

## CHAPTER 16

---

# The Learned Component of Language Learning

*Lila Gleitman*

Isolated infants and children have the internal wherewithal to design a language if there isn't one around to be learned (e.g., Senghas and Coppola 2001). Such languages exhibit categories and structures that look suspiciously like those of existing languages. There are words like *horse* and *think*. Not only that: the mapping between predicate type and complement structure is also quite orthodox, as far as can be ascertained. For instance, even in very primitive instances of such self-made languages, *sleep* is intransitive, *kick* is transitive, and *give* is ditransitive (e.g., Feldman et al. 1978). This fits with recent demonstrations – one of which I mentioned during the round-table discussion (see page XXX) – that even prelinguistic infants can discriminate between certain two- and three-argument events in the presence of the (same) three interacting entities (Gordon 2003). All of this considerable conceptual and interface apparatus being in place, and (“therefore”) language being so easy to invent, one might wonder why it's hard to acquire an extant language if you are unlucky enough to be exposed to one. For instance, only ten or so of the required 50,000 or so vocabulary items are acquired by normally circumstanced children on any single day; three or four *years* go by before there's fluent production of modestly complex sentences in all their language-particular glory. What takes so long?

The answer generally proposed to this question begins with the problem of word learning, and is correct as far as it goes: ultimately, lexical acquisition is accomplished by identifying concepts whose exemplars recur with recurrent phonetic signals in the speech or signing of the adult community. That is, we match the sounds to the scenes so as to pair the forms with their meanings. Owing to the loose and variable relations between word use and the passing scene – the “stimulus-free property of language use,” as Chomsky (1959c)

famously termed this – knowledge of these form/meaning relations necessarily accrues piecemeal over time and probabilistically over repeated exposures. But in the end (or so the story goes), *horse* tends to occur in the presence of horses, and *race* in the presence of racing, and these associations eventually get stamped in. Just so. (I will return presently to mention at least a few of the questions I am begging by so saying.)

Now here is a potentially hard question. Equating for frequency of utterance in caretaker speech, and presupposing the word-to-world pairing procedure just alluded to, some words are easier to acquire than others as indexed by the fact that they show up in the earliest vocabularies of infants all over the world. One general property of these novice vocabularies illustrates this point: The first-learned 100 or so words are – animal noises and ‘bye-bye’s excluded – mainly terms that refer in the adult language to whole objects and object kinds, mainly at some middling or “basic” level of conceptual categorization (Caselli et al. 1995; Gentner and Boroditsky 2001; Goldin-Meadow et al. 1976; Lenneberg 1967; Markman 1994; Snedeker and Li 2000). This is consistent with many demonstrations of responsiveness to objects and object types in the prelinguistic stages of infant life (Kellman and Spelke 1983; Needham and Baillargeon 2000).

In contrast, for relational terms the facts about understanding concepts do not seem to translate as straightforwardly into facts about early vocabulary. Again there are many compelling studies of prelinguistic infants’ discrimination of and attention to several kinds of relations including containment versus support (Hespos and Baillargeon 2001), force and causation (Leslie and Keeble 1987), and even accidental versus intentional acts (Woodward 1998). Yet when the time comes to talk, there is a striking paucity of relational and property terms compared to their incidence in caretaker speech. Infants tend to understand and talk about objects first. Therefore, because of the universal linguistic tendency for object concepts to surface as nouns (Pinker 1984; Baker 2001), nouns heavily overpopulate the infant vocabulary as compared to verbs and adjectives, which characteristically express events, states, properties, and relations. The magnitude of this noun advantage from language to language is influenced by many factors, including ratio of noun to verb usage in the caregiver input (itself the result of the degree to which argument dropping is licensed), but even so it is evident in child speech to a greater or lesser degree in all languages that have been studied in this regard (Gentner and Boroditsky 2001). In sum, verbs as a class are “hard words” while nouns are comparatively “easy.” Why is this so?

An important clue is that the facts as just presented are wildly oversimplified. Infants generally acquire the word *kiss* (the verb) before *idea* (the noun) and

even before *kiss* (the noun). As for the verbs, their developmental timing of appearance is variable too, with words like *think* and *know* typically acquired later than verbs like *go* and *hit*. Something akin to “concreteness,” rather than lexical class *per se*, appears to be the underlying predictor of early lexical acquisition (Gillette et al. 1999). Plausibly enough, this early advantage of concrete terms over more abstract ones has usually been taken to reflect the changing character of the child’s conceptual life, whether attained by maturation or learning. Smiley and Huttenlocher (1995: 20) present this view as follows:

Even a very few uses may enable the child to learn words if a particular concept is accessible. Conversely, even highly frequent and salient words may not be learned if the child is not yet capable of forming the concepts they encode . . . cases in which effects of input frequency and salience are weak suggest that conceptual development exerts strong enabling or limiting effects, respectively, on which words are acquired.

