

Words, Words, Words...

Lila R. Gleitman

Phil. Trans. R. Soc. Lond. B 1994 **346**, 71-77
doi: 10.1098/rstb.1994.0130

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: <http://rstb.royalsocietypublishing.org/subscriptions>

Words, words, words . . .

LILA R. GLEITMAN

Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104-6196, U.S.A.

SUMMARY

Traditional accounts of vocabulary acquisition assume that children succeed by aligning the utterance of words with their environmental contingencies, a word-to-world pairing. Experimental results suggest that such a procedure accounts for the acquisition of nouns but is insufficient for the acquisition of verbs. It is demonstrated that infants under two years of age systematically recruit the structural properties of sentences in which novel verbs occur to find their meanings: a sentence-to-world pairing procedure.

1. INTRODUCTION

Children face a number of problems in acquiring the vocabulary of their native tongue: they have to achieve knowledge of the concepts that words express; they have to parse their caretakers' speech to extract recurrent word-sized units; and they have to line up each concept with one of these words. The last of these topics is called *the mapping problem* for word learning, and it is the one taken up in this chapter. Assuming the (currently unknown) theory that allows infants at some stage of development to internally represent the concept 'elephant', how do they learn that in English this idea is expressed by the sound pattern /elephant/?

Traditional approaches assume that the mapping problem is solved by aligning the occurrence of single words with their contingencies in the world. Thus, the learners' task would be to discover that elephant is most often said in the presence of elephants and more rarely said in their absence (Locke 1690; and many modern sources).

Our recent investigations (J. Gillette & L. R. Gleitman, unpublished results) attempt to model the logic of this word-to-world pairing procedure, and show that it operates quite straightforwardly for concrete nouns. Adult subjects are shown videotapes of mothers playing with their infants (aged about 18 months) but with the audio turned off. These film clips are long enough for subjects to pick up the pragmatics of the conversation. We use mother-child interaction to generate the learning situation just because its topical and pragmatic contents may represent the simplest case, and the one in which real babies succeed.

The subjects are told that whenever the mother is uttering the target noun, a beep will sound. Their task is to identify the word. That is, they must search the visual scene to find the concept that maps onto the pronunciation /beep/. For each of the 20 most frequent nouns, each S views six of the conversational segments containing it. Ss are very good at this task.

Overall, each subject is correct for 45–80% of the targets.

This laboratory situation is radically reduced from the problem that infants confront in assigning interpretations to novel words. Ss are made aware in advance that the target is a noun. In light of the task as set for them, they also know that there exists in English a single common word that will fit their observations. Moreover, they are solving for one target noun at a time, so they have no memory problem. In contrast, a child learner might reencounter the new item only after the passage of considerable time, and mingled with other new words. Thus, these experiments model the child's vocabulary learning situation in at best a highly idealized fashion. But they do add to a literature demonstrating that maternal usage is sufficiently faithful to the 'here and now' to support learning by inspecting how the sounds of words match up with present scenes (Bruner 1974/75; Slobin 1975; Ninio 1980). Moreover, the principal question raised by Quine (1960), namely that there are myriad interpretations of any single scene, seems to pose little practical obstacle to human observers. We seem to share enough conversational-pragmatic and perceptual biases that there is significant consensus as to the 'salient' interpretation.

The apparent simplicity of mapping – once purged of the 'concept learning' issues – accounts for why it has been largely ignored in recent linguistic inquiry into language learning. The task has seemed devoid of any interesting internal structure, merely a matter of associating single words (*qua* phonological objects) with their standard contexts of use, as in the experiment just described. The burden of the present discussion is to show that, over the vocabulary at large, this word-to-world pairing procedure is too weak. Our claim is that word learning is in general performed by pairing a sentence (*qua* syntactic object) with the observed world. In this central sense, vocabulary acquisition is part of the same process that operates in the child's discovery of sentence structure.

