Indirect Evidence and the Poverty of the Stimulus: The Case of Anaphoric *One*

Stephani Foraker, Terry Regier, Naveen Khetarpal ({sforaker, regier, khetarpal}@uchicago.edu)
Department of Psychology, The University of Chicago, 5848 S. University Ave., Chicago, IL 60637 USA

Amy Perfors, Joshua B. Tenenbaum ({perfors, jbt}@mit.edu)
Department of Brain and Cognitive Sciences, MIT, Building 46-4015, 77 Massachusetts Ave., Cambridge, MA 02139 USA

Abstract

It is widely held that children's linguistic input underdetermines the correct grammar, and that language learning must therefore be guided by innate linguistic constraints. In contrast, a recent counterproposal holds that apparently impoverished input may contain indirect sources of evidence that allow the child to learn without such constraints. Here, we support this latter view by showing that a Bayesian model can learn a standard "poverty-of-stimulus" example, anaphoric *one*, from realistic input without a constraint traditionally assumed to be necessary, by relying on indirect evidence. Our demonstration does however assume other linguistic knowledge; thus we reduce the problem of learning anaphoric *one* to that of learning this other knowledge. We discuss whether this other knowledge may itself be acquired without linguistic constraints.

Keywords: language acquisition; poverty of the stimulus; indirect evidence; Bayesian learning; syntax; anaphora.

Introduction

Language-learning children are somehow able to make grammatical generalizations that are apparently unsupported by the overt evidence in their input. Just how they do this remains an open question. One influential proposal is that children succeed in the face of impoverished input because they bring innate linguistic constraints to the task. This argument from *poverty of the stimulus* is a long-standing basis for claims of innate linguistic knowledge (Chomsky, 1965). An alternative solution is that the learner instead relies on *indirect evidence* (e.g. Landauer & Dumais, 1997; Reali & Christiansen, 2005), rather than requiring innate linguistic constraints. We pursue this idea here, and focus in particular on what can be learned by noting that certain forms systematically fail to appear in the input. In exploring this idea, we assume a rational learner that is sensitive to the statistical distribution of linguistic forms in the input.

Noun phrase structure and anaphoric *one*

An established example of the argument from poverty of the stimulus concerns the anaphoric use of the word *one* (Baker, 1978; Hornstein & Lightfoot, 1981). The learning challenge is to determine the antecedent of *one* within a hierarchically structured noun phrase.[1] Concretely, the problem facing the learner is illustrated in (1), below. The antecedent of the anaphor *one* is ambiguous in that *one* could refer to 3 different levels of noun phrase structure, shown in Figure 1: (a) the upper N', referring to a yellow bottle, (b) the lower N', referring to a bottle of some unspecified color, or (c) N^0, also referring to a bottle of some unspecified color.

1. Here's a yellow bottle. Do you see another one?
   $[_{NP}$ a $[_{N'}$ yellow $[_{N'}$ $[_{N^0}$ bottle]]]]

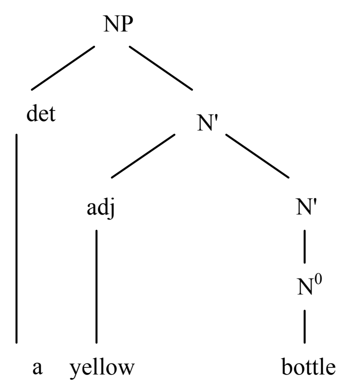

Figure 1: Structure of the noun phrase *a yellow bottle*.

What is the right answer, and how can a child acquire that knowledge? It is generally accepted that *one* can take any N' constituent as its antecedent (Radford, 1988). For instance, in example (1), *one* is usually taken to refer to a yellow bottle; thus the preferred antecedent here is the upper N' (but see below for more general treatment). Lidz, Waxman, and Freedman (2003) showed that 18-month-old infants, given sentences like (1), looked longer at a yellow bottle than at a bottle of a different color, which they interpreted to mean that the infants knew that *one* here took [upper N'] as its antecedent. Lidz et al. (2003) also searched through a child language corpus for input that, in a single occurrence, could unambiguously rule out incorrect hypotheses. They found that such input was effectively absent from the corpus. They argued that this poverty of the

