# Empirical assessment of stimulus poverty arguments[1]

GEOFFREY K. PULLUM AND BARBARA C. SCHOLZ

*Abstract*

*This article examines a type of argument for linguistic nativism that takes the following form: (i) a fact about some natural language is exhibited that allegedly could not be learned from experience without access to a certain kind of (positive) data; (ii) it is claimed that data of the type in question are not found in normal linguistic experience; hence (iii) it is concluded that people cannot be learning the language from mere exposure to language use. We analyze the components of this sort of argument carefully, and examine four exemplars, none of which hold up. We conclude that linguists have some additional work to do if they wish to sustain their claims about having provided support for linguistic nativism, and we offer some reasons for thinking that the relevant kind of future work on this issue is likely to further undermine the linguistic nativist position.*

1. This work was partially supported by grants from the Humanities Division at the University of California, Santa Cruz. Some of the material was presented in an earlier form by Pullum in the Parasession on Learnability at the Berkeley Linguistics Society's annual meeting in February 1996, and appeared as Pullum (1996). Early versions of some of the research were also presented to the Australasian Association for Philosophy (July 1996), a graduate seminar on learnability at UCSC (fall 1996), and the TABU Day conference at the Rijksuniversiteit Groningen, Holland (June 2000). For stimulating comments, questions, discussion, suggestions, assistance, encouragement, references, advice, and support we are grateful to many friends and colleagues, including Nameera Akhtar, Cher Bergeon, Melissa Bowerman, Joyce Tang Boyland, Dave Chalmers, Matt Chisholm, Eve Clark, Jane Edwards, Julia Fernandez, Gerald Gazdar, Nick Guilford, Chris Gunlogson, Stevan Harnad, Robert Levine, Brian MacWhinney, James McCloskey, Lise Menn, Frederick Newmeyer, Alva Noe, Daniel Osherson, Chris Potts, Nancy Ritter, Ivan Sag, Geoffrey Sampson, Pieter Seuren, Dan Slobin, Roger Wales, Rachel Walker, Tom Wasow, and Michael Wescoat. We have not taken all of their advice, and they bear no responsibility for the errors in which we have persisted. Mailing address for the authors: Geoffrey K. Pullum and Barbara C. Scholz, Stevenson College, UCSC, 1156 High Street, Santa Cruz, California 95064, USA. Email: pullum@ling.ucsc.edu, scholz@ling.ucsc.edu.

The Linguistic Review 19 (2002), 9–50 0167–6318/02/019-09

## 1. Introduction

It is widely believed, outside of linguistics as well as within it, that linguists have developed an argument for linguistic nativism called the 'argument from poverty of the stimulus'. Linguistic nativism is the view (putatively reminiscent of Cartesian rationalism and opposed to empiricism) that human infants have at least some linguistically specific innate knowledge. In this article, we separate out one particularly striking argument type from the many that have been associated with the appellation 'argument from poverty of the stimulus'. We subject the form and empirical warrant of the selected argument type to a close examination. The argument type in question turns on the claim that during the language acquisition process, children often come to know things about the language they are acquiring despite not having access to the crucial evidence that shows those things to be true of the language. We argue that linguists have not achieved what they are widely thought to have achieved. The empirical work that would be needed to support their claims has not been carried out.

The argument of this article casts doubt on a certain style of argument for linguistic nativism, and rebuts some particular instances of that argument. It does not challenge linguistic nativism directly, or claim that it is false. Linguistic nativism comes in many forms, some of them trivial, like the view that because frogs and spiders never learn human language "we must be endowed with a predisposition to learning language that these other organisms lack" (Smith 1999: 39). We have no quarrel with that platitude. (It is rather odd to find Smith counting it among Chomsky's "striking observations" and "radical suggestions": 1999: 38.) Certainly, humans are endowed with some sort of predisposition toward language learning. The substantive issue is whether a full description of that predisposition incorporates anything that entails specific contingent facts about natural languages.

We also caution the reader against reading this article as a defense of the claim that purely empiricist language learning, via domain-unspecialized algorithms for knowledge acquisition, can suffice for learning natural languages, given children's experience. It may well be that purely empiricist learning of natural languages can indeed be shown to be impossible.[2] That possibility is left open here. We are concerned not with whether empiricist claims about language learning are true, but we are concerned with whether linguists have established their falsity.

This article seeks to shift the onus, and show that there is work to be done to support the empirical premises of a certain style of argument for linguistic

2. Many seem to think that Gold (1967) has already shown this. We believe this grossly overstates what Gold shows; see pp. 194–196 of this issue for some discussion.

nativism. We do not suggest that such work is impossible. Indeed, we suggest a way of conducting it. But a program for research that might one day be undertaken should not be confused with one already completed. That is what has been happening up to now.

In Section 2 we identify the type of argument we are interested in and distinguish it from the plethora of other arguments that have been confused with it. In Section 3 we lay out explicitly what is required to provide empirical support for this type of argument. In Section 4 we work through four case studies from the literature in which we find linguists accepting the requisite burden of empirical proof, and show that in all those four cases they have failed. Two of these four cases are very famous, and are cited in an extremely broad range of work. If even these fail to show what they are supposed to show, then there is indeed work to be done. Section 5 offers a brief conclusion, and points out a way in which work on defending the argument for nativism is likely to be self-undercutting.

## 2. Will the real argument please stand up

The term "argument from poverty of the stimulus" first shows up in the linguistics literature in Chomsky (1980b: 34), where Chomsky reviews some very general considerations about learning being "better understood as the growth of cognitive structures along an internally directed course under the triggering and partially shaping effect of the environment," and remarks that he is giving "a variant of a classical argument in the theory of knowledge, what we might call 'the argument from poverty of the stimulus'." (See Thomas this issue, 51–71, for an historical account of the emergence of the poverty of stimulus argument.) Such is the prestige of Chomsky's pronouncements that this scare-quoted but undefined phrase was quickly picked up and repeated by others. The trend has intensified over the past decade, and now the term "argument from poverty of the stimulus" is everywhere: in *Behavioral and Brain Sciences* articles (Chomsky 1980a; Lightfoot 1989; Crain 1991) and articles in *Science* (Bates and Elman 1996; Seidenberg 1997); in books and journals on philosophy (Matthews 1984; Ramsey and Stich 1991; Lightfoot 1998; Cowie 1999), on connectionism (Elman et al. 1996), on biology (Wimsatt 1999), and on Noam Chomsky's works (Smith 1999; McGilvray 1999); it is beginning to turn up in entries in reference books (e.g., Garfield 1994 and half a dozen of the articles in Wilson and Keil 1999); and so on and so on. The putative argument has become a mainstream topic in cognitive science.

Yet no one attempts to *state* the argument. One might have thought that an article entitled 'On the argument from the poverty of the stimulus' would state it, for example. But Wexler's (1991) article under that title provides only a

rather garbled paraphrase of Chomsky's remarks, without any bibliographical reference (Wexler 1991: 253):

> How does the child construct her grammar? In other words, why is the adult output grammar the one that it is? Chomsky's answer notes that the attained grammar goes orders of magnitude beyond the information provided by the input data and concludes that much linguistic knowledge must therefore be innate.

Surely, *how* a grammar is constructed is not the same question as *why* that particular grammar was constructed, yet Wexler appears to conflate the two questions. And surely "orders of magnitude" is not the right phrase (input data sets are finite, and generative grammarians take grammars to generate not finite languages that are orders of magnitude bigger than input corpora, but infinite languages). Moreover, whatever the first two statements mean, it surely does not logically follow from them that "much linguistic knowledge must . . . be innate." Nevertheless, Wexler's claims are fairly typical.

### *2.1. Many premises in search of an argument*

The one thing that is clear about the argument from poverty of the stimulus is what its conclusion is supposed to be: it is supposed to show that human infants are equipped with innate mental mechanisms with specific linguistic content that assist the language acquisition process – in short, that the facts about human language acquisition support linguistic nativism. What is not clear at all is the structure of the reasoning that is supposed to get us to this conclusion. Instead of clarifying the reasoning, each successive writer on this topic shakes together an idiosyncratic cocktail of claims about children's learning of language, and concludes that nativism is thereby supported. Most of the frequently encountered claims are about children's observable accomplishments or aspects of the child's environment.

(1) *Properties of the child's accomplishment*

a. SPEED: Children learn language so fast.
b. RELIABILITY: Children always succeed at language learning.
c. PRODUCTIVITY: Children acquire an ability to produce or understand any of an essentially unbounded number of sentences.
d. SELECTIVITY: Children pick their grammar from among an enormous number of seductive but incorrect alternatives.
e. UNDERDETERMINATION: Children arrive at theories (grammars) that are highly underdetermined by the data.
f. CONVERGENCE: Children end up with systems that are so similar to those of others in the same speech community.

g. UNIVERSALITY: Children acquire systems that display unexplained universal similarities that link all human languages.

(2) *Properties of the child's environment*

a. INGRATITUDE: Children are not specifically or directly rewarded for their advances in language learning.
b. FINITENESS: Children's data-exposure histories are purely finite.
c. IDIOSYNCRASY: Children's data-exposure histories are highly diverse.
d. INCOMPLETENESS: Children's data-exposure histories are incomplete (there are many sentences they never hear).
e. POSITIVITY: Children's data-exposure histories are solely positive (they are not given negative data, i.e. details of what is *un*grammatical).
f. DEGENERACY: Children's data-exposure histories include numerous errors (slips of the tongue, false starts, etc.).

The following examples will serve to illustrate the kind of presentation we are alluding to (many more could be provided).

(i) Ramsey and Stich (1991: 290–91): "a child is exposed to only a very impoverished sample of often misleading linguistic data" that is "remarkably messy" [DEGENERACY]; a competent speaker knows a set of sentences "vastly larger" [PRODUCTIVITY] than "the idiosyncratic set of sentences to which children are exposed" [IDIOSYNCRASY]; and children "are rarely given any indication that certain queer and complex sentences are ungrammatical" [POSITIVITY].

(ii) Garfield (1994: 369)): "(1) the examples of the target language to which the learner is exposed are always jointly compatible with an infinite number of alternative grammars," [SELECTIVITY] "and so vastly underdetermine the grammar of the language" [UNDERDETERMINATION] "and (2) the corpus always contains many examples of ungrammatical sentences" [DEGENERACY] "and (3) there is, in general, no explicit reinforcement of correct utterances" [INGRATITUDE] and thus, "since it is impossible to explain the learning of the correct grammar – a task accomplished by all normal children" [RELIABILITY] "within a very few years" [SPEED] "– on the basis of any available data or known learning algorithms, it must be that the grammar is innately specified, and is merely 'triggered' by relevant environmental cues."

(iii) Seidenberg (1997: 1600): The input to the learner is "degenerate, consisting of both grammatical and ungrammatical sentences" [DEGENERACY], it is also "variable: children are exposed to different samples of utterances" [IDIOSYNCRASY] "but converge on the same grammar" [CON-

VERGENCE], and their exposure "does not include reliable negative evidence" [POSITIVITY].

(iv) Haegeman (1994: 10–11): "The problem of language acquisition has often been summarized in terms of the problem of the **poverty of the stimulus**. Our linguistic capacity, for instance our knowledge of English, goes beyond the evidence we have been exposed to in our childhood. The linguist wants to account for the fact that the linguistic competence is attained in spite of important inadequacies in the stimulus, the linguistic experience. Three types of inadequacies are standardly referred to in the literature. First, we do not just come across grammatical sentences: everyday use of language contains slips of the tongue, hesitations, incomplete sentences ..." [UNDERDETERMINATION; DEGENERACY]. "Second, the experience, i.e. the stimulus, is finite, and we end up being able to produce and process an infinite number of sentences" [FINITENESS; PRODUCTIVITY]. "Third, we acquire knowledge about our language for which we have no overt or **positive** evidence in our experience" [POSITIVITY].

There is little chance of a comprehensive critical discussion of the supposed argument in anything less than a full-length book. It would be like fighting the hydra to tackle such a many-headed beast.[3] Our plan in this article is to be selective, and to pick out a coherent task. We will concentrate on just one specific argument type. We believe it to be the strongest and most productive one for linguistic nativism. It is not actually represented in any of the above quotations, but is made explicit by Hornstein and Lightfoot (1981: 9):

> People attain knowledge of the structure of their language for which *no* evidence is available in the data to which they are exposed as children.

