Lidz, Waxman, and Freedman (2003) claim to have demonstrated that 18-month-old children know something about the syntax of noun phrases that they could not have learned from the linguistic input. Invoking the poverty of the stimulus argument, they conclude: “Our results provide clear support for the argument that the learner’s innate linguistic structure guides language acquisition.” (p.8)

What young children supposedly know is that in a sentence like

*I’ll play with this red ball and you can play with that one*

*one* refers not just to another ball of any type but to a red one. Someone who makes this linking thus knows that noun phrases have a nested structure [this [red [ball]]] rather than a flat structure [[this] [red] [ball]]. In searching through a corpus of child-directed speech, Lidz et al. found very few instances of utterances similar to the above example, that is, with a noun phrase containing an adjective as antecedent to the indefinite pronoun *one*. In most instances the antecedent was a determiner–noun sequence or pronoun with no adjective (e.g., *I’ll play with this ball and you play with that one*), which are not diagnostic. So children supposedly have no input that enables them to figure out, in the above example, that *one* refers back to “red ball” (N⁰) not just “ball” (N⁰).

To demonstrate that children nevertheless have such knowledge, Lidz et al. performed a preferential looking experiment with 18-month-old infants. In each of four trials, there was a familiarization phase followed by a test phase. In the familiarization phase:

an image of a single object (e.g. a yellow bottle) was presented three times...accompanied by a recorded voice that named the object with a phrase consisting of a determiner, adjective and noun (e.g. Look! A yellow bottle.)
Then in the test phase infants were presented with two pictures on opposite sides of a TV screen: one of a blue bottle and one of a yellow bottle, accompanied by a recorded voice saying: *Now look. Do you see another one?*

The finding was that in the test phase the infants looked longer at the yellow bottle, that is, the bottle that was the same color as the original object (whereas in a control condition, with neutral language at test, they looked more to the differently colored bottle). But note that the language here is very different from the original linguistic example above. It is different because the expression is now not just *one* but *another one*, an expression children experience many times in conjunction with certain nonlinguistic experiences (e.g. one grape and then another one). The current finding is that in a forced choice situation 18-month-olds think that *another one* goes best with another object almost identical to the one they have just seen (yellow bottle) rather than a differently colored one from the same category (blue bottle).\(^1\) This is an interesting finding, and it is not immediately clear why it should be so. But it has nothing to do with children’s understanding of the nested structure of noun phrases or innate linguistic knowledge; it has only to do with their understanding of the kinds of nonlinguistic experiences conventionally associated with the expression *another one* (i.e. the semantics of this expression). To put the point in the form of a challenge: if one were to do exactly the same experiment with no noun phrase during the familiarization phase (just attention-getters like *Look at this!*), the findings would be exactly the same. Children are matching *another one* to the visual stimulus that is “the same” across familiarization and test, with no attention to the noun phrase during familiarization at all.

The control experiment (designed to exclude a different alternative hypothesis) does not undermine this explanation. In that study children were presented with the same familiarization phase and the same two alternatives pictures at test, but in this case the language accompanying the choice at test was not *another one*, as in the main experiment, but either: (1) *another yellow bottle* (in which case they looked more to the yellow bottle), or (2) *another bottle* (in which case they looked equally to the two bottles). This shows simply that children of this age (1) associate the word *yellow* with yellow things, and (2) do not associate the phrase *another bottle* with the same-colored object across familiarization and test (as they do with the phrase *another one*). But again the noun phrase in the familiarization phase is very likely irrelevant in this study—children are just matching the language at test with what they are seeing and/or have seen.

And so, what the current studies tell us is how 18-month-old children react to *another one*, and similar phrases containing the word *another* when presented with one object followed by a choice of two other objects: one almost identical and one differently colored but from the same category. The findings are interesting and invite further inquiry from the point of view of children’s word meanings, categorization tendencies, and the like, but they do not inform us at all about their understanding of the syntax of noun phrases.

---

\(^1\) Note that I say “goes best with” rather than “means” (or similar words implying an interpretation) because in the preferential looking paradigm the child is not making an active choice of one object that excludes the other—such as pointing to one of the bottles—but merely looking 25% longer at one as compared with the other.
Discussion

Learning antecedents for anaphoric one

Nameera Akhtar\textsuperscript{a,}*\textsuperscript{,}, Maureen Callanan\textsuperscript{a}, Geoffrey K. Pullum\textsuperscript{b}, Barbara C. Scholz\textsuperscript{c}

\textsuperscript{a}Department of Psychology, University of California, Santa Cruz, CA 95064, USA
\textsuperscript{b}Department of Linguistics, University of California, Santa Cruz, CA 95064, USA
\textsuperscript{c}Department of Philosophy, San José State University, CA 95192, USA

Received 1 October 2003; accepted 16 December 2003

Abstract

Lidz et al. [Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn’t have learned: Experimental evidence for syntactic structure at 18 months. Cognition, 89, B65–B73.] claim experimental substantiation of an argument from the poverty of the stimulus, in the sense of Pullum and Scholz [Linguist. Rev. 19 (2002) 9]. They cite a specific feature of English—the assignment of appropriate antecedents for anaphoric one—that cannot possibly be learned from experience because the evidence needed is found only in utterances of a type too rare to be encountered. Their argument involves three empirical claims. In this note we dispute all three.

\textcopyright 2004 Elsevier B.V. All rights reserved.

Keywords: Language acquisition; Poverty of the stimulus; Linguistic input

1. Introduction

Lidz, Waxman, and Freedman (2003; henceforth LWF) claim to have provided experimental support for an argument from poverty of the stimulus (APS) in the sense of Pullum and Scholz (2002). Their conclusions (p. B72) are bold: that certain syntactic knowledge “must derive from linguistic structure inherent in the learners themselves because…the input to which infants are exposed does not unambiguously support the linguistic representations that they create,” which suggests that “the learner’s innate linguistic structure guides language acquisition”; specifically, “infants at the earliest stages of syntactic production share with mature speakers of English” a certain piece of...
specific information about noun phrase structure and the correct assignment of antecedents to anaphoric uses of *one*.

Standard analyses of the English noun phrase posit a category that we call a ‘nominal’, following Huddleston and Pullum (2002)—it is notated as N by LWF. The nominal is the constituent of the noun phrase containing the noun and its closest modifiers, but not the article. The knowledge infants are claimed to have at the earliest stages of learning is that there is such a constituent, and anaphoric *one* requires it as antecedent. This claim was first used as support for an APS by Baker (1978, see pp. 413–25; see also pp. 327–340).