2. INSUFFICIENCY OF OBSERVATION FOR VERB LEARNING

As a first demonstration of the structural requirement in word learning, we now reconsider the Gillette & Gleitman experiment as it pertains to the acquisition of verbs. The manipulation is the same. Ss are shown several silent videotapes of mother–infant interactions. At the moment that the mother was uttering some target verb, the beep sounds. Ss are correct only about 15% of the time, startlingly worse than the 45–80% results for nouns.

The difference between the noun and verb results is understated by pointing to the measure of percent correct. Every noun in the test set is correctly guessed by at least some subjects, and errors are almost invariably close to the mark. Thus, the false guesses to elephant when the scene shows a mother and child playing with an elephant puppet are trunk and puppet, i.e. close enough that one can envisage how further learning opportunities would solve residual problems in identification. In contrast, fully a third of the verbs are never guessed correctly by any S (this is so even in experimental variants in which 80 Ss provided a conjecture on each of six trials for the six [of 18] verbs never guessed correctly, i.e. out of 2880 chances, no one was ever correct).

Even worse, errors for verbs are nowhere near the semantic target. Consider 12 Ss' responses for the target come (one of the best cases for observational learning as it is correctly identified about a quarter of the time): come (3), look (2), fly, move, stop, do, put, take, watch (Rokhsar 1994). This scatter is characteristic of verb responses in all our experiments. The potential power of cross-situational observation (Pinker 1984) is that the learner can draw 'something recurrent' out of all the learning opportunities, preserving only what pertains to each of them. And indeed as we saw for nouns, even the incorrect guessers seemed to cull something 'elephant-related' from each observation (i.e. trunk, animal-puppet). But there is little related to coming in the false verb guesses (other than 'some action') that can be preserved to narrow the interpretive options on subsequent exposures.

Again, the experiment has various unrealistic elements, notably that the Ss know all these verbs in advance, and are told that they are to seek a verb. But this should make their task easier, not harder. One could also object, alluding to the results from a generation of developmental psycholinguists, that children are better language learners (including vocabulary learners) than adults and would therefore do better in this task; but this does not account for the massive difference for verbs and nouns in their tractability to this procedure.

If young children are like our adult Ss, they too should have more trouble learning verbs than nouns. And indeed they do. A robust generalization from the vocabulary learning literature is that early vocabularies (the first 50 words) contain few – often no – verbs; and nouns continue to outnumber verbs in productive vocabularies beyond their frequency

distribution in maternal speech until the child is past three years of age (Gentner 1978; Rescorla 1980).

How can we explain the special difficulty of verbs? It cannot be only that scenes viewed support many conjectures, for the same is true of nouns. One factor is that some of the verbs that mothers characteristically utter to their babies represent concepts that are not straightforwardly observable: perhaps surprisingly, such items as want and think are among the most frequent. This contrasts with the frequent maternal nouns, which are overwhelmingly often concrete and visible (a rare exception is the noun kiss, which is correspondingly hard to guess from observing the scene). Another factor is the temporal precision with which the environment matches the utterance. When elephant is uttered, almost always the elephant is being held, waved, even pointed at. Moreover, the caretaker will often label a new object while pointing, using such deictic locutions as 'This is an elephant', or 'See this elephant?'. But push is usually uttered well before or after the pushing event takes place (Lederer *et al.* 1991; Tomasello & Kruger 1992). And such verb-deictic expressions as 'See?', 'This is hopping' are virtually non-existent in the maternal speech that we have observed.

But there is a third factor, one that we believe holds the primary key to the lateness of verb learning: verb acquisition requires access to the phrase structure of the exposure language, and it takes the infant some time to get the relevant structural properties under control.

Eric Lenneberg (1967) provided indirect evidence suggesting a structure-sensitive model for verb learning (and perhaps all classes of words which do not typically express concrete object concepts): the 'explosion' of spoken vocabulary, including sudden increase in the range of lexical types, coincides with the appearance of rudimentary sentences at approximately the 24th month of life. Perhaps an ability to comprehend the spoken sentence is a requirement for efficient verb learning. It may be that once the child has learned some nouns via word-to-world pairing, she can ask not only 'What are the environmental contingencies for the use of a novel word?' but 'What are its environmental contingencies, as constrained by the structural positions in which it appears in adult speech?'

