6 Universal aspects of word learning

Lila Gleitman and Cynthia Fisher

> Suppose instead of saying *Dutch* we had said *Clashes with the wallpaper*; *I thought you liked abstract work*, *Never saw it before*, *Tilted*, *Hanging too low*, *Beautiful*, *Hideous*, *Remember our camping trip last summer*?
>
> (Chomsky 1959; Review of B. F. Skinner's *Verbal Learning*)

Most of the action in linguistic theory under Chomsky's aegis has focused on questions of how words are put together into sentences rather than on the words themselves. Fair enough: the universal hallmark of human language is its parametrically organized combinatorial structure. In linguistic systems, the individual word classes (for instance, the nouns, the verbs, the complementizers) play particular and crucial structural roles; rules and parameters directly implicate these classes. But what about the items that comprise these classes? It has been widely maintained that these individual atoms or particles of language play no central role in its core, or, to use recent terminology, in "language narrowly conceived" (Hauser, Chomsky & Fitch 2002). No distributional property of English worth its salt is dependent on whether some particular noun – let us say *elephant* rather than *rhinoceros* or *gnu* – appears in it.[1] Rather, what words seem to be relevant to is the conceptual system that interfaces with language. In this latter domain, relevant to "language broadly conceived," the difference between an elephant and a gnu really matters.

Similarly, the linguistic-theoretical study of language acquisition has focused primarily on the nature and setting of syntactic parameters rather than on the evolving character of the lexicon (e.g. Manzini & Wexler 1987; Lightfoot 1999; Baker 2001). Again, this is no surprise. Acquisition of the combinatorial features of a language poses a classical poverty-of-the-stimulus problem, requiring the learner to extract abstract organizing principles from input structured only (or almost only) as morpheme sequences (see, particularly, Chomsky 1986). In contrast, a first intuition is that acquiring words and their meanings can be fully accounted for by a procedure that associates the forms (say, the sound "elephant") with their meanings (here, the concept ELEPHANT) in consequence of observing the referential contingencies for the word's use (say, visible presence of an elephant). Here is this claim, as put by John Locke (1690):

(1) If we will observe how children learn languages, we shall find that . . . people ordinarily show them the thing whereof they would have them have the idea, and then repeat to them the name that stands for it, as "white," "sweet," "milk," "sugar," "cat," "dog," (Book 3, IX, 9)

In this chapter we will revisit lexical learning in aspects that seem to be relevant to Chomsky's program for understanding language and its users. We will focus on four broad issues. First, we show that the method of word-to-world pairing advocated by Locke in (1) is too weak, taken alone, to account for the robustness of word learning. Second, we discuss further data sources – *linguistic* data sources – that in principle can enrich and constrain this acquisition procedure. Third, we describe what is known of the circumstances in which children recruit and exploit these various data sources. Fourth and finally, we allude to the accumulating evidence that the word-learning procedure leaves its footprint in the mature mental representation of language: grammars are heavily lexicalized, in part because the learning procedure for words has – necessarily – built complex (syntactically organized) lexical structures.

Before beginning, we want to make clear the sense of "word learning" that we will discuss. We restrict ourselves to the *mapping problem*: how one comes to know, if exposed to French, that *chien* is the phonological form that expresses the concept DOG while *voir* expresses TO SEE; whereas if exposed to Italian, that it is *cane* and *vedere*. We leave aside altogether the question of where the concepts themselves come from, stipulating only that, to a useful approximation, these concepts are in place to support the word-learning procedure. That is, acquiring the meaning for *dog* and *see* requires that the learner antecedently be able to entertain these concepts[2] (Fodor 1983). Our question – the mapping problem for vocabulary acquisition – is how the child decides which sound goes with which meaning.

The robustness of learning to input variation

Locke's dictum (1) is that the correlation between word use and the specifics of the reference world is both necessary and sufficient to account for word learning. It follows from this that words could not be acquired (the mapping problem could not be solved) except under conditions in which words were uttered in the presence of their referents. We ask in this section whether this precondition is met in children's learning environment; and, if it is, whether this precondition is sufficient as well as necessary.

The contingencies for a word's use

Perhaps the central difficulty with Locke's approach was noted by Chomsky (1959), who challenged empiricist speculation on just this point; hence the

epigraph to the present chapter. The trouble is that one does not always or even particularly often say "Dutch" on viewing a Rembrandt painting. Conversation, even from mothers to babies, is not a running commentary on the objects, events, properties, and relations presently on exhibit in the world. No mother carefully utters "open" every time she opens the door; worse, "open" is frequently uttered – even systematically so – when the door is shut ("Help! I'm locked in this bathroom! Please someone come and open the door!"). In the next main section of this chapter we will document this "stimulus-free" property of language use. The problem is not that utterances have no relevance to matters at hand. It is rather that the concept *relevance* is so broad that it places very little constraint on what is likely to be said, given some circumstance. We should note that the issues here look even more ominous from the perspective of Quine (1960), who famously noticed that indefinitely many construals can be put on any observation of the world (say, a rabbit-observation). Chomsky's objections seem more to the point, as he grants constraints on interpretation deriving from human conceptual structure and from conversational relevance; and yet *still* the word-to-world contingencies appear to be so flexible as to render learning by observation intractable.

