LANGUAGE ACQUISITION, *1*(1), 3–55


# The Structural Sources of Verb Meanings

Lila Gleitman
*University of Pennsylvania*

> If we will observe how children learn languages, we will find that, to make them understand what the names of simple ideas or substances stand for, people ordinarily show them the thing whereof they would have them have the idea; and then repeat to them the name that stands for it, as 'white', 'sweet', 'milk', 'sugar', 'cat', 'dog'.
>
> (John Locke, 1690/1964, Book 3.IX.9)

Is vocabulary acquisition as straightforward as Locke supposes? Three hundred years after publication of *An Essay Concerning Human Understanding,* Locke's is still the dominant position on this topic for the very good reason that common sense insists that he was right: Word meanings are learned by noticing the real-world contingencies for their use. For instance, it seems obvious to the point of banality that the verb pronounced /run/ is selected as the item that means 'run' because this is the verb that occurs most reliably in the presence of running events.

Or is it? Who has ever looked to see? One trouble with questions whose answers are self-evident is that investigators rarely collect the evidence to see if they pan out in practice.

It is not my purpose in the present discussion to try to defeat the obviously correct idea that a crucial source of evidence for learning word meanings is observation of the environmental conditions for their use. I believe, however, that what is correct about such a position is by no means obvious, and therefore deserves serious study rather than acceptance as a background fact in our field.

I'll largely limit the discussion to the topic of acquiring verb meanings for two reasons: first, because the underpinnings of verb and noun learning are

Requests for reprints should be sent to Lila Gleitman, Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, PA 19104.

likely to differ significantly; and second, because it is in the former domain that I and my colleagues have some experimental evidence to offer in support of the position I want to adopt. Even within this subtopic, to begin at all I will have to make critical assumptions about some heady issues that deserve study in their own right. Particularly, I will not ask where the concepts that verbs encode come from in the first place, for example, how the child comes to conceive of such notions as 'run' (or 'think' or 'chase'). I want to look at the learner at a stage when he or she can entertain such ideas, however this stage was arrived at.[1] Second, I reserve for later discussion the question of how the child determines which word in the heard sentence is the verb—that it is the phonological object /run/, not /horse/ or /marathoner/—that is to be mapped onto the action concept.

The question that remains seems a very small one: How does the learner decide *which particular phonological object* corresponds to *which particular verb concept,* just Locke's topic. But I'll try to convince you that this question is harder than it looks. For one thing, matching the meanings to their sounds is the one part of acquisition that cannot have any very direct innate support. This is because the concept 'run' is not paired with the sound /run/ in Greek or Urdu, so the relation must be learned by raw exposure to a specific language. Moreover, it is not clear at all that the required pairings are available to learners from their ambient experience of words and the world.

In the first half of this article, I set out some of the factors that pose challenges to the idea that children can induce the word meanings from their contexts in the sense that Locke and his descendants in developmental psycholinguistics seem to have in mind. In this discussion, I will allude repeatedly to the work and theorizing of Steve Pinker, because he seems to me to be the most serious and acute modern interpreter of ideas akin to Locke's in relevant regards.[2] Then, in response to these challenges to the

[1]This is a large simplification of the learning problem for vocabulary, to be sure. It's not likely that learning in this regard is always and only a matter of mapping the words heard onto a preset and immutable set of concepts priorly available to the prelinguistic child. Rather, there is bound to be some degree of interaction between the categories lexicalized in a language and the child's conceptual organization; moreover, that conceptual organization is changing during the period of vocabulary growth, to some degree affecting the nature of lexical entries (for discussion, see Carey, 1985; Pinker, 1989). For present purposes, however, I abstract away from this intriguing class of problems.

[2]It is important to be clear about the sense in which many modern theorists seem close to Locke in their position: They believe the ambient environment in which words are heard is used as the primary—perhaps the sole—early basis for forming conjectures about the meanings. But it is just as important to note that Pinker and other recent commentators differ radically from the British Empiricists in almost all other respects. Particularly, Locke held—or is usually read to have held—that the vocabulary in which the description of the environment is couched is sensory. In contrast, modern perspectives often assert that children approach the task of interpreting the world equipped with a very smart perceptual system, as well as sophisticated

theory of learning by observation, I will sketch a revised position laid out by Landau and Gleitman (1985), illustrating it with some recent experimental evidence from our laboratory. The idea here is that children deduce the verb meanings in a procedure that is sensitive to their syntactic privileges of occurrence. They *must* do so, because either (a) there is *not enough* information in the whole world to learn the meaning of even simple verbs, or (b) there is *too much* information in the world to learn the meanings of these verbs.

## PART I: SOME DIFFICULTIES OF LEARNING BY OBSERVATION

At peril of caricaturing Locke—but who doesn't?—I select him as one who argued for a rather direct relation between knowledge and the experience of the senses. He frequently used the case of individuals born without sight as a testing ground for such a position. According to Locke, both sighted and blind people ought to be able to learn the meanings of such words as *statue* and *feel* and *sweet,* but the blind ought to be unable to acquire *picture* and *see* and *red,* for the concepts that these words express are primitive (i.e., not derivable from other concepts) or derivable from primitives that are available only to the eye.

Barbara Landau and I were directly inspired by Locke to study the acquisition of vision-related terms by blind babies (Landau & Gleitman, 1985). As our studies evolved, we realized that exactly the same conceptual issues about learning arise for sighted vocabulary learners as for blind ones, so I will move on to discussion of such normally endowed children. The blind population, which I discuss first, is perhaps special only as the biographical point of origin of our own thinking but will serve to dramatize some issues that seem less startling in the ordinary case. These have to do with how resistant the word-learning function is to the evidence of the senses.

### Locke's Idea: Differences in Experience Should Yield Differences in the Meanings Acquired

Landau and I were astonished to discover how much alike were the representations of vision-related terms by blind and sighted children at age

mental models of the current situation, a belief-desire psychology, a naive physics, more or less correct intuitive theories of semantics and pragmatics, and schemas for possible word meanings. And a couple of generations of inquiry in psychology generally support such an enriched view of the child's mental status as word learning begins. All this sophisticated representational apparatus in obvious ways puts the modern child in a vastly better position to fathom the world than Locke's child. Yet, in some perhaps less obvious ways that I'll be discussing, the increased representational power makes it harder rather than easier to learn the word meanings from observation of their environmental contingencies.

3, despite what would appear to be radical differences in their observational opportunities. For instance, all these babies showed by their comprehension performances that they took *look* and *see* as terms of perception, distinct from such contact terms as *touch*. As an example of this, a blind child told to "Touch but don't look at . . ." a table would merely bang or tap it. But if told "Now you can look at it," she explored all its surfaces systematically with her hands. Moreover, she understood *look* to be the active (or exploratory) and *see* the stative (or achievement) term in this pair. Just as surprising, blind children as well as sighted children understood that *green* was an attribute predicable only of physical objects (they asserted that ideas could not in principle be green while cows might be, for all they knew). Thus the first principle that a theory of observational learning must be subtle enough to capture is that the same semantic generalizations can be acquired in relative indifference to differing environmental experience, if the notion *experience* is cast in sensory-perceptual terms.

## Word-to-World Pairings and the Blind Child's Semantic Conjectures

While we found the surprising result that blind children shared much knowledge about vision-related terms with their sighted peers, we also achieved the unsurprising result that there were some differences in how these two populations understood these terms to refer to their own perceptions: Blind children think that *look* and *see* describe their own *haptic* perceptions, but sighted children think these same words describe their own *visual* perceptions. Thus blindfolded sighted children of 3 years turn their faces skyward if told to "Look up!" but a blind child of the same age holds her head immobile and searches the space above with her hands in response to the same command (see Figures 1 and 2).[3]

This outcome is of just the sort that is subject to seemingly obvious explanations involving the extralinguistic contexts of use: The difference in interpretation for blind and sighted seems to be directly attributable to differences in environmental contingencies for the words' use. Specifically, we reasoned (as does everyone to whom one presents this set of facts): A blind child's caretaker will use the terms *look* and *see* intending the child to

[3]A related difference holds for the color words. Sighted children of 4 and 5 years map the color words onto observed hues in the world while blind children ask for help. Perhaps they think the property is stipulative. Asked "Why are the flowers in the woods pink?" one blind child responded, "Because we *name* them pink!" They know these are attributes predicable only of physical objects (they say that an idea can't be green because "it's only in your head"), but they don't know what the real-world dimension may be. Interestingly, they avoid some choices that their extralinguistic experience appears to make available, for example, that color terms refer to sizes of objects (Landau & Gleitman, 1985, chapter 8).

FIGURE 1 A sighted blindfolded child's response to the command "Look up!" (Reproduced from Landau & Gleitman, 1985, p. 58, with permission of the artist, Robert Thacker.)

perceive in whatever ways her sensorium makes available. And since the blind child's way of discovering the nature of objects is by exploring them manually, the caretaker will surely use *look* and *see* to this child only when an object is near enough to explore manually. That is, the caretaker should say "Look at this boot" to her blind baby only if a boot is nearby, ready to be explored manually. The contexts of use for these words thus should include—among many other properties—conversationally pertinent objects that are near at hand. Had the caretaker instead rattled a boot noisily by the child's ear whenever she said "Look at this boot," the learner would have surmised that *look* meant 'listen'.

FIGURE 2 A blind child's response to the command "Look up!" (Reproduced from Landau & Gleitman, 1985, p. 56, with permission of the artist, Robert Thacker.)

So here we have a straightforward prediction from the environment of use to the formation of a semantic conjecture: By hypothesis, the blind learner decides that *look* involves *haptic* exploration because it is that verb which is used most reliably in contexts in which haptic exploration is possible and pertinent to the adult/child discourse. Landau and I decided to test that prediction to see if it was as true as it was obvious.

To do so we examined videotapes of a mother and her blind child recorded in the period *before* the child uttered any vision-related words or indeed any verbs at all (that is to say, during the learning period for these words). There were 1,640 utterances in the sample. We selected for situational analysis all verbs (excluding *be*) that occurred 10 or more times

TABLE 1
Spatial Analysis of the Mother's Use of Verbs to the Blind Child

| | *Proportion Used in Contexts* | | | | | | |
|---|---|---|---|---|---|---|---|
| | *In Hand or Near* | | *Far* | | *No Object* | | |
| *Verb* | *+Ling. Object* | *−Ling. Object* | *+Ling. Object* | *−Ling. Object* | *+Ling. Object* | *−Ling. Object* | *Total Cases*[a] |
| Perceptual | | | | | | | |
| *Look* | .50 | .22 | .00 | .08 | .14 | .06 | 34 |
| *See* | .33 | .06 | .44 | .11 | .00 | .06 | 18 |
| *Watch* | .56 | .00 | .44 | .00 | .00 | .00 | 17 |
| Nonperceptual | | | | | | | |
| *Come* | .00 | .05 | .00 | .32 | .00 | .63 | 19 |
| *Get* | .45 | .05 | .20 | .05 | .00 | .25 | 27 |
| *Give* | .97 | .00 | .03 | .00 | .00 | .00 | 21 |
| *Go* | .00 | .52 | .10 | .14 | .00 | .24 | 20 |
| *Have* | .53 | .00 | .33 | .14 | .00 | .00 | 11 |
| *Hold* | 1.00 | .00 | .00 | .00 | .00 | .00 | 10 |
| *Play* | .50 | .20 | .00 | .00 | .30 | .00 | 10 |
| *Put* | .97 | .00 | .00 | .00 | .03 | .00 | 61 |

*Note.* Reproduced from Landau and Gleitman, 1985, p. 214, with permission of the publisher, Harvard University Press.

[a]These total to $N = 248$, the number of utterances containing common verbs (those occurring 10 or more times in the corpus). The remaining rarer verbs (occurring fewer than 10 times) and 183 instances of *be* were excluded from the analysis.

in this corpus; the number of utterances including these common verbs was 248. We then coded all uses of these verbs according to whether they occurred when an object pertinent to the conversation (a) was NEAR enough to the child for her to explore it manually, i.e., within arm's reach, (b) was FARther away than that, or (c) when there was NO such pertinent OBJECT. Each of these three situational categories was further subdivided into cases where the "pertinent object" was specifically mentioned in the verb-containing sentence ("+Linguistic Context" in Table 1, e.g., a boot was in the child's hand when the mother said "Look at this boot") and cases where the pertinent object was not specifically mentioned but might have been inferred from the larger discourse properties ("−Linguistic Context," e.g., a boot in the child's hand when the mother said "See?" or "Look at this!"). The results, so coded, are shown in Table 1.[4]

[4]Notice that we couched the child's representation of the environmental distinction in sensory-perceptual terms (the object is "nearby" or "far away" as the action begins). But the child's representational terminology might instead—or in addition—be "object starts at a nearby/distant *source.*" That is, conceptions of these locations as sources and goals of the action rather than as physical locations and movements constituting the action might be closer to the child's real representations of the events perceived. Indeed, many others who have coded

We hypothesized that *look* and *see* would be among the verbs used most reliably in the NEAR condition accounting for why the child had assigned them the meanings 'explore/apprehend *haptically*' (while other verbs would be used less often in this condition and so would not be assigned this property of meaning). But inspection of Table 1 shows that this hypothesis fails to account for the child's haptic interpretation of *look* and *see*. *Put* and *give* and *hold* are the verbs used most reliably (over 95% of the time) under the NEAR condition, while *look* (72%, collapsing across the "+ and −Linguistic Object" cases) and especially *see* (39%) are not as reliably associated with this environmental condition.

What has gone wrong? Could it really be that the presence of pertinent objects near to hand had nothing to do with the blind child's interpretation of *look* and *see* as haptic? As I will show in Part II of this article, this conjecture about the experiential basis for this aspect of the words' meaning really does succeed, though not when used as in Table 1: in a procedure that maps isolated word forms against their extralinguistic contexts.

Reserving further discussion for later, it's worth noting here only that the nearbyness analysis of Table 1 cannot be written off as of some environmental property that is hopelessly irrelevant to the child's interpretation of events. For as it stands, this analysis extracts and explains important distinctions among verbs of physical motion that are in other respects semantically close, such as *give* versus *get*. The child is apparently told, sensibly enough, to *give* what she has in hand (this verb is used in the NEAR condition 97% of the time) but to *get* what she doesn't have (the relevant NEAR percentage for this verb is 50%).

## Latitude of the Hypothesis Space

If Table 1 mirrors the sole analysis that children perform in aid of learning the modality (hand or eye) implicated by *look* and *see,* they will clearly fail. As the blind child did learn, there must be something insufficient or wrong about this analysis. Of course this doesn't demonstrate that contexts of use can't account for this aspect of verb learning, or any other; rather, we might conclude that the idea of real-world context, to succeed, must be a good deal more subtle than we (and many others) originally supposed. That is, the response to the findings shown in Table 1 is usually, and perhaps should be:

---

maternal speech and its context have preferred this latter terminology, which will serve as well in our case, too. The point is that for present purposes the labeling doesn't matter at all, for the coding imposed will be the same in either case. Note also that the near/far analysis can succeed at all only if the child can determine the discourse addressee. This assumption is plausible because in these transcripts the mother's speech is over 90% about the "here-and-now," and in over 90% of instances the addressee is the child herself.