A complete APS requires that two things be specified first: something that children are claimed either to acquire or to antecedently know (Pullum and Scholz call this the *acquirendum*) and a class of sentences claimed to be necessary for learning the acquirendum from evidence but unavailable (the *lacuna*). Given these, the crucial psycholinguistic work involves providing three kinds of empirical evidence:

(i) **Inaccessibility** evidence: The claim that learners do not encounter utterances of sentence types in the lacuna must be defended.

(ii) **Indispensability** argument: It must be shown that learning the acquirendum from experience would require exposure to lacuna sentence types.

(iii) **Acquisition** evidence: The claim that the learner does indeed end up knowing the acquirendum must be substantiated.

The acquirendum for LWF is the adult pattern of assignment of antecedents to anaphoric *one*, i.e. that anaphoric *one* has a nominal antecedent. The lacuna is the set of sentences that contain an instance of anaphoric *one* for which the antecedent is unambiguously a nominal, i.e. an antecedent that could not be just the head noun of the noun phrase. LWF defend claims corresponding to the necessary inaccessibility, indispensability, and acquisition demonstrations. We challenge all three.

2. **Inaccessibility**

Baker (1978) was concerned with how the infant rules out an incorrect hypothesis: that *one* takes simply a noun as antecedent. He believed it would take an ‘unusual set of circumstances’ (p. 416). For brevity, let us call an utterance a *Baker event* if and only if, in context, it is compatible solely with a nominal antecedent, not with a noun antecedent. LWF report finding just two Baker events among the 792 examples of anaphoric *one* in the Adam and Nina corpora of the CHILDES database, and offer this as their support for the inaccessibility claim. They hold that ‘the data that the infant would need in order to learn the syntax of *one* occur at a rate (0.2%) that is indistinguishable from noise in the input’ (p. B68). The rate is not in fact 0.2%, but almost 0.253% (this difference matters, because we have to multiply by large factors). There is published evidence that even in linguistically impoverished households, 3-year-olds have heard roughly 10 million word tokens in context, with MLU = 4, hence around 2.5 million sentences (Hart & Risley, 1995). We can assume that 18-month-olds might well have heard a million sentences. To estimate the number of Baker events this would expose them to, we proceeded as follows.
The number of utterances by adults in the Adam and Nina corpora is 60,769. (Utterances by the children are irrelevant, as they never contain anaphoric one.) They contain 792 occurrences of anaphoric one by LWF’s count, so that is about one per 77 utterances, hence about 13,000 per million utterances. LWF’s count suggests that 0.253% = 33 of these would be likely to be Baker events. It is not at all clear to us that exposure to what could well be several dozen Baker events can be so easily dismissed as irrelevant to learning from experience.

3. Indispensability

In addition to (perhaps) not being all that rare, Baker events may not be indispensable for learning the acquirendum. Pullum and Scholz (2002, p. 35) observe: “[I]f learners do sometimes have to figure out from primary data that one has a multi-word antecedent, it does not follow that the context must entail that the antecedent is multi-word. Mere pragmatic implication or conversational implicature will do just as well.” The point is important. Suppose you are told “Hand me a yellow bottle,” and then you are told, “Now give me another one.” How do you respond? Does one stand for the nominal yellow bottle, or merely bottle? The answer is surely that if the modifier yellow was not relevant in the first utterance it would have been omitted, and given no further information it should be assumed to have continuing relevance, so one should be taken to stand for yellow bottle. But there need be no indication in the context that rules out the possibility of the other readings. Probability, given the relevance principle (see Sperber & Wilson, 1995), is enough.

Consider in this light the following conversation (from Callanan & Sabbagh, in press) between a parent (P) and a 20-month-old child (C):

P: There’s fishes on the wall. Fishes on the wall, fishes in the bag.
C: Bag.
P: Bag.
C: (walks over and gets a toy out of bag) Eee go-go eee. Eee. Fis.
P: A very big fish. A whale.
C: Ohh! Buh? (hands mother the whale)
P: Thank you.
C: (gets another toy and brings it to mother) Bih? Fis?¹
P: There’s another one. It looks like a shark.

It is not clear how LWF would code this particular example, but in the context, given the pragmatic principle of relevance, the infant could plausibly infer that the antecedent of one is very big fish rather than just fish (see Bloom, 1998, on the function of relevance in language acquisition). The correctness of this antecedent does not need to be deduced with full logical certainty.

¹ The child was probably attempting to repeat ‘big fish’ here, but without familiarity with this child’s phonological system it is difficult to be certain.
Note that the principle of relevance goes beyond specifically linguistic concerns. Anaphoric one can have antecedents that are merely inferred from deictic referents: if *Hand me that* is correctly understood and the next utterance is *Now give me another one*, no syntactic constituent serves as antecedent at all; but the relevance principle still suggests that one has, say, *yellow bottle* as antecedent rather than just *bottle*. LWF eliminate from consideration all deictic uses of one. But such uses may be highly instructive to the learner. In fact the anaphoric use of one may evolve from the exophoric use (where the reference is to something extralinguistically indicated).

If exophoric uses are relevant, the class of relevant events could be vastly larger than the class of Baker events under the original narrow definition. If we want to know more about how actual learners acquire the use of one, we need to analyze videotaped corpora of interactions in a way that is sensitive to pragmatics, rather than focus exclusively on syntactic properties of audio transcripts as LWF do (and as Baker tacitly suggested).

4. Acquisition

Finally, LWF do not convincingly demonstrate that 18-month-olds know and represent the antecedent of anaphoric one as a nominal. In the familiarization phase of the main experiment, infants were shown an image of an object (e.g. a yellow bottle) accompanied by the utterance “Look! A yellow bottle.” In the test phase, they were shown a pair of images (e.g. two bottles: one yellow, one blue). In the experimental condition, this pair of images was accompanied by “Do you see another one?” Infants looked longer at the (familiar) yellow bottle. Infants in the control condition, who heard “What do you see now?”, looked longer at the (novel) blue bottle. This is an interesting finding, but by itself it does not show that infants know and represent the structure of the antecedent of one. First, the word *now* (especially if it was stressed) in the control condition may well have biased the infants to look at what was now new. And second, infants’ looking longer at the yellow bottle in the experimental condition can be explained by their tendency to prefer what is familiar (Barrile, Armstrong, & Bower, 1999). LWF are aware they need to rule out this familiarity hypothesis to establish their conclusion, but their control experiment does not do so.

