How could a child use verb syntax to learn verb semantics?*

Steven Pinker

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, E10-016, Cambridge, MA 02139, USA

I examine Gleitman's (1990) arguments that children rely on a verb's syntactic subcategorization frames to learn its meaning (e.g., they learn that *see* means 'perceive visually' because it can appear with a direct object, a clausal complement, or a directional phrase). First, Gleitman argues that the verbs cannot be learned by observing the situations in which they are used, because many verbs refer to overlapping situations, and because parents do not invariably use a verb when its perceptual correlates are present. I suggest that these arguments speak only against a narrow associationist view in which the child is sensitive to the temporal contiguity of sensory features and spoken verb. If the child can hypothesize structured semantic representations corresponding to what parents are likely to be referring to, and can refine such representations across multiple situations, the objections are blunted; indeed, Gleitman's theory requires such a learning process despite her objections to it. Second, Gleitman suggests that there is enough information in a verb's subcategorization frames to predict its meaning 'quite closely'. Evaluating this argument requires distinguishing a verb's *root* plus its semantic *content* (what *She boiled the water* shares with *The water boiled* and does not share with *She broke the glass*), and a verb *frame* plus its semantic *perspective* (what *She boiled the water* shares with *She broke the glass* and does not share with *The water boiled*). I show that hearing a verb in a single frame only gives a learner coarse information about its semantic perspective in that frame (e.g., number of arguments, type of arguments); it tells the learner nothing about the verb root's content across frames (e.g., hot bubbling liquid). Moreover, hearing a verb across all its frames also reveals little about the verb root's content. Finally, I show that Gleitman's empirical arguments all involve experiments where children are exposed to a single verb frame, and therefore all involve learning the frame's perspective meaning, not the root's content meaning, which in all the experiments was acquired by observing the accompanying scene. I conclude that attention to a verb's syntactic frame can help narrow down the child's interpretation of the perspective meaning of the verb in that frame, but disagree with the claim that there is some in-principle limitation in learning a verb's content

from its situations of use that could only be resolved by using the verb's set of subcategorization frames.

1. Introduction: The problem of learning words' meanings

When children learn what a word means, clearly they must take note of the circumstances in which other speakers use the word. That is, children must learn *rabbit* because their parents use *rabbit* in circumstances in which the child can infer that they are referring to rabbits. Equally obviously, learning word meanings from circumstances is not a simple problem. As Quine (1960), among others, has noted, there are an infinite set of meanings compatible with any situation, so the child has an infinite number of perceptually indistinguishable hypotheses about meaning to choose among. For example, all situations in which a rabbit is present are also situations in which an animal is present, an object is present, a furry thing is present, a set of undetached rabbit parts are present, a something-that-is-either-a-rabbit-or-a-Buick is present, and so on. So how does the child figure out that *rabbit* means 'rabbit', not 'undetached rabbit part'?

Word learning is a good example of an induction problem, where a finite set of data is consistent with an infinite number of hypotheses, only one of them correct, and a learner or perceiver must guess which it is. The usual explanation for how people do so well at the induction problems they face is that their hypotheses are inherently constrained: not all *logically* possible hypotheses are *psychologically* possible. For example, Chomsky (1965) noted that children must solve an induction problem in learning a language: there are an infinite number of grammars compatible with any finite set of parental sentences. They succeed, he suggested, because their language acquisition circuitry constrains them to hypothesize only certain kinds of grammatical rules and structures, those actually found in human languages, and because the kinds of sentences children hear are sufficient to discriminate among this small set of possibilities.

In the case of learning word meanings, too, not all logically possible construals of a situation can be psychologically possible candidates for the meaning of a word. Instead, the hypotheses that a child's word learning mechanisms make available are constrained in two ways. The first constraint comes from the representational machinery available to build the semantic structures that constitute mental representations of a word's meaning: a Universal Lexical Semantics, analogous to Chomsky's Universal Grammar

(see, e.g., Moravscik 1981, Markman 1989, 1990; Jackendoff 1990). For example, this representational system would allow 'object with shape X' and 'object with function X' as possible word meanings, but not 'all the undetached parts of an object with shape X', 'object with shape X or a Buick', and 'object and the surfaces it contacts'. The second constraint comes from the way in which a child's entire lexicon may be built up; on how one word's meaning may be related to another word's meaning (see Miller 1991, Miller and Fellbaum 1992). For example, the lexicons of the world's languages freely allow meronyms (words whose meanings stand in a part–whole relationship, like *body–arm*) and hyponyms (words that stand in a subset–superset relationship, like *animal–mammal*), but do not easily admit true synonyms (Bolinger 1977, Clark 1987, Miller and Fellbaum 1991). A child would therefore not posit a particular meaning for a new word if it was identical to some existing word's meaning. Finally, the child would have to be equipped with a procedure for testing the possible hypotheses about word meaning against the situations in which adults use the words. For example, if a child thought that *pet* meant 'dog', he or she will be disabused of the error the first time the word is used to refer to a fish.

Although the problem of learning word meanings is usually discussed with regard to learning nouns, identical problems arise with verbs (Landau and Gleitman 1985, Pinker 1988, 1989; Gleitman 1990). When a parent comments on a dog chasing a cat by using the word *chase*, how is the child to know that it means 'chase' as opposed to 'flee', 'move', 'go', 'run', 'be a dog chasing', 'chase on a warm day', and so on?

As in the case of learning noun meanings (indeed, learning in general), there must be constraints on the child's possible hypotheses. For example, manner-of-motion should be considered a possible component of a verb's mental dictionary entry, but temperature-during-motion should not be. (See Talmy 1985, 1988; Pinker 1989, Jackendoff 1990, and Dowty 1991, for inventories of the semantic elements and their configurations that may constitute a verb's semantic representation.) Moreover, there appear to be constraints on lexical organization (Miller 1991, Miller and Fellbaum 1991). For example, verb lexicons often admit of co-troponyms (words that describe different manners of performing a similar act or motion, such as *walk-skip-jog*) but, like noun lexicons, rarely admit of exact synonyms (Bolinger 1977, Clark 1987, Pinker 1989, Miller and Fellbaum 1991). Finally, the child must be equipped with a learning mechanism that constructs, tests, and modifies semantic representations by comparing information about the uses of verbs by other speakers across speech events (Pinker 1989).

1.1. A novel solution to the word-learning problem

In recent years Lila Gleitman and her collaborators have presented a series of thorough and insightful discussions of the inherent problems of learning verbs' meanings (Landau and Gleitman 1985, Hirsh-Pasek et al. 1988, Gleitman 1990, Naigles 1990, Lederer et al. 1989, Fisher et al. 1991 and this volume). Interestingly, Gleitman and her collaborators depart from the usual solution to induction problems, namely, seeking constraints on the learner's hypotheses and their relation to the learner's input data as the primary explanation. Rather, they argue that the learner succeeds at learning verb semantics by using a channel of information that is not directly semantic at all. Specifically, they suggest that the child infers a verb's meaning by using the kinds of syntactic arguments (direct object, clause, prepositional phrase) that appear with the verb when it is used in a sentence. Such syntactic properties (e.g., whether a verb is transitive or intransitive) are referred to in various literatures as the verb's 'argument structure', 'argument frame', 'syntactic format', and 'subcategorization frame'. Indeed, Gleitman and her collaborators argue that information about a verb's semantics, gleaned from observing the circumstances in which other speakers use the verb (e.g., learning that *open* means 'opening' because parents use the verb to refer to opening things) is in principle inadequate to support the acquisition of the verb's semantics; cues from the syntactic properties of the verb phrase are essential.

This position has its roots in Brown (1957) and Katz et al. (1974), who showed empirically how children use grammatical information to help learn certain aspects of word meanings. But it was given a stronger form in Landau and Gleitman's (1985) book *Language and Experience: Evidence from the Blind Child*. Landau and Gleitman point out that a blind child they studied acquired verbs, even perceptual verbs like *look* and *see*, rapidly and with few errors, despite the child's severe impairment in being able to witness details of the scenes in which the verbs are used. Moreover, they noted that a sighted child's task in learning verbs is different from the blind child's task only in degree, not in kind. Since the learning of verbs like *see* and *know* cannot critically rely on information from vision, Landau and Gleitman presented the following hypothesis:

'In essence our position will be that the *set* of syntactic formats for a verb provides crucial cues to the verb meanings just because these formats are abstract surface reflexes of the meanings. ... there is very little information in any single syntactic format that is attested for some verb, for that format serves many distinct uses. However ... the *set* of subcategorization frames

associated with a verb is highly informative about the meaning it conveys. In fact, since the surface forms are the carriers of critical semantic information, the construal of verbs is partly indeterminant without the subcategorization information. Hence, in the end, a successful learning procedure for verb meaning must recruit information from inspection of the many grammatical formats in which each verb participates.' (1985: 138–139)

For example, here's how a child hearing the verb *glip* in a variety of syntactic frames could infer various components of its meaning from the characteristic semantic correlates of those frames. Hearing *I glipped the book* (transitive frame, with a direct object), a child could guess that *glipping* is something that can be done to a physical object. Hearing *I glipped that the book is on the table* (frame with a sentential complement), the child could infer that *glipping* involves some relation to a full proposition. Hearing *I glipped the book from across the room* (frame with an object and a directional complement) tells him or her that *glipping* can involve a direction. Moreover, the absence of *Glip that the book is on the table!* (imperative construction) suggests that *glipping* is involuntary, and the absence of *What John did was glip the book* (pseudo-cleft construction) suggests that it is not an action. With this information, the child could figure out that *glip* means 'see', because seeing is an involuntary non-action that can be done to an object or a proposition from a direction. Note that the child could make this inference without seeing a thing, and without seeing anyone seeing anything. In her 1990 paper laying out this hypothesis in detail and discussing the motivation for it, Gleitman calls this learning procedure 'syntactic bootstrapping', and offers it as a major mechanism responsible for the child's success at learning verb meanings.