> Oh, but the contextual analysis you imposed was so *feeble.* Showing that it failed is only showing the failure of Landau and Gleitman's imagination. The child surely imposes a richer analysis on the situation than that, and the only analysis relevant to the hypotheses under test is the one that the child herself imposes.

Fair enough. We limited the child to observing some perceptually salient features of the situation, features that the infancy literature tells us are available even to babies. This is because our aim was to see how far some small and independently documented set of observational primitives could get the learner in extracting simple meaning features for assignment to the verbs. These were that the world is populated with objects that endure over time (Spelke, 1982), and that move relative to each other (Lasky & Gogol, 1978) and with respect to the positions of the child's own body (Acredolo & Evans, 1980; Field, 1976). These assumptions put the child in a position to conceive of the situation as one of objects—in this case, objects whose noun names are known to the child—moving (as described by the verb) between sources and goals. For example, for *give* the object moves from NEAR as action begins to FAR when it ends, and in *get* the object goes from FAR to NEAR.

It can hardly be denied, in light of the infancy evidence, that youngsters do represent situations in terms of the positions and motions of pertinent objects. What is surely false, however, is that such categories are exhaustive among the child's extralinguistic analyses. Infants come richly prepared with means for picking up information about what is going on in their environment—looking, listening, feeling, tasting, and smelling; in fact these different sensory routes appear to be precoordinated for obtaining information about the world (Spelke, 1979). To take a few central examples, infants perceive the world as furnished with objects that are unitary, bounded, and persist over time and space (Gibson & Spelke, 1983; Spelke, 1985), and that cannot occupy two places at one time (Baillargeon, Spelke, & Wasserman, 1985). They distinguish among the varying properties of objects, for example, their rigidity or elasticity (Gibson & Walker, 1984), their colors (Bornstein, 1975), their movement or nonmovement (Ball & Vurpillot, 1976), their positions and motions relative to the child observer (Field, 1976), their animacy (Golinkoff, Harding, Carlson, & Sexton, 1984), causal roles (Leslie, 1982), and even their numerosity (Starkey, Spelke, & Gelman, 1983). If you think there's something that infants can't or won't notice, look in the next issue of *Developmental Psychology* and you will probably discover that someone proved they can.

Now that I have acknowledged something of the richness of infant perception, why not let the learner recruit this considerable armamentarium for the sake of acquiring a verb vocabulary? Why not assume that the child

encodes the situation not only in the restricted terms that yield Table 1, but in myriad other ways? For instance, over the discourse as a whole, probably the mother has different aims in mind when she tells the child to "look at" some object than when she tells her to "hold" or "give" it. The child could code the observed world for these perceived aims and enter these properties as aspects of the words' meanings. But also the mother may be angry or distant or lying down or eating lunch and the object in motion may be furry or alive or large or slimy or hot, and the child may code for these properties of the situation as well, entering them, too, as facets of the words' meanings.

The problems implicit in such an expansion of the representational vocabulary should be familiar from the literature on syntax acquisition: The trouble is that an observer who notices *everything* can learn *nothing,* for there is no end of categories known and constructable to describe a situation.[5] Indeed, not only learnability theorists but all syntacticians in the generative tradition appeal to the desireability of narrowing the hypothesis space lest the child be so overwhelmed with representational options and data-manipulative capacity as to be lost in thought forever. At least, learning of syntax could not be as rapid and uniform as it appears to be unless children were subject to highly restrictive principles of Universal Grammar, which rein in their hypotheses. As one famous example, learners are said to assume that all syntactic generalizations are structure-dependent rather than serial-order dependent (Chomsky, 1975; Crain & Fodor, in press). In fact, Universal Grammar is claimed to be as constrained as it is owing to the child's requirement that this be so (Wexler & Culicover, 1980).

I put it to you: Are these observations about the difficulties of learning when the hypothesis space is vast no less true of word learning than of syntax? In the domain of vocabulary acquisition as much as that of syntax acquisition, there is remarkable efficiency and systematicity of learning across individuals (and, as the blind children show, across learning environments): The rapidity and accuracy of vocabulary acquisition are jewels in the crown of rationalistically oriented developmental psycholinguistics

[5]As so often, Chomsky (1982) set the problem with great clarity:

> The claim we're making about primitive notions is that if data were presented in such a way that these primitives couldn't be applied to it directly, prelinguistically before you have a grammar, then language couldn't be learnt. . . . And the more unrealistic it is to think of concepts as having those properties, the more unrealistic it is to regard them as primitives. . . . We have to assume that there are some prelinguistic notions which can pick out pieces of the world, say elements of this meaning and of this sound. (p. 119)

The analysis of Table 1 is an attempt to see how far some small set of observational primitives, known experimentally to be available to infants, could get them in extracting a simple meaning feature ('haptic') for assignment to certain verbs.

(see particularly Carey, 1978). So just as in the case of syntax, we have initial grounds for claiming that a limit on the hypothesis space must be a critical source of sameness in the learning function. Bolstering the same view, languages seem to be as alike in their elementary vocabularies as they are in their syntactic devices (Talmy, 1975, 1985). But surprisingly enough, all the telling arguments invoked for syntax to restrict the interpretation of the input—that is, constraints on representations—that are to explain these samenesses in form, content, and learning functions are thrown out the window in most theorizing about the lexicon. *There* it is usually maintained that the child considers many complex, varying, cross-cutting, subtle conjectures about the scenes and events in view so as to arrive at the right answers, comparing and contrasting possibilities across many events, properties, discourse settings, and so forth. In other words, testing and manipulating an exceedingly broad and free-ranging hypothesis space.

In the domain of verb learning, a very few investigators have been responsive to the issues here. Pinker (1987), in a direct and useful discussion of the requirement to limit the space of observables that a learner will consider in matching the event to the unknown verb, wrote as follows:

> Verbs' definitions are organized around a surprisingly small number of elements: "The Main Event", that is, a state or motion; the path, direction, or location of an object, either literal spatial location or some analogue of it in a nonspatial semantic field; causation; manner; a restricted set of the properties of a theme or actor; temporal distribution (aspect and phase); purpose; coreferentiality of participants in an event; truth value (polarity and factivity); and a handful of others. (p. 54)

It is an open question whether Pinker's proposed list is narrow enough to meet the requirement for a realistic set of primitives upon which a verb-learning procedure can operate. Are purposes, truth values, causes, not to speak of "analogues of spatial location in nonspatial semantic fields" really primitives that inhere in the observations themselves? It seems highly unlikely that any choice of *perceptual* constraints will be restrictive enough to delimit the analyses a child performs in reaction to each word-to-world pair. Of course I'm not suggesting that there aren't principles of perception that are restrictive and highly structured (God forbid!). But they are likely not restrictive enough to account for vocabulary acquisition. How could they be? Perception has to be rich enough to keep the babies from falling off cliffs and mistaking distant tigers for nearby pussycats lest they all disappear from the face of the earth before learning the verb meanings. The very richness of perception guarantees multiple interpretive possibilities at many levels of abstraction for single scenes; but the problem for word learning is to select from among these options the single interpretation that is to map onto a particular lexical item.

Jerry Fodor has suggested to me, maybe seriously, that the problems of alternate encodings of the same scene go away because the caretaker and child are in cahoots, and they are mind readers. They are so attuned in discourse, being creatures of exactly the same sort, that the child zaps onto exactly the characteristics of the situation that the mother, just then, has in mind to express; and by the same token, the mother more or less unfailingly understands the intents of the child (see Bruner, 1974/1975, for a story about how the attentional conspiracy is to be set up by mother and child).

However, recent evidence leaves room for extreme pessimism concerning these telepathic capabilities in learner and tutor. Golinkoff (1986) examined communicative eposides between mothers and their 11–18-month-old children and found that, even in the later period, instances of immediate comprehension of the child's desires by the mothers constitute only about half the episodes. For the rest, the mother either initially misunderstands the child's desire or ignores his signals altogether. To be sure, the final outcome of these failed communications is rarely child or mother tantrum; usually they just give up and change the subject. Thus while affability is normally maintained, in practice communication with linguistic novices very often fails. This appears to dispose of the mind-reading solution. It seems that the multiply interpretable world poses a real problem for the language learner and teacher.

## Multiply Interpretable Events

The richness of perception is not the only, or even the major, problem faced by a hypothetical learner who tries to acquire verb meanings from observation. A more difficult problem is that even the homeliest and simplest verbs, though they refer to events perceivable, encode also the unobservable present interests, purposes, beliefs, and perspectives of the speaker. I turn now to this class of problems.

Consider the learning of simple motion verbs, such as *push* or *move.* In a satisfying proportion of the times that caretakers say something like "George pushes the truck," George can be observed to be pushing the truck. But unless George is a hopeless incompetent, every time he pushes the truck, the truck will move. So a verb used by the caretaker to describe this event may represent one of these ideas ('push') or the other ('move').

Moreover, every real event of the pushy sort necessarily includes, in addition to the thrust and goal, various values of trajectory, rate, and so forth, so that such ideas as 'slide', 'clank', 'roll', 'crawl', 'speed', and so on, are also relevant interpretations of a new verb then uttered. What is left open by the observation is whether that verb represents any or all of these manner differences: no, in the case of *push,* but yes in the case of *roll* or *speed.*

Note that the manner elements just mentioned do fall within the range encoded by verbs in many languages (Talmy, 1985) and are on the narrowed list of perceptual properties suggested by Pinker (1987). I leave aside various other interpretations often called less salient, that is, I ignore more general consideration of the "stimulus-free" character of language use (see Chomsky, 1959), especially the countless fanciful interpretations of this event that could be drawn by worried philosophers. Ignoring these, there are always many highly salient, linguistically sanctioned, interpretations of a single action scene. How is the child to decide which of these interpretations is truly encoded by the particular verb uttered in the presence of such a scene?

It is possible that these ambiguities are eliminated by looking at a verb's uses *across* situations. There will eventually be some instance of moving called /push/ in which the truck is moving rapidly, eliminating 'crawl' as a conjecture about the meaning of this item, and so on. By a process of cross-comparison and elimination, it has been proposed that each verb may eventually be distinguishable. In Pinker's (1987) words:

> the child could learn verb meanings by (a) sampling on each occasion in which a verb is used, a subset of the features . . . [the features are those mentioned in my earlier quotation of Pinker], (b) adding to the tentative definition for the verb its current value for that feature, and (c) permanently discarding any feature value that is contradicted by a current situation. (p. 54)

I discuss this general idea at some length in a later section of this article. But notice now that, as stated, the position is surely too strong. Even if mothers always and only refer in their speech to the here-and-now in the presence of a young child, it cannot be guaranteed, *pace* Fodor, that child and adult are always attentionally focused in the same way. After all, sometimes the mother is speaking of one thing ("Eat your peas, dear!") while the child is attending to something else altogether (say, the hungry dog under the table). So the learner had better not "discard permanently" any feature that contradicts the current situation *as the child is conceiving it.*

In fact, positive imperatives pose one of the most devastating challenges to any scheme that works by constructing word-to-world pairings, for the mother will utter "Eat your peas!" if and only if the child is not then eating the peas. Thus a whole class of constructions is reserved for saying things that mismatch the current situation.

It follows that the child's confirmation metric for a verb meaning cannot be so stringent as to exclude an interpretation "permanently" if it should mismatch even a very few scenes. The necessarily probabilistic nature of such a procedure complicates its operation to an unknown degree. Even more important, the burden of hypothesis testing for cross-situational analysis

becomes ominous as the comparison set (of verbs, properties, scenes, and discourse analyses) required to make it go through enlarges.[6]

## Paired Verbs That Describe Single Events

Difficult problems can be solved. Impossible ones are harder. Consider such verb pairs as *flee* and *chase, buy* and *sell, win* and *beat, give* and *receive,* and so on. Such pairs are common in the design of verb lexicons. The members of each pair allude to a single kind of event: Whenever the hounds are chasing the fox, the fox is fleeing from the hounds. If some hounds are racing, even with evil intentions, toward a brave fox who holds its ground, they cannot be said to be chasing him. The hounds are chasing only if the fox is fleeing. If the child selects a verb from the stream of speech accompanying such a scene, how then is she to decide whether it means 'chase' or 'flee'?

Such examples are thrusts to the heart of the observational learning hypothesis. As Pinker (1987) acknowledges,

> Basically, we need to show that the child is capable of entertaining as a hypothesis any possible verb meaning, and that he or she is capable of eliminating any incorrect hypothesis as a result of observing how the verb is used across situations. (p. 54)

But if *chase* and *flee* (and a host of similar pairs) are relevantly used in just the same situations, it follows that it *cannot* be shown that the child is

[6]Some ideas for pruning the observational data base into a more manageable form for learning have been suggested. Usually these involve ways of filtering out input that is complex by some semantic, structural, or processing criterion (for early attempts, see Newport, Gleitman, & Gleitman, 1977; Shipley, Smith, & Gleitman, 1969). However, the number and nontransparency of the categories that these preanalyses require often seem more troublesome than the problem that they were designed to simplify. Here is an example from Pinker (1984); the task discussed is discovery of exemplars of the (innate) property *subject* from their semantic/pragmatic environmental correlates; the problem addressed is that in many situations those correlates will be absent. I have italicized the linguistic and situational categories in terms of which, according to Pinker, the child is to construct a data base suitable for finding the subject exemplars.

> The semantic properties of subject hold only in basic sentences: roughly, those that are *simple, active, affirmative, declarative, pragmatically neutral,* and *minimally presuppositional.* . . . The parents . . . or the child might filter out nonbasic sentences from the input using various contextual or phonological diagnostics of nonbasicness such as *special intonation, extra marking of the verb, presuppositions set up by the preceding discourse or the context, nonlinguistic signals of the interrogative* or *negative illocutionary force of an utterance,* and *so on.* (p. 47)

capable of eliminating the incorrect hypotheses by cross-situational observation.

I think the problem is that words don't describe events *simpliciter*. If that's all words did, we wouldn't have to talk. We could just point to what's happening, grunting all the while. But instead, or in addition, the verbs seem to describe specific perspectives taken on those events by the speaker, perspectives that are not "in the events" in any direct way. How far are we to give the learner leave to divine the intents of his or her elders as to these perspectives? Are they talking of hounds acting with respect to foxes, or of foxes with respect to hounds? Speaking more generally, since verbs represent not only events but the intents, beliefs, and perspectives of the speakers on those events, the meanings of the verbs can't be extracted solely by observing the events.

