

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/242486501>

# Understanding Stimulus Poverty Arguments

Article *in* The Linguistic Review · January 2002

DOI: 10.1515/tlir.19.1-2.105

---

CITATIONS

34

---

READS

78

2 authors, including:



[Janet Dean Fodor](#)

CUNY Graduate Center

60 PUBLICATIONS 3,413 CITATIONS

SEE PROFILE

All content following this page was uploaded by [Janet Dean Fodor](#) on 17 September 2014.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

# Understanding stimulus poverty arguments<sup>1</sup>

JANET DEAN FODOR AND CARRIE CROWTHER

## *Abstract*

*The argument from the poverty of the stimulus as Pullum and Scholz define it (their APS) is undeniably true, given that all language learners acquire the ability to generate more sentences of the target language than they have heard. Uniformity across learners with respect to the additional sentences they project suggests that grammar induction is guided by general principles, which must be innate. What remains to be established is exactly which sentences can be projected on the basis of which others. The details of this are important to linguistic theory and to the psycho-computational modelling of natural language acquisition. They are not of great significance to the generic issue of nativism versus empiricism, except that they may clarify the extent to which the innate knowledge in question is specific to language. The argument for linguistic nativism appears to be solidly supported by the distinctive patterns of generalization that learners adopt in the absence of systematic negative evidence (a limitation that Pullum and Scholz exclude from APS). We argue that innate knowledge of how to represent natural language facts is necessary in order for learners to extract from their input the information that it does contain. Pullum and Scholz themselves rely on Universal Grammar in just this role when they make specific suggestions as to how learners arrive at the right generalizations.*

---

1. We thank Virginia Valian for her advice, and for the manuscript of her forthcoming book *Input and Innateness: Controversies in Language Acquisition*. We are grateful to Marcel den Dikken for approving our syntactic analyses. We have benefitted also from energetic discussion at the CUNY Cognitive Science Symposium.

### 1. A challenge

It is useful to be provoked occasionally into scrutinizing the deeper assumptions on which one's research is based, even if it can be irritating (as we discover) when the provokers deploy burden-of-proof arguments to ensure that they can say what work is needed, without themselves making a serious effort toward getting it done. Pullum and Scholz issue a challenge to believers in the poverty of the stimulus for language acquisition: Show one clear case in which some information necessary for acquisition is missing from the language input received by the learner. They imply that until a better demonstration is provided than any at present, the assumption of innate linguistic knowledge, which has been largely based on stimulus poverty arguments, is on shaky ground.

Pullum and Scholz cite four cases, one morphological and three syntactic, for which it has been claimed that children demonstrably know things about the target language which are not established by the language samples to which they are exposed. We will address only one of these four case studies, the case of auxiliary inversion in English. In responding to Pullum and Scholz's challenge we will comment on methods and research strategies, on what adults say to children, and on the generalizations that children formulate. Some of our remarks make very obvious points which other commentators will surely note also, such as the absurdity of using the *Wall Street Journal* as a source of information about language input to small children. But mostly we have tried to address the real issues that Pullum and Scholz raise while skirting the frivolity of some of the arguments they use in doing so.

A central theme that we will expand on in the discussion below is that the implications of missing positive exemplars for language acquisition cannot be judged in isolation from the question of missing negative evidence (evidence about what is not a sentence of the target language). Positive and negative facts together can go far to establish the grammar of a language, but one without the other is of limited value, as we will show. The consequences of missing positive data are therefore more far-reaching in the absence of reliable negative evidence. As Pullum and Scholz define the argument from the poverty of the stimulus (their APS), it includes in its scope only the absence or scarcity of positive exemplars; the paucity of negative data is specifically excluded from the discussion. To distinguish our broader perspective from Pullum and Scholz's approach, we will use our own abbreviation, POS, for the (purported) poverty of the stimulus for language acquisition. POS includes *every* respect in which learners' input underdetermines the adult grammar. In Section 3 below we indicate how POS relates to APS.

Before heading further into the skeptical arguments that Pullum and Scholz present, it is fair to ask whether they could possibly make any difference to views on innateness. After all, any recursive rule (or its equivalent in non-rule-

based grammars) suffices to establish that natural languages, since they are infinite, are necessarily underdetermined by finite input samples. Thus, the existence in adult grammars of a recursive rule of complement clause embedding is all it takes to establish stimulus poverty in Pullum and Scholz's sense, such that "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" (Section 1: 10). A typical adult English speaker knows (and will reluctantly admit) that a sentence with 29 embeddings along the lines of *Mary thinks that it is clear that John imagined that . . .* is well-formed in English. But no corpus study will show that English learners are typically exposed to sentences with 29 embeddings. So it would appear that the debate on stimulus poverty could stop right here. What do the other case studies presented in the literature and highlighted in Pullum and Scholz's article contribute that goes beyond this mundane example? Why have Chomsky and others emphasized them? And how much could skeptics gain by fending off those particular examples of stimulus poverty, if others are invulnerable?

The answer, we imagine, is that the more elaborate examples, though not crucial for disproving that acquisition is purely data-driven, do a better job of showing that the information missing from the input, which is reliably filled in by all normal learners, is specifically linguistic. Grammars are certainly underdetermined by input samples, but this establishes the existence of innate linguistic knowledge (equivalently: Universal Grammar, UG), only if it can be shown that grammars are underdetermined by input samples even in conjunction with general principles of induction. Since inductive principles must exist for other domains of knowledge, it is parsimonious to assume that the same or related principles are all that is at work in the case of language. To counter this, it needs to be shown that the only way to bridge the gap between input sample and grammar is with highly domain-specific information, such as facts about case marking, or the domains for anaphor binding, or the constraints on displacement of various syntactic categories. Chomsky signals this in the passage in which he first introduced the term *poverty of the stimulus*, where he refers to "the vast quantitative difference between the highly impoverished and unstructured environment . . . and the *highly specific and intricate structures* that uniformly develop" (Chomsky 1980: 34; our emphasis). The specificity and intricacy of what learners must project is part of the value of the celebrated examples, and from that point of view it would indeed set back the theory of linguistic innateness if these examples were to be undermined.

For this reason, Pullum and Scholz's challenge is worth taking up. We embark on the task with the following understanding. APS in its least interesting form is more or less incontrovertably true, as we have noted, and it would remain so even if every one of the four cases that Pullum and Scholz take on were to fall victim to corpus studies showing ample supplies of positive examples in

children's input. So if the issue is whether learners' evidence underdetermines their hypotheses in such a way that innate guidance must be assumed, it hardly matters how the corpus facts turn out. On the other hand, if these cases do survive Pullum and Scholz's investigation, they can contribute on two fronts. They can add richness of detail to fill in the particulars of a working model of UG and its deployment by learners. And in relation to the larger issue of nativism versus empiricism, these examples could help to show how much of what is innate is distinctively linguistic.

## **2. What do children hear?**

A first step is to set the record straight with respect to the kinds of sentences that children hear. They are very different from those in written texts for adults. Pullum and Scholz "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" (Section 4: 22–23). Nevertheless, such transcripts exist (most notably in the CHILDES database; [MacWhinney 1995](#)), and they seem like the sensible place to start.

The samples are limited. What is recorded varies from a few hours in total to an hour a week for several years. Only some of it is natural conversation, and only some is an adult speaking. What the adults (or the television) said in the many hours when the tape recorder wasn't running will never be known, so this is not a good method for tracking lacunae. We make no claims here about what linguistic constructions are absent from learners' input. But we can make two modest contributions: (a) some perspective on the relation between child-directed speech and other genres; and (b) a reminder that temporal correlations between the sentence types used by adults and by children may occur for reasons other than that the children are learning from what the adults say.

### *2.1. Comparison with other genres*

The basic facts of child-directed speech are well-documented in the literature, but are worth summarizing here. Adults (and older children) do characteristically speak to infants in a distinctive style (though there are some cross-cultural differences; see articles in Snow and Ferguson 1977). In support of the linguistic uniformity of different genres, Pullum and Scholz cite Hudson's (1994) finding that the percentage of nouns is constant across many styles of writing and speech. But this is a very simple yardstick; and even so, it may not hold for child-directed speech (see below). Certainly, other properties of child-directed speech do not conform to the norms for adult-to-adult speech or

writing. Speech to children is characterized by phonological clarity and shortness of sentences. Following [Newport et al. \(1977\)](#) it is fairly widely agreed that these properties have more to do with the communicative load of everyday adult-child interaction (getting children to do some things, preventing them from doing others), than with helping them acquire the language. But whatever their origin, the documented linguistic differences are at least grounds for caution, if the uniformity-of-genres thesis is to be extended to speech to children as they begin to acquire syntax.

Newport (1977) examined approximately 100 utterances by each of 15 middle-class mothers to her child, and approximately 50 utterances each by the mother to the experimenter in the same setting. She found not only very different sentence lengths between the two speech styles, but also a different distribution of syntactic constructions. The proportions of declaratives, interrogatives, imperatives, and deictic structures (such as *That's a rabbit*) differed, and so did the frequencies of particular types of transformation. Some of Newport's findings are summarized in Table 1 (drawn from Newport's Tables 4, 5 and 6; please see originals for more details). These are summary data over speech to children from 12 months to 27 months. The linguistic classifications are Newport's. All differences are statistically reliable except where indicated as not significant (ns).

Newport gives few examples. For benefit of readers who have not conversed with a toddler lately, we present in Table 2 the first 43 adult utterances in the first CHILDES file for Adam (adam01.cha), when he was 2;3, and about the same number of words uttered by adults to Adam one year later, when he was 3;3 (adam26.cha). For comparison, we present in Table 3 approximately the same number of words of continuous text from an arbitrarily chosen issue of the *Wall Street Journal* (February 27, 2001, starting at top left of page A1). Since the material from the Adam files consists of one side of a conversation, a better sense of how it differs from adult-directed language can be gained from comparison with an adult conversation. Adult-to-adult speech in the Adam files is too sparse to be useful, so we have drawn from the telephone conversations of the Switchboard corpus (accessible at [www.Talkbank.org](http://www.Talkbank.org), coordinated by Brian MacWhinney and Steven Bird).<sup>2</sup> The sample in Table 4 represents the first of the two speakers in the first file of this corpus (SW2019A).<sup>3</sup>

---

2. On the assumption that children learn primarily from speech addressed to them, the interest lies in ways in which the Adam samples are more impoverished than the Switchboard sample. But Pullum and Scholz suggest (Section 4) that children pick up a lot from overheard adult conversation. If that is so, the Switchboard sample should be viewed in another light, as potential input for acquisition, unlike the *Wall Street Journal*. (In that case, of course, proponents of data-driven learning cannot place too great reliance on the wellformedness of children's input.)

3. Breaks between utterances are marked with slashes. In the Switchboard sample, material

*Table 1.* Comparison of child-directed speech (CDS) and adult-to-adult speech (AAS) from Newport 1977

	CDS	AAS
MLU, mean length of utterance (in words)	4.24	11.94
Average number of clauses per utterance	1.16	2.66
Of total utterances:		
% Fragments and stock expressions	36	28 (ns)
% Unanalyzable	4	9
% Ungrammatical	0	5
% Fully grammatical	60	58 (ns)
% Declarative	38	87
% Yes–no questions	23	8
% Wh-questions	21	1
% Imperative	18	2
% Undeformed clauses	28	45
% Minor movement (e.g., particle-movement)	7	10 (ns)
% Aux inversion	14	6
% Wh-forms and wh-movement	16	10
% Deletion	33	26 (ns)
% Complex deformations (e.g., passive)	0	4

In Table 5 we summarize some of the properties of these four snippets, calculated in accord with Newport's criteria. (We found we had a few judgment calls to make in applying the criteria, but none that would greatly alter the profiles that emerge.) Clearly the samples are far too small for any kind of statistical analysis, so these numbers are not to be taken as reliably representative of the genres in question, but they suggest trends which might persist in a more substantial study. (Despite their small size they are at least not atypical. The speech to Adam at two years is quite similar in its properties to Newport's 1500-sentence adult-to-child corpus, while the Switchboard sample is more in keeping with Newport's adult-to-adult speech.) The percentage of nouns is reasonably constant across the four samples, as Table 5 shows, and is not too far from Hudson's range of 33–42% across genres, though the 51% in the first Adam file is pushing the envelope. (Hudson's sense of 'noun' is very inclusive, but we applied it consistently across our samples.) It may be that the proportion

---

within square brackets was supplied by the transcriber, not actually uttered. In counting words and assessing grammaticality for the Switchboard sample we set aside hesitations, duplications and false starts since these are not reported in the Adam files.