To effectively rule out the familiarity hypothesis, one would need a control condition that differentiates its predictions from those of LWF’s favored hypothesis. One such condition would involve infants hearing “Look! A bottle” during the familiarization phase. LWF would predict that when infants later heard “Do you see another one?”, they would show equal tendencies to look at either of the two bottles. In contrast, the familiarity hypothesis predicts a preference for what was seen during familiarization. LWF’s control experiment merely shows comprehension of the adjective, not of the whole nominal.
5. Conclusion

LWF should be commended for attempting an empirical defense of the APS based on anaphoric one, for the many linguists who have reiterated Baker’s argument (see Pullum & Scholz, 2002, p. 32) have done nothing to support it empirically. But there are weaknesses in LWF’s arguments for all three of the crucial propositions: inaccessibility, indispensability, and acquisition. That does not refute the notion that children are innately equipped with representations of the linguistic structure of noun phrases, nor does it refute linguistic nativism in general. But it does mean that we still lack adequate experimental evidence for this notion—and more generally, that no fully developed and convincing instance of the APS has been provided.

References

Discussion

Learning the unlearnable:
the role of missing evidence

Terry Regier,*, Susanne Gahl

Department of Psychology, University of Chicago, 5848 South University Avenue, Chicago, IL 60637, USA
University of Illinois at Urbana-Champaign, Urbana, IL, USA

Received 24 October 2003; accepted 16 December 2003

Abstract

Syntactic knowledge is widely held to be partially innate, rather than learned. In a classic example, it is sometimes argued that children know the proper use of anaphoric one, although that knowledge could not have been learned from experience. Lidz et al. [Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn’t have learned: Experimental evidence for syntactic structure at 18 months. Cognition, 89, B65–B73.] pursue this argument, and present corpus and experimental evidence that appears to support it; they conclude that specific aspects of this knowledge must be innate. We demonstrate, contra Lidz et al., that this knowledge may in fact be acquired from the input, through a simple Bayesian learning procedure. The learning procedure succeeds because it is sensitive to the absence of particular input patterns—an aspect of learning that is apparently overlooked by Lidz et al. More generally, we suggest that a prominent form of the “argument from poverty of the stimulus” suffers from the same oversight, and is as a result logically unsound.

Keywords: Language acquisition; Syntax; Innateness; Poverty of the stimulus; Emergence; Bayesian learning; Indirect learning

One of the core questions of cognitive science is whether human language relies on innate syntactic knowledge. On one influential view, at least some aspects of syntax must be innate, since the child possesses syntactic knowledge that could not have been learned from his or her impoverished linguistic input (Chomsky, 1981; Pinker, 1989). While this “argument from poverty of the stimulus” has generally met with wide acceptance, it has also recently been challenged. A growing number of researchers have suggested that

* Corresponding author.
E-mail address: regier@uchicago.edu (T. Regier).
the child’s linguistic input may suffice to allow general-purpose learning mechanisms to acquire syntactic regularities, without the benefit of specifically syntactic innate knowledge (Christiansen & Chater, 1999; Elman, 1993; Pullum & Scholz, 2002; Rohde & Plaut, 1999; Seidenberg, 1997; Tomasello, 2000).

Lidz, Waxman, and Freedman (2003) respond to these challenges, with an empirical investigation of young children’s syntactic knowledge and linguistic input. They conclude that specific aspects of children’s knowledge are not learnable from the input—and therefore must be innate.

We suggest that Lidz et al.’s innatist conclusion does not follow from their data. We support this claim by demonstrating that a simple Bayesian learning model can account for their findings, without the innate knowledge they propose. We suggest that the flaw in the argument is not theirs, however; rather, it was inherited from the “poverty of the stimulus” tradition. Like some (but not all) earlier work in this tradition, Lidz et al. overlook the fact that much may be learned by noting which patterns are absent from the input. Ultimately, we suggest that the argument from poverty of the stimulus is unsound if the role of missing evidence is ignored.

1. The syntax of anaphoric one

Lidz et al. (2003) approach the broad question of innate syntactic knowledge by examining a specific phenomenon: the anaphoric use of one, as in sentence (1).

(1) I’ll play with this red ball and you can play with that one.

Here, the word one refers anaphorically to red ball. Following Hornstein and Lightfoot (1981), the authors suggest that such sentences implicitly pose a learning problem concerning the structure of the antecedent noun phrase—here, this red ball. Such a noun phrase could in principle be analyzed in at least two ways, as illustrated in Fig. 1.

Lidz et al. argue that the nested structure must be the correct one: it is commonly assumed that an anaphoric element may substitute only for a constituent; here, one refers

(a) Flat structure (b) Nested structure

![Diagram of noun phrase structures](image)

Fig. 1. Two possible structures for noun phrases of the form determiner–adjective–noun, such as “this red ball.” (Adapted from Lidz et al., 2003)
anaphorically to red ball, which appears as a constituent (i.e. a node containing that string and nothing else) only in the nested structure, as the upper N. Given this assessment of adult grammar, Lidz et al. then ask how children could arrive at this knowledge. Their core learning problem is reproduced here, and marked (†) for future reference:

(†) Suppose that a learner is exposed to small discourses like [(1)] in which one is anaphoric to some previously mentioned discourse entity and that the learner has recognized that one is anaphoric. In order to understand this use of one, the learner must know that it is anaphoric to the phrasal category N, which is possible only under the nested structure hypothesis. However, the data to support this hypothesis are not available to the learner for the following reason. Every situation that makes one = [N red ball] true also makes one = [N ball] true [since any actual red ball, to which one might refer, is also a ball—TR & SG]. Thus, if the learner had come to the flat structure hypothesis or to the hypothesis that one is anaphoric to N and not N, evidence that this is wrong would be extremely difficult to come by. (p. B67)\(^1\)

Lidz et al. (2003) proceed to empirically demonstrate two points. The first is that children’s input almost completely lacks a particular form of evidence that, if present, would lead children to the correct hypothesis (i.e. [N red ball]). Children must instead learn on the basis of evidence such as (1), which is consistent with both correct and incorrect hypotheses. The second point is that children near the beginning of language learning nonetheless behave as if they know the correct hypothesis. Thus, the authors argue, children know the correct use of anaphoric one without ever having encountered evidence that would allow them to acquire that knowledge. They conclude that aspects of that knowledge are not acquired, but rather innate: children never entertain the incorrect hypothesis that one = [N ball] (p. B67).\(^2\)

The general logic is that of the argument from poverty of the stimulus (APS). However, the APS is not a single argument, but rather a family of related arguments (Pullum & Scholz, 2002). The specific version of the APS deployed by Lidz et al. makes a critical assumption: that evidence consistent with multiple hypotheses cannot discriminate among those hypotheses; for instance, that sentences such as (1) cannot discriminate between [N red ball] and [N ball]. This assumption may also be found in some earlier presentations of the APS (e.g. Baker, 1978:416; Hornstein & Lightfoot, 1981:18–20; Pinker, 1989:6).