## The Subset Problem

A related problem has to do with the level of specificity with which the speaker, by the words chosen, refers to the world. Consider the homely little objects in the world, the pencils, the ducks, the spoons. All these objects are supplied with more than one name in a language, for example, *animal, duck, Donald Duck*. I expect that the adult speaker has little difficulty in selecting the level of specificity he or she wants to convey and so can choose the correct lexical item to utter in each case. And indeed the learner may be richly pre-equipped perceptually and conceptually so as to be able to interpret scenes at these various levels of abstraction and to construct conceptual taxonomies (Keil, 1979). But, as usual, this very latitude adds to the mystery of vocabulary acquisition, for how is the child to know the level encoded by the as yet unknown word? The scene is always the same if the child conjectures the more inclusive interpretation (that is, if the first conjecture is 'animal' rather than 'duck'). For every time there is an observation that satisfies the conditions (whatever these are) for the appropriate use of *duck,* the conditions for the appropriate use of *animal* have been satisfied as well.

Analogous cases exist in the realm of verb meanings. To return to the instance dramatized by the blind learners, *perceive, see, look, eye* (in the sense of 'set eyes on'), *orient,* pose the same subset problem. There is no seeing without looking, looking without eyeing, eyeing without orienting, and so on. All this suggests that not only blind children, but sighted children as well, should have (essentially the same) difficulties in learning the meanings of *look* and *see,* because the distinction between the two words is not an observable property of the situations in which they are used. Yet, as I discussed earlier, it is just these unobservable properties that the blind and sighted 3-year-olds held in common.

Gold (1967) addressed a problem that seems related to this one. He showed formally that learners who had to choose between two languages, one of which was a subset of the other, could receive no positive evidence that they had chosen wrong if they happened to conjecture the superset (larger) language. This is because the sentences they would hear, all drawn from the subset, are all members of the superset as well. It has therefore been proposed that learners always hypothesize the smaller (subset) language; they initially select the most restrictive value of a parameter on which languages vary (Berwick, 1982; Wexler & Manzini, 1987).

But the facts about the lexicon do not allow us to suppose that the child has a solution so simple as choosing the least inclusive possibility; that is, to choose that interpretation which subsumes the smaller set of real-world referents (all the ducks rather than all the animals, or a limited aspect of perception rather than all of it). In the end, learners acquire words at all such levels of specificity. Moreover, neither the most inclusive nor the least inclusive possibilities seem to be the initial conjectures of learners; rather, some middle level of interpretation is the one initially selected, that is, *duck* and *look* (as opposed to *animal/mallard* and *examine/glimpse*) seem to be the real first choices.[7]

In sum, words that stand in a subset relation pose another serious problem for an unaided observation-based learning procedure. This is because the child who first conjectures a more inclusive interpretation can receive no positive evidence from word-to-world mappings that can dissuade him. And the idea that the child always begins with the least inclusive interpretation consistent with the data is falsified by the empirical facts.

Some plausible approaches to solution of this class of problems have been suggested in the literature, particularly with reference to the problem of learning noun meanings. To my knowledge, all of them invoke the idea mentioned earlier—that there is some middling level of abstraction (the "basic" level; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) in terms of which the child naturally parses the perceptual world. Assuming that we can make good on this initial assumption, one further postulate can help an observational learning theory go through. This is that there are no synonyms in the monomorphemic vocabulary of a language (Clark, 1988; Markman & Hutchinson, 1984). In that case, the child may step up or down within a cognitively ordered set of concepts at levels of abstraction higher and lower than the basic one in the cases where a new word is used to describe an entity for which the child already has a known name (for

[7]These results can't be written off on grounds of the differential frequency of these words in the input corpus, for if the frequencies are changed, the level of categorization does not. For instance, in some houses *Fido* is a more frequent word than *dog*, but in that case the youngest children think that the sound /fido/ means 'dog' (Rescorla, 1980).

discussion, see Carey, 1985; Jones, Landau, & Smith, 1986; Waxman & Gelman, 1986).

However, it is premature to be too optimistic about this sort of proposal, for it is not at all clear that the notion of "basic categories" can ever be brought to ground. This is because the set of elementary categories underlying the monomorphemic vocabulary may be so large that the constraints from this quarter could be quite insignificant in explaining how the child learns which word encodes which concept. The psycholinguistic literature to date cannot even account for the intuition that, while *grape* and *pea* are basic terms (with the superordinates *fruit* and *vegetable*), *bird* and *tree* seem to be basic (rather than the superordinates of *lark* and *elm*). Notice that if the idea of a basic conceptual level must allude to overall familiarity to repair such problems, it loses all explanatory force for answering to vocabulary acquisition issues. Note also that the descriptive problems for the idea of a basic level grow materially worse when more formal (e.g., *female* or *integer*) or functional (e.g., *equal* or *meet*) terms are considered (for discussion, see Armstrong, Gleitman, & Gleitman, 1983; Fodor, 1981).

More relevant to my present purposes, this class of solutions begins to invoke evidence that is not in the world of observation but rather resides in the design of language itself; in the present case, the child's assumption about the lexicon is that for all practical purposes it excludes synonyms. As I shall argue presently, quite sophisticated presuppositions about the structure of language appear to be necessary to account for the acquisition of vocabulary.

## Semantic Properties That are Closed to Observation

The verbs that most seriously challenge the observational learning hypothesis still remain to be discussed: These are the ones that don't refer to the observable world at all.

Locke noted that the meanings of many words involve properties that are not observable, but he did not consider this fact to be fatal to his overall position because his view, most likely warranted, was that those who used such abstract words didn't know what they were talking about half the time anyhow. Nevertheless, a key problem for observational learning is that many words are related to the real world only in the most obscure and invisible ways, if at all. Try, for example, to learn the meaning of the word *think* by titrating discourse situations into those in which thinking is going on, somewhere when you hear /think/ versus those in which no thinking is happening. Remember that there isn't always brow furrowing or a Rodin statue around to help. Keep in mind also that you are going to have to

distinguish as well among *think, guess, wonder, know, hope, suppose,* and *understand,* not to speak of—a few months or years later—*conjecture, figure, comprehend, discover, perceive,* and so forth. Many developmental psycholinguists rule such instances out of school on the grounds that these aren't words that children know very well at 2 and 3 years old, but this won't do. After all, we also want to understand the children who manage to survive to become the 4- and 5-year olds.

I don't really think this topic needs much more belaboring. If the child is to learn the meanings from perceptual discriminanda in the real world, the primitive vocabulary of infant perception has to be pretty narrow to bring the number and variety of data storing and manipulative procedures under control. But no such narrow vocabulary of perception could possibly select the thinkingness properties from events. I conclude that an *unaided* observation-based verb-learning theory is untenable because it could not acquire *think.*

## The Fitful Fit of Word to World

Earlier in this discussion, I claimed that a realistic observation-based procedure must operate in terms of probabilistic rather than absolute word-to-world matches, at least because child and caretaker cannot be assumed to be attending to the same aspects of the same scene on every occasion when some verb is uttered. Thus the wise child would not permanently give up on a conjectured verb meaning in the presence of a very small proportion of mismatches to the world. I now ask how serious this objection may be for the viability of such theories. In what proportion of cases, really, do the verbs uttered match up with the scenes in view?

### *The Relation Between Word and World is Probabilistic*

There has been almost no systematic work on this topic. The idea that word-to-world contingencies *must* be strong and stable is entrenched in three hundred years of empiricist speculation, and to a large extent this fixed belief has been a barrier to empirical inquiry. Table 1, in fact, describes one of the rare studies in which anyone has attempted to see just how the words line up with their contexts of use. And that analysis, as we've seen, yielded quite puzzling results.

A recent study by Beckwith, Tinker, and Bloom (1989) achieves findings at least as problematic as our own. Beckwith et al. are working with a very large maternal corpus of utterances to children in the age range of 13–23 months, with a view to understanding the acquisition of verb argument

structure. This sample includes about 8,000 verb-containing utterances. The assumption is that only when noun referents in these utterances were present (in the scene in view) would they be of any use to the child learner in acquiring the argument structures. Some 3,000 of the verb-containing utterances failed to meet this criterion and therefore were discarded. To take two specific examples: There were 566 sentences containing the verb *put;* in 55 cases (10%), the sentence did not refer to the here-and-now. There were 80 sentences with the verb *open,* of which 30 (37.5%) were not about the here-and-now. Thus, if these data are at all representative, a child who learns verb meanings by asking about their relations to ongoing scenes must be quite tolerant of counterexamples.

In fact, the prospects for observational learning may be materially worse than emerges in the analysis just described, for this tells us only that 67.5% of the time when /open/ is uttered, opening is happening—a somewhat ominous but not necessarily devastating proportion of fit of word to scene. But one must also ask the question in the opposing direction: What is the likelihood, given that an event of opening is in view and has captured the child's attention, that /open/ (rather than some other verb) will be uttered? Can one doubt that this relationship will turn out to be muddy in the extreme? For an ideal case, suppose the door to Alfred's house squeaks loudly, so his attention is invariably captured by the noise as it opens, and hence he invariably looks up and attends whenever it opens. When, every evening, Mother opens the door upon returning from work, what does he hear? I would venture that he rarely hears her say "Hello, Alfred, I am opening the door!" but very often hears "Hello, Alfred, whatcha been doing all day?" (and just as often hears Father say "Shut the door, it's freezing in here!"). In short, any scheme for learning from observation must have some machinery for dealing with the fact that caretaker speech is not a running commentary on scenes and events in view.

Beckwith et al.'s analysis does presuppose significant further machinery: As mentioned earlier, it summarily discards those utterances that don't refer to the ongoing event. But this is defensible only if it can be shown that the learner who doesn't know the word meaning, like the analyst who does, has some means for excluding these instances. After all, if the child truly believes that utterances refer to the here-and-now, he or she will simply form the wrong sound/meaning pairings when the adult speaks of things nonpresent. For instance, if the child hears "Let's get some duck for dinner tomorrow" while throwing a ball, she might assume that /get/ means 'throw' and that /duck/ means 'ball'. This problem seems especially acute for the mother/child discourse Beckwith et al. are studying, for the children are very young—on some theories, unable to understand the full sentences and thus really at the mercy of word-to-world pairings.

As no plausible theory is available for reducing the data base—in advance of learning—to one which reliably maps verb use onto scenes and events observed, the best guess is that the child acquiring meanings solely from word-to-world pairings must adopt an extremely liberal stance, accepting a meaning in the presence of a low proportion of situational "hits" and tolerating a large proportion of "misses."

### *Counterexamples and the Fitful Fit*

Such considerations bring me back to one crucial further point. I earlier asserted that cross-situational analysis as proposed by Pinker (1987) is insufficient to save the observation-based learning story, owing to examples such as *chase/flee, buy/sell,* and so on, whose real-world contingencies do not differ. But Pinker has pointed out to me that the claim of situational identity for such pairs is somewhat overdrawn, for one can think of some suitably arcane circumstances in which only one member of the pair applies—that is, situations in which one would utter *beat* but not *win,* and so forth. Here are two of Pinker's examples: It's possible for me to *flee the city* without it being implied that *the city is chasing me.* And it's at least somewhat more natural to say *I bought a Coke from the machine* than to say that *The machine sold me a Coke.* These examples defeat an absolute claim that there are no situations at all in which the meanings of such words can be disambiguated.

But these counterexamples must be evaluated in light of the child's confirmation metric for word-to-world relations, which I have tried to show must be tolerant of a significant proportion of mismatches.

For instance: Suppose Alfred has conjectured, based on some hundreds of uses of /flee/ in the presence of foxes/hounds, dogs/cats, mothers/errant children, and so on, that /flee/ means 'chase'. (Why not? All the contexts up to now fit 'chase' as well as they fit 'flee', and Alfred is among the 50% of children who guessed wrong). Now he hears "The boy took to his heels and fled the stable," with no bulls visibly in pursuit. What effect should this new data point have? Given Alfred's vexed interactions with /open/, and his consequently liberal evaluation criterion for word-to-world matches (67% hits must be good enough), this rare mismatch should have no effect at all on the prior conjecture. That is to say, overwhelmingly often when fleeing is around, chasing is around. No child who learned /open/ from its sometime relation to the world of scenes and events could be deflected by the vanishingly rare dissociation of chasing and fleeing events. As I next argue, a much more appealing procedure for dissociating the two verb meanings is by realizing that the subject noun phrase of /flee/ must represent the one who runs away and that the subject of /chase/ must be the entity in pursuit.

## Summary

I mentioned a number of problems for a theory that (solely or even primarily) performs a word-to-world mapping to solve the vocabulary-learning task. These are that:

1. Such a theory fails to account for the fact that children whose exposure conditions are radically different (the blind and the sighted) acquire much the same representations even of vision-related words.
2. Plausible, narrowly drawn candidates for event representation seem to be inadequate in accounting for the learning in certain apparently easy cases—such as expecting that words whose interpretation requires manual contact be uttered when one is in manual contact with something pertinent.
3. Broadening the hypothesis space so as to allow learners to distinguish among the many verb meanings may impose unrealistic storage, manipulation, and induction demands on the mere babes who must do the learning.

In addition, observational learning seems to fail in principle to the extent that:

4. Many verbs are identical in all respects except the perspectives that they adopt toward events (*chase, flee*) or
5. the level of specificity at which they describe a single event (*see, look, orient*) or
6. don't refer to events and states that are observable at all (*think, know*).

# PART II: NEW APPROACHES FOR VOCABULARY ACQUISITION

Since children learn verb meanings despite the apparently formidable problems of culling them from exposure to extralinguistic contexts, Landau and I conjectured (1985) that they have another source of information. This additional information derives from the linguistic (syntactic) contexts in which words occur in speech. Children's sophisticated perceptual and conceptual capacities yield a good many possibilities for interpreting any scene, but the syntax acts as a kind of mental zoom lens for fixing on just the interpretation, among these possible ones, that the speaker is expressing. To make use of this information source in acquiring the verb vocabulary, the learner must perform a sentence-to-world mapping rather than a word-to-world mapping.

To explain this position, I return first to the problem Landau and I faced in understanding the blind child's semantic achievements.

## How the Blind Child Might Have Learned the Visual Terms

Recall that the analysis of Table 1 was an attempt to explain only the most straightforward, perceptually transparent aspect of a blind learner's acquisition of *look* and *see;* namely, that if these verbs had to do with *haptic* perception, there must have been pertinent objects close to her hands when her mother said those words. Yet even this simple idea seemed to be falsified by our analysis. To find out why, our first step was to return to the data of Table 1 to see where and when the NEARNESS criterion had failed for so many uses of *look* and *see.* We found that the sentences that fit neatly with the object-nearby criterion were very simple ones: If the mother had said something like "Look at this boot!" or "See? This is a pumpkin," invariably the boot or pumpkin were NEAR, within the child's reach. But if the mother had said, "Let's see if Granny's home!" (while dialing the phone), "Look what you're doing!" (as the child spilled juice), "You look like a kangaroo in those overalls" (which had a pouch), or "Let's go see Poppy" (as they entered a car), the pertinent object was likely to be FAR or there was NO such pertinent OBJECT intended. Clearly, many of the sentences that tripped up our simple story were queer ones indeed. The mother didn't seem in most of these cases to mean 'examine or apprehend' either haptically or visually, but rather 'determine', 'watch out', or 'resemble'. Or else, as in the final example, a motion auxiliary (*go*) in the sentence transparently took off the NEARbyness requirement.

