

# Learnability, Hyperlearning, and the Poverty of the Stimulus\*

Geoffrey K. Pullum  
University of California, Santa Cruz

Presented at the Parasession on Learnability,  
22nd Annual Meeting of the Berkeley Linguistics  
Society, Berkeley, California, February 18, 1996.

+-----+  
| Copyright (c) 1996 by the Berkeley Linguistics Society. All rights |  
| reserved. This is a preliminary draft of a paper that will be |  
| be published in revised form by the Berkeley Linguistics Society. |  
| It is being made electronically available for purposes of private |  
| study only. Please do not cite or quote this draft in any published |  
| work. Correspond with the author to supply comments and/or to obtain |  
| the final version. Internet address: "pullum@ling.ucsc.edu"; postal |  
| address: Stevenson College, UCSC, Santa Cruz, California 95064, USA. |  
+-----+

+-----+  
| \* ACKNOWLEDGEMENT NOTE: I am grateful to Geoffrey Sampson for |  
| sending me a copy of Sampson (1989), which started me thinking |  
| about this topic, to Hinrich Schuetze for sending me a copy |  
| of his dissertation, to Dave Chalmers, Gerald Gazdar, Lotus |  
| Goldberg, Tom Wasow, and the students in my research seminar |  
| for useful comments and references, and particularly to Barbara |  
| C. Scholz, who subjected my first thoughts on this topic to |  
| much-needed penetrating interrogation. |  
+-----+

## 1. Stimulus poverty and hyperlearning.

Hornstein and Lightfoot (1981) make a bold and unambiguous claim concerning first language acquisition:

(1) "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)

Let me introduce the term "hyperlearning" to denote this feat. That is, I will say that one accomplishes hyperlearning if one acquires some piece of knowledge K without being exposed during the relevant learning period to any evidence that could rationally establish K.<sup>[1]</sup> I propose to question whether hyperlearning is in fact ever attested in the domain of human language acquisition.

This issue is of importance primarily because of its central relevance to the Argument from Poverty of the Stimulus (APS). The claim that rationalism' is confirmed over empiricism' by findings about language acquisition is the most celebrated of the claimed impacts of generative linguistic research on philosophy and psychology, and the APS is generally agreed to be the most important argument in this connection. I believe the argument is not sound.

An initial obstacle to addressing this issue is that the alleged argument, first dubbed the APS by Chomsky (1980:34), and espoused thereafter by many others, seems never to have been clearly stated. [2] Let me therefore begin by attempting to do that. We will need to distinguish between two ways in which an infant might in principle learn a language. (The reader will doubtless see in them the outlines of empiricism' and rationalism' as characterized in traditional epistemology, but in most of this paper I will avoid using those terms because of the baggage they carry.)

The first involves nothing but methods of rational belief fixation (inductive inference) applied to a corpus of observed utterances in natural contexts. Crucially, the learning is assumed not to be informed by preconceptions of any sort about what languages are like. I will call this "data-driven learning". It is important 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 do the same. Notice, therefore, that if this is the way first language acquisition is accomplished, we will never see an instance of hyperlearning in the first language acquisition domain.

The second way in which languages might be learned involves the learner being primed ab initio with special information, or endowed with special internal mechanisms that make available specific information about the domain. Call learning of this sort "innately-primed learning". By definition, innately-primed learning proceeds in a way that does not limit the learner's resources to the corpus of relevant observations. It is compatible with the occurrence of hyperlearning, because although the corpus might not suffice for learning what is learned, the sum of the contributions of the corpus and the innate priming might suffice.

Some may take it to be trivial that one or the other of these must be correct: either there is innate priming or there is not. Others claim that additional distinct positions lying between the two can be made out (see in particular Stich 1979). Garfield (1994: 367) notes that some regard "empiricism" and "rationalism" as merely two regions in a continuum, so that there are indefinitely many alternative positions. Ignoring these disputes, I propose simply to grant for the sake of argument what advocates of the APS assume: that data-driven learning and innately-primed learning are distinct, mutually exclusive, and jointly exhaustive. That yields a disjunction that constitutes the initial premise of the APS, which I can now set out as in (2).

