

# Searching for arguments to support linguistic nativism<sup>1</sup>

BARBARA C. SCHOLZ AND GEOFFREY K. PULLUM

Few arguments are more dangerous than the ones that “feel” right  
but can’t be justified.

Stephen Jay Gould, *The Mismeasure of Man* (1981: 157)

## 1. Introduction

This article is a reply to the foregoing responses to our “Empirical assessment of stimulus poverty arguments” (Pullum and Scholz, this special volume, hereafter *EASPA*). We first address certain philosophical themes that cut across all six responses. We correct the impression held by Lasnik and Uriagereka (L&U) and Crain and Pietroski (C&P) that *EASPA* owes the reader an alternative theory of language acquisition; we distinguish linguistic nativism from several alternatives, only one of them being anti-nativism as espoused by Sampson; we examine the claim of Thomas that there is an identifiable concept ‘poverty of the stimulus’ in the linguistics literature; we point out that Fodor and Crowther (F&C) appear to misunderstand certain mathematical learnability results; and we address a purported argument for nativism (quite distinct from the stimulus poverty argument we considered in *EASPA*) that is advanced independently by several respondents: F&C, L&U, and Legate and Yang (L&Y) – an argument based on the underdetermination of theory by evidence.

---

1. We are grateful to Chris Potts, who contributed substantively to the research for this article (especially sections 6 and 7), and gave us valuable comments throughout. We also thank John Colby and Pieter Seuren for useful discussion and observations. The arguments presented here were developed collaboratively, and both authors take responsibility for the way things are expressed in the final version, but it should be noted that sections 2, 3, and 5 are primarily the work of Scholz, and the rest is joint work.

Running through all of these commentaries is a tendency for the technical vocabulary of linguistic nativist theories not to be used in a univocal and coherent way. In fact our reading of these commentaries has suggested to us that one of the greatest obstacles to progress in the development of linguistic nativist theories of language acquisition is the absence of critical reflection on what the competitors to linguistic nativism are, on what ‘innate’ actually means, on what exactly a ‘poverty of the stimulus’ argument is, and on how the claim that theories are underdetermined by evidence could support a cogent argument for linguistic nativism. These general issues are taken up in Sections 2–5, because the unwarranted assumption that these notions are clear and coherent runs through many of the commentaries. Perhaps the most important contribution we can make to the literature on linguistic nativism is to highlight the fact that further reflection on these matters is necessary.

In the remaining sections, we address specific empirical claims made in several of the commentaries that are alleged to support some form of linguistic nativism. In this category are the commentaries by C&P, F&C, and L&Y.

## 2. Linguistic nativism and its rivals

Some of the responses to EASPA suggest that its argument fails because it does not present an alternative nonlinguistic nativist theory of language acquisition. For example, Lasnik and Uriagereka suggest that unless there are two competing theories of language acquisition under consideration, “it is pointless to engage in mathematical analyses of part of the data. What could that possibly decide if there is only one theory under scrutiny?” (p. 150). And Crain and Pietroski claim: “The challenge for Pullum and Scholz, therefore is to describe a plausible acquisition scenario” (p. 173). They go on:

Pullum and Scholz owe a plausible acquisition scenario of the same phenomena [...] Because Pullum and Scholz avoid the details of linguistic theory and their role in explaining linguistic generalizations, their rebuttal to poverty of the stimulus arguments misses the mark (p. 176).

L&U and C&P take theory preference to be a comparative matter, and we do not take issue with that. But their application of this platitude to EASPA betrays a misunderstanding both of our project and of the various competing linguistic-nativist theories of language acquisition.

Our project in EASPA was to examine the empirical evidence for what we think is the *best* stimulus-poverty argument for linguistic nativism. And although we found the empirical support inconclusive, we did not suppose this evaluation refuted linguistic nativism. Indeed, the aim of EASPA was to point out one way further empirical research could support linguistic nativism. But,

even if it turned out that stimulus-poverty arguments all fail to support linguistic nativism, it still would not follow that the other arguments for linguistic nativism in the literature (e.g., that of Fodor 1975; or the one implied in Pinker 1994: 33–39) also fail. In consequence, it is a mistake to suppose that what is at issue is a question of theory choice, as L&U and C&P seem to assume. It is possible to evaluate an argument for a view without raising questions of theory comparison.

There are other good reasons not to see the discussion in this volume as a matter of comparative theory choice. These have to do both with what the competitors of linguistic nativism are, and with the multiplicity of views that have been called ‘linguistic nativism’ in this volume and elsewhere. In EASPA, we were concerned only with the empirical premise in one kind of stimulus-poverty argument (the specific one we dubbed “the APS”). In consequence, we did not address the (obviously over-simple) disjunctive premise “infants learn their languages either by data driven learning or by innately primed learning”, where innately primed learning “calls upon domain-specific linguistic information”. There is really more than one kind of data-driven learning. To see this, consider the ways that linguistic nativism could be false. Linguistic nativism is the view stated in (1):

- (1) Children and adults have innate domain-specific linguistic knowledge.

Over the past forty years, (1) has been interpreted by its advocates in a variety of ways. The interpretations have centered on both what innate knowledge is and what domain-specific linguistic knowledge is. In the 1960s and 1970s, there were disputes between linguists and philosophers over the meaning of ‘knowledge’ in (1). Some philosophers argued that if the knowledge in question was propositional, as Chomsky had claimed, then it had to be justified, true belief. Chomsky claimed that innate knowledge was not justified, but was knowledge nonetheless. Chomsky (1980: 69) ceded the word ‘knowledge’ to the philosophers and proposed using ‘cognize’ wherever he had previously used ‘know’ a language. This suggestion has not been widely taken up, perhaps because it gave no positive content to what it is to cognize, rather than know, a language. EASPA accepted for the sake of argument that although the meaning of ‘knowledge’ in (1) has shifted over the past forty years, at any given time some sense and reference for it will predominate in the linguistic nativist literature. The article also accepted that ‘knowledge’ in the nativist’s sense is not consciously accessible to the knowing subject.