There are two ways to go now: One can claim that the NEARbyness environmental clue to the haptic interpretation was just a snare and delusion—but that is ridiculous. It just *has* to be right that this aspect of the environment was part of what licensed the child's haptic interpretation. The other choice is to find some nonquestion-begging way through which the child could have gotten rid of the sentences that otherwise would threaten the experiential conjecture. (The question-begging way, of course, is to say that the mother did not mean 'haptically explore' in the offending sentences.)

How can this be done? A potentially useful clue is that not only the meaning but the syntax too of these offending sentences is special, different from the syntax of sentences in which the child was *really* being told to explore and perceive nearby objects. This syntactic distinction may be available to the learner.

A syntactic partitioning of the verbs commonly used by the mother of the blind baby (based on the same corpus analyzed in Table 1) according to the subcategorization frames in which each verb appeared in the maternal corpus is shown in Table 2. The verbs of Table 1 appear as the columns in this table and the syntactic environments appear as the rows; the numbers

TABLE 2
Subcategorization Frames for the Common Verbs

| | *Perception Verbs* | | *Transfer Verbs* | | | | *Other* | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *look* | *see* | *give* | *put* | *get* | *hold* | *play* | *have* | *go* | *come* |
| *Look/see only* | | | | | | | | | | |
| V! | 8 | | | | | | | | | |
| V? | | 1 | | | | | | | | |
| V!, S | 10 | | | | | | | | | |
| V?, S | | 3 | | | | | | | | |
| V how S | 2[a] | | | | | | | | | |
| V S | | 5 | | | | | | 1[b] | | |
| V like NP | 5 | | | | | | | | | |
| come V NP | | 3 | | | | | | | | |
| *Exclude look/see* | | | | | | | | | | |
| V NP $PP_{loc}$ | | | 5 | 31 | 2 | | | | | |
| V NP $D_{loc}$ PP | | | | | | 1[c] | | | | |
| V NP $D_{loc}$ | | | | 28 | 2 | 6 | | | | |
| V $D_{loc}$ NP | | | | 1 | | | | | | |
| V NP NP | | | 16 | | 2 | | | | | |
| V NP where S | | | | 1[d] | | | | | | |
| V PP | | | | | | | 7[e] | | | |
| *Overlap look/see* | | | | | | | | | | |
| V $PP_{loc}$ | 3 | | | | 5 | | | | 2 | 2 |
| V $D_{loc}$ | 2 | | | | | | | | 10 | 13 |
| V ø | 2 | 3 | | | | | | | 8 | 4 |
| V NP | | 3 | | | 13 | 3 | 3 | 14 | | |
| V AP | 2 | | | | 3 | | | | | |
| Totals | 34 | 18 | 21 | 61 | 27 | 10 | 10 | 15 | 20 | 19 |

*Note.* Adapted from Landau and Gleitman, 1985, p. 112, with permission o f the publisher, Harvard University Press.

[a]E.g., "Look how I'm doing it." [b]"Let's have Barbara babysit" (causative) [c]"Hold the N up to me." [d]"Put it where it belongs." [e]*Play* with the reciprocal preposition *with*, for example, "You're not gonna play with the triangle, so forget it!"

in each cell are the number of instances of a verb in some particular syntactic environment. (Specifically, the rows of this table represent subcategorization frames, the sister nodes to the verb under the verb-phrase node.) Notice first that some of the typical syntactic environments for *look* and *see* are quite different from those for the other verbs in the set.

Moreover, we can—with only a little fudging—divide the environments of the vision-related verbs so as to pull apart those environments in which the NEARbyness contextual cue holds and those in which it does not. That analysis is shown in Table 3. Essentially, the top rows of Table 3 show the maternal uses of *look* and *see* in their canonical subcategorization frames

TABLE 3
Spatial/Syntactic Analysis of *Look* and *See*

| | *NEAR* | *FAR* | *NO OBJECT* | *"NEAR PROPORTION"* |
|---|---|---|---|---|
| Canonical sentence frames and deictic uses | | | | |
| Look at NP | 3 | 0 | 0 | |
| Look D | 2 | 0 | 0 | 1.00 |
| Look! | 8 | 0 | 0 | |
| Look!, This is NP | 10 | 0 | 0 | |
| See NP | 1 | 2 | 0 | |
| See? | 1 | 0 | 0 | .72 |
| See? This is NP | 3 | 0 | 0 | |
| Motion auxiliary | | | | |
| Come see NP | 0 | 3 | 0 | .00 |
| Other environments | | | | |
| Look AP | 0 | 1 | 1 | |
| Look like NP | 0 | 0 | 5 | .18 |
| Look how S | 0 | 2 | 0 | |
| Look ø | 2 | 0 | 0 | |
| See S | 2 | 3 | 0 | |
| See ø | 0 | 2 | 1 | .25 |
| Total (all environments) | | | | |
| Look | 25 | 3 | 6 | .73 |
| See | 7 | 10 | 1 | .39 |

*Note.* Reproduced from Landau and Gleitman, 1985, p. 115, with permission of the publisher, Harvard University Press.

(e.g., "Look at/see the frog," "Look up/down") and the deictic interjective uses that are the most frequent in that corpus (e.g., "Look!, That's a frog!" and "See?, That's a frog!"). When these syntactic types only are considered, the NEAR proportion of *look* rises (to 100%, from 72% in Table 1) and so does the NEAR proportion of *see* (to 72% from 39%). Thus if the learner can and does perform these analyses, the first result is that NEARbyness of the pertinent object becomes a much more reliable real-world clue than previously. But notice that the hypothesis now is that children perform a sentence-to-world mapping rather than the word-to-world mapping shown in Table 1: The children's interpretation of *extralinguistic* events has been

significantly modulated by their attention to *linguistic* events, namely the subcategorization frames.

Landau and I made yet another, and much stronger, claim based on the kinds of outcomes shown in Table 2. This was that the range of subcategorization frames has considerable potential for partitioning the verb set semantically, and that language learners have the capacity and inclination to recruit this information source to redress the insufficiencies of observation. This examination of structure as a basis for deducing the meaning is the procedure we have called *syntactic bootstrapping*.[8] This hypothetical procedure stands in contrast to a view that emphasizes observation as the main initial source of evidence for verb-meaning acquisition (*semantic bootstrapping*), devised by Grimshaw (1981) and considerably elaborated by Pinker (1984, 1987).[9] I turn now to a comparison of these two approaches.

## The Bootstrapping Proposals Compared

The two bootstrapping proposals are much alike in what they claim about correspondences between syntax and semantics, and they are also alike in proposing that the child makes significant use of these correspondences. First I sketch, very informally, the kinds of syntactic/semantic relationships that are crucially invoked in both proposals.

### *Syntactic/Semantic Linking Rules*

To an interesting degree, the structures in which verbs appear are projections from their meanings. To take a simple example, the different

[8]Once Landau and Gleitman embarked on this path, several colleagues (Adele Abrahamson, Paul Bloom, & Henry Gleitman, whom we thank for this observation) asked why we restricted ourselves to subcategorization frames as the source of linguistic evidence recruited by the child, rather than going whole hog for all the kinds of internal evidence potentially available across the sentence. For instance, the child could (and probably does) use selectional restrictions to narrow down the choice of verb meaning, for example, if you know that shrimp have veins, that might help achieve an interpretation of /devein/ in "Devein that shrimp!" (Compare *Devein that pencil!). Our answer was our usual one: For syntactic as well as for semantic categories, our aim was to see how far some extremely restrictive analysis could serve to handle the facts about verb learning. The most plausible choice was the subcategorization frames, which appear to vary with the meaningfully distinct predicates (for a useful discussion, see Wasow, 1985). As children open up to further data sources, they simultaneously increase the complexity of the data manipulations required. Nonetheless, I must agree that the kind of "linguistic inference" (Bloom's term) suggested by these commentators is sure to be part of the final story on vocabulary acquisition.

[9]Pinker (1984) actually reserved the term *semantic bootstrapping* for machinery that assigns words to lexical categories. For expository convenience, however, I take the liberty of using this expression to refer to his proposals at their broadest for extracting verb meanings from extralinguistic context.

number of noun phrases required by the verbs *laugh, smack,* and *put* in the sentences

1. Arnold laughs.
2. Arnold smacks Gloria.
3. Gloria puts Arnold in his place.

is clearly no accident but rather is semantically determined—by how many participant entities, locations, and so forth, the predicate implicates. Similarly, the structural positions of these noun phrases relative to the verb also carry semantic information. Thus, much more often than not, the subject noun phrase will represent the experiencer or causal agent (*Arnold* in sentence 1 and *Gloria* in sentence 3), and paths and goals will appear in prepositional phrases (*in his place* in sentence 3). These links of syntactic position and marking to semantic properties, although by no means unexceptional, typify the ways that English represents semantic-relational structure. In short, verbs that are related in meaning share aspects of their clausal syntax. Zwicky (1971) put the idea this way:

> If you invent a verb, say *greem,* which refers to an act of communication by speech and describes the physical characteristics of the act (say a loud, hoarse, quality), then you know that . . . it will be possible to greem (i.e., to speak loudly and hoarsely), to greem for someone to get you a glass of water, to greem at your sister about the price of doughnuts, to greem "Ecch" at your enemies, to have your greem frighten the baby, to greem to me that my examples are absurd, and to give a greem when you see the explanation. (p. 232)

### *Semantic Bootstrapping: Using the Semantics to Predict the Syntax*

As I mentioned earlier, both the bootstrapping proposals make critical use of these canonical relations between syntax and semantics. In the semantic bootstrapping procedure, the child first fixes the meaning of a verb by observing its real-world contingencies. I have argued at length that this hypothesis about verb-meaning extraction is too strong, for at least some features are unobservable. Yet no one can doubt that, at least sometimes, the context of use is so rich and restrictive as to make a certain conjecture about interpretation overwhelmingly likely.

Once the verb meaning has been extracted from observation, the semantic bootstrapping hypothesis invokes the linking rules (the canonical syntactic/semantic mappings) to explain how the child discovers the structures that are licensed for the use of these verbs, much in the spirit of Zwicky's

comments about the invented word *greem*. For instance, if a verb has been discoverd to mean 'give', then it will appear in three-argument structures such as *John gives the book to Mary*. This is because the logic of 'give' implies one who gives, one who is given, and that which is given, and each of these entities requires a noun phrase to express.

Not only is this position plausible. There is much evidence in its favor. Notably, Bowerman (1974, 1982) showed that children will make just such predictions based on their prior fixing of the verb meanings. That evidence came from instances where children's conjectures were evidently too bold or insufficiently differentiated, that is, where they were wrong—but still understandable. For instance, a child in Bowerman's study commanded "Don't eat the baby—she's dirty!" on an occasion when the mother was about to feed the baby (whose diaper needed changing). Presumably, the child had noted, implicitly of course, that an intransitive motion verb (e.g., *sink,* as in *The ship sànk*) could be uttered in a transitive structure (e.g., *The captain sank the ship*) to express the causal agent of this motion. If this is true of *sink* (and *open* and *melt,* etc.), why not of *eat*?

To summarize, the semantic bootstrapping procedure as developed by Grimshaw (1981) and Pinker (1984) works something like this: The child is conceived as listening to the words used and then trying to figure out their meanings by observing their situational concomitants, the word-to-world pairing that I have discussed. Quoting Pinker (1984) again:

> If the child deduces the meanings of as yet uncomprehended input sentences from their contexts and from the meanings of their individual words, he or she would have to have learned those word meanings beforehand. This could be accomplished by attending to single words used in isolation, to emphatically stressed single words, or to the single uncomprehended word in a sentence . . . and pairing it with a predicate corresponding to an entity or relation that is singled out ostensively, one that is salient in the discourse context, or one that appears to be expressed in the speech act for which there is no known word in the sentence expressing it. (p. 30)

According to this proposal, once the meanings have been derived from observation, the child can project the structures from (innate) knowledge of the rules that map semantic structures onto syntactic structures (variously termed *mapping rules, linking rules, projection rules,* or *semantic redundancy rules*). Perhaps so, but I have been arguing that entities and relations cannot in general be singled out ostensively, that "salience" and the question of what's "expressed in the speech act" are not so easily recoverable as this perspective must insist. For such reasons, Landau and I hypothesized an additional procedure, one that looks quite different from this.

### *Syntactic Bootstrapping: Using the Syntax to Predict the Semantics*

The syntactic bootstrapping proposal in essence turns semantic bootstrapping on its head. According to this hypothesis, the child who understands the mapping rules for semantics onto syntax can use the observed syntactic structures as evidence for deducing the meanings. The learner observes the real-world situation but also observes the structures in which various words appear in the speech of the caretakers. Such an approach can succeed because, if the syntactic structures are truly correlated with the meanings, the range of structures will be informative for deducing which word goes with which concept. This sentence-to-world mapping will be quite handy if, as I have argued, word-to-world mapping cannot succeed over the full range of meanings that we know are acquired.

The difference between semantic bootstrapping and syntactic bootstrapping, then, is that the former procedure deduces the structures from the word meanings that are antecedently acquired from the observation of events, while the latter procedure deduces the word meanings from the semantically relevant syntactic structures associated with a verb in input utterances. Note that although the hypothesized procedures are distinct, to hold that one of them is implicated in learning is not to deny that the other one is too. Quite the contrary. It is very likely that they operate in a complementary fashion.

Let us take the examples *put, look,* and *see,* which occurred in the corpus provided by the blind child's mother. Verbs that describe externally caused transfer or change of possessor of an object from place to place (or from person to person) fit naturally into sentences with three noun phrases, for example, *John put the ball on the table.* This is just the kind of transparent syntax/semantics relation that every known language seems to embody. It may therefore not be too wild to conjecture that this relationship is part of the original presuppositional structure that children bring into the language-learning task (Jackendoff, 1978, 1983; Pinker, 1984; Talmy, 1975). That is, 'putting' logically implies one who puts, a thing put, and a place into which it is put; a noun phrase is assigned to each of the participants in such an event. In contrast, because one can't move objects from place to place by the perceptual act of looking at them, the occasion for using *look* in such a structure hardly, if ever, arises (**John looked the ball on the table* sounds unnatural).[10] Hence the chances that /put/ means 'put' are raised and the

[10]The exceptions are (a) if you believe in psychokinesis such that your looking can move objects, or (b) if the rules of some game make it so that, in effect, an external agent *can* cause an object to move by looking at it, for example, *The shortstop looked the runner back to second base.* Once *look* does mean 'cause-to-move-by-perceptually-exploring', it becomes comfortable in this construction. That is, the subcategorization frames, just because they are

chances that /put/ means 'look' are lowered by the fact that the former and not the latter verb appears in three noun-phrase constructions in caretaker speech (see Table 2). Restating this more positively, the component 'transfer' is inserted into a verb's semantic entry in case it is observed to occur in three noun-phrase structures. This happens for /put/ but not for /look/ (see Table 2).

