young children seem determined to assign new meanings to new words. They do not expect there to be exact synonyms. (For this effect with nouns, see e.g., Clark 1987, and Markman, this volume; and for the analogous effects with verbs, Golinkoff et al., in prep.; Kako, in prep.)

We now ask whether the same pattern of syntactic effects on verb interpretation occurs for the phrasal responses taken alone. If so, it is unlikely that the children were merely 'filling in the blank' in the stimulus sentences with a known verb that fit, but were engaging in something more like the syntax-guided mapping procedure proposed here.

Table 5 shows the obtained response patterns for the phrasal responses, omitting all single-verb glosses. Three 3-year-olds and one 4-year-old contributed no data to this analysis. Given the reduction in the data entailed by recoding the responses in this way, the 3- and 4-year-olds' data were combined. Inspection of table 5 shows that the same pattern of structural effects on verb interpretation remains. This pattern was reliable in independent t-tests on the (arcsine transformed) proportion of X and Y responses. Response X was more likely in the context of sentence X than sentence Y ($t(30) = 2.94$, $p < 0.01$) and the 'No sentence' context ($t(32) = 4.11$, $p < 0.001$). Similarly, response Y was more frequent given sentence Y than either of the other two introducing contexts (sentence X: $t(30) = 6.59$, $p < 0.001$; no sentence: $t(32) = 3.61$, $p < 0.001$). In sum, even when the children evidently did not interpret the nonsense words as puppet synonyms for verbs in their known lexicon, there are strong and reliable effects of sentence structure on construal.

5. Discussion

We now outline a verb mapping procedure that comports with the findings just described, and with related effects in the language acquisition literature.

5.1. Extraction of linguistic formatives

A vexed problem in understanding language learning concerns how the child finds such units as word and phrase in the continuously varying sound stream: how Henny-Penny's listeners interpreted her as saying "The sky is falling" rather than "This guy is falling". A vast literature now supports the

Table 5
Proportion of X and Y responses by introducing context; phrasal responses only (3- and 4-year-olds together, $n=50$)

Scene	'No sentence' ($n=18$)		Sentence X ($n=16$)		Sentence Y ($n=16$)	
	X	Y	X	Y	X	Y
eat/feed	0.00	0.17	0.13	0.00	0.00	0.13
push/fall	0.33	0.06	0.75	0.00	0.31	0.50
flee/chase	0.00	0.17	0.38	0.06	0.00	0.56
ride/carry	0.11	0.28	0.38	0.06	0.25	0.50
give/take	0.00	0.11	0.13	0.00	0.00	0.19
put/cover	0.00	0.06	0.00	0.13	0.13	0.25
Mean	0.07	0.14	0.29	0.04	0.11	0.35

view that these segmentation decisions are made by infants based on prosodic and distributional cues in caretaker speech (for recent evidence and discussion, see Gleitman et al. 1988, Gerken et al. 1993, Fisher and Tokura, in press; Kelly, this volume; Cutler, this volume; Brent, this volume).[11] In what follows, we presuppose these approaches to solving the segmentation problem, concentrating attention on acquisition of the phrase structure (which requires labeling as well as segmentation of phrases) and the word meanings.

5.2. Word-to-world pairing and the acquisition of first nouns

We take as given that the human learner expects sentences to convey predicate/argument structure as organized by a phrase-structure grammar that conforms to X-bar principles. However, universal grammar leaves open some parameters of the phrase structure of the exposure language; these must be set by experience. Before this learning occurs, the novice can recruit only the exigencies of word use – the pairing of words to their extralinguistic contexts – to solve the mapping problem.

[11] For evidence on infant attention to clause bounding cues see Hirsh-Pasek et al. 1987; for phrase-bounding cues, see Jusczyk et al. 1992; and for word-bounding cues see Grosjean and Gee 1987, Kelly, this volume. There is also a literature on adult speech production and perception investigating the physical bases of these prosody-syntax mappings, too vast for us to cite here (but for seminal articles, see Klatt 1975, Cooper and Paccia-Cooper 1980, Lehiste et al. 1976, Cutler, this volume) and the availability of such cues in infant-directed speech (Fernald and Simon 1984, Fisher and Tokura, in press; Lederer and Kelly 1991).

By default, the effect should be that youngest children can acquire only object terms (nouns, in the adult language). This is because, as we described in introductory remarks, only the nouns occur in maternal speech in a tight time-lock with the situational contexts, and in ostensive contexts. And indeed one of the most striking findings in the language-learning literature is that first words are nouns despite the fact that from the beginning the learner is exposed to words from every lexical category. We assume that these first words are assigned to the formal category *noun* on a semantic basis, as conjectured by Grimshaw (1981) and Pinker (1984).

5.3. Setting the phrase structure, and first verbs

Our findings suggest that verb learning implicates a sentence-to-world pairing procedure and cannot in general be accomplished, as Pinker (1984) and others have advocated (see footnote 3), by pairing the isolated verb to its observational contingencies. But then how much grammatical knowledge is required as input to verb learning? And how might children acquire this?

We propose that the meanings of the first verbs, as well as relevant components of the phrase structure, are acquired by bootstrapping from a partial sentential representation (henceforth, *PSR*) that becomes available once some nouns have been learned: This consists of the known nouns and the unknown verb, as sequenced in the input sentence, e.g.

[... baby ... eat ... cookie ...]

By hypothesis, it is this richer-than-the-word, poorer-than-the-phrase-structure, representation that learners past the one-word (noun) stage first attempt to pair with the scene in view.