The claim that the innate knowledge is linguistic is typically understood to mean that what is known has a content that is specific and unique to the linguistic domain. Nativists have advocated quite distinct views about what this linguistic content must be like. Some assume innate knowledge of rules of Universal Grammar. Others maintain that experience sets certain parameters

of an innate parameter system. EASPA did not specify any particular view of domain specificity. This was to ensure that our net was cast as broadly as possible. All that EASPA required of any nativist view on this point was that there be some plausible sense in which the content of the innate knowledge is exclusive to language learning and other kinds of learning that subsequently depends on language learning.

But even if we fix the meaning of ‘knowledge’ and ‘linguistic domain-specificity’ to whatever is their current meaning in the contemporary linguistic nativist literature, there are at least four ways that (1) can be false, and hence four competing alternatives to it. The claim in (1) is false if any claim in (2) is true:

- (2)
  - a. *General skepticism*: There is no such thing as knowledge.
  - b. *Linguistic skepticism*: There is knowledge, but none of it is linguistic.
  - c. *Anti-nativism*: There is knowledge, but none of it is innate.
  - d. *General nativism*: There is innate knowledge, but none of it is linguistic knowledge.

*General skepticism* will be set aside here, since any such view is too far removed from disputes about linguistic nativism to be of interest.

*Linguistic skepticism* (and skepticism about rule-following in general) may be a plausible view, depending on what linguistic nativists, and others allege is known. If what we know comprises rules or principles that guide linguistic behavior in some sense, then skeptical questions attributed to the later Wittgenstein arise. The skeptic demands that some naturalistic fact (psychological, biological, genetic, or whatever) should exist about a speaker that determines that her future unconsidered behavior is guided by one rule *R* rather than another rule *R'* that covers all phenomena so far. A crucial issue is whether grammatical rules – the rules of the ‘psychogrammar’, in the terminology of George (1989) – are standards for correct performance. In other words, is the speaker *mistaken* when she is out of step with a rule of her psychogrammar, and makes a performance error? If genuine human performance errors do occur, then how can some naturalistic fact about the speaker be a standard of correct behavior? Linguistic competence cannot be both normative and non-normative (naturalistic) – or so the skeptic alleges. Some of these issues in relation to linguistics have been addressed in Chomsky (1986); George (1989); and Wright (1989). They were not pursued in EASPA because they were not germane.

*Anti-nativism* appears to be the view held by Sampson (1989, 1999, this volume). The view conflicts with linguistic and general nativism in that it maintains *no* knowledge is innate: anti-nativism holds that linguistic experience, when taken together with general innate capacities (not innate knowledge), learned abilities, and knowledge learned by using these general innate abilities,

is sufficient for language acquisition. The plausibility of the view is intimately tied to what ‘domain-specific linguistic knowledge’ means. For example, no anti-nativist need deny that very young infants have the capacity to detect facts about consonant feature distinctions or syllable sequence frequencies (in fact these abilities are to some extent shared with nonhuman primates, other mammals, and some birds, so *prima facie* they are not unique to human language). EASPA explicitly did not advocate anti-nativism (this seems to be a point of difference between us and Sampson), although anti-nativism is a form of what EASPA called “data-driven learning”, under the definition given there (Section 2.2: 17).

*General nativism* is the view that humans have what might be called ‘innate knowledge’, none of it being specifically linguistic. This innate nonlinguistic knowledge is used to acquire linguistic knowledge and linguistic abilities. The general nativist need not restrict her view to the claim that what is innate are only general deductive or inductive reasoning abilities, any more than the anti-nativist must. For example, suppose it turns out that humans have innate knowledge that permits detection of hierarchical structure in social systems, folk biological taxonomies, and other hierarchical relationships, syntactic and phonological ones being simply special cases. This would be a form of general nativism that is incompatible with general skepticism, anti-nativism, and linguistic nativism. In addition, if the general nativist accepts that linguistic performance is rule-guided, the view is incompatible with linguistic skepticism. EASPA did not advocate general nativism, although it too is a special case of EASPA’s ‘data-driven learning’.

So there are at least two or three plausible competitors to linguistic nativism that were not articulated in EASPA, because they were irrelevant to its argument – the issue there was not about theory choice, but about empirical support for linguistic nativism. But if we narrow our gaze from considering general nativist and non-nativist competitors, and examine only forms of linguistic nativism that are discussed in this volume, we still have many competing linguistic nativist views to choose from. EASPA took linguistic nativism as expressed by (1) to be, while perhaps not precisely delineated, at least univocal and coherent. But a reading of the responses to EASPA shows that it was a mistake to assume this. ‘Innate’ is used in this volume to mean *a priori*, and genetically inherited, and unlearned, and biologically determined:

- Lasnik and Uriagereka speak of “a rational conjecture” that “children come equipped with *a priori* knowledge of language” (p. 149, emphasis ours), explicitly taking innate knowledge to mean *a priori* knowledge.
- Sampson assumes ‘innate’ linguistic knowledge is both genetically inherited and known *a priori*: children “*genetically inherit* innate knowledge of language structure, so that for many aspects of the mother-tongue acquisition task they *do not need experience*” (p. 97, emphasis ours).

- Crain and Pietroski mention biological determination in their abstract (“one should conclude that there are substantive ‘universal’ principles of human grammar and, *as a result of human biology*, children can only acquire languages that conform to these principles” p. 163, emphasis ours), but elsewhere define linguistic nativism in terms of knowledge that is ‘unlearned’ (“The nativist claim is that [...] the grammatical principles that describe the space of possible human languages [...] are *not learned at all*”; p. 166, emphasis ours).

A priori knowledge is knowledge that can be known without a particular experience. Arithmetical knowledge is an example: the knowledge that adding two nonzero natural numbers always makes a number larger than either is acquired through learning, but not learning from any *particular* experience of the world. So L&U’s “a priori knowledge” is not necessarily C&P’s knowledge that is “not learned”, and C&P’s knowledge “as a result of human biology” is not necessarily the same as either. Recall, for example, a well-known behavioral property of rats: they are biologically determined never again to eat a food that made them sick. If a rat gets sick after eating poisoned corn, it will starve to death rather than eat corn again. The knowledge in this case (that it would perhaps be best never to eat corn again) is biologically determined knowledge; but it is not unlearned, and it is not a priori (a particular experience of eating corn was needed for it to be acquired).

At least three nonequivalent forms of linguistic nativism, then, are advocated in response to EASPA. It is an important question which of these views is best supported by the empirical evidence, and which should be preferred given some criteria of theory preference.

