macaques, *Journal of Neuroscience, 11,* 168–190 (1991); D. Boussaoud, L. Ungerleider, and R. Desimone, Pathways for motion analysis: Cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque, *Journal of Comparative Neurology, 296* 462–495 (1990).

25. C. Von Der Malsburg, *The Correlation Theory of Brain Function* (Internal Report 81-2) (Max Plank Institute for Biophysical Chemistry, Department of Neurobiology, Goettingen, Germany, 1981); F. Crick and C. Koch, Towards a neurobiological theory of consciousness, *Seminars in the Neurosciences, 2,* 263–275 (1990).

26. C.M. Gray, P. Konig, A.K. Engel, and W. Singer, Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties, *Nature, 338,* 334–337 (1989).

27. R. Baillargeon, Object permanence in 3.5- and 4.5-month-old infants, *Developmental Psychology, 23,* 655–664 (1987); R. Baillargeon, Young infants' reasoning about the physical and spatial properties of a hidden object, *Cognitive Development, 2,* 179–200 (1987).

A Picture Is Worth a Thousand Words, but That's the Problem: The Role of Syntax in Vocabulary Acquisition

Lila R. Gleitman and Henry Gleitman

Children acquire 5 to 10 words a day from about the 15th month through the 6th year of life. How is this astonishing feat accomplished? We discuss here the *mapping problem* for vocabulary acquisition: Granted that infants can entertain some concept, how do they discover which word in the target language expresses it? For example, granted that they can conceive of 'dog' and 'open', how do they learn that /dog/ is the label for 'dog' and /open/ for 'open'?

We focus particularly on the problem of verb acquisition. Verbs appear later than nouns in the infant's speech and are also understood later. This developmental priority of nouns is undoubtedly related to the fact that they typically label objects, while verbs label relationships among object concepts.[1] For example, *hit* expresses a relationship between two entities (the *arguments* of the verb), the hitter and the one hit. To understand *hit,* then, one must understand the type of relationship (the short, sharp contact) and the argument structure.

Lila and Henry Gleitman are Professors of Psychology at the University of Pennsylvania. Address correspondence to L.R. Gleitman, Institute for Research in Cognitive Science, University of Pennsylvania, 3401 Walnut Street, 4th Floor, Philadelphia, PA 19104.

WORD–WORLD PAIRING PROCEDURES

The standard solution to the mapping problem is by appeal to principles of perception and pragmatic inference. Infants are said to pair occurrences of, say, the sound /open/ with certain observed events because /open/ is the verb most likely for the caretaker to say when something is opening. In Locke's words, "People ordinarily show them the thing of which they would have them have the idea; and then repeat to them the name that stands for it."[2] Though in the end something of this sort must be part of the answer to the mapping issue, this explanation leaves some mysteries unsolved, a few of which we sketch below.

Imperfections of the Word–World Contingencies

A first problem in mapping verbs onto their real-world contexts is that caretaker speech is not a faithful running commentary on events in view. In fully a third of verb uses to children under 2 years of age, the act referred to is not taking place. For example, 'open' is often said when nothing present is opening.[3] And even more often, when something is opening, caregivers forebear from saying 'open': When returning home from work, a mother rarely greets her child by saying, "I am opening the door, Joey," but says, "Hello!" instead. Thus, the assumed contingency between opening events and the utterance of 'open' fails to a considerable degree in each direction. Even when the speech act is pertinent to current events, the listener's focus of attention may be elsewhere, as when the mother says, "Come take your nap!" while the child inspects a cat on a mat.

In the cases just discussed, the word–world contingency is exceedingly imperfect. In other cases, it is subtle and invisible: Many verbs understood adequately by 3- and 4-year-olds encode abstract mental acts and states that are not straightforwardly observable, such as *want, hope, and know.*[4]

Abstracting the Relevant World Event

A related problem is that the scene that accompanies utterance of a verb includes many events, only one of which is encoded by that verb. Consider the plight of the child to whom the mother says, "Do you want this ice cream cone?" The mother is speaking, smiling, holding and waving the cone, and perhaps pointing to it; the cone is observably something good to eat, dripping, melting, an object of present desire, and so forth. None of these aspects of the scene is irrelevant to the conversational intent, yet only one of them is correct to map onto the item *want.* A picture is worth a thousand words, but that is the trouble: A thousand words describe the varying aspects of any one picture.