This assumption is incorrect. Given evidence that is consistent with several hypotheses, a learner can come to discriminate among them, for principled, domain-general reasons. In particular, if one of the hypotheses predicts not only the input that is seen, but also input of

---

\(^1\) On this argument, the hypothesis one = [N red ball] (the lower N in the nested structure) should also be difficult to disconfirm, given only data such as (1). Our unverified intuition is that one = [N red ball] is false when (1) is spoken with neutral prosody, but true when it is spoken with emphasis on the word red, implying a contrast with the other (non-red) ball. Anaphoric one also refers to the lower N in a variety of other sentences. In this article, we restrict attention to sentences in which the correct hypothesis is the upper N rather than the lower N—as do Lidz et al. A fuller treatment of the acquisition of anaphoric one would need to also cover those cases in which one refers to the lower N.

\(^2\) We use the notations [N ball] and [N red ball] without loss of generality, to also refer to analogous structures in sentences with different adjectives and nouns.
another sort that is never seen, that absence can serve as evidence against the hypothesis. Some presentations of the APS explicitly consider this possibility (e.g. Chomsky, 1981:9).

How does this idea apply to anaphoric one? The hypothesis one = [N red ball] predicts that the referent of one will be red—this prediction is always confirmed. In contrast, the hypothesis one = [N ball] predicts that the referent of one will be a ball of any color—and thus we should expect that at least some of its uses will refer to balls that are not red. This will never happen since the hypothesis is incorrect. Thus, our expectations for this hypothesis are not fully met: we do not see the full range of expected referents. This absence can serve as implicit negative evidence against one = [N ball].

The same general intuition underlies Laplace’s (1825) law of succession, which may be used to estimate the probability that the sun will rise tomorrow, given evidence that it has risen every morning for some number of days. These successive sunrises are consistent with the correct hypothesis, namely that the sun rises without fail every day and therefore will tomorrow as well—but they are also consistent with false hypotheses such as that the sun has only a 50% chance of rising on any given day: each time we saw it rise, it may have just happened to rise. But on this latter hypothesis, we would also expect the sun to not rise on at least some of the days we have observed. This expected evidence never appears, and its absence undercuts the false “50%” hypothesis, despite the fact that the observed data are compatible with that hypothesis.

This mathematical parallel suggests that formal models may be helpful in addressing the role of absent evidence in the argument from poverty of the stimulus. It also suggests that there may be nothing particularly linguistic about the learning processes involved—the learning may fall out of domain-general considerations. We now turn to explore this possibility.

2. Models of indirect learning

Recently, a number of computational learning models have pursued probabilistic approaches to language learning (e.g. Brent & Cartwright, 1996; Eisner, 2002; Niyogi & Berwick, 1996). In addition, several learning models have demonstrated that indirect evidence may shape learning (e.g. Landauer & Dumais, 1997; Merriman, 1999; Regier, 1996; see Regier, 2003 for a review). The Bayesian learning model of Tenenbaum and Griffiths (2001) is particularly relevant for present purposes. This model and variants thereof have accounted for Shepard’s (1987) roughly exponential generalization gradient (Tenenbaum & Griffiths, 2001), and several aspects of word-learning: learning from a small number of examples (Tenenbaum & Xu, 2000), the interaction of syntactic and semantic knowledge (Niyogi, 2002), and lexical contrast (Regier, 2003). Critically, this model formalizes the idea of learning from the absence of expected data.

Tenenbaum and Griffith’s model assumes that learning is rational, governed by the normative standard of Bayes’ rule:

\[ p(H|e) \propto p(e|H)p(H) \]

---

3 This idea is distinct from the “subset principle” (Berwick, 1986; Pinker, 1995:172–175), which holds that children do not consider broad hypotheses until the input requires them to do so. We suggest, in contrast, that children initially consider all hypotheses, and learn to discard overly broad ones.
Here $H$ is a hypothesis in a hypothesis space, and $e$ is the observed evidence. Bayes’ rule determines the probability of each hypothesis in the hypothesis space given the observed evidence, as a function of the likelihood $p(e|H)$ (that is, the probability of observing that evidence, given that the hypothesis is true), and the prior probability $p(H)$ (that is, the a priori probability of that hypothesis being true).

At the heart of Tenenbaum and Griffiths’ model is their “size principle”. This principle holds that the likelihood of seeing a particular sort of evidence $e$, given that hypothesis $H$ is true, can be determined by considering the full range of different sorts of evidence that $H$ could give rise to, and assuming that $e$ was randomly selected from this set. Formally, if $H$ can give rise to $|H|$ different sorts of evidence, and “$e \in H$” means that $e$ is one of those sorts, then:

\[
(3) \quad p(e|H) = \begin{cases} 
\frac{1}{|H|} & \text{if } e \in H \\
0 & \text{otherwise}
\end{cases}
\]

This principle causes the likelihood $p(e|H)$ to be largest for those hypotheses that support the smallest range of possible evidence. For example, if we observe a dog barking, this principle would support the hypothesis “only dogs bark” more strongly than it would support the also-consistent hypothesis “animals of all sorts bark.” This follows since the hypothesis “animals of all sorts bark” can give rise to a broader range of possible evidence, making it less likely that a random selection from that range would yield a barking dog. In the absence of additional evidence such as a barking cat, this hypothesis will lose support. It is in this sense that the model formalizes the notion of learning from the absence of expected evidence.

Tenenbaum and Griffiths obtain the likelihood $p(e^n|H)$ of observing $n$ occurrences of evidence $e$ by assuming the observations are independent, and multiplying the likelihood for a single observation $n$ times:

\[
(4) \quad p(e^n|H) = \begin{cases} 
\frac{1}{|H|^n} & \text{if } e \in H \\
0 & \text{otherwise}
\end{cases}
\]

We may now reconsider the argument from poverty of the stimulus as applied to anaphoric one (†), through the lens of this model. We consider four hypotheses, each of which represents a possible node to which anaphoric one might refer. The initial word of each hypothesis’ name (“nested” or “flat”) indicates the structure from Fig. 1 in which the node resides. The hypotheses are:

i. nested: $[N^r \text{ red ball}]$
ii. nested: $[N^r \text{ ball}]$
iii. nested: $[N^r \text{ ball}]$
iv. flat: $[N^s \text{ ball}]$
The first hypothesis is the correct one for sentences such as (1). Critically, however, the hypothesis space also contains two hypotheses (nested: \( \text{N}^c \) ball] and flat: \([N^c] \) ball]) that Lidz et al. claim must be innately excluded from consideration for successful learning. On their view, it should be impossible to learn the correct hypothesis, given this hypothesis space and realistic input.