Verbs of perception and cognition are associated with some other constructions, as they should be. For example, if a verb is to mean 'see' (perceive, visually or haptically), it should appear with noun-phrase objects as in *John saw a mouse,* for noun phrases are the categories that languages select to describe such entities as mice. But since events as well as entities can be perceived, this verb should and does also appear with sentence complements, for clauses are the categories selected by languages for expressing whole events (e.g., *Let's see if there's cheese in the refrigerator*). The possibility that /see/ means 'see' is increased by appearance in this construction, and the likelihood that /put/ means 'see' is decreased by the fact that one never hears *Let's put if there's cheese in the refrigerator;* see again Table 2). That is, the component 'perceptual' (or more likely, 'mental') is added to the verb's entry when the sentential complement is observed.

Speaking more generally, certain abstract semantic elements such as 'cause,' 'transfer,' 'symmetry,' and 'cognition' are carried on clause structures, which are licensed by semantic information in the lexical entries of verbs. So these structures will be chosen for utterance only to the extent that they fit with the semantics of the verb items. It follows that the subcategorization frames, if their semantic values are known, can convey important information to the verb learner. To be sure, the number of such clause structures is quite small compared to the number of possible verb meanings: It is reasonable to assume that only a limited number of highly general semantic categories and functions are exhibited in the organization that yields the subcategorization frame distinctions. But each verb is associated with several of these structures. Each such structure narrows down the choice of interpretations for the verb. Thus these limited parameters of structural variation, operating jointly, can predict the

---

associated with particular truth values, are a prime linguistic vehicle for the extension of verb meanings and are so used by adults as well as by child learners. Of course, these simple examples vastly underestimate the detail required if these structural properties are to be used for learning purposes. One such problem is that the child must impose the proper parse on the sentence heard, lest *John saw the book on the table* be taken as a counterexample (that is, the analysis is to be of sister nodes under the verb phrase only, and a theory of how the child determines such configurations antecedently is a requirement of the position). Another real difficulty concerns how children should respond when they run into quirky constructions like *John saw his brother out of the room, John looked his uncle in the eye.*

possible meaning of an individual verb quite closely. Landau and Gleitman (1985) showed that the child's situational and syntactic input, as represented in Tables 1, 2, and 3, were sufficient in principle to distinguish among all the verbs commonly used in the maternal sample for the blind child. This general outcome is schematized in Figure 3.

The potential virtues of this syntactically informed verb-learning procedure are considerable:

1. It serves the local purpose of offering a nonmagical explanation for the blind child's acquisition of visual terms, as just described.
2. It points the way toward acquisition of terms when observation fails.

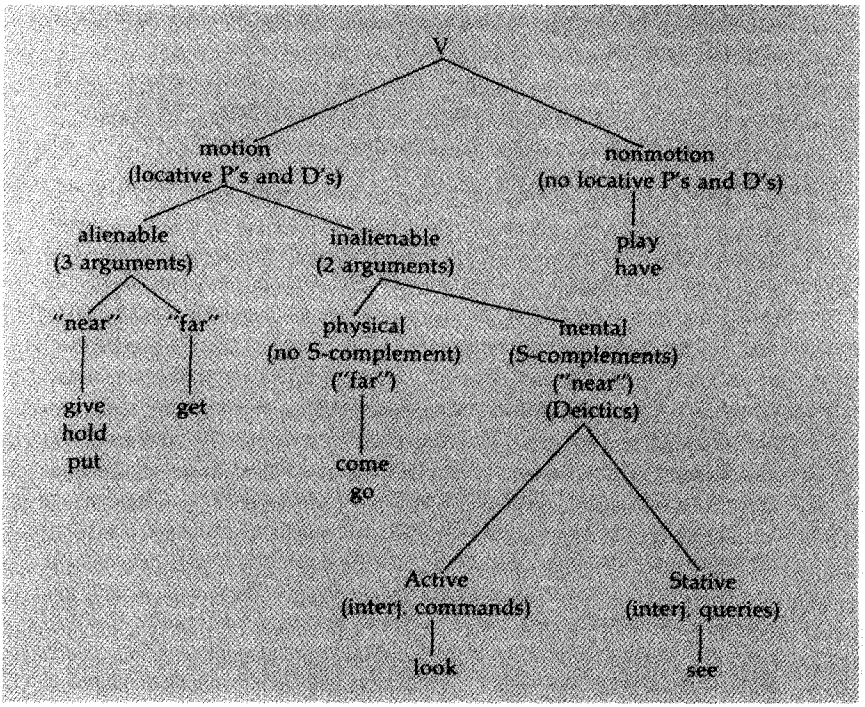


FIGURE 3 A summary of the verb subcategorization and spatial-situational clues to meaning in the maternal corpus (speech to the blind child). (Reproduced from Landau and Gleitman, 1985, p. 135, with permission of the publisher, Harvard University Press.) For expositional purposes only, the components of verb meanings (e.g., *motion, mental*) are organized in a tree diagram here, but it is likely that their real arrangement is as a cross-classification rather than a hierarchy. Postulated conceptual features such as *motion* are shown as the node labels in this tree and are assumed to be discovered through use of the syntactic and situational evidentiary sources listed in parentheses below each such feature.

This is because, for example, mental verbs such as *think* are unambiguously marked by the syntax (by taking sentence complements) even though their instances cannot be readily observed in the world.

3. It gives the child a way of learning from a very small data base. This is because the number of subcategorization frames associated with each verb is small (on the order of 10–20), and these are the data requirements for the procedure to work.

4. That data base is categorical rather than probabilistic. Though verb uses to the child are usually pertinent to what is going on in the here-and-now, sometimes they are not. For this reason, among the many described earlier, the child learning from observation must store the huge variety of situational contexts in which a word is used so as to evaluate what is "the same" about all of them. The daunting nature of such a procedure must be kept in mind, as I have tried to emphasize. The problem is that the learner can't know in advance which scene analysis is relevant to the verb meaning and so must store a multitude of these, awaiting the arrival of sufficient data (sample word-to-world pairs) for performing the cross-situational analysis. In contrast, mothers virtually never speak ungrammatically to their children—that is, use verbs in nonlicensed syntactic environments (Newport, 1977). Thus the child can take one or two instances of a verb in some frame as conclusive evidence that it is licensed in this syntactic environment.

5. What is used in this procedure for learning is part of what must be known by an accomplished speaker. Knowing the subcategorization privileges for each verb is part of what it means to know one's language. In contrast, many of the situational analyses constructed along the way by the semantic bootstrapper will not figure in the final definition of a verb.

In light of all these virtues, it would be nice if this theory turned out to be part of the truth about how the verb vocabulary is acquired. I provide some empirical evidence in its favor later. But first, some presuppositions of the position have to be defended before so apparently abstract a procedure can be considered viable at all. I turn now to such questions. But keep in mind that the approach here does not deny at all that observation of concomitant events is part of the answer to vocabulary acquisition. Rather, the idea is to remove part of the burden that a wide-ranging categorization of such events necessarily would impose and to make available a solution in the many cases where observation fails.

## Prolegomena to the Bootstrapping Hypotheses

The bootstrapping hypotheses involve a number of presuppositions that require demonstration in their own right, lest all learning questions be

begged. In company with all known theories of word learning, they presuppose that the human child, by natural disposition (or learning during the prelinguistic period, see footnote 1) is able to conceive of such notions as 'running' and 'looking' and implicitly understands that words can make reference to such acts and events. Past this background supposition, both semantic and syntactic bootstrapping procedures—but especially the latter—make very strong claims about the child's knowledge *as verb learning begins.* I now go through these claims, mentioning some of the experimental evidence that gives them plausibility.

### *Are the Rules Linking Semantics and Syntax Strong and Stable Enough to Support a Learning Procedure?*

If the syntactic structures associated with verbs are uncorrelated with—or hardly correlated with—their meanings, then the child can't learn much about the meanings by observing the structures. No one doubts the sheer existence of such form/meaning regularities, owing to the results achieved by a generation of linguists, notably Gruber (1968), Fillmore (1968a, 1968b), McCawley (1968), Vendler (1972), Jackendoff (1978, 1983), Levin (1985), Grimshaw (1983), and see the collection of papers in Wilkins (1988). But questions can be raised about the stability, degree, and scope of these relations. That is, how far can a syntactic analysis such as that in Table 2 succeed in partitioning the lexicon semantically for the learner?

I mention one line of investigation of these questions from our laboratory. Fisher, Gleitman, and Gleitman (in press), following Wexler (1970), reasoned as follows: If similarity in the range of subcategorization frames of verbs is correlated with similarities in their meanings, then subjects asked to partition a set of verbs (a) according to their meanings and (b) according to their licensed structures should partition the verb set in much the same ways. To test this idea, one group of subjects made judgments of meaning similarity for triads of verbs presented to them. Specifically, they chose the semantic outlier in each triad (e.g., shown *eat, drink,* and *sing,* they would probably choose *sing* as the outlier; but shown *eat, drink,* and *quaff,* they might choose *eat*). A semantic space for a set of verbs was derived from these data by tabulating how often two verbs stayed together (i.e., were not chosen as outlier) in the context of all other verbs with which they were compared. Presumably, the more often they stayed together, the more semantically similar they were. A second group of subjects gave judgments of grammaticality for all these same verbs in a large number of subcategorization frames. A syntactic space was derived in terms of the frame overlap among them. The similarity in the syntactic and semantic spaces provided by these two groups was then compared statistically.

The finding was that the frame overlap among the verbs is a very powerful predictor of the semantic partitioning. Verbs that behaved alike syntactically were, to a very interesting degree, the verbs that behaved alike semantically. For example, the semantic grouping of mental verbs (e.g., *think*) was predicted by acceptance of sentence complements, and the semantic grouping of transfer verbs (e.g., *give*) was predicted by acceptance of three noun phrases within the clause. Neatly enough, a semantic subgrouping of verbs of communication (or mental transfer, e.g., *argue, explain*) was predicted by acceptance of both the syntactic environments just mentioned, just as Zwicky proposed (see prior quotation). Thus taken jointly (i.e., in terms of the *range* of frames for a verb), the syntactic selections appear to have considerable semantic resolving power.

The strength of these results is particularly surprising considering the weakness and indirectness of the (triad) procedure used to construct the semantic similarity space and its heavy dependence on the choice of verbs considered. Thus these findings begin to show that a syntactic partitioning of the input can provide important evidence for a learner who is disposed to use such information—as was conjectured for the blind child (see Figure 3).

The subcategorization frames provide a relatively coarse-grained semantic partitioning of the verb set, quite obviously. Only a limited set of semantic properties are or could be encoded on the verb frames. According to Fisher et al. (in press), the semantic information in the verb frames is quite principled, limited to properties that (a) affect the argument structure, (b) are domain general (i.e., show up all through the lexicon), and (c) are closed to observation. This coarse partitioning is of considerable significance, however, for solving some of the problems posed in the first section of this article, for instance, deducing that *think* is a mental-state verb, distinguishing between *chase* and *flee,* and so forth, as I try to show later when I discuss our experimental findings. But keep in mind that the *syntax* is not going to give the learner information delicate and specific enough, for example, to distinguish among such semantically close items as *break, tear, shatter,* and *crumble* (Fillmore, 1968b). Luckily, these distinctions are almost surely of the kinds that can be culled from transactions with the world of objects and events.

### *Are the Semantic/Syntactic Relations the Same Cross-linguistically?*

The first proviso to the semantic usefulness of syntactic analysis for learning purposes is that the semantic/syntactic relations have to be materially the same across languages. Otherwise, depending on the exposure language, different children would have to perform completely different syntactic analyses to derive aspects of the meaning. And that, surely, begs the question at issue.

Recent theorizing in linguistics does support the idea that there are semantic/syntactic linkages that hold across languages. In a recent version of generative grammar (Government/Binding theory; see Chomsky, 1981), some of the relationships are stated as universal principles of language design. One example is the mapping of entities implied by the verb logic one-to-one onto noun-phrase positions in the clause: Every noun phrase in a sentence must receive one and only one thematic role (the *theta criterion*). Moreover, a related principle (the *projection principle*) states that the theta criterion will hold at every level of a derivation; in particular, that argument structure is preserved on the surface clause structures. This is just the organization required by a bootstrapper, semantic or syntactic.

Talmy (1975, 1985) investigated a number of typologically quite different languages and found a variety of striking similarities in how their syntax maps onto the semantics (though to be fair he's found some striking differences too). For those who prefer experimental evidence from linguistically naive subjects, Fisher et al. (in press)—in a very preliminary cross-linguistic foray with their method—showed that the relationship between being a verb of communication and accepting sentence complements and three noun phrases in the clause is as strong and stable in Italian as it is in English.

The two relationships just mentioned (that a noun phrase is assigned to each participant in the event and that verbs encoding the relation between an agent and a proposition accept sentence complements) are not only true cross-linguistically. They have a kind of cognitive transparency that makes them plausible as part of the presuppositional structure children might really bring into the language-learning situation. As Jackendoff (1978) put this point:

> In order to lighten the language learner's load further, it seems promising to seek a theory of semantics (that is, of conceptualization) in which the projection rules are relatively simple, for then the child can draw relatively straightforward connections between the language he hears and his conception of the world. The methodological assumptions for such a theory would be that syntactic simplicity ideally corresponds to conceptual simplicity; grammatical parallelisms may be clues to perceptual parallelisms; apparent grammatical constraints may reflect conceptual constraints. (p. 203)

From these and related arguments and demonstrations, I think the plausibility of the bootstrapping theories receives at least some initial defense.

### *Can the Learner Analyze the Sound Wave in a Way That Will Support Discovery of Syntactic Structure?*

There is a timing difference in the requirements of the semantic and syntactic bootstrapping approaches: for the latter, the learner has to be able

to parse the sentences heard in order to derive a syntactic analysis. Moreover, at least some of the mapping rules have to be in place before the verb meanings are known, or else the whole game is over. There is strong evidence supporting both these claims.