There is some evidence that the PSR can aid in verb identification in two ways. First, the identity of the nouns can provide information about the selectional properties of the verb. Lederer et al. (1991) showed that adults can identify about 28% of the verbs that mothers are uttering if, in addition to the scene information, they are also told which nouns occurred with the verb in the maternal utterances. This level of performance is not great, but is a significant improvement over the 7% success rate that subjects achieve if shown only the video-taped scene. It is easy to see why having the nouns is so helpful: If you are told that *baby* and *cookie* occurred in construction with the mystery verb, *eat* becames a plausible conjecture just because verbs that mean 'eat' should select for edibles.[12]

The PSR yields a second advantage for verb mapping, provided that the learner also has implicit access to the Projection Principle (roughly, that every argument position required by the verb will be reflected as a noun phrase in the surface sentence; Chomsky 1981). So armed, the learner can make a secure conjecture as between an intended unary relation (such as *fall*) vs. a binary relation (such as *push*), simply by counting the number of noun phrases in the sentence.[13]

Early use of this machinery is suggested by findings from Naigles (1990), with babies 23–25 months of age. They were shown a video-taped scene in which (a) a duck who by pushing on a rabbit's head forces the latter into a squatting position whilst (b) both the duck and the rabbit wheel their free arms in a circle. Half the subjects were introduced to the scene with the sentence "The duck is gorping the rabbit" and the other half heard "The duck and the rabbit are gorping". Thereafter, two new videos were shown, one to the child's left and one to her right, along with the prompt "Find gorping now!". One of the new videos showed the duck forcing the rabbit to squat (but no arm-wheeling) and the other showed the two side by side wheeling their arms (but no forcing-to-squat). The children who had been introduced to gorping within the transitive sentence now gazed longest at the causal scene while those who had heard the intransitive sentence looked longer at the noncausal scene. Here, as in the *push/fall* and *feed/eat* scenes investigated in the present experiment, the one-argument structure did not substain a causal interpretation despite the agency bias in event representation.

A related point is made by Fisher (1993), who showed unfamiliar agent–patient events to children aged three and five years but with the entities named only by pronouns. For instance, they saw one person causing another to rotate on a swiveling stool by alternately pulling on the ends of a scarf around the victim's waist. Half the subjects heard "She's blicking her around" and the other half heard "She's blicking around". The child's task was to point out, in a still photograph of the event, the one whose action was

[12] We must acknowledge, however, that the Gleitman-Landau archive of maternal speech includes many examples like "Don't eat that paper!", "We don't eat the book, Bonnie".

[13] Of course this can only work for simple sentences for, e.g., "*The rabbit* in *the grass* hopped away" contains two NPs within an argument position. But sentences to novices are characteristically short (approximately 5 words long, on average; Newport 1977) and so rarely embody this problem.

labelled by the novel verb ("Point to the one who's blicking the other one around" or "Point to the one who's blicking around"). The intent here was to put the children into the position of much younger learners who have access only to the PSR: They knew how many arguments were supplied to the new verb, but not which was which. Those who had heard the transitive frame confidently chose the causal agent as the blicker, while those shown the intransitive frame were willing to select the patient as blicker. Thus without being told which event participant has been cast as sentence subject, preschoolers interpreted a one-argument structure as incompatible with a causal interpretation.

The PSR taken together with the scene observed will allow learners to acquire a crucial aspect of the phrase structure itself. Suppose a child hears "kick" for the first time in the frame "The bunny kicked the monkey". The agency bias, as constrained by the minimal structure given by the PSR (namely, a 2-argument structure), will lead the learner to seek an agent–patient interpretation of the scene. Provided that she knows the nouns *bunny* and *monkey*, she can annotate the phrase structure as shown in figure 1, marking *bunny*, the first noun in the structure, as the agent (Joshi and Rambow, in prep.). This representation now matches two of the quasi-universal properties of the category *subject of transitive sentence* – there is one noun which is both the agent and the leftmost noun in a transitive structure.[14] There is a strong tendency for languages to place subjects before objects (Keenan 1976, Kayne 1992). Further phrase-structure options can then be set based on this initial assignment.

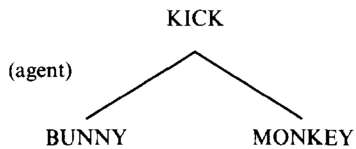

Fig. 1

[14] Notice that this claim differs from the subject-agent link which has been invoked in the literature to support an initially asyntactic verb-learning procedure (e.g., Grimshaw 1981, Braine and Hardy 1982, Pinker 1984, and others). In fact, subjects are often patients or experiencers. It is the *transitive* sentence whose agent just about universally surfaces as subject. To recognize this distinction requires, at minimum, PSR knowledge that will reveal the number of argument positions.

Notice finally that if there are languages in which objects precede subjects, the assignment of *subject of the sentence* (the NP immediately dominated by S) to the serially second NP is still possible, based on the PSR. In that case, the learner would have heard "Kicked the monkey the bunny" in the presence of a bunny-kicking-monkey scene. The observed scene identifies *bunny* as agent (hence subject of the transitive verb), consistent only with template (b) of figure 2.

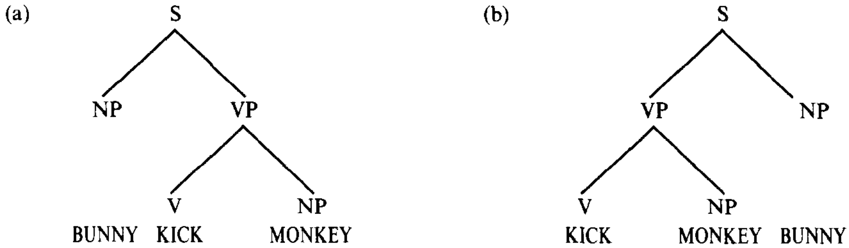

Fig. 2. (a) an SVO phrase-structure template; (b) a VOS phrase-structure template.

Several other kinds of cue to the phrase-structure of the exposure language have been suggested, and have plausibility. For instance, Mazuka (1993) suggests that branching direction in Japanese can be deduced by observing intonational markings of moved relative clauses. Further, subject and object NP's vary strongly in maternal speech (the former is usually a monosyllabic pronoun and is often omitted even in non-prodrop languages like English; Fisher and Tokura, in press; Gerken et al. 1993), and learners may be sensitive to such probabilistic patterns (for discussion, see Kelly, this volume).

Hirsh-Pasek and Golinkoff (1991) showed that 17-month-old children who utter only isolated nouns appreciate at least the rudiments of English word order and the way this maps onto thematic roles. If they heard a voice saying "Big Bird tickles Cookie Monster" they looked primarily at a video screen depicting this event, but if the voice said "Cookie Monster tickles Big Bird", they looked at another video screen which depicted this latter event. These toddlers evidently were sensitive to (something like) the idea that the doer of the action is the subject of the transitive sentence; and that in English the serially first noun is that subject.