Table 2. Adult-to-child speech

---

*Adam age 2;3.4 [adam01.cha in CHILDES Brown file]*

big drum./ horse./ who is that?/ those are checkers./ two checkers, yes./ play checkers?/ big horn?/ get over, Mommy./ shadow?/ are those your checkers?/ what's this?/ horse./ do you want to play with them?/ I like it./ why don't you read Shadow yourself?/that's a terribly small horse for you to ride./ why don't you look at some of the toys in the basket./ want to?/ do you want to see what I have?/what is that?/ can you put them together, Adam?/ not in your mouth./ let me put them together./ no, put floor./ you'll have to pick them up./ take two, all right?/ no, that's his pencil./ that's not Daddy, that's Colin./ I think perhaps he's going back to school./ not on the window sill, Adam./ you write on the paper./ my paper./ circus?/ sitting in Adam's chair?/ why don't you read it to me?/ dog./ that's Cathy./ horn?/ is this a horn?/ look what game is that?/ how many do you have like this?/ where are you going?/ how does a bunny-rabbit walk?/

*Adam age 3;3.0 [adam026.cha in CHILDES Brown file]*

it's a level./ it's a little game./ you can make very funny people./ shall we make a funny man with this?/ what?/ you're going to get what?/ that's a lock from a door./ d(o) you have one to put it on?/ what d(o) you have?/ tools?/ what d(o) you think it is?/ rain what?/ you used to wear this when you were a little baby?/ a paper punch./ a paper plunge to tell you when the paper comes./ what're you going to do with the crayons?/ where're you going?/ write a .../ a paper sponge .../ well, not with a screwdriver./ they're so tiny./ you can't color them. / what d(o) you mean, stick you in the feather?/ from Africa./ a what?/ it's really to measure things, Adam./ forty inches./ no, no, don't stand up with your shoes on the sofa./ what about opening your measuring stick some more?/ he's not a fish./ he's a little boy./ no, I'm not a fish./ no, I don't see any fish in here at all./ fish must swim in the lake./

---

Table 3. Written text

---

*Wall Street Journal February 27, 2001, page A1 column 1*

Last Labor Day, contractors were hard at work installing fiber-optic cable in central Dallas to feed America's craving for high-speed Internet service./ Then a drill bit struck a water main below Young Street./ With explosive force, water gushed into the streets and poured into a parking garage below a luxury apartment building, ruining two levels full of cars./ By the time the flooding over several square blocks receded, the damage topped \$4.5 million./ As telecommunications companies race to build the information highway of their visions, they are wreaking real havoc on the nation's streets./ All over America, crews are ripping out newly paved roads, clogging traffic and leaving plenty of potholes behind./ It's largely the result of the Telecommunications Act of 1996, which requires that cities give equal access to all who want into the communications business./ While cities can set a few rules— such as requiring contractors to post bond— they can't block one company from getting out its power drills this week simply because another company finished doing the same thing last week./

---

Table 4. Adult-to-adult speech

*Switchboard [sw2019-ms98-a] (telephone conversations)*  
 uh do you have a pet Randy/ a poodle miniature or uh full size/ I read somewhere that the poodle is one of the the most intelligent dogs uh a[round]- around/ oh uh-huh so you you've only known the dog wh[at]-/ how long did you say th[at]-/ oh well uh is it uh uh how old is the dog/ oh it's still just a pup/ yeah I have a uh well a mutt myself/ I call it a uh a chowpherd/ it's uh part Chow and part Shepherd and it as i understand it uh both sides of the were thoroughbreds so she's a genuine chowpherd/ she has the the color and the black t[ongue]- tongue of a Chow but uh she has the [cha[pe]-/shape] the shape of a uh uh Shepherd/ oh she weighs in at about fifty pounds so she's a medium size but she's big enough to be intimidating/ it is a fi[xed]- a fixed female by the way and right from day one she was teaching me/ she is she's the most intelligent dog I've ever seen/ course I'm little prejudiced of course/ no the first time I brought her home she was only uh was it six weeks old/

Table 5. Comparison of language samples in Tables 2–4

	Adam 2;3	Adam 3;3	Switchboard	WSJ
MLU (in words)	4.09	5.31	10.19	19.57
Average number of clauses per sentence	1.26	1.35	1.58	2.88
Of total utterances:				
% Fragments & stock expressions	35	24	6	0
% Unanalyzable	0	9	12	0
% Ungrammatical	2	0	6	0
% Fully grammatical	63	67	75	100
% Declarative	30	52	83	100
% Yes/no questions	18	13	8	0
% Wh-questions	37	31	8	0
% Imperative	15	4	0	0
Percentage of Nouns	51	41	37	39

of nouns is higher in speech to 2-year olds than in most genres, because of the number of deictic sentences and verbless sentence fragments used. Still, nouns aside, the Wall Street Journal sample clearly stands apart from the Adam samples in several structural respects: the percentage of complete and grammatical sentences, the average number of clauses per sentence, and the distribution of illocutionary sentence types.

The very evident contrasts here between the child-directed speech and the written material for adults should perhaps temper the methodological optimism

of Pullum and Scholz's default assumption that "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" (Section 4). With its internal negatives translated out, this sounds less plausible. It reads: "a construction that is reliably attested in text corpora at large is bound to be accessible to infants during the language acquisition process." This is indeed the right assumption to justify reporting data from a newspaper corpus in an investigation of first language acquisition, but it is probably not true. Certainly, the illustrations presented here suggest that it would be unwise to rely on it. Of course they do not show that any particular construction of the adult language is absent from what Adam hears. But they do reinforce commonsense in suggesting that it would be a good idea to check this on more appropriate material before drawing any conclusions that actually matter, such as whether linguistic nativism is true or false. If genres do differ significantly, then finding a learning-relevant example in the *Wall Street Journal* is not *almost* good enough; it is beside the point.

Newport et al. (1977) emphasized that child-directed speech is simple in some ways and complex in others. In accord with their findings, we see few embedded clauses in the sentences addressed to Adam, but the high proportion of imperative and interrogative speech acts means that phenomena such as subject deletion, auxiliary inversion and *wh*-movement are well-represented.<sup>4</sup> Newport et al.'s purpose was to assess the value of "Motherese" as a "simplified teaching language . . . presenting easy examples to little minds." Its transformational complexity was therefore an important finding. However, for purposes of studying the possibility of data-based learning, the paucity of subordinate clauses takes on greater significance. It entails that children at this age (roughly, in their second year) are not exposed to a high density of evidence concerning a host of syntactic phenomena in the adult language which they will one day have mastered. These include relative clause gaps (empty or filled with resumptive pronouns), the complements of subordinating adverbs (*after/despite the storm, after/\*despite it rained*, etc.), the legitimacy of long-distance *wh*-movement (acceptable in English, mostly not in Russian), extraposition of complement clauses (*That we're lost is obvious/\*seems, It is obvious/seems that we're lost*), the patterns of empty categories in purpose clauses or various infinitival complements (*He brought the chair in to stand on (it), The chair is not strong*

---

4. Multi-clause questions with auxiliary inversion, which are relevant to the fourth case study that Pullum and Scholz discuss (see section 4.2 below), lie athwart this classification of sentence complexity (number of clauses versus deformations of canonical clause structure), making it difficult to anticipate how well-represented they are likely to be in this age range. Specific data will be needed if this construction is to be brought to bear on questions of data-driven learning; see Legate and Yang's contribution to this volume.

*enough to stand on (\*it)/to support you, etc.*), relative pronoun deletion, and so on. We're not suggesting that 2-year olds know all of these things without having been exposed to them (though this is logically possible even for language-specific facts if parameters controlling complex sentences can be triggered by properties of simple sentences; see Lightfoot 1989). We cite the low rate of subordinate clauses for beginning learners in order to point up the fact that for evaluating the feasibility of data-driven learning, it is not good enough even to examine "language specifically addressed to or known to be encountered by children," as proposed by Pullum and Scholz (Section 4). The age of the children addressed matters too: a 2-year old lives in a different linguistic world than a 10-year old.

## 2.2. *Establishing cause and effect*

The factual issue to be settled is whether every child who has knowledge of a certain sentence type was exposed to exemplars of it before arriving at that knowledge. The total database for learning the language (say, all input from birth to 6 or 7 years – or to 10 or 11 years) is obviously irrelevant to this question. Since causes precede their effects, there is an end-point to the time interval during which a learner's input could influence a learner's knowledge. The end-point in each case is to be estimated by empirical studies like that of [Crain and Nakayama \(1987\)](#), who investigated the ability of children at different ages to perform subject-auxiliary inversion in two-clause sentences (see section 4.2 below for discussion). The age of mastery sets a bound on the corpora that it is pertinent to search. There would be no interest in showing, for instance, that auxiliary inversion in two-clause sentences is present in speech to children over 5 years, if 5 year olds already know how to do auxiliary inversion in two-clause sentences. None of this should need to be said. But it does need to be, because Pullum and Scholz (Section 4.4) cite as a relevant source of such sentences a children's encyclopedia which Sampson (1989: 223) describes as "aimed ... at the eight-to-twelve-year-old market". (In Sampson 1999: 42, this is amended without comment to "6- to 10-year-old", which would bring it a little closer to the 4;7–5;11 age range of the successful subjects in Crain and Nakayama's experiments, though perhaps not sufficiently so.) Only three of Pullum and Scholz's many examples of subject-auxiliary inversion are directed to a child documented as within the relevant age range. (Nina was 2;0.10 in the file *nina05.cha*.)

Equally, corpus studies both for and against data-driven learning must respect a lower age bound on the search for relevant exemplars. It is assumed in most current models of learning (other than very superficial parameter triggering) that children alter their grammars only in response to sentences whose

meaning they can grasp with the aid of the context, and whose structure they are capable of parsing (including, perhaps, the part that is novel and needs to be learned; for references to learning models, see section 2.4 below). For learning purposes, a sentence addressed to a child who is not yet capable of processing it might just as well not have occurred. Thus there is a tightly bracketed time interval (different from one language fact to another, and no doubt from child to child) which is the window of opportunity for input data to have a causal impact on acquisition. This might make little difference to language acquisition theory if input were homogeneous over time. But together with the indications that the linguistic make-up of speech to children changes as they grow older, it means that checking the right corpora is important.

However, even spotting instances of the relevant construction *in the relevant time range* may not securely demonstrate data-driven learning. There are at least two reasons for this, which may be noted as methodological warnings. First, there is a growing body of indications that one-exemplar learning, as in the concept of *triggering*, is incorrect. Rather, evidence for a construction may need to build up before learners act on it. Van Kampen (1997) has shown that learners may ignore substantial evidence for parameter setting for some time before adopting the right setting. Lightfoot (1998) has argued that language change occurs when a parametric trigger is present in learners' input at too low a frequency. Pinker (1989) has suggested that UG defines markedness scales such that learners demand a considerable number of exemplars before admitting a marked phenomenon into the grammar. What this weight-of-evidence approach implies for the mechanism for setting parameters is a question to be addressed elsewhere. Here, its relevance is that if examples are to contribute to learning, they may have to occur at a frequency robust enough to ensure not just that every child will encounter one in a suitable time frame, but that every child will encounter enough of them to build up the evidence to the threshold level for adopting that construction. If this is right, it means that corpus-based arguments concerning data-driven learning need to motivate a start-time, at which the learner's counter starts registering instances, as well as a density at which the exemplars must keep flowing in. How to establish these critical values isn't clear. We agree with Pullum and Scholz that it is difficult at present, in absence of a specific theory of data-use in learning, to decide when enough input is enough.