(2) 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 learn their first languages by means of data-driven learning, then hyperlearning will never be observed in this domain. [Immediate, from the characterization of data-driven learning methods.]
- c. Hyperlearning does in fact occur in the domain of first language acquisition by infants. [Empirical premise asserted in (1).]
- d. Human infants do not learn their first languages by means of data-driven learning. [From (b) and (c) by modus tollens.]
- e. Human infants learn their first languages by means of

innately-primed learning.  
[From (a) and (d) by disjunctive syllogism.]

This is a valid argument. My concern is with its soundness, and in particular with whether premise (2c) is true. My strategy will be to examine the strongest and best-known pillar of support for (2c) 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 (3):

- (3) 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 based on structural relations rather than just linear sequence: 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.

The first reference to the structure-dependence of auxiliary fronting -- the fact that it needs dominance as well as precedence to state it -- is in Chomsky (1965: 55 6). Chomsky (1968: 51 2) repeats the point, adding that the “language-learner knows” only to use structure-dependent operations. Chomsky (1971: 29 33) begins to expound an explicit case for hyperlearning, and this is echoed by similar passages in Chomsky (1975: 30 33; 153 154), Piattelli-Palmarini (1980: 114 5), and (adapting the data to Spanish), Chomsky (1988: 41 7). The claims have since been repeated by many others (e.g. Marcus 1993: 80, Pinker 1994: 40 42; 233 4).

Quite apart from its celebrity, there are reasons for taking this case to be potentially the strongest kind of case for the APS, because it is purely syntactic. A question like how the child learns that Mary appealed to the men to like each other is well-formed while \*Mary appealed to the men to like each other is not (Chomsky 1975: 140 153) cannot sensibly be considered in isolation from issues of how the child learns about the uses of reciprocals and the control patterns of verbs; and these issues seem intimately bound up with semantics. But in the auxiliary fronting case we are dealing with a purely syntactic device for signalling a difference in sentence type. It is purely syntactic evidence that is needed to identify which device is being used.

Chomsky never offered evidence for the empirical claim that children always do correctly identify that syntactic device. He just took it to be intuitively obvious. But Crain and Nakayama (1987) have examined the matter experimentally, and their results did support the claim, by showing that when children are manipulated into a situation where asking a question like (4a) is appropriate, they never utter strings like (4b) instead.

- (4) a. Is the boy who is in the corner smiling?  
b. \*Is the boy who in the corner is smiling?

Let us assume that Crain and Nakayama’s subjects were typical, and children do indeed always learn the correct generalization.[\[3\]](#) 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 are uniformly successful at discovering it. What is critical is an additional claim, the one that involves stimulus poverty.

## 2. Stimulus poverty and auxiliary fronting.

The claim about auxiliary fronting that permits it to be used in an instance of the APS is the very

specific one in (5).

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

Chomsky asserts this claim in quite extreme terms. He claims not just that the crucial kinds of example to distinguish structure-dependent from structure-independent formulations of auxiliary fronting have a lower frequency than examples with simple subject NPs (which is certainly true); in the paper and discussions published in Piattelli-Palmarini (1980), he makes the much stronger claims quoted in (6).

- (6)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 (6b), if true, undercuts (6a) and (6c). 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 many speakers 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. Here I prescind away from this suspicious epistemological aspect of the claim, and turn to the issue of its truth.

One scholar has suggested in print that the claims in (6) are empirically false. Sampson (1989) 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 poem 'Tiger', which no one seems to go through grade school without encountering, contains the line "Did He who made the lamb make thee?", also crucial positive evidence for the structure-dependent rule. These observations are anecdotal, certainly, but they certainly need some sort of response from defenders of the APS.

A few minutes of critical reflection suffice to raise some doubts. Could you really expect to live your whole life as an English speaker, or even reach kindergarten, without running into any sentences of the sort illustrated in (7)? (In these examples, the position in each string where the main clause auxiliary would be if it were not fronted is marked with underlining.)

- (7)a. Would anyone who is interested     see me later?  
b. Could the man who has lost his ticket     come to the desk?  
c. Can the people who are leaving early     please sit near the door?  
d. Will those who are coming     raise their hands?

- e. Can a helicopter that has lost its tail rotor \_\_ still fly?
- f. Will the owner of the car that is blocking the driveway \_\_ please move it?
- g. Is the boy who was bothering you \_\_ still here?
- h. Could a tyrannosaurus that was sick \_\_ beat a triceratops in a fight?

