Learnability, Hyperlearning, and the Poverty of the Stimulus
Author(s): Geoffrey K. Pullum
Proceedings of the Twenty-Second Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on The Role of Learnability in Grammatical Theory (1996), pp. 498-513

Learnability, Hyperlearning, and the Poverty of the Stimulus [1]

Geoffrey K. Pullum
University of California, Santa Cruz

> "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." (Hornstein and Lightfoot 1981:9)

1. Stimulus poverty and hyperlearning. The quote above makes a striking claim: that some knowledge about language structure is acquired by children on a basis that does not involve evidence. Let me use the term *hyperlearning* to denote knowledge acquisition that meets this condition. That is, I will say that hyperlearning occurs whenever an agent acquires a piece of knowledge p during a time interval t without being exposed at any time during or before t to any evidence that could establish p by ordinary methods of learning from examples — the methods that are the object of study for formal learnability theory (Gold 1967, Osherson et al. 1986).[2]

Hornstein and Lightfoot's claim that hyperlearning occurs when children learn their first languages is important. It is the central empirical premise in what has come to be called the Argument from Poverty of the Stimulus (APS). The APS is widely held to be the most important contribution of linguistics to changing the agenda in neighbouring disciplines like philosophy and psychology. In this paper I question its soundness. I show that the most frequently cited piece of support for its empirical premise is untrustworthy, which leaves the premise itself sorely in need of reliable empirical support.

My first task, however, is to state the APS. Anyone who consults the literature will find that no explicit statement of it has been published so far. The argument is scarcely even sketched. Broad hints of it begin to appear in such works as Chomsky (1975), but not under the name APS. Chomsky (1980:31–4), after reviewing some generalities 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," simply states 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'," as if the logic were self-explanatory. Hornstein and Lightfoot (1981) merely echo this language without analyzing it.

Philosophers like Stich (1979, 1981) and Garfield (1994) come a little closer to stating an explicit argument, but seem to miss the essential point: the point Lightfoot (1982b:428) stresses about the importance of "poverty of stimulus problems, i.e. where there are no data available to the child which will suffice to establish some rule or principle." They also tend to confuse it with very

different points (such as the problem of induction from a finite sample, the underdetermination of theories by evidence, the presence of errors in the child's corpus, and so on). Though article after article is devoted to discussing it (see, e.g., Demopoulos 1989, Matthews 1989, Wexler 1991), the APS never clearly emerges. Wexler (1991), for example, introduces the APS as follows:

> 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. (p. 253)

Having equated no less than four distinct questions — *how* grammars develop, *why* grammars are the way they are, *whether* hyperlearning takes place, and *what* is innate — Wexler adds (with no reference): "As Chomsky pointed out, this is an application of the classic rationalist argument from the poverty of the stimulus." And that is all he says about the content of the APS.

To set out an explicit statement of the APS, I need to distinguish between two ways in which an infant might in principle learn a language. The first I will call *data-driven learning.* This is what that the APS aims to challenge. Its defining characteristic is that it relies on attention to evidence, specifically, the corpus of utterances to which the child is exposed when they happen to come up in everyday contexts and are uttered in the child's presence. The general conditions on correct reasoning that are germane to learning anything else from evidence samples are deployed in data-driven language acquisition, but crucially, the learner is assumed *not* to be in prior possession of any information about what languages are like.

Notice that *if* an infant could acquire a language L by means of data-driven learning after being exposed to nothing more than a corpus C of observed utterance tokens, then any rational being could in principle also learn L if merely given access to C. In other words, if this is the way first language acquisition is accomplished, *we will never see an instance of hyperlearning in the first language acquisition domain.* Everything the child learns will be learnable from the evidence.

The second way in which languages might logically be learned, the one that generativists have been concerned to defend and develop, involves the learner being primed *ab initio* with special information about language, or endowed with special internal mechanisms that make available such specific information. Call language acquisition of this sort *innately-primed learning.* By definition, innately-primed learning proceeds in a way that does not rely solely on a corpus of observed utterance tokens.