The second concern cuts much deeper into the corpus-search program for defending data-driven learning. Children may be exposed to a linguistic construction shortly before first uttering it even if exposure to it is not necessary for knowing it and using it correctly. *Post hoc* is not *propter hoc*. This may seem a tediously obvious point; it could be said of any attempt to establish cause and effect. But in the case of language acquisition there are particular reasons for expecting exposure to precede production even when the former is

not causally necessary for the latter. A more advanced syntactic construction may presuppose greater conceptual sophistication, e.g., to formulate the sort of message whose expression calls for a relative clause. Consider a child who, for the first time, wishes to express such a message; and imagine for the sake of argument that he does project the correct relative clause sentence on the basis of experience of some other and simpler constructions, with the aid of UG (see section 3 below). (Fortunately it is not germane how we might come to know this fact about him, or whether this is a realistic example.) To make the case for POS, it would have to be shown that no-one had yet spoken a relative clause sentence to (or: in the hearing of) this child. But of course that's quite improbable. It requires that the child is ready to express a message more sophisticated than anyone has ever felt moved to express to him.<sup>5</sup> Note that this point is completely general, not limited to constructions that are especially frequent or infrequent. It is simply a matter of the *relative* likelihood of utterance of the construction by children versus adults. If (approximately) the propositions entertained by children are a proper subset of those entertained by adults, then it is highly likely that an adult will say it first – whatever it is. In other words, it is to be expected that exposure will precede production even in cases (if there are such) in which the exposure is not the source of the child's knowledge. It follows that even the clearest possible demonstration of exposure, in exactly the appropriate time period, would add no weight to the argument for a causal role for input. This suggests (a) that Chomsky's claims of delayed or non-existent exposure are likely to prove false in many cases, and (b) that their falsehood would not erode the case for nativism.

We have no remedies to offer for these methodological indeterminacies. They would not be put to rest by even the most complete record of every sentence a child says or hears from birth until learning is complete. Experimental studies, though themselves not fully decisive, may be essential to supplement corpus search. Elicitation experiments like those of Crain and Nakayama, or more recently by Thornton (1990) and others, may push down the age at which children's knowledge is demonstrated, by inducing them to express propositions that might not arise in everyday conversation. Perhaps some elicited constructions are rare enough that children will not have heard them from adults<sup>6</sup>.

---

5. Parents who are very attentive to their child's linguistic and conceptual development may try to limit the locutions they use to what (they believe) the child is ready to cope with. (See Newport et al. 1977 for discussion of whether parents do adapt their language to the developmental stage of the child.) Conceivably, in an extreme case, the parents might withhold all relative clauses until they hear one spoken by the child. Correct performance by this child would be splendid evidence for POS. But we must hope that such situations are not the only means of settling the issue.

6. See also Crain and Pietroski's article in this special issue: Children's elicited sentences may reveal the influence of UG constraints in constructions which could not have been modelled

Thus, a two-pronged approach combining corpora and experimental studies may be more informative than either one alone.

### 2.3. *Misleading input*

Language input to learners is defective in several ways. Not only may it lack relevant information of various kinds; it also contains misleading material. It is standard to cite the finding of Newport (1977) that adult speech to children is almost perfectly grammatical<sup>7</sup>. As summarized by Newport et al. (1977: 121): “the speech of mothers to children is unswervingly wellformed.” But though this is important, it is not the complete story. It probably underestimates the degeneracy of children’s language samples, in two ways.

First, potentially misleading information may include misparsings by children of perfectly grammatical adult utterances. If what the child *thinks* the adult said is ill-formed in the target language, that is just as bad as if the adult had *uttered* an ill-formed sentence. Since misperceptions are not overt events, they are obviously more difficult for researchers to track than overt speech errors by adults are. As a result, their incidence is not known. (Morphological misparsings are easier to detect and have sometimes been noted. We have first-hand experience of *weather report* misparsed as *weathery port*, as revealed in extension to *\*weathery man*; Pinker (1994) notes *behave* misparsed as *be + Adjective*, with the retort *I am heyyv!*.)

Second, Newport’s finding of well-formedness reflects only conversational speech by mothers to their children. But Pullum and Scholz quote sentences from children’s literature. The fairy tales, poems and nursery rhymes that are read aloud to (some) children contain archaic turns of phrase which could lead to some quite bizarre child language if taken as a basis for learning.

Consider William Blake’s poem *Tiger*, which Pullum and Scholz cite, following Sampson (1989). Sampson notes that it contains sentence (1), a well-formed two-clause fronted-auxiliary question, which could help to pin down the correct formulation of the auxiliary inversion rule. However, it also has (2) in the immediately preceding line, with object topicalization in an infinitival clause, which is totally ungrammatical in present-day English. And the second verse of the poem begins with (3), a *wh*-question in which the main verb has been fronted, without *do*-support – not a good model for learning about interrogative inversion in modern English.<sup>8</sup> (We follow here the orthographic conventions and poem title used in Sampson’s original discussion, 1989.)

---

on adult input because they are not part of the target language at all.

7. See Table 1 above. Newport’s analysis disregarded unanalyzable speech, and classified acceptable fragments as wellformed.

8. We find that Smith (in press) also notes example (3) in a commentary on Sampson’s argument.

- (1) *Did he who made the Lamb make thee?*
- (2) *Did he smile his work to see?*
- (3) *In what distant deeps or skies burnt the fire of thine eyes?*

Logic would seem to demand that if children in the 21st century don't grow up saying *thee*, topicalizing in infinitives, and inverting main verbs, then they also don't learn how to do auxiliary inversion on the basis of this poem.<sup>9</sup>

How do they know to exclude it from their database? This is something of a mystery. It suggests that children are learning about genres as they are learning about the language. Our colleague Barbara Bevington has suggested (p.c.) that children apply the principle "Never learn from anything that rhymes". This seems a wise strategy for several reasons. But whatever the means by which children succeed in filtering their input (for discussion see Grodzinsky 1989; Fodor 1994; Kapur 1994), it seems clear in general that for data-driven learning, either the style or the social context or the frequency must differ between input items that children do rely on and items which they provably reject as a basis for learning. Rare but acceptable phenomena (e.g., parasitic gaps<sup>10</sup>) must be evidenced more robustly in one of these respects than any construction that is not in the target language. This need not be so, of course, if innate guidance is allowed, for example in the form of a markedness ranking. As we noted in Section 2.2, Pinker has discussed a kind of markedness that he calls "strength-markedness" such that "a greater number of input exemplars is needed to learn

---

9. Are we attaching too much importance to this one example? In drawing attention to sentence (1), Sampson took care to put it in perspective. He wrote (1989: 223)

Of course, I do not suggest that experience of this particular sentence is actually decisive for adoption of [the correct hypothesis about auxiliary inversion] in the biography of the typical English child, but the rapidity with which I was able to dredge the example up from my far from voluminous mental store of quotations suggests that examples of the relevant kind are unlikely to be as rare as Chomsky suggests.

But note that the logic of the dredging-up argument applies equally to *thee* and *he who* in the very same sentence, not to mention (2) and (3) in the same poem. Thus there is a dilemma here for proponents of data-driven learning, even if (1) is regarded as only *typical* of what children hear and regardless of the *frequency* of such examples. The dilemma is: Either children's literature is off limits for normal language acquisition and shouldn't be cited as counterevidence to stimulus poverty claims, or else it must be explained how children could learn the right things from it without learning wrong things too.

10. Parasitic gaps would make a fifth case proposed in the literature, of language knowledge inadequately supported by positive evidence. Chomsky wrote: "the properties of gaps are intrinsically significant in that the language learner can confront little direct evidence bearing on them, so that it is reasonable to assume that they reflect deeper principles of UG ..." (1982: 19), and therefore "if languages appear to differ with respect to the existence or properties of parasitic gaps, these differences must be completely explained on the basis of other structural differences among the languages in question" (1982: 75).

a marked form to a given level of strength than is needed for an unmarked form” (1989: 117) (see also Fodor 1998). Hence a learner could reject an ungrammatical form in the input while acquiring an equally rare grammatical form, if the former is tagged by UG as linguistically less natural than the latter.

To summarize: Data-driven acquisition requires learners to engage in some rather subtle sorting of the language phenomena they are exposed to, in order to determine which are veridical and safe to adopt. This increases the difficulty of the learner’s task, and it needs to be taken into account in assessing the information value of typical learners’ input. Methodologically, the moral is clear: in searching for counterevidence to the claim that positive exemplars of a grammatical phenomenon are missing or rare in children’s experience, it is necessary to make comparison with a baseline frequency of occurrence for any ungrammatical forms which also occur and which learners have no independent grounds for excluding. Three, or even 300, examples of correct auxiliary inversion would count as equivalent to zero for a learner who encounters them intermingled with the same number of examples of incorrect auxiliary inversion. If innate linguistic knowledge exists, it may provide some guidance in the data sorting task, such as by requiring more environmental support for some kinds of phenomena than others. In that case, the extent of UG is underestimated by the tidy subtraction: UG contains everything that is in the adult language but not in learners’ input. Some aspects of UG may serve to limit the damage that nonveridical input could otherwise do.

#### 2.4. *Ambiguous input*

Recent work in the formal modelling of language acquisition has taken on the difficult problem of cross-grammar ambiguity: the fact that a sentence (in the sense of a string of words) may be compatible with more than one grammar. Gibson and Wexler (1994) gave the simple illustration of subject-verb-object (SVO) sentences, which occur in English and German and Swedish, though the grammars of these languages differ with respect to the underlying order of verb and object and the presence or absence of verb-second transformations. (In nontransformational theories these grammars also differ, though in different ways.) In some cases, cross-grammar ambiguity may be mitigated by prosody (which can provide cues to surface syntactic phrasing) and/or meaning (which can suggest which phrase is the subject, or which items belong to the same clause, etc.). But the example of SVO word order, though very simple, is of interest because it is *not* reliably disambiguated by either of these means.

Most current learning models assume that grammar acquisition is incremental: the child doesn’t accumulate a collection of sentences from which to infer the grammar, but hypothesizes a grammar after each sentence. Like other

assumptions in computational modelling, this may be too extreme. But if anything like it is true, then it can be a real handicap for a learner to encounter a sentence and not be able to tell what grammar rules (or principles and parameters) it was generated by. Gibson and Wexler's SVO example makes the point that indeterminacy of grammar choice is not just a theoretical possibility but a real-life occurrence. Since Clark (1988) drew attention to this as a problem for parameter setting, examples have multiplied. In all such cases, a positive instance of a target language construction may appear in a learner's input and yet provide very little information about which grammar to adopt. In other words: even when positive instances occur, they may not be informative.

It might seem that there is no problem as long as other sentences in the language supply the missing information. This could be so if learners were able to recognize when a sentence is ambiguous and so discard it, waiting for more reliable input instead. But ambiguity detection is not easy; see below. Much of the recent modelling of language acquisition by parameter setting has concerned itself with the damage that can result from wrong guesses that a learner may make when the input is ambiguous but not recognized as such. Clark (1988) showed that a learner that guesses wrong about one aspect of the target language may then misinterpret the grammatical implications of subsequent input and be led even further astray. An error about case assignment could lead to an error in setting the binding theory parameters. Gibson and Wexler considered as a palliative some constraints (the Single Value Constraint and the Greediness Constraint) to channel the search for the right grammar, but they demonstrated that these constraints could result, in some cases, in a total inability to recover from a wrong hypothesis.

