From Structuralism to Cognitive Science: Lila Gleitman's contributions to Language and Cognition

Jeffrey Lidz
University of Maryland

When Lila Gleitman enrolled as a graduate student at the University of Pennsylvania in the late 1950s the cognitive revolution was just picking up steam, but structuralist linguistics and its empiricist cousins in philosophy and psychology still had a very strong hold on how scholars and scientists thought about language and its acquisition. At the end of her career, now spanning 6 decades and still going strong, Lila presides over a field with very different contours. Lila raised some of the most powerful arguments against the orthodoxies of her day and ushered in a computational cognitive science of language, taking the in principle arguments of generative linguistics and demonstrating where the rubber meets the road. Lila innovated new ways of probing children's knowledge of language, new ways of thinking about that knowledge, and new ways of thinking about learning. Her perspective incorporated the obvious input-sensitivity of learners with a rationalist conception of mind, demonstrating how innate structure empowers learning. On the one hand, she showed that empiricist methods of language learning are unjustified, given the many ways that learners acquire knowledge richer than what can be found in the learning environment. On the other hand, she provided new ways of thinking about how learners do use their input to acquire a specific grammar. The overall rationalist conception of language acquisition that she offers reveals that there is more to language and its acquisition than meets the eye.

The historical context

Linguistics at Penn in the 1950s was very much a structuralist place. Lila gravitated towards Zellig Harris who, like Bloomfield, held the view that languages could vary without limit. The purpose of Harris' structuralism was to come up with flexible procedures for describing languages that would be robust to this endless variation. Harris (1951) describes the job of the linguist as identifying procedures that begin with the raw data of speech and end with a statement of grammatical structure, "a twice-made application of two steps: the setting up of elements, and the statement of the distribution of these elements relative to each other." But the aim of describing these procedures was limited in scope, unrelated either to the psychology of language acquisition or even to the problem of identifying a possible sentence of a language, let alone a possible language. Harris states, "Since the whole descriptive system is stated not for a language as a whole (whose complete stock of utterances cannot be listed), but for a corpus containing a closed utterance stock, the statistical problem of how this corpus is a sample of the language is dealt with outside the scope of descriptive linguistics, and not by its methods," (1951, p374).

Chomsky's *Syntactic Structures* (Chomsky 1957) represented a radical departure from this perspective. Chomsky imagined a linguistics that not only described sets of sentences, but inferred a constrained space of possible human languages, and explained its dimensions. Chomsky's review of Skinner (Chomsky 1959) displayed how behaviorist approaches to the acquisition of language were not equipped to capture either the infinite capacity of speakers to produce and understand new sentences or the stimulus-free nature of language use. Despite the force of Chomsky's ideas, it was not obvious how to operationalize them in a way that would allow for positive proposals about how language is acquired.

The chasm between structuralist and generative linguistics was exacerbated by a deep-seated anti-mentalism in linguistics, psychology and philosophy. Mental constructs, whether they represented phonemes, phrase structure rules, or meanings, were to be avoided. Bloomfield (1936) expressed the view that in order for linguistics to be a science, all talk of mental constructs should be replaced by reference to physical patterns: "Non-linguists (unless they happen to be physicalists) constantly forget that a speaker is making noise, and credit him, instead, with the possession of impalpable 'ideas'. It remains for the linguists to show, in detail, that the speaker has no ideas and that the noise is sufficient."

Quine (1951) gave an argument in favor of this anti-mentalist stance, rooted in an argument against the analytic/synthetic distinction. At stake was the intuition that there is a fundamental difference between sentences like (1a) and (1b).

(1) a. People who walk move their legs.
b. People who walk are in good health.

The traditional story goes that (1a) is “analytic,” because we know that it is true just by knowing what *walk* means; but we need to know more to know the truth of (1b), which is therefore "synthetic." Thus, one might say (1a) could be rewritten as (2), with *walk* replaced by its definition, making transparent why the former is uninformative (Frege 1884):

(2) People who transport themselves by moving their legs move their legs.

But Quine argued that the line dividing (1a) and (1b) is not so clear. How can we be sure that the meaning of “walk” has MOVE or LEGS as a part? Isn't it equally plausible that the distinction merely has to do with the strength of our belief that all walkings involve moving one’s legs? How can we distinguish between an analytic truth and a belief so strong that we can’t imagine letting it go? (See Rey 2003 for a fuller elucidation)

This argument led Quine to a kind of superficialism, whereby we could never know whether two words were synonymous, whether two people shared the same meaning for a term, or even whether it was possible to identify the meaning of a term outside of the entire set of assumptions the speaker has about the world. And more generally, what holds for meanings holds just as well for any aspect of a scientific theory of the mind. There could be no fact of the matter regarding mental representations because a different set of primitives and relations could yield the same pattern of behavior. Behavior, on this view, should thus be explained in terms of a simple stimulus-response psychology. The hope that there might be a uniform system underlying that behavior was misplaced. If psychologists or linguists posited any kinds of structures, these were to be seen as convenient ways for the scientist to talk about the behavior, though these had no claim to being a part of reality, psychological or otherwise.