Innately-primed learning is compatible with the existence of hyperlearning, because the corpus and the innate priming taken together might suffice to ensure acquisition even if the corpus alone would not have contained enough information. And the existence of hyperlearning is rather important epistemologically for those who believe in innate priming; if hyperlearning did not occur, so that all the learning that occurred was compatible with data-driven learning, it is not clear what kind of evidence for innate priming there could be.[3]

Some take it to be trivial that language learning must be either data-driven or innately primed, because they construe the issue as being simply about whether there is any innate priming or not. Others (like Stich (1979)) claim that a range of distinct positions lying between data-driven and innately-primed learning can be made out. As Garfield (1994: 367) notes, 'empiricism' and 'rationalism' (the positions in traditional epistemology with which the two methods are often respectively associated) can be regarded as merely two regions in a continuum, so that there are indefinitely many alternative positions arrayed between the extremes. But for the sake of argument, I propose here simply to grant everything that advocates of the APS need: that data-driven learning and innately-primed learning are well-defined, distinguishable, mutually exclusive, and jointly exhaustive of the possibilities. That yields a disjunction that constitutes the initial premise of the APS, which I can now set out as in (1).

(1) *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.]
   b. If human infants acquire their first languages via data-driven learning, then hyperlearning will never be observed in this domain. [Characterization of data-driven learning methods.]
   c. Hyperlearning does in fact occur in the domain of first language acquisition by infants. [Empirical premise.]
   d. Human infants do not learn their first languages by means of data-driven learning. [From (b) and (c) by modus tollens.]
   e. Therefore human infants learn their first languages by means of innately-primed learning. [From (a) and (d) by disjunctive syllogism.]

This is an entirely valid argument. My concern will be with whether it is *sound* — specifically, whether premise (1c) is true. My strategy will be to examine the strongest and best-known piece of support for (1c) that advocates of the APS have put forward, to show that it will not bear the load assigned to it.

The claim I have in mind concerns auxiliary fronting in polar interrogatives (e.g., *Are you happy?*, the polar interrogative corresponding to declarative *You are happy*). The crucial point about it is the one stated in (2):

(2) The rule for initial auxiliary verb position in polar interrogatives (and a variety of other construction types) in such languages as English or Spanish is *structure-dependent*: it is based on structural relations (dominance among constituents), not just temporal sequence (precedence among words). Specifically, it is the main clause auxiliary verb that is assigned initial position, not, e.g., whatever is the leftmost auxiliary in the corresponding declarative clause.

One reason for regarding this case as the strongest potential support that has been offered for the APS is that — unlike most of the very few other cases that have been suggested in the literature — it depends purely on syntax. Other cases discussed in the literature are bound up with meaning in a way that makes their direct relevance to the APS much less clear. For example, Chomsky (1975: 140–153) considers how the child might learn that *Mary appealed to the men to like each other* is well-formed while **Mary appeared to the men to like each other* is not; but this is bound up with issues about the uses of reciprocals and the control patterns of verbs that are thoroughly semantic (see Pollard and Sag 1991). Once the learner has grasped the concepts of appealing, appearing, and reciprocal liking, it is not clear that any syntactic triggering will be needed to learn that the latter is linguistically deviant.

Similarly, in a number of works[4] we find discussion of how the child can determine what are permissible antecedents for anaphoric *one*; but that topic is bound up with many questions about the contexts of use that provide the basis for learning the semantics and pragmatics of anaphors, as well as issues about whether the assumed anaphora generalization is in fact correct; see Wasow (1991: 636) for a brief comment).

What makes the auxiliary fronting case so suitable for supporting the APS is that it involves a purely syntactic matter. Initial positioning of the auxiliary signals certain English sentence types, notably polar interrogatives. It would seem that under a data-driven account, learning how these sentence types are constructed can be done, and must be done, simply from evidence about permissible word positions. Identifying the generalization about which auxiliary must be fronted to make the yes/no question corresponding to a certain statement is not dependent on learning some concept or meaning.[5] That makes it a lot easier to say exactly what a corpus of utterances would have to contain to provide necessary and sufficient evidence to permit the learning of the generalization.