The reason it matters that there is no univocal notion of nativism in the responses we are discussing is twofold. Since the advocates of (1) endorse nonequivalent and sometimes incompatible views, not all of the responses to EASPA can be right. Secondly, even if one of them presented a good argument for some form of linguistic nativism, it would not follow that the other incompatible forms of linguistic nativism were also supported. Griffiths (2002) notes a rampant polysemy of ‘innate’ as it occurs in the biological literature, and has concluded that use of the word should be banished from serious writing in the biological sciences. His positive suggestion is that authors instead say what they mean. Authors who mean, for example, ‘unlearned’, ‘shared by all members of the species’, ‘developmentally fixed’, or whatever else should use these more specific terms, not the polysemous term ‘innate’. If Griffiths’ suggestion was adopted in the linguistic-nativist literature, much confusion would be eliminated.<sup>2</sup>

---

2. Things are in no way improved by injecting the word ‘instinct’ into the debate as Pinker (1994) does; see Bateson and Martin (2000: 85–88) for a critique.

Legate and Yang apparently fail to understand that shifting and sliding from one sense of an expression to another in an attempt to support a view is bad reasoning. Their response to the observation in EASPA that there are many distinct stimulus-poverty arguments is to comment:

P&S's article contains more than the usual amount of rhetoric, which strikes us as largely empty. To take one example, P&S survey the arguments for the innateness hypothesis in the literature, identifying 13 separate arguments, and observing that authors choose to focus on a subset of these arguments. The natural conclusion from this would seem to be that the innateness hypothesis is convergently and thus strongly supported. (L&Y, p. 153, Fn. 2)

But we did *not* observe that authors “focus on a subset” of the arguments; we pointed out that they muddle them together. Notice that L&Y's “natural conclusion” is the wrong conclusion if the different arguments support entirely different or incompatible conclusions. It matters what is alleged to be impoverished, and what is alleged to be innate. And a good argument for one form of linguistic nativism from one kind of stimulus poverty is not necessarily a good argument for another form of linguistic nativism from another kind of stimulus poverty. What is required to show that something is biologically determined, for example, will not necessarily show that the same thing is unlearned or a priori.

Another reason it is important to see that there is not just one form of linguistic nativism or stimulus-poverty argument under discussion in this volume is that it sheds light on the kind work that needs to be done to either support or undermine linguistic nativism. It is clear that there is no single notion of innate linguistic knowledge advocated by generativists. What this means is that as such notions as ‘nativism’, ‘domain specificity’, and ‘cognizing’ are refined and developed; at least part of what has to be done – if phrases like ‘poverty of the stimulus’ and ‘innate ideas’ are to function as more than catchphrases – is to compare various linguistic nativist views and their empirical support. Thus the relevant context of theory comparison is *within* the linguistic nativist research program, comparing different linguistic nativist theories of language acquisition.

### **3. The history of stimulus poverty**

In an informative history of the use of the expression ‘poverty of the stimulus’ in the generative linguistics literature, Thomas attempts to relate it to a cluster of pre-1980 claims about what she takes to be related matters. Her discussion ranges widely, covering classical arguments in epistemology, empirical claims about children's experience, theoretical claims about adults' grammars,

methodological proposals for linguistic research, and problems of language acquisition. A variety of different alleged characteristics of children's linguistic experience (degeneracy, finiteness, idiosyncrasy, ambiguity, underdetermination, deficiency, etc.) are touched on. She states that "the term [poverty of the stimulus] participates in a complex network of related expressions and partial synonyms, each of which offers a distinctive prospect on the terms' common referential domain" (p. 52), as if the term might be implicitly defined by this network. But overall, Thomas's review convinces us that there is no unitary concept expressed by 'poverty of the stimulus'.

In Section 3 of her article, Thomas refers to a "central phenomenon to be explained" that has been "variously identified as 'the projection problem', 'the logical problem of language acquisition', 'Plato's problem', 'the learnability problem,' and 'the underdetermination problem.'" (p. 62). This list underlines the conceptual heterogeneity found in the literature. The problems are quite different, relating to transformational grammar, logic, epistemology, mathematical learnability theory, and scientific theories in general:

- (i) "The projection problem" is Peters' term for "the problem of providing a general scheme which specifies the grammar (or grammars) that can be acquired by a human on exposure to a possible set of basic data" (Peters 1972: 172). This problem is specific to generative grammar; in fact the main contribution of Peters' paper (1972: 182–184) concerns issues that arise specifically within transformational grammar.
- (ii) It is not clear what David Lightfoot's slogan "the logical problem of language acquisition" refers to: neither for the child nor for the linguist is language acquisition really a problem in logic as such. But if some problems of logic arise for either, those problems are certainly not specific to generative grammar.
- (iii) "Plato's Problem" is a name given by Chomsky (1986: xxv) to a question that he phrases in words borrowed from Bertrand Russell:<sup>3</sup> "How comes it that human beings, whose contacts with the world are brief and personal and limited, are nevertheless able to know as much as they do know?"; but this is not specifically about linguistic theory or language acquisition, it is a quite general question about knowledge.
- (iv) By "the learnability problem" we suspect that Thomas means the problem that many linguists see in the theorems of Gold (1967) and the many works that applied these results to issues of human language learning. But Gold's work is concerned purely with the mathematics of one very broadly applicable type of algorithm for rational inquiry. (We argue in

---

3. Chomsky gives no source, but the quote (italicized in the original) is from page *v* of Russell (1948).



Section 4 below that it is an error to suppose that Gold has shown language learning from positive input to be impossible.)

- (v) “The underdetermination problem” is not specific to linguistic theories, but to all of science, because all theories are underdetermined by the evidence. In Section 5 below, we study several attempts to draw from it some argument for linguistic nativism, and we find that they all fail.

These five problems, then, are in no way five facets of a single “phenomenon to be explained” (p. 62): the problems all belong to epistemology, but there is no *phenomenon* that unites them.