A quite different explanation for the changing character of child vocabularies, the so-called syntactic bootstrapping solution (Landau and Gleitman 1985; Gleitman 1990; Fisher 1996; Gleitman et al. 2005; Trueswell and Gleitman 2007), has to do with information change rather than conceptual change. The nature of the vocabulary at different developmental moments is taken to be the outcome of an incremental multi-cue learning procedure instead of being a reflection of changes in the mentality of the learner:

- (1) Several sources of evidence contribute to solving the mapping problem for the lexicon.
- (2) These sources vary in their informativeness over the lexicon as a whole.
- (3) Only one such source is in place when word learning begins: namely, observation of the word’s situational contingencies.
- (4) Other systematic sources of evidence have to be built up by the learner through accumulating linguistic experience.
- (5) As the learner advances in knowledge of the language, these multiple sources of evidence are used conjointly to converge on the meanings of new words. These procedures mitigate and sometimes reverse the distinction between “easy” and “hard” words.
- (6) The outcome is a knowledge representation in which detailed syntactic and semantic information is linked at the level of the lexicon.

According to this hypothesis, then, not all words are acquired in the same way. As learning begins, the infant has the conceptual and pragmatic wherewithal to interpret the reference world that accompanies caretaker speech, including the gist of caretaker–child conversations (to some unknown degree,

but see Bloom 2002 for an optimistic picture, which we accept). Words whose reference can be gleaned from extralinguistic context are “easy” in the sense we have in mind; that is the implication of point (3) above. By and large, these items constitute a stock of concrete nominals. Knowledge of such items, and the ability to represent the sequence in which they appear in speech, provides a first basis for constructing the rudiments of the language-specific clause-level syntax of the exposure language; that is, its canonical placement of nominal arguments and inflectional markings. This improved linguistic representation becomes available as an additional source of evidence for acquiring further words – those that cannot efficiently be acquired by observation operating as a stand-alone procedure. The primitive observation-only procedure that comprises the first stage of vocabulary growth is what preserves this model from the vicious circularity implied by the whimsical term “bootstrapping” (you can’t pull yourself up by your bootstraps if you’re standing in the boots), and is very much in the spirit of Pinker’s (1984) “semantic bootstrapping” proposal, with the crucial difference that by and large the initial procedure yields almost solely concrete noun learning. Structure-aided learning (“syntactic bootstrapping”), required for efficient acquisition of the verbs and adjectives, builds upward by piggybacking on these first linguistic representations. An important implication of the general approach is that word learning is subject to the same general principles over a lifetime (for laboratory demonstrations, see Gillette et al. 1999; Snedeker and Gleitman 2004). For the same reasons, these principles should and apparently do apply to vocabulary acquisition in later-learned as well as first languages (Snedeker et al. 2007).

For the rest of this paper, I’ll illustrate the explanatory power of this machinery for two kinds of case that pose principled problems for the idea that word-to-world pairing (though no doubt a necessary factor) is sufficient by itself as the information basis for vocabulary acquisition. The first case involves such perspective verb pairs as *give/get*, *chase/flee*, *buy/sell*, and the like, illustrated in Fig. 16.1. It depicts an action scene in which a dog is *chasing* a man. But literally by the same token it depicts a man who is *fleeing* (from) a dog. Suppose the adult utters a new verb – “Look! Pilking!” – in reference to such a scene. Is he or she speaking of chasing or of fleeing? Assuming that just these two interpretations come to mind, among the many that are really available and pertinent to the event, how is the listener to decide between them? At peril of belaboring the point, the next few hundreds of exposures to pursuit scenes are liable to embody the same ambiguity. Rarely do members of such pairs surface under real-world circumstances that differentiate between them. Which returns me to the problem that we are generally begging the questions at issue when we say that word-to-world pairing solves even the simplest cases of word learning –

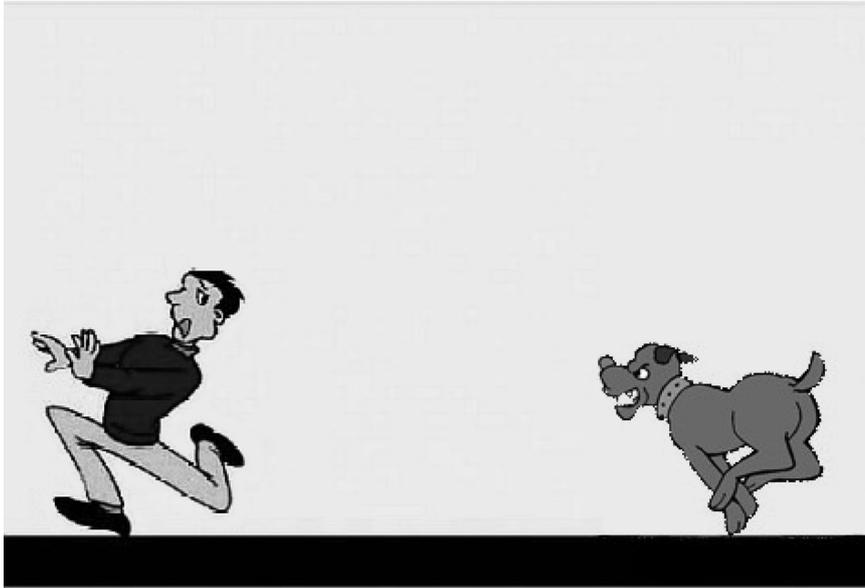


Fig. 16.1. Dual conceptions: chasing and fleeing.

that people acquire word meanings “from” observing the world. The difficulty from the outset is that, for word learning to occur, one has to conceive of the observed world *in the right way*, under the description that fits the word that is being used. But this requirement completes the circle.