In all his many discussions of the auxiliary-fronting example,[6] Chomsky never offered evidence for the empirical claim that children do unerringly pick the right subject-auxiliary inversion generalization. He just took it to be intuitively obvious. However, Crain and Nakayama (1987) eventually did examine the matter experimentally, and their results do confirm the claim: children front the main clause auxiliary, not whichever auxiliary would be first in the

declarative sentence. I will assume that Crain and Nakayama's subjects were typical, and that children do indeed always perform correctly. But this is not sufficient as a basis for the APS. No argument for innate priming can be developed merely from that fact that the generalization is structure-dependent and the fact that children's behavior conforms with it. This might be the result of uniform pressures from the linguistic environment or other aspects of the situation (most plausibly, species-invariant constraints on cognitive systems that make only hierarchically structured systems feasibly learnable by beings with tight limits on their ability to notice, store, and retrieve information). What is critical to the argument from poverty of the stimulus is the additional claim regarding stimulus poverty, and that is my focus here.

2. Stimulus poverty and auxiliary fronting. The claim about auxiliary fronting that figures as support for an instance of the APS is the very specific one that I set out in (3).

(3) *The Stimulus Poverty claim about auxiliary fronting:* The corpus of evidence presented to children during their language learning is insufficient to permit any data-driven learning method to accomplish selection of the correct auxiliary fronting generalization and elimination of all the alternatives.

Chomsky asserts this claim in quite extreme terms. He claims not just fairly low frequency for the crucial kinds of example that would distinguish structure-dependent from structure-independent formulations of auxiliary fronting; particularly in the paper and discussions published in Piattelli-Palmarini (1980), he makes much stronger claims:

(4) *Chomsky on stimulus poverty*

a. "A person might go through much or all of his life without ever having been exposed to relevant evidence, but he will nevertheless 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 in passing that (4b), if true, would undercut (4a) and (4c). 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.

Setting aside this slight inconsistency of Chomsky's claims and returning to the issue of their truth, we note that (as Freidin (1991, 618) remarks) they have never been established as true, or even seriously investigated by linguists. Few linguists seem to have paid much attention to them, in fact. However, one work, published by a linguist in a philosophy journal, has claimed that they are actually false. I refer to Sampson (1989).

Sampson cites only anecdotal evidence, but it is worth considering nonetheless. He notes that when he turned to the list of "wonder questions" in a children's encyclopedia he found crucial examples of the relevant sort within the first few questions; and he points out that William Blake's well-known poem 'Tiger' — which virtually every English speaker seems to encounter during schooling — contains the line *Did He who made the lamb make thee?*, also crucial positive evidence for the structure-dependent rule.[7] It seems to me that these observations merit some sort of response from defenders of the APS. They have not received one.

A moment of reflection should suffice to raise some suspicions that Sampson is 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 (5). (An underscore marks the position in each string where the main clause auxiliary would be if it were not fronted.)

(5) 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 boy who was hitting you _ 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, and thus the auxiliary that appears initially would not be the first auxiliary in the declarative. But of course the extra auxiliary does not need 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 (6) as well:

(6) 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 word counterpart to the string 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 anything to precede the first, is incorrect. That is crucial evidence confirming the structure-dependent generalization over any structure-independent one.

The range of relevant examples is yet wider once we notice that *wh*-movement questions in which the *wh*-phrase is a nonsubject 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 any evidence that we find examples like (7a) rather than (7b), and (7c) rather than (7d), is crucial evidence in favor of the structure-dependent auxiliary fronting hypothesis:

(7) 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?

The intuitively obvious grammaticality of these examples would suffice if the point were one of syntactic theory, but it is not sufficient to be relevant to the APS. The APS depends on a claim about *actual event probabilities*, not well-formedness. To evaluate the truth of the empirical premise of the APS we need to know whether such examples actually do turn up in everyday language use. Although generative linguists are much given to calling the problem of how infants acquire language "the logical problem of language acquisition" (the phrase that appears as the title of Baker 1981 and the subtitle of Hornstein and Lightfoot 1981), as if the issue were one to be settled by logical reasoning from known and uncontroversial data, this is not true. Whether sentences with a certain structural or semantic characteristic occur in the huge random assortment of fairly short utterances that an arbitrary human infant is likely to be exposed to during the critical years for language acquisition is not something to be settled by ratiocination. The matter at hand is entirely empirical, and its investigation will demand access to hard data of a sort with which we are

woefully underprovided at present.[8] More directly than most questions that arise in linguistics, this question seems to invite resolution through a computer search of a corpus of everyday utterances.