This point of view fit perfectly with Harris' structuralism. Harris saw it as a virtue that his distributional methods would identify the units of phonological and morphological analysis without making reference to meaning or mental states. His view was that "in general morphemes which differ in meaning will also differ in their environments, if we take sufficiently long envi-

ronments and enough of them," (Harris 1951, p189).[1] In other words, we can differentiate meanings by looking at relative distributions. No further elucidation of what meanings are (if they are anything) would be required.

This empiricist world view is essentially a "what you get is what you have seen" world view. If there is learning, then it consists of discovering the patterns that are evident in the environment, because the only source of knowledge is the environment. And if there is behavior, that behavior, not some unobservable mental system for producing it, is the target of study. When we look back over Lila's work, we see strong arguments against both of these positions. First, she argues that behavior is not a sufficient basis for conclusions about the structure of the mind. Observable behavior (in the domain of language) is only indirectly related to sources of that behavior. One should never confuse sentences, or even patterns of sentences, with grammar. And one should never confuse performance in some task with the knowledge that drives that performance. Second, her work argues that some knowledge of language comes from sources internal to the child, hence not from experience. Learners, on this view, do not learn languages by finding patterns and matching them in their own behavior. Rather, learning is inference. The patterns found in the world are evidence to be used by the learner in identifying the system responsible for those patterns. And sometimes, there is no evidence at all, and still learners acquire a grammatical system. To the extent that these kinds of ideas are well represented now in cognitive science and the psychology of language, it is in no small measure due to Lila's role in developing and defending them.

From the presidential questions to the competence-performance distinction

One of the key breaking points between generative grammar and its structuralist precursors lay in the notion of possible sentence. As noted above, Harris limited the scope of his distributional analyses to the description of a corpus, leaving the question of how the corpus related to the language as a whole to some other field. For Chomsky, however, the corpus was merely a sample, one manifestation of a particular abstract mental system. The object of study was this system, not the data in the corpus. The data was merely evidence. This point of view allowed for a broader set of methods for investigating grammar, most notably the introduction of acceptability judgments. One could ask not merely what sentences occurred but also what other sentences could have occurred but happened not to have been sampled. The linguist could then construct sentences and ask whether they were possible in the language, and use these judgments to identify the characteristics of the grammar.

In the early days of developmental psycholinguistics, many researchers seemed to have adopted this Chomskyan point of view. The data that a child produces is merely a shadow of the grammar underlying those sentences. Just as a shadow hides the complexity of its physical source, children's utterances do not reflect the full complexity of their grammatical source. The early goal, therefore, was to describe children's utterances in terms of the grammar responsible for them (Bellugi & Brown 1964, Bloom 1970, McNeill 1970, Brown 1973). As Lila put it, the goal was to answer "the presidential questions" - what do they know and when do they know it? But it didn't take much time for these early psycholinguists to discover that getting young children to provide explicit acceptability judgments was nearly impossible (but see chapters 4-5). And so, many of them relied on the methods of structuralist linguistics - distribu-

[1] The grip of this view about meaning can be seen in contemporary computational approaches to word meaning (Landauer & Dumais 1997, Bengio et al 2003, Mikolov et al 2013, inter alia).

tional analysis of a corpus (Braine 1963, Miller & Ervin 1964, Brown & Fraser 1964). And given that children produced "telegraphic speech" lacking in many functional elements, these studies came to the conclusion that children lacked the grammatical structures that in adult English license these forms.

Shipley, Smith and Gleitman (1969; chapter 3 here), however, took issue with this conclusion for two reasons. First, they recognized that a corpus cannot reveal the full range of grammatical structures in any language, because it does not sample from a wide enough range of the grammar. Any speaker's linguistic knowledge goes well beyond the sentences they happen to produce, and so corpus analysis is unlikely to reveal the breadth of structures possible in the language. Second, they recognized that many performance factors, external to the grammar proper, contribute to any corpus. Thus, telegraphic speech could be a reflection not of the grammar, but of factors having to do with speech planning and production, such as working memory, attention, and what we now think of as executive function, intervening between the grammatical structures licensed by the grammar and the sentences that children actually produce.

To test this hypothesis, Shipley, Smith and Gleitman (1969) presented children with commands either in grammatical adult English or in telegraphic speech. And, cleverly, they trained the mothers to give the commands as a way of ensuring that children would be comfortable with the experimenters. The key finding was that children who produced telegraphic speech responded more appropriately to adult speech than they did to telegraphic speech. The sentences most typical of children's speech were the least effective to them as commands. This result suggested that telegraphic speech is not a straightforward reflection of children's grammars. At some level, even children who produced telegraphic speech nonetheless recognized the appropriate forms in the target language. Thus, telegraphic speech derives from an interaction between the child's grammar and extralinguistic factors relating to sentence production.

The importance of this result cannot be overstated. This paper was the first demonstration of the centrality of the competence-performance distinction in explaining children's behavior. It set the stage for decades of research aiming to diagnose children's linguistic behavior as a reflection of an immature grammar or the interaction of an essentially adult-like grammar with extralinguistic systems that deploy that grammar in real time (Hamburger & Crain 1984, Bloom 1990, Grodzinsky & Reinhart 1993, Tomasello 2000, Omaki et al 2014, White, Baier & Lidz 2017, among many others).