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 (8) as well.

- (8)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?

The notion of fronting the wrong auxiliary might seem perverse here, because the preposed clauses precede the whole domain of the correct auxiliary fronting. But of course, the very idea of a structure-independent rule demands that we imagine ourselves with no access to such notions as clause boundaries. The question is whether there is naturally-occurring evidence against the hypothesis that a string like (9a) can be made into a polar interrogative by fronting the first auxiliary (word 3, rather than word 7). The answer is that corresponding to (9a) we find (8a) = (9b), rather than (9c), or for that matter (9d), the result of fronting the right auxiliary but fronting it too far in the string.

- (9)a. if1 you2 don't3 need4 this5 I6 can7 have8 it9
- b. if1 you2 don't3 need4 this5 can7 I6 have8 it9
- c. \*don't3 if1 you2 need4 this5 I6 can7 have8 it9
- d. \*can7 if1 you2 don't3 need4 this5 I6 have8 it9

This crucially confirms the structure-dependent generalization over the 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 (10a) rather than (10b) is crucial evidence in favor of the structure-dependent auxiliary fronting hypothesis:

- (10)a. How could anyone who was awake not hear that?
- b. \*How was anyone who awake could not hear that?

These examples all look plausible enough, but they are invented. What we need to know is whether such examples actually turn up in natural language use. This calls for a corpus search. Ideally, what we would want is a large machine-readable corpus containing a transcription of most of the utterances used in the presence of some specific infant over a period of years. By large I mean tens of millions of words. Less desirable but still of some use would be a large corpus of representative utterances used in natural contexts in the presence of a number of infants in their critical period for language acquisition. As far as I know, there are no such corpora in existence.

Lacking a corpus that was anywhere near ideal, I searched what I had: the text corpus on the CD-ROM made available by the ACL (Linguistic Data Consortium 1993). This contains about

forty million words of newspaper articles from the “Wall Street Journal” between 1987 and 1989. Now, let me be the first to concede that even bankers’ children do not spend their early years being read to from the Journal. But the exercise is not as misguided as it might superficially seem.

First, some data is better than no data at all. Second, the WSJ material is more suitable than one might think; it contains a lot of structurally simple colloquial speech in verbatim quotes from ordinary people interviewed in news stories, as well as ordinary English of every journalistic genre from news features to theater reviews to humorous essays. And third, note 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).

There are about 24,000 interrogatives in the WSJ corpus. (To be precise, there are 23,886 lines, nearly always one sentence per line, in which a question mark appears. This overcounts instances of interrogative syntax, because it counts irrelevant verbless headlines like “WARMING TREND?”, and it 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 explore using such a corpus is how many of these 24,000 questions one encounters before one comes up with a crucial example falsifying the hypothesis that auxiliary fronting is structure-independent. The answer is fifteen. The 15th question in the corpus is (11a).[\[4\]](#)

- (11)a. How fundamental are the changes these events portend?  
(LDC-93-T1:WSJ\1987\W7\_001:3963)
- b. \*How fundamental do the changes these events portend are?

The crucial evidence is supplied by the fact that (11a) occurs rather than (11b), with fronting of the supportive do that would be the auxiliary of the relative clause.

This example is a wh-interrogative rather than a polar interrogative, and it involves a supportive do rather than an auxiliary that makes a semantic contribution. Suppose one wanted to exclude such cases. Would it make much of a difference? The answer is that it would not. The 180th interrogative in the corpus[\[5\]](#) is the entirely unproblematic example (12a), actually involving two instances of the copula, exactly like Chomsky’s hypothetical examples. The crucial fact is that we do not find (12b) instead.

- (12)a. Is what I'm doing in the shareholders' best interest?  
(LDC-93-T1:WSJ\1987\W7\_003:2991)
- b. \*Am what I doing is in the shareholders' best interest?

It would be possible to engage in a certain amount of debate about what exactly one should count as a relevant example. (I will not illustrate this in detail here for reasons of space. The facts are readily accessible to anyone else who wants to take a look.) Suffice it to say that at least five crucial examples occur in the first 500 interrogatives; for example, the 387th interrogative in the corpus is (13), and the 456th is (14).