Cross-situational observation

The standard response to the problem just discussed is that the child doesn't have to learn from a single observation. Rather, over successive observations, probabilistic relations between word and world will converge to support word learning. After all, in the end *elephant* is uttered more in elephant situations than in zebra situations (Pinker 1984). One potential pitfall for a learning device reliant on these correlations is the ornate differentiating series of observations that would be required if it is not to go hopelessly wrong, a difficulty that Locke seems to acknowledge by his very choice of examples: in (1) whiteness is an attribute of both the sugar and the milk, and sweetness of both the sugar and (if one is lucky) one's mother. An even more difficult problem is that such a procedure does not seem to comport well with the actual learning facts. For one thing, a statistical learning procedure that maps between word sounds and events must be errorful over its course, involving significant backtracking and revision, and so predicts some proportion of howlers along the way. Children should sometimes mix up the milk and the sugar with their mothers, at least in speech acts. Moreover, vocabulary acquisition should overall be a slow process, since an accumulation of observations is required to warrant conclusions about each item (and the more variable the word-to-world circumstances, the slower should be the learning). Yet learning proceeds at rates up to ten new items per day, almost completely without the predicted howlers. Children seem to be drawing the right conclusions about the meanings and referents of words

based on one or a very few exposures. In Carey's (1978) terms, they are "fast mappers."

Populations deprived of information; colorless green ideas

Just as no child receives all the sentences of English as the condition for learning its grammar, so no child observes all the potential referents as a condition for acquiring word meanings. Totally disjoint dog observations underlie individuals' acquisition of the word *dog*. You see Fido and Spot, I see Rex and Ginger, yet for each of us a generalization results such that we *consensually* partition the whole world into the dogs and the non-dogs. The British Empiricists were among the first to ask just how far observational environments could diverge, to support the learning of the same word. They suggested that relevant test beds for the learning by observation hypothesis can come from populations who are systematically deprived of certain opportunities to observe the world. Here is this suggestion, as offered by David Hume:

> (2) . . . wherever by any accident the faculties which give rise to any impression are obstructed in their operations, as when one is born blind or deaf, not only the impressions are lost, but also their correspondent ideas; so that there never appear in the mind the least trace of either of them. (1739/1978: 49).

For example, the congenitally blind do not observe redness or seeing. Therefore by hypothesis they could not acquire the concepts RED and SEE. It follows, if word learning requires a mapping between instantiations of a concept and the hearing of a word (*qua* phonological object), then a blind child would have no basis for learning the meaning of words that express vision-related concepts. Two cases studied by Landau and Gleitman (1985) were vision verbs (*look*, *see*) and color nouns and adjectives (*color*, *green*, *red*). Sighted blindfolded three-year-olds told to "Look up!" turned their faces skyward, suggesting that they interpreted *look* to implicate vision in particular. This interpretation isn't true of the blind: a blind three-year-old given the same command raises her hands skyward instead of her face, suggesting that for her the term is connected to the manual sense.

So far so good for Locke and Hume: the difference in observational opportunities leads the blind and sighted to different interpretations of the same term. Successful communication from mother to blind child using this term often occurred just when the objects to be "looked at" were in the learner's hands, licensing a physical contact interpretation of blind looking. However, several common verbs used by the mother to the blind child shared the property of being uttered – and even more systematically than *look* – when the child had a relevant object in hand, including *hold*, *give*, *put*, and *play*.

Moreover, the blind child's interpretation of *look* goes beyond manual contact. An informative manipulation was to say "You can touch that table but

don't look at it." If *look* means *touch* to the blind, this command is incoherent and therefore cannot be obeyed; but instead the blind child gingerly taps or scratches at the table in response to this command. Subsequently told "Now you can look at it," the child explores the surfaces of the table manually. Based on this kind of evidence, Landau and Gleitman (1985) concluded that blind *look* (a) semantically differs from sighted *look* by implicating a different sense modality, but (b) semantically resembles sighted *look*, and differs from *hold*, *touch*, etc., in being a term of perception. To be sure, in order to look a blind person must touch; but that does not imply that *look* means – or "just means" – *touch*. (And predictably, "You can look at this table but don't touch it" elicits confused complaints from the blind child.)

Summarizing, we can easily account for the blind child's failure to map *look* onto the visual modality from an orthodox associative perspective on word learning that focuses on the necessity of extralinguistic observation. But this perspective cannot so easily explain how blind – and sighted – learners hit upon *looking* as a perceptual rather than as a contact term. Again, these findings suggest that word-to-world pairing is insufficient as a full explanation of word learning. Both populations know too much, from too little information about the world. Chomsky (1986) termed this "Plato's Problem."

The blind child's understanding of color terms offers a similar insight: Landau and Gleitman's (1985) blind preschool-aged informant knew that (a) *color* is the supernym for a subset of adjectival terms including *green* and *red*, but not *clean* and *happy;* and (b) the color terms apply only to concrete objects. Asked "Can a dog be blue?" the blind child at five years of age responded with different color terms: "A dog is not even blue. It's gold or brown or something else." Asked "Can an idea be green?" the child responded "Really isn't green; really just talked about – no color but we think about it in our mind." That is, *blue* may not be an actual attribute of dogs; but *green* is a category error as applied to ideas. These findings display the remarkable resilience of semantic acquisition over variations of input: lacking the ordinarily relevant observations that support solution of the mapping problem for visual terms, the blind are not helpless to do the same. But then what is the foundation for this learning?

Hard words

Many of the words that mothers frequently utter to their infants are so divorced from straightforward perception that it is hard to see how observation could possibly be available to support their acquisition. Even supposing (somewhat controversially) that the learner by the age of two or three is capable of understanding the concept we express as *think*, it is quite difficult to imagine circumstances, short of visits to the Rodin Museum, that would bring it to mind as what a conversation is about. If we point to some group of people who truly are

thinking, and contrast them with another set of people who truly are not thinking, what good can it do the learner to gaze upon these contrasting people-sets? Their thinking is happening inside their heads, invisible to any observer. Items of this sort have posed important challenges to the view that the mapping problem can be accomplished by a machinery responsive solely to the observable contingencies for a word's use. "Show them the thing . . ." seems a very much less compelling method for this kind of item than it does for the elephants.