The first direct demonstration that vocabulary acquisition is sensitive to linguistic context was from Roger Brown (1957), who showed that young children would interpret the relation of a novel word to a scene (in this case, a picture) differently, depending on available morphological cues to lexical category: if they heard 'Show me the gorp', they pointed to a visible novel object, but if they heard 'Show me gorping' they pointed to the implied action. The linguistic cues affected the relevant interpretation of the scene in view, to some extent reversing the causal chain suggested by common sense (namely, that the scene in view determines the interpretation of the linguistic object). These findings hint that learners expect there to be a link between formal properties of language and semantic interpretation. Landau &

Gleitman (1985) carried this line further, positing that children use linguistic cues in identifying novel word meanings within as well as across the major lexical classes; particularly, that they inspect syntactic structure to recover the argument structure of novel verbs. Below, we flesh out this perspective.

3. FORM-MEANING INTERACTIONS IN VERB LEARNING

Inspection of natural speech shows that different verbs characteristically occur with different complements, in accord with their differing argument structures (Fillmore 1968; Gruber 1968; Grimshaw 1983; Levin 1985). Thus inalienable (self-caused) actions typically are encoded with intransitive structures (Pinnocchio dances), acts that affect another's state with transitive structures (Gepetto kisses the puppet), transfer acts with ditransitive sentences (Gepetto gives the puppet a book), propositional attitudes with embedded sentential structures (Gepetto thinks that the puppet is alive), and so forth. This form-meaning correlation is usually described by asserting that the structure is a systematic projection from (aspects of) the verb's meaning (it's argument structure; Chomsky 1981). Two complementary approaches to verb learning have recently been developed, both taking advantage of such relations between meaning and structure.

(a) Bootstrapping complementation privileges from knowledge of verb meanings

If form-meaning correlations are systematic across the languages of the world, in principle learners could project the complement structures for a verb whose meaning they have acquired via event observation, rather than having to memorize these independently (Grimshaw 1981; Pinker 1984).

One kind of evidence in support of this hypothesis comes from studies of the invention of language by linguistically deprived youngsters (deaf children of hearing parents who are not exposed to sign language; Feldman *et al.* 1978). The self-invented gesture systems of these children embody many of the same form-meaning linkages that hold for the received languages: the children gesture one noun in construction with their invented gesture for laugh, two in construction with hit, and three with give. This suggests that aspects of the mapping of conceptual onto linguistic structure are not wholly learned from direct exposure, but are prefigured in the child's expectations about language design. Another kind of evidence comes from child errors in complementation, e.g. "Daddy giggled me" and "I poured it full of juice". These occur primarily where there are quirks and subtleties in the way the exposure language maps from argument structure to surface structure (Bowerman 1982). In sum, children use their knowledge of a verb's meaning as a basis for projecting the phrase structure of sentences in which it appears.

(b) Bootstrapping verb meaning from knowledge of complementation

A second learning hypothesis also exploits the fact that verb clause structure is a projection from verb argument structure. Hearing some new verb in a particular structural environment should constrain its interpretation (Landau & Gleitman 1985; Gleitman & Gleitman 1992). Thus, suppose John is observed to hit Bill and to smirk as he does so. There are two plausible interpretations of gorging. But if the adult says John is gorging, the probability that gorp means 'smirk' is increased and the probability that it means 'hit' is decreased. Hearing John gorps Bill should imply the reverse. Thus, knowledge of the semantic implications of the sentence structure in which a novel verb appears can narrow the search-space for its identification. It is this structurally derived narrowing of the hypothesis space for verb meaning on which we now concentrate.