This is not identical with any of the foregoing claims. It is different from incompleteness, which says that children are exposed to only a proper subset of the sentences in the language; it is different from finiteness, which says that the sentences they hear form a finite set; it is different from positivity, which says that they are supplied only with sentences and never with statements about which are the nonsentences. What Hornstein and Lightfoot claim is that some of the sentences children never hear are crucial evidence for learning from experience. That is, some aspects of languages are known to speakers despite the fact that the relevant positive evidence, *although it does exist*, is not accessible to learners during the acquisition process, because of its rarity: linguists can in principle discover it, but children will not.

This claim, if true, would refute non-nativist theories of language learning. If learners come to have linguistic knowledge without being exposed to evidence

3. Sampson (1989) makes an attempt at this, but clearly not an exhaustive one.

that would be crucially necessary for experience-based learning, then learning is not experience-based.

It could be objected that the claim Hornstein and Lightfoot make is too strong. Indeed, Sampson (1989, 1999a) has argued that it is self-contradictory. His argument is as follows. Consider the position of a linguist – let us call her Angela – who claims that some grammatical fact *F* about a language *L* has been learned by some speaker *S* who was provided with no evidence for the truth of *F*. Sampson raises this question: How does Angela know that *F* is a fact about *L*? If the answer to this involves giving evidence from expressions of *L*, then Angela has conceded that such evidence is available, which means that *S* could in principle have learned *F* from that evidence. That contradicts what was to be shown. If, on the other hand, Angela knows *L* natively, and claims to know *F* in virtue of having come to know it during her own first-language acquisition period with the aid of innate linguistically-specific information, then Angela has presupposed nativism in an argument for nativism. That is viciously circular. And there are no other cases to consider, so the argument refutes itself.

As far as we are aware, no defender of linguistic nativism has addressed this point. But what seems to us to be the right response is that a weaker version of Hornstein and Lightfoot's very strong claim would still have considerable interest: we replace total lack of evidence by *lack of evidence that is adequate to the task*. This weaker version is implied by Lightfoot (1982b: 428) when he says there are cases where "there are no data available to the child which will suffice to establish some rule or principle" but the rule or principle is acquired anyway. This is not fully explicit, of course: both "available to the child" and the "will suffice to establish" need clarification.[4] But the idea seems clear: there could be positive evidence for the correctness of some rule or principle, evidence that could be found in language use by an adult linguist doing a really intensive search, but would not emerge in conversational data anywhere near often enough to guarantee that any particular child would ever encounter it. This, we think, is the basis for the argument that most clearly deserves to be referred to as "the argument from the poverty of the stimulus" (henceforth 'APS').

It is important that we are setting aside arguments based merely on cases in which it appears that children would need negative data (definite information that some word sequence is not grammatical) in order to learn some fact. The

4. Lightfoot is surely not suggesting that there could be data that would provide a sufficient condition for knowing that a given grammar is the correct one (by the familiar dictum that theories are underdetermined by evidence, no data can provide that). We assume that to say a set of data "will suffice to establish" a rule or principle is to say that the data constitute an adequate sample for an evidence-based learning procedure to hypothesize that rule or principle. If that were not a logical possibility, there would be nothing to talk about, and Lightfoot would not have needed to make reference to any specific cases.

argument from the (presumed) fact that children get no (or very little) negative information from experience involves not so much stimulus poverty as stimulus absence. Stimulus poverty, as we construe it, involves the learning of facts about what the language is like from an environment that is typically not rich enough to provide crucial evidence that it is like that. Though there is evidence from sentences that can be found through intensive searching of a larger or more varied sample of linguistic material than an individual learner is normally exposed to, the typical learner will never come into possession of the crucial evidence.

Many discussions in the literature shift from this topic to a different one, namely why learners ever assume that anything is ungrammatical. If their environment is all but devoid of systematic information about what is prohibited, how do they come to know that anything is prohibited?

Haegeman, for example, immediately after the passage we quoted above in (iv), gives an example involving the *that*-trace effect, noting that children are not explicitly taught that sentences like **Who did they think that was available?* are ungrammatical. Only negative data, she asserts, could stop children from hypothesizing that every finite complement clause can optionally begin with *that*, which in this case is false for standard English. The puzzle is how they learn that this is *not* permitted from an environment that contains only information about what *is* grammatical.

Addressing this sort of case would take us away from the task we have set ourselves in this article. We can only sketch very briefly what we believe is the right line to take. We question whether children learn what transformational-generative syntacticians think they learn. If the types of constituent that can occur as complements to subordinating words like *that* and *than* are learned piecemeal from positive examples, then the pattern '*more than* + finite VP' will be learned (after examples like *They wanted more than was available*), but 'V + *that* + finite VP' will not be learned, because no examples of that sort will ever be encountered. On the other hand, 'N + *that* + finite VP' will be encountered, and so relative clauses of the type (*things that were available*) will be learned. In sum, piecemeal learning of constructions here would do better than rapid generalizations concerning the scope of *wh*-movement. Transformationalists always assume that the child's problem is that of learning how to constrain an over-hasty generalization concerning the scope of *wh*-movement; but that may not be true. After all, one lesson of Gazdar (1981) is that children do not necessarily learn grammars with any analog of *wh*-movement at all.

Lightfoot (1998: 585) similarly presents an argument from absence of stimulus rather than poverty of the stimulus. He asks how it is learned that reduced forms of auxiliaries like *is* do not occur in examples like **Kim's taller than Jim's* (compare the grammatical *Jim's happy* and *Kim is taller than Jim is*). Again, there is an issue about what is learned. Lightfoot does not go into de-

tail, but seems to assume that the child faces the problem of curbing an overhasty generalization concerning the applicability of a very general rule saying that *is* may be reduced to *'s* in any context. Lasnik (1989: 92–94) assumes the same thing in an argument about learnability based on auxiliary reduction, and Lightfoot may have that analysis in mind. Again, there are alternatives. One was offered by Pullum and Zwicky (1997). It would have very different implications for what was to be learned. Pullum and Zwicky give evidence that auxiliaries (like *is*) take their clitic forms (like *'s*) only when completely stressless. Certain syntactic constructions require the presence of weak stress on a lexical head. One of these is the construction in which a VP consists solely of a lexical head. Learning expressions like *Jim is* (with weak stress on the lexical head *is*) and learning expressions like *Jim's tall* (with zero stress on an auxiliary verb that has a complement) could proceed separately, neither of them implying that **Jim's* would be grammatical on its own.

For reasons of space, we do not attempt to give further details here. There is of course much more to be said about whether negative data are crucially necessary for learning natural languages from it. The topic raises a much bigger problem for generative linguistics, which we hope to address in future work. All we will say here is that for those readers who are disappointed that we will not treat the negative data issue further, and feel that the APS we do discuss is not the one that really deserves the name, we can only suggest that they reinterpret our abbreviation "the APS" to stand for "the Argument selected by Pullum and Scholz."

### *2.2. The crucial empirical premise*

It will be helpful if we now sharpen our discussion by explicitly representing the APS argument style in deductive form. Let us distinguish two ways children might in principle learn languages. The first, *innately primed* learning, calls upon inborn domain-specific linguistic information (called here *innate priming*).[5] The second, *data-driven learning*, does not; rather, it relies entirely on generalization from experience by the ordinary methods that are also used for learning other (nonlinguistic) things from experience. Suppose we simply assume for the sake of argument that these are the only possible alternatives.[6]

---

5. It is seldom explicitly noted that the innate priming hypothesis, at least as defended by Chomsky, posits not just some kinds of innate linguistic knowledge but innate knowledge of *contingent* linguistic facts or principles. Definitional truths about languages, or consequences of laws of other sciences, do not suffice. What linguists are seeking, and claim in some cases to have found, is evidence that some things about languages *that could have been otherwise* are known to the human infant at birth.
6. Stich 1979 may be read as denying this premise by sketching several views that fall between

Then we can state the APS as follows:

(3) *The Argument from Poverty of the Stimulus (APS)*

a. Human infants learn their first languages either by data-driven learning or by innately-primed learning. [Disjunctive premise; by assumption.]
b. If human infants acquire their first languages via data-driven learning, then they can never learn anything for which they lack crucial evidence. [By definition of data-driven learning.]
c. But infants do in fact learn things for which they lack crucial evidence. [Empirical premise.]
d. Thus human infants do not learn their first languages by means of data-driven learning. [From (b) and (c), modus tollens.]
e. Conclusion: human infants learn their first languages by means of innately-primed learning. [From (a) and (d), disjunctive syllogism.]

Determining the soundness of this (clearly valid) argument turns on determining whether the empirical premise (3c) is true. Some researchers outside of linguistics are convinced that linguists have established this beyond doubt. For example, Ramsey and Stich (1991: 296) assure their audience of philosophers and connectionists:

> For the argument [from poverty of the stimulus] to be persuasive there must be a substantial number of examples in which the choice between two equally simple and natural grammars can be made only by appealing to the sort of abstruse evidence that is unlikely to be found in the primary linguistic data.[7] There is by now a substantial collection of plausible cases in the literature.

And Lightfoot says much the same thing about the achievements of linguists:

> The linguistic literature is full of arguments of this type, arguments from the poverty of the stimulus … (1998: 585)

If the the literature were indeed as full of them as such statements suggest, then the APS might deserve its place in reference books and science journals.

---

the 'rationalist' and 'empiricist' extremes of full innate priming for language acquisition and pure data-driven learning. We concede the (possibly false) dichotomy for the sake of argument.

7. Ramsey and Stich's use of the words "equally simple and natural" places a further stringent restriction on what would count as confirming their claim. We will ignore that issue here, for reasons of space, though in a fuller account it would deserve some discussion. Our interest here lies in the claim about needing "abstruse evidence that is unlikely to be found in the primary linguistic data."

Our aim is to determine whether, in the domain of syntax on which most of the discussion has focussed, the "substantial collection" really exists. We begin with an analysis of how to provide support for a particular instance of the APS.

## 3. How to provide empirical support for the APS

The APS depends on the existence of attested cases of rules or principles of grammar that are known despite lack of access to crucial evidence. Call the item of knowledge in question the *acquirendum* (under the empiricist view it is to be acquired on the basis of evidence). To substantiate the claim that some learner knows some acquirendum *A* without having encountered crucial evidence, the linguist needs to identify a set of sentences such that (i) if learners were found to have access to them it would support the case that *A* is acquired from experience, and (ii) access to at least some of them would be necessary to learn *A* from experience; and then it must be shown that the learner has acquired *A* without such access. We will call the set of crucial sentences the *lacuna* (because from the learner's point of view it represents a gap in experience). So the task breaks down into five parts:

(4) *APS specification schema*

a. ACQUIRENDUM CHARACTERIZATION: describe in detail what is alleged to be known.
b. LACUNA SPECIFICATION: identify a set of sentences such that if the learner had access to them, the claim of data-driven learning of the acquirendum would be supported.
c. INDISPENSABILITY ARGUMENT: give reason to think that if learning were data-driven then the acquirendum could not be learned without access to sentences in the lacuna.
d. INACCESSIBILITY EVIDENCE: support the claim that tokens of sentences in the lacuna were not available to the learner during the acquisition process.
e. ACQUISITION EVIDENCE: give reason to believe that the the acquirendum does in fact become known to learners during childhood.

These five parts of the job are rather different in status and disciplinary connections. Acquirendum characterization falls within descriptive linguistics, and involves providing in exact terms some fragment of a grammar, or statement entailed by a grammar, for some actual language. Lacuna specification is a task in formal language theory: giving a definition of some set of sentences. Giving an indispensability argument is in principle a mathematical task, falling

within formal learnability theory, which has provided numerous interesting results about what sorts of input permit what sorts of languages to be identified (see Jain et al. 1999 for an introduction to this rich literature). Providing inaccessibility evidence is an empirical matter of identifying the content of the actual sequence of utterance tokens observed (or probably observed) by the learner during the first language acquisition period.[8] And finally, obtaining acquisition evidence is a matter of experimental developmental psychology.