- (13) Is a young professional who lives in a bachelor condo as much a part of the middle class as a family in the suburbs?  
(LDC-93-T1:WSJ\1987\W7\_006:2813)
- (14) 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?  
(LDC-93-T1:WSJ\1987\W7\_006:16426)

Finding five crucial examples in the first 500 interrogatives suggests that there may be as many as one crucial auxiliary fronting case for every hundred cases of interrogative mood. And focussing more closely on polar interrogatives seems to push the number up rather than down; my examination of the full corpus of polar interrogatives in the corpus suggested that about 12 percent of the examples crucially confirmed the structure-dependent regularity over the structure-independent one.

Thus Sampson's suspicions seem to be borne out.<sup>[6]</sup> Chomsky's assertion that "you can go over a vast amount of data of experience without ever finding such a case" is unfounded hyperbole. And this is apparently the best candidate for a case of attested hyperlearning ever put forward -- the candidate Chomsky has been using to support the APS for over twenty years. I conclude that the defenders of the hypothesis that there is a specialized language-acquisition brain module pre-programmed with universals of language are sorely in need of a new well-confirmed case of hyperlearning.

### **3. Implications.**

I am not aiming to present here an argument for adopting an empiricist' view of language acquisition. However, casting doubt on the Stimulus Poverty claim about auxiliary fronting does remove a key reason for thinking that the success rate of children at learning the structure-dependency of auxiliary fronting cannot be explained in terms of data-driven learning. The utterance tokens that could provide the crucial data apparently make up between 1% and 10% of interrogatives. A child obviously hears hundreds of thousands of sentences while engaged in language acquisition, and thus will 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 greatly weakens the support for it. Children could be learning in a data-driven way that auxiliary fronting is structure-dependent. The APS, then, loses its vital premise (2c), and cannot at present be shown to be sound.

Searching for new cases of linguistic hyperlearning will involve generative linguists in intimate involvement with two things toward which they have typically shown considerable antipathy: research results on formal learning theory, and the methods of corpus linguistics. The relevance of formal learning theory (see Osherson et al. 1986 for an introduction) 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. Thus generative linguists, if they are going to develop serious conclusions about language acquisition using the APS, are going to have to become more broad-minded in these two respects.

Serious work on establishing the soundness of the APS is almost certainly going to be partially self-undercutting. Since it will involve close study of the capabilities of data-driven learning procedures, it is highly likely that improvements in the success rate of such procedures will be attained in the course of research. Such successes will eliminate many apparent cases of hyperlearning. I have no space here to expand on this point, but already Brent (1993) has demonstrated that an "unsupervised" algorithm taking as input just raw, untagged text plus extremely rudimentary knowledge about grammatical cues (e.g., "if X occurs and Xing occurs, then X is probably a verb"; "if a determiner occurs before X then X is probably not a verb"; "the word THE is a determiner") can identify the verbs of a language and their subcategorization

frames.

Even more remarkably, Schuetze (1995) develops methods for deducing syntactic category, semantic class, word sense for ambiguous forms, and subcategorization information from raw text alone, on the basis of no syntactic information except the distributional information in the corpus. In short, more is learnable from an unanalyzed corpus than most linguists think.

A serious defense of the APS will have to be based on a study of work of this sort. The strategy will involve establishing known limits on what data-driven learning algorithms can do, and then searching for instances of children acquiring knowledge about a language that such algorithms provably could not induce (within a reasonable time) from a corpus that could plausibly have been presented to a child under normal child-rearing conditions. Nothing like this has yet been attempted. I have shown that the case of the structure-dependence of auxiliary fronting is probably not a good place to look.

It may indeed be the case that something construable as a system of Cartesian innate ideas is in play when human infants learn natural languages. But as yet, there is no support for that view to be garnered from the Argument from Poverty of the Stimulus.

## REFERENCES

- 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, Massachusetts: MIT Press.
- Chomsky, Noam. 1968. *Language and Mind*. New York: Harcourt, Brace and World.
- Chomsky, Noam. 1971. *Problems of Knowledge and Freedom*. London: Fontana.
- Chomsky, Noam. 1975. *Reflections on Language*. New York: Pantheon.
- Chomsky, Noam. 1980. *Rules and Representations*. Oxford: Basil Blackwell.
- Chomsky, Noam. 1988. *Language and Problems of Knowledge*. Cambridge, Massachusetts: MIT Press.
- 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.
- Garfield, Jay L. 1994. Innateness. Samuel Guttenplan, ed., *A Companion to the Philosophy of Mind*, 366-374. Oxford: Basil Blackwell.
- Gold, E. Mark. 1967. Language identification in the limit. *Information and Control* 10, 447-474.
- Hornstein, Norbert, and David Lightfoot. 1981. Introduction. *Explanation in Linguistics*, 9-31. Longmans, London.