(i) The zoom lens hypothesis

According to our hypothesis, the first use of structural information is as an on-line procedure for interpreting a novel verb. Though there may be quite a few salient interpretations of the scene, the learner 'zooms in' on one (or at least fewer) of these by demanding congruence of the conjecture not only with the visual-perceptual information but also with the semantic implications of the sentence form: much as the movie camera focuses the attention of the viewer by narrowing the visual frame. On this view, the input to verb learning is: (i) the extralinguistic event, as represented by a perceptually and pragmatically sophisticated observer; paired with (ii) the linguistic event, represented as a novel verb positioned within the parse tree constructed from the adult utterance.

The learner exploits the semantically relevant structural information in the latter point to choose among the several interpretations that may be warranted by the first point.

An early demonstration is from Naigles (1990). She investigated responses to novel verbs as a function of linguistic introducing circumstances in children under two years of age, who had no, or few verbs in their spoken vocabularies. In the learning phase of the experiment, the children were shown videotaped action scenes that had two novel salient interpretations. For example, they saw a rabbit pushing down on a duck's head, thus forcing the duck to bend over; simultaneously, both the duck and the rabbit were wheeling their free arm in a broad circle. Whilst watching this scene, half the babies heard 'The rabbit is gorging the duck' while the other half heard 'The rabbit and the duck are gorging'. Then 'gorging' might plausibly refer to forcing-to-bend or to arm-wheeling. Subsequently, the scene disappeared and a voice said 'Find gorging now!, Where's gorging?'. At this point, new action scenes appeared, one on a videoscreen to the child's left, the other on a screen to her right. The one on the left showed the rabbit forcing the duck to bend, but with no arm wheeling.

The one on the right showed rabbit and duck wheeling their arms, but with no forcing to bend. The measure of learning was the child's visual fixation time on one or the other screen during a six second interval. Twenty-three of 24 infants tested looked longest at the videoscreen that matched their syntactic introducing circumstances. Evidently the transitive input biased subjects toward something like the cause-to-bend interpretation while the intransitive input biased them toward arm wheeling. Though we cannot know from this manipulation exactly what the children learned about 'the meaning of gorp', their interpretation of what they were (relevantly) perceiving during the training phase was clearly affected by the syntax, for the subjects' situations differed in no other way.

More direct evidence of the effect of syntactic context on verb identification comes from studies with three-year-old learners. This age group is the one in which the verb vocabulary (and complex sentence structure) burgeons. These subjects are also useful because they can answer questions about the meanings of novel verbs that they encounter. Fisher *et al.* (1994) investigated the acquisition of perspective verbs (e.g. chase/flee, lead/follow) with children of this age.

Principled difficulties for observation-based learning arise for these items, for they come in pairs that vary primarily in the speaker-perspective on a single action or event, and thus their situational concomitants are virtually always the same. This makes them a good testing ground for proposed learning procedures that rely on word-to-world contingencies only. Consider give and get. Both these verbs describe the same intentional transfer of possession of an object between two individuals. Disentangling them based on the pragmatics of the conversation would require the listener to gain access to the mental perspective of the speaker, whether she is likely to be referring to Mary's volitional act of passing the ball to John or John's consequent act of getting the ball from Mary.

Children perhaps can do some mind-reading of just this sort by attending to the gist of conversation. But additional information can come from inspecting the presence and structural positioning of (known) nouns in the sentence heard and comparing these against the scene in view. If that scene shows the ball moving from Mary to John, then an adult utterance like:

- (1) 'Look!, Ziking!',

provides no differentiating information, but if one hears:

- (2) 'Mary zikes the ball to John',

zike likely means give (or throw, hand, etc.). In contrast, if the sentence is:

- (3) 'John zikes the ball from Mary',

then zike likely means get (or take, receive, catch, etc.). The potential clues for disentangling this pair are the choice of nominal in subject position, and the choice of a goal (to) *vs* source (from) preposition.