A deliberately implausible example will help illustrate. Take the target language to be standard English, let the acquirendum be that *By and large, things are fine* is grammatical, and let the lacuna be the set of English sentences containing the phrase *by and large*. This curious idiom has an unprecedented structure, apparently something like $[_{\text{PP}}[_{\text{P}}\ by\ ][_{\text{Adj}}[_{\text{Cnj}}\ and\ ][_{\text{Adj}}\ large\ _{\text{Adj}}]\ _{\text{Adj}}]\ _{\text{PP}}]$. Obviously it could never be learned from observations of expressions that do not contain it, since its structure is unparalleled by that of any other phrase. That is, no procedure can learn from linguistic data that some sentences in English containing *by and large* are grammatical if the input completely lacks examples of that expression.[9] Suppose some learner were definitely known never to have been exposed to even a single utterance containing *by and large*, and yet when asked about *By and large, things are fine* this learner reported that it is grammatical (and perhaps, for good measure, could say what it meant, and judged expressions containing **by and small* or **of and large* to be ungrammatical as well). We would then have everything we needed to support an instance of the APS style. On the other hand, if the learner had definitely heard instances of the *by and large* construction, that would support the idea that data-driven learning might have taken place.

This example is silly, because *by and large* is a unique asyntactic idiom, and no one would be tempted to imagine it could be learned by someone who had never heard it; it would not be expected to occur in English on the basis of the rest of English syntax. However, linguists who accept the APS claim there exist other instances of the same general form that are not silly, instances in which the claim of learning without evidence is actually true.

---

8. It is most curious that generative linguists call the problem of how infants acquire language "the logical problem of language acquisition" (the phrase appears as the title of Baker 1981 and the subtitle of Hornstein and Lightfoot 1981, and is repeated in many other places in the literature). Whether sentences with a certain structural characteristic occur in the child's experience cannot be settled by mere ratiocination. Working on the question demands access and attention to concrete empirical data about utterance events and infants' perceptions of them and reactions to them.
9. Notice that there are infinitely many sentences of English containing *by and large*, so their absence from the learner's input would be a worse defect in experience than what is called "incomplete text" in the learnability literature (Jain et al. 1999: 168).

## 4. Empirical linguistic testing of inaccessibility claims

The ideal way to test inaccessibility claims directly would of course be to conduct an examination of every utterance used in the presence of some specific infant during the first five years or so of the infant's life. This is not feasible at present; no one has yet captured in machine-readable form such a massive corpus for any learner. Indeed, Chomsky, with perhaps just a touch of hyperbole, questions the sanity of anyone who would think of undertaking the project:

> Of course, in order to demonstrate that there is no relevant experience with respect to some property of language, we really would have to have a complete record of a person's experience – a job that would be totally boring; there is no empirical problem in getting this information, but nobody in his right mind would try to do it. So what we can do is to find properties for which it is very implausible to assume that everyone has had relevant experience. (Piattelli-Palmarini 1980: 113).

We ourselves would not charge someone with being out of their mind for attempting a full input corpus for a designated infant; it would be an extremely useful contribution to science. But Chomsky's last point is nonetheless correct: useful work can still be done by looking for cases where it is implausible to assume accessibility. The lack of a complete corpus of the input for any given learner will not be an insuperable obstacle provided only that corpora of English text do not vary dramatically with respect to the constructions that they contain. Our default assumption will be this: *a construction inaccessible to infants during the language acquisition process must be rare enough that it will be almost entirely absent from corpora of text quite generally*.

Now, this assumption could be wrong. We must stress that our focus is on method, and on implications for further research. The corpus work that we are suggesting is necessary has not yet been undertaken, and we do not claim to have undertaken it. We will cite some results from corpus searches of various sorts, be we do not think our samplings constitute definitive experiments; they are merely suggestive pilot studies.[10]

We do make some efforts to consider different kinds of text, and to make sure that language specifically addressed to or known to be encountered by

10. The pre-pilot reported in Pullum (1996) has unfortunately been cited outside of linguistics as straightforward disconfirmation of APS claims (see e.g. Cowie 1999; Osherson 1999; Sampson 1999a, 1999b), while within linguistics it has not been cited at all. Pullum's paper contains only very preliminary empirical work, intended to encourage nativists to go beyond self-congratulatory assurances and vague promissory notes and collect some data. Like the present article, Pullum (1996) should be read as calling for empirical work on the question of whether nativist claims are true, not as answering that question.

children is included. But, although it is possible that there are major differences between the relevant syntactic properties of different kinds of text, and that the texts we have had access to are therefore not suitable ones, we think it would be premature to let that deter us from taking a look at available corpus data. There are some good reasons for expecting considerable constancy. Consider the surprising result of Hudson (1994) that about 37 % of the word tokens in running text are nouns. This odd finding seems to hold essentially independently of genre, style, modality, source, or even language (Hudson's evidence goes beyond English). Such a robustly unchanging figure suggests great stability in the syntactic characteristics of utterances across genres.

We do not know how much remains constant across genres in this way. Frederick Newmeyer (personal communication) has pointed out to us that there are certainly some differences that are of sufficient magnitude to be relevant. For example, Biber (1988) finds that about 31 past participial reduced relatives are found per 10,000 words of journalists' prose, but only one occurs in 10,000 words of typical face-to-face conversations. And Thompson (1988) found that of more than 300 adjectives in spontaneous discourse, only one occurred prenominally in a definite NP, whereas the rate is much higher in printed prose.

On the other hand, we have noticed that some features of English have the opposite property, being much more frequent in spoken language than in printed prose. For example, it would appear that consecutive synthetic and analytic negations (e.g. *We can't not go*) are edited out of printed sources such as journalistic prose. They are extremely rare in print. This is not because of any rarity of *n't* tokens: among more than 222,000 tokens of *n't* in a diverse collection of machine-readable texts, we found just two occurrences of *n't not*, both in quotes from spontaneous speech. For a construction like this, it would be easy to *under*estimate the frequency of examples in the spoken language if one looked only at written sources. We have found it easy to spot further examples in spontaneous conversation, film dialogue, and the like (there are several instances in the screenplay of *Sliding Doors*, for example). PPs as subjects (*Under the sink needs some attention*) and direct objects (*I gave behind the cooker a new coat of paint*) also appear to be almost impossible to find in printed sources, but they occur conversationally and can be found in film scripts. Such extreme rarity in print would lend a spurious plausibility to the idea that these constructions were impossible to learn from examples in a typical linguistic environment. Hence the genre argument cuts both ways.

We do acknowledge that genre differences represent a real issue. Too little is known about the direction and effect of differences in construction frequency across genres. Here we tentatively assume that differences between styles of speech or contexts of utterances are not so enormous as to make tape recordings of spontaneous conversation in front of children the only relevant kind of

material. We assume that research on one kind of text will tell us at least something about the syntactic structures likely to be found in others. This seems to us a sensible assumption because the main determinant of how often particular constructions turn up will be the random exigencies of what needs to be said, and we take it that these random exigencies will be roughly constant across many genres. This may be a false assumption. But it is important to see that it is the defenders of the APS who need to establish its falsity. Moreover, they have to show quantitatively that it is false to a degree that makes a difference in the relevant cases. It is part of our aim to highlight the need for such work, and to encourage people to undertake it.

We also need to know much more about what sort of utterances constitute the typical input to children, and what they pay attention to in that input. The work begun by Snow and Ferguson (1977) needs to be extended, particularly through large corpus collection. But it cannot be assumed without argument that children only get some debased input consisting of snippets of baby talk. At least one researcher who has spent time looking at the way adults speak to children reports that "the input to young children is neither so depleted nor so uniform as some have suggested" (Berman 1990: 1160). Berman also notes that speech addressed to the learner is not the only evidence about the language that the learner has; children learn from at least some utterances that are not addressed to them. And indeed, Weddell and Copeland (1997) found that even children between ages 2 and 5 pick up and understand much more (e.g., about news events) from the language they hear on television than their parents typically realize.

One other point is that we doubt whether ordinary speakers of English have enough conscious control over the syntactic properties of their utterances to enable them to switch their syntactic complexity level in any really significant way as they move from boardroom to street to kitchen to classroom to church to school. Certainly there are constructions limited almost entirely to written prose; but these are not relevant to early language acquisition, since young children do not learn them early on.

What we claim will not be found is a construction that is learned early and occurs centrally and productively in adult speech but slumps to zero frequency in performance when young children are around.

We now proceed to consider the four best and clearest candidates we know of for membership in the "substantial collection of plausible cases" with which linguists are thought to have provided support for the APS, and we will examine the extent to which they make good on establishing all five components of the APS specification schema in (4).

### *4.1. Plurals in noun-noun compounding*

The majority of this article concerns the APS in the domain of syntax, but it will be useful to first take a look at an example involving word formation set out in Gordon (1986), because it constitutes an interesting attempt to flesh out a specific instance of the APS. This work has been cited in popular books like Gazzaniga (1992: 92–93) and Pinker (1994). Gordon performed some experiments that appeared to show that children between the ages of 3 and 6 use irregular plurals as first elements of compounds to form novel items like *teeth-eater* and *mice-eater* but do not use regular plurals in a similar way to form *toys-eater* or *rats-eater*. Gordon used an analysis of the million-word Brown corpus to verify that compounds of the former sort, with irregular plural forms of the words *mouse*, *man*, *tooth*, *foot*, and *goose* have a very low text frequency (3 tokens, compared to 153 tokens with the singular forms of these words). Gordon takes this to be too low a frequency of occurrence for the child to be able to form, on the basis of the evidence, a belief that such compounds are possible. Gordon's instance of the APS, then, involves these elements:

(5) *APS specification for regular plurals in compounds*

- a. ACQUIRENDUM: irregular but not regular plurals can be used as the non-head (non-rightmost) part of a compound.
- b. LACUNA: the set of all sentences containing compounds with irregular plurals as the non-head parts of compounds.
- c. INDISPENSABILITY ARGUMENT: the acquirendum cannot be learned without access to sentences containing compounds with irregular plural non-heads.
- d. INACCESSIBILITY EVIDENCE: the text frequency of compounds with irregular plurals as non-head components is extremely low.
- e. ACQUISITION EVIDENCE: purportedly achieved by the experiments reported in Gordon (1986).

We have some concerns about Gordon's experimental design. Gordon trained children to answer questions like *What would you call an* X *eater?* (for specific values of *X*) by giving them mass noun values of *X* (since with those the issue of pluralization does not arise), and one of the three mass nouns used was *rice*. Thus the children who were being tested to see if they would coin a compound like *mice-eater* had been induced to utter *rice-eater* just minutes before. The likelihood of some priming effect – a greater likelihood of *mice-eater* being uttered because the rhyming term *rice-eater* was still fresh in the child's memory – seems rather high.

But this is not worth worrying about compared to the major problem with Gordon's work, which is that the acquirendum he is testing for is incorrect. Regularly inflected plurals do occur as non-heads in compounds.

Gordon's conception of compounds is very broad, because he mentions (p. 76) not only N + V +*-er* forms like *rat-eater* but also compounds like *tooth-brush* (which, counter to the generalization he is suggesting children know, is much more acceptable than ?*teethbrush*), *mouse-trap*, *mice-infested*, etc. But such details make little difference. Gordon is assuming the truth of level-ordered morphology (Kiparsky 1982). We will not summarize this hypothesis in detail; it suffices that it entails that irregular plurals will appear as non-head parts of compounds but regular plurals will not. The trouble is that the relevant claims do not hold for any class of compounds that we have been able to identify. Among compounds of the type Noun[plural] + Verb + *er*, compounds like *chemicals-maker*, *forms-reader*, *generics-maker*, and *securities-dealer* are quite common. And these are not pluralia tantum cases (which behave differently, as seen from *jeans-maker*, *clothes-dryer*, *scissors drawer*, etc.): it is perfectly possible to have a factory making just one chemical, but a chemicals factory makes more than one, just as a forms reader reads arbitrary different forms rather than just one form, and so on.