Usually, investigators have solved these problems only by begging the questions they raise. For example, children are said to ignore adult utterances whose meanings are unrelated to the present context. But then, how do children avoid positing a different (and incorrect) pairing of the verb heard with some event that *is* present in the environment? That is, if the learner assumes that speech maps onto current events, and does not know the meaning of some utterance heard, she must attempt a pairing between that utterance and whatever is happening. Woe to such a learner, for in a case we mentioned earlier, she *must* then try to pair up "Come take your nap" with the cat-on-mat observation, a decidedly false step.

As for abstract verbs, the problems they pose for observational learning are often waved aside because their acquisition is relatively late compared with the acquisition of action verbs—as if this lateness answers the question of how they *are* learned, in the end.

Verb Doublets

Other attempts to circumvent these difficulties involve a probabilistic procedure in which the mapping choice is based on the most frequent word–world match, across situations. But this approach faces another problem. Many verbs 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. Consider *give* and *get.* Both these verbs describe the same transfer of possession between two parties. If John gives a book to Mary, Mary necessarily gets the book from John. Movie directors make an art of distinguishing between such notions. The camera can zoom in on the grateful recipient, the giver out of focus or off the screen altogether. Using the word *get* rather than *give* is a linguistic way of making the same distinction. But only for a listener who understands the meanings of the two words. Without a zoom lens, how is the child who knows neither word to guess whether the adult meant 'give' or 'get'? This problem holds for *chase/flee, buy/sell, lead/follow,* and hundreds of other such doublets.

To get around this difficulty, proponents of the word–world approach appeal to biases in event representation. For instance, in an experiment we will describe more fully below, both 3-year-olds and adults presented with a nonsense verb and shown a scene of a giving-getting event almost always guessed that the nonsense verb meant 'give' and not 'get' or 'take'. This interpretive bias is very real, but it will not solve the mapping problem because mothers cannot be relied on to utter 'give' in the presence of such a scene: As it happens, *get* is more frequent in maternal usage than *give.* The same bias in event representation that would help children learn *give* should undermine them in learning *get,* for if they encode an event as 'give' and hear "get," their hypothesis must be that "get" means 'give.' Yet the learning function is not characterized by confusions among such words.

A THEORETICAL ALTERNATIVE: SENTENCE–WORLD PAIRING

We claim that these problems can be solved if learners perform a sentence–world pairing rather than a word–world pairing, taking advantage of the clues to interpretation that reside in the structure of the sentence heard.[5] The nature of the relationship expressed by the verb is spelled out across the clause structure within which the verb occurs. For example, the fact that *hit* is a two-argument verb (one that expresses a relationship between a hitter and a thing hit) predicts that it will occur in transitive structures, such as *John hit Bill,* whereas *laugh,* a one-argument verb, will occur in intransitive structures, such as *John laughs.* More generally, because every argument of the verb requires one noun-phrase position in a syntactic structure, *biff* heard in a sentence like *John biffs* is unlikely to be describing a two-place relation such as 'hit' even if perceptual biases commend such a construal. This option eliminated by attention to the structure, the new verb must be describing some other aspect of the scene in view: Perhaps John is convulsed with mirth while he hits Bill.

To narrow the search-space for the correct mapping, learners must take advantage of these argument-taking properties of verbs, and the way these arguments are expressed in sentences. In this sense, the structure of the sentence that the child hears can function like a mental zoom lens that cues the aspect of the scene the speaker is describing. Consequently, verb learning begins just when the child shows rudimentary sensitivity to syntactic structure and has a working vocabulary of simple nouns.

In principle, a systematic mapping between syntax and semantics could be one of the presuppositions that learners bring into the learning situation, part of the innate processing system that makes it possible to learn language. One assumption implicit in this hypothesized language-learning machinery is that some specific mappings between verb meanings and their syntactic expressions hold universally across languages. Another is that very young children can analyze the structure of a sentence heard and use this information to derive the argument structure of the verb. At first glance, such a procedure seems too abstract to grant to babies. Yet the operation of this kind of mechanism has been documented in a variety of experiments with children as young as 16 months.

One such experiment investigated whether 16- to 18-month-old children who utter only isolated nouns are attending to the structure within which a new verb occurs, and can assign semantic roles to the participant entities based on their structural positions.[6] Two videotaped scenes were shown to the infants, while half of them heard, "Big Bird is tickling Cookie Monster," and the other half heard, "Cookie Monster is tickling Big Bird." The infants looked longest at the videoscreen that matched the syntax. We can conclude that even before any verbs are uttered, infants understand the semantic implication of subject versus direct object (or, at minimum, of serially first versus second noun) in English sentences. As we explain next, if this machinery is in place, it supports deductions about which action is being described by a novel verb.