The goal of the present paper is to examine the general question of how a child could use the syntactic properties of a verb to figure out its semantic properties. I will discuss several kinds of mechanisms that infer semantics from syntax, attempting to distinguish what kinds of inputs they take, how they work, what they can learn, and what kind of evidence would tell us that children use them. I will focus on Gleitman's (1990) thorough and forceful arguments for the importance of syntax-guided verb learning. After she puts these arguments in particularly strong form in order to make the best case for them and to find the limits as to what they can accomplish, Gleitman settles on an eclectic view in which a set of learning mechanisms, some driven by syntax and some not, complement each other. I agree with this eclectic view and will try to lay out the underlying division of labor among learning mechanisms more precisely. In doing so, I will, however, be disagreeing with some of the particular strong claims that Gleitman makes about syntax-guided learning of meaning in the main part of her paper.

2. What is learned from what: Two preliminary clarifications

Sentences contain a great deal of information, and the child is learning many things at once from them. To understand how syntax can help in learning semantics, it is essential to be clear on what kinds of information in a sentence are and are not 'syntactic', and what kinds of things that a child is learning are and are not 'semantic'. Before examining Gleitman's arguments, then, I make some essential distinctions, without which the issues are very difficult to study.

2.1. Linguistically-conveyed semantic content is not the same as syntactic form

Gleitman's hypotheses literally refer to the acquisition of verb meanings via the use of *syntactic* information, specifically, the syntactic properties of the arguments that the verb appears with (e.g., whether it takes a grammatical object, a prepositional object, a sentential complement, or various combinations of these arguments in different sentences). Note that this is not the same as claiming that the child uses *semantic* information that happens to be communicated by the linguistic channel.

Sentences, obviously, are used to convey real-world information, and children surely can infer much about what a verb means from the meanings of the other words in the sentence and from however much of the sentence's structure they are able to parse. For example, if someone were to hear *I glipped the paper to shreds* or *I filped the delicious sandwich and now I'm full*, presumably he or she could figure out that *glip* means something like 'tear' and *filp* means something like 'eat'. But although these inferences are highly specific and accurate, no thanks are due to the verbs' syntactic frames (in this case, transitive). Rather, we know what those verbs mean because of the semantics of *paper, shreds, sandwich, delicious, full*, and the partial syntactic analysis that links them together (partial, because it can proceed in the absence of knowledge of the specific subcategorization requirements of the verb, which is the data source appealed to by Gleitman). In other words, inferring that *tear* means 'tear' from hearing *paper* and *shreds* is a kind of cognitive inference using knowledge of real-world contingencies, the same one that could be used to infer that *tear* means 'tear' when seeing paper being torn to shreds. It is not an example of learning a verb's meaning from its syntactic properties, the process Gleitman is concerned with. For this reason, a blind (or sighted) child can learn a great deal about a verb's meaning from the sentences the verb is used in, without learning anything about the meaning from the verb's syntax in those sentences.

Moreover, some of the information about how a verb is used in a sentence is based on universal features of semantics. For example, the sentence *I am glipping apples* could inform a learner that *glip* can't mean 'like', because the progressive aspect marked on the verb is semantically incompatible with the stativity of liking. Here, too, one can learn something about a verb's meaning from the sentence in which the verb is used, as opposed to the situation in which the verb is used, but the learning is driven by semantic information (in this example, that liking does not inherently involve changes over time), not syntactic information.

Gleitman (1990) does not contest this distinction; in footnote 8 on p. 27 and in footnote 26 (p. 379) of Fisher et al. (1991), she states that her arguments are not about the use of linguistically-conveyed information in general, but about the use of the syntactic properties of verbs per se. Nonetheless, the distinction has implications that bear on her arguments in ways she does not make explicit.

First, the distinction blunts the intuitive impact of two of Gleitman's recurring arguments for the importance of syntactic information: that blind children learn verbs' meanings without seeing their referent events, and that parents do not invariably use verbs in unique situations (e.g., they do not say *open* simultaneously with opening something). These phenomena suggest that children must attend to what parents say, not just what they do. The phenomena do not, however, lead by some process of elimination to the hypothesis that children are using the syntactic subcategorization properties of individual verbs. The children may just be figuring out the content of the sentences, and inferring a verb's semantics from its role in the events conveyed.

Second, many of the supposedly syntactically-cued inferences that Gleitman appeals to may actually be semantically cued in the same sense that hearing a verb used with *sandwich* suggests that it involves eating. The 'subcategorization frames' that Landau and Gleitman (1985), Gleitman (1990), and Fisher et al. (1991) appeal to are distinguished more by the semantic content of particular words in them than by their purely syntactic (i.e., categorical) properties. Indeed, most of the entries are not syntactically distinct subcategorization frames in the linguist's sense at all. Of the 33 entries listed in Appendix A of Fisher et al. (1991), two thirds are actually not syntactically distinct subcategorization frames. Seventeen frames are syntactically identical V-PP frames differing only in the choice of preposition (e.g., *in* NP versus *on* NP). (Fisher et al. did, to be sure, collapse these prepositions into a single frame type in the data analysis of their study.) Three are V-S'

frames differing only in the choice of complementizers (e.g., *that* S versus *if* S). There are V-NP-PP frames differing only in the choice of preposition (e.g., NP *to* NP versus NP *from* NP; these were, however, collapsed in the analysis). And three are not subcategorization frames at all but the morpho-syntactic constructions imperative, progressive, and pseudo-cleft, which are syntactically well-formed with any verb (though some are awkward because of semantic clashes, such as involuntary verbs in the imperative). The problem is that even if learners can use verbs' patterning across these linguistic contexts, it is misleading to say that they would be relying on syntactic information. In most modern theories of verbs' compatibility with prepositions and complementizers (see Jackendoff 1987, 1990; Pinker 1989, Grimshaw 1979, 1981, 1990), the selection is made on semantic grounds: for example, verbs involving motion in a direction can select any preposition that involves a direction. There are verb-specific idiosyncrasies, to be sure (such as *rely on* and *put up with*), but even these may be treated as involving idiosyncratic *semantic* properties of the verb. Thus if a child notices that a verb takes *across* and *over* but not *with* or *about*, and infers that the verb involves motion, the child is not using syntactic information, but figuring out that an event involving the traversal of paths (inherent to the meaning of *across* and *over*) is likely to involve motion, just as an event that involves sandwiches and hunger is likely to involve eating.[1]

2.2. The term 'syntactic bootstrapping' and the opposition of 'syntactic' and 'semantic' bootstrapping are misleading

It is unfortunate that Gleitman chose the term 'syntactic bootstrapping' to refer to the process of inferring a verb's meaning from its set of subcategoriza-

---

[1] Note that some of the other linguistic contexts that Landau and Gleitman call 'subcategorization frames' are not subcategorization frames either, but frozen expressions and collocations that are probably idiosyncratic to English and hence no basis for learning. These include *Look!*, *See?*, *Look! The doggie is running!*, *See? The doggie is running!*, *Come see the doggie*, and *look like* in the sense of 'resemble'. Since *look* and *see* are the only two verbs that Landau, Gleitman, and their collaborators discuss in detail, if their learning scenarios for these two verbs adventitiously exploit particular properties of English, one has to be suspicious about the feasibility of the scenario in the general case. More generally, Fisher, Gleitman, and Gleitman's claim that there are something like 100 distinct syntactic subcategorization frames, hence, in principle, 2^{100} syntactically distinguishable verbs, appears to be a severe overestimate. I think most linguists would estimate the number of syntactically distinct frames as an order of magnitude lower, which would make the estimated number of syntactically distinguishable verbs a tiny fraction of what Fisher et al. estimate.

tion frames. She intended the term to suggest an opposition to my 'semantic bootstrapping' (Pinker 1982, 1984, 1987, 1989), and one of the sections in her 1990 paper is even entitled 'Deciding between the bootstrapping hypotheses'. Though the opposition 'semantic versus syntactic bootstrapping' is catchy, I suggest it be dropped. The opposition is a false one, because the theories are theories about different things. Moreover, there is no relationship between what Gleitman calls 'syntactic bootstrapping' and the metaphor of boot-straps, so the term makes little sense.