5.4. Structural information narrows the search space for verb mapping

We have so far argued that PSR taken together with scene representation allows the child a bootstrap into the phrase structure of the exposure language. Once the full phrase structure is acquired, the learner can approach the perspective-changing verbs that we have studied. Disentanglement of the members of these pairs requires more than counting NP positions (which are the same for both interpretations) and fitting these to the logic of the observed situation (which suits either choice).

Once the phrase structure has been bootstrapped from PSR, the learner can make this decision by inspecting the geometry of the tree to determine which noun is sentence subject. If the plausible agent appears as subject with the *give/receive* scene, then the situational and syntactic cues converge on *give*. But if the plausible agent appears in nonsubject position then it is not the agent, despite appearances. In the present experiment, we showed child responsiveness to these implications of structure for verb interpretation even in cases where they had to overcome a semantic bias in event interpretation to use it.

5.5. The informativeness of multiple frames

In principle, attention to the licensed *range* of syntactic environments for a verb can provide converging evidence about its interpretation, just because these several environments are projections from the range of argument structures associated with that verb. This feature of syntactic bootstrapping is controversial (as opposed to the 'zoom lens' notion which appears to have gained wide currency). So after describing the potential usefulness of frame-range information for solving the mapping problem, we will discuss available experimental evidence in its favor. Specifically, we will discuss the experimental manipulations deemed critical by some skeptics (particularly, Pinker, this volume) for confirming the hypothesis.

5.5.1. The resolving power of frame ranges for verb mapping

In very many cases, a surface-structure/situation pair is insufficient or even misleading about a verb's interpretation. One such case is the *eat* example that we mentioned earlier. The phrase structure is the same when the adult says "Did you *eat* your cookie?" as when he says "Do you *want* the cookie?", and the two verbs are used by caretakers in situations where their interpretations can easily be mistaken. Subjects always come up with an action term that fits the observed scene and the structure, and guess *eat* instead of *want*. In response to the fact that the next scene observation does not support the *eat* conjecture (it may show, say, the mother offering a toy rather than a cookie to the child), subjects now come up with yet another physical term (e.g., *take*). They are mulishly resistant to conjecturing any mental term. While successive observations force them to change their minds about *which* physical-action term is the right one, they never seem to get the idea that this bias should be overridden altogether. This effect was shown by Lederer et al. (1991) with adults, and by Gillette (1992) with children.

Examination of the further syntactic privileges of *eat* and *want* can resolve this problem. *Eat* occurs intransitively and in the progressive form while *want* does not. *Want* also occurs with tenseless sentence complements (*Do you want to eat the apple?*). These distinctions are sufficient to disentangle the two verb construals, for only mental activity verbs license these constructions. (Note that *force* verbs, which also accept sentence complements, require an additional nominal position, e.g., *Make him eat the apple!*, but not **Make eat the apple!*). As we will discuss presently, subjects seize upon this disambiguating structural information to find the right construal.

A second example of residual problems unresolved by single frames, even though these are paired with differing events, concerns a blind child's learning of the distinction between *touch* and *see*. Blind learners receive observational evidence about the meanings of both terms though it is perforce haptic and not visual. The blind child's first uses of *see* at age two were in the sense 'touch', e.g., she commanded "Don't see that!", while pushing her brother away from her record-player. The confusion arose, doubtless, because every scene in which the blind child can see ('ascertain by perceptual inspection') is a scene in which she can touch. And both verbs occur most often in maternal speech as simple transitives. Further syntactic experience ("Let's see if there's cheese in the refrigerator") can account for how the blind child could, as she did, come to distinguish between the two construals by age three (Landau and Gleitman 1985).

As a more general example of the convergence that frame ranges make available for verb mapping, consider the four verbs *give*, *explain*, *go*, and *think*. These verbs are cross-classified both conceptually and syntactically. *Give* and *explain*, different as they are in many regards, both describe the transfer of entities between two parties. Accordingly, they can appear in structures with three noun-phrase positions:

(7) Ed gave the horse to Sally.
(8) Ed explained the facts to Sally.

In (7), a physical object (the horse) is transferred from Ed's to Sally's hand and in (8) abstract objects (the facts) are transferred from Ed's to Sally's mind. A noun phrase is required for each of the entities involved: the giver, the receiver, and that which is transferred between them. It is this similarity in their meanings that accounts for the similarity in the structures that they accept. Verbs that describe no such transfer are odd in these constructions:

(9) *Philip went the horse to Libby.
(10) *Philip thought the facts to Libby.

But there is another semantic dimension for these four verbs in which the facts line up differently. *Explain* and *think* concern mental events while *give* and *go* concern physical events. There is a typical surface reflex of this distinction also, namely, mental verbs accept sentence complements (express a relation between an actor and a proposition):

(11) Jane thinks/explains that there is a mongoose in the parlor.
(12) *Jane goes/gives that there is a mongoose in the parlor.

The learner who appreciates both of these mapping relations can deduce from the range of syntactic environments that *give* expresses physical transfer while *explain* expresses mental transfer (that is to say, communication; Zwicky 1971, Fisher et al. 1991). Potentially there can be a rapid convergence on the meaning of a verb from examination of the several structures in which it appears in speech. Though there are hundreds of transfer verbs and scores of cognition-perception verbs, there is a much smaller number of verbs whose meanings are compatible with both these structures, and which therefore can express communication (e.g., *tell*, *shout*, *whisper*). Thus across uses, the syntax can significantly narrow the hypothesis space for the verb meaning.