Fisher *et al.* (1994) showed such scenes/sentences to

three- and four-year-olds in a context where a puppet was uttering the sentence. The children were asked to help the experimenter understand some 'puppet words' (e.g. zike). If the input sentence to the child was uninformative of the give/get distinction (e.g. sentence 1), then child and adult subjects showed a bias in interpreting the scene. They were likely to say that it described something like giving rather than getting. This 'agency bias' (whoever was agent of the action is subject of the transitive verb) characterized the set of five scenarios tested. If the input sentence was (2), which matches the bias as to how to interpret the scene, the tendency to respond with a verb that meant something like give was further enhanced; in fact, almost categorical. But if the input sentence was (3), which mismatches the perceptual/ conceptual bias toward give, subjects' modal response became get (or one of its relatives, e.g. take).

In sum, structural properties of the sentence heard influence the perception of a single scene even in cases where the bias in event representation, taken alone, leads in the opposite direction. Such findings begin to explain why children rarely confuse the perspective verbs despite the fact that they occur in very similar extralinguistic contexts. Their caretakers speak to them in whole sentences, and they exploit the logic of sentence structure as well as salient aspects of the observed world.

Of course, the structure cannot identify a unique verb concept, rather (at best) only a class of such concepts that are united in argument structure. For instance, give, hand, throw share the semantic property of source-to-goal transfer and hence the ditransitive structure and choice of preposition (to). The observed structure evidently focuses attention on this aspect of the scene in view; but then the particular mode or manner of the transfer must be culled directly from observing details of the observed event. It is the surface structure/situation pairing that does the work.

(ii) *The multiple frames hypothesis*

In many cases, a structure/situation pair is insufficient or even misleading about a verb's interpretation. For instance, the phrase structure and the typical situation in adult-child discourse are often the same when the adult says 'Did you eat your cookie?' as when he says 'Do you want a cookie?'. In principle, examination of the further syntactic privileges of eat and want can cue distinctions in their interpretations. For example, want but not eat also occurs with (tenseless) sentence complements ('Do you want to eat the cookie?'), suggesting a mental component of its meaning. More generally, the range of syntactic frames can provide convergent evidence on the meaning of a verb. John is ziking the book to Bill suggests an active verb of transfer (progressive, ditransitive). This would include a broad range of verbs such as bring, throw, explain, etc. But then John is ziking that the book is boring narrows the interpretive range to mental verbs. Taken together – and examined against the accompanying scenes – these joint structural privileges of zike imply mental transfer, whose local interpretation is commu-

nication (e.g. explain, shout; Zwicky 1971; Fisher *et al.* 1991).

There is evidence that the linguistic information provided by mothers to their young children is refined enough to support learning from frame ranges. Lederer *et al.* (1994) examined lengthy conversations of mothers with 18-month-old babies. For the 24 most common verbs in these mothers' child-directed speech, a verb by syntactic-environment matrix was developed. Within and across mothers, each verb was found to be unique in its syntactic range. Using a procedure devised by Fisher *et al.* (1991), it was shown that degree of overlap in syntactic range predicted the verbs' semantic overlap to a striking degree. Partial replications of these findings in Hebrew (Geyer 1991) and in Mandarin Chinese (Li 1994) achieve closely related results.

4. THE POTENCY OF VARIOUS EVIDENTIARY SOURCES

So far we have mentioned some demonstrations with children and adults suggesting that they can use syntactic evidence to aid in the mastery of new verbs. The question remains how much of the burden of verb identification the structure bears; particularly, the multiple-frame evidence. After all, even if syntactic constraints will affect the observer's interpretation in some carefully constructed laboratory situations, in real life the evidence from extralinguistic cues may be so decisive that syntactic deductions rarely come into play.

One suggestive kind of evidence that multiple-frame information is recruited by young children comes from correlational studies (Naigles & Hoff-Ginsberg 1994). The idea behind this work is to inquire how well maternal usage at some point in learning ('time 1') predicts learning, by testing the child's progress after some suitable interval ('time 2'). Specifically, they investigated the use of common verbs in the speech of mothers to one- and two-year-olds, and then the children's subsequent use of these verbs. The diversity of syntactic frames in which verbs appeared in maternal speech at time 1, with verb frequency in maternal speech partialled out, significantly predicted the frequency with which these verbs appeared in child speech ten weeks later.