To escape from this circularity there has to be a way for the learner to focus (“zoom in” is our own favored technical term) on the right description (representation) of the scene without presupposing knowledge of the word whose acquisition is at issue. How could attention be focused on just one of these interpretations in the case of perspective verbs? For comparison, first consider another famous ambiguity, the duck-rabbit in Fig. 16.2. Perception psychologists

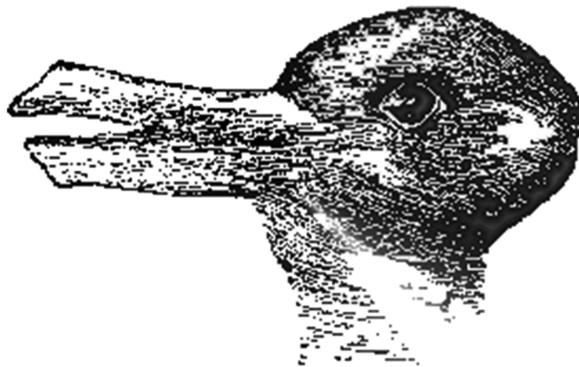


Fig. 16.2. Dual perceptions: duck and rabbit.

Georgiades and Harris (1997) showed that the chances of a naïve observer reporting seeing the duck versus the rabbit can be influenced by a subliminal visual attention-capture cue judiciously placed on such a figure. Perhaps more surprisingly, the same is true of chasing and fleeing depictions and other cases interpretable as one of two paired perspective verbs, including *give* vs. *get*, *win* vs. *lose*, and *buy* vs. *sell* (Gleitman et al., in press). Following Georgiades and Harris, we captured our subjects' visual attention on such pictures by briefly (60–80 msec) flashing a square on the computer screen just prior to the onset of the picture. This square was aligned with the upcoming position of one or the other character. Typically this caused eye movements to that character, even though the subjects were not able to report noticing the flashed square. Fig. 16.3 exemplifies the procedure using the intended contrast *win/lose*. This manipulation reliably influenced the speaker's tendency in describing the scene. For the *chase/flee* case, the tendency to describe the scene as one of chasing was enhanced when the flash was where the dog subsequently appeared, and as one of fleeing if it was on the man. So how the speaker “attentionally approaches” an event like this does seem to affect its description and, consequently, verb choice.

It remains to ask how speaker choice might be related to the learning situation for such cases. We know from the work of Dare Baldwin (1991) that infants will attend to the direction of the speaker's gaze as a cue to the reference of a new noun. In preliminary studies we have shown adult subjects a version of these verbs in which a cartoon character (“John”) is looking down on the scene. “John's” eyes are directed either to the chaser or the flier, as shown in Fig. 16.4; and again this influences the subject's report of what she thinks John would say, to describe the scene (Gleitman et al., in press). So here we have a hint that social-attentive cues from the speaker might direct the listener-learner toward a

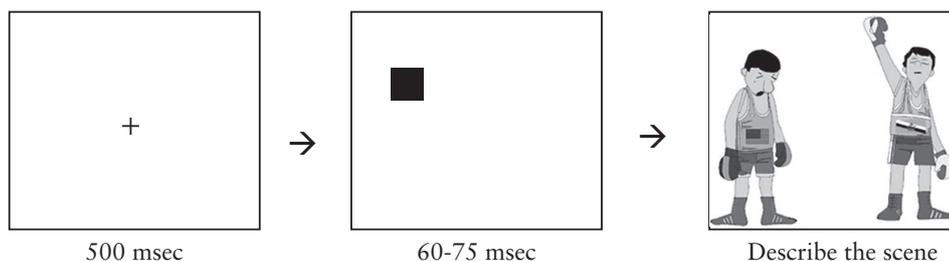


Fig. 16.3. A subliminal attention manipulation: After visual fixation (panel 1), a block is briefly flashed, situated where (on different trials) the winner, the loser, or a place in between, will subsequently appear (panel 2). The picture then appears and the subject describes what is seen (panel 3).