This problem of cross-grammar ambiguity in the positive data for learning is compounded by the fact that current models of learning are generally unable to tell, by inspection of a target sentence, even which grammars *could* have generated it. To do so would seem to demand unrealistic computational resources: it requires that every grammar be tested for compatibility with that sentence. Nor can these learning models tell, in case a sentence is completely unambiguous with respect to which grammar created it, what that unique grammar is. So the problem is clearly not just cross-grammar ambiguity. It is that these learning models find it difficult to *decode* the grammatical properties of their input. Even if a sentence is a completely unambiguous trigger for some grammar, these learning models may hypothesize an entirely different grammar that has no relation to that sentence at all. If this were true of human learners too, then clearly language acquisition would be even harder than just reading off whatever information the input provides.

Fortunately, this seems to be – at least to some extent – a problem about the models, not about real children. We believe that most current models do not do justice to the ability of human infants to extract information from their

input. Our research group has been trying to establish that decoding the grammatical implications of an input sentence is computationally possible and is not beyond the bounds of psychological feasibility even given human processing resource limitations (Sakas and Fodor 2001; Fodor 2001a). On the other hand, in order to respect reasonable resource limits on parsing, the decoding algorithm that we have developed can work only up to a point. Realistically, it can deliver only one licensing grammar per input sentence. Hence it *can* recognize the unique grammar that generates an unambiguous input, but it *cannot* report *all* licensing grammars in case there is more than one. In such cases, the most it can reasonably do is alert the learning mechanism to the fact that cross-grammar ambiguity is present, and compute one of the grammars that *might* have generated the sentence. The learner must then decide whether to set aside the ambiguous data, or to go along with the candidate grammar that is offered. Neither procedure is ideal. The former may entail long waits between usable inputs. The latter is chancy, though perhaps more stable if guided by a universal, innate ranking of grammars as provided by a theory of markedness.

We have no formal proof, but we do not believe that any humanly feasible decoding system can do better than this. So we conclude that the *effective* poverty of the stimulus is greater even than may appear at first sight. Positive evidence about the target language is not always computable by learners even when it is present in the sample. (It may be accessible to linguists, whose research methods are not resource-limited in the way that infant sentence processing routines must be. Linguists often *can* identify several different grammars that a given sentence is compatible with.) Of particular interest to the innateness debate is that even this much decoding is possible only if the number of alternative grammars is finite, or else (perhaps) if the domain of grammars is strongly ordered by an innate evaluation metric or scale of markedness. This is because the decoding routine works only if the alternatives are prefigured, ready to be drawn on by the learner's sentence parsing routines whenever needed. Thus – and this will become a recurring theme as our discussion proceeds – it appears that strong innate constraints on possible grammars are a precondition on being able to make good use of the data the input provides.

To summarize: Poverty of the stimulus is not limited to a lack of positive examples, or the presence of ungrammatical examples, or the absence of negative evidence (see Section 3 below). There is also the danger that, even though collectively the whole body of evidence may be decisive, an incremental learner may find few unambiguous examples to rely on. UG cannot resolve these cross-grammar ambiguities, since they represent choices among UG-approved alternatives. But UG might help by ranking the alternatives, or by limiting the alternatives to a finite number which it is feasible to check out on-line, so that learners have a chance of extracting as much as possible of the information that is present in the sentences they hear.

### 3. Implications of poverty, positive and negative

We turn now to the logic of the argument from stimulus poverty to the existence of innate linguistic knowledge. We suggest that Pullum and Scholz's APS does not capture its full power. It sets aside the paucity of negative evidence (evidence about what is not a sentence of the target language), though this is widely regarded as more extreme than the incompleteness of positive evidence. Pullum and Scholz do sketch very briefly an approach to solving the negative evidence problem (which we discuss in Section 3.3 below), but they isolate it from the issue of missing positive evidence, and thereby overlook the interplay between the two which is central, we believe, to the argument for innateness. Part of their reason for omitting the negative data problem was practical: to limit their investigation to a manageable size. But they offer also the justification that "the 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" (Section 2.1: 15–16). This is a curious excuse. It is like an economist saying that a study of poverty should include only people with insufficient money, not those with none at all. It is not a convincing way to divide up the topic, and it may distort the conclusions drawn.

Poverty of positive and negative evidence impact learning in different ways. But also, as we will show, they interact in a way that necessarily leaves much of the work of grammar selection to the learner. In order to discuss this we need now to divide POS, the poverty of the stimulus, into POPS and PONS. The focus of Pullum and Scholz's article is POPS, the poverty of the positive stimulus. This is more or less coextensive with their APS.<sup>11</sup> PONS, the poverty of the negative stimulus, has had the greater influence on current theories of language development, because without negative evidence it is impossible to acquire constraints. (Constraints are negative grammar statements about what does not occur in a language; this includes all filters, statements of obligatoriness, context restrictions on rules, and various other familiar descriptive devices – anything such that, if it were dropped from the grammar, the language

---

11. That APS excludes PONS is not evident in the official statement of APS (Pullum and Scholz' (3) in Section 2.2: 18), where the central premise (3c) is that "infants do in fact learn things for which they lack crucial evidence". The lack of crucial evidence here *could* include lack of negative evidence, if the "things" that are learned can include negative facts (such as the fact that *Jumped she into the water?* is not a sentence of English, or that relative clauses in Japanese may not follow their head nouns.) But that seems not to be the intention. The end of Pullum and Scholz's Section 2.1 makes it clear that negative data issues are excluded, and this policy is evident in the corpus search methodology throughout the article; see also their informal characterization in Section 2.1: 15 of what they consider "the basis for the argument that most clearly deserves to be referred to as 'the argument from the poverty of the stimulus' (henceforth 'APS')".

generated would be larger.) What can't be learned must be innate. So to the extent that negative evidence is lacking, this has seemed to call for a theory of UG in which all constraints are innate. Much work has gone into developing a theory of innate, hence universal, constraints which are nevertheless compatible with the non-universal facts of particular natural languages.

This is the fruit of PONS. For this purpose, PONS need not entail that learners receive no negative evidence at all. It is usually understood to say that learners *cannot rely on* negative evidence, that they have access at best to a small and unsystematic fraction of the negative facts that hold true of the target language. This is sufficient to rule out any acquisition model which requires every successful language learner to be exposed to many or all negative properties of the target. Pullum and Scholz don't dispute the facts about the poverty of negative data in this sense, so for space reasons we will not survey the empirical evidence, which we believe is sufficiently sturdy to need no further advocacy. We refer readers to Marcus (1993) and references there. Issues of *indirect* negative evidence are discussed in part in Section 4.2 below, and also in Fodor and Crain (1987).

We will argue that POPS and PONS conspire to create a degree of underdetermination for learning which is qualitatively more severe than if either occurred alone, and which cannot be overcome except by means of principles internal to the learner, which shape the generalizations that are formulated about the language. This argument rests on the fact that for virtually every syntactic generalization that holds of an adult language, there is a space between a lower bound that is set by learners' (incomplete) positive data, and an upper bound that is set by learners' (incomplete) negative data. Inside that no-man's land, data-driven learning is impossible; hence the learner's language faculty must supply the hypotheses. This information-gap will shrink as more positive evidence is accumulated. But we will show in Section 4.2 that if PONS is true, surprisingly *little* is gained by adding more positive exemplars.

### 3.1. Generalizations are not capped by positive data

In Section 1 we observed that POPS must be true, to some degree, as long as children are exposed to a finite number of sentences and project an infinite language. This is a very familiar point, but it has important consequences. It entails that children create sentence *structures* they have not heard. That is, when they project new sentences on the basis of positive exemplars, they do not limit themselves to slotting new lexical items into the same tree structure. They also expand the structure, by inserting additional clauses, or substituting a phrase where the exemplar had a single word. That learners do this is no surprise if UG exists, and if it makes available only a limited set of rules, including recur-

sive rules for some structures. Suppose, for example, that only a recursive rule (or set of rules) exists for embedding a relative clause. Once a relative clause (RC) has been observed and the rule for it has been put into the grammar, that will automatically generalize far beyond what was observed. It will do so in a way that is unlimited in scope and yet quite specific (e.g., the learner doesn't hypothesize free word order on observing RC word order different from main clause order). The same is true if UG offers principles and parameters rather than rules, and RC structure is not recorded as such but results from the confluence of various more abstract facts that the grammar does represent. But without UG of some form or other, it is quite unclear how all children would reliably and uniformly project relative clauses in just this way and no other.

Even more interestingly, there are properties of root clauses which children do not project onto non-root clauses, and vice versa. For example, by the time they start producing subordinate clauses, German learners know that the finite verb is final, not in second position as it is in root clauses (Clahsen and Smolka 1986). So it seems that though children generalize from 1-embedded contexts to *n*-embedded contexts, there are at least some phenomena (e.g., verb-second) that they do not generalize from root (0-embedded) contexts to embedded contexts. This suggests that UG insists that natural language grammars make a root/non-root distinction, while allowing all levels of embedding beneath the root to be treated alike. Could the facts be explained without invoking UG? Perhaps it's just that children are unable to keep track of multiple levels of embedding well enough to differentiate them consistently. But, though it sounds sensible, this latter explanation won't do, because learners do acquire long-distance movement, which can involve keeping track of the relation between a gap (a trace) and its antecedent in clauses separated by multiple levels of embedding.

The claim that children generalize RCs does not rest on observations of many, or any, children. It is supported by (a) adults' knowledge of the unbounded distribution of RCs, and (b) the logical impossibility of a child having heard RCs at every level of embedding in the language. This is how it differs from the four arguments from POS that Pullum and Scholz address. For those cases (e.g., subject-auxiliary inversion) it is a factual question how the sentences of the language that is acquired relate to the sentences from which it was acquired. That is: for non-recursive phenomena, the fact that learners spontaneously project beyond their input has to be established by studying learners rather than just the adult language they end up with, since it's logically possible, though it seems unlikely, that these phenomena are acquired item by item. Of course, the cases in the standard POS literature are not the only ones worth considering. There are many less dramatic cases in which it seems likely that learners have advance information on how to project from what they hear. Without empirical results in hand, we may do no more than

speculate. But we commend the plausibility of supposing, for instance, that a child who encounters a fronted *wh*-phrase doesn't wait to observe fronting for every combination of singular and plural, animate and inanimate, masculine and feminine *wh*-phrases. Instead, learners may be confident that the fact that *which actress* contains a *wh*-operator is relevant to its frontability but the fact that it has a feminine head noun is not. Careful empirical testing may show that learners do not take advantage of these reliable facts (though why, if so, would they persist in natural languages?), but these are among the many cases in which learners plausibly do project beyond their data. They contrast with cases where it seems clear that learners do not – they know somehow that here they have to stick close to what they observe. For instance, there are widespread restrictions on the grammatical roles of phrases that can be fronted in RCs; and many languages limit the bridge verbs across which long-distance movement is permitted. This suggests that learners limit *wh*-movement by the case or theta role of the *wh*-phrase, and by the identity of an intervening verb, until or unless they encounter further positive data that motivate a broader generalization. There is indeed some evidence of conservative learning in this regard; see Keenan and Comrie (1977).

In short, children do generalize; they do not limit their language to the sentences they have actually heard. But they don't generalize indiscriminately. So for each sentence they do hear, we want to know *how* a learner generalizes beyond it. What change is made to the grammar? What additional sentences are introduced into the learner's language? These questions would arise even if children, like linguists, stored all their input and contemplated it as a whole, but they are even more pressing for incremental learning. Whatever the answers, we know they are fairly constant across learners. The shape and the extent of the patterns that learners project from their positive exemplars are not random; they are replicated reliably (with only minor occasional slippages) by generation after generation of children. But *they do not follow from the exemplars themselves*. The exemplars must be included in the language; but other than that, the shape of the correct generalization is left wide open. In this respect, positive evidence vastly underdetermines almost everything learners end up knowing.