Thomas herself adds one confusing notion to the literature. She claims that what is really at issue in the poverty-of-stimulus literature is the “incommensurability” of knowledge and experience (pp. 65, 68). We do not understand what sense of ‘incommensurability’ could be relevant here. Incommensurability involves lack of comparability due to the absence of any uniformly applicable measurement system (a cubic centimeter cannot be said to be either larger or smaller than a gram). For example, it is frequently claimed that from finite evidence the child develops a command of language that characterizes infinitely many sentences. But that implies commensurability: there *is* a measure of how many sentences the child receives as input, and how many sentences the acquired generative grammar defines as well-formed, and clearly the generativist’s claim – that the latter is bigger than the former by a factor of infinity – presupposes that.

Thomas’s examples on pp. 67–68 suggest that she sees incommensurability as a relation holding between (i) the knowledge that a certain word-sequence is not grammatical and (ii) experiencing grammatical utterances as having various properties. Certainly, there are differences between (i) and (ii), just as there are various differences between (i’) the knowledge that a certain face is not that of your mother and (ii’) visually experiencing your mother’s face as having certain characteristics. But if this is supposed to show “incommensurability” between (i) and (ii), we do not see how.

Yet Thomas thinks EASPA errs by failing to appreciate that it is knowledge-experience incommensurability, and not stimulus poverty, that is the issue in the case of the acquisition of structure dependence. She writes:

Chomsky’s point [...] balances not so much on the scarcity of models of structure dependence as on whether children’s grammars rule out sentences which are not attested at all, namely ungrammatically structure-independent ones.

(p. 67)

This is just false. The point made in the three Chomsky quotations given in (23) of EASPA (pp. 39–40) is about the extreme rarity of sentences that falsify the first-auxiliary generalization about the grammar of auxiliary-initial clauses. Thomas brushes aside the textual support concerning a specific, testable claim,

and declares that the real issue is something else. The focus in EASPA was on whether there is a good argument that the child has innate linguistic knowledge. We focused on one that depends on a claim that certain evidence is too rare to help the typical child with the acquisition process. Thomas does not address that at all, but simply changes the subject.

Thomas also indicts EASPA for not addressing Crain (1991), in which, she suggests, Crain has “demonstrated children’s knowledge of what is inadmissible, under the assumption that what is inadmissible is not signaled in the properties of the input” (p. 68). But we accept that children learn that some possible sound-meaning correspondences are not permissible. If children didn’t do that much, no language learning could occur. Crain’s experimental work on acquisition evidence – experimental confirmation of claims that children have actually learned certain acquirenda as characterized by linguists – is important work that most linguists have neglected so far. But it does not bear on the questions of inaccessibility (sparseness of evidence in experience) which was EASPA’s main concern.

Thomas remarks in conclusion (p. 69) that the phrase ‘poverty of the stimulus’ has the status of “a kind of touchstone of generative orthodoxy”, but ‘shibboleth’ would actually be a better term. A touchstone is a genuine test of elemental purity; a shibboleth is an expression with which social group membership can be tested. ‘Poverty of the stimulus’ does indeed have a sociological function: it is used by authors to identify themselves as being generative linguists who endorse some nativist line on acquisition.<sup>4</sup> But, the fact that the phrase has that social function does not show that it expresses some definite concept.

#### **4. Negative evidence and learnability**

Theorists of language acquisition have frequently exaggerated the importance of the fact that learning is mostly accomplished on the basis of positive evidence. The standard view is that little or no negative evidence, i.e., explicit information about which sentences are not in the target language, is supplied to the typical infant learner. Some linguists and philosophers seem to think it is has been established that it is impossible to learn a natural language from a finite sequence of purely positive data. Typical of the arguments they give is the one from Haegeman (1994: 10–11) that we mention in EASPA: Haegeman

---

4. For example, Epstein et al. (1998: 4), discussing the framework of classical transformational grammar, say: “Given the poverty of the stimulus, explanatory adequacy was not attained (and perhaps was unobtainable within such a framework).” Nothing is added to the truth conditions of this statement by the first six words, but by using them, the authors identify themselves as nativists.

points to the salient contrast between the grammatical (3a) and the ungrammatical (3b), and asks, in effect, how the child could ever come to know from mere presentations of grammatical utterances that (3b) is not grammatical.

- (3) a. *Who did they think was available?*  
 b. *\*Who did they think that was available?*

The underlying idea seems to be that learning from positive data is impossible in principle if the language is infinite and some strings are not in it.

Indeed, Fodor and Crowther (pp. 106–107) actually seem to think that the infiniteness of languages alone establishes that they are unlearnable from positive data: they note that “any recursive rule (or its equivalent in non-rule-based grammars) suffices to establish that natural languages, being infinite, are necessarily underdetermined by finite input samples”, and as we shall see in Section 5 below, they believe that this underdetermination, combined with the fact that only very limited negative evidence is available, forces the conclusion that there is innate linguistic knowledge.

It is important to see that learnability theory does not support the mathematical impossibility of language learning from positive examples, or what Gold (1967) calls *identification in the limit from text*,<sup>5</sup> even for infinite languages in which some strings are ungrammatical. To assume otherwise would be to assume that the only class of languages over a given vocabulary that is identifiable in the limit from text is the singleton class containing just the language in which every string over the vocabulary is grammatical. But not only are classes other than this identifiable in the limit from text, there are infinite classes of infinite languages that are identifiable (Jain et al. 1999 provide many examples).

Indeed, there exists a class  $\mathcal{L}$  of languages (sets of strings) such that (i) it can reasonably be assumed that the stringsets of all human languages are in  $\mathcal{L}$ , and (ii)  $\mathcal{L}$  is identifiable in the limit from text. The results of Shinohara (1994) entail that for any fixed number of context-sensitive<sup>6</sup> rules  $k$ , the class  $\mathcal{L}_\Sigma^k$  of all languages over a specified vocabulary  $\Sigma$  that have a context-sensitive grammar with not more than  $k$  rules is identifiable in the limit from text (i.e., learnable solely from a finite sequence of positive instances; see Gold 1967 for details of the learning algorithm). It is relatively uncontroversial that all natural language syntactic phenomena can be described by context-sensitive grammars. Given

---

5. A class of languages is identifiable in the limit from text if there exists a learning algorithm that for any language  $L$  in the class will guarantee successful learning of  $L$  from a finite sequence of positive data from  $L$ , settling permanently upon a correct grammar for  $L$  after some finite number of trials.