Let the evidence \( e^n \) consist of \( n \) observations of one referring to a red ball while (1) is uttered—this is the same as the scenario envisaged in (†). Assume that the world contains balls of \( c \) different colors, including red. Under these circumstances, the likelihood \( p(e^n|H) \) is shown below for each of the above four hypotheses.

\[
p(e^n|\text{nested : } [N^c \text{ ball}]) = \frac{1}{1^n} = 1 \\
p(e^n|\text{nested : } [N^c \text{ ball}]) = \frac{1}{c^n} \\
p(e^n|\text{nested : } [N^c \text{ ball}]) = \frac{1}{c^n} \\
p(e^n|\text{flat : } [N^c \text{ ball}]) = \frac{1}{c^n}
\]

These values are derived from (4), which is based on the size principle. For each hypothesis, \(|H|\) is the number of different colors that balls may appear in, when referred to by one. While the evidence is consistent with all four hypotheses, it supports the correct one more strongly than the others.

Fig. 2 shows the results of applying this model to the learning problem described in (†), under the assumption that \( c = 2 \) (e.g. there are only red balls and blue balls). Higher values of \( c \) yield qualitatively similar results, although with faster learning. A uniform prior was assumed, such that the four hypotheses were equally probable before observing evidence: \( p(H) = \frac{1}{4} \) for each. We combined the prior \( p(H) \) and likelihood \( p(e^n|H) \) to obtain the posterior probability \( p(H|e^n) \) for each hypothesis, for \( n = 0 \) through \( n = 5 \) exposures to the evidence. The \([N^c \text{ red ball}] \) hypothesis is found to be quite probable, and the three \([ball] \) hypotheses quite improbable, demonstrating that this allegedly unlearnable knowledge is in fact learnable, from evidence of the sort described in (†).

Does the child receive enough data to support such learning? Lidz et al. found 31 adult utterances of the form shown in (1) (i.e. anaphoric one with an antecedent noun phrase of the form determiner—adjective—noun) in the pooled Adam (Brown, 1973) and Nina (Suppes, 1974) corpora in the Childes database (J. Lidz, personal communication). The Adam corpus spans 116 h of caretaker—child interaction, while the Nina corpus spans 48 h, for a total of 164 h. If we take these numbers to be representative, we may assume that the child receives a sentence of the form shown in (1) every 164/31 = 5.3 h. The simulation shown here required only five exposures for nearly complete learning; these exposures could be supplied in 26\( \frac{1}{2} \) h of interaction—that is, a few days. Thus, even very young children may receive enough input to allow them to learn the knowledge in question.
3. Discussion

We have shown that a simple Bayesian learning model can learn syntactic knowledge that was claimed to be unlearnable. We anticipate two potential objections to such a demonstration.

The first potential objection is that the model is unrealistic. It assumes perfect memory, a capacity that both children and adults lack. The model also fails to capture the full range of uses of anaphoric *one*: it falsely assumes that *one* will always refer to the same node in the antecedent noun phrase. In actuality, *one* will refer to either the upper or the lower \( N \) in the nested structure, depending on the sentence. We followed Lidz et al. in focusing on the restricted case in which *one* refers to the upper \( N \)—and we suggest that the broader lessons of this demonstration are not undermined by that restriction, nor by the idealization of perfect memory. The broader point that the model illustrates, despite its shortcomings, is that evidence that is consistent with multiple learning hypotheses can sometimes discriminate among those hypotheses in a principled manner. Since some prominent versions of the argument from poverty of the stimulus assume that such learning is not possible, the model highlights a flaw in these arguments.

The second potential objection is that the model’s hypothesis space is very small. How, one may ask, can such a constrained model challenge the proposition that syntax learning must be constrained? We would respond that the model does not challenge that rather general proposition. Instead, it challenges the more specific proposal that particular sorts of syntactic knowledge must be innate (e.g. that *one* \( \neq [N^e \text{ ball}] \)). The model is also constrained in preferring narrow hypotheses over broad ones—but this is not a language-specific constraint, nor an arbitrary one. Rather, as we have seen, it emerges
from the general process of determining how likely a particular observation is, given a hypothesis.

More generally, learning of any sort is impossible without constraints of some kind—stemming from the structure of the hypothesis space, the nature of the learning process, or both. The central question then should not be whether syntax learning is constrained; it must be, like any other form of learning. Instead, in our view, the central question should be whether the constraints that govern syntax learning are themselves syntactic in nature—part of a specifically syntactic predisposition for language—or whether they emerge from more general aspects of cognition. A fuller answer to this question will require continued investigation, to determine how much of syntax learning can be explained in domain-general terms.

Acknowledgements

This work was supported by NIH grant DC03384. We thank Larry Hedges, Jeff Lidz, and Josh Tenenbaum for helpful discussions.

References

Discussion

Reaffirming the poverty of the stimulus argument: a reply to the replies

Jeffrey Lidz, Sandra Waxman

Abstract

Lidz, Waxman, and Freedman [Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn’t have learned: Evidence for syntactic structure at 18-months. Cognition, 89, B65–B73.] argue that acquisition of the syntactic and semantic properties of anaphoric one in English relies on innate knowledge within the learner. Several commentaries have now been published questioning this finding. We defend the original finding by identifying both empirical and logical flaws in the critiques.

Keywords: Universal grammar; Poverty of the stimulus; Language acquisition

1. Introduction

Fluent use of human language calls for a seamless integration of knowledge from a vast range of sources, both linguistic (e.g. phonology, morphology, syntax, semantics and pragmatics) and nonlinguistic (e.g. conceptual and perceptual representations of objects, events, and ideas; an appreciation of others’ minds and intentions). Surely, in acquiring language, learners must coordinate information from these diverse sources. But just as surely, different aspects of human language depend on these sources to different degrees. For example, sensitivity to the intentions of others likely plays a larger role in determining the meaning of a novel word than in determining the characteristic stress pattern of an exposure language. By the same token, sensitivity to amplitude in the speech signal likely

* Corresponding author.
E-mail address: jlidz@northwestern.edu (J. Lidz).
plays a larger role in determining stress patterns than in ascribing meaning to a novel word. The trick—for the learner and for the researcher—is to consider the most appropriate sources for each of the various aspects of language.