More generally, the trouble with verbs and other predicate terms is that they're abstract, and therefore much less obviously displayed in the flow of events. Verb meanings depend not only on the events in view, but also on a choice of perspectives on events (Clark 1990; Gleitman 1990; Gleitman, Gleitman, Miller & Ostrin 1996; Pinker 1989, *inter alia*). At the extreme, perspective-changing verbs like *chase* and *flee* pose a problem of principle for any theory of lexical acquisition that relies solely on word-to-world mapping. Verbs within these pairs describe the same events, differing only in their focus on the perspective of one or the other participant in that event. This focus difference is unobservable, residing only in the speaker's head. As further examples, one cannot say of a scene that one of its participants is *giving* if nobody *gets*, and every time one *puts* the cup on the table, the cup also *lands* on the table (Gleitman 1990).

Inconvenient facts about word learning

Children do not learn every word they hear. Input frequency does not even begin to explain this fact. Perhaps we need no complex theory to explain why no child's first word is *the* despite its frequency in maternal speech. But a timing difference for object terms versus action terms (and its surface correlate, nouns versus verbs) requires a little further theorizing. Many studies show that children's early production vocabulary is dominated by concrete nouns – names for objects and people in particular (see for reviews Gentner & Boroditsky 2001; Woodward & Markman 1998). This is true in languages other than English, even in languages like Italian that possess surface properties conducive to verb learning, including the omission of inferable noun phrases. The same bias toward learning object names is present in the earliest language comprehension as well. Novel words presented in object-manipulation contexts cause one-year-olds to focus on the kinds of similarity across objects that can indicate shared category membership (Waxman & Markow 1995). When a new word is presented, the object-kind interpretation is often so salient that it's difficult to get children to arrive at any other interpretation (Bloom 2000; Gentner 1982).

At first glance, this bias does not seem to be in conflict with the observational theory of word learning. It could be – this surely is true in the limit – that some concepts are more difficult to entertain than others, and therefore are simply unavailable to the infant mind. The relational notions that verbs typically express

might be harder to grasp, or less salient to infants, than are notions of object kind. If the relevant world observations are potentially available but uninterpretable, then the words can't be learned. This account of the late appearance of verbs relative to nouns would thus accord perfectly well with the Lockian prediction. However, there is a different account of this striking, universal, input–output disparity for the efficient acquisition of nouns and verbs: the acquisition of nouns and verbs may require different kinds of information, and the information sources themselves may become available at different developmental moments. We turn now to evidence for just such a position.

Linguistic and conceptual supports for vocabulary acquisition

How can the robustness of word learning be understood in the face of the vagaries of ordinary experience and the vast array of reasonable interpretations of what has been said? We will argue that vocabulary acquisition is not of a piece. Some words are *necessarily* learned before others. The initial learning of concrete nouns sets in motion a process that makes possible the efficient learning of less concrete words; bits of the lexicon and grammar of the exposure language are acquired in a succession of causally interlocking steps. Learners construct the linguistic ladder, so to speak, while they are climbing it.

Word-to-world mapping: showing them "the thing" suffices for some words, but not for think – *or* thing

Some words are less obscure than others in the flow of experience made available for our inspection. Evidence for systematic variations in the recoverability of various words' meanings from world context alone comes from studies by Gillette, Gleitman, Gleitman, and Lederer (1999). Their interest was in understanding how information structure in the input influences solution of the mapping problem, apart from whatever role developmental differences in mentality may play. Therefore they used *adults* to simulate vocabulary learning *under various informational circumstances*.

The first step in these investigations was to understand the limits of word-to-world pairing: solving the mapping problem by using observed scenes as the sole clue. To do so, the investigators showed their adult subjects brief videoclips (about 45 seconds in length) of mothers and toddlers playing with toys and conversing. The soundtracks were removed from these video clips, and a "beep" was inserted in each clip at the moment when the mother had uttered a particular target word. The targets were the 24 nouns (e.g. *ball*, *hand*, *hat*) and 24 verbs (e.g. *push*, *come*, *look*) most frequently produced by mothers in the corpora from which these clips were drawn. For each word, the adult observers were told they would see six videotaped occasions of the same word's use in a row.

Thus they had some opportunity for cross-situational observation. Their task was to identify each "mystery word" using this accumulating evidence. With only these scene observations as evidence, adults correctly identified three times as many of the mothers' nouns (about 45 percent) as of their verbs (about 15 percent).

Success rates in this task could be predicted by other adults' judgments of the imageability (concreteness) of each word. On average, the common nouns in the mothers' speech were judged more imageable than the common verbs, and variability in judged imageability was a better predictor than the noun/verb distinction of which words were successfully induced from observation of the scenes. The most concrete of the target verbs (e.g. *throw*) were identified more frequently than the most abstract. Those judged most abstract, including *think* and *know*, were never guessed correctly by any subject. Subsequent studies have begun to refine the notion of concreteness that determines which words are relatively easy to learn from observation alone. For example, Kako and Gleitman (in prep.) found that words for basic-level categories of whole objects (e.g. *elephant*) are strikingly easier to identify based on observations of their circumstances of use alone than are abstract nouns (e.g. *thing*) or part terms (e.g. the elephant's *trunk*).

These findings (see also Snedeker & Gleitman in press) yield a simple explanation for the probabilistic noun advantage in infants' first vocabularies. The adult subjects in Gillette et al.'s (1999) studies had already grasped the concepts lexicalized by all the English words to be guessed in the study. Nevertheless, only the most concrete words were successfully identified from observing the extralinguistic contexts alone. The most concrete words, including a useful vocabulary of names for things, are just those for which linguistically unaided observation is likely to be informative.