*Can infants parse?* Once upon a time not so very long ago, it was believed that babies could divide up the sound wave into words but not into phrases. This perspective necessitated complex theories for how learners could derive phrasal categories from the initial wordlike representations (Pinker, 1984). In retrospect, these ideas were somewhat improbable. For one thing, there is evidence that infants are sensitive to such physical properties of the wave form as change in fundamental frequency, silent intervals, and syllabic length, all of which are universal markers of phrase boundaries (Cooper & Paccia-Cooper, 1980; Klatt, 1975; Streeter, 1978; and see Fernald, 1984; Fernald et al., in press; and Schreiber, 1987, for relevant developmental evidence). As Gleitman and Wanner (1982) pointed out, the physical correlates of word segmentation are far more subtle and less reliable (see Echols & Newport, 1989, for an analysis).[11] More generally, Gleitman and Wanner's reading of the cross-linguistic facts about language learning led them to propose that the infant's analysis of the wave form was as a rudimentary phrase structure tree.

In a similar vein, Morgan and Newport (1981; Morgan, Meier, & Newport, 1987), showed in a series of artificial language-learning experiments that adults could learn phrase structure grammars if provided with phrase-bracketing information but not if provided only with word-level information. This finding led these investigators independently to the same proposal as Gleitman and Wanner's about the child's initial representation of the input wave forms. Hirsh-Pasek et al. (1987) and Jusczyk et al. (1988) showed that prelinguistic infants listen to maternal speech doctored so as to preserve phrase- and clause-bounding information in preference to speech doctored so as to becloud this information (see Gleitman, Gleitman, Landau, & Wanner, 1987, for a review of the evidence and its interpretation for a language acquisition theory).

The evidence just cited is not precise enough to give a detailed picture of the infant's phrasal parse. But even so, it is strong enough to support the

[11]Notoriously, word segmentation in a language like English is so fraught with ambiguity that new pronunciations (e.g., *nother* and *apron* replacing *other* and *napron*) are quite common. Moreover, there are long-lasting segmentation errors by children, for example, one 6-year old wrote, "The teacher said, Class be smissed!" The phrasal parses suggested by Gleitman and Wanner were rudimentary to the extent that the unstressed elements in the phrases were presumed to be less well analyzed than the stressed elements, and the phrases were unlabeled (but see Joshi & Levy, 1982, for evidence that much of labeling, or its equivalent, can be derived from "skeletal" representations in which there are configurations but no overt labels).

view that children, even in the prelinguistic period, impose an analysis on the wave form sufficient for partitioning it into phrases. It is incontrovertible that the 2- and 3-year-olds who are the real verb learners can achieve the analyses of input shown in Table 2, and which are a requirement for achieving the semantic partitioning of the verb set shown in Figure 3.

*Does the learner know the correspondence rules?'* A crucial further requirement for the bootstrapping hypotheses is that the child understand the semantic values of the subcategorization frames. A child who recovers the meaning from observation and who is to deduce the structures therefrom has to know what the semantics of the verb implies about the syntactic structures licensed. And a child who recovers the syntactic structures licensed for verbs from the linguistic contexts in which they are heard has to know what semantic elements are implied by participation in these structures. As Jackendoff emphasized, the burden of learning would certainly be reduced for a child in possession of such information. But do real learners actually have it? There is striking evidence that they do.

Golinkoff, Hirsh-Pasek, Cauley, and Gordon (1987) developed a very useful paradigm for studying very young children's comprehension. Essentially, they adapted the preferential looking procedure designed by Spelke (1982) for studying infant perception. The setup for the language case is shown in Figure 4. The child sees different scenes displayed on two video screens, one to the left, one to the right. The scenes are accompanied by some speech stimulus. The mother wears a visor so that she cannot observe the videos and so cannot give hints to her child. Hidden observers are so positioned that they cannot observe the video, but they can observe which way the child is looking and for how long. It turns out that children look sooner and longer at the video that matches the speech input.

In a first demonstration relevant to the syntactic bootstrapping hypothesis, Hirsh-Pasek, Golinkoff, Fletcher, DeGaspe Beaubien, and Cauley (1985) showed that 17-month-old children—many of whom had never put two words together in an utterance and knew few if any verbs—understand some facts about the semantic values of English constructions. Two simultaneous videos showed cartoon characters known to the children interacting. For some subjects, the stimulus sentence was *Big Bird is tickling Cookie Monster.* For the others, it was *Cookie Monster is tickling Big Bird.* The children demonstrated by their preferential looking that they knew which sentence described which observed event: They looked longer at the screen showing Big Bird tickling Cookie Monster when they heard the former sentence and at the screen showing Cookie Monster tickling Big Bird when they heard the latter sentence. That is, these children recognize the order of phrases (or something approximating phrases) within the heard

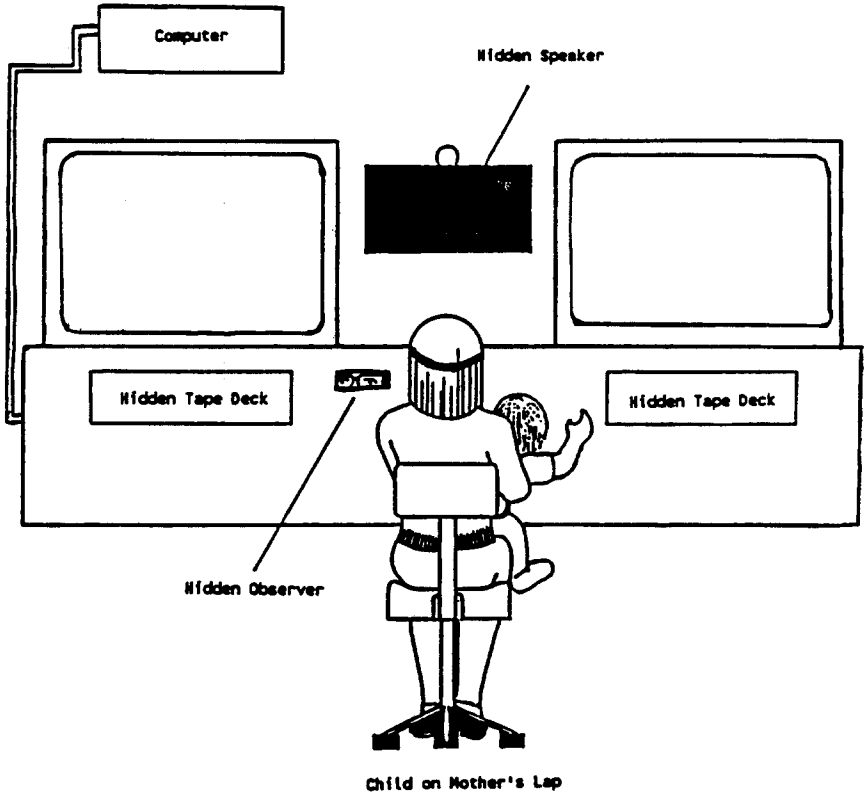


FIGURE 4 Apparatus for the preferential looking experiments. (Reproduced from Naigles, in press, with permission of *Journal of Child Learning,* published by Cambridge University Press.)

sentences and also understand the semantic significance of the ordering for the propositional interpretation of English speech (see Slobin & Bever, 1982, for cross-linguistic evidence on this topic). Note that in this and all other experiments I'll be describing, all the depicted participants are animate, so there's no room for trivializing interpretations such as the strategy of assigning the animate entity to the subject position.

My colleagues and I (Hirsh-Pasek, Gleitman, Gleitman, Golinkoff, & Naigles, 1988) used this same procedure to investigate one more property of the mapping rules, namely the causative structure for which Bowerman (1974) had found many innovative uses by youngsters: Roughly, intransitive motion verbs (e.g., *Big Bird turns*) can be "transitivized" in English and then will express the causal agent as well (*Cookie Monster turns Big Bird*).

To study this question using the preferential looking method, it is necessary that both entities appear in the stimulus sentence; otherwise the children may use the relatively trivial strategy of looking at the stimulus showing Big Bird if and only if Big Bird is mentioned. Hence, the stimuli used were, for example, *Big Bird is turning Cookie Monster* and *Big Bird is turning with Cookie Monster.* One video showed the two characters turning side by side, and the other video showed one character physically causing the other to turn. In addition to verbs like *turn* that were known to the 2-year-old subjects, unknown ones (by maternal report) were also used. For example, the characters were shown crossing their arms back and forth, or one crossing the arms of the other, along with the stimuli *Big Bird is flexing with Cookie Monster* and *Big Bird is flexing Cookie Monster.* At age 27 months, almost every child tested showed the effect of the structure by looking longest at the syntactically congruent screen.

The conclusions to be drawn are very important ones for the syntactic bootstrapping hypothesis. The paired *actions* are the same, for example, both are of turning in a circle or both are of crossing the arms. What differs is whether a causal agent of that action is also present in that scene. The children seem to know that only the transitive use of the verb can be expressing that cause. More strongly, that causal agent cannot be in an oblique argument position (the *with* phrase). Most strongly of all, they appear to realize that the *with* phrase excludes a transitive reading. This implies that toddlers who are primitive in their own speech are doing an astonishing amount of parsing of the speech of others and are interpreting the structures semantically.

Prior demonstrations of knowledge of mapping rules have generally been with much older children. For instance, Bowerman noted that most spontaneous overgeneralizations of the causative structure ("Don't eat the baby!") are later, in the 3- to 5-year-old period. Pinker and his colleagues have offered many compelling demonstrations of a variety of mapping rules but again mainly with 3- to 5-year-olds (e.g., Pinker, Lebeaux, & Frost,

1987).[12] These findings give general support to the idea that learners recruit the semantic/syntactic correlations somewhere during the course of learning. But the early appearance of these skills is crucial as support for the notion that the child has the mapping rules under control early enough for them to contribute to the acquisition of the verb meanings themselves. As just described, we have documented that 27-month-olds have these capabilities.

## Investigations of Syntactic Bootstrapping

So far I've tried to show that a number of presuppositions of syntactic bootstrapping are reasonable: The language does exhibit strong and stable syntactic/semantic correlations, and these powerfully predict adult classificatory behavior; infants in the prelinguistic period can and do parse sentences to recover the analyses required for extracting subcategorization frame information; such phrasal information is a requirement for language learning, at least for adults in the artificial language-learning laboratory; children at a very young age and language-learning stage understand the semantic values of at least some syntactic frames.

All of these findings were prolegomena to the syntactic bootstrapping approach. They were adduced because it is critical to determine that the child can come up with the analyses that the position presupposes. But now that I have presented at least some preliminary support that children can meet these prior requirements, the next question is: Do they use syntactic evidence to decide on the meaning of a new word?

### *Basic Findings*

The first, and justly famous, work on this topic was done by Roger Brown (1957). He showed 3- to 5-year-olds a picture in which, say, spaghettilike stuff was being poured into a vessel. This scene was always the same one, but some of the children were asked to show *some blick,* others *a blick,* and still others *blicking.* The childrens' choices were, respectively, the spaghetti, the vessel, and the action. Evidently, the semantic core of the

[12]But see also Naigles, Gleitman, and Gleitman (in press) for a demonstration that 2-year-olds understand the significance of new motion transitives, even though they may not be brave enough to invent any until they are 3. The children here were asked to act out scenes using a Noah's Ark and its animal inhabitants. For instance, the child might be told to act out "Noah brings the elephant to the ark." But some of the stimuli were more unusual, for example, "Noah comes the elephant to the ark" or "The elephant brings to the ark." The children by their acting-out performances showed that they thought transitive *come* means 'bring' and that intransitive *bring* means 'come.'

word classes affects the conjecture about the aspect of the scene in view *that is being labeled linguistically.*

Brown's results, though alluded to respectfully, just sat there for 20 years or so because, in this respect as in many others, Brown was a theorist ahead of his time. Eventually MacNamara took up and advanced these ideas: In his important 1972 paper, he argued forcefully for the place of language structure in language acquisition. Experimentally, Katz, Baker, and MacNamara (1974) showed that children as young as 18 months used the structure in which new nouns appeared (*a gorp* vs. *Gorp*) to decide whether a new word encoded a class or an individual (i.e., a doll of the gorpish sort or some doll named Gorp). Thus the lexical category assignments of words were shown to carry semantic implications, and these were evidently recruited by learners to deduce the aspect of the world being encoded by the new word.

Naigles (in press), working in my lab and also in the labs of Hirsh-Pasek at Temple University and of Golinkoff at the University of Delaware, extended this kind of demonstration to the case of verb learning (i.e., to the usefulness of syntax for drawing semantic inferences within a single lexical category), thus giving the first direct demonstration of syntactic bootstrapping at work.

Children (mean age 24 months) were again put into the preferential looking situation. This time, however, their task was to decide between two utterly disjoint interpretations of a new verb. In the training (learning) period, they saw a single screen and the following mad event: A rabbit is pushing a duck down into a squatting position with his left hand (these were people dressed up as rabbits and ducks so they did have hands). The duck pops up, and the rabbit pushes him down again, and so on. Simultaneously, both rabbit and duck are making big circles in the air with their right arms. Some children heard a voice say *The rabbit is gorping the duck* and other children heard *The rabbit and the duck are gorping* as they watched this scene.

Subsequent to this observation, two new videos appeared on two screens, as shown in Figure 4. On one screen, the rabbit was pushing the duck down (but with no arm-wheeling). On the other screen, rabbit and duck were wheeling their arms (but with no squatting or forcing to squat). The child then was cued by the voice saying the (syntactically uninformative) sentences *Where's gorping now? Find gorping!* The child's looking time at the screens as a function of her syntactic introducing circumstances was recorded (double-blind as usual, i.e., neither the mother nor the experimenters knew which event was being depicted on the child's left and which was on her right during the test).

Naigles's result was that virtually every infant tested—and there were many, this being a PhD thesis—showed the effect of the syntactic introducing circumstance. Those who heard the transitive sentence apparently

concluded that *gorp* means 'force-to-squat'. Those who heard the intransitive sentence decided that *gorp* means 'wheel the arms'.[13]

What shall we conclude from this experiment? Clearly the child uses the event context in some way to license conjectures about a verb meaning. But in this case, "The Main Event" is ambiguous not only in principle but in fact. Under these trying circumstances, at least, the learner attends to the information potential of the semantically relevant syntactic evidence. The position I have tried to defend is that the zoom-lens effect of the structural context is critical for vocabulary learning in the real world of multiply interpretable scenes and events.

Notice also what *should not* be concluded from this experiment. Whatever the real power of syntactic bootstrapping when the child is provided with a set of frames for some verb, that full power was not exploited in the present experiment. Only the usefulness of a single syntactic property as disambiguator was tested. Therefore, even if (as I doubt very much) there is enough information in the subcategorization frames of a language to distinguish between 'squat' meanings and 'wheel' meanings, there certainly is not enough evidence in one or two frames to make this distinction. The verb meanings, insofar as they were acquired at all in this experiment, were learned by inspecting the real-world contingencies, much as Pinker has suggested. But as so often—just about always, if I'm right—there was a choice in this situation for how to conceive the scene semantically. How is this choice adjudicated? What Naigles showed was that the syntactic evidence guides the child observer, determining the choice among situationally available options.