When we turn to compounds with past participles (like *rat-infested* and *mice-infested*), we find far more cases of regular plurals in the first element. In modern journalistic prose such compounds are extremely common, comprising an open class: *issues-oriented*, *citizens-sponsored*, *profits-based*, *affiliates-led*, etc.[11] Such compounds suggest that Gordon has neither characterized the acquirendum correctly nor supported the inaccessibility claim properly.

So far we have quite artificially restricted ourselves to examples from printed sources in which plurals precede hyphens, signalling the compound status orthographically. But hyphens do not have to occur in compounds (there is a great deal of variation on this point among writers). Most compounds with regular plural non-heads do not have the hyphen. Thus we have so far ignored a large number of further cases, indistinguishable in speech from hyphenated ones: *drinks trolley*, *rules committee*, *publications catalogue*, *parks commissioner*, *programs coordinator*, *buildings inspector*, *faces research*, *letters policy*, *complaints department*, *claims backlog*, *counterexamples list*, and many others that Gordon admits (p. 88) are clearly further embarrassment to the level-ordering hypothesis that he takes to be known but not learnable from the evidence.

Also ignored in our discussion so far are cases in which the first part of the compound is itself complex. *Vital records protection plan* occurred in the

11. The *Wall Street Journal* 1987–89 (Linguistic Data Consortium 1993), henceforth *WSJ*, contains *admissions-related*, *citizens-sponsored*, *commodities-oriented*, *features-loaded*, *futures-led*, *imports-fed*, *issues-oriented*, *materials-related*, *metals-related*, *numbers-oriented*, *operations-oriented*, *options-backed*, *orders-related*, *premiums-associated*, *products-oriented*, *profits-based*, *ratings-obsessed*, *receivables-backed*, *results-based*, *rules-based*, *sales-linked*, *services-oriented*, *shares-valued*, *systems-oriented*, *weapons-related*, and scores of others.

widely syndicated *Dilbert* comic strip that appeared in most U.S. newspapers on October 1, 2000. It exemplifies a type also found in *new books shelf*, *equal rights amendment*, *displaced persons camp*, and so on. These are very common. Children's reactions to them are discussed in Alegre and Gordon (1996).[12] Evidence is cited to show that children apparently grasp the principle that a new books shelf is a shelf of new books, not a new shelf of books (specifically, that a red rats eater is not typically red but rather typically eats red rats). But the semantic speculations and theoretical complications involved in this article are unconvincing. A lot more is going on in the data than the simple generalization that Gordon (1986) studied. The net effect of Alegre and Gordon (1996) is to undermine the specification of the acquirendum in Gordon (1986).

There is some evidence that no universal generalization can describe which plurals can occur in which types of compound. Johansson (1980), cited by Sampson (1999a: 173n), studies plural nouns as non-head elements in compounds ("attributive nouns") in detail, and notes that British dialects favor regular plural non-head elements considerably more than American dialects. British English seems more inclined to employ compounds of this sort: where National Public Radio in the USA will use the term *a drug problem*, the BBC is likely to use *a drugs problem*. This is not compatible with the view that universal grammar dictates the principle involved. Even if the relevant kinds of example were completely absent from American English, which they certainly are not, the fact of their occurrence in British English would refute Gordon's claim that the root of children's acquisition of the American pattern lies in constraints of universal grammar that cannot be learned.

Until a much clearer account is given of why the many occurrences of regular plurals in compounds do not falsify Gordon's alleged acquirendum, his work does not support the APS. This is not to say that we have an alternative theory of how children learn compounds. There is obviously a great deal that is not understood, and Gordon's ingenious experiments may have a role to play in our ultimately understanding it. But Gordon's research does not, at its present state of development, support innate priming.

Rather, we think the importance of Gordon's program lies in the fact that it is exactly the sort of empirical research that must be done to support the APS, including research to support inaccessibility claims. It is to Gordon's credit that he did check corpus frequencies for the supposedly crucial data, namely irregular plurals in compounds, albeit from the admittedly small Brown corpus (see Gordon 1986: 76).

---

12. Thanks to Nameera Akhtar for referring us to this article.

### *4.2. Auxiliary sequences*

Kimball (1973: 73–75) offers a clear and instructive instance of the APS in syntax over twenty years before Gordon's paper, antedating by some years the first appearance of the term 'poverty of the stimulus'.[13] Kimball presents the familiar facts of English auxiliary sequences:

(6) a. *It rains*.
b. *It may rain*.
c. *It may have rained*.
d. *It may be raining*.
e. *It has rained*.
f. *It has been raining*.
g. *It is raining*.
h. *It may have been raining*.

He takes these to be correctly described by a rule of the sort advocated in Chomsky (1957):

(7) $Aux \rightarrow T(M)(have + en)(be + ing)$.

This is in fact not a phrase structure rule but a finite rule schema; parentheses mark optional elements: '$A \rightarrow B(C)D$' abbreviates the ordered set $\{A \rightarrow BCD, A \rightarrow BD\}$. Kimball notes that if example (6h) were not grammatical, the rule schema would have had to be more complicated, and would not be statable in one line: the optional material following $T$ would have had to be broken down into three subcases, one covering $(M)(have + en)$, one covering $(M)(be + ing)$, and one covering $(have + en)(be + ing)$.

Kimball then raises the question of how the child learning English learns that the simpler rule is the right one, which entails learning that sentences like (6h) are indeed grammatical:

> It may be thought that the language learner acquires the full system of the auxiliary by hearing examples of sentences of each type, i.e., sentences like each of those listed... However, sentences in which the auxiliary is fully represented by a modal, perfect, and progressive are vanishingly rare. One computerized sample of more than a million sentences from educated, written text yielded no sentence like [*It may have been raining*]. Further, one observer remarked that after eight years of attention to the problem, he had found [such] sentences ... fewer than a dozen times in conversation.
> Thus, the evidence indicates that a great many English-speaking children will acquire the full auxiliary system ... without having heard sentences directly illustrating each of the rules. (Kimball 1973: 74)

13. Thanks to Jim McCloskey for reminding us of the case discussed in this section.

Kimball is saying that children know that sentences with a finite auxiliary plus two nonfinite auxiliaries plus a nonauxiliary verb in sequence are grammatical, and they know what the right order is out of the logically possible orders, without being presented with crucial evidence. He offers an innate priming hypothesis to explain this: learners evaluate the complexity of candidate grammars containing rule schemata using the parenthesis notation. He conjectures that learners acquire some subset of the rules by reference to actually observed patterns in utterance tokens, "and then deduce the existence of the remaining rules on the grounds of simplicity." We can summarize the evidence for innate priming that he presents as in (8).

(8) *APS specification for auxiliary sequences*

- a. ACQUIRENDUM: the rule schema (7).
- b. LACUNA: the set of all sentences exhibiting the sequence 'tensed modal auxiliary + perfect *have* + progressive *be* + present participial verb' (hereafter, MHBV sequences).
- c. INDISPENSABILITY ARGUMENT: without hearing examples containing an MHBV sequence it is not possible to learn that such sequences are grammatical.
- d. INACCESSIBILITY EVIDENCE: clauses containing an MHBV sequence are "vanishingly rare".
- e. ACQUISITION EVIDENCE: trivial, since it is undisputed that everyone who speaks English knows that MHBV sequences are grammatical.

We consider first the inaccessibility claim, which is unfortunately left quite vague. When Kimball was writing we did not have the tools we have today for searching corpora. Looking in the relatively small and in some ways unsuitable texts available to us (and we do not intend to suggest in any way that our preliminary testing represents a definitive result), we find hundreds of examples. There are some in almost every novel accessible to us in machine-readable form. Here is a brief sampling, with the auxiliary sequences boldfaced:

(9)
- a. (i) *"You should have been attending to the lesson"*
  (ii) *"… when I should have been studying"*
  (Lucy Maud Montgomery, *Anne of Green Gables*)
- b. (i) (from Jonathan Harker's diary) "*I must have been sleeping soundly*"
  (ii) (from Lucy Westenra's diary) "*I must have been falling asleep*"
  (Bram Stoker, *Dracula*)
- c. (i) *"… might have been exchanging jokes"*
  (ii) *"… must have been thinking of himself"*

(iii) *"… might have been alluding"*
(iv) *"… might have been discussing"*
(v) *"… might have been hiding in the bushes"*
(Joseph Conrad, *Lord Jim*)

d. (i) *"… must have been drinking"*
(ii) *"… must have been dreaming"*
(Herman Melville, *Moby Dick*)

e. (i) *"I should have been dressing the children"*
(ii) *"your sister would have been living now"*
(Emily Brontë, *Wuthering Heights*)

f. *"We stand together again at the steps of this symbol of our democracy – or we would have been standing at the steps if it hadn't gotten so cold. Now we are standing inside this symbol of our democracy."*
(President Ronald Reagan's inaugural address [delivered inside the Capitol instead of the usual place, the steps of the Lincoln Memorial])

Given the application to language acquisition, it might be suggested that only material definitely addressed to children is relevant. But examples such as the following, from classic children's books, show that this challenge can be met:

(10) a. *"must have been dreaming"*
(Lewis Carroll, *Through the Looking Glass*)

b. *"You must have been thinking again"*
(L. Frank Baum, *The Wonderful Wizard of Oz*)

c. (i) *"It would not have been fighting fair"*
(ii) *"Oh, surely she must have been dreaming"*
(J. M. Barrie, *Peter Pan*)

How many cases does one have to find to challenge Kimball's claim that these constructions are "vanishingly rare"? The question is not just rhetorical; there is an empirical issue here. It is necessary for the APS advocate to tell us how many examples it is supposed to take before a child is likely to pay attention, and then compare that with how many such examples turn up in actual language use. This involves knowing something about both input (what utterances occur in the child's vicinity) and uptake (what the child pays attention to).

Phrases like *must have been drinking* and *must have been dreaming* are very common. These very word sequences turn up several times each in the small sample of novels that we quote above, and they are not at all literary in flavor. It is probable that almost everyone hears such phrases at some time early in life. How many would be enough? We do not know. It is simply not clear what

would support and what would undermine Kimball's inaccessibility claim.[14] On the one hand, we do find scores of solid, unremarkable, attested examples of Modal–*have*–*be*–V+*ing*. On the other hand, they certainly occur in only a small number of the sentences in those corpora we have examined – our rough estimate is that one such sequence will be encountered every 3,000 to 4,000 sentences. If that is not enough, the APS advocates should tell us why not.

Kimball's work illustrates one other thing, a common shortcoming of APS cases in the literature: his acquirendum is arguably not what learners actually come to know. His argument depends entirely on the assumption that the rule schema (7) is the acquirendum, and this is no longer compatible with what most syntacticians believe.

Analyses of the auxiliary that involve a rule schema like (7) had been challenged before Kimball's (1973) publication, in Ross (1969). Those challenges have since been significantly strengthened (Pullum and Wilson 1977; Gazdar, Pullum and Sag 1982; Falk 1984; etc.). It is now the consensus that all (or at least most) English auxiliaries are verbs taking complements (either VPs or subjectless clauses). The evidence for this view is strong, and so are the arguments that analyses involving rule schemata like (7) fail to describe English syntax.

There are significant consequences for learnability. If auxiliaries are complement-taking verbs there is no need to assume that strings like *may have been writing* must be heard before their grammaticality can be known. No one thinks that an example like *is able to avoid appearing to be drunk* must be heard before similar structures are known to be grammatical. The adjectival lexeme *able* can occur after the copula form *is* because any predicative adjective can occur after any form of *be*; *to* can occur after *able* because *able* selects infinitival complements; *avoid* can appear after *to* because it is the bare infinitive form of a verb, namely *avoid*; *appearing* can appear after *avoid* because *avoid* takes complements with present participle (*ing*-suffixed) verbs; and so

14. The problem of apparently quantitative claims that actually have no numerical content has appeared before in linguists' writings. Compare the acquisition speed claim in (1a). As Sampson (1989) notes, we are entitled to ask how fast is fast. And those who have claimed children's language acquisition is fast have failed to do any benchmarking and comparison with other tasks. The conclusions of the thorough literature review in Singleton (1989) are not supportive: "It seems unlikely, in view of the evidence, that there is a *particular* point in a child's post-natal development where language acquisition can be said to truly begin. It is also unlikely that there is a particular critical point where the capacity to learn language goes into a general and rapid decline." Singleton also finds some reason to believe that "first language acquisition normally continues, in some of its aspects at least, right into old age" (p. 265). Our point does not presuppose that Singleton is correct; the point is that linguists who repeat the dogma about the early onset and rapid course of first language acquisition never even allude to evidence, or acknowledge that it would be useful to undertake a survey of the kind Singleton has undertaken.

on. All of this can be learned from examples containing one item acting as head of the complement of another.