These data confirm that the solution to the mapping problem may start with Locke's procedure but cannot end there. Observation of the thing is sufficient for the acquisition of some (e.g. *elephant*) but not all of our words (e.g. *thing*, *trunk*, *think*). The true beginner can only try to observe elements in the world that systematically covary with the use of particular words. This leads to success in those cases in which the word's meaning is concrete enough to be readily observable in the flow of events: mostly nouns, but also a heterogeneous set of other words.

Sentence-to-world mapping

How does the child move beyond an initial concrete, largely nominal, vocabulary? To learn less concrete (less observable) terms, the learner needs other kinds of evidence – linguistic evidence, bootstrapped from (*grounded by*) the previously acquired vocabulary of concrete words.[3]

The view known as *syntactic bootstrapping* proposes that the interpretation of verbs and other predicate terms is guided by information about the structure of the sentence in which the verb appears (Landau & Gleitman 1985; Gleitman 1990; Fisher 1996). Most generally, this view proposes that word learning after the first steps proceeds by sentence-to-world pairing rather than merely by word-to-world pairing.

To illustrate, let us return to the Gillette et al. (1999) "human simulations" described earlier. These investigators repeated their experiment, asking adults to identify verbs spoken to young children based on various combinations of linguistic and extralinguistic information. Adults were much more accurate in guessing which verb the mother said to her child when given information about the sentences in which the verb had occurred. When given a list of the nouns that occurred in the sentence (alphabetized to remove word-order information), along with the scene in which the verb was produced, subjects' guesses were significantly more accurate than when given the scene alone. Subjects also profited from more explicit syntactic information about the verbs' original contexts of use, even when denied observation of the scene: presented only with a set of sentences in which all the content words were replaced with nonsense words (e.g. Can ver GORP litch on the fulgar?; much as in Carroll's poem *Jabberwocky*), subjects were significantly more accurate in guessing the verbs than when they saw the scenes, or even when they saw the scenes plus an alphabetized list of cooccurring nouns. When presented with the complete sentence contexts, with only the verb replaced by a nonsense word (e.g. Can you GORP Markie on the phone?), subjects' guesses were quite accurate even without access to the scenes, and nearly perfect with both sentence and scene.

Why would syntactic information so strongly guide semantic inferences? Verbs vary in their syntactic privileges (i.e. the number, type, and positioning of their associated phrases). These variations are systematically related to the verbs' meanings (Chomsky 1981a; Fisher, Gleitman & Gleitman 1991; Gleitman 1990; Grimshaw 1990; Jackendoff 1983; Rappaport, Hovav & Levin 1988; Pinker 1989, *inter alia*). A verb that describes the motion of an object will usually occur with a noun phrase that specifies that object; a verb that describes an action on an object will typically accept two noun phrases (i.e. be transitive); a verb that describes the transfer of an object from one position to another will take three arguments. Similarly sensible patterns appear for argument type: *see* can take a noun phrase as its complement because we can see objects, but also can take a sentence complement because we can perceive states of affairs.

Such syntactic–semantic correspondence patterns show striking regularities across languages (Baker 2001; Croft 1990; Dowty 1991). These crosslinguistic regularities have long been taken to be primary data for linguistic theories to explain, leading to principles such as the *theta criterion* and the *projection principle* (Chomsky 1981a), which jointly state that the nouns in sentences must

be licensed by the right kind of predicate (one that can assign them a thematic or "theta" role), and that clause structure must be projected from lexical entries. Similarly, unlearned constraints linking thematic roles such as *agent* and *theme* to grammatical functions like *subject* and *object* have been proposed to explain crosslinguistic regularities in the assignments of semantic roles to sentence positions. Causal agents, for example, overwhelmingly appear as grammatical subjects across languages (Baker 2001; Keenan 1976).

Based on these systematic links between syntax and meaning, the adults in Gillette et al.'s (1999) studies, or a suitably constructed young learner, can consult each verb's sentence structure to glean information about its meaning. The observed sentence structure, by specifying how many and what types of arguments are being selected by the verb, provides a kind of "linguistic zoom lens" to help the learner detect what is currently being expressed about an ongoing event or a state or relation. The set of such structures associated with a verb, across usages, is a complex function of its full expressive range (we discuss these issues further on p. 138, below).

Linguistic evidence aids identification of abstract nouns as well. Kako and Gleitman (in prep.) found that the inductive advantage for basic-level object kinds was reduced when linguistic information was added to or substituted for the scene information (*Sorf the RENCK's reb?* or *See the RENCK's trunk?*). Consistent with this finding, several prior studies have found that abstract, superordinate, or part nouns are typically introduced into the conversation in informative *linguistic* contexts (e.g. *This is the bear, here are his ears*; *Here's a dog, a cat, and a horse, they're all animals*; see Shipley, Kuhn & Madden 1983; Callanan 1985).

Inferences from syntax to meaning will presumably differ in their mechanics for abstract nouns and for verbs. To a considerable degree, however, sentence structures will be informative insofar as they convey information about the predicate-argument structure of their meanings, for argument-taking nominals of various sorts (e.g. *John's shoe; the fact that Bill likes ham*), nominal arguments of known verbs (e.g. *feeding the ferret*), unknown verbs (e.g. *she adores ham*), or other argument-taking predicates (e.g. *the cat is on the mat*).