### *A Question of Scope*

So far the experiments I have mentioned have lingered nervously around a few constructions, for example, the lexical causative in English, which is a notorious focus of syntactic extension by adults as well as children. Even if it is accepted that children sometimes do use syntactic evidence to bolster their semantic conjectures, how broad can the scope of such a procedure be? Maybe its role is just to clean up a few little details that are hard to glean from the world—just reverse linking, as Pinker has sometimes put the

[13]In the present experiment, the intransitive sentence contained a conjoined nominal (*The duck and the rabbit*) and this might be seen as a defect: Maybe the child knows the difference between a preverbal and a postverbal nominal rather than the difference between a transitive and an intransitive structure. This interpretation is effectively excluded by the version presented earlier (Hirsh-Pasek et al., 1988) in which the two noun phrases appear in different argument positions, one serially before and one after the verb (*Big Bird is turning with Cookie Monster*). For elegance, however, it certainly would be nice to redo the present experiment with the stimulus type used in the former one.

matter. To investigate the real scope of children's exploitation of the syntactic environment in learning new verb meanings, my colleagues and I have now studied 3- and 4-year-old learners. Let me first suggest why we've now turned to this older population.

The studies I've described so far, performed with children 2 years old and younger, yield evidence that satisfies an explanatory demand of this approach: The bootstrapping procedure has to be able to operate very early in the child's linguistic life, else its role is restricted to a late and ancillary method for refining the observation-based conjectures. But the preferential looking paradigm (which is one of very few that work with toddlers) is too much of a straightjacket to be the only vehicle for investigation of this approach. It is tedious in the extreme to set up (requiring the preparation of movies, etc.), takes hoards of infants to carry out (for some scream or sleep or worse and have to be removed from the premises; and only a few trials can be presented even to the more docile infants). Moreover, it is likely that children's knowledge of the linking rules expands as their language knowledge grows, creating more latitude within which they can learn new meanings from linguistic evidence. (After all, in the end we can do it by looking in the dictionary.)

So now that it has been shown (in Naigles's work) that the use of syntax in verb learning begins very early, certainly by 24 months, it is reasonable to refine and expand such findings in studies of older—but still very young—learners. Specifically, Fisher, Hall, Rakowitz, and Gleitman (1989) asked whether 3- and 4-year-olds would give us meanings in response to linguistic-situational stimuli upon request. The idea derived from a manipulation attempted by Marantz (1982). He had asked whether children are as quick to learn noncanonical as canonical mappings of semantics onto syntax. He introduced children to novel verbs as they watched a movie. For instance, one movie showed a man pounding on a book with his elbow. Marantz's question was whether children were as quick to learn that *The book is moaking Larry* (the noncanonical mapping) was a way of describing this scene as that *Larry is moaking the book* (the canonical mapping) was a way of describing the scene.

Although the manipulation was an interesting one, unfortunately Marantz never asked the children how they interpreted the scene, so his results are not really relevant to understanding the child's perception of syntactic/semantic correlations. That is, Marantz *presupposed* that a scene viewed has only a single interpretation, an idea I have strenuously opposed throughout this discussion. We now revised this experiment, changing the measure so we could find out about the child's comprehension in these circumstances. In essence, we asked how the nonsense word is interpreted within differing linguistic environments. As a first step, we showed the *moaking* scene (in which Larry pounds the book with his elbow) to adults.

If we said, "This scene can be described as a moaking scene" and then asked them what /moak/ meant, they said 'pounding'. And if instead we showed them the scene and said, "This is Larry moaking the book," they still asserted that /moak/ means 'pound'. But when we showed them the scene and said, "This is the book moaking Larry," they answered that /moak/ means 'hurt'.

This suggests that adults make use of the fact that particular surface syntactic structures are associated with particular semantic values. They seem to bootstrap the meaning from examination of the scene taken together with its syntactic expression, just as the syntactic bootstrapping procedure claims. To be sure, the contextless presentation of /moak/ with this scene irresistably yields the concept 'pound' as its interpretation. So there is much to be said for the idea of *salience* in the interpretation of events (though, to be sure, no one knows *what* exactly). But the important point is that there is a categorical shift in interpretation of the same scene—to a less salient, but still possible, interpretation—in response to its linguistic setting; namely, 'pound' if Larry is in the subject position, but 'hurt' if the book is in that position.

Fisher et al. (1989) now adapted this procedure for children. We took advantage of the idea, popularized by such Penn developmentalists as Waxman, Gelman, Macario, and Massey, that preschoolers will do just about anything to help out a puppet. We introduced a puppet saying, "This puppet sometimes talks puppet-talk so I can't understand him; can you help figure out what he means?" Our sixteen 4-year-old subjects were happy to oblige. They were shown videotaped scenes in which animals were performing certain acts. For example, a rabbit appeared, looked to the left, and then ran rapidly off the screen toward the right. Directly behind the rabbit ran a skunk, also disappearing at the right. So this scene is one that can be interpreted as either one of chasing or of fleeing. Then the child would hear the puppet say either "The rabbit is gorping the skunk" or else "The skunk is gorping the rabbit."

The scenes/structures we investigated were designed to ask whether children are sensitive to a variety of syntactic cues to interpretation. These are shown in Table 4. The first property investigated was the number of argument positions (Stimuli 1 and 2). For instance, rabbit and elephant are shown eating/feeding and the puppet says either "The rabbit moaks" or "The elephant moaks the rabbit." The second property was canonical structural positions of agent and patient (Stimuli 3 and 4, e.g., *ride/carry*), and the third was the structural positions taken together with prepositional markers of the oblique roles (Stimuli 5 and 6, e.g., *give/take*). Thus we now began to investigate the scope of the structural/semantic linkages to which learners may be sensitive.

The pairs chosen were designed to be revealing of solutions to the

TABLE 4
Scenarios and Their Sentential Descriptions

| *Scenario* | *Sentence* |
|---|---|
| 1. (a)Rabbit eating. | The rabbit moaks. |
| (b)Elephant feeding rabbit. | The elephant moaks the rabbit. |
| 2. (a) Monkey pushing elephant. | The monkey pumes the elephant. |
| (b) Elephant falling. | The elephant pumes. |
| 3. (a) Monkey riding elephant. | The monkey gorms the elephant. |
| (b) Elephant carrying monkey. | The elephant gorms the monkey. |
| 4. (a) Rabbit fleeing skunk. | The rabbit zarps the skunk. |
| (b) Skunk chasing rabbit. | The skunk zarps the rabbit. |
| 5. (a) Rabbit giving a ball to elephant. | The rabbit ziffs a ball to the elephant. |
| (b) Elephant taking a ball from rabbit. | The elephant ziffs a ball from the rabbit. |
| 6. (a) Skunk putting blanket on monkey. | The skunk is biffing a blanket on the monkey. |
| (b) Skunk covering monkey with a blanket. | The skunk is biffing the monkey with a blanket. |

*Note.* All children were exposed to the same six scenes (each scene has two plausible interpretations, called (a) and (b) in the left-hand column). Along with these scenes, half of the children heard (a) stimulus sentences and half heard (b) stimulus sentences (with appropriate counterbalancing across children and stimuli).

problem that I have discussed throughout: Single scenes, multiply interpretable, are shown but accompanied by a novel verb; this verb is introduced to half of the children in one construction and to the other half in another construction. The question is whether the introducing syntactic environment enables the observing child to fix on a single meaning for the novel verb.

The outcomes of this experiment were extremely strong. Not every young child responded to each scene/sentence example (sometimes they said something irrelevant or just looked piteously at the experimenter). But when they did respond, their guess was guided heavily by the syntactic frame. For instance, consider the scene in which the rabbit appears to flee, pursued by a skunk. Six (of eight) children who heard the puppet say "The rabbit zarps the skunk" said that /zarp/ means 'run away', while only one guessed 'chase'; the eighth child did not respond. Symmetrically, all eight who heard "The skunk zarps the rabbit" said that /zarp/ means 'chase.' Of the 84 relevant responses made by these children, 71 were congruent with the semantic value implied by the syntactic structure and only 13 were inconsistent with the structural information, a statistically highly reliable

result. Moreover, *for each child, for each scene, and for each syntactic type,* the number of syntactically congruent responses was greater than the noncongruent responses. The level of congruence (about 85%) was approximately the same for all three semantic/syntactic relations studied.

One might object that these children were merely paraphrasing verbs that they previously knew to occur in these syntactic environments. That is true, but it does not take away seriously from our interpretation of these findings: Evidently, the children knew that the appropriate meaningful verb had to be one that fit both with the scene and with the sentence structure heard. This is the reverse of Pinker's claim that the verb meanings must be acquired by extralinguistic observation *in advance of,* and as the basis for, deducing their appropriate syntactic structures. But the results are exactly those expected in the syntactic bootstrapping approach. The syntax guides the choice of interpretive options in ambiguous observational circumstances. As just about all observational circumstances are ambiguous, I believe this is saying a lot about the explanatory value of the learning procedure proposed.

### *The Input*

One of several holes in our present evidence has to do with the characteristics of caretaker speech. I have presented a single example corpus (Table 2) tending to support the idea that caretaker speech is rich enough to yield quite a full range of structures to support a strong variant of the syntactic bootstrapping procedure. And this corpus was for a mother speaking to a blind child, whose word-learning situation may be quite special. We are now analyzing an extensive corpus of mother/child speech in a naturalistic setting (originally collected by Landau and Gleitman) to see whether children characteristically receive the range of structures adequate to support a realistic syntax-based procedure (Lederer, Gleitman, & Gleitman, 1989). So far, the prospects from this larger data base look good. Lederer et al. found that each of the 24 verbs most often used by these mothers to their children has a distinctive syntactic distribution. When the usages are pooled across mothers, these distinctions are preserved unmuddied.

The next question is whether the syntactic distributions culled from maternal speech map coherently onto the target semantic space (namely, the semantic space as known by adults). As independent assessment of the *adult* semantic relations among these verbs is required as the evidence. As a first pass, Lederer et al. (1989) investigated these verbs in the kind of manipulation employed by Fisher et al. (in press), namely asking adult subjects for judgments of the semantic outlier in all triads of these verbs. The question of interest, of course, concerns the correlation between the semantic

similarity space as it emerges from these adult triad judgments and the overlaps and nonoverlaps in the syntactic behavior of the verbs in the maternal corpora (both these similarity spaces are extracted from the data by a cluster analysis). These correlations turn out to be massive and highly reliable, with the maternal subcategorization patterns accounting for about 50% of the variance in the adult triad patterns. Considering the roughness of the semantic analysis to which the maternal speech was submitted in this first test, I consider these findings to be the strongest evidence thus far in demonstrating the general feasibility and power of syntactic bootstrapping.

## PART III: CONCLUSIONS

I began discussion by acknowledging the intuitive power of Locke's view that words are learned by noticing the real-world contingencies for their use. Then I tried to show that such a word-to-world mapping, unaided, was in principle insufficiently constrained to answer the question of how the child matches the verb items with their meanings. The solution that I and my colleagues have offered is that semantically relevant information in the syntactic structures can rescue observational learning from the sundry experiential pitfalls that threaten it. This theory, of course, is the very opposite of intuitive. But when probable solutions fail, less probable ones deserve to be considered. I therefore sketched a rather wide-ranging empirical review that we have undertaken to see whether, after all, children might not be deducing some of the meanings from their knowledge of structural/semantic relations. I believe that the evidence we now have in hand materially strengthens the plausibility of the viewpoint.

Still, the conclusions that can be drawn currently about the generality and pervasiveness of syntactic bootstrapping must be exceedingly tentative, on a variety of grounds. Some of these I have discussed: No one has more than a glimmer of an idea about just how the verb lexicon is organized semantically, and therefore we cannot be very precise about the semantic information potential of the frame specifications. Also, we have at present only the most meager data concerning the orderliness and richness of the child's syntactic input. Facts about cross-linguistic similarities in the syntax/semantics correspondences are even more fragmentary.

Moreover, the position I have tried to defend is that the *range* of frames associated with each verb, operating jointly, narrow the hypothesis space for the verb meaning to such a degree that the faltering and probabilistic observational mapping of words to their meanings can succeed. But the experiments with children that I have reported show only the effects of single frames in the presence of multiply interpretable scenes. These demonstrate the focusing (zoom-lens) power of the syntax for disambigu-

ating aspects of those scenes. But the stronger version of this hypothesis—that the meaning of the verb falls out directly from the range of frame specifications—has yet to receive direct experimental review and confirmation. The accumulating power provided by joint operation of frame/verb relations was inferred for the blind child only by showing that the data base provided by the mother was rich and restrictive enough to support such an analysis (see Figure 3), should the child have been inclined to perform it; Lederer's studies are designed to generalize such a conclusion about the data base. In addition, the triad studies with adults (Fisher et al., in press) show that the range of frames associated with verbs is powerfully correlated with a global semantic space that people construct when asked to sort verbs according to their semantic similarity.

Despite these encouraging initial results, what is still required is direct evidence of the semantic resolving power of the complete frame sets associated with particular verbs, for children and adults. Though inquiries on this matter are on our experimental agenda, evidence is not now available, except in the form of parlor games popular in our lab: For intellectuals playing games, at least, and for selected verbs, it's possible to guess which verb an individual has in mind by inspection of a set of frames presented as a sequence of phrasal category labels (under some strong assumptions, i.e., that the frames stand in entailment relations to each other). The appropriate experimental review has yet to be carried out in these terms, so it remains in question just how much of the burden of observational learning can be reduced by the learner's attention to syntactic evidence.

In addition, there are numerous problems with our analyses of input corpora that I have altogether skirted so far. For example, it is not an easy task to decide which structures co-occurring with verbs should actually be considered part of the frame specifications and which are merely adjuncts. To construct Table 2 (and in Lederer's ongoing work) we had to make some choices, but some of them may be wrong. And if we had these problems in assigning structural descriptions to the mother's utterances, isn't the learner similarly beset?[14]

Another problem is the idiomatic verb uses that I mentioned in passing (footnote 10, e.g., *John saw his victim out of the room, John looked his enemies in the eye*). It may be significant that these monstrosities are just

[14]Lederer and Kelly (1989) are now testing whether prosodic distinctions typically disambiguate the readings of sentences in this regard. Pilot laboratory results suggest that native speakers distinguish their pronunciations of ambiguous sentences depending on whether the adjunct or complement reading is intended; native listeners correctly guess which reading was intended about 80% of the time. As "motherese" is characterized by exaggerated intonation contours, and infants show strong preference for this style of speech (Fernald et al., in press), it is likely that children have a physical basis for distinguishing these boundary types.

about totally absent from the maternal corpora we have examined, but absence in fact rather than in principle is a pretty weak reed on which to build so strong a position as the one I've tried to defend.