Gleitman uses the term 'semantic bootstrapping' to refer to the hypothesis that children learn verbs' meanings by observing the situations in which the verbs are used. But this is not accurate. 'Semantic bootstrapping' is not even a theory about how the child learns word meanings. It is a theory about how the child begins learning syntax. 'The bootstrapping problem' in grammar acquisition (see Pinker 1987) arises because a grammar is a formal system consisting of a set of abstract elements, each of which is defined with respect to other elements. For example, the 'subject' of a sentence is defined by a set of formal properties, such as its geometric position in the tree with respect to the S and VP nodes, its ability to force agreement with the verb, its intersubstitutability with pronouns of nominative case, and so on. It cannot be identified with any semantic role, sound pattern, or serial position. The bootstrapping problem is: How do children break into the system at the very outset, when they know nothing about the particular language? If you know that verbs agree with their subjects, you can learn where the subjects go by seeing what agrees with the verb – but how could you have learned that verbs agree with their subjects to begin with, if you don't yet know where the subjects go? How can children 'lift themselves up by their bootstraps' at the very outset of language acquisition, and make the first basic discoveries about the grammar of their language that are prerequisite to any further learning?

Pinker (1982), following earlier suggestions of Grimshaw (1981), suggested that certain contingencies between perceptual categories and syntactic categories, mediated by semantic categories, could help the child get syntax acquisition started. For example, if the child was built with the universal linking rule that agents of actions were subjects of active sentences, and they could infer from a sentence's perceptual context and the meanings of some of its content words that a particular word referred to the agent of an action, the child could infer that that word was in subject position. Once the position of the subject is established as a rule or parameter of the child's nascent grammar, further kinds of learning can proceed. For example, the child could now infer that any new word in this newly-identified position must be a subject, regardless

of whether it is an agent; he or she could also infer that verbs must agree in person and number with the element in that position. See Pinker (1984) and (1987) for a more precise presentation of the hypothesis.

The semantic bootstrapping hypothesis does require, as a background assumption, the idea that the semantics of at least *some* verbs have been acquired without relying on syntax. That is because the theory is about how syntax gets 'bootstrapped' at the very beginning of learning; if all word meanings were acquired via knowledge of syntax, and if syntax were acquired via knowledge of words' meanings, we would be faced with a vicious circle. The semantic bootstrapping hypothesis is agnostic about *how* children have attained knowledge of these word meanings. Logically speaking, they could have used telepathy, surgery, phonetic symbolism, or innate knowledge of the English lexicon, but the most plausible suggestion is that the children had attended to the contexts in which the words are used. Gleitman takes this latter assumption (that the child's first word meanings are acquired by attending to their situational contexts), generalizes it to a claim that all verb meanings are acquired by attending to their situational contexts (i.e., even verbs acquired after syntax acquisition is underway), and refers to the generalized claim as 'semantic bootstrapping'. But this is a large departure from its intended meaning.

And what Gleitman calls, in contrast, 'syntactic bootstrapping', is not a different theory of how the child begins to learn syntax. Thus it is not an alternative to the semantic bootstrapping hypothesis. (The only reason they could be construed as competitors is that semantic bootstrapping assumes that at least *some* verb meanings can be acquired before syntax, so a very extreme form of Gleitman's negative argument, that *no* verb meaning can be learned without syntax, is incompatible with it.) More-over, since 'syntactic bootstrapping' is a theory of how the child learns the meanings of specific verbs, and since it can only apply at the point at which the child has *already* acquired the syntax of verb phrases, it is not clear what it has to do with the 'bootstrapping problem' or the metaphor of lifting oneself up by one's bootstraps. For these reasons, I suggest that the term be avoided.

Here is a somewhat cumbersome, but transparent and accurate set of replacements. 'Semantic cueing of syntax' refers to the semantic bootstrapping hypothesis. 'Semantic cueing of word meaning' refers to the commonplace assumption that meanings are learned via their semantic contexts (perceptual or linguistic). 'Syntactic cueing of word meaning' is the hypothesis defended by Gleitman and her collaborators.

Now, in some contexts Gleitman does present a genuine alternative to the semantic bootstrapping hypothesis. She suggests that the child can use the prosody of a sentence to parse it into a syntactic tree. Though she never specifies exactly how this could be done, presumably the child would assume that pauses or falling intonation contours signal phrase boundaries. Having thus inferred a syntactic tree, the child could infer a verb's meaning from the trees it appears in. Note, though, that the information that the child uses to get syntax acquisition started is not itself syntactic, but prosodic; the hypothesis can thus sensibly be called 'prosodic bootstrapping'. If both prosodic bootstrapping, and syntactic cueing of word meaning were possible, semantic bootstrapping would be otiose.

But while it is plausible that the infant uses prosodic information to *help* in sentence analysis at the outset of language acquisition (e.g., to identify utterance boundaries), it is completely implausible that this information is *sufficient* to build a *full syntactic tree* for an input sentence (see Pinker 1987). The prosodic bootstrapping hypothesis, taken literally, is quite extraordinary. It is tantamount to the suggestion that there is a computational procedure that can parse sentences from any of the world's 5,000 languages when the sentences are spoken from behind a closed door (i.e., the sentences are filtered so that only prosodic information remains). Among the surprising corollaries to this claim is that it should be fairly easy for a person or machine to give a full parse to an English sentence heard from behind a closed door, because the listener can use both the universal and the English-specific mappings between prosody and syntax, whereas the child supposedly is capable of doing it using only the universal mappings. If, on the contrary, we, knowing English, cannot parse a sentence from behind a closed door, it suggests that the young child, not knowing English, is unlikely to be able to do so either. Thus the claim that infants can bootstrap syntax *from* prosody must be viewed with considerable skepticism.[2]

Overview of Gleitman's arguments for the syntactic cueing of verb semantics. With these independent issues out of the way, we can now turn to Gleitman's arguments for the importance of the syntactic cueing of verb meaning. These arguments fall into three categories. There are *negative arguments*: verb meanings cannot be learned from observation of situational contexts alone; therefore some other source of information is required. There is a *positive*

---

[2] Moreover, many of the 'syntactic' frames that Gleitman assumes the child is discriminating in order to infer verbs' meanings are prosodically identical, such as frames differing only in the specific prepositions or complementizers they contain, like *in* versus *on* or *that* versus *if* (see, e.g., Gleitman 1990: table 2).

hypothetical argument: verb meanings *could* be learned from verb syntax; therefore verb syntax probably is that other source. And there are *empirical arguments*: Children in fact learn verb meaning from verb syntax. I will examine these arguments separately.

3. The negative arguments: Verb meanings can't be learned from observation

Gleitman presents six arguments why attending to the situations in which a verb is used (what she calls 'observation') is in principle inadequate to learn the verb's meaning. I believe that none of the arguments establishes her main point, that there is an in-principle gap in observational learning that only syntactic subcategorization information can fill. There are two problems in the argument.

3.1. Arguments directed against 'observation learning' only refute learning by associative pairing

The first problem is that Gleitman's arguments are not aimed at 'observation' in general. They are aimed at a particular straw theory of observation. This foil, a version of one-trial associative pairing, has the child identify a verb's meaning with the sensory features activated by the situation at the moment when a verb is uttered. But no one believes this particular theory, so refuting it is ineffective in establishing in-principle limitations on observation; a few uncontroversial assumptions make Gleitman's objections moot. Let me examine the arguments in order.

3.1.1. Multiply-interpretable events

Any single event is multiply-ambiguous as to which verb it exemplifies. Gleitman (1990) notes, for example, that most situations of *pushing* also entail *moving*. If a situation is described as (say) *The boy is pushing the truck*, the child cannot know whether *push* means 'push' or 'move'.

This point, however, only shows that children cannot learn the meaning of a verb from a single situation. But no one, not even the British associationists and their descendants, has ever suggested they do. Simply allow the child to observe how a verb is used across multiple situations (see, e.g., Pinker 1989: ch. 6), and the problem disappears. Sooner or later, *push* will be used for instances of pushing without moving (e.g., pushing against a wall, or pushing

someone who holds his ground), and *move* will be used for instances of moving without pushing (e.g., sliding or walking). To take another one of Gleitman's examples (1990: 14), even though a single event may be describable as *pushing*, as *rolling*, and as *speeding*, most events are not. The child need merely wait for an instance of rolling without pushing or speeding, speeding without pushing or rolling, and pushing without rolling or speeding. See Gropen et al. (1991a) for experimental demonstrations that children use this kind of information.