It can be verified from simple examples that verbs like *may* occur with bare infinitival complements, *have* occurs with past participial VP complements, *been* is the past participle of *be*, *be* occurs with present participial VP complements, and *writing* is the present participle of *write*. If phrases can be classified into types and heads can be identified as selecting certain types, then the sort of thing seen in the sequencing of the auxiliaries can be learned from examples that do not include any with three auxiliaries in a row, whether such examples occur or not. And it is not denied by APS advocates that subcategorization of heads for certain complement types might be learnable from positive examples. Indeed, there does not seem to be an alternative, since subcategorization is a highly parochial matter.

To posit learning from experience for facts of complement type selection is no more than would be assumed for learning any other structure of parts recombining to make particular types of wholes. We learn – at a very young age – that houses contain rooms, and rooms are of different types such as kitchens and bedrooms; that nearly all houses have both kitchens and bedrooms; that cookers and sinks are found in kitchens, and sinks have taps, and faucets produce water; that dressers are found in bedrooms, and dressers have drawers, and drawers have knobs; and so on. Our experience does not establish for us definitively that dressers with faucets are impossible, or that bedrooms do not contain cookers; yet we come to believe these things too – defeasibly (perhaps some people might have a cooker in their bedroom), and in a way that is dependent on our environment (we would not have learned the same generalizations growing up in a rural African village), but without explicit negative evidence. It would be rather radical to claim that learning facts of this sort from experience is impossible, so that there must be innate domain-specific knowledge about architecture, furnishings, and appliances.

The deepest fault with Kimball's instance of the APS, therefore, is that access to sentences in the alleged lacuna does not appear to be indispensable for learning what actually has to be learned: the acquirendum to consider is probably not the rule schema (7), but rather a small collection of specific facts about what sort of complement verb inflection is encountered with what sort of matrix verb, and no case has been made that information of that sort is not accessible from the millions of tokens of auxiliary combinations that are encountered in a few years of linguistic experience.

### *4.3. Anaphoric* one

A more fully elaborated instance of the APS (also antedating the first appearance of the term 'poverty of the stimulus') is offered by Baker (1978: 413–425; see also 327–340). It concerns the question of how the child can come to know the correct anaphoric use of the English word *one*.[15] Baker's argument has been repeated approvingly in a number of subsequent publications (for example, Hornstein and Lightfoot 1981: 18 ff; Lightfoot 1982a: Chapter 4; Lightfoot 1989: 322 f; Crain 1991: 609 ff; Ramsey and Stich 1991: 295). However, it has received remarkably little empirical scrutiny.

Baker claims that *one* is used as an anaphor whose antecedent must be a constituent of the type he labels Nom (also labelled N′ or $\bar{\text{N}}$ in the subsequent linguistic literature; we will follow Baker in using 'Nom').[16] A Nom constituent will generally contain a noun, and will often contain nothing else (as in *This box is bigger than the other one*, where *one* means *box*). But Baker observes that *one* can have more than a noun as its antecedent. This is shown by such examples as (11), where *one* may have either of two Nom constituents as antecedent: *glass* (a Nom containing just a head noun) or *blue glass* (a Nom modified by an adjective to make a larger Nom).

(11) *John has a blue glass but Alice doesn't have one.*

Indeed, Baker claims that the antecedent *cannot* be just a noun (as opposed to a Nom containing a noun and nothing else). He claims this is shown by the ungrammaticality of such strings as (12), where *student of chemistry* is a Nom but *student* is a noun taken separately from its PP complement.

(12) **The student of chemistry was more thoroughly prepared than the one of physics.*

---

15. The matter is more complex than is revealed in Baker's discussion, because *one* has several different functions. In examples like *He made one small mistake*, *one* functions as determiner (compare **He made small mistake*, ungrammatical because there is no determiner). In *the one little mistake he made*, or *his one and only mistake*, or *one or other of them*, it seems to be an adjective. And in *the other ones* it is clearly a noun. Baker's discussion mixes cases where *one* is probably a determiner (*Alice has one*, where it might be reasonable to posit a null head noun for one of the readings) with cases where it is definitely not a determiner (*the English one*). Since he does not distinguish these for purposes of his argument, we will not attempt to do so either, though conceivably it could make a difference to go back over the facts with these distinctions in mind.
16. Since Baker is concerned only with matters of choice of antecedent when there is an antecedent, we can ignore the fact that introduction of *one* by an antecedent-sensitive transformational rule (which is what he assumes) is certainly not a workable analysis, and we can also ignore the fact that (as pointed out by Hankamer and Sag 1976) *one* can be used deictically, with no linguistic antecedent: *The green one is mine*.

Two different acquirenda are thus implicit in Baker's discussion: (i) that a single noun (like *student* in (12) taken without its PP complement *of chemistry*) *cannot* be an antecedent for anaphoric *one*, and (ii) that a multi-word phrase of Nom type *can* be.

After much reflection concerning (i), we are unable to frame the purported acquirendum in a general way that makes it true. Is the definite article in (12) supposed to be crucial, for example? Examples like the following do not seem at all ungrammatical to us:

(13) a. *I'd rather teach linguistics to a student of mathematics than to one of any discipline in the humanities.*
b. *An advocate of Linux got into a heated discussion with one of Windows NT and the rest of the evening was nerd talk.*

Even if we do focus on instances of *the one* P NP, it is clearly not true that *one* can never have a complemented noun on its own as antecedent. The following examples are taken from *WSJ* (Linguistic Data Consortium 1993; example identifiers are file names), and the final PP in each case is a complement:

(14) a. *an agreement to replace the one* with Olivetti [`/wsj/w7_002`]
b. *… certain exemptions from the gas tax, including the one* for gasohol [`/wsj/w7_034`]
c. *(The bid …) is higher than the one* for Delta II. [`/wsj/w7_048`]
d. *the raid on Dieppe, even more disastrous than the one* on Zeebrugge. [`/wsj/w7_050`]
e. *I like the one* of Monty … [`/wsj/w7_108`]

These final PPs are complements to nouns (note the subcategorization of the related verbs: *agree with* NP, *exempt* NP *from* NP, *bid* NP *for* NP, *raid* NP, *picture* NP). With depictive complements as in (14e), *the one of* NP is quite common. Is it really the case that with some verb-related nouns, specifically *student*, the construction is ungrammatical? This partly depends on the status of examples such as (15):

(15) *Who is more culpable: the manufacturer of cigarettes who sells them knowing they cause cancer or the one of handguns who sells them knowing they are mainly used to kill?*

Indeed, is it even the case that the construction is impossible with *student*? More appropriately contextualized examples than Baker's (12) seem to have little wrong with them:

(16) *He told endless horror stories about the current state of higher education, from the student of literature who had never heard of Chaucer*

> *to the one of chemistry who could not name five elements from the periodic table.*

We are simply not convinced that such examples are syntactically ill formed, and thus regarding Baker's first suggested acquirendum we are not persuaded that there is anything to acquire.

But it is Baker's second acquirendum that is relevant to our argument, because, like the others we have considered, it involves positive but inaccessible evidence: Baker considers the fact that a multi-word phrase of Nom type as its antecedent *can* be the antecedent for anaphoric *one*, and asks: "Under what sorts of conditions could a child conceivably obtain the information that *one* can stand for sequences consisting of more than just a noun alone?" (p. 416).

The answer, he suggests, is that the child would have to be exposed to an "unusual set of circumstances" such as hearing (11), *John has a blue glass but Alice doesn't have one*, in a context where it is quite clear that Alice does have a glass, though it is not a blue one. He takes it to be extremely improbable that children have such experiences often enough to ensure that they would all learn from the examples they had heard that *one* could have a phrasal antecedent. The specification for this case of the APS is thus as follows.

(17) *APS specification for anaphora with* one

a. ACQUIRENDUM: the fact that anaphoric *one* can have Nom antecedents that are larger than just a single noun.
b. LACUNA: the set of all utterances containing anaphoric *one* that reveal in context that the antecedent of *one* cannot be just a noun (i.e., examples like (11) in a context where Alice does have a glass).
c. INDISPENSABILITY ARGUMENT: in order to learn that an instance of anaphoric *one* can can have a multi-word antecedent, the learner must experience cases in which an instance of *one* has a multi-word antecedent.
d. INACCESSIBILITY EVIDENCE: it is only in an "unusual set of circumstances" (Baker asserts) that the learner will encounter *one* in a context where the antecedent is clearly more than just a noun.
e. ACQUISITION EVIDENCE: none cited (the issue is nontrivial).

Concerning the unsupplied acquisition evidence, note that Baker has set an epistemological puzzle for himself. Suppose it were true that contexts of the specified sort (where a noun-only antecedent for *one* cannot be correct) are rare to the point of nonexistence (or at least inaccessibility). It would then be extraordinarily difficult to tell that someone had learned the wrong generalization: situations that would force speakers to reveal that they *did* interpret *one* as always having single-word antecedents would be just as vanishingly rare as the

ones allegedly needed for data-driven learning. That is, there could be speakers, interacting with us undetected, who have actually learned the wrong generalization. To have fully convincing acquisition evidence, Baker would need some kind of evidence that there are no such people.

But 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. Baker apparently does not see this, yet it makes a real difference. If all that is needed is for it to be implied in the context that the antecedent is more than a mere noun, there is more crucial evidence for the hypothesis of multi-word antecedents. Looking through examples containing *one* in the first corpus we examined immediately revealed cases of the relevant sort. In the first two 1987 files of *WSJ*, we find examples like these:

(18) a. *"You can teach what happens in a firing episode," he says. "But each one is so different that the advice may be misleading."* (WSJ7_001:12759-12760)

b. *On the other hand, Mr. Fletcher notes the "amazing success" of Werner Forssman, who pioneered the use of cardiac catheters in surgery by inserting one into his own body, a chilling feat that won him a 1956 Nobel Prize in medicine.* (WSJ7_002:10454)

"Each one" in (18a) obviously means 'each firing episode,' since 'each episode' is too general to make any sense. And in (18b), "inserting one" has to be inserting a cardiac catheter (inserting just any kind of catheter into any part of his body would not have won Dr. Forssman a Nobel).

Nothing here depends on the material searched being prose by journalists. The issue is about how rare the contexts Baker alludes to can be, if finding them in texts is this easy.

And it is easy. Similar examples can be found in almost any kind of text. The second corpus that came to hand was Oscar Wilde's comedy *The Importance of Being Earnest* (about 100 KB in length), and we rapidly found Lady Bracknell saying this of her late brother-in-law:

(19) *I cannot at the present moment recall what the General's Christian name was. But I have no doubt he had one.*

It is certainly a Christian (first) name she means, not a name, since she knows his last name is Moncrieff.

The point of this example is that it was so easy to find in the first text that came to hand; the source is of no significance. Whether children watch Oscar Wilde plays during the learning period (sadly, they generally do not) is irrelevant. If we turn to a corpus of speech from TV programs often watched by

children, we find the same thing. The transcribed TV dialogue available on the CHILDES database CD ROM (see MacWhinney 1995) is rich in brief, casual, and elliptical sentences, with low syntactic complexity (mean sentence length is less than 7 words). Yet even in the tiny amount of material we examined, we immediately noted this exchange, from the daytime soap "One Day At A Time":

(20) A: *"Do you think you will ever remarry again? I don't."*
B: *"Maybe I will, someday. But he'd have to be somebody very special. Sensitive and supportive, giving. Hey, wait a minute, where do they make guys like this?"*
A: *"I don't know. I've never seen one up close."*

Speaker A has surely seen a guy up close at some time, but not a sensitive, supportive, generous one; so the antecedent must be more than just the noun *guy*.