The lesson that children's utterances were often an indirect indicator of their grammatical knowledge, however, opened the floodgates for further explorations about the link between knowledge and behavior more generally. We can see these explorations in Lila’s work on prototypes (chapters 19-20), the Sapir-Whorf hypothesis (chapters 23-24), and in her work on syntactic bootstrapping (chapters 9-14).

Prototypes and experimental credulity

Quine’s arguments against the analytic-synthetic distinction were taken by Fodor (1975; Fodor, Bever & Garrett 1975, et seq) to be arguments against definitions in general (though, of course, Fodor rejected Quine’s behaviorist conclusions). In general, neither concepts nor word meanings could be identified with necessary and sufficient conditions on their true application. And if meanings were not definitions, they had to be something else. While Fodor adopted the view that lexical meanings/concepts were unstructured atoms, others in the field pushed for a dif-

ferent alternative – namely, that meanings were prototypes – sets of features that were neither necessary nor sufficient for class membership, but which defined a family resemblance structure that shaped the category. Robins were more birdlike than penguins because they had more of the relevant features of bird-hood.

The central empirical evidence for the prototype theory came from Rosch (1978) who showed that people made graded judgments of category membership in ways that aligned with the perceived structure of the prototype. They judged penguins to be less good examples of birds and it took longer for them to make these judgments. Armstrong, Gleitman and Gleitman (1983; chapter 19 here), however, questioned the inference. They noted that concepts could *have* prototypes without *being* prototypes. And they showed that this was the case by asking people to make typicality judgments about concepts that everyone agreed did have necessary and sufficient conditions: concepts of arithmetic and geometry. Four was a better even number than 34 and a square was a better quadrilateral than a trapezoid. The behavior initially taken to be evidence for the prototype structure of concepts was therefore evidence about something other than the concept, since it applied even where it should not. Here again, Lila's work demonstrated that prima facie evidence about some aspect of the mind could be seen through a different lens. Our theory of the mind would need to rely on careful consideration of alternatives and less direct evidence. Behavior is not always a direct expression of the mental representations and computations that give rise to that behavior. More generally, this work provided a warning against experimental credulity, and this caveat pervades all work in psycholinguistics to this day.

Symmetry

A related issue arose in the domain of symmetrical predicates, a topic that came up several times over the 50-plus years of Lila's career. The issue first arose in the context of Harris' syntactic theory, in which transformations related sentences with different form to the same kernel sentence. Thus, as in the generative theories that immediately followed, passives and actives were derived from the same source:

(3) Kim kicked the ball
(4) The ball was kicked by Kim

Similarly, sentences with conjoined subjects, as in (5), could be derived from two kernel sentences with the same predicate (6).

(5) Kim and Chris fell.
(6) Kim fell and Chris fell.

Gleitman (1964) observed that the latter analysis would fail for symmetrical predicates like *meet*. The ungrammaticality of (8) indicates that it could not be the source for (7).

(7) Kim and Chris met.
(8) * Kim met and Chris met.

Lakoff and Peters (1964) proposed to solve this problem by treating *meet* (and similar predicates) as applying collectively to its plural subject. If certain predicates were inherently symmetrical, and hence would apply to their subjects as a group, then there would be no need to derive (7) from (8).

Tversky (1977) revived the issue in a somewhat different context when he argued that the concept SIMILAR was not symmetrical. By symmetry in this context, Tversky meant that R(x,y) entailed R(y,x) and vice versa. Thus, if if Kim was similar to Chris, then Chris was similar to Kim. However, Tversky observed that these entailments are not always readily detected. Experimental participants judged (9) and (10) to have different meanings:

(9) North Korea is similar to China
(10) China is similar to North Korea.

Gleitman et al (1996, chapter 21 here) took up this challenge, arguing that while the judgments were robust and could be found in a whole range of symmetrical predicates, the proper interpretation of these effects was not to abandon symmetrical concepts. Rather, the solution was to partition the judgment into two parts: the symmetrical part that came from the predicate, and an asymmetrical part that could be attributed to the syntactic structure. The idea was that the asymmetry between subject and object that holds in the general case of transitive clauses carried over to color the interpretation of symmetrical relations. Thus, the perceived asymmetry between (7) and (8) came not from the predicate, but from the syntactic structure. This idea was supported by showing that even if the arguments were replaced by nonsense words, participants made judgments indicating that the subject was the figure to be compared against the object/ground. When presented with sentences like *the zum is identical to the gax*, participants judged the gax to be older, more famous, bigger and more important. Thus, the judgments of asymmetry came from the syntactic position and not from the meanings of the words themselves. Here again, splitting the explanatory pie in a way acknowledges the multiple contributors to a given behavior leads to a more satisfying explanation than one which takes the behavior at face value.