The largest problem of all is how learners acquire the semantic/syntactic linking rules in the first place. Bowerman's evidence, and all the findings I have just discussed, are understandable only (so far as I can see) by asserting that learners are in possession of such linking rules. But where did they come from? In the present discussion, I have subscribed to a version of Jackendoff's hypothesis that the linking rules are somehow cognitively transparent to the child. But because there is at least *some* cross-linguistic variance in such syntactic/semantic regularities (see Talmy, 1985), I admit that I'd be happier to find that they could be derived from some more primitive categories or functions. The problems here cry out for serious investigation.

In light of the various issues just mentioned, one must remain agnostic about both of the bootstrapping proposals, at present. But I hope I have persuaded you that the prospects they open for explanation of the verb-learning feat are enticing enough to make continued investigation seem worthwhile.

It remains to point out that, by their nature, both semantic and syntactic bootstrapping are perilous and errorful procedures, and their explanatory power must be evaluated with this additional proviso in mind. Bowerman's children, drawing syntactic conclusions from meaningful overlap, are sometimes wrong; for instance, one can't, but children sometimes do, say, "Daddy giggled me." To take another kind of case, *exit, enter, reach,* and *touch* differ from most verbs describing directed motion through space in not requiring prepositional phrases to express the motion paths (compare *come into the room* but *enter the room*). One outcome of this varying mapping of meaning onto form is errorful learning (the child may say, "I touched on your arm") and its end point, language change (while *exit the stage* was the more common in Shakespeare's time, *exit from the stage* is now on the ascendancy). But syntactic bootstrapping is subject to related kinds of error. For instance, children in the learning period may exchange *push* and *pull,* and *infer* and *imply* have come to be used interchangeably by many adults, perhaps because their syntactic (as well as situational) overlap is misleading. Short of changing the language, how do learners recover from such errors?

The position I have been urging is that children usually succeed in ferreting out the forms and the meanings of the language just because they can play off these two imperfect and insufficient data bases (the saliently interpretable events and the syntactically interpreted utterances) against each other to derive the best fit between them. Neither syntactic nor semantic bootstrapping works all the time, nor taken together do they

answer all the questions about how children acquire the verb vocabulary and argument structures. But I have tried to show that each of these procedures works very well indeed when it does work, so the wise child should, and probably does, make use of both of them.

## ACKNOWLEDGMENTS

I thank quite a few colleagues who helped me in the preparation of this article. The first is my husband, Henry Gleitman, who—as always—contributed a large share of the ideas and most of whatever organization and coherence this article contains. Anne Lederer has also been a crucial aid in offering significant ideas, and also helped with a good deal of bureaucratic labor. I had the advantage of discussing the ideas contained herein with Paul Bloom, Noam Chomsky, and Steven Pinker. I also thank several colleagues who read earlier drafts of the article and gave me important comments. These were Adele Abrahamson, Steven Crain, Cynthia Fisher, Roberta Golinkoff, Kathy Hirsh-Pasek, Barbara Landau, Letitia Naigles, Elissa Newport, Thomas Roeper, and Kenneth Wexler. My thanks go as well to Richard Beckwith, who kindly sent me preliminary data and commentary from his work for use in my discussion. The approach described herein was developed in collaboration with a number of colleagues whose work is cited throughout the text. I want also to express appreciation for a University of Pennsylvania Biomedical Research Grant (sponsored by the National Institute of Health under Grant #2-S07-RR-07803-23), which underwrote the more recent experimental work.

## REFERENCES

Acredolo, L., & Evans, D. (1980). Developmental changes in the effects of landmarks on infant spatial behavior. *Developmental Psychology, 16,* 312–318.

Armstrong, S., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13,* 263–308.

Baillargeon, R., Spelke, E., & Wasserman, S. (1985). Object permanence in 5-month old infants. *Cognition, 20,* 191–208.

Ball, W., & Vurpillot, E. (1976). Perception of movement in depth in infancy. *L'Année Psychologique, 76,* 383–391.

Beckwith, R., Tinker, E., & Bloom, L. (1989, October). *The acquisition of non-basic sentences.* Paper presented at the Boston Child Language Conference, Boston.

Berwick, R. C. (1982). *The acquisition of syntactic knowledge.* Cambridge, MA: MIT Press.

Bornstein, M. H. (1975). Qualities of color vision in infancy. *Journal of Experimental Child Psychology, 19,* 401–419.

Bowerman, M. (1974). Learning the structure of causative verbs: A study in the relationship

of cognitive, semantic, and syntactic development. *Proceedings of Research in Child Language Development, 8,* 142–178.

Bowerman, M. (1982). Evaluating competing linguistic models with language acquisition data: Implications of developmental errors. *Quaderni di Semantica, III,* 5–66.

Brown, R. (1957). Linguistic determinism and parts of speech. *Journal of Abnormal and Social Psychology, 55,* 1–5.

Bruner, J. S. (1974/1975). From communication to language: A psychological perspective. *Cognition, 3,* 255–287.

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

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: MIT Press, Bradford Books.

Chomsky, N. (1959). Review of B. F. Skinner, *Verbal behavior. Language, 35,* 26–58.

Chomsky, N. (1975). *Reflections on language.* New York: Random House.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht, Holland: Foris.

Chomsky, N. (1982). *Noam Chomsky on the generative enterprise: A discussion with R. Hybregts & H. van Riemsdijk.* Dordrecht, Holland: Foris.

Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language, 15,* 327–335.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech.* Cambridge, MA: Harvard University Press.

Crain, S., & Fodor, J. D. (in press). Competence and performance in child language. In E. Dromi (Ed.), *Language and cognition: A developmental perspective.* Norwood, NJ: Ablex.

Echols, C. H., & Newport, E. L. (1989). *The role of stress and position in determining first words.* Unpublished manuscript, University of Rochester, Rochester, NY.

Fernald, A. (1984). The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, & R. Golinkoff (Eds.), *The origins and growth of communication.* New Brunswick, NJ: Ablex.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (in press). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language.*

Field, J. (1976). Relation of young infant's reaching to stimulus distance and solidity. *Child Development, 50,* 698–704.

Fillmore, C. (1968a). The case for case. In E. Bach & R. Harms (Eds.), *Universals of linguistic theory* (pp. 1–91). New York: Holt, Rinehart & Winston.

Fillmore, C. (1968b). Lexical entries for verbs. *Foundations of Language, 4,* 373–393.

Fisher, C., Gleitman, H., & Gleitman, L. (in press). Relationships between verb meanings and their syntactic structures. *Cognitive Psychology.*

Fisher, C., Hall, G., Rakowitz, S., & Gleitman, L. (1989). *Verb-meaning acquisition is guided by syntactic evidence.* Unpublished manuscript, University of Pennsylvania, Philadelphia.

Fodor, J. A. (1981). The present status of the innateness controversy. In J. A. Fodor (Ed.), *Representations* (pp. 257–316). Cambridge, MA: MIT Press.

Gibson, E. J., & Spelke, E. (1983). The development of perception. In J. H. Flavell & E. Markman (Eds.), *Cognitive development: Vol. III of P. Mussen (Ed.), Handbook of cognitive psychology* (pp. 1–76). New York: Wiley.

Gibson, E. J., & Walker, A. S. (1984). Intermodal perception of substance. *Child Development, 55,* 453–460.

Gleitman, L., Gleitman, H., Landau, B., & Wanner, E. (1987). Where learning begins: Initial representations for language learning. In E. Newmayer (Ed.), *The Cambridge Linguistic Survey* (Vol. II, pp. 150–193). New York: Cambridge University Press.

Gleitman, L., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 1–50). New York: Cambridge University Press.

Gold, E. M. (1967). Language identification in the limit. *Information and Control, 10,* 447–474.

Golinkoff, R. M. (1986). "I beg your pardon?": The preverbal negotiation of failed messages. *Journal of Child Language, 13,* 455–476.

Golinkoff, R., Harding, C., Carlson, V., & Sexton, E. (1984). The child's perception of causal events: The distinction between animate and inanimate objects. *Advances in Infancy Research, 3,* 145–165.

Golinkoff, R., Hirsh-Pasek, P., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14,* 23–46.

Grimshaw, J. (1981). Form, function, and the language acquisition device. In C. L. Baker & J. J. McCarthy (Eds.), *The logical problem of language acquisition* (pp. 183–210). Cambridge, MA: MIT Press.

Grimshaw, J. (1983). Subcategorization and grammatical relations. In A. Zaenen (Ed.), *Subjects and other subjects.* Bloomington: Indiana University Linguistics Club.

Gruber, J. S. (1968). Look and see. *Language, 43,* 937–947.

Hirsh-Pasek, K., Gleitman, H., Gleitman, L. R., Golinkoff, R., & Naigles, L. (1988, October). *Syntactic bootstrapping: Evidence from comprehension.* Paper presented at Boston Language Conference, Boston.

Hirsh-Pasek, K., Golinkoff, R., Fletcher, A., DeGaspe Beaubien, F., & Cauley, K. (1985, October). *In the beginning: One word speakers comprehend word order.* Paper presented at Boston Language Conference, Boston.

Hirsh-Pasek, K., Kemler-Nelson, D. G., Jusczyk, P. W., Wright, K., Cassidy, K., Druss, B., & Kennedy, B. (1987). Clauses are perceptual units for young infants. *Cognition, 26,* 269–286.

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

Jackendoff, R. (1983). *Semantics and cognition.* Cambridge, MA: MIT Press.

Jones, S., Landau, B., & Smith, L. (1986, April). *The importance of shape in early lexical learning.* Paper presented to the Eastern Psychological Association, New York.

Joshi, A. K., & Levy, L. S. (1982). Phrase structure trees bear more fruit than you would have thought. *American Journal of Computational Linguistics, 8,* 1–12.

Jusczyk, P., Hirsh-Pasek, K., Kemler-Nelson, D., Kennedy, L., Woodward, A., & Piwoz, J. (1988). *Perception of acoustic correlates of major phrasal units by young infants.* Unpublished manuscript, University of Oregon.

Katz, N., Baker, E., & MacNamara, J. (1974). What's in a name? A study of how children learn common and proper names. *Child Development, 45,* 469–473.

Keil, F. C. (1979). *Semantic and conceptual development.* Cambridge, MA: Harvard University Press.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics, 3,* 229–240.

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

Lasky, R. E., & Gogol, W. C. (1978). The perception of relative motion by young infants. *Perception, 7,* 617–623.

Lederer, A., Gleitman, L., & Gleitman, H. (1989, October). *Input to a deductive verb acquisition procedure.* Paper presented at the Boston Child Language Conference, Boston.

Lederer, A., & Kelly, M. (1989). *Prosodic marking of the adjunct/complement distinction.* Unpublished manuscript, University of Pennsylvania, Philadelphia.

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

Levin, B. (1985). Lexical semantics in review: An introduction. In B. Levin (Ed.), *Lexical semantics in review* (Lexicon Project Working Papers). Cambridge, MA: MIT Center for

Cognitive Science.
Locke, J. (1964). *An essay concerning human understanding.* Cleveland: Meridian Books. (Original work published 1690)
MacNamara, J. (1972). Cognitive basis for language learning in infants. *Psychological Review, 79,* 1–13.
Marantz, A. (1982). On the acquisition of grammatical relations. *Linguistische Berichte, 80,* 32–69.
Markman, E., & Hutchinson, J. (1984). Children's sensitivity to constraints on word meaning: Taxonomic vs. thematic relations. *Cognitive Psychology, 16,* 1–27.
McCawley, J. D. (1968). The role of semantics in a grammar. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory* (pp. 125–170). New York: Holt, Rinehart & Winston.
Morgan, J., Meier, R., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of intonational and morphological marking of phrases to the acquisition of language. *Cognitive Psychology, 19,* 498–550.
Morgan, J., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior, 20,* 67–85.
Naigles, L. G. (in press). Children use syntax to learn verb meanings. *Journal of Child Language.*
Naigles, L. G., Gleitman, L., & Gleitman, H. (in press). Children acquire word meaning components from syntactic evidence. In E. Dromi (Ed.), *Language and cognition: A developmental perspective.* Norwood, NJ: Ablex.
Newport, E. L. (1977). Motherese: The speech of mothers to young children. In N. Castellan, D. Pisoni, & G. Potts (Eds.), *Cognitive theory* (Vol. II). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Newport, E. L., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and noneffects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 109–150). Cambridge: Cambridge University Press.
Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.
Pinker, S. (1987). Resolving a learnability paradox in the acquisition of the verb lexicon (Lexicon Project Working Papers 17, pp. 1–100). Cambridge, MA: MIT Center for Cognitive Science.
Pinker, S. (1989). Review of D. I. Slobin, *The crosslinguistic study of language acquisition. Journal of Child Language, 16,* 429–475.
Pinker, S., Lebeaux, D., & Frost, A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition, 26,* 185–267.
Rescorla, L. (1980). Overextension in early language development. *Journal of Child Language, 7,* 321–336.
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.
Schreiber, P. A. (1987). Prosody and structure in children's syntactic processing. In H. Horowitz & S. Samuels (Eds.), *Comprehending oral and written language* (pp. 243–270). New York: Academic.
Shipley, E., Smith, C., & Gleitman, L. R. (1969). A study in the acquisition of language: Free responses to commands. *Language, 45,* 322–342.
Slobin, D. I., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition, 12,* 229–265.
Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 25,* 626–636.
Spelke, E. S. (1982). Perceptual knowledge of objects in infancy. In J. Mehler, E. Walker, &

M. Garrett (Eds.), *Perspectives on mental representations.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Spelke, E. S. (1985). Perception of unity, persistence, and identity: Thoughts on infants' conceptions of objects. In J. Mehler & R. Fox (Eds.), *Neonate cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of 1-1 correspondences by human infants. *Science, 210,* 1033–1035.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustic Society of America, 64,* 1582–1592.

Talmy, L. (1975). Semantics and syntax of motion. In J. Kimball (Ed.), *Syntax and semantics 4* (pp. 181–238). New York: Academic.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and language description.* Cambridge, England: Cambridge University Press.

Vendler, Z. (1972). *Res cogitans.* Ithaca, NY: Cornell University Press.

Wasow, T. (1985). Postscript. In P. Sells (Ed.), *Lectures on contemporary syntactic theories* (pp. 193–206). Stanford: Center for the Study of Language and Information.

Waxman, S., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development, 1,* 139–156.

Wexler, K. (1970). *Embedding structures for semantics.* Unpublished manuscript, University of California, Irvine.

Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition.* Cambridge, MA: MIT Press.

Wexler, K., & Manzini, R. (1987). Parameters and learnability in binding theory. In T. Roeper & E. Williams (Eds.), *Parameter setting* (pp. 41–76). Dordrecht: Reidel.

Wilkins, W. (Ed.). (1988). *Syntax and semantics. Vol. 21: Thematic relations.* San Diego, CA: Academic.

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