Perhaps the most important reason for postulating this cross-sentence procedure has to do with 'open roles'. Though *eat* is logically a two-argument predicate, with an eater and an eatee, still one can say *The baby is eating*. It is often supposed that learners would unerringly interpret a co-occurring scene as one of eating, and simply refuse this scene/sentence as a learning opportunity because the scene doesn't line up with the required argument structure for *eat* (assuming, of course, that the child can't 'hear traces'). But unfortunately, scenes are complex and therefore almost always support false construals if a single scene/sentence pair is to be decisive. For instance, the baby is sitting, smiling, and so forth, while she eats. Why not map one of these inalienable acts onto the observed intransitive sentence? The advantage of cross-sentence analysis taken together with cross-scene analysis is that it can reveal the argument structure associated with the verb overall. Attention to several structural environments becomes an even more important capacity of the learning device when we consider languages such as Chinese, which allows rampant omission of arguments in the surface structure.[15]

We have now hypothesized that the lexical entry for the verb is derived from observing its range of syntactic environments (which reveal the verb's argument-taking properties) taken together with the observational environments (which reveal 'everything else'.) This does not mean that each use of the verb instantiates each argument-taking component of the lexical entry. For instance, verbs like *open* and *sink* express a causal relation in some environments but not others (*Carol opens the door* vs. *The door opens*). The gloss assigned to /open/ in the lexicon must be one that is compatible with both licensed semantic-syntactic environments. In contrast, *die* is noncausal (intransitive) only and *kill* is causal (transitive) only. Following Grimshaw (1992), we might render the entry for *open* as

```
  (x      CAUSE      (y      OPEN))
   |                  |
AGENT               THEME
```

'decomposing' the meaning only to the level required to state the argument structure and thus to predict the surface structures. The semantic distinction between opening and closing is then derived from examining situational factors, though only God and little children know just how. There is no opening-vs.-closing syntactic reflex to aid them.

This scheme certainly does not imply that *open* is interpreted as meaning 'an event which is causal and noncausal at the same time'. If *that* were true, then the larger a verb's syntactic range the less it would mean. Rather, the lexical description predicts that *open* is causal when transitive, noncausal when intransitive. The interpretive choice among those made available by the lexical entry is (on any single use of the verb) derived computationally from the truth value of the sentence structure.

[15] To mention one more example pertinent to the verbs studied in our experiment, note that both *eat* and *get* can occur in transitive environments. But a clue to the transfer-of-possession sense of *get* is manifest in ditransitive *Emily gets Jacques an ice cream cone* vs. **Emily eats Jacques an ice cream cone*. Similarly, there is a semantic correlate of object dropping (*eat* vs. *want*), see Resnik (1993).

5.5.2. Documentation of the use of frame ranges in acquisition

Theory all aside, the questions remain whether speech to novices, which tends to be quite simple, is sufficiently rich in structural information to support learning; and whether observers will use such information even if it is there in the input. The answer to both questions appears to be yes.

The database provided by mothers to their young children is refined enough to support learning from frame ranges. Lederer et al. (in press) examined lengthy conversations of 8 mothers with their young children (MLU < 2.0) to find the verbs used most frequently within and across mothers. For the 24 most common verbs, a verb by syntactic-environment matrix was developed for each mother's speech. Within and across mothers, each verb was found to be unique in its syntactic range. Using a procedure devised by Fisher et al. (1991), it was determined that overlap in the syntactic environments predicted their semantic overlap to a striking degree.

Much more important, there is also evidence that observers (at least adults in the laboratory) will use this cross-sentence information for verb identification. This was shown by Lederer et al. (1991) with a version of the experiment that Pinker (this volume) has acknowledged would put frame-range learning to the crucial test.

Adult subjects were shown lists of actual syntactic structures used by mothers to their infants (MLU < 2.0) with all nouns and verbs converted to nonsense (e.g., "Rom GORPS that the rivenflak is grum", "Can vany GORP the blicket?"). Thus the input stimuli were designed to permit a strong test of the frame range hypothesis – no scenes, no content words, just syntactic frames. The subjects correctly identified 52% of the maternal verbs under this condition. This level of performance compares very favorably to the 7% correct performance based on scene observation, the 13% level achieved by knowing only the nouns in the sentence, and the 28% correct performance achieved in the presence of nouns-plus-scenes. Not only does this show that some 'ideal' frame range is informative for verb identification (as in Fisher et al. 1991). It shows – by using as the stimulus set real maternal sentences whose content words have been converted to nonsense – that the frame range provided by mothers to infants under age two is sufficient to support a good measure of verb identification. In particular, subjects provided with frame-range information had no difficulty in conjecturing mental verbs. For these (e.g., *want*), the structural information is particularly useful while, as we discussed earlier, the scene information is misleading owing to the bias to interpret scenes as depicting concrete actions.

A difficulty with interpreting these results onto the child learning situation is that these subjects (when correct) were identifying old verbs that they knew, by definition: Perhaps they just looked up the frame ranges for these known verbs in their mental lexicons rather than using the frames to make semantic deductions. Because of this possibility, the pertinence of the findings is much more easily interpreted by inspecting the 48% of cases where the subjects *failed* to identify the maternal verb, guessing something else.[16] The finding is that false guesses given in response to frame-range information are semantically close to the actual verb the mother said (as assessed by the Fisher et al. semantic-similarity procedure) while false guesses in response to scenes were semantically unrelated to the verb the mother actually uttered. As syntactic bootstrapping predicts, the frame range put the subjects into the 'semantic neighborhood' even when they did not allow convergence to a unique verb construal.

Note that 52% percent correct identification, while a significant improvement over 7% or 28%, is not good enough if we want to model the fact that verb learning by three-year-olds is a snap. They do not make 48% errors so far as we know, even errors close to the semantic mark. But as we have repeatedly stressed, syntactic bootstrapping is not a procedure in which the child is assumed to forget about the scene, or the co-occurring nominals, and attend to syntax alone (as Lederer et al. forced their subjects to do in this manipulation by withholding all other evidence). It is a sentence-to-world pairing procedure. Indeed, adding the real nouns to the frames without video in this experiment led to over 80% correct verb identification; adding back the scene yielded almost perfect performance. So if the child has available (as she does, in real life) multiple paired scenes and sentences, we can at last understand why verb learning is easy.

Collateral evidence from children for the use of multiple frames is at present thin, largely because it is difficult to get young children to cooperate while a lengthy set of structures/scenes is introduced. (For this reason, the experiment presented in this article settled for studying the effect of single structures on the interpretation of single scenes, though we interpret the findings as a snapshot of an iterative process.) Supportive evidence comes from Naigles et al. (1993), who found that young children will alter their interpretation of known verbs in response to hearing them in novel syntactic

[16] This is analogous to the findings of our experiment as organized in table 5. There we looked only at instances where the children did *not* come up with the known verb, but indicated some new construal through a paraphrase. The findings were the same as for the single-word glosses.

environments, while older children and adults usually will not (for a replication and extension, see Naigles et al. 1992). Evidently, expansion of the frame range is taken as evidence for alteration of the construal early in the learning process for that word, but after extensive experience the word's meaning is set and the syntax loses its potency to change the construal.