### 3.2. *Generalizations are not capped by negative data*

So far we have observed that the top limit on a linguistic generalization is not set by positive acquisition data: learners do not hypothesize syntactic generalizations which subtend all and only instances they have heard. This wouldn't be problematic (it wouldn't call for innate assistance) if the top limits on generalizations were set by negative data. Acquiring a language by noting what

sentences it contains is only one method, after all. In principle, it would work equally well for learners to suppose that every word string (over some vocabulary) is a sentence of the language, except those that are excluded by constraints in the grammar, learnable from negative evidence. Note that on this approach, learners would require no positive exemplars at all. So: if there were an unlimited supply of negative evidence for learners, POPS/APS would be completely unimportant.

In fact, the top limit on learners' generalizations is almost certainly *not* set by negative data, since it is generally agreed that there is far too little of that, i.e., too little to set a cap on every potential generalization a learner might contemplate. Parents do not typically volunteer information about where a generalization must be reined in, and children do not utter enough ungrammatical sentences for any feedback they may receive to eliminate all wrong generalizations. This is just PONS, and we are presupposing its truth for present purposes. When combined with POPS, it entails that learners' generalizations are not dictated by the totality of their input. Somewhere between what positive data attests to, and what negative data (if any) rules out, lies the correct grammar hypothesis; but nothing tells the learner just where.

Together, therefore, POPS and PONS force learners to rely on their own inner resources in finding the right generalizations to hypothesize. This is what we have always thought of as the argument from the poverty of the stimulus, and it still seems irresistible. The question is then not *whether* there is UG but only how much of it there is and exactly what it consists in.

### 3.3. *How does UG compensate for POS?*

Several more brushstrokes are needed to give this argument from POS its proper place in a step by step account of how language acquisition proceeds. In this section, we consider how innate knowledge could interact on-line with learners' input to shape their grammar hypotheses. In Section 4, we will examine the question of whether the inner guidance which compensates for stimulus poverty is truly linguistic.

In our own work on modeling syntax acquisition, we have found it necessary to shake off a common picture of adult language knowledge as consisting of a set of innate facts and a set of acquired facts. Rather, what is innate helps learners to acquire what needs to be acquired. Before learners can derive grammars from the sentences they hear, they have to work out what sentences they are hearing, i.e., what the structures are: the phrases, the movement chains, the licensing relations among the elements. If a sentence is mentally represented with the wrong structure, a wrong grammar is likely to result. In fact, once sentences have been structurally represented, a large part of formulating gen-

eralizations about them is already done. The representation determines what sentence type the input string is an instance of (i.e., what rules or principles it reflects), and hence what counts as another instance of the same sentence type. Thus UG could guide grammar hypotheses if it were to provide a restrictive format for representing sentences, defining the possible phrase structures, chains, bindings, null categories, argument structures, and syntactic features that sentences may or must exhibit. A simple example: With [+root] and [−root] as UG-defined features, but not [1-embedded], [2-embedded] and so forth, learners would be able (perhaps forced) to acquire root phenomena and non-root phenomena separately, but would be unable to acquire different properties for non-root clauses at different depths of embedding (as noted in Section 3.1 above).

Note that if this is how UG works, discovering generalizations is not the goal of the learning mechanism (certainly not an active goal of the child). No attempt need be made to detect what's in common across exemplars. The child's goal can be simply to understand what people are saying, i.e., to process the sentences in the environment, novel ones as well as familiar ones. Extension to new cases is the fall-out from this immediate practical goal. In processing a novel construction, the learning mechanism uses some of the representational resources made available by UG, and in consequence some more new sentences enter the language along with it, while others do not.

This is only one account of how UG could engage with learners' input to mould the resulting grammar, though it is one that we think is on the right track. It accords UG a crucial role, but it doesn't say that input is unnecessary or that learning doesn't occur. It also doesn't slice adult knowledge neatly into innate facts and learned facts. Instead, it puts UG to work to help with the enormous amount of data-based learning that evidently does occur.<sup>12</sup> Interestingly, Pullum and Scholz take essentially this tack themselves in arguing that sequences of three auxiliary verbs can be acquired without exposure to any instances.

In their second case study, they report the proposal by [Kimball \(1973\)](#) (following Chomsky 1965: 44) that a learner who has heard sequences of two auxiliaries as in (4) and (5) can generalize to 3-auxiliary examples as in (6), *if* the learner represents the 2-auxiliary sentences as economically as possible in the format of context free phrase structure schemata with parentheses to indicate optional items in the expansion.

---

12. We are firm believers in the vast amount of data processing that even UG-guided learning involves; see Fodor (1994, 2001b). Many aspects of 'non-core' constructions, especially, couldn't possibly be anticipated by learners. But this doesn't entail, or even suggest, that language learning must be *totally* data-driven. To the contrary, the more massive the data processing problem, the more welcome innate assistance would be.

- (4) *It may have rained.*
- (5) *It has been raining.*
- (6) *It may have been raining.*

In that format, it was claimed, the 2-auxiliary sentences *constitute* evidence for the grammaticality of certain (but not just any) 3-auxiliary sequences. Pullum and Scholz (Section 4.2) propose a solution of the same logical type as Kimball's, though clothed in a different linguistic vocabulary: it is suggested that a number of individual facts are acquired from simpler sentences, which then automatically combine in such a way that they license the full 3-auxiliary sequence. The learner's grammar records the facts that "... *may* occur[s] 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*" (Section 4.2: 31). Let us suppose for now that this way of capturing the facts is linguistically correct. What matters for present purposes is how it is supposed to come about. There is no reason to think that this is how all learners would choose to represent what they heard *unless* they were guided by some innate knowledge of what grammars should be like. Pullum and Scholz recognize that their account of the learning of auxiliary sequences presupposes that "phrases can be classified into types and heads can be identified as selecting certain types" (p. 31). But in fact, if the adult grammar is to be attained, and overgeneralization avoided, then phrases and heads *must* be so classified by learners. And the particular types of phrases and heads employed in the classification matters too; if learners didn't also have guidance on that, they could misgenerate all sorts of other quite unacceptable verbal sequences. Of course children can't know without experience what sort of VP is selected by the English verb *have*. But they *could* know in advance of experience, and it is hard to see how they could cope if they didn't, that [+/-finite] and [+/-participle] are relevant features for classifying the VPs that heads select, but other properties (such as [+/-transitive]) are not. What is presented in the literature as data-driven learning may thus embody in fact a significant amount of innate linguistic knowledge.

In their brief discussion of absent negative evidence Pullum and Scholz provide more examples of this. They regard PONS as not a serious problem as long as learners engage in "piecemeal learning of constructions" rather than "rapid generalizations" which may later have to be pruned back (Section 2.1: 16). Of course this is true, in a general programmatic sense; it has been a cliché for years that conservative learning is the solution to PONS. This is why the Subset Principle has occupied a central position in learnability research (see Section 4.1 below). But Pullum and Scholz overlook the consequences of this for the agnostic stance they wish to take on innate knowledge. They might have

asked what it is that holds learners to this conservative policy of piecemeal learning. It is clearly not the input itself. In Section 4, we consider the possibility that conservatism is the consequence of domain-general (non-linguistic) principles of induction; if so, our own proposal that it is UG which restrains the boldness of learners wouldn't go through. But before turning to that, let's consider here how UG might serve as the enforcer of the conservative piecemeal learning that Pullum and Scholz need to postulate. As it happens, Pullum and Scholz themselves propose UG solutions to both cases of PONS that they mention, one concerning auxiliary reduction from Lightfoot (1998), and one concerning the *that*-trace effect from Haegeman (1994).

In response to Lightfoot, they propose that learners avoid overgeneralizing auxiliary reduction as in (7) to VP ellipsis contexts such as (8), because "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" (Section 2.1: 17).

(7) *Jim's happy.*

(8) \**Kim's taller than Jim's.*

Clearly, these facts about clitic forms and weak stress and lexical heads would be powerless unless learners had knowledge of them. And since it seems truly unlikely that they could induce these generalizations prior to analyzing a particular instance in the target language, it may be concluded that this knowledge is innate. So this is a beautiful example of the argument from PONS to the existence of innate knowledge. Indeed, if this analysis is correct, it adds support to the idea that the innate knowledge that learners possess must be specifically linguistic, i.e., UG. The particular pattern of conservative generalization that learners exhibit here, if Pullum and Scholz are right, is exquisitely refined and apparently quite peculiar to language.

This account of the acquisition of auxiliary reduction illustrates just how much of the explanatory work UG must do. So does the account that Pullum and Scholz propose of how the *that*-trace effect (which rules out examples like (9)) could be learned without negative data.

(9) \**Who did they think that was available?*

They propose (in essence though not in these words) that learners acquire selection restrictions which limit the complements taken by subordinating words like *that* and *than*. Pullum and Scholz suggest that "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." (Section 2.1: 16). What is doing the

work here is the learner's presumed knowledge (which could only be innate or derived from something that is) that this kind of fact *must* be recorded in the lexical entry for the subordinating word, i.e., separately for *that* and for *than*. It must not be recorded as a general syntactic fact, nor must it be recorded under a categorial description such as *Complementizer* or *Subordinator* or *th-word* (like *wh-word*). If it were, significant overgeneralization would result.

On the other hand, learners presumably don't write *every* property of a sentence they hear into the lexical entries for all the words the sentence contains. They know, it seems, that at least some of the facts about a phrase containing the word *cat* don't need to be stored in the lexicon as facts about *cat*, but can be recorded under the generic category Noun. If they didn't, it would take a very long time indeed to acquire the general patterns of distribution of nouns in the language. For words like *than*, learners presumably do not record in its lexical entry such co-occurrence facts as that the finite VP that follows it contains a transitive or an intransitive verb. Somehow they know that that's irrelevant; it is a point that can be generalized over, and does not need to be attended to here (though it must be elsewhere). Other properties of this construction do matter, and learners seem to be forewarned to take note of them. For example, Pulum and Scholz's analysis presupposes that children know to retain the feature [+finite] on the complement of *than*, as in (10a). If they were to omit it, as in (10b), they could misgenerate (10c) on the basis of hearing (10d).

- (10) a. *more than* + finite VP  
 b. *more than* + VP  
 c. \**They wanted more food than to be available.*  
 d. *They wanted more food than was available.*

Also, learners don't coalesce the co-occurrence patterns for extraction and non-extraction contexts. If they did, the sequence *think* + finite VP (11a) that is attested in questions such as (11b) would be mis-learned as generally acceptable, creating errors such as (11c).

- (11) a. *think* + finite VP  
 b. *Who do you think is clever?*  
 c. \**You think are clever.* (Meaning 'You think you are clever', with recoverable subordinate subject)

Discriminations such as these between the generalizable and the non-generalizable aspects of an observed sentence are what make it possible for learners to acquire some facts conservatively without slowing all of language acquisition to a crawl while every conceivable constraint on every construction is checked out. But these discriminations are not given in the learner's data. Just possibly they might be discernable if the data were all laid out simultaneously

to compare and contrast, as linguists can do. But this assumption was rejected above as unreasonable. There is no cumulative memory for input sentences; learning is incremental, i.e., grammar hypotheses are retained or updated after each input. The conclusion is thus the same as we have already seen: In order for data-driven learning to produce reliable results, it must apparently be laced with a considerable admixture of innate linguistic knowledge – in this case about the scope of selection constraints, the difference between functional and lexical categories, and which features of which words within a construction control its grammaticality.

The theme of Pullum and Scholz's article is that stimulus poverty arguments for UG may be unsound, because as yet there is no good demonstration that the stimulus for language acquisition is too impoverished to learn from without help from UG. Yet, it is beginning to look as if appeals to UG may be unavoidable. When Pullum and Scholz make positive proposals about how learners might arrive at the correct generalizations, they too apply the UG remedy, presupposing some highly specific constraints on how input information is mentally recorded. Implicitly, it seems, they recognize that without these constraints, the same input could have led to completely wrong generalizations.