Such easily-found examples tell us that Baker's empirical inaccessibility claim, so often quoted but never tested, is almost certainly false. Like Kimball, Baker suggests no numerical threshold, so knowing what percentage of occurrences of *one* were of the requisite sort would not reveal anything helpful. But the examples we have given were found immediately in the first few texts we scanned for the word *one*. The ease with which they were found suggests that the context often suggests pragmatically (though may not entail semantically) that an instance of anaphoric *one* has a multi-word antecedent phrase. This is sufficient to cast grave doubt on Baker's implicit claim that learning to use *one* involves syntactic learning in the absence of evidence.

### *4.4. Auxiliary-initial clauses*

The apparently strongest case of alleged learning from crucially inadequate evidence discussed in the literature, and certainly the most celebrated, concerns auxiliary-initial positioning in polar interrogatives in languages like English and Spanish: sentences like *Are you happy?*, the polar interrogative (yes/no question) corresponding to the declarative *You are happy*. The generalization concerning the formation of such sentence types is *structure-dependent*: it is based on structural relations (dominance among constituents), not just temporal sequence (precedence among words). Specifically, it is the matrix-clause auxiliary verb that is assigned initial position, not, e.g., whatever is the leftmost auxiliary in the corresponding declarative clause. Chomsky (1971b: 29–33) gives these examples:

(21) a. *The dog in the corner is hungry.*

b. *Is the dog in the corner hungry?*
c. *The dog that is in the corner is hungry.*
d. *Is the dog that is in the corner hungry?*
e. **Is the dog that in the corner is hungry?*

In an example as simple as (21a) the auxiliary that is positioned initially in the interrogative (21b) is the only auxiliary in the declarative (21a), so the two hypotheses ("the first auxiliary in the declarative is fronted" and "the main clause auxiliary of the declarative is fronted") make the same prediction. But for (21c) the predictions diverge: one hypothesis predicts (21d) as the question, while the other incorrectly predicts (21e).

(22) *APS specification for auxiliary-initial clauses*

a. ACQUIRENDUM: the structure-dependent generalization about auxiliary-initial clauses, i.e., that auxiliary-initial clauses differ from subject-initial clauses in the position of the main clause auxiliary.
b. LACUNA: the set of all sentences in which the initial auxiliary in an auxiliary-initial clause is not the first auxiliary in the corresponding declarative clause.
c. INDISPENSABILITY ARGUMENT: evidence that distinguishes the structure-dependent and structure-independent hypotheses, i.e. evidence of auxiliary-initial sentences in which the clause-initial auxiliary is not the first auxiliary in the related declarative, is needed to learn the acquirendum, because on other types of sentence the two have the same predictions.
d. INACCESSIBILITY EVIDENCE: evidence has not been offered, but Chomsky asserts that utterances in which a clause has an initial auxiliary that would not be the first auxiliary in the corresponding declarative are so rare that "A person might go through much or all of his life without ever having been exposed" to them (see Piattelli-Palmarini 1980: 40), and "you can go over a vast amount of data of experience without ever finding such a case" (see Piattelli-Palmarini 1980: 114–115).
e. ACQUISITION EVIDENCE: not attempted by Chomsky, but Crain and Nakayama (1987) provide experimental confirmation that children do learn the acquirendum.[17]

17. We will assume here that Crain and Nakayama's subjects were typical, and the acquirendum here is indeed one that first-language learners always learn correctly. But it should be noted that one study looked at the issue in connection with second language acquisition, with not entirely unequivocal results. White (1989: 63–66) summarizes work by Y. Otsu and K. Naoi

Two points should be noted in passing. First, Chomsky believes that the real issue is a much more general item of knowledge than the English-specific acquirendum in (22), namely that all generalizations in the grammars of natural languages are based not on string manipulation and word positions in surface strings but on constituent structure. This may indeed be true. But it is not reasonable to take that as the acquirendum, first because it is not precisely enough characterized, and second because there is no reason any child ever has to know it: to assume that a learner of English has to know that in all natural languages constituent-order relationships between sentence types are structure-dependent is to beg the question in favor of innate knowledge of universal grammatical principles.

Second, it should be noted that sentences like (21d) actually only support the claim that sentences formed under the main-clause-auxiliary hypothesis are grammatical. They do not show that sentences formed under the first-auxiliary hypothesis are *un*grammatical. Sentences like (21d) are compatible with the idea that the correct principle allows *any* auxiliary to be fronted. We ignore this interesting point (which is due to Freidin 1991), because our concern here is to assess whether the inaccessibility claim is true.

One reason for seeing this case as offering a particularly clear and interesting instance of the APS is its purely syntactic character. The cases discussed in the last two sections are bound up with meaning in a way that makes their direct relevance to the APS much less clear. But auxiliary-initial positioning is a purely syntactic matter, a formal device used to mark (among other things) interrogative sentence type. Under a data-driven account, learning how these sentence types are constructed must apparently be done entirely from evidence about permissible word positions. Identifying the generalization about which auxiliary must be fronted in yes/no questions corresponding to a certain statement is not dependent on learning some concept or meaning from the context. [18]

---

in which an attempt was made to assess by forced question construction whether the structure-dependent generalization would be employed by Japanese subjects of 14 to 15 years old who had learned simple polar interrogatives in English but had not yet learned relative clauses, and whose native language did not use a structure-dependent constituent-order variation to signal interrogative sentence type (Japanese merely adds a question particle to the declarative). In these experiments, one of the eleven subjects did produce errors of the 'impossible' structure-independent type (and another three remained neutral by finding ways to avoid constructing strings of the relevant sort). This is unexpected if innate priming defines the possibilities for second language learners as well as first language learners, but perhaps the defenders of innate priming would say it does not.

18. Auxiliary-initial syntax is actually not quite innocent of semantic entanglements, as noted by Joseph Emonds and discussed in Chomsky (1971a: 209–210). But, if special meanings are in some cases associated with auxiliary-initial sentences so that they have to be learned independently of the strings from which they are putatively derived, it becomes *harder* to show that a syntactic generalization about structure has been acquired without evidence, not

That makes lacuna characterization and arguing for their indispensability a lot easier.

This instance of the APS has an extraordinarily high citation index. Chomsky has written about it eight or nine times. He first refers to the structure-dependence of the auxiliary-fronting generalization in Chomsky (1965: 55–56), and takes it up again in Chomsky (1968: 51–52); adding the remark that the "language-learner knows" that only structure-dependent operations are available. The idea that it shows something about the innate basis that must exist for language acquisition is expounded more fully in Chomsky (1971b: 29–33) and developed further in Chomsky (1972) and Chomsky (1975: 30–33, 153–154). Further amplification then appears in Chomsky's contributions to the debate reported in Piattelli-Palmarini (1980: 114–115). It is referred to once again (and claimed to be one of a "great many examples") in Chomsky (1986: 7–8), and it is adapted to Spanish in Chomsky (1988a: 41–47). And Chomsky's claims have been subsequently reiterated and endorsed by many other writers, in both technical articles and introductory texts, by both linguists (e.g. Lightfoot 1991: 3–4; Uriagereka 1998: 9–10; Carstairs-McCarthy 1999: 4–5; Smith 1999: 53–54; Lasnik 2000: 6–9) and psychologists (e.g. Crain 1991: 602; Marcus 1993: 80; Pinker 1994: 40–42, 233–234). It is not an exaggeration to say that this instance is being passed around and repeated over and over again; APS enthusiasts never tire of citing it.

The lacuna characterization is quite clear: the sentences in the lacuna are those auxiliary-initial sentences in which it is the main clause auxiliary that is fronted and there is some other auxiliary, which would be the first auxiliary in the corresponding declarative sentence, which is not fronted.

The indispensability claim is also fairly clear. At least, we are not going to challenge it. It is quite reasonable to assume that the learner must have evidence about sentences in the lacuna to learn the acquirendum.

Our focus will be on inaccessibility. The claim made is that the child is deprived of crucial positive evidence that supports the main-clause-auxiliary generalization over the incorrect first-auxiliary generalization. Chomsky asserts this in quite extreme terms. He claims more than just low frequency for the crucial kinds of example that would distinguish structure-dependent from structure-independent formulations of auxiliary fronting. In the paper and discussions published in Piattelli-Palmarini (1980), he makes much stronger claims:

(23) a. "A person might go through much or all of his life without ever having been exposed to relevant evidence, but he will neverthe-

---

easier. We set aside this sort of evidence here in order to give the proponents of the APS the benefit of the doubt.

less unerringly employ [the structure-dependent generalization], on the first relevant occasion" (Chomsky, in Piattelli-Palmarini 1980: 40)

b. "the more complex cases that distinguish the hypotheses rarely arise; you can easily live your whole life without ever producing a relevant example to show that you are using one hypothesis rather than the other one." (Chomsky, in Piattelli-Palmarini 1980: 114–5)

c. "The examples cited are the only kind for which the hypotheses differ, and you can go over a vast amount of data of experience without ever finding such a case. Thus in many cases the Martian scientist could not know by passive observation whether the subject is using the first hypothesis or the second one." (ibid.)

Note here a consistency problem for Chomsky, of a very similar sort to the epistemological problem Baker sets for himself (see previous section). If (23b) is true, then it undercuts the force of (23a) and (23c). If people so rarely produce utterances that exhibit their grasp of the structure-dependent character of the auxiliary fronting generalization, then there could well be speakers around who have acquired an "incorrect" structure-independent generalization instead, but who are never detected because of the rarity of the crucial situations in which they would give themselves away.

But let us set aside this problem and return to the empirical issue. Freidin (1991: 618) notes that the APS instance based on auxiliary-initial clauses "is based on an empirical assumption that children encounter (or pay attention to) simple sentences prior to those with subjects containing finite relative clauses." He adds: "Whether this is actually the case has not been established." This gets to the heart of the matter. Sampson (1989), of which Freidin was perhaps unaware, explicitly suggests that it is probably not true that children are deprived of access to questions with auxiliary-containing subjects (like (21d)) during the acquisition process. He takes William Blake's well-known poem 'Tiger' to be typical of language encountered by almost every English child during their early years, and notes that the line *Did He who made the lamb make thee?* offers crucial positive evidence for the structure-dependent rule (under the first-auxiliary account it would have been **Did He who make the lamb made thee*).[19] He finds another relevant example by scanning a list of questions in an encyclopedia for young children. His point is that with no effort at all he could immediately identify a couple of sentences that numerous children

19. This might seem an implausibly literary example. But of course Sampson is thinking of his own middle-class upbringing in England over five decades ago, and how *he* learned English. Your linguistic experience may differ. But does it differ enough?

would have heard and that had the relevant characteristic. How rare they could be, he is asking, if it is that easy to come up with them?

These observations surely merit some sort of response from defenders of the APS. But over a decade passed after the publication of Sampson's article without any response being offered.

A moment of reflection should suffice to raise some suspicions that Sampson must be right. Surely it is implausible that one could expect to live one's whole life as an English speaker, or even reach kindergarten, without running into *any* sentences of the sort illustrated in (24). (An underscore marks the position in each string where the main clause auxiliary would be if it were not fronted.)