Ideally, what we need to settle the question is a large machine-readable corpus — some tens of millions of words — containing a transcription of most of the utterances used in the presence of some specific infant (less desirably, a number of infants) over a period of years, including particularly the period from about one year (i.e., several months earlier than the age at which two-word utterances start to appear in children's speech) to about 4 years (i.e., the age at which we can take the SAI pattern to have become established in the child's speech). I have not as yet obtained access to any extensive computer corpora that are fully appropriate to the task. However, even without a corpus that is anything like ideal we can do some useful preliminary testing.

One of the most accessible large text corpora is the forty million words of newspaper articles from the *Wall Street Journal* between 1987 and 1989 included on the CD-ROM made available by the ACL (Linguistic Data Consortium 1993). To be sure (and let me be the first to concede this), even bankers' children do not get their main exposure to English through being read to from the *Journal*. But looking at the *WSJ* corpus is not as misguided as it might at first be thought to be. The *WSJ* material contains a lot of structurally simple colloquial speech in verbatim quotes from ordinary people of all education levels who are interviewed in news stories, as well as ordinary English of every journalistic genre from news features to theater reviews to humorous essays. The notion that the questions in the *WSJ* corpus might be of high syntactic complexity is soon dispelled by browsing through them; extremely simple structures in which the auxiliary and subject are combinations like *do you* or *is it* predominate overwhelmingly. If the claim made had been merely that one-word (usually pronominal) subjects are much more frequent in polar interrogatives than anything else, it might have been tenable. But it was not. The claim we are concerned with is that subjects containing auxiliaries will essentially not be found at all.

It is highly relevant here that many statistically defined syntactic properties of running text vary little from genre to genre (recall the surprising result of Hudson 1994 that about 37% of the word tokens in running text are nouns, regardless of genre, style, modality, source, or even language). Researchers who do look at adults' speech to or in the presence of children report that "the input to young children is neither so depleted nor so uniform as some have suggested" (Berman 1990:1160). We have no reason to assume that we will get an unrepresentative sample of the syntactic types of questions that would come up in natural contexts in front of children if we simply look for question marks in the *WSJ* corpus. Speakers of English simply do not have enough conscious control over the syntactic properties of the questions they

ask to make such a source unrepresentative: the questions that get asked, whether in the boardroom or the street or the kitchen, are determined by the random informational needs and situations that come along.

In the *WSJ* corpus there are 23,886 lines containing question marks. This overcounts instances of interrogative syntax somewhat, because it counts irrelevant verbless headlines like "WARMING TREND?". It also includes large numbers of constructions irrelevant to auxiliary fronting, e.g. the extremely frequent simple subject *wh*-interrogative constructions like *Who cares?*. The obvious question to ask is how many of these one has to go through before one comes up with a crucial example falsifying the hypothesis that auxiliary fronting is structure-independent. The answer, despite the overwhelming predominance of simple structures that provide no such evidence, is only fifteen. The 15th question in the corpus[9] is (8a):

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

The critical evidence is that (8a) occurs rather than (8b), with fronting of the supportive do that would be the auxiliary of the relative clause.

Several other such examples occur within first five hundred interrogatives. One is (9a), where again I contrast what we actually find with what we would find if the structure-independent generalization were correct:

(9) a. Why did The Cosby Show's Lisa Bonet, who has a very strong screen presence, think that participating in a graphic sex scene would enhance her career as a legitimate actress? (W7_006:16426)[10]
b. *Why has The Cosby Show's Lisa Bonet, who __ a very strong screen presence, thought that participating in a graphic sex scene would enhance her career as a legitimate actress *t*?

These examples are both *wh*-interrogatives rather than polar interrogatives. If one decided to exclude such cases and concentrate purely on 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 (10a), not (10b).

(10) 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?