[1] A separate interesting question is how the child comes to know in the first place that language is hierarchically, not just sequentially, structured. Perfors, Tenenbaum, and Regier (2006) presented a Bayesian model that acquired this general knowledge from child-directed speech, without any prior bias favoring hierarchical structure. We pursue the same general principles here in learning *which part* of this hierarchical structure is the antecedent of anaphoric *one*.

stimulus implicated innate syntactic knowledge: children know something about language that they couldn't have learned from the input, so at least part of the knowledge must be innate. In particular, they argued that learning is innately constrained to exclude the [N^0] hypothesis from consideration.

However, Regier and Gahl (2004) showed that a simple Bayesian model could learn the [upper N'] solution for such sentences, given only input of the form shown in (1), without this constraint. Thus, after learning, their model qualitatively matched the behavior of the children Lidz et al. had tested. Their demonstration relied on a simple domain-general principle: hypotheses gradually lose support if the evidence they predict consistently *fails to appear*. Here, if either [N^0] or [lower N'] were the correct choice, we would expect to see utterances like (1) sometimes spoken in contexts in which *one* referred to a non-yellow bottle. But since the correct antecedent for this utterance is the upper N', such evidence will not appear, and that *absence* of evidence can drive learning, in a gradual rather than one-shot fashion. Regier and Gahl (2004) argued on this basis that it is not necessary to posit an innate exclusion of the [N^0] hypothesis, contra Lidz et al. More broadly, they argued that investigations of the poverty of stimulus should attend closely to what is absent from the input, as well as what is present (Chomsky, 1981:9).

In response, Lidz and Waxman (2004) argued that Regier and Gahl's model is inadequate in two important ways. First, the input it received was not realistic: it was given only determiner-adjective-noun NPs as input, whereas the vast majority of uses of anaphoric *one* have an antecedent NP that does not contain an adjective. Second, and more fundamentally, they argue that the model learned the wrong thing. The model learned to support the [upper N'] hypothesis only, whereas as stated above, more generally anaphoric *one* can substitute for *any* N' constituent. For instance, as they noted, and as shown here in (2), *one* can also substitute for the lower N':

2. Here's a yellow bottle. Do you see a blue one?
   $[_{NP}$ a $[_{N'}$ yellow $[_{N'}$ $[_{N^0}$ bottle]]]]

This critique represents a slight shifting of the goalposts, since the Lidz et al. (2003) experiment itself, to which Regier and Gahl responded, suggested that children interpret *one* as anaphoric to the upper N' in the context of (1) – and the experiment did not speak to the more general [any N'] hypothesis. Still, since adults do know that the correct answer is [any N'], it is reasonable to require that an adequate model explain how that knowledge is learned.

This requirement is especially troublesome for Regier and Gahl's (2004) model, since that model based its discrimination among hypotheses on *referential* grounds, by noting the color of the real-world bottle when sentences like (1) were uttered. And referentially, nothing distinguishes a situation in which the antecedent of *one* is the lower N' from a situation in which it is N^0, since the referenced object can be a bottle of any color. This is a problem because the lower N' situation is consistent with the correct hypothesis [any N'] while the N^0 situation is not.

Since referential evidence will not suffice to learn the correct [any N'] hypothesis, what sort of evidence might? We know that *one* cannot be anaphoric to N^0 because, as shown in (3), it is ungrammatical for *one* to be anaphoric with a complement-taking noun (*side*) without its complement (*of the road*, a prepositional phrase in argument slot, Radford, 1998; Lidz & Waxman, 2004).

3. *I'll walk by the side of the road and you can walk by the one of the river.
   $[_{NP}$ the $[_{N'}$ **[**$_{N^0}$ side**]** $[_{PP}$ of the road]]]
   $[_{NP}$ the $[_{N'}$ **[**$_{N^0}$ one**]** $[_{PP}$ of the river]]]

Such unacceptable complement structures contrast with modifiers. In (4), which has a noun with a post-nominal modifier, it is grammatical for *one* to be anaphoric with the noun (*ball*) without its modifier (*with stripes*). In syntactic structure this is reflected by the modifier attaching to the lowest N' rather than N^0.