(24) a. *Will those who are coming ___ raise their hands?*
b. *Can the people who are leaving early ___ please sit near the door?*
c. *Is the man who was showing you the pictures ___ still here?*
d. *Would anyone who is interested ___ see me later?*
e. *Can a helicopter that has lost its tail rotor ___ still fly?*
f. *Will the owner of the bicycle that is chained to the gate ___ please move it?*
g. *Could the girl who has lost her ticket ___ come to the desk?*
h. *Could a tyrannosaurus that was sick ___ kill a triceratops?*

These examples have an auxiliary verb within the subject NP. But strictly it is not necessary for the auxiliary to be *in* the subject NP in order for there to be a contrast between fronting the main clause auxiliary and fronting the first auxiliary. All that is needed, as Sampson recognizes, is for any auxiliary to precede the main clause auxiliary. And that condition would be met in examples like the ones in (25) as well:

(25) a. *If you don't need this, can I have it?*
b. *Since we're here, can we get some coffee?*
c. *When you're done, could I borrow your pencil?*
d. *Given that I'm not needed, can I go home?*
e. *While you're getting cigarettes, could you get some more milk?*
f. *Though you won't like me asking, did you brush your teeth?*

For example, to make the polar interrogative counterpart to the string (26):

(26) $if_1$ $you_2$ $don't_3$ $need_4$ $this_5$ $I_6$ $can_7$ $have_8$ $it_9$

the seventh word has to be repositioned to precede the sixth; repositioning the third to precede the second, or any word to precede the first, is incorrect. That is trivial if one understands where the clause boundaries are, of course (the initial auxiliary is supposed to be initial in the main clause, not in a fronted

subordinate clause); but structure-independent operations are blind to clause boundaries. Thus the sentences in (25) also confirm the structure-dependent generalization over any structure-independent one.

The range of relevant examples widens further once we notice that non-subject *wh*-movement questions always incorporate an auxiliary-fronting construction. (We find, for example, strings of the form *WX* where *W* is a *wh*-word and *X* is a string of the sort that instantiates auxiliary fronting.) Thus utterance tokens like (27a) or (27c) would be crucial evidence in favor of the structure-dependent auxiliary fronting hypothesis (given, of course, the complete absence of utterances like (27b) or (27d)):

(27) a. *How could anyone who was awake ___ not hear that?*
b. **How was anyone who ___ awake could not hear that?*
c. *When will the man who is in charge ___ be back?*
d. **When is the man who ___ in charge will be back?*

Of the roughly 23,000 questions in *WSJ*, one must look through only 15 before hitting an example in the (alleged) lacuna. Despite the admittedly overwhelming predominance of simple structures that provide no evidence either way, the 15th question mark in *WSJ* is in example (28a), which has the crucial structure. In (28b) we show the non-occurring string that would occur if English auxiliary-initial clauses were formed by fronting the supportive DO that would be the auxiliary of the relative clause if it had one.

(28) a. *How fundamental are the changes these events portend?* (`W7_001:3963`)
b. **How fundamental do the changes these events ___ portend are* t*?*

Several other such examples occur within the first five hundred interrogatives in *WSJ*.

If one decided, for some reason, to exclude *wh*-interrogatives and look for yes/no questions, as in Chomsky's examples, little would change. Within the first 500 questions in the 1987 files of the *WSJ* database we find such cases, for example (29a) – rather than (29b).

(29) a. *Is a young professional who lives in a bachelor condo as much a part of the middle class as a family in the suburbs?* (`W7_006:2813`)
b. **Does a young professional who ___ live in a bachelor condo is as much a part of the middle class as a family in the suburbs?*

If we narrowed the criteria yet further and insisted on examples with two overt auxiliaries in the corresponding declarative, rather than examples involving supportive *do* like the examples seen so far, it would not make a major

difference. The 180th interrogative in *WSJ* is (30a), with two instances of the copula, one inside the (headless relative) subject NP, exactly like Chomsky's hypothetical examples; (30b) shows what we would get by fronting the first auxiliary in the related declarative.

(30) a. *Is what I'm doing in the shareholders' best interest?* (W7_003:2991)
b. **Am what I doing is in the shareholders' best interest?*

*WSJ* happens not to contain any examples with two occurrences of the word form *is*, one fronted across the other, as in Chomsky's invented examples in (21). However, even those can readily be found. Scanning the script of *The Importance of Being Earnest* for 'is … is …?', we found an example immediately: (31a), contrasting with (31b).

(31) a. *Who is that young person whose hand my nephew Algernon is now holding in what seems to me a peculiarly unnecessary manner ___?*
b. **Who is that young person whose hand my nephew Algernon ___ now holding in what seems to me a peculiarly unnecessary manner is?*

The speaker is Lady Bracknell again – not an ideal exemplar of modern everyday colloquial speech, granted. But the relevant question is not whether her style is ponderous. Her question could have been phrased as *Who's the woman whose hand Algernon's holding?*, or *Who's the woman who's holding Algernon's hand?*; these are still crucial evidence. The question is whether we can expect situations calling for questions with this sort of content to arise frequently in real life. We only searched *Earnest* because we happened to have an electronic copy of it, and it is only 100 KB, yet it happened to contain an 'is … is…?' just like the ones Chomsky says one may not encounter in a lifetime. Perhaps it is a highly unusual text. But that is an empirical claim that must be established by empirical work, not just reasserted on the strength of previous unsupported assertions.

It is highly unlikely that the claim of unusualness is true. Situations calling for examples of the crucial type arise with enough frequency that we were able to find some even in the small corpus in the CHILDES database of transcripts from TV programs aimed at or popular with children, mentioned earlier. We found two cases of the relevant sort in a *Mork and Mindy* script. One was (32a). Notice that the declarative form under Chomsky's assumptions (with *where* replaced by *there*) is something like (32b). What we would get from fronting the auxiliary from the leftmost slot would be (32c).

(32) a. *Where's the application Mork promised to fill out?*

b. *the application Mork* [$_{\text{Aux}}$ PAST] *promise to fill out* [$_{\text{Aux}}$*is*] *there*
c. **Where did the application Mork promise to fill out is?*

Another such example from the same source is (33a). Again, we give the corresponding declarative and what would be the result of fronting the wrong auxiliary.

(33) a. *Where is that girl that I had to teach you tennis?*
b. *that girl that I* [$_{\text{Aux}}$ PAST *have to teach you tennis* [$_{\text{Aux}}$*is*] *there*
c. **Where did that girl I have to teach you tennis is?*

If these were not enough of an indication that children do hear utterances of the sort in question, another source we can search is the corpus of utterances addressed to a child named Nina by her caregivers when she was between 1;11 and 3;3, collected by Patrick Suppes in 1973 (see MacWhinney 1995: 322), also available on the CHILDES database CD ROM. It did not take long to find three examples in the file NINA05.CHA in the DATABASE/ENG/SUPPES/ directory, which contains material from a session when Nina was just over two years old:

(34) a. *Where's the little blue crib that was in the house before?*
b. *Where's the other dolly that was in here?*
c. *Where's the other doll that goes in there?*

Thus, if evidence were needed that two-year-olds were receiving relevant input, it is available. (Notice the importance of sentences of the form *Where is the* NP*?* where NP contains a relative clause, as in both (33) and (34). These seem quite common, and offer an excellent chance to see that complex NP's invert with auxiliaries just as simple NP's do.)

Taken together, this evidence seems sufficient to show that Sampson's anecdotally illustrated suspicions are justified: Chomsky's assertion that "you can go over a vast amount of data of experience without ever finding such a case" is unfounded hyperbole. We have found relevant cases in every corpus we have looked in.

Have we found enough of them? Some relevant evidence may be found in the remarkable work of Hart and Risley (1995). These authors documented the vocabulary development of 42 children from age 1 to age 3, taking note not only of their production but also of their verbal environments. Their results give us some quantitative guidelines. For example, extrapolating from their measured averages, they suggest that by the age of 3, a child being raised in a professional household might have heard 30 million word tokens.

One rather startling result of Hart and Risley's concerns the magnitude of the difference in language experience between classes: it is calculated that a child in a working-class family will have heard 20 million word tokens by the

age of 3, and a child being raised in a family on welfare will have heard only 10 million (p. 132). Nonetheless, even in a welfare household, the cumulative exposure to language use of a 3-year-old amounts to ten times the entire extant corpus of Old English literature, or ten times the size of the Brown corpus on which the celebrated Kučera and Francis (1967) study was based.

Hart and Risley offer statistics that are relevant to evaluating the significance of the results of this section. Questions turn out to be dramatically more frequent in everyday speech to children than any other speech act type (this is one way in which printed texts like newspapers, naturally, differ from everyday interaction, as noted above; this makes support from sources like *WSJ* the more remarkable: speech is much richer in interrogatives). Hart and Risley report (p. 129) that about 30 percent of parent utterances while interacting with a child were questions. Suppose, therefore, that we consider a welfare child who has been exposed to only 10 million words of language use. Hart and Risley find the mean length of utterances to be 4.0 words, which would mean about 2.5 million utterances, therefore about 750,000 questions.

We do not know how many of those questions should be assumed to be of the crucial form, but note that expressions as short as 4 words can actually provide crucial evidence: consider *Has whoever left returned?* which, if the first auxiliary were fronted, would have been **Did whoever left has returned?*. Even the mean utterance length, therefore, is not so low as to exclude the crucial type of interrogative.

Our preliminary investigations suggest the percentage of relevant cases is not lower than 1 percent of the interrogatives in a typical corpus. By these numbers, even a welfare child would be likely to hear about 7,500 questions that crucially falsify the structure-independent auxiliary-fronting generalization, before reaching the age of 3. But assume we are wrong by a whole order of magnitude on this, so there are really only 750. Would exposure to 750 crucial examples falsifying the structure-independent generalization not be enough to support data-driven learning? Would a thousand be enough? Our point is that the advocates of the APS – defenders of the claim that the child could not possibly be exposed to enough data to learn the generalization in a data-driven manner – must develop some explicit and quantitative answers to such questions if they wish to support their claim.

In the absence of any answers, or any research that is even directed at trying to develop answers, we conclude that the APS instance that Chomsky has used to support the APS for over twenty-five years has no empirical support and offers no support for the existence of innately primed learning.

## 5. Implications and conclusions

The extensive literature on linguistics, learnability, and language acquisition contains many further interesting arguments. This article is far from exhaustive. But on the basis of the four cases reviewed here, which include what are unquestionably the most celebrated and widely cited instances of the APS, we believe it must be concluded that the defenders of the APS are still searching for a well-confirmed case of the type they need.

This does not show that some particular data-driven theory of language acquisition is correct. But casting doubt on alleged APS cases, in the way that we have done, undercuts skepticism about data-driven learning.

Before concluding, we want to point out that searching for a sound instance of the APS is a task that will, if undertaken with seriousness, bring generative linguists into contact with two lines of work toward which they have traditionally been rather antipathetic: mathematical learning theory and corpus linguistics. Mathematical learning theory studies the a priori limits on data-driven learning, and without clear results on such limits, indispensability arguments cannot be clearly established. (We have not been very stringent about indispensability arguments in this article.)

And corpus linguistics is relevant because inaccessibility claims have entailments about the contents of corpora: to claim that children could not possibly have access to certain evidence is to say that in typical corpora (from the right kind of language use, whatever that is) crucial evidence will not be found. Generative linguists interested in defending the APS must embrace both of these lines of work if they are going to accomplish the task they have set themselves.

There is an interesting reason why work aimed at supporting the APS is almost bound to be self-undercutting if it is pursued with any seriousness. Such research will have to involve close study of what data-driven learning procedures can do. And such study is almost bound to yield at least some improvements in success rates of such procedures. (This follows simply because if no serious experiments are done on data-driven learning, there will be a zero success rate; doing at least some experimentation will almost certainly yield better than zero success.) But each such success is likely to eliminate or weaken some particular apparent APS instance, by revealing that more can be learned in a data-driven way than was previously thought.

To some extent this has already begun to happen. The literature on "unsupervised learning" in computational linguistics (Brent 1993 and Schütze 1995 are two examples) is demonstrating that algorithms can be learned much more from text than most linguists would have thought possible. A substantive defense of the APS will have to address this sort of research.

We hazard no guess about whether domain-specific innate mechanisms specialized for first language acquisition will turn out to be nonexistent, or sur-

prisingly few, or surprisingly rich and detailed. Although this article will no doubt be misread as a defense of 'empiricism', that is *not* an accurate characterization of it. We are agnostic at this point, perfectly prepared to accept that perhaps at some time in the future data-driven algorithmic techniques for language learning will hit a serious roadblock, or that credible evidence for innate task-specific language acquisition mechanisms will turn up. We are also willing to accept it as a possibility that once linguists and philosophers get straight on what is required to establish an instance of the APS, the distinction between data-driven and innately-primed learning will be less absolute. However, this much is clear: until data-driven learning is investigated in more detail, linguists will remain ill-equipped to do more than fantasize and speculate on the matter.