So far, all the examples cited involve an interaction with supportive *do* rather than having two overt auxiliaries in the corresponding declarative. But if one insists on paying attention only to examples with two overt auxiliaries in

the declarative, it makes no difference. The 180th question-containing line in the corpus is the entirely unproblematic example (11a), with two instances of the copula, one inside the (headless relative) subject NP, exactly like Chomsky's hypothetical examples. The crucial fact is that we do not find (11b) instead.

(11) 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?

As far as I have been able to find out, the *WSJ* corpus 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. But this is not so surprising. English has as many as 50 morphologically distinct auxiliary verb word forms (the number varies slightly between dialects). This yields 2,500 distinct ordered pairs of auxiliary word forms. Thus in only 0.4% of two-auxiliary cases composed at random (assuming word forms are equiprobable) will we find *is* in both auxiliary positions. If only one in each hundred questions has the right structural characteristics to count as crucial, we would expect only one double-*is* example per 25,000 questions, and the *WSJ* corpus contains fewer questions than that.

Nonetheless, I found a double-*is* example on my first scan of the only electronic text I have personally prepared, a file containing the script of Oscar Wilde's comedy *The Importance of Being Earnest*, around 100 KB. Searching for the sequence ' is ... is ...?' in the file immediately brought up (12a), which Lady Bracknell addresses to Jack Worthing in the last act. She does not use (12b).

(12) a. Who$_i$ is that young person whose hand my nephew Algernon is now holding in what seems to me a peculiarly unnecessary manner __ t_i?
b. *Who$_i$ is that young person whose hand my nephew Algernon __ now holding in what seems to me a peculiarly unnecessary manner is t_i?

Granted, Lady Bracknell is not an ideal exemplar of modern everyday colloquial speech; but her question could have been asked with less verbosity, as *Who is the girl whose hand Algernon is holding?*, or *Who is the girl who is holding Algernon's hand?*). What is really at issue here is how often in everyday life we can expect situations to arise in which questions like this are asked. Chomsky's claim that many people will *never* hear any is surely much weakened by finding attested examples so easily in any kind of English one takes the trouble to search.

One could debate further concerning what exactly is a relevant example. But it seems to me that even the preliminary results reported here hold some implications for the stimulus poverty claim. Relevant examples seem to occur

at least once per 500 interrogatives in the *WSJ* corpus. Focusing more closely on polar interrogatives increases the ratio: an examination of the corpus of polar interrogatives in the *WSJ* corpus suggests that up to 12% of the polar interrogative examples confirm the structure-dependent regularity over the structure-independent one (my vagueness being due to various questions about what sort of example should count as relevant).

Thus Sampson's anecdotally-illustrated suspicions are decisively borne out by a text search. Chomsky's assertion that "you can go over a vast amount of data of experience without ever finding such a case" is unfounded hyperbole. The putative case of hyperlearning with which Chomsky has been supporting the APS for over twenty years provides no warrant for his hypothesis that there is a specialized language-acquisition brain module pre-programmed with universal grammar. Those who seek to develop such a warrant are sorely in need of a new well-confirmed case of hyperlearning.

3. Implications and conclusions. It is not my aim to present here an argument for adopting an 'empiricist' view of language acquisition. However, it is true that casting doubt on the Stimulus Poverty claim about auxiliary fronting removes a key reason for scepticism about data-driven learning. The utterance tokens that could provide the crucial data apparently make up something between 1% of interrogatives and over 10% of polar interrogatives in running text. Rough calculations based on plausible active hours, utterance frequency, and speech rates suggest that a child would hear hundreds of thousands of questions during the language acquisition period, and thus must hear thousands of examples that crucially confirm the structure-dependence of auxiliary fronting. This does not show that there is no innate priming, but it does show that the best and most often-repeated claim in support of the empirical premise of linguists' central argument for innate priming is false.

Searching for new cases of linguistic hyperlearning to support the APS will bring generative linguists into contact with two things toward which they have traditionally shown considerable antipathy: research results on formal learning theory, and the methods of corpus linguistics. The relevance of formal learning theory is that it is the mathematical study of the limits on data-driven learning, and without clear results on that, hyperlearning cannot even be characterized. And corpus study is relevant because a claim that hyperlearning occurs will incorporate a specific claim about what occurs in typical corpora of material available to infants during their critical period for language acquisition. Generative linguists interested in defending the APS are going to have to become more broad-minded about these matters if they are going to accomplish the task they have set themselves.