Summary: the information base for word learning

The "Human simulation" studies just discussed tell us about the information structure of the input: more than one kind of information is available for vocabulary learning, and these information sources are more or less informative depending on the kind of word being learned. For the case of basic-level names for things, reference is fairly easy to determine from unaided inspection of the scene, whereas there is almost no information to be gained from these words' licensed positions in sentences (other than the fact that they are nouns). It is easy

to see why. Tens of thousands of English nouns appear in the linguistic context "the gorp," so this context hardly narrows the search-space for mapping; in contrast, it is relatively easy to observe, say, a horse or a flower or a fork in the situational context. The meanings of more abstract words, including most verbs, are harder to identify in the flow of events, but have more informative linguistic contexts. At the far end of this abstractness continuum are the credal verbs, such as *think* and *know*, for which the situational observation is of almost no value, while the linguistic-syntactic information is hugely informative for these very cases (Gillette et al. 1999; Snedeker & Gleitman in press).

Children's use of multiple cues in word learning

Important as it is to determine the potential informativeness of the multiple cues to word meaning available in the input, it still remains to demonstrate how and when learners are responsive to them.

The meanings to be communicated, and their systematic mapping onto linguistic expressions, arise independently of exposure to any language

In advance of language learning, infants during the first year of life naturally factor their representations of events into conceptual predicates and arguments (Bloom 2000; Fisher & Gleitman 2002). Some of the most striking evidence that the structure of human cognition yields a language-appropriate division of our thoughts into predicates and arguments comes from learners who are isolated from ordinary exposure to a language and therefore have to invent one on their own.

Most deaf children are born to hearing parents who do not sign, and therefore the children may not come into contact with gestural languages for years (Newport 1990). Deaf children with no available language model spontaneously invent gesture systems called "Home Sign" (Feldman, Goldin-Meadow & Gleitman 1978; Goldin-Meadow 2003). Remarkably, though these children are isolated from exposure to any conventional language, their home sign systems partition their experience into the same pieces that characterize the elements of sentences in Italian, Inuktitut, and English. Specifically, home sign systems have nouns and verbs, distinguishable from each other by their positions in the children's gesture sequences and by their distinctive iconic properties. Moreover, and especially pertinent to the issues that we have been discussing, sentence-like combinations of these gestures vary in both the number and positioning of the nouns as a function of what their verbs mean. Systematically appearing with each verb in a child's home sign system are other signs spelling out the

thematic roles required by the logic of the verb: the *agent* of the act, the *patient* or thing affected, and so forth.

The nature of this relationship is easy to see from a few examples: Because "crying" involves only a single participant (the crier), a verb with this meaning appears with only one nominal argument. Because "tapping" has two participants, the tapper and the thing tapped, such verbs appear with two nominal arguments. Because "giving" requires a giver, a getter, and a gift, this verb shows up with three nominal phrases. As mentioned earlier, these semantic functions of the nouns vis à vis the verbs are known as their *thematic* or *semantic* or *theta roles* (Chomsky 1981a). The same fundamental relationships between verb meaning and nominal arguments surface in much the same way in the speech of children who are acquiring a conventional language, and in the gestures of linguistically isolated children who must invent one for themselves.[4]

In addition, the nouns occurring with each verb do not appear haphazardly to either side of the verb. The isolated deaf children adopt systematic gesture orders, such as routinely signing undergoers immediately before verbs, (transitive) agents following verbs, and intransitive actors before verbs. Thus, a home signer who produced "Snack$_{<\text{theme}>}$-Eat-Susan$_{<\text{agent}>}$," might also produce "Susan$_{<\text{actor}>}$-Move Over" and "Cheese$_{<\text{theme}>}$-Eat" (Goldin-Meadow 2003). Apparently, just as no child has to learn to factor experience into predicates and arguments, no child has to learn to use word order systematically to specify the semantic role played by each element.

In sum, linguistically isolated children construct, out of their own thoughts and communicative needs, systems that resemble the languages of the world in at least the following universal regards: all have words of more than one kind, at minimum nouns and verbs, organized into sentences expressing predicate-argument relations. The number of noun phrases is predictable from the meaning of the verb; the positioning of the nouns expresses their semantic roles relative to the verb. Thus, the fundamental structure of the clause in both self-generated and more established communication systems derives from the non-linguistic conceptual structures by which humans represent events, coupled with strong preferences for "flattening" these conceptual structures into linguistic expressions. This "cognitivist" interpretation of the origin of language in child conceptual structure motivates all modern linguistic treatments of verb semantics that we know of (Baker 2001; Chomsky 1981a; Dowty 1991; Fillmore 1968; Jackendoff 1983; Rappaport-Hovav & Levin 1988). The same cognitivist approach figures in most psychological theories about learning of both syntax and lexicon, whatever the other disagreements of their proponents (e.g. Gleitman, Gleitman et al. 1988; Pinker 1989; Slobin 2001; Tomasello 2000). Indeed, both "nativist" and "learning-functionalist" wings of the language-learning investigative community have seized upon the transparency and universality of such form-to-meaning correspondences in language acquisition as uniquely supporting their

learning positions in particular. (After the battle, the opposing generals retreated to their tents to celebrate their victory.)

Young children use the structure of a sentence to guide interpretation of new verbs

Here we discuss evidence that young language learners, much like the infant language inventors that we just discussed, exploit form-to-meaning correspondences as a rich source of evidence about the lexicon.

The case of argument number Particularly well studied has been early sensitivity to noun-phrase number as a cue to verb interpretation (Fisher 1996, 2002; Naigles 1990; Lidz et al. 2003; Naigles & Kako 1993). For example, Naigles (1990) showed that children as young as 25 months of age interpret new verbs in accord with the number of their noun-phrase arguments. The children watched a video-taped event in which two actions occurred simultaneously: in one composite display, a bunny pushed a duck into a bending posture, while the bunny and duck bent their free arms at the elbow. Each child heard this display described by either a transitive ("The bunny is gorping the duck!") or an intransitive sentence ("The bunny and the duck are gorping!"). Following this training, the two subevents of the composite scene were shown separately on two side-by-side monitors, and the children were exhorted to "Find gorping!" One screen showed the causal event in which the bunny bent the duck; the other showed the non-causal event in which both animals bent their arms. Children who had heard the transitive training sentence looked longer at the causal event, while children who had heard the intransitive sentence looked longer at the non-causal event.