- Hudson, Richard A. 1994. About 37% of word-tokens are nouns. *Language* 70, 331-339.
- Lightfoot, David. 1982. Review of Geoffrey Sampson, *Making Sense*. *Journal of Linguistics* 18, 426-431.
- Linguistic Data Consortium (1993) CD-ROM no. LDC-93-T1. University of Pennsylvania.
- 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, MA: The 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.
- Sampson, Geoffrey. 1989. Language acquisition: growth or learning? *Philosophical Papers* 18, 203-240.
- Schuetze, 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.
- Wexler, Kenneth. 1991. The argument from poverty of the stimulus. Asa Kasher, ed., *The Chomskyan Turn*, 252-270. Cambridge, Massachusetts: Basil Blackwell.
- White, Lydia. *Universal Grammar and Second Language Acquisition*. Amsterdam: John Benjamins.

## FOOTNOTES

[1.](#) I am not suggesting that this is an adequate exact definition. For one thing, I am being deliberately vague about the character of K here; we can take it to be propositional knowledge, so that we can talk about such things as inferring K, but that is not ultimately essential, and one might discuss such things as skill acquisition in similar terms. And for another, there is really a different definition of hyperlearning for each fully precise definition of data-driven learning procedures. However, I think the rough definition offered in the text suffices to gloss the shorthand term that is all I need for present purposes.

[2.](#) Chomsky (1980:31 4) gives some general discussion of how learning is “better understood as

the growth of cognitive structures along an internally directed course under the triggering and partially shaping effect of the environment” (p.33), and then 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.’” Hornstein and Lightfoot (1981) take up the phrase but do not spell out an argument; Lightfoot (1982:428) insists on the importance of the “poverty of stimulus problems, i.e. where there are no data available to the child which will suffice to establish some rule or principle,” but also states no explicit argument; Sampson (1989) critiques the argument but offers only inexplicit quotations from Chomsky to exhibit it; whole articles such as Demopoulos (1989), Matthews (1989), and Wexler (1991) are devoted to discussing the argument without clearly stating it. (Let me quote Wexler (1991:253) introducing the topic, for example: “Chomsky... asked the question of explanation: 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.” And having thus confounded four different questions (how grammars develop, why grammars are the way they are, whether hyperlearning takes place, and what is innate), Wexler simply announces that “As Chomsky pointed out, this is an application of the classic rationalist argument from the poverty of the stimulus.” But no argument is given or referenced.)

Stich (1979, 1981) and Garfield (1994) come a bit closer to offering a statement of the argument, but miss the central point about learning things for which there is no evidence in the corpus, and confuse it with other points (underdetermination of theories by evidence; occurrence of errors in the corpus; etc.) from which Sampson (1989) is careful to distinguish it. Space precludes a detailed discussion here.

**3.** I am aware of one study that looked at the issue in connection with second language acquisition. White (1989:63-66) summarizes work by Y. Otsu and K. Naoi 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 desired sort). This is not particularly strong support of the claim that innate priming defines the possibilities.

**4.** The example is in w7\_001, the first file in the 1987 directory. This UNIX command will pick it out:

```
fgrep "?" w7_001 | head -15 | tail -1
```

Sentences quoted from the Wall Street Journal corpus are given with an identifier of the form X:Y:Z where X is the Linguistic Data Consortium catalog number for the CD-ROM, Y is the MS-DOS path name of the file on the disk, and Z is the line number in the file.

**5.** The reader can verify this with the following UNIX command on the files in the 1987 directory of the WSJ corpus:

```
fgrep "?" * | head -180 | cat -n | tail
```

**6.** Sampson also has 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 the evidence is available; if the warrant is held to be the result of knowledge gained via innate priming they have committed the fallacy of *petitio principii*; and there are no other cases. This may be a valid additional objection to the APS, but I have not appealed to it here.