3.1.2. Paired verbs that describe single events

Gleitman (1990: 16; see also Fisher et al. 1991: 380) suggests that there are pairs of verbs that overlap 100% in the situations they refer to. For example, there can be no *giving* without *receiving*, no *winning* without *beating*, no *buying* without *selling*, and no *chasing* without *fleeing*.

In fact, I doubt that pairs of verbs that refer to *exactly* the same set of situations exist (or if they do, they must be extremely rare.) Such pairs would be exact synonyms, and there is good reason to believe that there are few if any exact synonyms (Clark 1987, Bolinger 1977, Miller and Fellbaum 1991). To take just these examples, I can *receive* a package even if no one *gave* it to me; perhaps I wasn't home. John, running unopposed, can *win* the election, though he didn't *beat* anyone, and the second-place Celtics *beat* the last-place Nets in the standings last year, though neither *won* anything. Several of my gullible college friends *sold* encyclopedias door to door for an entire summer, but in many cases, no one *bought* any; I just *bought* a Coke from the machine across the hall, but no one *sold* it to me. If John *fled* the city, no one had to be *chasing* him; Bill can *chase* Fred even if Fred isn't *fleeing* but hiding in the garbage can.

I would certainly not claim that the learning of all these distinctions awaits the child's experience of the crucially disambiguating situation. But a lot of it could, and more important, the *in-principle* arguments for an alternative that are based on putative total overlap among verb meanings are not valid if meanings rarely overlap totally.

3.1.3. The subset problem

In some cases, Gleitman suggests, verb learning is impossible even if verbs do not totally overlap in the situations to which they refer. If the situations referred to by Verb A are a *superset* of the situations referred to by Verb B, a

child who mistakenly thought that Verb B had the same meaning as Verb A could never reject that hypothesis by observing how Verb B is used; all instances would fit the A meaning, too. The only disconfirming experience would be overt correction by parents, and there is good reason to believe that children cannot rely on such corrections. This argument is parallel to one commonly made in the acquisition of syntax (see, e.g., Pinker 1984, 1989; Wexler and Culicover 1980; Berwick 1985, Marcus 1993). For example, *move*, *walk*, and *saunter* are in a superset relation; any child that thought that *saunter* meant *walk* would do so forever, because all examples of sauntering are also examples of walking.

But this is only a problem if the child is allowed to maintain synonyms in his or her vocabulary. If children do not like to keep synonyms around (see Carey 1982, Clark 1987, Markman 1989, for evidence that they do not), then if they have a verb A (e.g., *walk*), and also a verb B (*saunter*) that seems to mean the same thing, they know something is wrong. They can look for additional meaning elements from a circumscribed set to make the meaning of B more specific (like the manner of motion). Pinker (1989: ch. 6) outlines a mechanism for how this procedure could work.

3.1.4. The poor fit of word to world

Gleitman suggests that even when a verb corresponds in principle to a unique set of situations, it is not, in practice, reliably used in that set of situations, so the child has no way of figuring out a verb's meaning based on the situations it actually is used in.

For example, Landau and Gleitman showed that the blind child they studied learned haptic equivalents of the verbs *look* (roughly, 'palpate' or 'explore haptically') and *see* (roughly, 'sense haptically'). But, they found, her mother didn't use *look* and *see* more often when object was near than when object was far.

The point of this argument is unclear. Of course, the mother didn't necessarily use *look* when an object was near. *Look* doesn't mean 'an object is near'; it means 'look'. The lack of correlation between some easily sensed property like nearness and use of a verb is only relevant if the child is confined to considering lists of sensory properties as possible verb meanings. If children can entertain the concept of looking, in something like the adults' sense (and Gleitman 1990: 4, assumes they can), it doesn't matter how many sensory properties a verb fails to correlate with if those properties define only a crude approximation of the verb's actual meaning. (This is a problem, for

example, with the conclusions drawn by Lederer et al. 1989.) All that matters is whether a child can recognize situations in which that correct concept applies.

Gleitman (1990) then turns to a stronger argument. Even when one examines genuine instances of the concept corresponding to a verb's meaning, one finds a poor correlation with instances of the parent uttering the verb. For example, in one study *put* was found to be used 10% of the time when there was no putting going on. Similarly, *open* was used when there was no opening 37% of the time. As Gleitman notes, this is not a surprise when one realistically considers how parents interact with their children. When a mother, arriving home from work, opens the door, she is likely to say, *What did you do today?*, not *I'm opening the door.* Similarly, she is likely to say *Eat your peas* when her child is, say, looking at the dog, and certainly not when the child is already eating peas. Indeed, Gleitman (1990: 15) claims that 'positive imperatives pose one of the most devastating challenges to any scheme that works by constructing word-to-world pairings'.

The problem with this argument is that it, too, only refutes the nonviable theory of learning by associate pairing, in which verb meanings are acquired via temporal contiguity of sensory features and utterances of the verb. It doesn't refute any reasonable account, in which the child keeps an updated mental model of the current situation (created by multi-sensory object- and event-perception faculties), including the likely communicative intentions of other humans. The child could use this knowledge, plus the lexical content of the sentence, to infer what the parent probably meant. That is, children need not assume that the meaning of a verb consists of those sensory features that are activated simultaneously with a parental utterance of the verb; they can assume that the meaning of a verb consists of what the parent probably meant when he or she uttered the word. Thus imperatives, where the child is not performing the act that the parent is naming, are not 'devastating'. Certainly when a parent directs an imperative at a child and takes steps to enforce it, the child cannot be in much doubt that the content of the imperative pertains to the parents' wishes, not the child's current activities.

3.1.5. Semantic properties closed to observation

Gleitman considers this the 'most serious challenge' to the idea that children learn verb meanings by attending to their nonsyntactic contexts. Mental verbs like *think, know, guess, wonder, know, hope, suppose,* and *understand* involve private events and states that have no external perceptual

correlates. Therefore children could not possibly infer their meanings observationally.

One problem I see with this argument is that although children may not be able to observe other people thinking and the contents of others' beliefs, they *can* observe themselves thinking and the contents of their own beliefs. Similarly, children may not know what their mothers are feeling, but they certainly know what they are feeling. And crucially, in many circumstances so do their mothers. When a parent comments on what a child is thinking or feeling, that constitutes information about the meanings of the mental state verbs they use.

Moreover, there surely are ways to infer a person's mental state from his or her behavior. Indeed, the standard way that humans explain each other's behavior is to assume that it is caused by beliefs and intentions, which can only be inferred. This must be how adults, during ordinary speech production, know when to *use* mental verbs based on their own mental state or guesses about others', even though there is no obvious referent event. There is no principled reason that children could not *infer* meanings of *new* mental verbs using exactly the same information that adults employ to *use* existing mental verbs accurately.

3.1.6. Does a richer system of mental representation hurt or help the child?

Gleitman suggests that if children are not temporal contiguity associators – if they can entertain hypotheses about causes, mental states, goals, speakers' intentions, and so on – their learning task is even harder. For the very richness of such representational abilities yields a combinatorial explosion of logically possible hypotheses for the child to test.

This argument, however, seems to conflate two ideas: 'a rich set of hypotheses', and 'a set of rich hypotheses'. Gleitman correctly points out that a rich (i.e., numerous) set of hypotheses is a bad thing if you're a learner. But replacing her associative-pairing mechanism with a cognitively more sophisticated one results in a set of rich (i.e., structured) hypotheses, not a rich set of hypotheses. And a set of rich hypotheses may in fact be fewer in number than a set of impoverished ones (e.g., combinations of sensory features) in any given situation: creatures with complicated human brains see the world in only a few of the logically possible ways. Presumably there are many more hypotheses for a learner who considers all subsets of patches of color and bits of fur and whisker than there are for a learner with a sophisticated object-recognition system who obligatorily perceives these patches as a single

'rabbit'. The whole point of a rich computational apparatus is to reduce the interpretations of a scene to the small number of correct ones. This is exactly what is needed to help solve the learning problem.

3.2. Problems in understanding observational learning do not constitute evidence for syntactic cueing

In much of her discussion, Gleitman attempts to place the burden of proof on anyone who believes that verb learning depends on observation, by identifying many areas of ignorance and difficult puzzles regarding how it could work. Indeed, anyone who thinks that a child can infer what a parent means from the situation and the nonverb content of the sentence must propose that a heterogeneous collection of not-very-well specified routes to knowing — indeed, the entirety of cognition — is available for use in the learning of verb meanings. Moreover, any such proposal must deal with the fact that even the most perceptive child and predictable parent cannot be expected to be in perfect synchrony all the time.

Gleitman's discussion contains penetrating and valuable analyses that clearly define central research problems in how children learn the meanings of words. But to support the alternative claim that verb subcategorization information is crucial, it is necessary to show that *no* theory of inferring communicative intent could ever be adequate, not that we currently don't have one that is fully worked out.