But, observing that the asymmetries observed in symmetrical predicates came from the syntax leaves open the question of how we know which predicates are symmetrical to begin with. How do we learn that predicates are symmetrical if their use in transitive clauses typically conveys asymmetry? To address this question, Gleitman et al (2019, chapter 22) examined the development of Nicaraguan sign language. This signed language emerged over the last 50 years in a school for children born deaf to hearing parents in Nicaragua. Because there was no signed language in Nicaragua, the children who attended this school developed their own language spontaneously. Gleitman et al compared NSL descriptions of possibly symmetrical events (e.g., two people high-five each other) which physically very similar events that might have been construed asymmetrically (two people punch each other). In NSL, there is a constraint against having two equally animate Noun Phrases as subject and object of the same verb/clause. Thus, to say that John punched Bill requires two verbs, roughly *John Bill punched got-punched*. So, if symmetrical predicates were understood not as involving two events, but rather as a single event where the two participants participated collectively, then these predicates should not require this construction. This is what they found. NSL signers produced multiple verbs for asymmetrical transitives to describe reciprocal action, but only a single verb for symmetrical predicates. Because this language was not learned from the environment, but rather was invented by its first generation of speakers, this shows that the conceptual difference between symmetrical and asymmetrical predicates is part of the basic logic with which we approach language and the world.

Where does grammatical knowledge come from?
With the observation that children's knowledge cannot be equated with their performance, the question of what that knowledge is and how it arises becomes more pressing. In an important and obvious sense, knowledge of language comes from the environment. Children exposed to English learn English, those exposed to Mandarin learn Mandarin, and so on for the 8000 or so languages on the planet. But this simple observation belies the complexity of the task because children's knowledge goes far beyond the simple corpus of sentences they are exposed to. And, as Chomsky (1965) emphasized, echoing Quine 1960 and Goodman 1955, any set of sentences is compatible with a vast space of possible grammars. Speakers of a language come to have judgments about the form and meaning of uncounted sentences that they have never been exposed to. This gap between what children are exposed to and what their grammars ultimately represent has come to be known as the poverty of the stimulus (Chomsky 1971).

In the 1970s the idea arose that parents shaped the corpus that children were exposed to by limiting what they said and how they said it. The hope was that this restricted form of child directed speech, what Henry Gleitman dubbed "motherese", would by itself allow children to successfully navigate the space of possible grammars and land on the correct one. The specific mechanisms by which motherese might shape grammatical knowledge in a way that would divert children from making incorrect generalizations have never been fully specified. How *could* restricting the corpus that children were exposed to lead them to make just the kinds of generalizations that would cover those parts of the grammar that were unsampled by this corpus? But despite this unclarity, the motherese hypothesis carried quite a lot of currency (Snow 1972, Bruner 1974). In all likelihood, the idea was popular because it offered a way around the idea that an innate grammatical architecture offered a solution to the poverty of the stimulus problem.

It is at this point that Newport, Gleitman and Gleitman (1977; chapter 6 here) appeared, providing the richest analysis at the time of the link between parental speech and developing child speech. If parents were implicit tutors, then we would expect variation in their speech to predict children's learning path. The results were not encouraging. For the most part, parents' speech did not predict children's linguistic growth. Variation across mothers was not predictive of variation in learning rates for various aspects of grammar across children. In short, properties of maternal speech were not shaped by an implicit desire to direct the learning path, but rather by the need to communicate with the child and get her to do what the mother wanted. The conclusion of this work, then, was that internal properties of children have more impact on grammar development than maternal speech does.

The idea that children contribute more to language acquisition than the environment was further amplified in the work of Feldman, Goldin-Meadow and Gleitman (1978; chapter 7 here). Here, Lila and her colleagues examined the development of home sign systems – communication systems created in the home by deaf children who were not exposed to a signed language. The key finding here was that the gestural communication systems developed in these homes were richly structured in ways that mirrored universal aspects of grammatical structure. For example, they found evidence for the separation of subject and predicate, revealed in the likelihood of the sign expressing the agent of an action. Specifically, the number of terms found inside the predicate predicted the likelihood of the subject appearing. As the predicate phrase got longer, the agent/subject was increasingly likely to be dropped. Ditransitive verbs like *give* were less likely to express their agents than transitive verbs like *kick*, which in turn were less likely to do so than intransitive verbs like *sleep*. Because this structure could not have come

from the exposure language, since there was none, it must have come from internal to the child.

Children's capacity to go beyond their input was further explored in the domain of word meaning by Landau and Gleitman (1985; reviewed in chapters 8-9 here). They found that blind children learned the meanings of *look* and *see*, despite not having access to the visual modality. This led to the realization that learning the meanings of words isn't really about pairing up the sound with some part of the external environment. It dealt a powerful blow to the empiricist world view of Quine and Skinner. There was just no way these words were learned by simple association between the sound and the extralinguistic context of use.

In sum, Lila's work on these topics was aimed at showing the empirical shortcomings of structuralist and empiricist conceptions of linguistic behavior and its acquisition. Behavior, whether children's utterances or adults' prototype judgments, was a poor indicator of underlying knowledge. When it came to acquisition, the input is a poor guide to the acquisition of a language. Children are sometimes indifferent to properties of the environment and can build linguistic structure even without a language model.