#### 4. Strategies of the learning mechanism cannot compensate for POS

Our argument has been that encountering an example of a grammatical phenomenon, or even many examples, is only a small part of discovering the grammar by which it was generated. There is also the *projection problem*: what general pattern is this an exemplar of? So far we have considered how UG could constrain learners' generalizations, thereby supplying implicitly the information that no collection of exemplars can provide. However, to make the case for innate *linguistic* knowledge, it must be shown that nothing other than UG can do this. In this section, we examine the possibility that the learning routines have principles or strategies which dictate how to project from examples to other cases. If this were so, it would shift innate language knowledge out of the 'grammar component' and into the learning mechanism. This would modify the linguistic innateness hypothesis but not thereby seriously undermine it. However, if the learning strategies were so general that it became plausible to suppose that linguistic regularities can be attained by inductive processes not specialized for language, that could scupper linguistic innateness altogether. Auxiliary inversion and Subjacency would be on the same footing as "All swans are white" and there would be no need for UG.

We can't do justice to this possibility here. We take it seriously and believe it deserves more attention than it standardly receives in the linguistics literature on acquisition. Here we will consider just the Subset Principle and the Unique-

ness Principle, as prime candidates for general inductive principles that might direct the hypothesization of grammars (see Berwick 1985 for others). These principles offer guidance on how to generalize from instances, they apply extremely broadly, and they incorporate no specifically linguistic concepts (such as auxiliary or clitic or lexical head).

#### 4.1. *The Subset Principle*

The *Subset Principle* (e.g., Berwick 1985; Manzini and Wexler 1987) requires (approximately) that learners select a grammar that generates a less inclusive language over one that generates a more inclusive language, if both are compatible with the input. The Subset Principle has seemed to be a necessary condition on successful learning in the absence of negative evidence: overgeneral hypotheses which subsume the target language must be avoided, because if adopted they would be uncorrectable without negative data to motivate retreat. Now we must ask: How much of the work that might be attributed to UG can be done by the Subset Principle alone? We believe it is very little. The Subset Principle must work *with* UG if it is to work at all. However indispensable it may be, the Subset Principle is demonstrably *false* of human language acquisition if there is no UG.

Without UG, the Subset Principle predicts that each new sentence a learner hears will be added individually to the currently hypothesized language, without being generalized in any way. That is what would yield the least inclusive language, the smallest subset of sentences compatible with the input. But this is not what children do. If they did, they would acquire finite languages, and as adults they would be unable to comprehend or produce any sentences they hadn't encountered during the learning stage. So the Subset Principle taken literally is false. On the other hand, it's hard to deny that the Subset Principle is a necessary antidote to PONS, given the impossibility of motivating retreat from overgeneralizations without negative data (unless the Uniqueness Principle can do this; see Section 4.2 below). Problem: How can the necessity for the Subset Principle be reconciled with the fact that general learning would be paralyzed if it were literally applied?<sup>13</sup>

---

13. Sampson (2001) has suggested that a simplicity measure would disfavor the adoption of very particular grammars and thus could offset the limiting influence of the Subset Principle. We cannot take up this idea here, but we note that simplicity metrics have a bad history of biting the hand that applies them. As long as there is a positive correlation between grammar simplicity and generality (size of language generated), a simplicity criterion would have to be very carefully tailored indeed to avoid its overwhelming the Subset Principle and selecting incurably overgenerating grammars. See [Fodor and Crain \(1987\)](#), Fodor (2001b) for discussion.

The standard answer is that some imaginable languages have grammars that are never considered by human learners. The idea is that it would be impossible in many cases to add *just* the current input sentence to the language, without adding other sentences as well, because those other sentences are included in the smallest language which is compatible with the input *and which has a grammar accessible to the human language faculty*. The language projected by the learner as time goes by is pictured as swelling outward, not in tiny increments of one sentence at a time, but in great leaps and bounds, with only a few possible stepping stones between the initial state and the final grammar. Chomsky (1965 and since) has advocated this as the most effective way to solve problems of grammar selection: enrich UG so that it radically limits the number of intervening grammars that need to be considered and eliminated by learners en route to the target grammar. Perhaps there are comparable limits on the set of humanly-accessible hypotheses about swans and other non-linguistic matters. But even so, it seems clear (in view of facts such as in Section 3.3 above) that the particular hypotheses along the route from least to most inclusive are domain-specific: the leaps that learners make are peculiar to language in the one case, and to swans or perhaps to biological kinds in the other. If that is correct, then it follows that the only way the Subset Principle can be effective without being crippling is in conjunction with innate constraints on the class of possible grammars.

#### 4.2. *The Uniqueness Principle*

The *Uniqueness Principle* (e.g., [Wexler and Culicover 1980](#); [Pinker 1984](#); [Berwick 1985](#)) comes in several forms, but we'll construe it here as imposing a default: that there is only one right way to express any given proposition. This can create useful negative evidence out of positive data. Hearing an adult express a proposition in one form of words would instantly notify a learner that a great many other potential ways of saying it are wrong. (Because this is only a default, it can be overridden in those cases where a language does have alternating forms.) We maintained above that positive evidence cannot do more than set a lower limit on the extent of any linguistic pattern a child should project. But together with the Uniqueness Principle it could set a top limit too. Learners could afford to generalize unconservatively, even irresponsibly, and the Uniqueness Principle with positive data could allow them to retreat to the correct grammar where necessary.<sup>14</sup>

---

14. This is sometimes referred to as indirect negative evidence. Another possible source of indirect negative evidence that is sometimes cited (e.g., Chomsky 1981, Ch.1) is the realization by the learner that some construction that is licensed by the current grammar has not been

It is not known to what extent children do rely on the Uniqueness Principle for syntax learning.<sup>15</sup> It is not without disadvantages: it is inapplicable in some cases, and cumbersome to apply in others. It is inapplicable where there is no pre-empting form to drive out a wrong form. This is so in a wide range of cases, such as violations of island constraints on movement, as in (12), and incomplete paradigms as in (13).

- (12) \**Who did John overhear the statement that Mary had dinner with yesterday?*
- (13) a. *How good a dancer is he?*  
b. \**How good dancers are they?*

Exactly what counts as a competing form can be crucial, and it is unclear how learners could get this right without having recourse to UG. For instance, if an ungrammatical passive such as (14) has no competitor in the relevant sense, then the Uniqueness Principle cannot avert the danger of overgeneralizing the passive to *have*. If (15) with *own* is (14)'s competitor, then (14) will be blocked, but so will members of other near-synonymous passive pairs, such as *was hit* and *was struck*. If the existence of the active sentence (16) were sufficient to exclude (14), then active forms would exclude passives for all verbs. In that case the Uniqueness Principle would need to be overridden by positive data: every acceptable passive would have to be subsequently re-acquired on the basis of sufficient positive evidence for both forms. Though not impossible, the relearning of forms improperly excluded by the Uniqueness Principle would slow down the acquisition process.

- (14) \**Two houses are had by Susan.*
- (15) *Two houses are owned by Susan.*
- (16) *Susan has two houses.*

Let us apply these considerations to the case of auxiliary inversion that Pullum and Scholz discuss in their fourth case study. By itself, a one-clause exemplar of inversion such as (17) shows that inversion occurs in English, but it doesn't tell a learner a great deal about the extent of the phenomenon. It shows only (at

---

attested in the input. We believe that the mechanics of this are psychologically unfeasible, and we will not consider it further here. It is discussed in Fodor and Crain (1987).

15. Preemption principles like the Uniqueness Principle have often been called on in other domains such as phonology and morphology, and may work better there than for syntax. See, for example, Pinker (1999).

best) that it is acceptable in a question to invert a single auxiliary verb if there is one.<sup>16</sup> (Examples from Crain and Nakayama 1987.)

(17) *Is the boy asleep?*

A sentence like (17) does not reveal whether inversion also occurs in negative questions, in passive questions, or echo questions or *wh*-questions. Context restrictions such as these are not too far-fetched. For example, Rizzi (1982) noted that learners of French must discover that stylistic inversion “is only permissible (in a first approximation) when the COMP is filled by a WH pronoun or trace” (1982: 448). Sentence (17) also doesn’t reveal whether inversion is obligatory in any context, even in (17) itself. It doesn’t show whether an auxiliary can be fronted within a subordinate clause, or whether it could be moved out of one clause into another one. It doesn’t show whether, in a sentence with more than one auxiliary, both/all of them could move, or any one of them, or any two, or just the final one, and so forth. Thus, ungrammatical sentences like (18), in which the subordinate clause auxiliary has been moved to the front of the main clause, are perfectly compatible with the evidence provided by (17).

(18) *\*Should the boy who be working is asleep?*

Yet as Crain and Nakayama (1987) showed, children learning English seem to know that sentences like (18) are not good. Even during their fourth or fifth year when they still make many errors in transforming embedded questions without inversion into direct questions, they don’t make errors like (18).

Pullum and Scholz are concerned to show, contra Chomsky (1971), that children learning English encounter not only simple questions like (17) but also two-clause questions like (19) with inversion of the auxiliary of the second/main clause. (As noted in Section 2, pertinent experiences must precede mastery of the correct rule.) The implication is that hearing (19) would resolve the question of how learners could know to treat (17) as an instance of fronting the *highest* auxiliary in the tree structure, rather than as an instance of fronting the *first* auxiliary in the word string, and thus avoid potential errors like (18).

(19) *Is the boy who should be working asleep?*

---

16. Though it is standardly referred to as subject-auxiliary inversion it is not always an auxiliary that inverts. In (17), and (19) below, it is the copular *be* introducing an adjectival predicate. In Experiment 1 of Crain and Nakayama (1987), the inverting verb was the copula in four of six test examples (progressive *be* in the other two). Somehow English learners avoid inverting ‘real’ main verbs like *sleep*, though this is compatible with UG and occurs in other languages. Thus, further questions arise about how learners determine which verbs undergo inversion (see Crain and Nakayama 1987: 540 for comment). For convenience in what follows we will refer to the copula as an auxiliary.

In fact, without UG assistance an example like (19) is not all that informative. It doesn't narrow the class of viable hypotheses down to the correct one, or even down to a small set. It does not even exclude all grammars that generate the ungrammatical (18) (e.g., Move any auxiliary to the top of the main clause).

On hearing (19) a learner can eliminate only grammars which prohibit fronting of the auxiliary of the main/second clause of a sentence in which a subordinate clause precedes the matrix auxiliary.<sup>17</sup> This does eliminate the structure-independent rule that Chomsky's argument focussed on: To form a question, obligatorily front the first auxiliary in the word string. And it excludes other rules too, such as: Obligatorily move the last auxiliary in each clause to the top of the main clause; or move every auxiliary to the front of its own clause. But this still leaves lots of possibilities open. Like (17), example (19) is compatible with inversion being optional or being obligatory, and with it occurring only in non-negative interrogatives, or only in present tense sentences, and so on. Also, (19) is compatible with there being obligatory movement of the last auxiliary in the word string, or of any one or two or three auxiliaries in the sentence; and with there being a free choice between movement and copying; and with deletion of an auxiliary plus initial adjunction of *is*; and with fronting of main verbs when no auxiliary is present. Thus, a host of hypotheses compatible with (17) remain in the running even when (19) is added. Some are structure-dependent and some are not; quite a few would license the ungrammatical (18). Thus additional positive data may bring about relatively little reduction in uncertainty compared with the total number of hypotheses in the pool that a data-driven learner would be selecting from.