5.3. What semantic clues reside in the syntax?

We have suggested that the formal medium of phrase structure constrains the semantic content that the sentence is expressing, thus providing crucial clues to the meaning of its verb. One such clue resides in the number of arguments: A noun phrase position is assigned to each verb argument; this will differentiate *push* from *fall*. Another concerns the positioning of the arguments: the subject of transitives is the agent, differentiating *chase* from *flee*. The case-marking and type of the argument also matters, e.g., verbs whose meaning allows expression of paths and locations typically accept prepositional phrases (*The rabbit puts the blanket on the monkey*; Jackendoff 1978, Landau and Jackendoff, in press), and verbs that express mental acts and states accept sentential complements (*John thinks that Bill is tall*, Vendler 1972).

Of course one cannot converge on a unique construal from syntactic properties alone. Since the subcategorization properties of verbs are the syntactic expressions of their arguments, it is only those aspects of a verb's meaning that have consequences for its argument structure that *could* be represented in the syntax. Many – most – semantic distinctions are not formally expressed with this machinery. An important example involves the manner in which an act is accomplished, e.g., the distinctions between *slide*, *roll* and *bounce*, which are not mapped onto differences in their syntactic behavior (Fillmore 1970). All these verbs require as one argument the moving entity and allow the causal agent and path of motion as other arguments; hence, *The ball slid, rolled, bounced (down the hill); Kimberley slid, rolled, bounced the ball (down the hill)*. The specific manners of motion are expressed within the verb rather than surfacing as distinctions in their syntactic ranges.

In sum, it is only the meaning of a verb as an argument-taking predicate that can be represented by the surface phrase structures (Rappaport et al. 1987, Fisher et al. 1991, Fisher, in press). The structures can therefore reveal only certain global properties of the construal, such as whether the verb can express inalienable (intransitive), transfer (ditransitive), mental/perceptual (inflected sentence complement), and symmetrical (sensitivity of the frame to the number of one of its arguments) contents, and whether it expresses an activity (progressive) or a state (simple present).

Overall, our view is not that there are 'verb classes', each of which has semantic components and (therefore) licenses certain structures. Rather we suggest that verb frames have semantic implications (truth values), and verbs have meanings. Owing to the meaning of the verb, it will be uncomfortable – and thus rarely or never uttered – in some frame, e.g., we don't say "Barbara looked the ball on the table" because no external agent can cause a ball to move just by looking at it (that would be psychokinesis). If the circumstances warrant, however, *look* can and will be used unexceptionally in this frame; for example, the rules of baseball make it possible to say (and sports announcers do say) "The shortstop looked the runner back to third base". As for learners, we believe they note the frame environments in which verbs characteristically occur, and thus the argument structures with which their meanings typically comport. These ranges of 'typical structures' are compatible with only small sets of verb meanings.

Because the formal medium of phrase structure is revealing only of a restricted set of semantic properties, we cannot and have not argued that the verb mappings are learned 'from' the syntax. Indeed we have just made clear that what most people think of as the 'meaning' (that *open* concerns being ajar while *close* concerns being shut) is nowhere to be found in the syntax of sentences. Rather, we have shown that the initial narrowing of the search-space for that meaning, by attention to the argument structure as revealed by the syntax, is the precondition for using the scene information efficiently to derive the meaning. When babies do not appear to know the phrase structure, they learn few verbs; when adults and young children are required to identify verbs without phrase structure cues (as when told "Look! Ziking!" or when presented with silent videos of mother–child conversation) again they do not converge to a unique interpretation. We conclude that the phrase structure is the learner's version of a zoom lens for verb vocabulary acquisition.[17]

[17] Pinker (1984) hypothesized that, in the relatively advanced child, phrase structural information could be used for another purpose: to assign abstract words, those that are neither things nor acts (e.g., *situation, know*) to lexical categories such as noun and verb: a new item that occurs in a verb position in the structure is, in virtue of that position, a verb; and so forth. Pinker termed this procedure 'structure dependent distributional learning'. This seems plausible. But this procedure will give no clue to the verb meaning, other than that the word means 'something verby'. In contrast, the position we adopt allows semantic distinctions within the verb class to be extracted. *Know* can be assigned to the class of 'mental' verbs just because it accepts tensed sentence complements. This gross semantic classification accomplished, the burden on observation is still to distinguish among *think, know, realize* and so forth.

5.4. *Quirks, provisos, and limitations*

The usefulness of form-to-meaning correspondences for verb learning is limited by several factors:

5.4.1. *Language-specific linkages of syntax to semantics*

By no means all functions from syntax to semantics are universal. There are within-language quirks in these mappings; for example, paths are usually encoded via prepositional phrases in English (e.g., *come into the room*) but occasionally not (e.g., *enter the room*); see Gruber 1965. Moreover, there are some systematic differences in the mappings across languages (Talmy 1985). It would be incoherent to suppose that language-specific correspondences (which themselves must be learned) could serve as input to verb learning at early stages. But linguistic-descriptive findings due to Gruber (1965), Fillmore (1968), Levin (1985), Jackendoff (1972, 1978, 1990), Talmy (1985), Pinker (1989), and experimental evidence from Fisher et al. (1991) and Geyer et al. (forthcoming) suggest that the correspondence rules are broad in scope and sufficiently stable cross-linguistically to support a good measure of verb-vocabulary learning.

5.4.2. *Arguments, adjuncts, and phrase boundaries*

Another severe problem for our approach has to do with the syntactic analysis that children could perform on the sentences heard. After all, *I saw the book on the table* must be analyzed differently from *I put the book on the table*. Otherwise *see* and *put* will be assumed to have the same number of argument positions, falsely suggesting a similarity in their conceptual structure. How is the correct parse of an utterance to be achieved? Practically speaking, the very short sentences used to novices contain such structural ambiguities only very rarely. Moreover, there is evidence that height of attachment is robustly cued by prosody in speech to novices (Lederer and Kelly 1991). All the same, the principled difficulty of deriving the intended parse from observation of a spoken sentence certainly complicates the proposed learning procedure (see Grimshaw, this volume).