Bootstrapping

Now Lila was in a bit of a pickle. She had played a major role in highlighting the wrongheadedness of empiricist approaches to language acquisition, but a theory of how the child exposed to English actually acquired English remained at some distance. We had learned that children bring substantial resources of their own to language acquisition, that they were indifferent to some aspects of experience and when there was no experience, they nonetheless acquired structures and meanings that obeyed basic principles of language organization. Here, is where Lila made her most serious contributions to the field, as she developed positive proposals about how children use the basic principles of language organization to guide their acquisition of a particular language.

Pinker (1984) took the first big swing at this problem. He offered a picture in which learners took advantage of principled mappings between form and meaning. By making observations about meaning and surface form, learners could deduce aspects of the abstract grammar that connected these together. For example, upon seeing an event with an agent and a patient, and hearing a transitive clause, children could use their knowledge of a principle mapping agents to subject position to identify which Noun Phrase was the subject and to then look for other surface features (e.g., word order, agreement, etc) to determine how subjects were marked in the language.

There was just one problem with this approach, that Lila identified. It relied on the child knowing the meanings of the sentences without knowing the meanings of the words. But the meanings of sentences are not the kind of thing a learner can observe directly. Sentences present a perspective on the world, and a language learner cannot know what perspective the people around them are taking. The very same event could be described with any of the sentences in (11a-f), depending on the perspective the speaker is taking on the event (to say nothing of sentences like (11g), which might also be triggered by this event).

(11) a. April is walking quickly down the street.
b. April is going home.
c. April is late.

d. I think April is late for dinner.
e. April wants to get home soon.
f. They expected April to be home already.
g. Remember when we went sailing?

The problem is made most apparent by pairs of verbs that seem to describe the same events in the external world. Every scene that can be described as a selling, could also be described as a buying; every scene that can be described as a chasing can also be described as a fleeing. And such pairs of events have distinct agents and patients. The character fleeing is every bit as agentive as the character chasing. Thus a pairing of a sentence and a scene does not determine the perspective on the event, or the thematic relations to that event. Consequently, a learning procedure based on the supposition that learners can observe the meanings of sentences could not be supported.

Landau and Gleitman (1985) highlighted a further problem with a semantic bootstrapping approach to the acquisition of syntax, namely, that blind and sighted children seemed to acquire both the grammar and the meanings of words in the same way. But surely, the experience of the external world, and the evidence for what events are being talked about are distinct in these two populations. The fact that this difference in nonlinguistic experience did not translate into differences in language acquisition, suggested further that such experience was not the primary driver of language acquisition.

So, if Pinker's semantic bootstrapping was untenable, an alternative was needed. Landau and Gleitman (1985) took an essentially opposite point of view, reviving in a modern context Harris' idea that meaning and distribution were related. Unlike Harris though, they recognized, from the work of linguists like Fillmore (1970) and Jackendoff (1972), that meaning and syntax were systematically related. Patterns of distribution were related to whole classes of words with similar meanings. And given these principled relations, learners could use the syntax as one piece of evidence about the meanings of the words.

As noted above, Landau and Gleitman (1985) showed that a blind child learned the meanings of *look* and *see*, despite lacking visual inputs. Instead, they seemed to map these words onto the haptic domain. Specifically, *look*, but not *see*, was used for haptic exploration. So, they reasoned that these verbs would be used by the child's parents when the object to be apprehended was near to the child, and hence available for manual inspection. But dividing verbs along this situational dimension did not distinguish these verbs from other verbs that had nothing to do with perception. However, if you first used the syntactic distributions to distinguish perception verbs from other verbs (using features like compatibility with a clausal complement), then the near/far distinction became informative. *Look* was used much more often when the object was nearby than *see* was. Thus, the syntactic context provided the information that the verbs had something to do with perception and the extralinguistic context provided further information about distinguishing perceptual exploration from mere perceptual apprehension. Thus, armed with knowledge of how meaning projects onto syntax, a learner could use the syntax of a word to zoom in on a specific semantic domain and then use the extralinguistic context to further specify the word's meaning. The extralinguistic context provided information only after the initial syntactic/semantic partition.

This work began an industry of work (sampled in chapters 9-14) showing that syntax provided an important information source about word meaning and that adults, children and even infants

could use that information in the acquisition of a range of meanings across a range of languages.

Syntactic bootstrapping was actually two ideas. The first idea was that syntax was a constraining factor in narrowing reference. Given the lack of information in a scene by itself about a verb's meaning, the syntactic structure of a clause could direct a learner to have a particular perspective on the scene. For example, the image in figure 1 (From Nappa et al 2009) might be described as in (12a) or (12b).

(12) a. The bunny is chasing the elephant
b. The elephant is fleeing the bunny

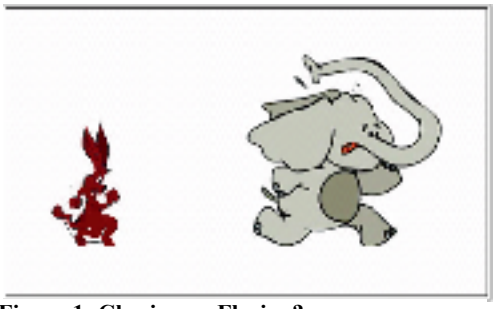

Figure 1: Chasing or Fleeing?