Are there sentence types that would be more helpful than (19), and exclude more of these hypotheses? In fact, as long as optional rules are permitted, positive data of any kind excludes hardly anything; all it demands is a rule that generates the observed instances among others. But here is where the Uniqueness Principle steps in. Derivational optionality is excluded by the Uniqueness Principle, which allows only one way to express each proposition. (The Subset Principle can also rule out optionality, but needs to collaborate with UG, as we have seen.) So under the Uniqueness Principle, inversion will be treated by learners as obligatory where it is permitted. Then any potential type

---

17. Without UG, even this broad an inference is not in fact warranted. At most, (19) tells the learner that in a question whose underlying structure is comparable to that of (19) in all relevant respects, it is permissible to invert the auxiliary that corresponds to the *is* in (19). Then the question is: what does a human learner take to be the relevant respects in which some novel sentence must match the exemplar (19), and why? We can't pursue this here. It is far too tedious to describe all the cautious inference steps that would be involved in truly conservative data-based learning. So we will cut expository corners here and take for granted that learners know innately that the relevant respects have to do with phrase structure, and constituent order, but not with consonants and vowels or the number of words in the sentence.

of inversion that observably does not occur where it might have occurred can be assumed to be ungrammatical. So now learners *can* deduce from (19) that movement of a subordinate clause auxiliary into the main clause, for instance, is prohibited. Also excluded as ungrammatical are: auxiliary movement within the subordinate clause only; movement of every auxiliary within its own clause; auxiliary copying; *is*-adjunction without any movement; and so forth. Even the Uniqueness Principle can't rule out *all* wrong hypotheses on the basis of (19), e.g., Front all the auxiliaries in the main clause, as in (20a), or front the main verb when no auxiliary is present, as in (20b). But a little more positive data such as (21a, b) will set these points straight, given Uniqueness.

- (20) a. \**Will be the boy asleep soon?*  
 b. \**Slept the boy?*
- (21) a. *Will the boy be asleep soon?*  
 b. *Did the boy sleep?*

Thus, positive examples carry a lot of information as long as the Uniqueness Principle is operative.

To summarize so far: The search for sentences like (19) in learners' input is pointless as a means of explaining how children know that (18) is ungrammatical, unless the assumption is made that something like the Uniqueness Principle creates negative data from positive examples. This is because positive data alone do not greatly limit the class of hypotheses. A new positive example nudges the lower bound of the generalization upward a little, but without establishing the total shape of the phenomenon. But if Uniqueness is operative, much more information can be extracted from positive examples – they provide both positive and negative evidence. In that case, it looks as if a data-driven approach to learning auxiliary inversion could get by very well without calling on UG. In fending off arguments for linguistic nativism, therefore, Uniqueness could be a powerful tool.

However, for reasons we will now consider, the Uniqueness Principle does not always perform well when needed, and it does not work here. In the present case, the problem is that auxiliary inversion is *optional* in English yes/no questions. Children know this. They hear inverted questions but also questions marked only by intonation, like *You're drawing one for me?* and *You were just sleeping?* (from file adam18.cha), and once they have auxiliaries, they produce some recognizably uninverted questions (e.g., *It's hot?* and *We can use all?* from file adam25.cha).<sup>18</sup> So English yes/no questions will have to be a case

---

18. We relied on the transcriber's question mark to identify questions, but the contexts in which these examples occur support their interrogative status, and also make clear that they are not echo questions.

where the Uniqueness Principle default is overridden by positive evidence of two acceptable forms: with and without inversion. But once the learner has recognized that inversion and non-inversion can co-exist, exactly the same problems arise as before the Uniqueness Principle was applied. How far should the observed facts be generalized? Which of the many potential rules is correct? If learners know that inversion is optional, then observing an auxiliary that is not inverted is no guarantee that it *cannot* be inverted in that context. So then (17) and (19) are once again compatible with a free choice of which auxiliary to invert, or how many to invert, or within which clauses inversion can occur, and so on. (Of course, the learner might decide that inversion couldn't be optional except for the highest auxiliary in the sentence. That would get things right. But nothing in the input demands this restriction, so it's no more explicable in data-driven terms than the puzzle we began with, for which Uniqueness was to be the solution.<sup>19</sup>)

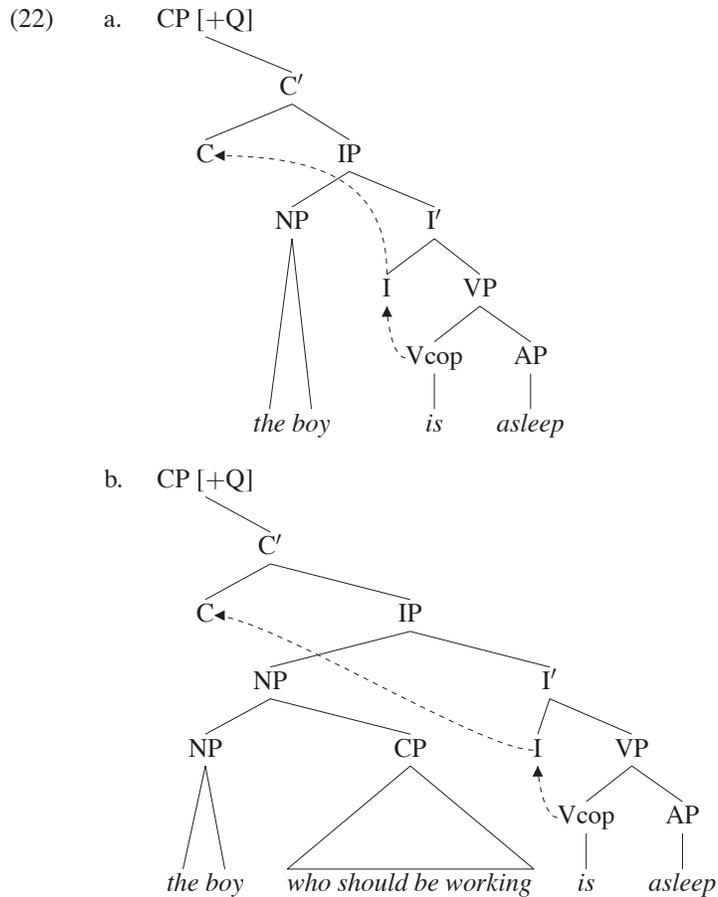
In short: The Uniqueness Principle can indeed provide some powerful information when it happens to be true, but it is only a default principle and must sometimes be overridden by learners. And in recovering from its overly severe consequences in cases where it happens to be false, the same two theoretical extremes arise as without the Uniqueness Principle: learners surely don't add legitimate sentences to their language one by one, but if they do generalize beyond the immediate datum, it's hard to see how they would know how to select among the myriad potential hypotheses without UG to guide them. We conclude, then, that the Uniqueness Principle, like the Subset Principle, can supplement a rich set of innate linguistic principles, but cannot substitute for it.

#### 4.3. *Getting auxiliary inversion right*

That conclusion is all very well as long as learners can select among the myriad potential hypotheses if they do have UG to guide them. In fact, current theories of UG rule out almost all of the incorrect generalizations we have been considering. For convenience, in (22a, b) we illustrate an analysis of (17) and (19) in the style of late Government Binding theory (Chomsky 1986a, b), using simplified tree diagrams loosely based on those in Haegeman (1994). For explanation and motivation, see Haegeman's text.

---

19. It is imaginable that the Subset Principle steps in here to limit the spread of optionality observed in main clauses, once the Uniqueness Principle has been put out of action. Complex interactions between the Subset Principle and the Uniqueness Principle have not been explored sufficiently for it to be clear how closely they could tailor a learner's generalization tendencies to the patterns that are in fact acquired. Clearly, if this approach were adopted in the present case, the need for an innate definition of the right 'grain' for the Subset Principle would arise again, as it did in Section 4.1.



Consider (22a) first, with no subordinate clause inside the subject phrase, which corresponds to the one-clause sentence (17). We've represented the copula here as the head of a VP (Verb Phrase), with an adjective phrase as its complement. If a verb moves, it must obey the rigorous constraints on head movement. The primary constraint is the *Empty Category Principle*, also known in this context as the *Head Movement Constraint*, which requires (approximately) that a head may move only from one head position to the next highest one, within the extended verbal projection (VP, IP, CP) that it originated in. The copula verb in (22a) can thus move up to the head position of the IP (Inflection Phrase) and from there it can move to the head of the CP (Complementizer Phrase). This is what we've been calling "subject-auxiliary inversion".

A child who has already acquired basic phrase structure, and who hears (17), could represent it with some confidence as in (22a). Not with complete confi-

dence, since there is an alternative structure which has the copula in the head of IP and the subject still in the VP where it originated. (Subject raising from VP is assumed but not shown in (22).) This latter structure, perhaps preferred because it has less movement, could only be ruled out by other sentences. (For example, given that negation is projected above VP, it would be ruled out by the fact that the subject and copula precede *not* in sentences like *The boy is not sleeping*.) For a learner who has established that the subject is above the VP, (22a) would be the only acceptable structure for (17). Something is needed to associate this structure with interrogative force; we've assumed here a feature (traditionally [+Q]) in the CP, which needs the C head to be filled to license it.

Now consider (22b), which includes an embedded clause as in sentence (19) above. The same verb movement operation is applicable. The feature [+Q] would be in the matrix CP, because of the semantic scope of the question. The strict locality constraint on head movement entails that only the matrix copula can move up to the head of the CP[+Q]. Auxiliaries in the subordinate clause cannot escape their own verbal projection; they are trapped beneath the subordinate CP node. Thus the interrogative form (19) is predicted, and (18) is impossible. There is no need to observe (19) or any other multi-clause question to know this. Among the many other patterns excluded by the locality constraint on head movement is movement of all the auxiliaries, or of any auxiliary other than the highest, in a matrix clause with more than one.

Chomsky (1971) proposed only that UG would direct learners toward a structure-dependent auxiliary-inversion transformation rather than a transformation defined over linear positions in the word string. But that still left many hypotheses for learners to evaluate. The richer theory of UG that has since been developed goes well beyond that; it selects the *right* structure-dependent auxiliary-inversion transformation. There is more to be said about the linguistics, of course, and more work to be done to apply it to learnability problems. For example, can learners predict that auxiliaries don't invert in embedded questions? Perhaps not, since this differs across languages (at least for subjunctives). But non-inversion in relative clauses seems to be universal so it may be guaranteed by UG. A puzzle remains about optionality. In the Minimalist Program (Chomsky 1995), no derivational operations are optional, so an explanation would be needed for the fact that a movable auxiliary can optionally remain in situ in English. As for the Uniqueness Principle: it is not needed in this framework to create negative evidence from examples like (19), since those facts follow from the way that (17) is analyzed. But there's no harm in learners' assuming Uniqueness in contexts where UG leaves alternatives open.

#### 4.4. *Summing up*

The defense of POS that we have developed here is a little different from the well-known treatments in the literature. It is less extreme than some of the claims that Pullum and Scholz take issue with in their article. It is grounded in the experience of trying to build a learning model that actually works. For this we have found UG indispensable, regardless of whether certain construction types are systematically withheld from learners. This is why we have not defended POS by maintaining, as Chomsky did, that children standardly hypothesize the correct rule for auxiliary inversion without having heard any examples like (19). We have thus downplayed the importance for nativism of the classic cases which have been claimed to demonstrate that children predict the properties of a target construction before (if ever) they experience it. But we have done so not because we doubt that children make such predictions, and not to diminish the argument from POS, but to emphasize that there is poverty aplenty to support innateness even without these cases. There are kinds of sentences that *nobody* hears, because they are too long or too difficult to process. There are ambiguous sentences. There is the lack of any distinguishing mark which separates non-sentences from good sentences that just happen not to have occurred yet. And there is the low informativeness of positive examples, even the ones that are often regarded as decisive, if they are not accompanied by substantial negative evidence of some kind – or by guidance from UG in analyzing them. Even given some experience of those sentences, learners must make a leap beyond them, and it must be the very same leap as every other learner. How that's done is what needs explaining.