Interestingly, work on establishing the soundness of the APS is almost certainly going to be self-undercutting to some degree. The research will necessarily involve close study of what data-driven learning procedures can do.

It is highly likely that research on that topic will yield some improvements in the success rates of such procedures. Those successes are likely to eliminate apparent cases of hyperlearning by revealing that more can be learned in a data-driven way than was previously thought. To some extent this is happening already. Brent (1993) has demonstrated that an "unsupervised" algorithm taking raw text as input and using rudimentary facts about grammatical diagnostics (like what the verbal affixes are) can identify the verbs of a language and their subcategorization frames. And even more remarkably, Schütze (1995) develops methods for deducing syntactic category, semantic class, word sense for ambiguous forms, and subcategorization information from raw text alone, using nothing more than the distributional information in the corpus.

A substantive defense of the APS will have to be based on study of this sort of research. The strategy would be to establish known limits on what such data-driven learning algorithms can do, and then search for instances of a child engaging in hyperlearning in the specific sense of acquiring knowledge about a language that such algorithms provably could not induce (within a reasonable time) from the corpus presented to that child. Nothing like this has so far been attempted.

I hazard no guess about whether innate mechanisms for first language acquisition will turn out to be surprisingly minimal or surprisingly rich. Perhaps, at some time in the future, data-driven algorithmic techniques for language learning will hit a serious roadblock, or perhaps some kind of genuine evidence for innate task-specific language acquisition mechanisms will turn up. But linguists have been paying virtually no attention to data-driven learning for the last forty years, so they are currently ill-equipped to speculate on the matter.

Some philosophers have accepted linguists' claims about the APS, and treated it as an established philosophically relevant contribution of linguistic research to the cognitive sciences. Fodor (1981, 258), for example, a year after Chomsky introduced the phrase 'poverty of the stimulus', described the APS as "*the* existence proof for the possibility of cognitive science". That was somewhat premature. Fifteen years later the APS still provides no such existence proof. Its soundness as an argument depends on finding empirical confirmation of hyperlearning in first language acquisition. Such confirmation is still awaited.

Footnotes

[1] This work was partially supported by a grant from the Humanities Division to the Mind and Meaning Research Group at the University of California, Santa Cruz. Thanks to Geoffrey Sampson for sending me a copy of Sampson (1989), which started me thinking about this topic; to Hinrich

Schütze for sending me a copy of his dissertation; to Melissa Bowerman, Dave Chalmers, Eve Clark, Herb Clark, Gerald Gazdar, Lotus Goldberg, Lise Menn, Cris Moddeé, Daniel Osherson, Jaye Padgett, Pete Schult, Dan Slobin, Rachel Walker, Tom Wasow, and Kathryn Zally for useful comments and discussion; to Cher Bergeon for a fine copy-editing job (remaining errors are the fault of the typesetter and author, who are unfortunately token-identical), and particularly to Barbara C. Scholz, who read several drafts and supplied incisive and useful criticisms.

[2] Strictly, there is a different notion of hyperlearning for each of the indefinitely many fully precise definitions of data-driven learning procedures that can be given, but the approximate general characterization offered in the text will suffice for me to make the point I want to make in this paper.

[3] It should be clear that neurophysiological evidence would not suffice. The point is not about brain mechanisms, but rather about the interpretation of their operations in terms of specificially linguistic knowledge states, and we have no idea of how to glean that sort of information through dissection of neural tissue. We cannot even do it for CPU chips, since one cannot identify data inscribed in silicon without knowing something about the encoding.

[4] Among others, see Hornstein and Lightfoot (1981: 18ff), Lightfoot (1982a, chapter 4), Lightfoot (1989: 322f), and Crain (1991: 609ff).

[5] Actually, it is not quite correct to say that auxiliary-initial syntax is innocent of semantic entanglements. There is interesting evidence, originally pointed out by Joseph Emonds and discussed in Chomsky (1971a, 209–210), of meaning differences between auxiliary-initial constructions and the subject-initial structures from which they are derived in transformational accounts (note that *I shall go* makes a definite future prediction, but *Shall I go?* does not simply ask about the truth value of that prediction, but rather asks if the speaker *should* go, or whether the hearer would like the speaker to go). But my assumption is that if special meanings are 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 via hyperlearning, not easier. I set aside this sort of evidence here in order to give the proponents of the APS the benefit of the doubt.