As physical events, chasings and fleeings are indistinguishable. Where they differ is in the perspective taken towards the physical event. Thus, a transitive clause that identifies the agent and patient as subject and object can provide a learner with the appropriate perspective on the event to make the clause and verb meanings more accessible. Hearing a sentence like (12a) would require a perspective on the event that makes the rabbit the agent. Consequently the verb in that sentence should describe the event as a chasing. By the same token, hearing (12b) would provide a perspective in which the Elephant is the agent, hence a fleeing but not a chasing (Nappa et al 2009, Fisher et al 1994).

More generally, learners are isolated from the perspective of the speakers around them with respect to sentence and verb meaning. And syntax provides a source of evidence to relieve that isolation. And, much work has since demonstrated that children are indeed sensitive to the perspectival information that can be conveyed in a syntactic frame (Gleitman 1990, Chapter 9 here; Naigles 1990; Fisher et al 1994; Lidz, Gleitman & Gleitman 2003, chapter 12 here; see Fisher et al 2010 for review).

The second way to see syntactic bootstrapping is not with respect to reference, but to distribution. The idea was that the set of frames that a verb occurs in provides information about its semantics. A single frame paired with a scene might provide some information on what event is being talked about, but this information is weak; finer grained aspects of verb meaning can be seen only in patterns of distribution. Fisher, Gleitman and Gleitman (1991, chapter 10 here) showed that these patterns are predictive of meaning. They asked English speakers to make two kinds of judgments - meaning similarity judgments and syntactic acceptability judgments. They found that patterns of acceptability of verb-frame pairings were significantly correlated with judgments of verb similarity. Using a relatively small number of verbs and a fairly coarse-

grained notion of syntactic frame, they found that verbs with similar meanings showed similar distributions. Recently, this work has been extended to show that even within the narrow class of mental state verbs, syntactic distribution is a strong predictor of semantic similarity (White et al 2018). This is shown in figure 2, where acceptability of a verb-frame pair is indicated by the darkness of the cell.

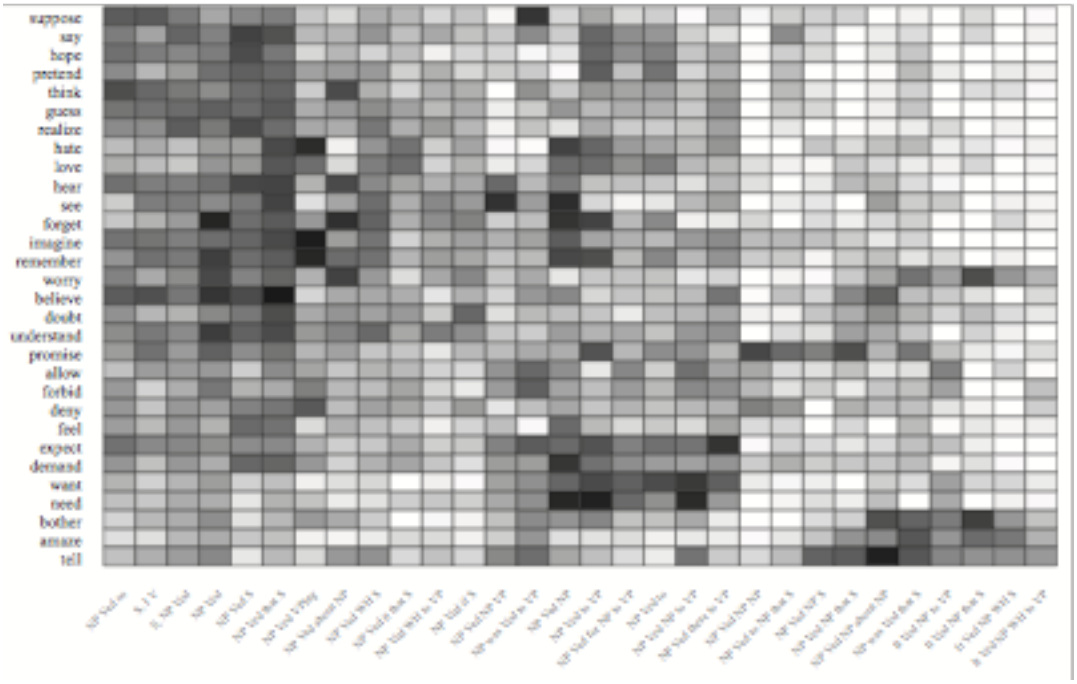

Figure 2: Verb-frame acceptability ratings. Darker shades indicate higher acceptability. Verbs with similar meanings occur in similar frames. From White et al, 2018.

These two ideas about syntactic bootstrapping are linked together by the fundamental idea that the less information there is in the world about a verb's meaning, the more information will be found in the syntactic distribution (Snedeker & Gleitman 2004). At the far extreme are mental state verbs, whose meanings are hidden from observers because their contents reside inside the mind. For such verbs, the syntactic environment is highly informative both as evidence about the verb in a single sentence and about the verb across many different sentence contexts (see chapters 10 and 13).