Source: Gleitman et al. 2007

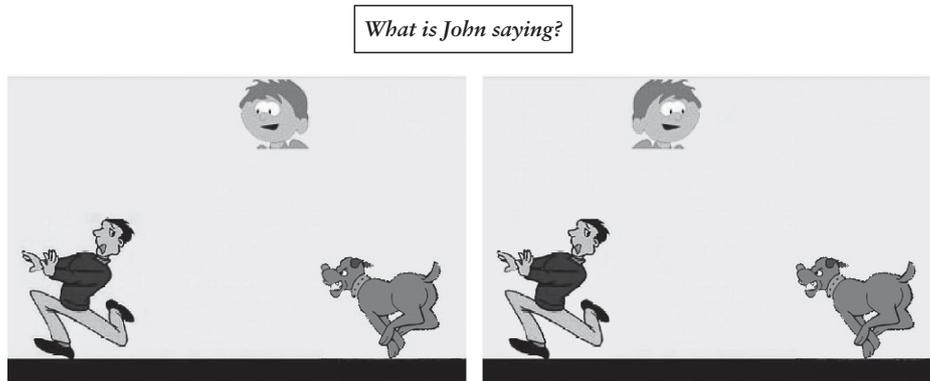


Fig. 16.4. Visual cueing of chase and flee: John's gaze direction influences subjects' utterance of chase (left image) or flee/run away (right image).

particular choice of interpretations even in these cases where on the surface the scene itself seems to provide no basis for disambiguation.

The effects of speaker-gaze direction on disambiguation of these pairs are by no means categorical even in this laboratory situation, and even with adult subjects. So I turn now to another attentional cue, evidently a more powerful one. In this experiment (Fisher et al. 1994) we showed 3- and 4-year-old children (and adult controls) videotaped puppet shows designed to exemplify perspective verbs, and we introduced an extra hand-held puppet, telling the children that it was a Martian puppet that talks Martian talk. We asked them if they could help us figure out what Martian words this puppet was saying. One third of the children heard the Martian (who, in company with the child subjects, was viewing the puppet show) say *Look, gorpington!*; the next third of the subjects heard *Look, the skunk is gorpington the rabbit!*; and the final group heard *Look, the rabbit is gorpington the skunk!* The results are rendered in Fig. 16.5, collapsing across several of the scenes that the children saw and responded to. Notice first that there is a cognitive bias in all these results toward source-to-goal interpretations. This shows up strongly for both the children and adults in the no-sentence (*Look! Gorpington!*) condition which does not linguistically bias the subject. For instance, *give* is heavily preferred to *get*, *chase* is preferred to *flee*, and so forth. For the subjects who heard instead *The skunk is gorpington the rabbit*, this effect is enhanced – it becomes essentially categorical because the form of the sentence supports the cognitive bias. But for those subjects who heard *The rabbit is gorpington the skunk*, the results reverse. The adults shift completely to the goal-to-source verb (*flee* or *run away*) dispreferred by the prior subjects. You still see the residue of the cognitive bias with the children, but the modal response has for them too now shifted

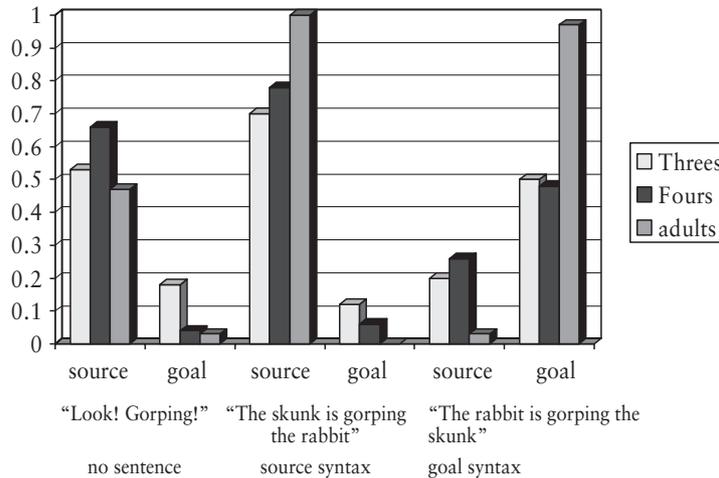


Fig. 16.5. Source versus goal by syntactic introducing context: The source is the preferred subject (e.g., the giver or chaser is preferred to the getter or evader) if the syntax is neutral. This tendency is enhanced or diminished in both adults and young children as a function of syntactic information.

Source: (Fisher et al. 1994)

to the goal-to-source interpretations.<sup>1</sup> This pattern would be expected if the structural configuration chosen by a speaker is understood by the listener to reflect the speaker's attentional stance. Research on discourse coherence strongly suggests that subject position is often used to denote the current discourse center and to mark transitions from one center to another (e.g., Gordon et al. 1993; Walker et al. 1998). This is why Fisher et al. (1994) described their effect as a "syntactic zoom lens" in which the structural configuration of the words in the utterance helps the child take the perspective necessary to achieve successful communication, and to infer the meaning of unknown elements in an utterance.