### 5. Research strategies and truth

To end, we would like to distinguish between POS claims as the *outcome* of research, and POS claims as the *impetus* for research. We think POS can be defended in both roles, but the arguments are tidier when they are separated out.

To keep things simple let us suppose that there are just two broad research strategies for investigating human language acquisition, which divide on how impoverished they take the stimulus to be. These we will call A: the *innateness strategy*, and B: the *data-driven strategy*. Both are currently being pursued, and both, we believe, can yield useful results. Though their starting points and their methods differ, they are not incompatible, and their outcomes ought eventually to meet. Arguably the surest way to the truth is to try both.

A. The *innateness strategy* makes the working assumption that everything that might be innately known, is so. This includes all universal properties of

natural languages. It may also include some non-universal phenomena, to the extent that these can be argued to represent choices among innately constrained candidates (i.e., the parameters of the principles and parameters theory).

What recommends this research strategy is (a) that there are other quite different considerations which suggest that the human brain is innately prepared for language (e.g., neurological facts, deprivation facts), and (b) that it has proven very difficult to explain how language acquisition is possible even with a massive assist from innate knowledge. The idea is to assume that language development involves a mix of innate and acquired information, until we have devised a learning model that works; *then* an attempt can be made to relinquish some of the innate knowledge that was presumed along the way. Indeed, this is a pact one makes with oneself, or with the deities of scientific truth, in adopting research strategy A: that one will periodically try to strip away various aspects of UG, to see whether, after all, they are unnecessary for successful language acquisition or normal adult language use. In a sense, then, Pullum and Scholz have simply reminded us that such an exercise is overdue.

B. The *data-driven strategy* makes the opposite starting assumption: that no language knowledge is innate. It puts its ingenuity into finding ways in which all adult knowledge could be induced from the available input. (See Elman et al. 1996, Brent 1996, and references there.) The merits of this approach include (a) openness to advances in brain science and computer science which may suggest powerful and previously unknown mechanisms for learning, and (b) independence from the current state of understanding in linguistics, such as the difficulty of distinguishing true universals from what is merely common to the languages that are known.

The intellectual pact in this case is to be vigilant for any scraps of unlearned information that may have crept into the learning program. The tally should include even apparently innocuous presuppositions such as that there are words; that the same word may vary its phonological shape in different circumstances; that the order in which words are pronounced in an utterance matters; and so on.<sup>20</sup> We have seen in the discussion above how easy it is to import into what is intended to be a data-driven learner, some subtle innate guidelines as to how human languages are structured. Once these have been recognized as such, they may be eliminable. Or perhaps instead they may have to be acknowledged as essentially innate, either because they can't be learned, or because they need to be in place before a data-based learning algorithm can benefit from the information the input does contain.

---

20. These fundamental facts might be learnable, but that would need to be shown, especially since it appears that other species do not arrive at them as unerringly as human children do. See, for example, the discussion by Valian (in press) of bonobos' lack of interest in word order.

Which of these approaches to studying language development is more legitimate? Neither one of them can claim any a priori advantage that we can see. Certainly it's no use pretending that in linguistics, as in the mature sciences, new investigations are always grounded in securely established facts. That would be wonderful. In reality all that anyone can do at present is explore the consequences of certain premises and see where they lead. It's not necessary to believe the premises in order to do a good job of evaluating their merits (though perhaps it adds zest to demonstrating the value of a research framework if we are so committed that we fail when it does). There may even be interesting cross-overs between working assumptions and eventual conclusions. For example, a vigorous attempt to model purely data-based learning of language might uncover the clearest evidence yet for innate linguistic knowledge.

To evaluate POS as part of investigative strategy A is thus quite different from evaluating it as an empirical truth (see Harman 1986: 46–47). In the former role it has paid for its keep many times over. It has been the driving force for most of what has been found out about language acquisition in the last two or three decades, ranging from the work of Crain and colleagues on how much children know about their language how early in life, to mathematical theorems such as Osherson et al.'s (1984) demonstration that (on certain assumptions) if learning succeeds when there is noise in the input, it follows that there is only a finite number of human languages. We have reported briefly above on our own investigations of how innate knowledge can interact with knowledge derived from experience to solve problems of cross-grammar ambiguity brought to attention by Clark and by Gibson and Wexler. This line of work has uncovered some new facts about the capabilities of different learning algorithms, as well as enough tricky new problems to fuel research for the next decade.

However, the fact that the practical successes of a research strategy can be decoupled from the truth of its working assumptions cuts both ways. It's all very well for strategy A to be productive, but still we do want to know whether POS is true – not just generically, in order to score points in the unending debate between rationalism and empiricism, but in order to continue to craft the very best model of human language development that we can. As we get closer to that goal, the details start to matter more. Soon we are going to need to know *exactly* which of the language facts known by adults are and are not missing from the language sample that children learn from. Working out just what sorts of interactions between an infant and the environment could result in knowledge of a human language does require the kind of careful empirical study of the input that Pullum and Scholz are calling for. On the other side, there is no reason why research strategy B should hold itself back while corpus studies are being conducted. If a human language really can be learned from scratch without any innate guidance, it would be good to see a working model.

## References

- Berwick, Robert (1985). *The Acquisition of Syntactic Knowledge*. Cambridge, MA: MIT Press.
- Berwick, Robert and Partha Niyogi (1996). Learning from triggers. *Linguistic Inquiry* 27: 605–622.
- Brent, Michael (1993). From grammar to lexicon: unsupervised learning of lexical syntax. *Computational Linguistics* 19: 243–262.
- Brent, Michael (ed.) (1996). *Computational Language Acquisition*, special issue of *Cognition* 61, 1–2: 1–194.
- Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- (1971). *Problems of Knowledge and Freedom*. London: Fontana.
  - (1980). *Rules and Representations*. New York: Columbia University Press.
  - (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
  - (1982). *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.
  - (1986a). *Barriers*. Cambridge, MA: MIT Press.
  - (1986b). *Knowledge of Language*. New York: Praeger.
  - (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Clahsen, Harald and Klaus-Dirk Smolka (1986). Psycholinguistic evidence and the description of V2 in German. In *Verb Second Phenomena in Germanic Languages*, H. Haider and M. Prinzhorn (eds.), 137–167. Dordrecht: Foris.
- Clark, Robin (1988). On the relationship between the input data and parameter setting. *Proceedings of NELS* 19: 48–62.
- 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.
- Fodor, Janet Dean (1994). How to obey the subset principle: binding and locality. In *Syntactic Theory and First Language Acquisition: Cross-Linguistic Perspectives – Volume 2*, Barbara Lust, Gabriella Hermon and Jaklin Kornfilt (eds.), 429–452. Hillsdale, NJ: Lawrence Erlbaum.
- (1998). Parsing to learn. *Journal of Psycholinguistic Research* 27, 3: 339–374.
  - (2001a). Setting syntactic parameters. In *The Handbook of Contemporary Syntactic Theory*, Mark Baltin and Chris Collins (eds.), 730–767. Oxford: Blackwell.
  - (2001b). Parameters and the periphery: reflections on *Syntactic Nuts*. *Journal of Linguistics* 37: 367–392.
- Fodor, Janet Dean and Stephen Crain (1987). Simplicity and generality of rules in language acquisition. In *Mechanisms of Language Acquisition*, Brian MacWhinney (ed.), 35–56. Hillsdale, NJ: Lawrence Erlbaum.
- Gibson, Edward and Kenneth Wexler (1994). Triggers. *Linguistic Inquiry* 25: 115–126.
- Grodzinsky, Yosef (1989). The language learner: A trigger-happy kid? *Behavioral and Brain Sciences* 12: 342–343.
- Haegeman, Liliane (1994). *Introduction to Government and Binding Theory*, 2nd Ed. Cambridge, MA: Blackwell.
- Harman, Gilbert (1986). *Change in View*. Cambridge, MA: MIT Press.
- Hudson, Richard (1994). About 37 % of word-tokens are nouns. *Language* 70: 331–339.
- Kampen, Jacqueline van (1997). *First Steps in Wh-Movement*. Delft: Eburon.
- Kapur, Shyam (1994). Some applications of formal learning theory results to natural language acquisition. In *Syntactic Theory and First Language Acquisition: Cross-Linguistic Perspectives – Volume 2*, Barbara Lust, Gabriella Hermon and Jaklin Kornfilt (eds.), 491–508. Hillsdale, NJ: Lawrence Erlbaum.

- Keenan, Edward L. and Bernard Comrie (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8: 63–99.
- Kimball, John P. (1973). *The Formal Theory of Grammar*. Englewood Cliffs, NJ: Prentice-Hall.
- Lightfoot, David (1989). The child's trigger experience: degree-0 learnability. *Behavioral and Brain Sciences* 12: 321–334.
- Lightfoot, David (1998). *The Development of Language: Acquisition, Change and Evolution*. Oxford: Blackwell.
- MacWhinney, Brian (1995). *The CHILDES Project: Tools for Analyzing Talk*, 2nd Ed. Hillsdale, NJ: Lawrence Erlbaum.
- Manzini, M. Rita and Kenneth Wexler (1987). Parameters, binding theory and learnability. *Linguistic Inquiry* 18, 3: 413–444.
- Marcus, Gary F. (1993). Negative evidence in language acquisition. *Cognition* 46: 53–85.
- Newport, Elissa (1977). Motherese: the speech of mothers to young children. In *Cognitive Theory, Vol. II*, N. John Castellan, David B. Pisoni, and George R. Potts (eds.), 177–217. Hillsdale, NJ: Lawrence Erlbaum.
- Newport, Elissa L., Henry Gleitman and Lila R. Gleitman (1977). Mother, I'd rather do it myself: some effects and non-effects of maternal speech style. In *Talking to Children: Language Input and Acquisition*, Catherine Snow and Charles Ferguson (eds.), 109–149. Cambridge: Cambridge University Press.
- Osherson, Daniel, Michael Stob, and Scott Weinstein (1984). Learning theory and natural language. *Cognition* 17: 1–28.
- Pinker, Steven (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
- (1989). Markedness and language development. In *Learnability in Linguistic Theory*, Robert J. Matthews and William Demopoulos (eds.), 107–127. Dordrecht: Kluwer.
  - (1994). *The Language Instinct*. New York: W. Morrow and Company.
  - (1999). *Words and Rules: The Ingredients of Language*. New York: Basic.
- Pullum, Geoffrey K. and Barbara C. Scholz (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19, 1–2.
- Rizzi, Luigi (1982). Comments on Chomsky's chapter, "On the representation of form and function." In *Perspectives on Mental Representations: Experimental and Theoretical Studies of Cognitive Processes and Capacities*, Jacques Mehler, Edward C. T. Walker and Merrill Garrett (eds.), 441–445. Hillsdale, NJ: Lawrence Erlbaum.
- Sakas, William G. and Janet Dean Fodor (2001). The structural triggers learner. In *Language Acquisition and Learnability*, Stefano Bertolo (ed.), 172–233. Cambridge: Cambridge University Press.
- Sampson, Geoffrey (1989). Language acquisition: growth or learning? *Philosophical Papers* 18: 203–240.
- (1999). *Educating Eve*. London: Continuum.
  - (2001). *Empirical Linguistics*. London: Continuum.
- Smith, Neil. (in press). Backlash. *GLOT International*.
- Snow, Catherine E. and Charles A. Ferguson (1977). *Talking to Children: Language Input and Acquisition*. Cambridge: Cambridge University Press.
- Thornton, Rosalind (1990). Adventures in long distance moving: the acquisition of complex wh-movement. PhD. dissertation, University of Connecticut, Storrs.
- Valian, Virginia (in press). *Input and Innateness: Controversies in Language Acquisition*. Cambridge, MA: MIT Press.
- Wexler, Kenneth and Peter Culicover (1980). *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.