In our original article (Lidz, Waxman, & Freedman, 2003, henceforth LWF), we focused on the acquisition of the syntactic element known as anaphoric one. We argued that to interpret anaphoric one, the learner must depend on certain aspects of phrase structure. More specifically, we provided experimental evidence that by 18 months of age, infants have knowledge of the internal phrase structure of Noun Phrases, and we provided a corpus analysis of child-directed speech to show that this knowledge could not have been gleaned from the available input. These two pieces of evidence led us to conclude that the syntactic knowledge underlying the interpretation of anaphoric one depends more on the child’s representational presuppositions about syntax than on their experience with their language environment.

Clearly, this conclusion has struck a chord, sparking spirited responses from a strong set of commentators. These responses posed a number of interesting challenges, beckoning us to consider alternative interpretations. After a thorough consideration of each alternative, however, we find ourselves just where we began, asserting that infants at 18 months have an articulated, nested structure for the Noun Phrase and know that one is anaphoric to phrasal categories.

2. An overview. Poverty of the stimulus arguments

Although each commentary presents a unique perspective, they all touch upon the nature of poverty of the stimulus (henceforth POS) arguments. We therefore outline the logical form of POS arguments, place each commentary within this framework, and then respond to each in turn.

In general, the logic of the POS argument is to show that a piece of linguistic knowledge is not sufficiently triggered by the environment and hence involves some amount of innate structure in the learner. Any POS argument requires four parts. To illustrate these parts, we follow the terminology of Pullum and Scholz (2002). First, the acquirendum identifies a particular piece of syntactic knowledge. Second, the indispensability piece identifies what kind of input would be necessary for the learner to acquire the acquirendum. Third, the inaccessibility piece demonstrates that the indispensable evidence is unavailable to the learner. Fourth and finally, the acquisition piece of a POS argument shows that nonetheless the syntactic knowledge is present at the earliest possible age. Together, these pieces support the conclusion that learners have succeeded in acquiring a piece of syntactic knowledge that could not have been extracted from the environment without some inherent constraints on the hypothesis space.

Our original argument, following Baker, 1978, went like this. The acquirendum is the knowledge that one is anaphoric only to syntactic constituents larger than N° (i.e. the phrasal categories N’ or NP). The indispensable evidence was utterances compatible only with the \( one = N’ \) hypothesis; utterances that are consistent with both \( one = N° \) and \( one = N’ \) are treated as irrelevant to learning (since they do not decide the issue). We established the inaccessibility piece of the argument by documenting in a corpus analysis that the indispensable evidence was essentially unavailable. Finally, we showed that by 18-months,
infants have indeed acquired the acquirendum; infants’ behavior in our preferential looking experiment revealed that they know that one is anaphoric to phrasal categories.

The three responses to our original paper can be integrated within the logic of the classic POS argument. The arguments in these responses fall into two classes. One class focuses on acquisition. Engaging the work primarily on a descriptive level, Tomasello (2004) and Akhtar, Callanan, Pullum and Scholz (2004) (henceforth ACPS) ask “Do the results reveal that infants have the syntactic knowledge that LWF claim they have? Or could these results have been achieved through non-syntactic means?” A second class of challenge focuses primarily on the indispensability and inaccessibility pieces. Engaging the work on a deeper level, ACPS and Regier and Gahl, 2004 (henceforth R&G) pose questions that get at the issues of explanation and theory. Here they ask, “How did the babies develop the complex syntactic knowledge that LWF have demonstrated? Could this syntactic knowledge have been learned from the input?” We first address the descriptive challenge, defending our finding and clarifying its force. This clears the deck for a fuller treatment of the rich theoretical issues that lie at the heart of the POS argument.

3. Descriptive challenges. Have we really documented syntactic knowledge in infants?

Both Tomasello and ACPS consider whether infants’ performance in our experiments really depends upon syntactic knowledge. ACPS raise this concern within the context of a nuanced discussion that also involved issues of explanation and theory (see below). But Tomasello, who focuses exclusively on this issue, asserts that infants’ performance on our task reveals nothing in the way of syntactic knowledge on the part of our infant subjects.

Tomasello’s argument is based on the supposition that another one is an undecomposable expression (i.e. it is an unanalyzed whole) that is “conventionally associated” with “certain types of nonlinguistic situations” and that it therefore has a meaning that is not determined by its syntax. More precisely, he asserts that the expression another one simply refers to any object that is similar to an object already under consideration. To illustrate, Tomasello offers a hypothetical scenario in which a child is playing with a red block (which is not referred to linguistically) and then someone says, “Here’s another one”. In Tomasello’s view, the expression another one could only refer to another red block. He then asserts that scenarios like these provide a sufficient basis for acquiring the meaning of another one. Our objection takes three parts.

First, although the expression another one is not uncommon in the input to children (in the corpus used in LWF, it accounted for 33% (or roughly 4%) of the 792 anaphoric uses of one), only rarely is it used without a linguistic antecedent (only 3 of these 33 cases had no antecedent, or 0.37% of the anaphoric uses). This observation blunts Tomasello’s intuition that such contexts are sufficient to the acquisition of another one.

Second, our intuitions regarding such hypothetical scenarios does not match Tomasello’s. In our view, in such scenarios, the expression “Here’s another one”, could just as easily refer to a blue block, or to another toy altogether, depending upon the objects available and the goals at hand.

Third, and perhaps most telling, Tomasello’s intuition is contradicted directly by empirical evidence. Waxman and Markow (1998) presented this very scenario to
21-month-old infants in a control condition for a word-learning experiment. An experimenter introduced control infants to a target object (e.g. a yellow car) while saying, “See this. Look at this.” The experimenter then presented two test objects (e.g. a different yellow car and a green car) and asked, “Can you find another one?” As in Tomasello’s scenario, there was no linguistic antecedent for one. If Tomasello was correct, then infants in this condition should pick the test object that most closely resembled the target object; in this case, they should reveal a clear preference for the yellow car. But this was not the case. Infants had no such preference. In fact, they chose the yellow car at a rate that was indistinguishable from chance. This stands as empirical evidence that infants do not interpret another one as referring to the most similar object available in the context.

In contrast to these infants in Waxman and Markow’s control condition, infants in the LWF experiment did hear a linguistic antecedent during familiarization and did pick the most similar object at test. This indicates that LWF’s infants attended to the structure of the NP presented during familiarization, and that their appreciation of the internal syntactic structure of that NP led them (but not Waxman and Markow’s control infants) to pick the most similar object at test.