6. A context-sensitive rule has the form  $\phi\alpha\psi \rightarrow \phi\beta\psi$ , where  $\alpha$  is a string containing at least one nonterminal;  $\beta$ ,  $\phi$ , and  $\psi$  are strings of nonterminals and/or terminals; and  $\beta$  is nonempty. A context-sensitive grammar comprises a finite set of nonterminals, a finite set of terminals, a start symbol selected from the nonterminals, and a finite set of context-sensitive rules.

nothing more than some suitable choice of  $k$  (possibly quite large, so that the general cognitive capacities of human beings would make it impossible for them to utilize grammars that were any larger) and some appropriate linguistic description vocabulary  $\Sigma$ , we can define a class  $\mathcal{L}_{\Sigma}^k$  of which the natural languages are plausibly a subclass. So learning natural language syntax from text is a mathematical possibility.

Shinohara's results reveal that the worst-case scenario for psycholinguistics is not as bad as many think: they show that languages could be learned even from an absurdly weak input consisting of a finite sequence of uninterpreted positive example strings, and that this holds for large and rich classes of languages. If richer input is assumed, or if weaker assumptions than Gold's are made about what counts as learning success, language learning becomes even easier. For example, identifying a function  $\theta$  from a presented sequence of pairs  $\langle a, \theta(a) \rangle$  is trivially easy: every recursive function is identifiable in the limit from a finite sequence of such pairs (see Jain et al. 1999: Section 3.9). As has been pointed out by a number of researchers, this means that identifying a function that pairs sounds with meanings given a sequence of phonetic/semantic representation pairs is vastly easier than identifying a stringset.

Likewise, if the definition of success is weakened from exact identification to various kinds of approximate identification, or to the 'strongly approaching' relation of Feldman (1972), the effect on the ease of language identification is similarly dramatic: all the recursively enumerable sets become identifiable in the limit from text.

Many factors can affect learnability under this mathematical idealization: whether the 'learner' is taken to have a finite limit on memory, whether the input is presumed to contain a finite quantity of bad data, etc. Many surprising results have been proved about what does or does not make a difference to mathematical learnability; whether a mathematically defined learning procedure can be shown to yield guaranteed learning success on a certain defined class of languages depends on various technical details. But it does not depend on assistance from some innate component of mind that is language-specialized. The 'learner' is assumed to follow some algorithmic procedure, but it can be a procedure that is used for learning anything; that is, the conception of learning involved is as compatible with general nativism as with linguistic nativism.<sup>7</sup>

From the observation that children do not appear to be supplied with negative evidence, therefore, it does *not* follow that learning is impossible. Any claim that it is impossible for children to learn a language from text, and thus that linguistic nativism must be true, needs to be supported in some other way.

---

7. It may be compatible with some forms of anti-nativism too; that would depend on whether using a learning procedure entails having knowledge of that procedure.

## 5. Underdetermination of theories by evidence

A common thread running through several of the articles in this volume, particularly Fodor and Crowther, Lasnik and Uriagereka, and Legate and Yang, is that linguistic nativism can be supported by arguments based on the premise that *theories are underdetermined by evidence*. Since the argument addressed in EASPA has no such premise, these authors are changing the subject. But they appear to think that they are changing the subject to a better argument for linguistic nativism, one that we overlooked. EASPA did not overlook arguments based on underdetermination theses: they were alluded to in separating the different arguments found in the literature. We will now discuss underdetermination arguments more fully. We argue that they hold no promise of supporting linguistic nativism.

There are various underdetermination theses that are familiar from the philosophy literature. Some refer to deductive inference and others to ampliative (inductive) inference. We consider these separately in Section 5.1 and Section 5.2, and then briefly deal with another argument in Section 5.3.

### 5.1. Deductive underdetermination and linguistic nativism

A common and uncontroversial formulation of deductive underdetermination (usually credited to Hume), says:

- (4) *Deductive Underdetermination*  
Any given finite body of evidence is compatible with indefinitely many mutually contrary theories.

The truth of (4) depends on what we understand by ‘theory’ and ‘evidence’.<sup>8</sup> Let us stipulate first that all theories contain at least one universally quantified hypothesis with an unrestricted domain of quantification, and second that no universally quantified statements with unrestricted domains are evidence.<sup>9</sup> We accept (4) on this understanding; we assume that all parties do.

---

8. (4) is not strengthened if we add that the domain of quantification of the theory is infinite. All that is required is that the domain does not have some definite upper bound that exactly matches the evidence. Fodor and Crowther seem to think the claim that languages have infinitely many expressions adds something to (4). In truth, it does not alter the situation.

9. This last restriction may be too strong, but accepting it only strengthens the linguistic nativist’s case. The point is to avoid trivialization. For example, clearly deductive underdetermination would not be true if we could take as our set of observations  $\{F(a), F(b)\}$  and deduce the ‘theory’  $\{F(a) \wedge F(b)\}$ , or if we could take  $\forall x[F(x)]$  as an ‘observation’ and let the deduced theory be identical with that.

Notice that deductive underdetermination is not unique to linguistic theories. If it is true, then the linguist is in the same epistemic situation as all other scientists: linguistics is no easier than any other science. And linguistic nativists appear to assume that (4) also holds for child language learners too.

If (4) is true, no unique theory  $\Theta$  can be deduced from a finite body of evidence  $E$ , because no theory incompatible with  $\Theta$  but compatible with  $E$  will be deductively ruled out by  $E$ , and there will always be such additional theories. No conclusion about innate knowledge – linguistic or otherwise – follows directly from (4), since it says nothing about knowledge. But Legate and Yang state an argument for nativism (which they say is the argument from poverty of the stimulus) that appeals to underdetermination:

The logic of the APS is quite simple: if you know  $X$ , and  $X$  is underdetermined by experience, then knowledge of  $X$  must be innate. (p. 152)

The idea is that given some acquisition evidence supporting the claim that some  $X$  is known by a subject, and assuming that (4) holds for the subject's knowledge of  $X$ , one can conclude that the subject knows  $X$  innately. But there are some suppressed premises: this conclusion follows only if

- (i)  $X$  is a theory (that is, it contains at least one universally quantified hypothesis), and
- (ii)  $X$  is known by being deduced from a finite body of evidence.