5.4.3. *The problem of polysemy*

If a single phonological object has more than one, and unrelated, senses, this creates another problem for the procedure we have envisaged. Consider *relate* in the senses ‘tell’ and ‘conceptually connect’. Such an item will have a variety of subcategorization frames, but some of these will be consequences of one of the meanings and others the consequence of the other. Putting them together as a single frame-range should lead to chaos. The degree to which polysemy reduces the plausibility of the use of multiple frames is unknown in detail (see Grimshaw, this volume, for a pessimistic view). However, there is some suggestive evidence that the frame-ranges of verbs are well-correlated with their meanings in the general case, despite this problem.

The manipulations of interest in this regard were carried out in English by Fisher et al. (1991) and in Hebrew by Geyer et al. (forthcoming). One group of subjects provided the frame ranges for a set of common verbs (they gave judgments of grammaticality of all the verbs in various syntactic environments). A second group of subjects provided semantic-relatedness judgments for these verbs presented in isolation (with no syntactic context). The question was whether the overlap in frame-ranges predicted the semantic relatedness among the verbs. The answer is yes, massively – and in materially the same way for English and for Hebrew. The more any two verbs overlapped in their syntactic privileges, the closer they were judged to be in their meanings. Evidently, overlap in frame range provides a guide to semantic relatedness that (though probabilistic) is stable enough to contribute to the verb-learning feat.

The three problems just described – variability of the mapping relations, alternate parses for input sentences, and polysemy – limit or at least complicate the potential effectiveness of the procedure we have called syntactic bootstrapping. Thus any linguist or psychologist worth his or her salt can find counterexamples to the claim that frame-information, or frame-range information, always and perfectly predicts the relevant (argument-taking) properties of the verbs – just as counterexamples to the usefulness of situational information are easy to find.

We must suppose, in consequence, that the learner draws on convergent cues from prosody, syntax, and situation, as available, jiggling them all across instances to achieve the best fit to a lexical entry. That is, the internal structure of the child's learning procedure is likely to be quite mixed in the information recruited and probabilistic in how such information is exploited, sad as this seems. In the work presented, we could show only that syntactic evidence is on theoretical grounds crucial for working out certain mapping problems (those that involve perspective-taking verbs) and indeed is used by youngsters solving for these under some exquisitely constrained laboratory conditions. We take the outcomes to lend plausibility to the overall approach.

6. Conclusions

We have proposed a learning procedure for verbs which requires that children be armed with at least some innate semantic/syntactic correspondence rules, and considerable abilities and dispositions to perform formal analyses on the speech they hear. As such, these ideas have often been rejected as too formidable to be used by babies – sometimes by the same commentators who invoke highly abstract formal principles to account for the child's acquisition of syntax. In Pinker's (1989: 263–264) words, if the syntactic bootstrapping 'mechanism is used at all, it is used ... as a sophisticated form of cognitive problem solving rather than a general interpretive linguistic scheme', 'a kind of riddle solving'. In contrast, there is something so tangible and appealing to introspection about the idea of parsing of ongoing events that this is widely accepted as a sufficient basis for lexical learning.

But these intuition-derived theoretical biases cannot so lightly be accepted. They have fooled us before. For example, there is strong cross-linguistic evidence that two-year-olds are at least as quick – probably quicker – to extract the formal aspects of gender as to extract their semantic aspects (Levy 1983). Gordon (1985) has shown that young children are more attentive to the formal distinction between mass and count nouns in English than to the semantic correlates of this distinction.

The present experiment documented only the focusing ('zoom lens') aspect of the syntactic bootstrapping procedure. The successive narrowing of the semantic conjecture that derives from observation of a verb's several licensed structural environments was not tested, though prior experimentation we have cited demonstrates both the strength of these relations and use of this procedure by adults and children in verb identification.

It is rather more surprising that there is little systematic evidence for the word-to-world pairing procedure either. An interesting exception is Gropen et al. (1991). Children in this experiment heard new motion verbs while shown several example scenes along with the syntactically uninformative sentence "This is pilking". Most of them learned the verb meanings (though some did not). Unfortunately, the teaching procedure included negative instances ("This is *not* pilking") and specific correction whenever the children erred. So far as we know, such explicit negative evidence is not usually available to learners.[18] Still, no one can doubt that a salient motion (e.g., zigzagging in

[18] For Gropen et al.'s purposes this unusual teaching environment did not matter. Their aim was to show that if the children *did* learn the verb, they could project its argument structure and

this experiment) can sometimes be mapped against a spoken verb from the bare evidence of observation.

It is the overwhelming fallibility of such a word-to-scene procedure that we have emphasized in this article. Therefore we have challenged the logic of observation alone as the input to verb learning. One such challenge is that too many verbs come in pairs that are just about always mapped onto the same situations, so cross-situational observation will never distinguish between them. Another is that some verbs encode concepts that are not observable at all, e.g., *know* or *want*. Another is that considerations of salience (the agency bias) are fatal to the possibility of verb learning in all the cases where the caretaker happens to utter some word which is not 'the most salient of all' for the scene then in view.

To help redress these logical and practical problems we have shown that children can make significant use of structural information. At the same time their construals of new verb meanings are affected by biases as to how to represent an event. But this influence from plausibility considerations acknowledged, the influence of structure is materially stronger and wins out in the majority of cases. Toddlers know that it is better to receive than to give when Santa is the indirect object of the sentence.

References

Armstrong, S., L.R. Gleitman and H. Gleitman, 1983. What some concepts might not be. Cognition 13, 263–208.

Bates, E. and B. MacWhinney, 1982. Functionalist approaches to grammar. In: E. Wanner, L.R. Gleitman (eds.), Language acquisition: The state of the art, 173–218. New York: Cambridge University Press.

Beckwith, R., E. Tinker and L. Bloom, 1989. The acquisition of non-basic sentences. Paper presented at the Boston University Conference on Language Development, Boston.

Bowerman, M., 1976. Semantic factors in the acquisition of rules for word use and sentence construction. In: D.M. Morehead, A.E. Morehead (eds.), Normal and deficient child language. Baltimore, MD: University Park Press.

Bowerman, M., 1982. Reorganizational processes in lexical and syntactic development. In: E. Wanner, L.R. Gleitman (eds.), Language acquisition: The state of the art, 319–346. New York: Cambridge University Press.