One further piece of evidence is relevant here. Although we did not include this evidence in the original report for lack of space, we did run a control condition in which we left out another in the test question. As in the original study, infants were familiarized to an object (e.g. a yellow bottle) which was labeled (here’s a yellow bottle…). However at test, these infants heard “do you see one now?” If Tomasello’s intuition is right, then infants in this condition should show no preference, since on this view the preference in our original experiment was due to the noncompositional meaning of the expression another one. Following Tomasello’s logic, one by itself is not conventionally associated with the kinds of nonlinguistic experiences that another one is and so does not uniquely identify either of the bottles. But, as in the original experiment, infants showed a reliable preference for the yellow bottle. This suggests that performance in our original experiment cannot be explained by appealing to the “kinds of nonlinguistic experiences conventionally associated with the expression another one.” Rather, infants’ interpretation of one reflects their knowledge that this expression is anaphoric to a phrasal category.

4. Theoretic and explanatory challenges. How do infants acquire the syntax of anaphoric one?

The other two commentaries were concerned more with the question of how children manage to acquire the syntactic knowledge that supports the appropriate interpretation of anaphoric one. At issue is in these commentaries is whether we successfully established the indispensability and inaccessibility portions of our argument.

4.1. Akhtar, Callanan, Pullum and Scholz (ACPS)

ACPS question our conclusions from multiple perspectives, challenging both the indispensability and inaccessibility portions of our POS argument. And within the context of these questions, they challenge the acquisition piece of the argument, asking whether
we have in fact demonstrated that infants have syntactic knowledge of the structure of NP or the properties of anaphoric *one*.

4.1.1. The indispensability argument

We argued that the only evidence that would lead the learner to the correct hypothesis regarding anaphoric *one* (nested structure in NP, *one* = N₁) would be events in which *one* is compatible with this hypothesis alone. ACPS take a broader view of the input, arguing that there are two additional kinds of indispensable evidence (namely, syntactically ambiguous utterances and exophoric reference) that could support learning of anaphoric *one*. Their argument is predicated on the view that comprehension is driven by the relevance principle (Sperber & Wilson, 1995).

ACPS argue that “if learners do sometimes have to figure out from primary data that *one* has a multi-word antecedent, it does not follow that the context must *entail* that the antecedent is multi-word. Mere pragmatic implication or conversational implicature will do just as well.” In other words, by virtue of the principle of relevance, the use of *yellow* in an antecedent noun phrase makes the property *yellow* relevant to all subsequent utterances. It therefore follows that learners assume this property (even in situations which do not entail that it is included in the antecedent) and interpret *one* accordingly. In essence, then, the argument is that a pragmatically driven assumption can then support the nested structure grammar in which *one* is anaphoric to N₁.

But this argument runs aground because the assumption in this relevance-oriented view is too broad. The assumption is that the adjective in the antecedent should make the property relevant not only in cases of anaphoric *one* (as in our experimental condition: *do you see another one*?), but also in a broad range of other subsequent utterances¹ (as in our control condition: *what do you see now*?). For both types of utterances, a relevance-based approach predicts (a) that the color of the object presented during familiarization was relevant (otherwise, why would we have mentioned it) and (b) that participants would look to the yellow bottle in both conditions. But the data from LWl do not support this prediction. Infants in the experimental condition preferred the yellow bottle at test, but those in the control condition performed at chance. This difference between the experimental and the control conditions demonstrates that infants’ interpretation of *one* cannot be reduced to the principle of relevance.

ACPS also argue that exophoric uses of *one* (i.e. uses in which the referent of *one* is inferred from context alone) could serve as indispensable evidence to the acquisition of anaphoric *one*. They state, “if *Hand me that* is correctly understood (to mean the yellow bottle) and the next utterance is *give me another one*, no syntactic constituent serves as antecedent at all. But the relevance principle still suggests that *one* has, say, *yellow bottle* as antecedent rather than *just bottle*...Such uses may be highly instructive to the learner.”

This challenge parallels the concern raised by Tomasello. To revisit that issue briefly, recall that the evidence from Waxman and Markow (1998) showed that exophoric uses are

¹ Unless, of course, the subsequent utterance contradicts the first utterance, as in “You have a yellow bottle and I have a blue one.”
not informative in this situation. This renders implausible the assertion that exophoric uses of *one* form the basis for learning the anaphoric uses.

4.1.2. *The inaccessibility argument*

In LWF, we used an analysis of a CHILDES corpus to address this issue of inaccessibility. Our analysis revealed that technically speaking, the input does contain some situations which could, in principle, support the acquisition of anaphoric *one*. The trouble is that such utterances occurred at rate of 0.2%, which is the very same rate as ungrammatical uses of *one*. The crux of our argument was this: Certainly learners hear examples that they could learn from, but they hear these at the same rate as ungrammatical examples that they shouldn’t learn from.

ACPS take a different direction, arguing that the 0.2% of situations that would allow learning would amount to dozens of examples accumulated over the first 2 years of life. We concur, but our claim was not about raw percentages or even projections from the corpus to the actual input a child might receive over a more extended period of time. For if the number of good examples increases *n*-fold over a period of time, so does the number of ungrammatical uses. Unfortunately, the input to the learner does not come marked as to whether and how it should be used by the learning algorithm. The raw input is not marked with regard to grammaticality, felicity, truth value, etc. By logic, then, we can assume that if learners do not acquire the syntactic properties of anaphoric *one* from the 0.2% of the data that is ungrammatical (of which we are certain, since mature speakers of English clearly have command of the syntax), then neither could they acquire the syntactic properties from the 0.2% of the data that is crucially informative.

4.2. *Regier and Gahl (R&G)*

R&G’s criticism takes yet a different tack, focusing on the indispensable (and consequently, inaccessibility) parts of our argument. They correctly point out an important component of any learnability problem, namely, that indirect negative evidence can have a powerful influence on acquisition (Chomsky, 1981). R&G say,

> Given evidence that is consistent with several hypotheses, a learner can come to discriminate among them, for principled, domain-general reasons. In particular, if one of the hypotheses predicts not only the input that is seen, but also input of another sort that is never seen, that *absence* can serve as evidence against the hypothesis.

R&G’s point is that learners may be able to use the absence of certain kinds of data as evidence against hypotheses that would have predicted such data to occur. We agree, in principle. The trouble is that the solution proposed by R&G fails. To show why this is so, we outline three problems inherent in their proposal which, singly and in combination, lead us back to our original point—that the acquisition of anaphoric *one* cannot be learned from the available input (even if the learner considers the full range of negative and positive evidence).