But (4) cannot be accepted together with both (i) and (ii), because (4) entails that no theory can be deduced from a finite body of evidence. Given (4), (i) and (ii) cannot both be true.<sup>10</sup>

We suspect L&Y may hold the view that there are two kinds of theories. The kind of theories found in the sciences that are deductively underdetermined by the evidence, so they are not known by means of deductive inference. The other kind includes theories of the structure of individual languages that are learned by children – in short, grammars – which *are* acquired by deductive inference yet are *not* underdetermined. Given (4), L&Y reason, learning a grammar depends on some innate knowledge which, together with the evidence, deductively entails a speaker's grammar.<sup>11</sup> In other words, they assume that (4) does not hold for children's grammars. But then, if they assume that children's grammars are not underdetermined, their argument is not based on (4), because they deny that (4) is true for children's grammars. Moreover, in assuming that children's grammars are not underdetermined, their argument tacitly *assumes*

---

10. Linguistic nativism would follow from the conjunction of (i) and (ii) with (4), but only because anything follows from a contradiction.

11. Note the remarks of Chomsky (1980: 75) about the possibility of “a theory of universal grammar from which we can, let us say, literally deduce the grammar that gives a first-order account of the facts, given boundary conditions set by experience.”

nativism, and thus begs the question: they posit something innate in order to make sure that something together with the evidence entails the grammar that is acquired by the child.

It is worth noting that even if L&Y's argument were not question-begging, it would do no more than support some form of nativism (i.e., the view that there is some innate knowledge), not the desired conclusion that there is innate *linguistic* knowledge. So the argument could not support linguistic nativism over general nativism.

Fodor and Crowther also claim (4) must be false for children's grammars of a particular language. They remark that positive and negative linguistic evidence

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. (p. 123)

Temporarily putting aside the question of the role of negative evidence (addressed below in Section 5.3), this appears to mean that (4) is false of "the generalizations that are formulated about the language", because language learners have "internal principles" from which, together with the evidence, they deduce these generalizations. Again, nativism is being *assumed*. And again, the argument would support whatever form of nativism is assumed. So at best, the arguments of L&Y and F&C do not support linguistic nativism over general nativism (one of the ways it can be false).

In fact at one point F&C actually seem to admit that linguistic nativism comes in as a necessary assumption rather than as a conclusion: "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" (p. 108).

What makes it so hard to see that linguistic nativists like F&C and L&Y are presupposing the thesis of linguistic nativism, rather than giving an argument for it, is that they appear to accept (4). For example, L&Y, when they start amplifying their argument sketched above, say that they are proposing a specific instantiation of the argument from poverty of the stimulus that has this structure:

- [G]iven language data  $D$ , and a simple but incorrect hypothesis of  $D$ ,  $H$ ,
- a. the child behaves as though he/she does not entertain  $H$
  - b. the evidence necessary to rule out  $H$  is not available to the child
  - ∴ the child possesses innate knowledge excluding  $H$  from the hypothesis space. (p. 152)

Their premise (b) is true not just for the child language learner but for anyone, given (4). So L&Y seem to assume (4) for the child. But their conclusion only

follows if their argument has a suppressed premise that states:

- c. Every child has innate knowledge sufficient to rule out every incorrect contrary hypothesis that is compatible with the evidence.

In this case the conclusion follows, but only because it was assumed in (c). Or at least, L&Y's argument assumes that for every incorrect *H* that the child fails to entertain, *something other than the evidence* rules it out. If that something is innate knowledge, nativism is being assumed; if it is something else, then no argument for nativism has been presented. And even in the best case, the argument could only support the form of nativism that is assumed in (c). But then it may not support linguistic nativism over general nativism.

A similar mixture of views – an apparent appeal to (4) coexisting with talk about deductive hypothesis elimination – is found in Lasnik and Uriagereka:

Not even the fact that (1b) [*Is the dog that is in the corner hungry?*] is grammatical *proves* that something with the effect of hypothesis B ['front the auxiliary in the matrix Infl'] is correct (and the only possibility), hence [that fact] does not lead to adult knowledge of English. More generally, no amount of positive evidence, 'exotic' or not, would suffice. The poverty of the stimulus is thus extreme. (p. 148, emphasis added)

L&U seem to assume that for any finite body of evidence there are indefinitely many contrary hypotheses that are all compatible with it; that is, they assume (4). And they seem to assume that adult knowledge of language must be knowledge of a universally quantified hypothesis. They also assume that "adult knowledge of English" deductively rules out all hypotheses but one (it is "correct", and it is "the only possibility"). So they also assume (4) is false of adult knowledge of English. And from the claim that there are indefinitely many theories that are compatible with any given finite body of evidence they conclude that "the poverty of the stimulus is ... extreme". This actually puts the boot on the other foot: they offer not an argument from the poverty of the stimulus to a conclusion for some kind of nativism, but instead they present an argument that assumes (4) plus adult knowledge of English, and concludes that the stimulus must therefore be regarded as impoverished!

In the face of the extreme poverty of linguistic experience L&U purport to have established, they take linguistic nativism to be a "rational conjecture" about language learning. And so it is. But general nativism is also a rational conjecture, as is the view that social factors in conjunction with the evidence could determine grammars. Nothing in L&U's argument supports one rational conjecture over the others.

What is constant across F&C, L&U, and L&Y is that none of them gives a good argument for linguistic nativism based on (4). F&C and L&Y assume



that indefinitely many incorrect hypotheses are ruled out or overcome by innate knowledge, but that does not show that the posited innate knowledge is necessarily linguistic knowledge. Their arguments beg the question in favor of nativism, and fail to support it over its most plausible competitor – general nativism. L&Y also assume a contradiction, namely that the universally quantified statement (4) is true but that an instance of it (the case of the acquisition of grammars by infant learners) is false.

The argument presented in EASPA has linguistic nativism as a conclusion (unlike L&U's); it does not assume what it concludes; and it does not assume a contradiction. So it is plausibly a stronger argument for linguistic nativism than the ones presented by F&C, L&U, and L&Y.