Moreover, Gleitman's attempt to shift the burden of proof ultimately fails, because she herself, at the end of the 1990 article and in Fisher et al. (1991 and this volume), concedes (in response to some of the points I elaborate on in the next section) that some form of observational learning in indispensable. She notes that information about manner of motion, type of mental state, nature of physical change undergone, and so on, are simply not available in the syntax of subcategorization: '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* ... Luckily, these distinctions are almost surely of the kinds that can be culled from transactions with the world of objects and events' (Gleitman 1990: 35).

This concession, however, completely redirects the force of Gleitman's criticisms of observational learning. For the meaning components that Gleitman agrees are learned by observation are the very components that she, earlier in the article, claimed that observation cannot acquire! For example, the fact that *open* is often used when opening is not taking place (e.g.,

imperatives), and that *open* is not used when opening is taking place (e.g., when someone enters the house), if it is relevant at all, pertains in full force to the 'delicate and specific' aspects of the meaning of *open* (i.e., those aspects that differentiate it from syntactically identical *close*). Similarly, parents surely cannot be counted on to use *break* or *tear* when and only when breaking or tearing are taking place, respectively. Nonetheless, Gleitman concedes that the meanings specific to *open*, *break*, and *tear* are somehow learned by observation. Thus it is not true, as she suggests (1990: 48), that 'semantically relevant information in the syntactic structures can rescue observational learning from the sundry experiential pitfalls that threaten it'. There are pitfalls, to be sure, but for most of the ones Gleitman originally discussed, syntax offers no rescue. What we need is a better, non-associationist theory of observational learning.

3.3. Conclusions about Gleitman's arguments against observational learning

Gleitman convincingly refutes a classical associationist theory of semantic learning, in which word meanings are acquired via temporal contiguity of sensory features of the scene and utterances of the word. She also convincingly shows that to explain verb learning, we need a constrained representational system for verbs' meanings, principles constraining how one verb is related to another in the lexicon, a learning mechanism that can construct and modify semantic representations over a set of uses of the verb, and a greater understanding of how children interpret events, actions, mental states, and other speakers' communicative intentions. But the arguments do not show that the full set of semantic cues to semantics is so impoverished in principle that the child must use sets of syntactic subcategorization frames as cues instead, nor that syntactic cues provide just the information that semantic cues fail to provide. Rather, Gleitman herself assumes that there exists some form of observational learning powerful enough to acquire aspects of meaning that her own arguments show to be hard to acquire.[3]

---

[3] Paul Bloom has pointed out to me that arguments similar to Gleitman's were originally made by Chomsky (1959) in his review of Skinner's *Verbal Behavior*. For example, Chomsky showed that noun meanings could not in general be learned by hearing the nouns in the presence of their referents. But Chomsky used examples like *Eisenhower*, a proper name, whose meaning could not possibly be distinguished using syntactic cues from the thousands of other proper names that must be learned (e.g., *Nixon*). This suggests that observation and syntactic cues are not the only possible means of learning. See Bloom (this volume) for discussion of similar issues in the learning of noun meanings.

4. The positive hypothetical argument: Semantic information in subcategorization frames

Gleitman and her collaborators give a few specific examples of how a learner might use a verb's syntactic properties to predict aspects of its meaning. Unfortunately, they do not relate these examples to a general theory of the relation of syntax to semantics in verbs' lexical entries and of how a learner could exploit them. In this section I will attempt to fill this gap by laying out the logic of verbs' syntax and semantics and the implications of that logic for learning.

4.1. Verb roots versus verb frames

The first question we must ask is, what do we mean by 'a verb'? The term is ambiguous in a critical way, because in most languages a verb can appear in a family of forms, each with a distinct meaning component, plus a common meaning component that runs throughout the family. For example, many verbs can appear in transitive, intransitive, passive, double-object, prepositional object, and other phrases. In some cases the verb actually changes its morphological form across these contexts, though in English only the passive is marked in this way. Following standard usage in morphology, we can say that all the forms of a given verb share the same verb *root*. We can then call the syntactically distinct forms of a given root its *frames*. For example, consider the matrix of verb forms in figure 1, where the existence of a given root in a given frame is marked with an 'x'.

The meanings of the *x*'s differ along two dimensions. Let me use the term *root meaning* to refer to the aspects of meaning that are preserved in a given root across all the frames it appears in; that is, whatever aspects of meaning *The water boiled* and *I boiled the water* have in common, and fail to share with *The door opened* and *I opened the door*. Let me use the term *frame meaning* to refer to the orthogonal dimension: the aspects of meaning that are shared across all the roots that appear in a given frame; that is, whatever aspects of meaning differentiate *The water boiled* from *I boiled the water*, and that *The water boiled* and *The door opened* have in common.

Note that root meanings are much closer to what people talk about when worrying about acquisition of word meaning. That is, the main problem in learning *boil* is learning that it is about hot liquid releasing bubbles of gas. This is the aspect of *boil* that is found in both its transitive and intransitive uses, that is, its root meaning. The root meaning corresponds to what we

Frames

Roots	NP_	NP_NP	NP_S	NP_PP	NP_NP-PP	NP_NP-S
eat	×	×				
move	×	×		×	×	
boil	×	×				
open	×	×				
kill		×				
die	×					
think	×		×	×		
tell		×			×	×
know		×	×	×		
see		×	×	×	×	
look		×		×		

Fig. 1

think of as the *content* of a verb. The frame meaning – the fact that there must be an agent causing the physical change when the verb is used in the transitive frame, and that the main event being referred to is the causation, not the physical change – is just as important in understanding the sentence, but it is not inherently linked to the verb root *boil*. It is linked to the transitive syntactic construction, and would apply equally well to *melt*, *freeze*, *open*, and the thousands of other verb roots that could appear in that frame. This is a crucial distinction.

4.2. Learning about a verb in a single frame

The first question that follows is, What can be learned from hearing a verb in one frame? Something, clearly, for frame semantics and frame syntax are highly related. For example, it is a good bet that in *A glips B to C, glip* is a verb of transfer. The regularities that license this inference are what linguists call *linking rules* (Carter 1988, Jackendoff 1987, 1990; Pinker 1989, Gropen et al. 1991a). For example, if *A* is a causal agent, *A* is the subject of a transitive verb. Linking rules are an important inferential mechanism in semantic bootstrapping (semantic cueing of syntax at the outset of language acquisition), in predicting how one can use a verb once one knows what it means, and in governing how verbs alternate between frames (see Gropen et al. 1991a for discussion).

One might now think: If syntax correlates with semantics, why not go both ways? If one can infer a verb's syntax from its semantics (e.g., in semantic bootstrapping), couldn't one just as easily infer its semantics from its syntax? As Gleitman puts it (1990: 30):

'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 on to 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.'

I believe this argument is problematic. The problem is that a *correlation* is not the same thing as an *implication*. 'Correlation' means 'many X's are Y's or many Y's are X's or both'. 'Implication' means 'if X, then Y, though not necessarily vice-versa'. The asymmetry inherent in an implication is crucial to understanding how it can be used predictively. For example, if I feed two numbers (e.g., 3 and 5) into the *sum-of* function, the value must be 8. But if I guess which inputs led to a value of 8, I cannot know that they were 3 and 5.

Linking rules are implications. They cannot straightforwardly be used in the reverse direction. If a verb means 'X causes Y to shatter', then X is the subject of the verb. But if X is the subject of a verb, the verb does not necessarily mean 'X causes Y to shatter'. This asymmetry is inherent to the design of language. A grammar is a mechanism that maps a huge set of semantic distinctions onto a small set of syntactic distinctions (for example, thousands of kinds of physical objects are all assigned to the same syntactic category 'noun'). And because this function is many-to-one, it is not invertible.

Now, if one casts away most of the meaning of a verb (e.g., the part about shattering), there may remain some abstract feature of meaning that could map in one-to-one fashion to syntactic form. To the extent that that can be done, one could learn some things about a verb form's meaning from the frame that the verb appears in. First, one can learn how many arguments the verb relates in that form, as in the difference between *The water boiled* (one argument) and *She boiled the water* (two arguments), or the difference between *die* (one argument) and *kill* (two arguments). Second, one can infer something about the logical type of some of the arguments, like 'proposition' (if the verb appears with a clause) versus 'thing' (if the verb appears with an NP) versus 'place/path' (if the verb appears with a PP). That is, the syntax can help one distinguish between the meaning of *find* in *find the book* and *find*

that the book is interesting; between *shoot the man* and *shoot at the man*; perhaps even between *think, eat,* and *go*. Third, the syntax of a sentence can help identify which argument can be construed as the agent (viz., the subject) in cases where the inherent properties of the arguments (such as animacy) leave it ambiguous, for example, in *kill* versus *is killed by*, and *chase* versus *flee*. Similarly, syntactic information can distinguish the experiencer from the stimulus in 'psych-verbs' with ambiguous roles, such as *Bill feared Mary* and *Mary frightened Bill*. Fourth, syntactic information can help identify which argument is construed as 'affected' (viz., the syntactic object) in events where several entities are being affected in different ways. For example, in *load the hay* and *load the wagon*, on cognitive grounds either the hay or the wagon could be interpreted as 'affected': the hay, because it changes location, or the wagon, because it changes state from not full to full (similar considerations apply to the pair of verbs *fill* and *pour*. The listener has to notice which of the two arguments (content or container) appears as the direct object of the verb to know which one to construe as the 'affected' argument for the purpose of understanding the verb in that frame. Gleitman and her colleagues give many examples of these forms of learning, which I have called 'reverse linking' (see Pinker 1989 and Gropen et al. 1991a, b for relevant discussion and experimental data).