I want to emphasize a couple of points in wrapping up this part of the discussion. First was the idea that the word-to-world pairing procedure that is in place from earliest infancy is effective primarily for whole-object terms (Markman 1994; Gillette et al. 1999), accounting for the noun-dominated character of the novice vocabulary. My next ambition in this paper was to show how linguistic structure itself acts to redress these limitations once the novice (whether an infant or older language learner) has acquired its rudiments by considering the sequence of nouns against their contexts of use. To expose the problem and elements of the solution, I showed you how children and adults

<sup>1</sup> The proportions in the figure don't add up to 100% in any condition just because of the indeterminacy of what's said, given a situation. Thus children and even adults sometimes respond "They're having fun!" rather than "He's chasing him," in response to some of these scenes.

discover the interpretation of novel terms – here, the perspective verbs – whose reference is just about always ambiguous and which therefore cannot be wholly explained as observation-based learning. Because the solution to this problem must be (somehow) to draw the observer’s attention toward one of the two primary interpretations, it is reassuring that attentional cues of varying kinds, including subliminal flashes but also eye-gaze direction of a cartoon figure, materially influenced these interpretations in the laboratory. Perhaps more surprising, especially in its influence on young preschoolers, is that the strongest cue of all was implicit and linguistic. They interpreted the scene in accord with the semantic information latent in the structure of the introducing sentence, specifically, according to which character captured the subject position.

It remains to say that no one of these cues (situation or syntax) can be sufficient. Obviously the subjects couldn’t have learned (and therefore didn’t) the meaning of “gorping” solely by hearing it used in an appropriately structured sentence, any more than they could have disambiguated, say, *chase* from *flee* solely by observing the puppet shows. What does the trick for learning is the two cues working conjointly. The argument structure is revealed by the syntax, to be sure, but simultaneously the sentence is interpreted against the world to which it refers. This use of multiple cues lies at the heart of the syntactic bootstrapping procedure. With acquisition of the language-specific grammar, the learner is able to bring to bear a linguistic representation that matches in sophistication, and dovetails with, his or her natural ability to impose a predicate–argument interpretation on events. Given this narrowing of the hypothesis space to fit the argument-structure framework, the observed world more efficiently fills in the richer semantic content of the novel predicate.

I mentioned back at the beginning of this paper that I was going to motivate the syntactic bootstrapping approach in terms of two kinds of lexical item that pose a principled difficulty for lexical learning models that rely solely, or even very heavily, on word-to-world pairing. The first were these perspective verb pairs. Now I want to turn to the second case, which looks even harder. This is acquisition of verbs that describe unobservable acts and events, such as *think* and *believe*. Here the world is of very little value, or so it seems at first glance. You can’t see thinking. And the literature tells us that these items indeed appear relatively late in the infant’s verb-learning career. Even though children produce verbs describing actions or physical motion very early, often before the second birthday (Bloom et al. 1975), and appear to understand them well, they do not use mental verbs as such until about two and a half years of age (Bretherton and Beeghly 1982; Shatz et al. 1983) and do not fully distinguish them from one another in comprehension until around age 4. These facts are often adduced as rather straightforward indices of concept attainment (e.g., Dromi 1987;

Huttenlocher et al. 1983), put forward to support the view that conceptual change is what's accounting for the trajectory and contents of early vocabularies. In particular, the late learning of credal ("belief") terms is taken as evidence that the child doesn't have control of the relevant concepts, in this case the ability to entertain concepts that refer to one's own or others' mind, aka Theory of Mind. As Gopnik and Meltzoff (1997: 121) put this:

the emergence of belief words like "know" and "think" during the fourth year of life, after "see", is well established. In this case... changes in the children's spontaneous extensions of these terms parallel changes in their predictions and explanations. The developing theory of mind is apparent both in semantic change and in conceptual change.

And in this case too I'm going to try to convince you that there is another potential explanation for why these terms are late acquired, short of saying that they are too "abstract" for young ears and minds. Specifically, I suggest that the child's problem isn't the inability to think about thinking, but only to find the evidence that the sound/word *think* is the item that expresses the concept 'think' in English: a mapping problem rather than a conceptual problem. It simply is harder to glean, by observation alone, that thinkers are thinking than that, say, jumpers are jumping. Not only is thinking invisible in the first place. Even more important, the difficulty is that it is actions that people, young and old, are inclined to think about when they interpret the world, rather than the thoughts of those performing the actions (alas, perhaps, but true nonetheless).

Now here is a parade case to introduce this topic: When one shows Rodin's famous statue, The Thinker, even to museum-knowledgeable adults and asks "What's going on here?" the respondents are disinclined to say "That's a thinker thinking." Even though, if *anything* is, this *is* a thinker thinking. They are inclined to respond instead: "He's resting his head," or "He's scratching his chin," in short to offer just about any overt act in preference to an internal one, in describing this scene – though I grant that Rodin himself was an exception to this generalization. In short this is a case of massive insalience of a concept. Nobody thinks about thinking even though it's always going on when people are around. Even here in this room, most of you are emphatically thinking, but thinking is not what you're thinking *about*. If children are to learn the word *think*, there must be circumstances in which the concept it encodes comes readily to mind.