And as for the philosophers and cognitive scientists who have accepted linguists' claims about the APS, and treated it as an established contribution of linguistic research to the cognitive sciences, they have some reassessment to do. For example, Fodor (1981: 258) had better be wrong in describing the APS as "*the* existence proof for the possibility of cognitive science", because the APS still awaits even a single good supporting example. Cognitive science needs a better proof of its possibility than this. The advocates of the APS must shoulder the burden of participating in empirical work to support instances of the argument, and thus to begin to vindicate Fodor's belief that cognitive science is possible. It is a burden that as yet linguists have neither shouldered nor even recognized.

*Stevenson College, UCSC*

## References

Alegre, Maria A., and Peter Gordon (1996). Red rats eater exposes recursion in children's word formation. *Cognition* 60: 65–82.

Baker, C. L (1978). *Introduction to Generative-Transformational Syntax*. Englewood Cliffs, NJ: Prentice-Hall.

Baker, C. L. (ed.) (1981). *The Logical Problem of Language Acquisition*. Cambridge: MIT Press.

Bates, Elizabeth, and Jeffrey Elman (1996). Learning rediscovered. *Science* 274 (13 December): 1849–1850.

Berman, Ruth (1990). On acquiring an (S)VO language: subjectless sentences in children's Hebrew. *Linguistics* 28–6 [310]: 1135–1166.

Biber, Douglas (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Bod, Rens (1998). *Beyond Grammar: An Experience-Based Theory of Language*. Stanford, CA: CSLI Publications.

Brent, Michael (1993). From grammar to lexicon: unsupervised learning of lexical syntax. *Computational Linguistics* 19: 243–262.

Carstairs-McCarthy, Andrew (1999). *The Origins of Complex Language: An Inquiry into the Evolutionary Origins of Sentences, Syllables, and Truth*. Oxford: Oxford University Press.

Chomsky, Carol S. (1969). *The Acquisition of Syntax in Children from 5 to 10*. Cambridge: MIT Press.

Chomsky, Noam (1957). *Syntactic Structures*. The Hague: Mouton.

— (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

— (1968). *Language and Mind*. New York: Harcourt, Brace and World. [Enlarged edition 1972. Page references are to first edition.]

— (1971a). Deep structure, surface structure, and semantic interpretation. In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*, Danny Steinberg (ed.), 183–216. Cambridge, England: Cambridge University Press.

— (1971b). *Problems of Knowledge and Freedom*. London: Fontana.

— (1972). Universal principles of language structure. [Interview.] *Indian and Foreign Review* 10, no. 5 (15 December): 20–22. Reprinted in Chomsky 1988b: 151–157.

— (1975). *Reflections on Language*. New York: Pantheon.

— (1980a). Rules and representations. *Behavioral and Brain Sciences* 3: 1–61.

— (1980b). *Rules and Representations*. Oxford: Basil Blackwell.

— (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

— (1988a). *Language and Problems of Knowledge*. Cambridge, MA: MIT Press.

— (1988b). *Language and Politics*, ed. by Carlos P. Otero. Montreal: Black Rose Books.

Cowie, Fiona (1999). *What's Within? Nativism Reconsidered*. New York: Oxford University Press.

Crain, Stephen (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 14: 597–612. [Commentaries, response, and bibliography 612–650.]

Crain, Stephen, and Mineharu Nakayama (1987). Structure dependence in grammar formation. *Language* 63: 522–543.

Elman, Jeffrey L., Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Falk, Yehuda (1984). The English auxiliary system. *Language* 60: 483–509.

Fodor, Jerry A (1981). *Representations*. Cambridge: MIT Press.

Freidin, Robert (1991). Linguistic theory and language acquisition: a note on structure-dependence. *Behavioral and Brain Sciences* 14: 618–619.

Garfield, Jay L. (1994). Innateness. In Guttenplan (ed.), 366–374.

Gazdar, Gerald (1981). Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12: 155-184.

Gazdar, Gerald, Geoffrey K. Pullum, and Ivan A. Sag (1982). Auxiliaries and related phenomena in a restrictive theory of grammar. *Language* 58: 591–638.

Gazzaniga, Michael S. (1992). *Nature's Mind: The Biological Roots of Thinking, Emotions, Sexuality, Language, and Intelligence*. New York: Basic Books.

Gold, E. Mark (1967). Language identification in the limit. *Information and Control* 10: 447–474.

Gordon, Peter (1986). Level ordering in lexical development. *Cognition* 21: 73–93.

Guttenplan, Samuel (ed.) (1994). *A Companion to the Philosophy of Mind*. Cambridge, MA: Basil Blackwell.

Haegeman, Liliane (1994). *Introduction to Government and Binding Theory.* 2nd edition. Oxford, U.K. and Cambridge, MA: Basil Blackwell.

Hankamer, Jorge, and Ivan A. Sag (1976). Deep and surace anaphora. *Linguistic Inquiry* 7: 391–428.

Hart, Betty, and Todd R. Risley (1995). *Meaningful Differences in the Everyday Experiences of Young Children*. Baltimore: Paul H. Brookes.

Hornstein, Norbert, and David Lightfoot (1981). Introduction. In *Explanation in Linguistics: The Logical Problem of Language Acquisition*, 9–31. London: Longman.

Hudson, Richard A. (1994). About 37 % of word-tokens are nouns. *Language* 70: 331–339.

Jain, Sanjay, Daniel Osherson, James S. Royer, and Arun Sharma (1999). *Systems That Learn* (2nd ed.). Cambridge, MA: MIT Press.

Johansson, Stig (1980). *Plural Attributive Nouns in Present-day English*. Lund Studies in English, 59. Lund, Sweden: Gleerup.

Kimball, John P. (1973). *The Formal Theory of Grammar*. Englewood Cliffs, NJ: Prentice-Hall.

Kiparsky, Paul (1982). From cyclic phonology to lexical phonology. In *The Structure of Phonological Representations*, part 1, Harry van der Hulst and Norval Smith (eds.), 131–175. Dordrecht: Foris.

Kučera, Henry, and W. Nelson Francis (1967). *Computational Analysis of Present Day American English*. Providence, RI: Brown University Press.

Lasnik, Howard (1989). On certain substitutes for negative data. In Matthews and Demopoulos (eds.), 89–105.

— (2000). Syntactic Structures *Revisited: Contemporary Lectures on Classic Transformational Theory*. Cambridge: MIT Press.

Lightfoot, David (1982a). *The Language Lottery: Toward a Biology of Grammars*. Cambridge: MIT Press.

— (1982b). Review of Geoffrey Sampson, *Making Sense*. *Journal of Linguistics* 18: 426–431.

— (1989). The child's trigger experience: degree-0 learnability. *Behavioral and Brain Sciences* 12: 321–334 [commentaries, response, and bibliography 334–375].

— (1991). *How to Set Parameters*. Cambridge, MA: MIT Press.

— (1998). Promises, promises: general learning algorithms. *Mind and Language* 13: 582–587.

Linguistic Data Consortium (1993). CD-ROM LDC-93-T1. University of Pennsylvania.

MacWhinney, Brian (1995). *The CHILDES Project : Tools for Analyzing Talk*. 2nd ed. Hillsdale, N.J.: Lawrence Erlbaum.

Marcus, Gary F. (1993). Negative evidence in language acquisition. *Cognition* 46: 53–85.

Matthews, Robert J. (1984). The plausibility of rationalism. *Journal of Philosophy* 81: 492–515. Reprinted in Matthews and Demopoulos (eds.), 51–75.

Matthews, Robert J., and William Demopoulos (eds.) (1989). *Learnability and Linguistic Theory*. Dordrecht: Foris.

McGilvray, James A. (1999). *Chomsky: Language, Mind, and Politics*. Cambridge, MA: Polity Press.

Osherson, Daniel (1999). Learning systems. In Wilson and Keil (eds.), 461–463.

Piattelli-Palmarini, Massimo (ed.) (1980). *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky.* Cambridge: Harvard University Press.

Pinker, Steven (1994). *The Language Instinct*. New York: William Morrow.

Pullum, Geoffrey K. (1996). Learnability, hyperlearning, and the poverty of the stimulus. In *Proceedings of the 22nd Annual Meeting: General Session and Parasession on the Role of Learnability in Grammatical Theory*, Jan Johnson, Matthew L. Juge, and Jeri L. Moxley (eds.), 498–513. Berkeley, CA: Berkeley Linguistics Society.

Pullum, Geoffrey K., and Deirdre Wilson (1977). Autonomous syntax and the analysis of auxiliaries. *Language* 53: 741–788.

Pullum, Geoffrey K., and Arnold M. Zwicky (1997). Licensing of prosodic features by syntactic rules: the key to auxiliary reduction. Annual meeting of the Linguistic Society of America, Chicago, Illinois. Abstract available at http://ling.ucsc.edu/~pullum/locker/lsa1997abstr.html.

Ramsey, William, and Stephen P. Stich (1991). Connectionism and three levels of nativism. In Ramsey, Stich, and Rumelhart (eds.), 287–310.

Ramsey, William, Stephen P. Stich, and David E. Rumelhart (eds.) (1991). *Philosophy and Connectionist Theory*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Ross, John Robert (1969). Auxiliaries as main verbs. *Studies in Philosophical Linguistics* 1: 77–102. Evanston, IL: Great Expectations Press.

Sadler, Louisa, and Douglas J. Arnold (1994). Prenominal adjectives and the phrasal/lexical distinction. *Journal of Linguistics* 30: 187–226.

Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport (1996). Statistical learning by 8-month-old infants. *Science* 274 (13 December): 1926–1928.

Sampson, Geoffrey (1989). Language acquisition: growth or learning? *Philosophical Papers* 18: 203–240.

— (1999a). *Educating Eve*. London: Cassell.

— (1999b). Collapse of the language nativists. *The Independent*, April 9, 1999: 7.

Schütze, Hinrich (1995). Ambiguity in language learning: computational and cognitive models. Unpublished PhD. dissertation, Stanford University.

Seidenberg, Mark S. (1997). Language acquisition and use: learning and applying probabilistic constraints. *Science* 275 (14 March): 1599–1603.

Singleton, David (1989). *Language Acquisition: The Age Factor*. Philadelphia, PA: Multilingual Matters.

Smith, Neil (1999). *Chomsky: Ideas and Ideals*. Cambridge: Cambridge University Press.

Snow, Catherine, and Charles A. Ferguson (1977). *Talking to Children: Language Input and Acquisition*. Cambridge: Cambridge University Press.

Stich, Stephen P. (1979). Between Chomskyan rationalism and Popperian empiricism. *British Journal for Philosophy of Science* 30: 329–347.

Thompson, Sandra A. (1988). A discourse approach to the cross-linguistic category 'adjective'. In *Explaining Language Universals*, John Hawkins (ed.), 167–185. Oxford: Basil Blackwell.

Uriagereka, Juan (1998). *Rhyme and Reason: An Introduction to Minimalist Syntax*. Cambridge, MA: MIT Press.

Weddell, Cassandra, and Janice Copeland (1997). "What's on the box?": How do parents, teachers, and children respond to T.V.? Ms (23 pp), Queensland University of Technology, Brisbane, Qld., Australia. Presented at the 21st Triennial Conference, Australian Early Childhood Association, Melbourne, Australia. [Summary entitled 'A chance to look and learn' published in *The Age*, Melbourne, September 23, 1997.]

Wexler, Kenneth (1991). On the argument from the poverty of the stimulus. In *The Chomskyan Turn*, Asa Kasher (ed.), 252–270. Oxford: Basil Blackwell.

White, Lydia (1989). *Universal Grammar and Second Language Acquisition*. Amsterdam: John Benjamins.

Wilson, Robert A. and Frank Keil (eds.) (1999). *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press.

Wimsatt, William C. (1999). Generativity, entrenchment, evolution, and innateness: philosophy, evolutionary biology, and conceptual foundations of science. In *Where Biology Meets Psychology: Philosophical Essays*, Valerie Gray Hardcastle (ed.), 137–179. Cambridge, MA: MIT Press.