Unfortunately, while one can learn *something* about a verb form's meaning from the syntax of the frame it appears with, especially when there are a small number of alternatives to select among, one cannot learn *much*, relative to the full set of English verbs, because of the many-to-one mapping between the meanings of specific verbs and the frames they appear in. For example, one cannot learn the differences among *slide, roll, bounce, skip, slip, skid, tumble, spin, wiggle, shake,* and so on, or the differences among *hope, think, pray, decide, say,* and *claim*; among *build, make, knit, bake, sew,* and *crochet*; among *shout, whisper, mumble, murmur, yell, whimper, whine,* and *bluster*; among *fill, cover, tile, block, stop up, chain, interleave, adorn, decorate* and *face,* and so on. Indeed, Gleitman herself (1990: 35) concedes this point in the quote reproduced above.

In sum, learning from *one frame* could help a learner distinguish *frame meanings,* that is, what *the water boiled* has in common with *the ball bounced* and does not have in common with *I boiled the water*. But it does *not* distinguish *root meanings,* that is, the difference *the water boiled* and *the ball bounced.* And the root meanings are the ones that correspond to the 'content' of a verb, what we think of as 'the verb's meaning', especially when a given verb root appears in multiple frames.

The frame meanings (partly derivable from the frame) are closer to the 'perspective' that one adopts relative to an event: whether to focus on one actor or another, one affected entity or another, the cause or the effect. Indeed in some restricted cases, differences in perspective are most of what distinguishes pairs of verb roots, such as *kill* and *die, pour* and *fill,* or Gleitman's example of *chase* and *flee.* Gleitman (1990) and Fisher et al. (this volume) adopt a metaphor in which the syntax of a verb frame serves as a 'zoom lens' for the aspects of the event referred to by the verb. This metaphor is useful, because it highlights both what verb syntax can do and cannot do. The operation of lens when aimed at a given scene gives the photographer three degrees of freedom, pan, tilt, and zoom, which have clear effects on the perspective in the resulting picture. But no amount of lens fiddling can fix the vastly greater number of degrees of freedom defined by the potential contents of the picture – whether the lens is aimed at a still life, a nude, a '57 Chevy, or one's family standing in front of the Grand Canyon.

So I have no disagreement with Gleitman's arguments that a syntactic frame can serve as a zoom lens, helping a learner decide which of several perspectives on a given type of event (discerned by other means) a verb forces on a speaker. But because this mechanism contributes no information about a verb's content, it cannot offer significant help in explaining how children learn a verb's content despite blindness, nor in explaining how children learn a verb's content despite the complexity of the relationship between referent event and parental usage.

4.3. Learning about a verb from its multiple frames

Gleitman recognizes the limitations of learning about a verb's meaning from a single frame:

'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 possible meaning of an individual verb quite closely.' (Gleitman 1990: 30–32)

The claim that inspection of multiple frames can predict a verb's meaning 'quite closely' appears to contradict the earlier quote in which Gleitman notes that syntactic information in general is not 'delicate and specific enough to

distinguish among ... semantically close items'. To see exactly how close the syntax can get the learner to a correct meaning, we must ask, 'What can be learned from hearing a verb in *multiple* frames?' In particular, can a root meaning – the verb's content – be inferred from its set of frames, and if so, how?

Unfortunately, though Gleitman and her collaborators give examples of how children might converge on a meaning from several frames, almost always using the problematic example of *see* (see fn. 1), they never outline the inferential procedure by which children do so in the general case. In Fisher et al. (this volume) they suggest that the procedure is simply the zoom lens (single-frame) procedure applied 'iteratively'. They give the procedure as follows: 'In assigning a gloss to the verb, satisfy all semantic properties implied by the truth conditions of all its observed syntactic frames'. But this cannot be right, for reasons they mention in the next paragraph. The truth conditions (what I have been calling 'frame meaning') that belong to a verb form in one frame do *not* belong to it in its other frames. So satisfying all of them will not give the root meaning or verb's content. If we interpret 'satisfying all semantic properties' as referring to the *conjunction* of the frame meanings, we get the meaning of its most restrictive frame, which will be incompatible with its less restrictive frames. For example, the truth conditions for transitive *boil* include the presence of a causal agent. But presence of a causal agent cannot be among the semantic properties of *boil* across the board, for its intransitive version (*The water boiled*) is perfectly compatible with spontaneous boiling in the absence of any agent. But if we interpret 'satisfying all semantic properties' to be the *disjunction* of frame meanings, the aggregation leads to virtually no inference at all. Consider again the frame involved in *The water boiled*. This intransitive frame tells the learner that the meaning of *boil* in the frame consists of a one-place predicate. Now consider a second frame, the one involved in *I boiled the water*. This transitive frame tells you (at most) that the meaning of *boil* in the frame consists of causation of some one-place predicate. What do they have in common? 'One-place predicate'. Which is not very useful. It says nothing whatsoever about the root meaning of *boil*, that is, that it pertains to liquid, bubbles, heat, and so on.

This is a problem even for verbs that appear in many frames, for which the syntax would seem to provide a great deal of converging information (see Levin 1985, Pinker 1989). For example *sew* implies an activity. *Sew the shirt* implies some activity performed on an object. *Sew me a shirt* implies an activity creating an object to be transferred to a beneficiary. *Sew a shirt out*

of the rags implies an activity transforming material into some object. What do these frame meanings have in common? Only 'activity'. *Not* 'sewing'.

The conclusion is clear: you can't derive a verb's root meaning or content by iterating the zoom lens procedure over multiple frames and taking the resulting union or intersection of perspectives.

4.3.1. Can anything *be learned from multiple frames?*

I do not wish to deny that there is some semantic information implicit in the set of frames a verb appears with, nor that an astute learner could not, in principle, use this information. The example Gleitman uses most often, *see*, has clear intuitive appeal. But which general procedure is driving the inference about *see* and other such cases? I can think of two.

According to Gleitman, a set of argument frames implicitly poses the question, 'What notion is compatible with involving a physical object, involving a proposition, and involving a direction?' The child deduces the response 'seeing'.[4] In other words, this is a kind of cognitive riddle-solving (Pinker 1989); it involves all of a learner's knowledge, beliefs, and cognitive inferential power.

I am not arguing either that children can or cannot solve such riddles. I am simply pointing out what would be going on if they could do so. In particular, note what they would not be doing. They would not be relying on any *grammatical* principle, and hence would not be enjoying the putative advantages of universal constrained linguistic principles to drive reliable inferences. That is, if guessing a verb's meaning from its set of frames succeeds at all, it does so by virtue of the child's overall cognitive cleverness, and hence could suffer from the same unreliability of overall cleverness as inferring a speaker's likely meaning from the knowledge of the situation. It is not a straightforward mechanical procedure that succeeds because the frames 'are abstract surface reflexes of the meanings' (Landau and Gleitman 1985: 138)

---

[4] Actually, the question and answer should be stated in terms of 'a family of notions', not 'notion', because verbs like *see* that can take either objects or clausal complements do not exhibit a single content meaning across these frames: '*see* NP' does not mean the same thing as '*see* S'. The latter is not even a perception verb: *I see that the meal is ready* does not entail vision. (Clearly not, because you can't visually perceive a proposition.) Similarly, *I feel that the fabric is too smooth* does not entail palpation; it's not even compatible with it. And *Listen! I hear that the orchestra is playing* is quite odd. (These observations are due to Jane Grimshaw.) Clearly there is a commonality running through each of these sets, but it is a metaphorical one; 'knowing' can be construed metaphorically as a kind of 'perceiving'.

or because 'much of the [semantic] information can be read off from the subcategorization frames themselves by a general scheme for interpreting these semantically' (Landau and Gleitman 1985: 142). Moreover, the premises that would drive this riddle-solving are far more impoverished than the premises derived from inferring a speaker's meaning from the context. The latter can include any concept the child is capable of entertaining (sewing, boiling, and so on); the former are restricted to a smaller set of abstract concepts like causability and taking a propositional argument.