[6] The first reference in the linguistics literature to the structure-dependence of auxiliary fronting is in Chomsky (1965: 55–56). Chomsky (1968: 51–52) repeats the point, adding that the "language-learner knows" that only structure-dependent operations are available. But over the last twenty-five years, Chomsky has taken things further, and developed an explicit argument that hyperlearning occurs. This begins with Chomsky (1971b: 29–33) and

continues in Chomsky (1975: 30–33; 153–154), Piattelli-Palmarini (1980: 114–115), and (with the data adapted to Spanish) Chomsky (1988: 41–47). The claims have since been repeated by many others (e.g., Marcus 1993: 80, Pinker 1994: 40–42; 233–234).

[7] Sampson also gives an epistemological argument against the APS: he argues that anyone defending the stimulus poverty claim by exhibiting a fact F about a language L that could not be induced from the evidence of ordinary use of L must face the question of how they know that F is a fact. If the warrant they offer for F comes from evidence of use of L, they have contradicted themselves by conceding that such evidence is available; if the warrant is held to be the result of knowledge gained by an adult native speaker investigator via her innate priming, they have committed the fallacy of *petitio principii*; and there appear to be no other cases. This may be a valid additional objection to the APS (whether or not it is depends on whether it can be made clear that the linguist has, and standardly uses, some third way of finding out about what the rules of a grammar are, one that depends neither on mere facts of what sentences are grammatical nor on the putative fact that the linguist has been handed the evidence by an innate mechanism), but I do not take a position on that here.

[8] Despite the promising start provided through the CHILDES database; see MacWhinney (1995).

[9] All examples from the *Wall Street Journal* corpus in this paper are taken from the 1987 directory, **/WSJ/1987**. I append to each an identifier of the form *file-name:line-number*. The example just cited is in the file **w7_001**. The UNIX command `fgrep "?" w7_001 | head -15 | tail -1` will find it.

[10] I have silently removed some strange and typographically incorrect double quotes that appear around the phrase *The Cosby Show's* in the original; they are irrelevant to the point at hand.

References

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

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

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

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge: MIT Press.

Chomsky, Noam. 1968. *Language and Mind.* New York: Harcourt, Brace and World.

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

Chomsky, Noam. 1971b. *Problems of Knowledge and Freedom.* London: Fontana.

Chomsky, Noam. 1975. *Reflections on Language.* New York: Pantheon.

Chomsky, Noam. 1980. *Rules and Representations.* Oxford, U.K.: Basil Blackwell.

Chomsky, Noam. 1988. *Language and Problems of Knowledge.* Cambridge: MIT 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.

Demopoulos, William. 1989. On applying learnability theory to the rationalism-empiricism controversy. Matthews and Demopoulos, eds., 77–88.

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. Guttenplan, ed., 366–374.

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

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

Hornstein, Norbert, and David Lightfoot. 1981. Introduction. *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.

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

Lightfoot, David. 1982b. Review of Geoffrey Sampson, Making Sense. *Journal of Linguistics* 18, 426–431.

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

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. 1989. The plausibility of rationalism. Matthews and Demopoulos, eds., 51–75.

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

Osherson, Daniel, Michael Stob, and Scott Weinstein. 1986. *Systems That Learn.* Cambridge: MIT Press.

Piattelli-Palmarini, Massimo. 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.

Pollard, Carl, and Ivan A. Sag. 1991. An integrated theory of complement control. *Language* 67, 63–113.

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

Schütze, Hinrich. 1995. *Ambiguity in Language Learning: Computational and Cognitive Models.* Ph.D. dissertation, Stanford University.

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

Stich, Stephen P. 1981. Can Popperians learn to talk? *British Journal for Philosophy of Science* 32, 157–164.

Wasow, Thomas. 1991. Debatable constraints. *Behavioral and Brain Sciences* 14, 636–637.

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