Similar syntactic evidence can persuade young children to alter their interpretation of a familiar verb. Naigles, Gleitman, and Gleitman (1992) asked preschoolers to act out sentences using a toy Noah's Ark and its associated characters. The informative trials were those in which a verb was presented in a new syntactic environment, as in *Noah brings to the ark* or *Noah goes the elephant to the ark*. Young children adjusted the interpretation of the verb to fit its new syntactic frame, for example acting out *go* as "cause to go" (a.k.a. "bring") when it was presented as a transitive verb.

Compare these results with the innovations of the deaf home signers who invented their own manual communication systems. In both cases, children map participants in a conceptual representation of an event one-to-one onto noun arguments in sentences. Elsewhere we have proposed (Fisher 1996, 2000a; Gillette et al. 1999) that children might first arrive at this structure-sensitive interpretation of a sentence in a simple way – by aligning a representation of a sentence with a structured conceptual representation of a relevant situation.

In this way a child might infer that a sentence with two noun arguments must encode some conceptual relationship between the referents of the two nouns, while a sentence with only one noun argument might describe a state, property, or act of its single referent. This simple structure-mapping could take place as soon as the child learns to identify some nouns, and can represent them as parts of a larger utterance.

The centrality of argument number as a learning cue is further clarified by recent studies that emphasize two important issues. One is that sensitivity to argument number makes its appearance astonishingly early in language acquisition, often before the child has uttered a single verb. The other is that this sensitivity can be demonstrated even in stripped-down experimental settings that remove all alternative evidence.

In a recent series of studies (Fisher 1996, 2002) children aged two, three, and five years heard novel verbs in the context of (videotaped) unfamiliar causal events; the verbs were presented either transitively or intransitively. The sentences contained only ambiguous pronouns, as in *She's pilking her over there* versus *She's pilking over there* so that the sentences differed only in their number of noun phrases. The children's interpretations of the novel verbs were tested by asking them to point out, in a still picture of the event, which character's role the verb described (*Who's pilking (her) over there?*). Children at all three ages were more likely to select the causal agent in the event as the subject of the transitive verb. Just as for the adult judges in the Gillette et al. (1999) studies, these findings provide evidence that the *set of noun phrases* in the sentence – even without information about which is the subject – influences young children's interpretations of verbs. Recently these findings have been extended to children as young as 21 months of age. Fisher and Snedeker (2002) showed 26- and 21-month-olds side-by-side videotaped events. One screen displayed a novel caused-motion event involving two people, and the other displayed a novel independent motion event involving only one person. As they watched each pair of scenes, the children heard either a transitive (*He's pilking him!*) or an intransitive sentence (*He's pilking!*). Children who heard the transitive sentence looked longer at the two-participant caused-motion event, while children who heard the intransitive sentence tended to look equally at the two events (both of which displayed possible referents for an intransitive verb).

The findings reported so far are consistent with the view that there is a bias to map one-to-one between the set of arguments of the verb and the set of participants in the event, in children acquiring an established language as well as for linguistic isolates inventing their own sign systems. But perhaps, in the case of children learning an established language, the early honoring of this simple mapping from participant number to noun-phrase number is an effect of language learning rather than the reflection of some unlearned bias. Do children simply exploit the most stable cues to mapping made available in

the language they hear, rather than relying on an unlearned bias for one-to-one mapping?

To investigate this issue, Lidz, Gleitman, and Gleitman (2003) asked preschoolers to act out novel combinations of verbs and syntactic structures in two languages: English (as in Naigles et al. 1992) and Kannada, a language spoken in southwestern India. Kannada permits pervasive argument dropping, rendering the relationship between argument number and noun-phrase number relatively variable in typical input sentences. Kannada also has, however, a causative morpheme that only occurs with causative verbs. The critical sentences pitted argument number (two nouns vs. one) against causative morphology (explicitly marked as causal or not). Kannada-speaking three-year-olds ignored the presence or absence of the causative morpheme, relying only on the number of noun-phrases in the sentence they heard. In contrast, Kannada-speaking adults' enactments were influenced by both morphology and argument number. The adult findings again demonstrate that language learners ultimately acquire whichever cues to sentence meaning the exposure language makes available. But strikingly, they also show that children are not totally open-minded: they appear to find some formal devices (argument number) more potent than others (inflectional morphology).

It is important to notice that the count-the-nouns procedure taken by itself is coarse at best and fallible at worst. This is because *nouns in the sentence* and *arguments of the verb* are by no means the same thing. Often, for example, there are too few nouns to match up with the event participants. This is for several systematic reasons, including incorporation phenomena of many kinds, and the possibility of argument omission. In many languages, sentence subjects can be omitted if they are recoverable from context and prior discourse; in some languages, including Chinese, Japanese and Korean, a verb's direct objects can be omitted as well. Violations of any simple noun-counting principle are also obvious in the reverse direction, for example when a language (like English) requires a subject even for argumentless predicates (*It is raining*). And in any language, complex noun-phrases (*John's sister*, *a horse of a different color*) contain more than one noun, and sentences can contain adjunct phrases (*with Ginger*, *in the morning*), again yielding more noun phrases than argument positions.