In several experiments, we asked whether young children would distinguish between two different interpretations of a novel verb based on the number and positions of noun phrases used in construction with that verb.[7] In one such study, a videotaped scene shown to infants aged 22 to 24 months depicted two actions simultaneously: A duck forced a rabbit to squat by pushing on its head, while the duck and rabbit each wheeled its free arm in a circle. While watching this scene, half the subjects heard a voice say, "The duck is biffing the bunny"; the other half heard, "The duck and the bunny are biffing." This video was then removed, and the voice said, "Find biffing now!" Two new videos then appeared. In one, the duck was forcing the rabbit to squat (but there was no arm-wheeling). In the other, the duck and rabbit were side by side wheeling their arms (but the duck was not forcing the rabbit to squat). Fifteen of 16 infants looked longest at the scene that matched the syntax in which *biffing* was introduced: Those exposed to the intransitive sentence watched the arm-wheeling scene, and those exposed to the transitive sentence watched the forcing-to-squat scene. Clearly, the sentence structure was decisive in cuing which aspect of the complex initial scene was relevant to the interpretation of /biff/.

Another experiment directly pitted scene-interpretive biases against syntactic clues.[8] Three- and 4-year-olds were shown videotaped scenes that could be labeled by different verbs of the doublet variety we discussed earlier (e.g., *chase/flee, give/get*). The children were introduced to a puppet who spoke "puppet language" and asked to tell the experimenter what the puppet meant as they watched these scenes. For example, in one scene, a rabbit ran across the screen from left to right; as the rabbit was disappearing, a skunk appeared at the left and ran across. If the puppet said, "Look! Biffing!" about 80% of the subjects responded, "He's chasing him," or "He's running after him." Without syntactic clues, then, the scene was more likely to be interpreted as depicting *chasing* than *fleeing*. This bias was enhanced (held for well over 90% of subjects) when supported by syntactic evidence, that is, if the puppet said, "The skunk is biffing the rabbit." But this conjecture, evidently the most natural visual-cognitive interpretation of the scene, was offered by fewer than 30% of children who heard, "The rabbit is biffing the skunk." The modal response to this latter sentence was "He's running away from him," or "He's trying to get away from him." These results held as strongly when the number rather than the position of noun phrases was the available syntactic clue to interpretation (e.g., *push* vs. *fall, feed* vs. *eat*).

To be sure, under all these presentation conditions, the subjects were guessing the verb meaning by inspecting the situational context. But what they thought was happening was powerfully—albeit implicitly—affected by the linguistic context. Despite biases in event representation, the learner can and does avoid errors by respecting the semantic implications of structural formats. For this simple case, the listener posits that the subject of the sentence is the "doer of the action."

THE RELATIONS BETWEEN SENTENCE FRAME AND VERB SEMANTICS

These findings in hand, we can look a little more closely at a theory that can explain them. Our view is that the several syntactic structures in which particular verbs can appear determine certain aspects of the verbs' meanings, specifically, those aspects related to the argument-taking properties of the verbs. Within the coarse-grained constraints on meaning imposed by these structural facts, the search-space for verb meaning is narrowed to allow observation to do its work.

Consider as examples the verbs *give, go, think,* and *explain*. Overall, these verbs are very different in their meanings, yet they share certain semantic properties. Both *give* and *explain* describe the transfer of some entity between two parties, and for this reason, they can both appear in three-noun-phrase structures, such as *John gives the pencil to Bill* and *John explains the facts to Bill*. In the *give* sentence, an object goes from John's to Bill's hands, and in the *explain* sentence, the facts go from

John's to Bill's mind. A noun phrase is required for the source (John), the goal (Bill), and that which moves between them (the pencil or the facts). Verbs that describe no such transfer sound odd in this sentence frame (e.g., *John goes the pencil to Bill, John thinks the facts to Bill*). This first structural distinction thus partitions the four verbs into a transfer set (*give, explain*) and a nontransfer set (*go, think*).

But there is another semantic–syntactic property that lines up these four words differently: Although *go* and *give* pertain to physical actions, *think* and *explain* pertain to mental states or acts. There is a syntactic correlate of this semantic property also, namely, appearance with tensed sentence complements: *John thinks that Bill is tall* and *John explains (to Mary) that Bill is tall.* Verbs that describe physical acts are anomalous in this construction (*John goes that Bill is tall, John gives (to Mary) that Bill is tall*). This distinction partitions the four verbs into a mental set (*think, explain*) and a physical set (*go, give*).