4. I want the ball with stripes and you can have the one with dots.
   $[_{NP}$ the **[**$_{N'}$ $[_{N^0}$ ball]**]** $[_{PP}$ with stripes]]
   $[_{NP}$ the **[**$_{N'}$ $[_{N^0}$ one]**]** $[_{PP}$ with dots]]

Thus, if the language-learning child had grasped the distinction between complements and modifiers, that distinction could serve as a basis for learning about anaphoric *one*.

This idea inverts a standard linguistic test (e.g. Radford, 1988: 175), in which the acceptability or unacceptability of substituting *one* in an NP is used to determine whether a given post-nominal phrase within the NP is a complement or a modifier, as in examples (3) and (4). There is an apparent circularity in this: we can use the acceptability of substituting anaphoric *one* to determine whether a phrase is a complement or modifier – but we need the distinction between complements and modifiers to learn the correct use of anaphoric *one* in the first place. This circularity is only apparent, however, since complements may be distinguished from modifiers on semantic and conceptual grounds, as we discuss below. We assume that the child is able to use semantic/conceptual information to begin distinguishing between complements and modifiers.

In this paper, we address both of the criticisms that Lidz and Waxman (2004) directed at the Regier and Gahl (2004) model. We show that a different Bayesian model can learn the correct [any N'] hypothesis given realistic input, without innately excluding the [N^0] hypothesis. In doing so, we provide further support for the central claim of Regier and Gahl (2004): that by relying on indirect negative evidence, a child can learn the knowledge governing anaphoric *one*, without the allegedly necessary innate linguistic constraints.

Model

We assume a rational learner that assesses support for hypotheses on the basis of evidence using Bayes' rule:

$$p(H \mid e) \propto p(e \mid H) p(H)$$

Here H is a hypothesis in a hypothesis space, and e is the observed evidence. The likelihood $p(e|H)$ is the probability of observing evidence e given that hypothesis H is true, and the prior probability $p(H)$ is the a priori probability of that hypothesis being true. To flesh this general framework out into a model, we need to specify the sort of evidence that will be encountered, any general assumptions, the hypothesis space, the prior, and the likelihood.

Evidence

The model observes a series of noun phrases, drawn from child-directed speech. Each noun phrase is represented without hierarchical structure, as a sequence of part-of-speech tags (e.g. *the big ball* would be coded as "determiner adjective noun"), supplemented with a code for a modifier or complement, if any (e.g. *side of the road* would be coded as "noun complement"). This source of evidence was chosen because (1) children receive a steady supply of such input, and (2) following our discussion above, it is linguistic data of this sort, rather than the real-world objects to which anaphoric *one* may refer, that can discriminate among the relevant hypotheses.

Assumptions

We made three assumptions about the knowledge available to the language-learning child. First, following Lidz et al. (2003), we assumed that the child is able to recognize anaphoric uses of *one*. Second, we assumed that the child knows a fundamental fact about pronouns generally, including *one*: that a pronoun effectively *substitutes* for its antecedent, and must therefore be of the same syntactic type as the antecedent. Thus, if the pronoun occupies an N' position within its noun phrase, the antecedent must similarly occupy an N' position in its noun phrase, for otherwise the pronoun would not be able to substitute for the antecedent. By the same token, if the pronoun occupies an N^0 position, the antecedent should, too. Critically, given this assumption, the problem of determining whether the antecedent of *one* is N' or N^0 reduces to the problem of determining whether *one* itself, within its own NP, takes the role of N' or N^0. We felt justified in making this assumption since the knowledge we assumed concerns pronouns generally, and could be learned by observing the behavior of pronouns other than *one*.