There is a second way that sets of syntactic frames could assist semantic learning. That is via *narrow argument-structure alternations*. Often the verbs that can appear with a particular set of syntactic frames have surprisingly specific meanings (see Levin 1985, in press; Pinker 1989, for reviews). For example, in English, the verbs that can appear in the double-object form but not the *to*-object form are verbs of implied deprivation like *envy*, *bet*, and *begrudge* (e.g., *I envied him his good looks/*envied his good looks of him*). Similarly, verbs of manner of motion can alternate between causative-transitive and inchoative-intransitive forms (e.g., *I slid the puck/The puck slid*), but verbs of direction of motion cannot (e.g., *I brought the book/*The book brought*). An astute learner, in principle, could infer, from hearing *I glipped him those things* and from failing to hear *I glipped those things of him*, that *glip* involves some intention or wish to deprive someone of something. But note that these regularities are highly specific to languages and to dialects within a language. (For example, **I suggested her something* is grammatical in Dutch, and **I pushed him the box* is grammatical in some dialects of English; see Pinker 1989.) Exploiting them requires first having acquired these subtle subclasses and their syntactic behavior in the dialect, presumably by abstracting the subclasses from the semantics and syntax of individual verbs, acquired by other means. This kind of inference depends on a good deal of prior learning of verbs' meanings in a particular language, and thus is most definitely not a case of 'bootstrapping' performed by a child to acquire the meanings of the verbs to begin with.

In general, learning a verb's content or root meaning from its set of syntactic frames ('syntactic bootstrapping') is fundamentally different from learning its perspective or frame meaning from a single frame ('zoom lens'). Thus I disagree with Gleitman's (1990) suggestion that they are versions of a single procedure, or Fisher et al.'s suggestion (this volume) that one is simply the iteration of the other. There is a clear reason why they are different. While there may be a universal mapping between the meaning of a frame and the syntax of that frame (allowing the lens to zoom), there is no universal

mapping between the meaning of a root and the *set* of frames it occurs in (see Talmy 1985 and Pinker 1989 for reviews). For example, universal linking rules imply, roughly, that an inchoative verb can appear in an intransitive frame, and a causative verb can appear in a transitive frame. And it's clearly possible for some roots to be able to have both causative and inchoative meanings (and hence to appear in both frames). But it's an accident of English that *slide* appears in both frames, but *come* and *bring* appear in one each. Thus the kinds of learning that are licensed by universal, reliable, grammatical linking regularities are restricted to differences in perspective. A verb's content is not cued by any one of its syntactic frames, and at best might be related to its entire set of frames in a tenuous, language-specific way.

5. Experimental evidence on children's learning of verb meanings from verb syntax

Gleitman (1990) reviews three ingenious experimental studies intended to support the claim that children can and do use a verb's syntax to learn its meanings. Though the experiments are important contributions to our understanding of the acquisition of verb meaning, they do not speak to Gleitman's proposals about 'syntactic bootstrapping', because none of the experiments involved the presentation of multiple frames to a child. Let me review them.

5.1. Hirsh-Pasek et al. (1988)

One set of studies (Hirsh-Pasek et al. 1988) presented 27-month-olds with two video screens. One showed a pair of puppets performing a simple action like flexing their own arms; the other showed one puppet causing the second to perform the action, by manipulating its arms. A recorded voice used a verb unknown to the child in one of two syntactic frames, either the intransitive 'Big Bird is flexing with Cookie Monster' or the transitive 'Big Bird is flexing Cookie Monster'. Children looked at the screen showing the noncausal scene more often when the sentence was intransitive, and looked at the screen showing the causal scene more often when the sentence was transitive.

Note that this study provided children with only a *single* frame, not multiple frames. And note as well that what the children were sensitive to was a frame-specific perspective (causation by an agent), not a root-specific

content (flexing). The content (flexing) was acquired through observation, not syntax; it was depicted on the video screen, and the child was watching it. Thus at best the children were demonstrating use of the zoom lens procedure; there was no opportunity for multiple frames to cue the verb's content. (At worst, the children were not acquiring any information about the verb at all, but were ignoring the verb and merely responding to the transitive and intransitive sentence frames themselves when directing their attention.)

5.2. Naigles (1990)

A second study appears to show children learning a verb's content from a single frame. In Naigles (1990), 24-month-olds first saw a video of a rabbit pushing a duck up and down, while both made large circles with one arm. One group of children heard a voice saying 'The rabbit is gorping the duck'; another heard 'The rabbit and the duck are gorping'. Then both groups saw a pair of screens, one showing the rabbit pushing the duck up and down, neither making arm circles, the other showing the two characters making arm circles, neither pushing down the other. In response to the command 'Where's gorping now? Find gorping!', the children who heard the transitive sentence looked at the screen showing the up-and-down action, and the children who heard the intransitive sentence looked at the screen showing the making-circles action.[5]

What have the children learned, and from what source? Clearly, they learned most of the verb root's content – that *gorp* means pushing and/or making circles, but not sliding or boiling or killing or dancing – not by attending to the syntax, but by observing the scene; that's what the video depicted. Without the video, the children would have learned little if anything. What the children learned from the sentence syntax was, once again,

---

[5]   Unfortunately, the Naigles experiment had a confound, commented on by Gleitman, 1990, in footnote 13 on p. 43. The difference between the *gorps* could have been cued by the conjoined versus singular subjects, disregarding the verb syntax entirely. That is, the difference between *The rabbit is gorping* and *The rabbit and the duck are gorping*, with identical verb syntax, could have been sufficient for children to pick out a screen to look at. In the first case, the children would look at what the rabbit alone was doing; in the second case, they could look at what the rabbit and duck were doing simultaneously. This would be sufficient to direct their attention in testing to the push-down screen in the first case, the make-circles screen in the second. This confound, a good example of the difference between using linguistically-conveyed content as opposed to verbal syntax, could be eliminated by using the sentences of Hirsh-Pasek et al. The voice-over would say either 'the rabbit is gorping the duck' or 'the rabbit is gorping near the duck'. Children in the first condition, if they could use verb syntax as a cue, would still find *gorping* in the up-and-down screen, children in the second would find it in the make-circles screen.

the meaning of the verb in its particular frame. The sentence played to the first group of children told them that *gorp* is a two-place predicate, presumably a causative. It means either 'cause to pop up and down' or 'cause to make arm circles'. The sentence played to the second group told them that *gorp* is a one-place predicate, presumably an activity. It means either 'make arm circles' or 'pop up and down'. Once again, the syntactic frame cued only the coarse information of how many arguments were immediately related by the verb; the rest came from observation.

One might object at this point that Naigles's experiment has demonstrated what I have been arguing is impossible: children appear to have learned about a verb's content (in the case, up-and-down versus make-circles) from the sentence in which it is used, and could not have learned that content from observation alone. But this is misleading. Success depended completely on the fact that Naigles engineered an imaginary world in which perspective and content were confounded, so that when children were using syntax to choose the right perspective, they got the right content, too, by happy accident. Note there is no grammatical constraint forcing or preventing either 'popping up and down' or 'making arm circles' from being exclusively transitive, exclusively intransitive, or alternating. Nor is there any real-world constraint that could cause creatures to make arm circles and to pop up and down in tandem. But Naigles's teaching example exemplified both such constraints: only popping up-and-down was causable, and such causation took place in the presence of arm circles. It was only these artificial contingencies that made the forms learnable by syntax rather than observation. Consider what would have happened if the children had been shown a scene depicting circles without pushing up and down or vice-versa. In that case, observation would have been sufficient to distinguish the two actions, with no syntax required. Now consider what would happen if the children had been shown an arm-circling rabbit causing the duck simultaneously to pop up and down *and* to make arm circles. This is no more or less bizarre than the conjunction that Naigles did show children, where causation of popping up and down was simultaneous with uncaused arm-circling. In that case, neither the sentence *The duck is gorping the rabbit* nor the sentence *The duck and rabbit are gorping* would have distinguished the two kinds of motion. This shows that syntax is neither necessary nor sufficient to distinguish the alternative content meanings of *gorp*, across all the different scenes in which it can be used; observation, in contrast, *is* sufficient. In sum, Naigles simply selected a contrived set of exposure conditions that penalized observation while letting syntax lead to the right answer by coincidence.

5.3. Fisher et al. (1994)

The third experiment is reported by Fisher et al. (this volume). Here 4-year-olds viewed scenes that were consistent with two verbal descriptions, like *eat* and *feed*, *chase* and *flee*, *ride* and *carry*, *give* and *take*, or *put* and *carry*. A puppet described the action using a novel verb root like *zorp*. For each sentence, different children heard the verb used in one of two syntactic forms. Depending on the action, the forms differed in transitivity, in which animal was named as the subject or object, or in whether an animal or a thing was named as the direct object. The child was asked to describe the event for the puppet by providing it with the English translation of the puppet word. The children tended to provide the English word that corresponded to the syntactic frame that the puppet used.