Lederer *et al.* (1991) have examined the potential information value of various properties of mothers' speech to infants: its (multiple) extralinguistic contexts, nominal co-occurrences, selectional, and syntactic properties. Which of these attributes of adult speech, taken singly or in various combinations, provide enough information for solving the mapping problem for verbs?

The method was to provide (adult) subjects with a large number of instances (usually, about 50) of the use of some target verb by mothers to 18-month-olds, but blocking out one or several potential sources of information. For example, some subjects saw 50 or so videotaped film clips of mothers uttering a single common verb but without audio; the procedure was repeated for the 24 most common verbs in these

mothers' child-directed speech. Other subjects were told the nouns that occurred with the target verb in each of the 50 maternal sentences. A third group was shown the list of 50 sentences that the mother actually uttered but with all nouns as well as the verb converted to nonsense (e.g. Rom GORPS that the rivenflak is grum, can vany GORP the blicket?)

The first finding was that, just as in the Gillette & Gleitman experiment cited earlier, subjects systematically failed to guess the verb from observing its real-world contexts of use (7% correct identification). In the second condition, subjects did not see the video but were told the co-occurring nouns for each sentence in which the mother uttered that verb. After all, if a verb regularly occurs with nouns describing edibles, maybe it means eat. Subjects identified the verb from this kind of information in about 13% of instances.

It is surprising that subjects' mapping performance was so dismal in both the scene and noun-context conditions. Further, when new subjects were given both these kinds of information (that is, shown the videos and told the co-occurring nouns) they still hit upon the target verbs only 28% of the time. But when subjects were provided with frame-range information – no scenes, no real nouns or verbs, just the set of syntactic structures that the mothers used, with their content-bearing words converted to nonsense – the subjects identified 52% of the verbs correctly. It appears that syntactic range information is highly informative.

A difficulty with interpreting these results onto the child-learning situation is that these subjects (when correct) by definition were identifying old verbs that they knew. Perhaps they just looked up the frame-ranges for these known verbs in their mental lexicons rather than using the frames to make semantic deductions. Because of this possibility, the pertinence of the findings to the real learning situation is more easily evaluated by inspecting the 48% of instances where subjects failed in this condition (and the 93% of cases where they failed in the scene condition, etc.). The finding is that false guesses given in response to frame-range information were semantically close to the actual verb that the mother had said (as assessed by the Fisher *et al.* semantic-similarity procedure), e.g. for think, the only false guess was believe. In contrast, the false guesses offered in response to looking at the scenes in which think was actually said were semantically unrelated to this verb (including run, catch, go, look, etc.). The frame-range information puts the subjects into the 'semantic neighborhood' of the target verb while unintended interpretations of scenes don't get the subject close to the speaker's intention.

Note that 52% correct identification in the presence of syntactic frame-range information only, while a significant improvement over 7%, is not good enough if we want to model the fact that verb learning by three-year-olds is a snap. They do not make 48% errors, even errors close to the semantic mark. But as we have stressed, our hypothesis is not that the child ignores the scene or the co-occurring nominals, and

attends to syntax alone (as Lederer *et al.* forced their subjects to do in the experiment just described); rather, the interpretations congruent with the syntax are matched to visual observations. Indeed, adding the nouns to the frames without video led to over 75% verb identification; adding back the scene yielded almost perfect performance.

Summarizing, visual information and noun contextual information taken alone are quite uninformative while structural information is highly informative. In the presence of structural information, the scene and noun-context information significantly improve performance, i.e. the effects of combining information sources are better than additive. The reasons why are easy to see. Consider the noun contexts: It doesn't much help in verb identification to know that one of the words in the utterance was hamburger. But if this word is known to surface as direct object, the meaning of the verb might well be 'eat'. That is, the structural information converts co-occurrence information to selectional information. Similarly for the videotapes. Once the structure of the sentences uttered in their presence is known, the subject can zoom in on fewer interpretations of the events and states that might be pertinent for the mother to have said of them. So if the child has available – as she does, in real life – multiple paired scenes and sentences, we can at last understand why verb learning is easy.