Finally, we assumed that the child is able to recognize and distinguish between complements and modifiers when they appear in the child's linguistic input. To our knowledge, there are no studies that have tested whether young children are indeed sensitive to this distinction, but we felt justified in making this assumption since the core distinction between complements and modifiers can be captured in semantic or conceptual terms, and thus could in principle be learned without innate specifically syntactic constraints. A complement is necessarily conceptually evoked by its head. For instance, *member* necessarily evokes the organization of which one is a member; so in *member of congress*, the phrase *of congress* is a complement. In contrast, a modifier is not necessarily evoked by its head. The word *man* does not necessarily evoke conceptually where the man is from, whether he has long hair, etc.; so in *man from Rio*, the phrase *from Rio* is a modifier, not a complement (Baker, 1989; Bowen, 2005; Keizer, 2004; Taylor, 1996). While there are more subtle intermediate cases, this is the conceptual core of the distinction.

We return in the discussion to the question of just how much of our argument hangs on these assumptions.

Hypothesis space

We assumed a hypothesis space containing two hypotheses which addressed the question "Which of the constituents of the NP does anaphoric *one* take as its antecedent?" The two hypotheses are [any N'], and [N^0]. Thus, we chose the simplest possible hypothesis space that includes both the correct answer [any N'] and the hypothesis that Lidz et al. argue must be innately *excluded* if learning is to succeed [N^0]. If a rational learner can learn the correct answer given this hypothesis space and realistic input, that outcome will indicate that the posited innate exclusion of [N^0] is unnecessary.

Each hypothesis takes the form of a grammar that generates a string of part-of-speech tags corresponding to a noun phrase. The two grammars are identical except for one rule. Each grammar contains the following productions, with options separated by "|":

NP → Pro | Nbar | Det Nbar | Poss Nbar
Poss → NP ApostropheS | PossPronoun
Nbar → Poss Nbar | Adj Nbar | Nbar Mod
Nbar → Nzero | Nzero Comp
Det → determiner
Adj → adjective
PossPronoun → possessive-pronoun
ApostropheS → apostrophe-s
Mod → modifier
Comp → complement
Nzero → noun
Pro → pronoun

In addition to these productions, the [any N'] hypothesis contains the production:

Nbar → anaphoric-*one*,

while the N^0 hypothesis instead contains the production

Nzero → anaphoric-*one*.

Thus, the two grammars embody, in their last production, the link between *one* and either N' or N^0. The grammars were designed to be able to parse noun phrases in a child-language corpus. Each production in each grammar has a production probability associated with it, and these probabilities may be adjusted to fit the observed corpus.

Prior

The two grammars are equally complex: they differ only in one production, which is itself equally complex in the two cases. Since grammar complexity gave no grounds for assigning either hypothesis greater prior probability, we assigned the two hypotheses equal prior probability $p(H)$: $p(N^0) = p(\text{any } N') = 0.5$. Thus, all discrimination between hypotheses was done by the likelihood.

Likelihood

Given a hypothesis *H* in the form of a grammar, and evidence *e* in the form of a corpus of noun phrases, we used the inside-outside algorithm[2] to obtain a maximum likelihood fit of the grammar to the corpus. This algorithm iteratively reestimates production probabilities in the grammar so as to maximize the probability $p(e|H)$ that the corpus *e* would be generated by the grammar *H*. Given the prior and likelihood, we then obtained the probability of each grammar given the corpus, $p(H|e)$, using Bayes' rule.

Both grammars were designed to be consistent with all noun phrases in our corpus. What differs between the grammars is the *expected* observations given that a hypothesis is true. To see why this is the case, consider the interaction of two rules from the N^0 grammar: [Nbar $\rightarrow$ Nzero Comp], and [Nzero $\rightarrow$ anaphoric-*one*]. Together, these two rules produce strings of the form "*one* + complement", as in (3) above. Thus the N^0 hypothesis predicts that such strings will be encountered in the input. But since such strings are ungrammatical, that expectation will not be fulfilled. In contrast, the N' hypothesis does not give rise to this false expectation, since it lacks the second rule. This difference between the two grammars is captured in their likelihoods. If no instances of "*one* + complement" appear in the input, the N^0 grammar will progressively lose support, and the learner will select the N' grammar as the correct hypothesis.