There is a problem, though, with interpreting this experiment as evidence for syntactically-cued verb learning: the children didn't learn any verbs. They simply had to select *existing* English verbs that were the translation of the puppet words (this is literally how the task was described to the children). That is, the children weren't learing what *eat* or *feed* mean; they were simply demonstrating that by the time they had walked into the lab, they had previously learned (through whatever means) that *eat* can have a sole, actor argument whereas *feed* can have an agent and a beneficiary/goal argument. The task was: 'which English verb is consistent with the videos and has an agent subject and a beneficiary/goal object?' not, 'what does a verb mean if it has an agent subject and a beneficiary/goal object?' It's not surprising that children can be shown to know the answer to the first question, because that's exactly the knowledge they depend on every time they use the verb correctly in their speech. So the experiments show what children already know about English verbs, not how they learn English verbs.

Gleitman (1990) and Fisher et al. (this volume) recognize this problem but suggest that it provides indirect evidence in favor of their position. I think it might show little or nothing about verb learning at all, because the knowledge being used could be specific to what the children know about English. Suppose someone were to advance the (obviously wrong) hypothesis that people learn what a word means by its sound – words beginning with 'p' are causative, words beginning with 'f' are inchoative. To test it, they show a scene that is ambiguous between falling and pushing, and say 'the puppet calls this f....' or 'the puppet calls this p....'. Subjects guess *falling* and *pushing* accordingly. But obviously they do it by tapping particular facts of English acquired beforehand, with no utility in learning any other language. That is,

there is no reason to think that the *retrieval* cue that the experimenter provides now, for existing knowledge, was ever used as a *learning* cue, in order to acquire that knowledge originally. Substitute 'transitive syntax' and 'intransitive syntax' for 'p....' and 'f....' and one has the Fisher et al. experiment (this volume) – a test of whether children can use transitivity correctly as a retrieval cue for previously learned words when the content of the words is available observationally.

5.4. What experiment would show syntactic cueing of verb semantics?

There is an extremely simple experiment that could test whether children can learn a verb root's semantic content from multiple frames. There could be no TV screen, or content words, just syntactic frames. For example, children would hear only *She pilked*; *She pilked me something*; *She pilked the thing from the other things*; *She pilked the other things into the thing*; *She pilked one thing to another*, and so on. If children can acquire a verb's content from multiple frames, they should be able to infer that the verb basically means 'create by attaching' (Levin 1985). (Of course, one would have to ensure that the child was learning a new meaning and not simply using the frames to retrieve an existing word, for reasons mentioned in the preceding subsection.) Lest one think that this set of inputs is way too impoverished and boring for a child to attend to, let alone for the child to draw semantic conclusions from, in the absence of perceiving some accompanying real-world event, recall that this is exactly the situation that Landau and Gleitman assume the blind child is in. It would be an interesting finding if children (or adults) could learn significant aspects of a verb's content from syntactic cues, as this experiment would demonstrate. If Gleitman and her collaborators are correct, they should be able to do so.

6. Conclusions

I have gone over Gleitman's arguments against the sufficiency of learning verb semantics by observation of semantic cues in the situations in which a verb is used, and her arguments for the utility and use of syntactic sub-categorization information. I suggest that a careful appraisal of these arguments leads to the following conclusions.

As Gleitman shows, temporal contiguity between sensory features and verb usages cannot explain the acquisition of verb meaning. What this suggests is

that the explanation of verb learning requires a constrained universal apparatus for representing verb meanings, principles governing the organization of the lexicon, a perceptual and conceptual system acute enough to infer which elements of verb meanings an adult in a situation is intending to refer to, and a learning procedure that can compare hypothesized semantic representations across situations.

Gleitman has also convincingly demonstrated that single syntactic frames provide information about aspects of the meaning of the verb *in that frame* (the 'zoom lens' hypothesis). This information is largely about the perspective that a verb forces a speaker to take with regard to an event. It includes the number of arguments, the type of argument, a focus on the cause or effect, and the choice of agent and affected entity when more than one is cognitively possible. As Gleitman points out, these are exactly the kinds of information that are difficult or impossible to infer from observing the situations in which a verb is used.

I disagree, however, that multiple syntactic frames provide crucial information about the semantic content of a verb root across its different frames (what Gleitman calls 'syntactic bootstrapping'). There is no syntactically-driven general inferential scheme by which such learning could work; there is no empirical evidence that children use it; and it does not make up for any of the problems Gleitman notes in understanding how children learn about a verb's meaning from observing the situations in which it is used. Indeed, the suggestion is incompatible with one of the basic design features of human language: a vast set of concepts is mapped onto a much smaller set of grammatical categories.

References

Berwick, R.C., 1985. The acquisition of syntactic knowledge. Cambridge, MA: MIT Press.
Bloom, P., 1994. Possible names: The role of syntax-semantics mappings in the acquisition of nominals. Lingua 92, 297–329 (this volume).
Bolinger, D., 1977. Meaning and form. London: Longman.
Brown, R., 1957. Linguistic determinism and the part of speech. Journal of Abnormal and Social Psychology 55, 1–5.
Carey, S., 1982. Semantic development: The state of the art. In: E. Wanner, L.R. Gleitman (eds.), Language acquisition: The state of the art, 347–389. New York: Cambridge University Press.
Carter, R.J., 1988. On linking: Papers by Richard Carter. (Lexicon Project Working Paper 25). B. Levin, C. Tenny (eds.). Cambridge, MA: MIT Center for Cognitive Science.
Chomsky, N., 1959. A Review of B.F. Skinner's 'Verbal behavior'. Language 3, 26–58.

Chomsky, N., 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Clark, E.V., 1987. The principle of contrast: A constraint on language acquisition. In: B. Mac-Whinney (ed.), Mechanisms of language acquisition. Hillsdale, NJ: Erlbaum.

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

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

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

Gleitman, L.R., 1990. The structural sources of verb meaning. Language Acquisition 1, 3–55.

Grimshaw, J., 1979. Complement selection and the lexicon. Linguistic Inquiry 10, 279–326.

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

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

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

Gropen, J., S. Pinker, M. Hollander and R. Goldberg, 1991b. Syntax and semantics in the acquisition of locative verbs. Journal of Child Language 18, 115–151.

Hirsh-Pasek, K., H. Gleitman, L.R. Gleitman, R. Golinkoff and L. Naigles, 1988. Syntactic bootstrapping: Evidence from comprehension. Paper presented at the 13th Annual Boston University Conference on Language Development.

Jackendoff, R.S., 1987. The status of thematic relations in linguistic theory. Linguistic Inquiry 18, 369–411.

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

Katz, B., G. Baker and J. Macnamara, 1974. What's in a name? On the child's acquisition of proper and common nouns. Child Development 45, 269–273.

Landau, B. and L.R. Gleitman, 1985. Language and experience. Cambridge, MA: Harvard University Press.

Lederer, A., H. Gleitman and L.R. Gleitman, 1989. Input to a deductive verb acquisition procedure. Paper presented to the 14th Annual Boston University Conference on Language Development.

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

Levin, B., in press. English verb classes and alternations: A preliminary investigation. Chicago, IL: University of Chicago Press.

Marcus, G.F., 1993. Negative evidence in language acquisition. Cognition 46, 53–85.

Markman, E.M., 1989. Categorization in children: Problems of induction. Cambridge, MA: MIT Press/Bradford Books.

Markman, E.M., 1990. Constraints children place on word meanings. Cognitive Science 14, 57–77.

Miller, G.A., 1991. The science of words. New York: Scientific American Library.

Miller, G.A. and C. Fellbaum, 1991. Semantic networks of English. Cognition 41, 197–229.

Moravcsik, J.M.E., 1981. How do words get their meanings? The Journal of Philosophy 78, 5–24.

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

Pinker, S., 1982. A theory of the acquisition of lexical interpretive grammars. In: J. Bresnan (ed.), The mental representation of grammatical relations. Cambridge, MA: MIT Press.

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

Pinker, S., 1987. The bootstrapping problem in language acquisition. In: B. MacWhinney (ed.), Mechanisms of language acquisition. Hillsdale, NJ: Erlbaum.

Pinker, S., 1988. Resolving a learnability paradox in the acquisition of the verb lexicon. In: M. Rice, R. Schiefelbusch (eds.), The teachability of language. Baltimore, MD: Brookes.

Pinker, S., 1989. Learnability and cognition. The acquisition of argument structure. Cambridge, MA: MIT Press/Bradford Books.

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

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

Talmy, L., 1988. Force dynamics in language and cognition. Cognitive Science 12, 49–100.

Wexler, K. and P. Culicover, 1980. Formal principles of language acquisition. Cambridge, MA: MIT Press.