### 5.2. *Ampliative underdetermination and linguistic nativism*

Fodor and Crowther make it clear that at least some of the time they do not have (4) in mind when they appeal to the underdetermination of theories by evidence. For example, they write:

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. (p. 107)

F&C concede that “inductive principles must exist for other domains of knowledge” (i.e., they seem to accept that either general nativism or anti-nativism is true for inductive reasoning abilities), and they claim that underdetermination of theory by linguistic input can establish the existence of innate linguistic knowledge only if grammars are underdetermined by ampliative inductive inference. A “gap between input sample and grammar” must be established, and it must be shown that “highly domain-specific information” is needed to “bridge the gap”.

In a trivial sense there is by definition a gap between premises and conclusion in even fully cogent inductive arguments. Induction contrasts with deduction in that inference is ampliative: the conclusion of an inductive argument always contains more information than its premises. So inductive arguments are non-demonstrative: the truth of their premises does not guarantee the truth of their conclusions. The conclusions of even the best inductive inferences are

*fallible* in a way that the conclusions of sound deductive inferences are not. In addition, inductive arguments can be strengthened (or weakened) as additional premises are added to them in a way that deductive arguments cannot be. In consequence, we should regard the conclusions of even the best inductive arguments as *revisable*, as additional relevant premises are added to the argument. There is nothing unique to linguistics here, if we accept that children fallibly induce a revisable grammar.

Suppose, then, that from the linguistic input, using inductive principles of inference, children (fallibly) infer some (revisable) grammar. There is no gap needing to be bridged, other than the gap that is present by definition in all inductive inferences, unless it can be shown that in the case of inductive inference of a grammar there is some special “gap” that needs to be bridged between the child’s input and the child’s linguistic knowledge at a given time. We might be able to establish the existence of such a gap for the child if we used the method suggested in EASPA: characterize some acquirendum, specify a lacuna, establish the indispensability of the lacuna sentences for inductive inference of the acquirendum, and present the inaccessibility evidence and acquisition evidence. But F&C just assume what they need to show: that there *is* some special gap between input and grammar.

A passage in Lasnik and Uriagereka offers a hint of assistance for F&C on this point. L&U write:

In sum, [Pullum and Scholz] are missing Freidin’s point: (1b) [*Is the dog that is in the corner hungry?*] is not *direct* evidence for anything with the effects of hypothesis B [‘Front the auxiliary in the matrix Infl’]. At best, (1b) is evidence for something with the logical structure of ‘A [‘Front any auxiliary’] or X’, but certainly not for anything having to do with B, implying such notions as Infl and matrix clause (or, another possibility suggested by Freidin, the Head Movement Constraint). Being compatible with hypotheses A or B does not mean providing the elements to form either hypothesis. Crucially for the issue at hand, a learner needs more to acquire hypothesis B. (p. 149)

L&U’s point here is that no linguistic input contains “elements” like ‘Infl’ or ‘matrix clause’, but the acquired knowledge of language – e.g., hypothesis B, which says ‘Front the auxiliary in the matrix Infl’ – does include such “elements”. So L&U suggest there is a special gap between the raw linguistic data of the input and a linguistic concept that is acquired, a special gap that is bridged by innate language-specialized concepts like ‘Infl’ and ‘matrix clause’.

Clearly, L&U cannot intend the trivial claim that children’s typical linguistic experience does not include explicit information about the concepts ‘matrix clause’, ‘clause boundary’ and other expressions that appear in syntactic theories. The absence of this explicit information in the input will not show that there is a special gap between the totality of the information in the input and

the ultimately acquired grammar, because learners can and do inductively infer concepts that are merely implicit in their experience. For example, children who learn how to count come to know (implicitly)<sup>12</sup> that ‘successor of’ is a function, not merely a relation, even though the information that counting relies on its being a function does not explicitly appear in their mathematical experience.

For this kind of special gap to exist there must be nothing that is detectable by the child from which she could (and does) induce the relevant linguistic concepts. There must be no detectable signs of clause boundaries or of matrix clauses in the input, for example, and there must be nothing detectable by the child in the input from which these concepts can be inferred. So to learn the structure-dependence generalization about auxiliary-initial clauses, F&C and L&U need to show *but not assume* that children cannot detect clause boundaries and matrix clauses, or whatever else is necessary for inducing a structure-dependent generalization. If the child can detect clause boundaries and clause relationships, and can induce the relevant revisable linguistic concepts, then there is no special gap between the input and the concepts needed to formulate general linguistic principles.

At this point, there are two questions that the advocate of this kind of argument for nativism must answer: first, how can the advocate of nativism show that there is such a gap and not just assume it? And second, what kind of inductive underdetermination is it that can permit a mere concept to bridge the demonstrated gap?

In relation to the first, if the linguist assumes “that everything that might be innately known, is so,” as in the nativist research strategy described by F&C (p. 141), then she will not find evidence for a special gap between input and acquired grammar, because the whole point of the innate linguistic concepts is that they make sure *there is no gap*.

In relation to the second, if the gap between premises and conclusion is just the one that is expected in all inductive inference, no mere *concept*, innate or not, can bridge the gap. What is needed is some further principle of inductive inference. On the other hand, if for inductive learning of grammars there is some special gap that can be bridged by an innate concept, then we need to be told exactly what that is, and its existence needs to be supported.

The aim of the experimental schema suggested in EASPA was to show that there *is* a special gap. The idea is to show that evidence about certain construction types is inaccessible to the child, yet the child comes to know the properties of those construction types nevertheless. Far from it being “pointless to engage in mathematical analyses of [...] the data” to verify inaccessibility, as L&U

---

12. Skeptics about mathematical knowledge will not agree with this, of course, but we are setting aside skeptical positions.

assert (p. 150), this is exactly what the nativist must do (in addition to other things) to get the project off the ground.

### 5.3. *Between POPS and PONS: a third underdetermination argument*

Many other arguments from underdetermination might in principle be constructed. We cannot deal with every possibility here. But we will deal with one, offered by Fodor and Crowther. After referring to the “qualitatively more severe” underdetermination of children’s grammars that is the result of the combined effects of limitations on positive and negative evidence, they say:

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. (p. 123)

The form of this argument appears to be as follows:

- (5)
- a. The positive data the child receives cannot determine a unique grammar because positive data convey no information about what is *not* in the language.
  - b. The negative data the child receives is incomplete, so it does not precisely delimit what is not in the language generated either.
  - c. For strings that are not included in the positive data and not ruled out by the negative data, there is an information gap: no data received by the child have any bearing on their status.
  - d. But children learn the status of such sentences anyway (they arrive at knowledge of a grammar that generates some and does not generate others).
  - e. So there must be something internal to the learner that overcomes the underdetermination that the information-gap represents.