Data

We selected as our input data source the Nina corpus (Suppes, 1974) in the CHILDES database (MacWhinney, 2000), which Lidz et al. (2003) had consulted. The corpus was collected while the child was 23-39 months old, and consists of just over 34,300 child-directed mother utterances, containing approximately 60,000 noun phrases, which we identified preliminarily using a parser.[3] From those noun phrases we selected a random 5%, 10%, and 15% cumulative sample for further coding by one of the authors (SF). These percentages yielded samples large enough for our purposes, yet small enough to code by hand.

We coded each noun phrase as a sequence of part-of-speech tags, without hierarchical structure, e.g. "determiner adjective noun", "determiner noun", "pronoun", etc., of the sort generated by the above grammars. We used these corpus data in two ways. First, we designed both grammars to accommodate these data. Second, we provided the data as input to the model. While both grammars were designed to be *consistent* with the data, we were interested in finding which grammar fit the data most closely.

We also coded whether a complement or modifier (or neither) was present. Thus, for example, the noun phrase *a piece of cheese* would be coded "determiner noun complement", while *crackers with cheese* would be coded "noun modifier". We limited ourselves to post-head complements and modifiers in the form of prepositional phrases or clauses (Bowen, 2005; Keizer, 2004; Radford, 1988). To identify a complement or a modifier we used the conceptual intuition described earlier, identified by several sources. The head noun that takes a complement presupposes some other entity which must be expressed (Huddleston & Pullum, 2002:221; Taylor, 1996:39, see also Fillmore's "inalienably possessed nouns", 1965) or inferable from context (Bowen, 2005:18, Keizer, 2004). To classify post-head strings that followed anaphoric *one*, such as *the one in the picture*, we identified the head noun of the antecedent NP from the transcript, and applied the same test as for nouns. Note that the head noun is the same regardless of whether N' or N^0 is the correct hypothesis, so this coding does not depend on knowing the correct hypothesis.

We did not use substitution of anaphoric *one* as a test for classifying the post-head forms, to avoid the circularity alluded to above. We restricted ourselves to the conceptual distinction that could lead a child to the complement-modifier distinction without requiring prior syntactic knowledge of anaphoric *one*. However, we did find post hoc that the anaphoric *one* test for count nouns[4] yielded results consistent with the criteria we adopted.

Table 1: Frequency counts of post-head structures in the input for 5%, 10%, and 15% cumulative samples

Noun phrase forms	5%	10%	15%
noun	1253	2478	3752
pronoun	1605	3213	4784
anaphoric *one*	13	32	49
noun + complement	29	47	65
noun + modifier	66	113	161
noun + complement + modifier	0	1	1
anaphoric *one* + modifier	3	3	4

[2] We used the code made publicly available by Mark Johnson at http://www.cog.brown.edu/~mj/Software.htm

[3] We used the Stanford Parser, version 1.5.1, available at http://nlp.stanford.edu/downloads/lex-parser.shtml.

[4] Anaphoric *one* does not substitute for mass nouns.

Table 1 shows a summary of the types of noun phrase forms available in the input, focusing on the post-head structure (none, complement, modifier). Following a head noun we found complements (piece *of a puzzle,* your side *of the road*) and modifiers (food *for the llamas,* the ocean beach *with big waves*), while following anaphoric *one* we found only modifiers (the one *with the girl,* the other ones *you like*), consistent with an adult-state grammar. The complements were all prepositional phrases, while the modifiers consisted of prepositional phrases or subordinate clauses.

To explore what data are critical to learning anaphoric *one*, we also created two variants of the 15% sample: the first variant (the "no-ones" corpus) was stripped of all NPs containing anaphoric *one*, while the other variant (the "no-complements" corpus) was stripped of all NPs containing a complement. We refer to the 15% sample as "the full corpus". The no-ones, no-complements, and full corpora contained 8763, 8750, and 8816 NPs respectively. Thus, the manipulations removing instances of *one* or complements each eliminated only a very small proportion (0.6% and 0.8%, respectively) of the "full corpus".

Methods

We first calculated the probability of each hypothesis ([N^0], [any N']), given the 5, 10, or 15% samples, corresponding to 4.2, 8.4, and 12.6 hours of mother input, respectively. We then re-calculated the same probabilities on the no-ones and no-complements corpora.