First, the data presented to the R&G model bears little relation to the input seen by actual children. Their model saw only examples of *one* being used with an antecedent containing
an adjective (e.g. here’s a red ball. do you see another one?). Their model learned quickly that one = [N red ball]. But this not altogether surprising since there were no exposures in which the actual referent was not red. Child learners are not so fortunate. In the corpus we examined, examples like those presented to the R&G simulation (with one being used with an antecedent adjective) were rare. Less than 5% of the utterances containing one were preceded by antecedents containing adjectives; 95% of the data they were exposed to had an antecedent that did not contain an adjective. That is, the vast majority of cases were more like “here’s a bottle. do you see another one?” This raises questions concerning the degree to which the R&G model fits the child’s case in acquiring anaphoric one.

Second, even if we accept R&G’s model and apply it to the data that children actually do receive, we suspect that the simulation will run aground. This is because in cases with no adjective in the antecedent, there is no a priori reason for the learner to privilege any one particular property of the referent (e.g. color) over another. That is, in cases like “here’s a bottle” (where the bottle happens to be yellow), there is no reason to interpret “get another one” (anaphoric one) as referring to a yellow bottle, because that referent (the yellow bottle) is also a full bottle, a warm bottle, your sister’s bottle, etc. In fact, learners should conclude that one in such cases makes no commitment to any of a host of potential properties of the referent. Put simply, in the face of input that more closely matches that provided to actual children, one would most plausibly be interpreted as referring to the object category (bottle), making no commitment to its various properties. And because utterances of this type are so frequent in the actual input to infants, this will have a consequence for learner’s interpretation when the antecedent NP does mention color (e.g. here’s a yellow bottle), size, shape, texture, attractiveness, temperature, or any other of its properties. Now, if for some reason, the learner were paying attention to the various properties of the bottle in the cases where it was not mentioned, she might decide that properties of the antecedent (like color) are altogether irrelevant for assigning an antecedent to one, thereby swamping the potential effect of indirect negative evidence in the cases when the adjective is present and therefore potentially relevant. Our point here is not that indirect negative evidence is unavailable to a learner. We believe that it is. Rather, we claim that the indirect negative evidence that kids really do get could not solve the learning problem in this instance, given the overwhelming proportion of data which are consistent with the one = N hypothesis and the vast range of properties that should correctly be ignored almost always when assigning an antecedent to anaphoric one.

A third issue brought up by R&G’s model concerns the details of the grammar of anaphoric one. Our original discussion was primarily concerned with the possibility of one taking a multi-word antecedent, e.g. yellow bottle in our example. It is this possibility that provides the evidence against a flat structure grammar. R&G’s discussion, however, assumes that it is necessary for one to have a multi-word antecedent if one is available. This is transparently not the case, as shown by examples like (1a) and (1b).

(1) a. I have a yellow bottle and you have a blue one.
   b. I like the yellow bottle but you like that one.

Now, these cases do involve anaphora to N, but they differ from our original example in that here one is anaphoric to the lower of the two N’s, i.e. the one containing only
the string bottle. Consider the structure in (2).

```
(2)                NP
     Det N'          
                    |     N'
                    |     N
                 AdjP
                    the
               yellow
              bottle
```

Here, there are two nodes labeled N' and either one of these is a potential antecedent for anaphoric one.\(^2\) The fact that one can refer to either of two N' nodes is important because it shows that R&G's learning model is too powerful. The conclusion of their model is that one must refer to the largest available N', namely the one containing the string yellow bottle. But this conclusion is too strong, because one can refer to lower instances of N', as we have seen. It is simply not an entailment of sentences like “I like this red ball and you like that one” that one be anaphoric to red ball. Although this reading may be the most felicitous one in some contexts, in other contexts and utterances, one can indeed refer to the lower N'. If we accept R&G's indirect evidence model, then such contexts either would be ruled impossible (which would be an error in learning) or would constitute evidence that one can refer to balls of any color, even when the antecedent mentions color, and would thus blunt the effectiveness of any indirect negative evidence of the sort described by R&G. In short, the use of indirect negative evidence in the fashion proposed by R&G

\(^2\) We know that the nonbranching N' dominating bottle is present because of the impossibility of one being anaphoric to complement taking nouns:

(i)    a. I met the member of congress...
       b. * ...and you met the one of the rotary club
       c. [NP the [N [S member] [S of congress]]]
(ii)   a. I reached the conclusion that syntax is innate...
       b. * ...and you reached the one that learning is powerful
       c. [NP the [N [S conclusion] [CP that syntax is innate]]]

These contrast with cases in which what follows the head noun is an adjunct/modifier. Here, one can substitute for what appears to be only the head noun.

(iii)  a. I met the student from Peoria...
       b. ...and you met the one from Podunk.
       c. [NP the [N [S student] [S from Peoria]]]
(iv)   a. I met the student that you invited to the party
       b. ...and you met the one that Bill invited
       c. [NP the [N [S student] [S that you invited to the party]]]

What these cases tell us is that one can take a single word as its antecedent only when that single word is dominated by a nonbranching N'. In other words, in cases like (1a), it must be the case that ball = N', as in the structure in (2). If it weren’t, we would have no way to distinguish this case from one in which one cannot substitute for a single word, as in (i) and (ii).
leads to a grammar in which it is never possible for one to refer to the lower of two N’s, contrary to fact. Their model is simply too restrictive.

A fourth, more general, issue brought up by R&G’s reply concerns the constraints on the model. We argued that two aspects of syntactic structure guide the acquisition process: (a) that NPs have a nested (X-bar theoretic) structure, and (b) that anaphora to X° be excluded from the hypothesis space. R&G argue that these particular pieces of linguistic knowledge need not be innate. However, R&G do have a massively restricted hypothesis space (one containing only four hypotheses). So, we are in agreement that the hypothesis space for syntactic acquisition must be severely restricted. We differ only in which hypotheses are represented in that space. If it turns out, as R&G suggest, that the learner can work through that hypothesis space using domain general learning mechanisms, that would be a very interesting finding, but it is important to emphasize, as R&G do themselves, that all learning must be constrained. Without a severely restricted hypothesis space, the learner cannot even get started.

5. Conclusion

We are pleased that our research has sparked such enthusiastic discussion from such a diverse source of commentators. The commentaries have forced us delve deeper into the nature of the POS and to consider alternatives to our particular instantiation of the POS argument. But, having visited these alternatives, we end right where we began. By the time they are 18-months old, infants have an articulated, nested, structure for the Noun Phrase and they know that one can be anaphoric only to phrasal categories. As far as we can tell, this grammatical knowledge could not have come from the input that they were exposed to (even if they use pragmatic information and/or negative evidence to evaluate that input). The alternative, then, is that this knowledge must have come from within the infants themselves.

References