I'll now describe just one experiment in a line we have been pursuing, focusing on this vexing class of words (Papafragou et al. 2007; and for a theoretical review, Gleitman et al. 2005). The idea again is to assess the contribution both of syntactic cues and cues from observation. Pilot findings had provided us with the intuition that for the case of mental verbs, *it's not the truth that sets you free*. Instead, people think about thinking under circumstan-

## THE LEARNED COMPONENT OF LANGUAGE LEARNING 249

1. Lo PILK what fimmet wifs.
2. Well, bo PILK what mippy rucky zavvy smegs  
are so far, don't bo?
3. Po PILK lo pung mo.
4. Lo PILK what lo can wif with ti?
5. Do lo PILK where the kax's lif is?
6. Do lo PILK where a fimmit is in mippy runk?

Fig. 16.6. What does PILK mean? The range of syntactic environments is revealing of the verb interpretation.

Source: Snedeker and Gleitman 2004.

ces where someone is in a state of *false* belief. Moreover, just as was the case for the perspective verbs, there are characteristic structural environments in which such verbs leap immediately to subjects' minds. Fig. 16.6 shows an example from a study by Snedeker and Gleitman (2004). It is constructed from a random sample of mothers' natural usages of a credal verb in sentences uttered to their 18–24-month-old children, but the experimental version of these that you see here is doctored and disguised. We leave enough of the closed-class material in place so that the subject can recover the structure spontaneously, that is, without explicit instruction from us. All the other words are replaced by nonsense words. The “mystery word” (the verb) is also replaced by nonsense (in caps) and the subject's task is to recover and report its meaning, given these half-dozen *Jabberwocky*-like exemplars. People are very good at this task, evidently using the appearance of sentential complements as a giveaway clue for a credal verb interpretation.

In the Papafragou et al. (2007) study, 4-year-old children and adults watched a series of videotaped stories with a pre-recorded narrative. At the end of each clip, one of the story characters described what happened in the scene with a sentence in which the verb was replaced by a nonsense word. The participants' task was to identify the meaning of this mystery word. The stories fully crossed type of situation (true vs. false belief) with syntactic frame (transitive frame with direct object vs. clausal that-complement) as shown in the design diagram (Fig. 16.7). For instance, in one of the false-belief stories inspired by the adventures of Little Red Riding Hood, a boy named Matt brought food to his grandmother (who in reality was a big bad cat in disguise); in the true-belief variant of the story, Matt accompanied by the big cat brought food to his real grandmother. At the end of the story, the cat offered one of these two statements:

- (a) [Complement Clause condition] “Did you see that? Matt GORPS that his grandmother is under the covers!”
- (b) [Transitive condition] “Did you see that? Matt GORPS a basket of food!”

	NP comp	S-comp
True belief		
False belief		

Fig. 16.7. Scene type X syntax type: This illustrates the design of an experiment in which the child hears a story in which a true or a false belief figures prominently, crossed by a verbal description in the form of a transitive construction (e.g., “The boy is eating his snack”) or in a sentence-complement construction (e.g., “The boy thinks that this is his snack”).

Source: Papafragou et al. 2007

It was hypothesized that false-belief situations would increase the salience of belief states and acts and would make these more probable topics for conversation, thereby promoting mentalistic conjectures for the novel verb. It was also hypothesized that sentential complements would prompt mentalistic interpretations for the target verb. It was expected that situations where both types of cues cooperate (i.e., in the false belief scenes with a sentential complement) would be particularly supportive of mentalistic guesses. Finally, syntactic cues were expected to overwhelm observational biases when the two conflicted (e.g., in false-belief scenes with a transitive frame).

These predictions were borne out. Scene type had a major effect on the verb guesses produced by both children and adults. Specifically, false-belief scenes increased the percentage of belief verbs guessed by the experimental subjects, compared to true-belief scenes (from 7.4% to 26.5% in children’s responses and from 23.5% to 46.3% in adults’ responses). The effects of syntax were even more striking. Transitive frames almost never occurred with belief verbs, while complement clauses strongly prompted belief verbs (27.2% and 66.2% of all responses in children and adults, respectively). When both types of supportive cue were present (i.e., in false-belief scenes with complement clause syntax), nearly half (41.2%) of children’s responses and an overwhelming majority (85.5%) of adults’ responses were belief verbs.

Similar effects were obtained in a further experiment with adults, which assessed “pure” effects of syntactic environment (minus supporting content words) in the identification of mental verbs. True- and false-belief scenes were paired up with transitive or complement clause structures from which all content words had been removed and replaced with nonsense words (e.g. *He glorps the fleep* vs. *He glorps that the fleep is glexing*). Again syntax proved

a more reliable cue than even the most suggestive extra-linguistic contexts. Furthermore, the combination of clausal and scene (false belief) information again resulted in an overwhelming proportion of mental-verb guesses.