---

hence its probable surface structure privileges. The answer was positive: Knowledge of meaning predicts surface structure (though errorfully, see Bowerman 1982) just as the latter (again, errorfully) predicts the former.

Braine, M. and J. Hardy, 1982. On what case categories there are, why they are, and how they develop: An amalgam of *a priori* considerations, speculation, and evidence from children. In: E. Wanner, L.R. Gleitman (eds.), Language acquisition: The state of the art, 219–239. New York: Cambridge University Press.

Brown, R., 1957. Linguistic determinism and the part of speech. Journal of Abnormal and Social Psychology 55, 1–5.

Bruner, J.S., 1975. From communication to language: A psychological perspective. Cognition 3, 255–287.

Bruner, J.S., 1983. Child's talk. New York: Norton.

Chomsky, N., 1957. Syntactic structures. New York: Mouton Publishers.

Chomsky, N., 1981. Lectures on government and binding. Dordrecht: Foris.

Clark, E.V., 1987. The principle of contrast: A constraint on language acquisition. In: B. MacWhinney (ed.), Mechanisms of language acquisition, 264–293. Hillsdale, NJ: Erlbaum.

Clark, H.H. and J.S. Begun, 1971. The semantics of sentence subjects. Language and Speech, 14, 34–46.

Cooper, W. and J. Paccia-Cooper, 1980. Syntax and speech. Cambridge, MA: Harvard University Press.

Dowty, D., 1991. Thematic proto-roles and argument selection. Language 67(3), 547–619.

Dromi, E., 1987. Early lexical development. New York: Cambridge University Press.

Fernald, A. and T. Simon, 1984. Expanded intonation contours in mothers' speech to newborns. Developmental Psychology 20(1), 104–113.

Fillmore, C.J., 1968. Lexical entries for verbs. Foundations of language 4, 373–393.

Fillmore, C.J., 1970. The grammar of hitting and breaking. In: R. Jacobs, P. Rosenbaum (eds.), Readings in English transformational grammar, 120–133. Waltham, MA: Ginn.

Fisher, C., in press. Structure and meaning in the verb lexicon: Input for a syntax-aided verb learning procedure. Language and Cognitive Processes.

Fisher, C., 1993. Preschoolers' use of structural cues to verb meaning. Paper presented at the 60th Annual meeting of the Society for Research in Child Development, New Orleans, LA.

Fisher, C. and H. Tokura, in press. Prosody in speech to infants: Direct and indirect cues to syntactic structure. In: J. Morgan, C. Demuth (eds.), Signal to syntax. Hillsdale, NJ: Erlbaum.

Fisher, C., H. Gleitman and L.R. Gleitman, 1991. On the semantic content of subcategorization frames. Cognitive Psychology 23, 331–392.

Fodor, J.A., 1981. The present status of the innateness controversy. In: J.A. Fodor (Ed.), Representations. Cambridge, MA: MIT Press.

Fritz, J.J. and G. Suci, 1981. Facilitation of semantic comprehension at the one-word stage of language development. Journal of Child Language 9, 31–39.

Gentner, D., 1978. On relational meaning: The acquisition of verb meaning. Child Development 49, 988–998.

Gentner, D., 1982. Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In: S.A. Kuczaj (ed.), Language development, Volume 2: Language, thought and culture, 301–334. Hillsdale, NJ: Erlbaum.

Gerken, L., P.W. Jusczyk and D.R. Mandel, 1993. When prosody fails to cue syntactic structure: Nine-month-olds' sensitivity to phonological vs. syntactic phrases. Ms., State University of New York at Buffalo.

Geyer, J., L.R. Gleitman and H. Gleitman, forthcoming. Subcategorization as a predictor of verb meaning: Evidence from modern Hebrew. Ms., University of Pennsylvania.

Gillette, J.A., 1992. Discovering that /think/ means 'think': The acquisition of mental verbs. Ms., University of Pennsylvania.

Gillette, J.A. and L.R. Gleitman, forthcoming. Effects of situational cues on the identification of nouns and verbs.

Givón, T., 1986. The pragmatics of word order: Predictability, importance, and attention. Amsterdam: Benjamins.

Gleitman, L.R., 1990. The structural sources of verb meanings. Language Acquisition 1, 1–55.

Gleitman, L.R., H. Gleitman, B. Landau and E. Wanner, 1988. Where learning begins: Initial representations for language learning. In: E. Newmayer (ed.), The Cambridge Linguistic Survey, Vol. III: 150–193. New York: Cambridge University Press.

Goldin-Meadow, S., M. Seligman and R. Gelman, 1976. Language in the two-year-old. Cognition 4, 189–202.

Golinkoff, R.M., R.C. Jacquet and K. Hirsh-Pasek, in progress. Lexical principles underly the learning of verbs. Ms., University of Delaware.

Gordon, P., 1985. Evaluating the semantic categories hypothesis: The case of the count/mass distinction. Cognition 20, 209–242.

Grimshaw, J., 1981. Form, function and the language acquisition device. In: CL. Baker, J.J. McCarthy (eds.), The logical problem of language acquisition, 165–182. Cambridge, MA: MIT Press.

Grimshaw, J., 1990. Argument structure. Linguistic Inquiry Monograph 18. Cambridge, MA: MIT Press.

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

Grosjean, F. and J.P. Gee, 1987. Prosodic structure and spoken word recognition. Cognition 25, 135–155.

Gruber, J.S., 1965. Studies in lexical relations. Ph.D. dissertation, MIT.

Hall, D.G., 1993. Basic-level individuals. Cognition 48, 199–221.

Hall, D.G. and S. Waxman, 1993. Assumptions about word meaning: Individuation and basic level kinds. Child Development 64, 1550–1570.

Hirsh-Pasek, K. and R.M. Golinkoff, 1991. Language comprehension: A new look at some old themes. In: N. Krasnegor, D. Rumbaugh, M. Studderty-Kennedy, R. Schiefelbusch (eds.), Biological and behavioral aspects of language acquisition, 301–320. Hillsdale, NJ: Erlbaum.

Hirsh-Pasek, K., H. Gleitman, L. Gleitman, R. Golinkoff and L. Naigles, 1988. Syntactic bootstrapping: Evidence from comprehension. Paper presented at the Boston University Conference on Language Development, Boston.