Despite the complexity of the relationship between nouns in sentences and the subcategorized arguments of the verb, several sources of evidence suggest that ordinary sentences provide strong probabilistic information about the participant structures of verbs. For example, in the human simulations of Gillette et al. (1999), adults benefited from simply being given an alphabetized list of the nouns in each sentence in which the mothers had produced a particular verb. In this case the adults (like the hypothetical learner) could not tell which nouns were arguments of a verb and which were adjuncts, yet this linguistic hint aided recovery of verb meanings from scenes. Li (1994; see also Lee &

Naigles 2002) analyzed speech to young children in Mandarin Chinese, and found that although mothers often did omit noun phrases in sentences, maternal utterances still supported a systematic distinction among semantically and syntactically distinct classes of verbs. Though arguments can be omitted, transitive verbs still occur with two nouns in the sentence more often than intransitive verbs do, and systematically so.

Beyond argument number Experimental studies of novel verb learning by young children have focused on argument number, in part because this is an easily detectable cue to sentence interpretation. But clearly there's more linguistic evidence for lexical learning than argument number. Landau and Gleitman (1985) argued, based on analysis of a blind child's lexical development, that the child's observation that *look* and *see* appeared in sentence complement structures made sense of her seemingly effortless acquisition of the perceptual nature of these verbs, while purely observational evidence yielded no clear way to discriminate these from object-contact verbs like *touch* and *hold*. One can, for example, *see that the sky is falling*, and *look how I'm doing this*, but not *touch that the sky is falling*. Argument type, like argument number, provides a powerful source of information for lexical learning.

Differences in argument type and number systematically map onto a semantic cross-classification of the verb lexicon, as revealed by naïve adults' judgments of semantic relatedness among verbs (Fisher et al. 1991). Verbs that accept sentences as their complements describe relations between their subjects and an event or state; these include verbs of cognition (*know*, *think*), perception (*see*, *hear*), and communication (*explain*, *say*). Verbs that take three noun-phrase arguments describe relations among the referents of those three noun phrases, typically transfer of position (*put*, *drop*), possession (*give*, *take*), or information (*explain*, *argue*). Later studies using the Fisher et al. procedure documented that these regularities could be recovered from a sample of English sentences produced in spontaneous child-directed speech in English (Lederer, Gleitman & Gleitman 1995) and Mandarin Chinese (Li 1994). Verbs' syntactic behavior, including both argument type and number, thus provides a source of information that systematically cross-classifies the set of verbs in much the same way within and across languages, pointing to dimensions of semantic similarity. Indeed, it is this *cross-classification* – the set of structures associated with a single verb or small verb class – that accounted for subjects' accuracy in the *Jabberwocky* (syntax-only) condition of the Gillette et al. (1999) human simulation experiments.

Can young children, like these adults, profit from the full range of syntactic structures they might observe with each verb? Considerable evidence tells us that they are quite good at learning about the sentence structures in which particular verbs occur (Gordon & Chafetz 1990; Snedeker, Thorpe & Trueswell 2001;

Tomasello 2000); such findings suggest that they may well be capable of taking advantage of probabilistic evidence, presented across multiple sentences, for the range of sentence structures assigned to each verb. Moreover, a computer simulation of syntactic learning from a sample of child-directed English (Brent 1994) suggested that subcategorization frames for verbs could be recovered, based on very little prior syntactic knowledge (a few function words): an analysis of verbs' lexical contexts provided useful information for distinguishing among verbs that are transitive or intransitive, or that take verbal or sentential complements (as in *John likes to fish*).

Beyond the first primitive mapping of two-noun sentences onto two-participant relations, mapping rules that are language-specific also come into play, further enriching the informational base and thus further increasing the efficiency and precision of predicate-term acquisition. The earliest-appearing of these is probably the interpretation of word order in multi-argument sentences: Hirsh-Pasek and Golinkoff (1996) reported that English-learning 17- to 19-month-olds were sensitive to word order in transitive sentences containing familiar verbs (*Cookie Monster is tickling Big Bird* vs. *Big Bird is tickling Cookie Monster*): they looked longer at a video screen on which the subject of their test sentence was the agent of the target action. Children aged 21 and 26 months show the same sensitivity to English word order when presented with made-up verbs and unfamiliar actions (Fisher 2000b). Young children acquiring a free word-order language quickly acquire the semantic implications of case-marking morphology (e.g. results for Turkish learners reported in Slobin 1982).

Children also develop more subtle language-specific expectations about the meanings of classes of words in their language; Slobin (2001) has termed this phenomenon "typological bootstrapping." For example, Talmy (1985) described systematic differences in the typical meanings of motion verbs in languages like Spanish (verb-framed languages) and English (satellite-framed languages). Spanish motion verbs tend to encode direction (*enter*, *ascend*), while in English, path information is relegated to a prepositional phrase, and verbs are more likely to encode manner (*walk in, run up the hill*). Native speakers of Spanish and English learn these tendencies, and develop slightly different expectations for the likely semantic content of a new verb (Naigles & Terrazas 1998; for further discussion see Choi & Bowerman 1991; Fisher & Gleitman 2002; Landau & Gleitman 1985).

The particular nouns that typically occur with each verb also undoubtedly guide the child's interpretation: *drink* and *eat* are not only transitive verbs; they systematically select animate subjects and different direct-object nouns (the potable and edible items). Data from "human simulations" suggest that this sort of information is helpful in both noun and verb learning (Gillette et al. 1999; Kako & Gleitman in prep.; see Pinker 1989 for the initial statement of this

proposal). Quite young children have access to this sort of information as well: two-year-olds look at a picture of a glass of juice (rather than of a non-potable object) when they hear the familiar verb *drink* (Fernald 2003), and successfully induce the referent of a new word introduced in an informative context as the object of a familiar verb (e.g. *She's feeding the ferret!*) (Goodman, McDonough & Brown 1998).

Summary

In the course of lexical development, children have opportunities to observe each verb's typical subcategorization frames, its typical nominal arguments, and the kinds of scenes or events that pertain when the verb is invoked. Although much work remains to be done to specify how each source of information is detected and used by children, and how multiple sources of information interact in development, we argue that all of these sources of information converge to make vocabulary learning efficient and nearly errorless.