Results

Figure 1 shows that the probability of the correct [any N'] hypothesis is equal to that of the incorrect [N^0] hypothesis prior to observing any data (0 hours of input). However, as more data are seen, the probability of the correct hypothesis given those data grows steadily higher, at the expense of the incorrect hypothesis. This indicates, contra the poverty-of-stimulus argument, that a rational learner can discover that the antecedent of *one* is N', even though N^0 is not innately excluded from consideration during learning.

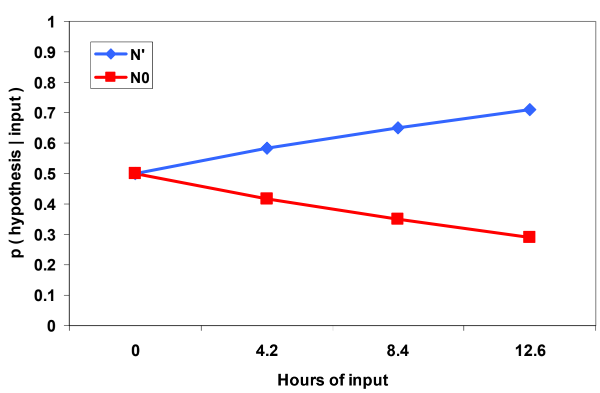

Figure 1: Probability of hypotheses given varying amounts of mother input.

Why does this happen? The N^0 hypothesis falsely predicts that the input will include strings containing "*one* + complement", while the N' hypothesis does not. Thus, the likelihood of the actual observed data is higher for the N' hypothesis. As we have seen, the false N^0 prediction arises from the interaction of two rules in the N^0 grammar: [Nbar → Nzero Comp], and [Nzero → anaphoric-*one*]. The first production is shared with N', while the second is not. The probabilities of both productions must be substantial, since the corpus contains complements, which require the first rule, and instances of *one*, which require the second. However, if the corpus lacked either *one* (e.g. the no-ones corpus), or complements (e.g. the no-complements corpus), the corresponding rule would receive 0 probability in a maximum likelihood fit to the data. In such cases the N^0 hypothesis would *not* make the false "*one* + complement" prediction, and there would be nothing to distinguish N^0 from N'. These expectations were confirmed, as shown in Figure 2.

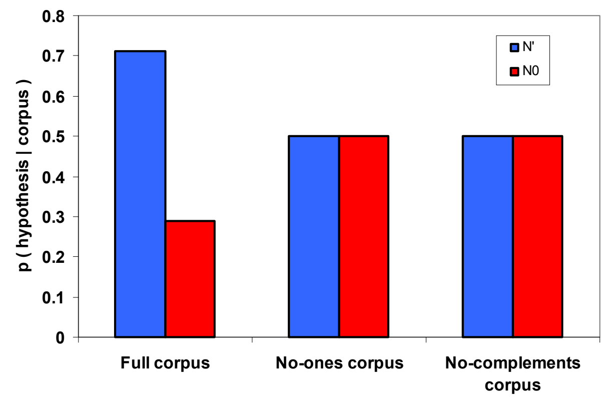

Figure 2: Probability of hypotheses given the full corpus, and same corpus stripped of *one*, or of complements.

Thus, the successful learning we see on the full corpus is dependent on the interaction of anaphoric *one* and complements. When both appear in the corpus, even in very modest quantities as was true here, the N^0 hypothesis falsely predicts the unattested "*one* + complement" pattern, and is penalized for its absence. This interaction supports learning without the innate exclusion of N^0.

Discussion

We have shown that a rational learner can learn the behavior of anaphoric *one* without a linguistic constraint that has been held to be necessary, and necessarily innate. Our demonstration relies on learning from the absence of predicted input patterns, a form of indirect evidence that is broadly consistent with other recent work emphasizing the power of indirect evidence in countering standard poverty-of-stimulus arguments (Landauer & Dumais, 1997; Reali & Christiansen, 2005).