Hirsh-Pasek, K., D. Kemler-Nelson, P. Jusczyk, K. Cassidy, B. Druss and L. Kennedy, 1987. Clauses are perceptual units for young infants. Cognition 26, 269–286.

Jackendoff, R., 1972. Semantic interpretation in generative grammar. Cambridge, MA: MIT Press.

Jackendoff, R., 1978. Grammar as evidence for conceptual structure. In: M. Halle, J. Bresnan, G. Miller (eds.), Linguistic theory and psychological reality, 201–228. Cambridge, MA: MIT Press.

Jackendoff, R., 1990. Semantic structures. Cambridge, MA: MIT Press.

Joshi, A. and O. Rambow, in progress. Dependency parsing for phrase structure grammars. Ms., University of Pennsylvania.

Jusczyk, P.W., K. Hirsh-Pasek, D.G. Kemler-Nelson, L.J. Kennedy, A. Woodward and J. Piwoz, 1992. Perception of acoustic correlates of major phrasal units by young infants. Cognitive Psychology 24, 252–293.

Jusczyk, P. and D. Kemler-Nelson, in press. Syntactic units, prosody, and psychological reality during infancy. In: J. Morgan, C. Demuth (eds.), Signal to syntax. Hillsdale, NJ: Erlbaum.

Kako, E., in progress. Preschoolers assign novel verbs to novel actions. Ms., University of Pennsylvania.

Kayne, R., 1992. A restrictive theory of word order and phrase structure. Paper presented at the Annual Meeting of GLOW, Lisbon, Portugal.

Keenan, E.L., 1976. Toward a universal definition of 'subject'. In: C.N. Li (ed.), Subject and topic, 305–333. New York: Academic Press.

Klatt, D.H., 1975. Vowel lengthening is syntactically determined in a connected discourse. Journal of Phonetics 3, 129–140.

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

Landau, B. and R. Jackendoff, in press. 'What' and 'where' in spatial language and spatial cognition. Brain and Behavioral Sciences.

Lederer, A., and M. Kelly, 1991. Prosodic correlates to the adjunct/complement distinction in motherese. Papers and Reports in Child Language Development 30, Stanford, CA..

Lederer, A., H. Gleitman and L.R. Gleitman, 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., H. Gleitman and L.R. Gleitman, in press. The syntactic contexts of maternal verb use. In: M. Tomasello, W. Merriman (eds.), Beyond names for things: Young children's acquisition of verbs. Hillsdale, NJ: Erlbaum.

Lehiste, I., J.R. Olive and L. Streeter, 1976. The role of duration in disambiguating syntactically ambiguous sentences. Journal of the Acoustical Society of America 60(5), 1199–1202.

Lenneberg, E.H., 1967. Biological foundations of language. New York: Wiley.

Leslie, A.M., 1982. The perception of causality in infants. Perception 11, 173–186.

Leslie, A.M. and S. Keeble, 1987. Do six-month-old infants perceive causality? Cognition 25, 265–288.

Levin, B., 1985. Lexical semantics in review: An introduction. In: B. Levin (ed.), Lexical semantics in review. Lexicon Project Working Papers, 1: 1–62. Cambridge, MA: The MIT Center for Cognitive Science.

Levy, Y., 1983. It's frogs all the way down. Cognition 15, 75–93.

Mandler, J., 1991. Prelinguistic primitives. In: L.A. Sutton, C. Johnson (eds.), Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society, 414–425. Berkeley, CA: Berkeley Linguistics Society.

Markman, E., 1989. Categorization and naming in children: Problems of induction. Cambridge, MA: MIT Press.

Mazuka, R. 1993. How can a grammatical parameter be set before the first word? Paper presented at the Signal to Syntax Conference, Brown University, Providence, RI.

Michotte, A., 1963. The perception of causality. London: Methuen.

Naigles, L., 1990. Children use syntax to learn verb meanings. Journal of Child Language 17, 357–374.

Naigles, L., A. Fowler and A. Helm, 1992. Developmental shifts in the construction of verb meanings. Cognitive Development 7, 403–427.

Naigles, L., H. Gleitman and L.R. Gleitman, 1993. Children acquire word meaning components from syntactic evidence. In: E. Dromi (ed.) Language and development, 104–140. Norwood, NJ: Ablex.

Nelson, K., 1973. Structure and strategy in learning to talk. Monographs of the Society for Research in Child Development 38.

Nelson, K., 1974. Concept, word, and sentence: Interrelations in acquisition and development. Psychological Review 81, 267–285.

Newport, E.L., 1977. Motherese: The speech of mothers to young children. In: N. Castellan, D. Pisoni, G. Potts (eds.), Cognitive Theory, Volume 2. Hillsdale, NJ: Erlbaum.

Ninio, A., 1980. Ostensive definition in vocabulary acquisition. Journal of Child Language, 7, 565–573.

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

Pinker, S., 1989. Learnability and cognition. Cambridge, MA: MIT Press.

Quine, W.V., 1960. Word and object. Cambridge, MA: MIT Press.

Rappaport, M., B. Levin and M. Laughren, 1987. Levels of lexical representation. Lexicon Project Working Papers #20. Cambridge, MA: MIT Center for Cognitive Science.

Resnick, P., 1993. Selectional preference and implicit objects. Paper presented at the CUNY Sentence Processing Conference, Amherst, MA.

Schlesinger, I., 1988. The origin of relational categories. In: Y. Levy, I. Schlesinger, M. Braine (eds.), Categories and processes in language acquisition, 121–178. Hillsdale, NJ: Erlbaum.

Slobin, D.I., 1975. On the nature of talk to children. In: E.H. Lenneberg, E. Lenneberg (eds.), Foundations of language development. New York: Holt, Rinehart and Winston.

Talmy, L., 1985. Lexicalization patterns: Semantic structure in lexical forms. In: T. Shopen (ed.), Language typology and syntactic description, Volume 3: Grammatical categories and the lexicon. New York: Cambridge University Press.

Tomasello, M. and A.C. Kruger, 1992. Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. Journal of Child Language 19, 311–333.

Vendler, Z., 1972. Res cogitans: An essay in rational psychology. Ithaca, NY: Cornell University Press.

Zwicky, A., 1971. In a manner of speaking. Linguistic Inquiry 11(2), 223–233.