More generally, this work presented a perspective on learning that diverged enormously from the structuralist and behaviorist assumptions of Lila's youth. The observation that language is learned from those around them invites the conclusion that children learn the patterns by observing the patterns. If you are learning meaning, then you make observations about meaning. If you are learning syntax, then you make observations about syntax. The perspective offered by syntactic bootstrapping represents a radical departure from this view of learning because it involved cross-domain inferences. To learn about meaning, you could make observations about syntax. This perspective, separating the properties of the input from the conclusions we draw about the language generating that input, has now permeated the field of developmental psycholinguistics. To learn about syntax, you could make observations about semantics (Pinker 1989), or prosody (Christophe et al 2008) or even word frequency (Hochmann et al 2010).

Such cross-domain inferences are possible only in a mind that has rich expectations about how levels of analysis relate to each other. This kind of learning mechanism implies a rationalist perspective on the mind in which the point of learning is not to match the data, but to explain it. A learner makes observations, in this case about sentences, and makes inferences about the system/grammar that produced those observations. We explain the syntactic distribution by positing meanings that would give rise to that distribution, given general principles mapping between form and meaning. These kinds of learning mechanisms are simply incompatible with the structuralist and empiricist perspectives in which Lila initially found herself. Lila's ideas provided developmental psycholinguists with a whole new approach to thinking about learning that recognized the importance of both distributional analysis of the environment and the prior structure of the learner.

Association and Epiphany

The idea of syntactic bootstrapping first arose in response to the observation that the world provides little information about the meaning of a word. As Quine (1960) famously observed, the extralinguistic environment accompanying the use of a word leaves open the particular concepts in the mind of the speaker that condition that use. A situation in which a speaker uses a word to refer to a rabbit might be identical to one in which he intended to refer to the rabbit's fur, the speed at which it is moving, or to the memory of a delicious stew. Indeed, the fact that we are free to talk about anything that crosses our minds and not just our immediate environment, makes a learning procedure based on pairing words with the contingencies of use of questionable utility (Chomsky 1959). Finally, even if a speaker is talking about the here and now, and even if a learner could pinpoint precisely the part of the world being picked out by a novel word, recognizing, for example, that the speaker was referring to the rabbit, this reference would be consistent with a broad range of concepts with overlapping extensions (e.g., Pete the rabbit, rabbits or black holes, rabbits with more than one ear, things physically identical to rabbits, etc.; Goodman 1955).

Linguists and cognitive scientists have responded to these philosophical problems by noting that humans have perceptual, conceptual and linguistic abilities that can help. These abilities lead learners to parse the world into the same pieces and to relate those pieces to the same concepts as the speakers producing those words (Gleitman 1990, Spelke 1990, Markman 1990, Waxman & Lidz 2006, Carey 2009). Moreover, these conceptual burdens may be significantly reduced by learners' ability to track the goals and intentions of their interlocutors (Baldwin 1991, Bloom 2001, Clark & Amaral 2010).

But even granting learners these constraining capacities, the world still vastly underdetermines the concepts that learners might invoke in explaining the use of a novel word. The insufficiency of the world is perhaps made most clear in the Human Simulation Paradigm (Gillette et al 1999, chapter 11 here). In this paradigm, adults are shown videos of parent-child interactions with the sound turned off and are then asked to guess what word the parent said at a particular point in the video [even when the words are selected from the 25 most frequent Nouns and Verbs that children hear]. Adults were remarkably bad at this task. For nouns, participants guessed the correct word 44% of the time on average (range: 3.6-100%) and for verbs, they guessed correctly only 15% of the time (range: 0-86%).

One might expect that the weakness of the world as an information source could be overcome by multiple exposures, over which the learner could find a common thread. Pull on that thread and the meaning will be revealed. However, participants in these experiments actually had 6

attempts to make their guesses, and improvement across trials was very weak. This lack of improvement suggests that integrating information across occurrences does not help much. Indeed, the lack of similarity across contexts led to a significant rise in guesses like "toy" or "look" as the trials progressed. One might expect, then, that a learning procedure based in cross-situational comparison would end with every word having an extremely weak meaning so that all contexts would fall under the word's extension.

In more recent work (chapters 15-18), in collaboration with John Trueswell and others, Lila has shown not just that cross-situational comparison is a bad idea, but also that learners just don't do it. The idea of cross-situational word learning is that learners retain associations between a word and all possible referents for that word, with the expectation that the true association will emerge over many exposures. This kind of theory predicts that learners gradually zero-in on a word's meaning, getting closer and closer across time. Medina et al (2011; chapter 15 here) looked for precisely this improvement in an experimental context. Participants were shown natural scenes in which a parent uttered a word to their child and were asked to guess the word's meaning after each scene. They found no improvement across trials, and hence no evidence of cross-situational comparison. Instead, their data was more consistent with a theory in which learners guessed a meaning on the first trial and then sought to confirm that guess.

This theory was further tested in Trueswell et al (2013; chapter 16 here). Participants were shown several images while they heard a novel word and were asked to guess the meaning of the novel word. A cross-situational learner would predict gradual improvement across trials and would predict no real difference in the information retained between trials in which learners guessed correctly and those in which they guessed incorrectly. These researchers found that if participants had guessed correctly, then on subsequent trials, they were above chance at choosing the correct object. However, if they guessed incorrectly, they remained at chance, showing no evidence of having retained relevant information about the context. In sum, learners did not show any evidence of retaining information about un-chosen candidate word meanings. In all of these manipulations, the key finding was that learners do not retain lots of information about the context of use. Instead, they make a targeted guess about the word's meaning and forget everything else about the context.