Central to this so-called *syntactic bootstrapping* view is the interaction of multiple cues for word learning, trading off in different ways for different classes of verbs. This position does not *replace* observation of situations with linguistic observations. Ultimately, word learning is a mapping problem. The learner must identify what (in the world, or at least in a human's conception of the world) the surrounding community of speakers means by each word. Though our arguments focus on the ambiguity of referential settings – and thus the need for linguistic evidence to make vocabulary learning stable – the observations that give semantic content to words are observations of the non-linguistic world. Sentence structures are relevant only to a subset of the dimensions of verb meaning, those that affect the number and type of arguments associated with the verb, and the temporal structure of the event it names (Fisher et al. 1991; Grimshaw 1990; Rappaport-Hovav & Levin 1988). In contrast, the various manners of motion encoded by *slide*, *roll*, and *bounce* have no direct reflection in sentence structure (Fillmore 1968). When *push* comes to *shove*, only observations of the manner (and comparative violence) of actions in the world will suffice to differentiate syntactically and semantically related verbs.

Lexical learning and the structure of linguistic knowledge

We have summarized evidence that, in order to acquire word meanings, child learners amass all sorts of specific knowledge about individual words: their contexts of use, their lexical-distributional properties (e.g. the association of *bake* with *cake*) and the full range of their syntactic behavior. All this evidence is required, in different degrees for different words, to converge on their meanings. The result is a knowledge representation in which detailed syntactic and

semantic information is linked at the level of the lexicon. What happens to all this information, collected to solve the mapping problem? Is it disassembled, rather as a building scaffold is dismantled once the beams and bricks are in place? We think not.

Experimentation on sentence comprehension in older children and adults suggests the continued linking of linguistic distributional knowledge to particular lexical items. Native speakers learn not only which sentence structures each verb can grammatically combine with, but also how often each verb occurs in each structure. Adults retrieve this information as soon as they identify a verb, and use it to bias online sentence interpretation (Garnsey et al. 1997; Trueswell & Kim 1998). Snedeker et al. (2001) demonstrated that both children and adults resolved the ambiguity of such sentences as *Tickle the frog with the feather* and *Choose the frog with the feather* as a function of the frequency with which these verbs naturally occur with noun-phrase versus verb-phrase modification. Online parsing decisions by adults and by children as young as five are influenced by detailed and frequency-sensitive knowledge about the syntactic behavior of each verb.

These findings from the psycholinguistic literature mesh naturally with computational approaches to parsing that also represent syntactic representations as strongly lexicalized: in Lexicalized Tree Adjoining Grammar, for example, the syntactic possibilities of a language are represented by a finite set of tree structures that are linked with individual lexical items, and a small set of operations by which trees can be joined (Joshi & Srinivas 1994). This apparatus permits the statement of syntactic dependencies (such as subcategorization) and semantic dependencies (such as selection restrictions), and yields a natural treatment of non-compositional idioms (*kick the bucket*). Such approaches are based on a claim similar to the one we derive from examination of the learning procedure: an adequate description of the syntactic combinatorial principles of a language is unworkable if kept separate from the lexicon and lexical learning. Similarly, independent evidence from crosslinguistically based theoretical linguistics supports a view of language in which significant structural properties reside in the lexical component of the grammar (Borer 1984; Chomsky 1995c).

General and particular, and the requirement for universal grammar

The manifest specificity of lexical organization that we have discussed throughout has often been taken as supporting a picture of language and its learning that can avoid appeals to unlearned constraints on the construction of grammars. For example, Tomasello (2000) and Goldberg (1995) have proposed construction-based accounts of language representation and acquisition, suggesting that children simply learn, word by word and construction by construction, how to

express each idea. According to this view, a more general and flexible grammar grows slowly from this piecemeal knowledge through general cognitive principles of induction and generalization. Proponents of such theories emphasize variability and exceptions in the syntax–semantics mapping rules, and argue that no constraints beyond those of the human cognitive/conceptual apparatus are needed to account for the nature of human languages, and the facts of language development.

However, we believe that this is a mistake. Although syntax must be represented in the lexicon – to explain how we know, with such exquisite detail, which structures each verb appears in, and how we learn the meanings of the verbs in the first place – strong universal constraints on the alignment of syntax and semantics are needed to explain the full set of facts. One issue, compellingly discussed by Mark Baker (2001), is the uniformity of clause-level structures within languages:

> We do not find languages in which the verb meaning "hit" comes before the object, English-style ("The child might hit his parent"), and the verb meaning "kiss" comes after it, Japanese-style ("The child might his parent kiss"). The word order of the object and the verb is thus not learned purely by learning . . . individual verbs but must be somehow keyed into the process of learning the verbs as a class. (2001: 80)

If the organizational structure for predicate–argument structures could be anything at all, with the learner simply picking up these facts on an item-by-item basis, it would be hard to explain why each language organizes these structures so regularly, across verbs. The same is true for the basic semantic–syntactic linkages, even in languages at the extremes of linguistic diversity, as Baker shows by comparing languages such as English and Mohawk. In all the languages of the world, though with occasional quirks and exceptions, not only do all the core participants in the action denoted by the verb get expressed grammatically (the "theta criterion"), but the causal verbs put their agents in subject position, the undergoers systematically surface as objects, the complements of mental verbs surface as clauses, and so forth. These linguistic properties shared across cultures and language families, however otherwise diverse, imply strong restrictions on how we factor experience into predicates and arguments, and what aspects of the conceptual predicates and arguments are reflected in the organization of the clause. Children, being creatures like us, expect language to be organized in accord with such principles and therefore they can learn it.