We anticipate a number of objections to our demonstration. First, our hypothesis space is very restricted, containing just two hypotheses. One might concede our point that N^0 need not be excluded from consideration, contra the standard argument, but then counter that we ourselves use a very constrained space. Thus, perhaps the fundamental "constrained space" idea is correct, even if the excluded-N^0 proposal was not. We consider it self-evident that the space must be constrained; the critical question is whether the constraints are specifically linguistic. Will our demonstration scale up in a hypothesis space that is constrained only by non-linguistic general cognitive considerations? We consider that an open and interesting question.

A related possible objection is that we have assumed a good deal of linguistic knowledge: for instance, knowledge that pronouns substitute for their antecedents, that language is hierarchically structured, and knowledge of the complement-modifier distinction. This is true. Our primary contribution has been to *reduce* the problem of learning anaphoric *one* to the problem of learning this other knowledge. The critical subsequent question is whether this knowledge that we have assumed can itself be learned without innate linguistic constraints (e.g. Perfors et al., 2006 show that the hierarchical structure of language may be learned without prior bias). If this knowledge we have assumed can be so learned, a standard poverty-of-stimulus example will have been shown to be learnable without specifically linguistic constraints. If not, the example of anaphoric *one* will retain its status as an argument for innate linguistic knowledge – but we will have shown that the critical linguistic constraints lie elsewhere than traditionally imagined.

Perhaps the broadest potential objection is that it may seem wrong-headed – or paradoxical – to argue against the nativist poverty-of-stimulus claim while using structured linguistic representations of exactly the sort commonly proposed by nativists. We see no problem here. We consider ourselves to be working "from the inside out." We start with linguistic representations that a nativist should recognize, and show that domain-general principles support learning of the nominally correct grammar, contra specific unlearnability claims in the literature. This allows us to engage the poverty-of-stimulus argument in its own representational terms, while working "outwards" to domain-generality. In contrast, connectionist studies that also question the poverty of the stimulus (e.g. Reali & Christiansen, 2005) work "from the outside in." They start with domain-general representations, and learn linguistic behavior similar to that of a grammar. The two approaches complement each other: the starting-point for connectionist studies is undeniably domain-general, while in our case that which is learned is undeniably a grammar.

Acknowledgments

We thank Susanne Gahl and four anonymous reviewers for helpful comments. This research was partially supported by NIH fellowship HD049247 to SF.

References

Baker, C. L. (1978). *Introduction to generative transformational syntax*. Englewood Cliffs, NJ: Prentice Hall.

Baker, C. L. (1989). *English Syntax*. Cambridge, MA: MIT Press.

Bowen, R. (2005). Noun complementation in English: A corpus-based study of structural types and patterns. *Gothenburg Studies in English 91*. Göteborg, Sweden: Acta Universitatis Gothoburgensis.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures.* Berlin: Mouton de Gruyter.

Fillmore, C. J. (1965). *Indirect object constructions in English and the ordering of transformations.* The Hague: Mouton.

Hornstein, N., & Lightfoot, D. (1981). Introduction. In N. Hornstein, & D. Lightfoot (Eds.), *Explanation in linguistics*. London: Longman.

Huddleston, R., & Pullum, G. K. (2002). *Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Keizer, E. (2004). Postnominal PP complements and modifiers: a cognitive distinction. *English Language and Linguistics,8,* 323-350.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104(2)*, 211-240.

Lidz, J. & Waxman, S. (2004). Reaffirming the poverty of the stimulus argument: a reply to the replies. *Cognition, 93,* 157-165.

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.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah, NJ: Erlbaum.

Perfors, A., Tenenbaum, J., & Regier, T. (2006). Poverty of the stimulus? A rational approach. *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 663-668). Mahwah, NJ: Erlbaum.

Radford, A. (1988). *Transformational Grammar: A First Course*. New York: Cambridge University Press.

Reali, F. & Christiansen, M. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science 29*, 1007-1028.

Regier, T., & Gahl, S. (2004). Learning the unlearnable: The role of missing evidence. *Cognition, 93,* 147-155.

Suppes, P. (1974). The semantics of children's language. *American Psychologist, 29,* 103-114.

Taylor, J. R. (1996). *Possessives in English: An exploration in cognitive grammar*. Oxford, UK: Oxford University Press.