5. HOW THE STRUCTURES OF SENTENCES CAN AID VOCABULARY ACQUISITION

We have suggested that the formal medium of phrase structure constrains the semantic content that the sentence is expressing, thus providing clues to the meaning of its verb. One such clue resides in the number of arguments. A noun-phrase position is assigned to each verb argument; this will differentiate push from fall in a scene that shows both events. Another concerns the positioning of the arguments. The subject of transitives surfaces as the agent, differentiating chase from flee. The case marking and type of the argument also matters, e.g. spatial verbs which allow expression of paths and locations typically accept prepositional phrases, and verbs that express mental acts and states appear with sentence complements.

As mentioned earlier, one cannot converge on the unique construal of a verb from syntactic properties alone. Because the sub-categorization properties are the syntactic expressions of their arguments, it is only those aspects of a verb's meaning that have consequences for its argument structure that could be represented syntactically. Many – most – semantic distinctions are not formally expressed with this machinery. The role of the syntax in this regard is only to narrow the search-space for the meaning, as this latter is revealed by extralinguistic context.

What our experimentation suggests is that this initial narrowing of the hypothesis space by attention to structure is the precondition for using the scene information efficiently to derive the verb's meaning. When babies do not appear to know the phrase

structure, they learn few if any verbs. When adults and young children are required to identify verbs without phrase structure cues (as when told 'Look! Ziking!' or when presented with silent videos of mother-child conversations) again they do not efficiently identify target verbs. But the observation of scenes taken together with observation of structures is sufficient to the task. We conclude that verb learning by infants operates by an implicit but surprisingly sophisticated deductive sentence-to-world pairing procedure.

I am grateful to Professors B. Butterworth, J. Morton, M. J. Snowling and E. K. Warrington for the opportunity to participate in this discussion. Thanks also go to Steven and Marcia Roth for a grant that supported the writing of this article, and to an STC Center Grant from the National Sciences Foundation. This article is based on a prior discussion by Gleitman & Gillette (1994).

REFERENCES

- Bowerman, M. 1982 Evaluating competing linguistic models with language acquisition data: Implications of developmental errors. *Quad. Semant.* **III**, 5–66.
- Brown, R. 1957 Linguistic determinism and parts of speech. *J. Abnorm. Soc. Psychol.* **55**, 1–5.
- Bruner, J.S. 1974/5 From communication to language: a psychological perspective. *Cognition* **3**, 255–287.
- Carey, S. 1978 The child as word learner. In *Linguistic theory and psychological reality* (ed. M. Halle, J. Bresnan & G. Miller), pp. 264–293. Cambridge, Massachusetts: MIT Press.
- Chomsky, N. 1981 *Lectures on government and binding*. Dordrecht: Foris.
- Feldman, H., Goldin-Meadow, S. & Gleitman, L.R. 1978 Beyond Herodotus: The creation of language by linguistically deprived deaf children. In *Action, symbol, and gesture: the emergence of language* (ed. A. Lock), pp. 238–295. NY: Academic Press.
- Fillmore, C. 1968 The case for case. In *Universals of linguistic theory* (ed. E. Bach & R. Harms), pp. 191. New York: Holt, Rinehart and Winston.
- Fisher, C., Gleitman, H. & Gleitman, L. 1991 On the semantic content of sub-categorization frames. *Cogn. Psychol.* **23**, 331–392.
- Fisher, C., Hall, G., Rakowitz, S. & Gleitman, L. 1994 When it is better to receive than to give: Semantic and conceptual constraints on vocabulary growth. *Lingua* **92**, 153–185.
- Gentner, D. 1978 On relational meaning: The acquisition of verb meaning. *Child Devel.* **49**, 988–998.
- Geyer, H. 1991 *Sub-categorization as a predictor of verb meaning: evidence from modern Hebrew*. Unpublished manuscript, University of Pennsylvania.
- Gleitman, L.R. 1990 The structural sources of verb meanings. *Lang. Acquis.* **1**, 3–55.
- Gleitman, L.R. & Gillette, J. 1994 The acquisition of verbs and their arguments. In *The handbook of child language* (ed. P. Fletcher & B. MacWhinney). Oxford: Blackwell. (In the press.)
- Gleitman, L.R. & Gleitman, H. 1992 A picture is worth a thousand words, but that's the problem: the role of syntax in vocabulary acquisition. *Curr. Direct. Psychol. Sci.* **1**, 31–35.
- Grimshaw, J. 1981 Form, function, and the language acquisition device. In *The logical problem of language*