These cross-classifications severely constrain the overall verb interpretation. Though there are hundreds of mental verbs and hundreds of transfer verbs, the number of verbs that, like *explain,* have both these properties is small, including such additional items as *shout, tell,* and *argue*. And these, like *explain,* describe mental transfer, specifically, communication. Thus, the several clause structures that are fitting to a single verb provide convergent evidence as to its meaning. The specific manner of communication—whether shouting or whispering or explaining—has to be derived by inspecting the scene, to be sure, for such manner distinctions (as opposed to argument-taking distinctions) are not formally encoded in the sentence structures. But the structural factors have narrowed the search-space for the verb meaning enough for observation to work without an endless proliferation of categories coming to mind.

In recent work, we have shown the power and scope of these semantic–syntactic relations in the verb lexicon. We asked two sets of (adult) subjects to partition a set of verbs, one group providing a semantic partitioning, the other a syntactic partitioning.[9] If the structures are predictive of the semantics, then the more two verbs' syntactic ranges overlap, the more the verbs should overlap in their construals. We found that the two partitionings were powerfully correlated. Furthermore, the same correlations were found (and involved the same syntactic properties) when the experiment was replicated in Italian and in Hebrew. For example, relations between an actor and an event are described by verbs that accept sentence complements across as well as within languages. That there be such universal syntactic–semantic mappings is an obvious requirement for a learning machinery that uses these relations as input to verb learning.

In related work, we have shown that the actual usage of mothers to children under age 2—when verb learning begins—provides the syntactic information required by this model.[10] Each of the 24 verbs most frequently used by a sample of mothers occurred in a distinctive range of structures, and the semantic relatedness among these verbs was closely predicted by the degree of overlap in their syntactic ranges.

Finally, given the findings that children as young as 16 months—and adult controls as well—actually make use of the structural evidence to derive verb construals, Lederer, with the present authors, is using adult subjects to get a broader idea of the information potential of both situational and structural supports for verb learning. To model a word–world pairing procedure (in which the learner has no access to structural information but has many exposures to scenes in which some verb is uttered), one group was shown dozens of videotaped scenes (but with no audio) in each of which a mother was uttering some single verb to her child. This procedure was repeated for the 24 most frequent verbs in these mothers' speech. The subjects could not come within calling distance of guessing what the mystery verbs might be (guessing right in only 7% of trials). The set of observed scenes vastly underdetermined the verb construals.

In a second condition, subjects were told all the other content words that occurred in each sentence with each mystery verb (but not their syntactic positioning and without the video). It might seem that this co-occurrence information would help; for example, knowing that such nouns as *candy, ice cream,* and *hamburger* occurred with some verb might suggest that it means 'eat.' But this information improved subjects' performance only to 13% correct, and (as in the first condition) their false guesses were wildly unrelated to the target verbs (e.g., for the target *want,* guesses included *go, catch,* and *eat*). This pathetic performance hardly mirrors the learning characteristics of real children.

In a third condition, subjects were shown the dozens of structures in which each mystery verb occurred in the mothers' speech, again without the videotaped context, and with all content nouns converted to nonsense. This procedure tests for the informativeness of syntactic ranges. These subjects correctly guessed the target verbs over 50% of the time. Moreover, their false guesses fell into the right semantic neighborhoods (e.g., for the target *want,* false guesses were other mental verbs, such as *expect* and *hope*). Thus, exposure to several structures is more informative than exposure to several scenes.

A final group of subjects received the structural information, as just described, but the nouns were the real ones that the mothers had used. These subjects guessed the target verbs correctly over 80% of the time, though they saw no scenes. This result reveals that the nouns co-occurring with a verb can add significant information about its meaning if the syntactic positioning (and, hence, the semantic role) of these nouns is known. Knowing that *ice cream* and *hamburger* occurred "somewhere" in the mother's utterance (as in the second condition) is not too informative: After all, the sentence might be "Ice cream *ruins* your appetite" or "The hamburger *fell* on the floor." But knowing that these "edibles" occurred as the direct object of the mystery verb is a good clue that it might mean 'eat.'

We conclude that structural information is a requirement for efficient verb learning. The frame ranges provide strong cues to interpretation. In the presence of this structural information, the complement selection (the syntactically positioned nouns) provides significant further clues. In contrast, the set of scenarios taken alone—or even taken in combination with (asyntactic) knowledge of all co-occurring words—leaves too much latitude to allow verb identification.