Taken together, these experiments demonstrate that the syntactic type of a verb's argument (e.g., whether the object of a transitive verb is a noun phrase or a tensed sentence complement) helps word learners narrow down their hypotheses about the possible meaning of a new word. Furthermore, this type of syntactic cue interacts overadditively with cues from the extralinguistic environment (e.g., the salience of a mental state). We interpret these findings to support the presence of a learning procedure with three crucial properties: (1) it is sensitive to different types of information in hypothesizing the meaning of novel words; (2) it is especially responsive to the presence of multiple conspiring cues; (3) it especially weights the language-internal cues when faced with unreliable extralinguistic cues to the meaning of the verb.

To summarize some of the effects I've been discussing, the first general finding is that not all words are learned from the same kind of information. Certain items, including words encoding the basic-level object terms, appear early. This is one of the most robust effects in the literature of language learning, and is seen again and again cross-culturally and cross-linguistically. A popular explanation for why these items are so rapidly and uniformly learned is that they instantiate just about the only concepts that infant minds can entertain. But I have argued instead that it is these words' tractability to the first-available property of the learning procedure, word-to-world pairing, that explains why they are learned first. As support for this view, we have shown in several experiments that when adults are by experimental artifice reduced to this same information – roughly, if they are exposed to single “mystery words” in context, rather than to whole sentences in context – they too are capable of little lexical learning beyond the basic-level nominals. The information for acquiring the noun *apple* and such physical-action verbs as *jump* or *hit* resides largely in the observable world, as interpreted by both adults and very young children.

In contrast, words that describe unobservable mental states and acts are cued almost exclusively by information that resides in the semantics of syntactic structures (see Fig. 16.8, from a verb-identification task with adult subjects, which shows this effect). These adults identify action verbs largely by examining the scenes in which these are uttered, but they identify mental verbs largely from hearing nonsense-containing structures in which these occur (Snedeker and Gleitman 2004). Because children acquire the requisite (language-particular) aspects of the grammar only during the second and third years of life, they are limited until then in their word learning largely to

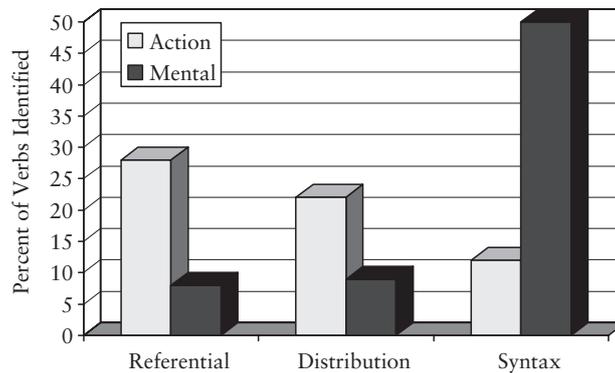


Fig. 16.8. Different verbs require different kinds of information to acquire: Referential information (the visual-situational context) provides the lion's share of information for identifying action verbs such as *jump* or *put*, but syntactic information is far more informative for mental verbs such as *think*, *see*, and *want*.

Source: Snedeker and Gleitman 2004

lexical items whose meaning can be wrested more or less directly from transactions with the referential world.

More generally, my colleagues and I have tried to explain word learning as a mapping process, one which matches sounds to their meanings. To be sure, the mapping procedure is a complex one, requiring the recruitment and integration of several kinds of linguistic and extralinguistic information. Word learners, in the special case where they are young children, may also be undergoing significant conceptual change. Even if so, these changes in mentality do not seem to be the chief limiting factors in vocabulary growth.

### *Discussion*

PARTICIPANT: I have two questions. I think there is an important difference between saying that you need a particular structural context, a sentential complement, to solve the mapping problem for propositional attitude verbs, and saying that you need particular kinds of structural arrangements to acquire the concepts. So there are two problems: first, to solve the mapping problem for propositional attitude verbs, and for this you need a particular structural context (syntax, you said, is needed). So that is one problem. The other problem is to ask to what extent you need sentential complements – a certain structural context – to have propositional attitude concepts in the first place. To what extent is the structure actually instrumental for having the concepts in the first place? I of course would go for the latter alternative, and I was wondering about your view on that. Related to that, if you go against the conceptual change view

of Gopnik (Gopnik and Meltzoff 1997) and Carey,<sup>2</sup> so you posit belief-type verbs in the biology (the evolution), it is obviously just pushing the problem. It's not solving any problem, I would say. So here my question is: how little Platonism do you get away with?

GLEITMAN: You are correct that there are two problems here. One has to do with where the concept *think* (or any other) comes from, how these ideas get into the mind. The second has to do with identifying the word in the exposure language that encodes each such idea; for instance, learning that *think* is pronounced /think/ in English. You, along with many others, find it congenial to suppose that hearing the sound /think/ (in some sentential context) is what – or part of what – causes the concept to grow in the mind. I myself find that position hard to understand, it seems to imbue words with some magical property. But we can't argue from what is a congenial or intuitively plausible approach on these matters, at least we won't get far that way. So, congeniality aside, what I tried to do in my talk is to show you some evidence to the effect that infant and adult word learning look very much alike. This suggests that both populations are solving the same problem, namely the mapping problem (which sound encodes the concept *think*) rather than one population solving this problem (the adults) while the other (the children) is solving two problems at once – the mapping problem and the concept acquisition problem. I tried to show you that when by experimental artifice one reduces the information that the adult has – his or her evidentiary sources for word learning – the learning trajectory and contents for child and adult look much alike. By exhibiting such laboratory effects, I invited you to consider whether information availability rather than concept availability might not hold the major key in explaining word learning.