- acquisition (ed. C. L. Baker & J. J. McCarthy), pp. 183–210. Cambridge, Massachusetts: MIT Press.
- Grimshaw, J. 1983 Sub-categorization and grammatical relations. In *Subjects and other subjects* (ed. A. Zaenen), pp. 1–44. Evanston: Indiana University Linguistics Club.
- Gruber, J.S. 1968 Look and see. *Language* **43**, 937–947.
- Landau, B. & Gleitman, L. 1985 *Language and experience: evidence from the blind child*. Cambridge, Massachusetts: Harvard University Press.
- Lederer, A., Gleitman, L.R. & Gleitman, H. 1991 *The informativeness of cross-situational and cross-sentential evidence for learning the meanings of verbs*. Paper presented at the Boston University Conference on Language Development, Boston, October, 1991.
- Lederer, A., Gleitman, L. & Gleitman, H. 1994 Verbs of a feather flock together. In *Beyond names for things: young children's acquisition of verbs* (ed. M. Tomasello & W. Merriman). Hillsdale, New Jersey: Erlbaum. (In the press.)
- Lenneberg, E.H. 1967 *Biological foundations of language*. New York: Wiley.
- Levin, B. 1985 Lexical semantics in review: an introduction. In *Lexical semantics in review. Lexicon Project Working Papers* (ed. B. Levin). Cambridge, Massachusetts: MIT Center for Cognitive Science.
- Li, P. 1994 *Maternal verb usage in Mandarin Chinese*. Unpublished manuscript, University of Pennsylvania.
- Locke, J. 1690 *An essay concerning human understanding* (ed. A. D. Woozley), pp. 1–165. Cleveland: Meridian Books, 1964.
- Naigles, L.G. 1990 Children use syntax to learn verb meanings. *J. Child Lang.* **17**, 357–374.
- Naigles, L.G. & Hoff-Ginsberg, E. 1994 Input to verb learning: verb frame diversity in mothers' speech predicts children's verb use. *Devl Psychol.* (In the press.)
- Ninio, A. 1980 Ostensive definition in vocabulary teaching. *J. Child Lang.* **7**, 565–573.
- Pinker, S. 1984 *Language learnability and language development*. Cambridge, Massachusetts: Harvard University Press.
- Quine, W.V.O. 1960 *Word and object*. Cambridge, Massachusetts: MIT Press.
- Rescorla, L. 1980 Overextension in early language development. *J. Child Lang.* **7**, 321–336.
- Rokhsar, R.M. 1994 *Effective cues in the acquisition of nouns and verbs*. Unpublished manuscript, University of Pennsylvania.
- Slobin, D.I. 1975 On the nature of talk to children. In *Studies in child language development* (ed. E. H. Lenneberg & E. Lenneberg), pp. 58–76. New York: Holt, Rinehart and Winston.
- Tomasello, M. & Kruger, A.C. 1992 Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *J. Child Lang.* **19**, 311–333.
- Zwicky, A. 1971 In a manner of speaking. *Linguist. Inquiry* **11**, 223–233.