All in all, our work suggests that verbs' meanings cannot be extracted by a procedure that pairs single words to their observational contingencies. This is because verbs do not as a rule directly encode actions and events. If they did, grunting and pointing could substitute for elaborate human language systems. Instead, verbs encode acts and states of the world and of the mind under particular (and invisible) stances toward these adopted by the speaker. A further data source is therefore required to rein in the hundreds of salient interpretive choices made available by perception and pragmatic inference as to the speaker's intent. It is the infant's natural appreciation of syntactic structure and its mapping onto conceptual structure that provides this additional data source.

Acknowledgments—This work was supported by a grant to L.R. Gleitman from Steven and Marcia Roth, whose aid we acknowledge with gratitude.

Notes

1. D. Gentner, Why nouns are learned before verbs, in *Language Development: Vol. 2. Language, Thought and Culture*, S. Kuczak, Ed. (Erlbaum, Hillsdale, NJ, 1981).
2. J. Locke, *An Essay Concerning Human Understanding* (Meridian Books, Cleveland, 1964; original work published in 1690), Book 3.IX.9.
3. R. Beckwith, E. Tinker, and L. Bloom, *The acquisition of nonbasic sentences*, paper presented at the Boston Child Language Conference, Boston (1989, October).
4. One particularly startling example of children's competence in this regard is that *see* was the first verb uttered by a congenitally blind 2-year-old; B. Landau and L. Gleitman, *Language and Experience: Evidence From the Blind Child* (Harvard University Press, Cambridge, MA, 1985).
5. For a fuller discussion, see L. Gleitman, Structural sources of verb learning, *Language Acquisition, 1*, 3–55 (1990).
6. K. Hirsh-Pasek and R. Golinkoff, Language comprehension: A new look at old themes, in *Biological and Behavioral Aspects of Language Acquisition*, N. Krasnegor, D. Rumbaugh, M. Studdert-Kennedy, and R. Schiefelbusch, Eds. (Erlbaum, Hillsdale, NJ, in press).
7. L. Naigles, Children use syntax to learn verb meanings, *Journal of Child Language, 17*, 357–374 (1990). See also L. Naigles, H. Gleitman, and L. Gleitman, Children acquire word meaning components from syntactic evidence, in *Linguistic and Conceptual Development*, E. Dromi, Ed. (Ablex, New York, in press).
8. C. Fisher, G. Hall, S. Rakowitz, and L. Gleitman, *When it is better to receive than to give: Syntactic and conceptual supports for verb learning*, unpublished manuscript, University of Pennsylvania, Philadelphia (1991).
9. C. Fisher, H. Gleitman, and L. Gleitman, On the semantic content of subcategorization frames, *Cognitive Psychology, 23*, 331–392 (1991); H. Geyer, L. Gleitman, and H. Gleitman, *Semantic/syntactic linkages in the Hebrew verb lexicon*, unpublished manuscript, University of Pennsylvania, Philadelphia (1991).
10. A. Lederer, L. Gleitman, and H. Gleitman, Input to a deductive verb-acquisition procedure, *Lingua* (to appear).

Neural Foundations of Visual Motion Perception

J. Anthony Movshon and William T. Newsome

The detection and analysis of motion is one of the fundamental tasks of vision, because practically everything of interest in the visual world moves. Although motion analysis of a high order is evident in such simple visual systems as the fly's, it is only in primates that a well-defined anatomical division of the central visual pathways can be seen to be specialized for the analysis of motion.[1] The best defined area in this pathway is an extrastriate area known as MT (or V5). Unlike other areas of the monkey's extrastriate visual cortex, almost all neurons in MT are direction selective, meaning that they typically respond best to motion within a given range of directions, and respond not at all (or with inhibition) to motion in the opposite direction. The activity of MT neurons has been linked to a variety of motion-related tasks, including the analysis of the motion of complex patterns, the detection of target motion relative to the background, and the generation of signals for smooth pursuit eye movement.[2]

J. Anthony Movshon is Professor in the Center for Neural Science and the Department of Psychology at New York University and an Investigator of the Howard Hughes Medical Institute. **William T. Newsome** is Associate Professor in the Department of Neurobiology at Stanford University School of Medicine. Address correspondence to J. Anthony Movshon, Center for Neural Science, New York University, New York, NY 10003.