Finally, Cartmill et al (2013; chapter 17 here) shows that children's vocabulary growth is not determined simply by the number of words they hear, but also by the quality of the contexts in which they hear them. Repeated exposure to a word does generally lead to higher likelihood of acquiring it. But this effect is not due to strengthening association. Rather the repeated exposures mean that the learner has a higher chance of encountering the word in a context where its meaning is easy to guess.

While much work continues to be done in this area, the overall lesson is the failure of simple associative learning mechanisms even in the domain where they would seem to have the best chance of success. Learners do not simply pair up words with experiences and let these pairings drive word learning. Instead, they actively make guesses and then check the accuracy of those guesses over time. And these guesses are driven by sophisticated mechanisms that allow learners to have a sense of whether a given situation warrants a guess. Learning a word's meaning does not begin gradually; rather it begins with an epiphany.

Conclusions

In this introductory chapter, I have tried to put Lila's work in its historical context, highlighting the intellectual climate in which she first started working and taking us through the key insights that

have emerged from the body of work she has produced. In doing so, I hope to have conveyed the distance that the field has come in its understanding of language and its acquisition. Lila's work was the first to demonstrate experimentally the distance between grammatical knowledge and linguistic behavior, in syntax and semantics, and in children and adults. Her work provided some of the most creative demonstrations of children's partial indifference to their linguistic environment, from their insensitivity to parental speech to learning in the blind and deaf. And she showed that when children are sensitive to their environment, they capitalize on correlations between levels of linguistic analysis in order to acquire syntax and semantics. And, finally, she has given us new ways to think about how words and their referents come to be associated. The anti-mentalist milieu in which Lila first emerged is overturned by her work, replaced by a view of linguistic structure and its learning in which learners use prior knowledge of how languages might be to determine how the particular language they are learning actually is. The intellectual import of this work cannot be overstated as it continues to offer insights not just for historians of the field but also for current practitioners.

References

Baldwin, D. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, 62, 875–890.

Bengio, Y., R. Ducharme, P. Vincent & C. Jauvin. A neural probabilistic language model. *Journal of Machine Learning Research*, 3(Feb), 1137-1155.

Bloom, P. (2001). *How children learn the meaning of words*. Cambridge, MA: The MIT Press.

Bloomfield, L. (1936). Language or ideas? *Language*, 12, 89-95.

Bruner, J. (1974). From communication to language - a psychological perspective. Cognition, 3(3), 255-287.

Carey, S. (2009) *Origins of Concepts*. Oxford University Press.

Chomsky, N. 1957. Syntactic Structures. Mouton: The Hague.

Chomsky,N. 1959. Review of Verbal Behavior by B.F. Skinner. *Language*, 35(1), 26-58.

Chomsky, N. 1971. Problems of Knowledge and Freedom. Pantheon: NY.

Christophe, A., S. Millotte, S. Bernal & J. Lidz (2008) Bootstrapping Lexical and Syntactic Acquisition. *Language and Speech*, 51, 61-75.

Clark, E. V. & Amaral, P. M. (2010). Children build on pragmatic information in language acquisition. *Language & Linguistics Compass*, 4(7), 445–457.

Fillmore, C. 1970. The grammar of hitting and breaking. in R. Jacobs & P. Rosenbaum (eds). *Readings in Transformational Grammar*, 120-133. Ginn: Waltham, MA.

Frege, G. 1884.The foundations of arithmetic. Breslau: W. Koebner. Translated by J. L. Austin, 1974. Oxford: Blackwell.

Goodman, N. (1955). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Harris, Z. 1951. *Methods in structural linguistics*. Chicago: University of Chicago Press.

Hochmann, J.R., A. Endress & J. Mehler (2010) Word frequency as a cue for identifying function words in infancy. *Cognition*, 115, 444-457.

Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.

Landau, B. & Gleitman, L. (1985). *Language and Experience. Evidence from the blind child.* Cambridge, MA: Harvard University Press.

Landauer, T.K. & S.T. Dumais 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, 14, 154–173.

Mikolov, T., K. Chen, G. Corrado & J. Dean (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:*1301.3781.

Pinker, S. 1989. *Learnability and Cognition*. Cambridge: MIT Press.

Rey, G. 2003. The analytic/synthetic distinction. in E. Zalta (ed.). *The Stanford Enyclopedia of Philosophy*. Stanford: Metaphysics Research Lab.
Quine, W.V.O. (1951). Two dogmas of empiricism. *Philosophical Review*, 60(1): 20–43.
Quine, W.V.O. 1960. *Word and Object*. Cambridge: MIT Press
Snow, C. 1972. Mothers speech to children learning language. *Child Development*, 43, 549-565.
Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, 14(1), 29–56.
Waxman, S.R. & Lidz, J. (2006). Early word learning, in Kuhn, D. & Siegler, R. (Eds). *Handbook of Child Psychology: Cognition, Perception and Language*. New York, NY: John Wiley & Sons, 299–335.