No one doubts that there are conceptual-sophistication differences between, say, an average 3-year-old and Noam Chomsky or even the college sophomores to whom we teach *gorping* and *pilking* in the laboratory. It is the sameness in learning properties, once the task is equated for information, across these individuals and populations, that suggests that the mapping problem rather than the concept-learning problem is the chief limiting factor in word learning. But what I most wanted to show you is that observation of “the world” is insufficient as the input basis for acquiring the word /think/ – for anyone, child or adult. Even to use the situation as a constraint, one needs to narrow the search space by being told the argument-taking properties of the novel predicate. That is what the syntax does for you, and it does so for 2- and 3-year olds as well.

---

<sup>2</sup> Carey (1985, 2001).

PARTICIPANT: I'm not a linguist and I really want to comment on the question of language acquisition. From an interdisciplinary approach, I wanted to offer a possible alternative way of thinking about it. When for example a parent gives a child a stuffed animal, and the parent utters the sound *elephant*, the child has an experience of the joy of the moment, of possibly understanding that they are getting something and it's a toy and it's fun. Later, the parent sees an elephant on television and utters the same sound. So at this point, the child has to negotiate for a distinction. Now in another theory you look for distinctions between phenomena, but you also want to find the categories of representation. In the first case, the stuffed animal resembles an elephant – to the parent. To the child, those distinctions don't yet exist, so it could be a cat, it could be a puppy, it's a stuffed animal to the child, whatever that means to the child. The television representation actually points to an elephant in the world somewhere. So there you have this index to something in the world. Then you have a third scenario: the parent takes the child to the zoo and suddenly the child hears this same utterance while experiencing this huge object in front of him, the actual elephant in nature. It is at this point that I believe Peirce used the term *abduction*. The child is confronted with a sign, the sound *elephant*, which is used in three different contexts as a reference to an object in the world, and the child then has to negotiate the initial meaning of the sound associated with this stuffed animal, with the TV image, and now this massive object in nature. So this is where this abductive reasoning is a partial explanation of what I believe Peirce meant by abduction. This is a partial explanation of abduction where the child then has to negotiate the semantics.

GLEITMAN: Your suggested solution is a very sensible one. Your idea is to redress the insufficiencies of any one situational observation by comparing across many such observations, parsing out of scenes in which, say, /elephant/ is uttered, that which is common to all these otherwise quite variable scenarios. This cross-situational observation solution has commended itself to everyone from John Locke and David Hume to modern connectionist modelers. And as I mentioned, surely such a procedure *must* play a role, your various elephant-scenarios are probably a good sample of how this goes. Yet among the many problems of trying to do the whole job of word learning using this situation-observing procedure are the ones I concentrated on in my talk – you can't easily tell *chase* from *flee* this way because they map onto about the same scenes, and it is hard to "observe" thinking in any literal or straightforward way, no matter how many thinking scenes/utterances you are exposed to.

But there is a greater problem and that is the infeasibility of your suggested model given the rate and relative errorlessness of actual word learning. The

child is learning about ten words a day. This is a very, very large number. In light of it, there doesn't seem to be enough time and varied, yet systematic, scene-observations for such a model to work, *unaided*. In fact there's considerable evidence that children are correctly inducing the meanings of words from one or a very few instances, rather than pursuing a compare/contrast procedure across many observations. And this "fast mapping" of new words goes on for a long time, probably until you're about 30–35 years of age, so you get a vocabulary of maybe 75,000 words. Though then, as we elders can tell you, it plummets [laughter]. Luckily Noam and I started with a big vocabulary [laughter].

But seriously: the speed and accuracy and persistence of word learning is something which I think influences how plausible various models should look to you. Another feature of acquisition that might influence you in this regard is the sameness of word meanings acquired by learners whose observational circumstances are wildly different, for instance, deaf, blind, and even deaf-blind persons. I and my many colleagues have offered a different solution. Though of course it involves information gleaned from word-to-world correspondences, it is not limited to this evidentiary source, at least not after the child is 18 or 24 months old and has gained some principled linguistic (as well as world) experience. What this model substitutes for sole use of a multitude of cross-cutting situational observations is a small set of exposures to a novel word, but with most such exposures *simultaneously offering evidence of different kinds*. Observations of a word's fit with the passing scene, yes, but also observations of its structural environment, its morphology, and its co-occurrence with other words (e.g., *cake* occurs more often with *bake* than with *wake*). These cues trade and conspire to overdetermine interpretation based on very small numbers of incidents during which a novel word is heard.