It is certainly true that no unique set  $L$  is determined by choosing a finite set  $P$  of positive evidence (a collection of sentences that are in  $L$ ) and a disjoint finite set  $N$  of negative evidence (a collection of sentences that are not in  $L$ ). That is, for any vocabulary  $\Sigma$ , there are infinitely many infinite sets  $L$  of strings over  $\Sigma$  such that for finite  $P$  and  $N$  we have

- (6)
- a.  $P \subset L$
  - b.  $N \cap L = \emptyset$

We can prove this simply by giving an example of an infinite collection  $\mathcal{L}$  of distinct infinite sets satisfying the stated conditions:

- (7) Let  $\Sigma$  be a finite vocabulary of symbols, let  $\Sigma^*$  be the set of all strings over  $\Sigma$ , and let  $P$  and  $N$  be fixed finite sets of strings in  $\Sigma^*$ . Let ' $A \setminus B$ ' stand for the set  $A$  with all members of the  $B$  removed. Then let
- $$\mathcal{L} = \{L \subseteq \Sigma^* \mid \exists x [L = (P \cup (\Sigma^* \setminus N)) \setminus \{x\}]\}$$

That is,  $\mathcal{L}$  is the set of all those sets of strings  $L$  over the vocabulary  $\Sigma$  such that for some string  $x$  over  $\Sigma$ ,  $L$  contains all the strings in  $P$  plus every other string over  $\Sigma$  that is not in  $N$ , except for  $x$ .  $\mathcal{L}$  is infinite, because  $N$  is finite, which means the complement of  $N$  in the set of all strings over  $\Sigma$  is infinite, so there are infinitely many possible values for  $x$ .

F&C want to conclude that since choosing a finite  $P$  and  $N$  does not uniquely determine a particular  $L$ , language learning is impossible. But the collection  $\mathcal{L}$  defined in (7) is identifiable in the limit from text in the sense of Gold (1967)!<sup>13</sup> Under this classic mathematical formalization of learning from positive examples, at least, it can be demonstrated that the positive and negative evidence do *not* conspire to make acquisition impossible. Data-driven learning is possible, and the “qualitatively more severe” underdetermination of which F&C speak says nothing to support linguistic nativism.

It is interesting to reflect on why F&C think otherwise. It apparently has to do with the kind of grammars they think a child has to acquire. They assume a child must learn a generative grammar that contains “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 generated would be larger” (p. 122). They believe that without negative evidence it is impossible to acquire such negative grammar statements. Their assumptions have in effect built linguistic nativism into their grammatical theory. If some particular form of grammatical theory requires that vast amounts of evidence about what is *not* in the language is needed for a grammar to be acquired, and everyone agrees such evidence is *not* available to children, a sensible conclusion to draw would surely be that the grammatical theory in question should be discarded. But F&C choose instead to hold onto a gerrymandered distinction between positive and negative grammar statements, and assume that whatever is needed to learn the latter is already built into the brain of the learner. They are entitled to their view. But they are not entitled to say that linguistic evidence has been provided in its support.

Why do F&C believe that the only way learners could ever discover that a syntactic configuration is grammatically disallowed is to receive explicit negative information to that effect? One reason may be their apparent antipathy

---

13. The proof is trivial – the example is basically identical to the one discussed as an introductory example by Jain et al. (1999: Section 1.3: pp. 5–8).

to the idea of statistical learning. They ignore very widely promulgated developments in stochastic grammar. In particular, they ignore the idea that when attempting to select a grammar on the basis of a given corpus of positive examples, a learner might strive for simultaneous maximization of (i) the probability of the grammar given the corpus and (ii) the probability of the corpus given the grammar.

Even without a detailed understanding of stochastic grammars, which explicitly assign a probability figure to every constituent described, or of the mathematics of measuring the probability of a corpus relative to a given grammar, the point should be intuitively clear. Take a classic example of a negatively formulated constraint like the one from Postal (1974) saying that no constituent may linearly intervene between the verb of a clause and its direct object. A learner who has tentatively hypothesized a grammar that leaves manner adjunct position in the VP free will be expecting verb phrases of the form (8b) to be just as frequent as those of the form (8a).

- (8) a. [<sub>VP</sub> call the police immediately]  
 b. \*<sub>[VP</sub> call immediately the police]

If the latter type actually turns out to have a vastly lower (virtually zero) frequency, then it is clear that the grammar-corpus fit could be greatly improved. A much better fit would be achieved, in fact, if the grammar were revised to forbid adverbs between verb and direct object altogether. No negative evidence is needed to make that determination.

Linguists who are seriously interested in how grammars might be acquired should not ignore the developing literature on statistical learning and how it might apply to grammar induction, especially in view of the firmly established fact that prelinguistic human infants only a few months old become sensitive to differential frequencies of syllable sequences they are exposed to, on the basis of just minutes of exposure (Saffran, Aslin, and Newport 1996).

## 6. Arguments from genres and corpus properties

An obvious strategy for responding to EASPA would be to argue that its critiques of inaccessibility claims are based on inappropriate corpus material. Fodor and Crowther take up this line, albeit in unhelpfully polemical mode.<sup>14</sup>

---

14. For example, they surely make too much of EASPA's brief mention of Blake's line, *Did He Who made the lamb make thee?* Sampson (1989) had mentioned that since many children hear this in childhood, that settles the question of whether children ever hear sentences refuting the first-auxiliary-fronting generalization. F&C respond by worrying about whether infants learn from 18th century poetry, how they must screen out archaic forms, and so on. They even cite

One of their charges is that EASPA represented the *Wall Street Journal* as a “source of information about language input to small children” (p. 106). This claim is false, as the reader of EASPA can verify. We did use the *Wall Street Journal* corpus (*WSJ*) for preliminary testing of the occurrence of certain constructions (mainly because the price was right: in the 1990s the Linguistic Data Consortium and the Association for Computational Linguistics sold a CD ROM containing *WSJ* at a price that works out to millions of words per dollar, a real bargain). But we used *WSJ* only for initial checks on frequency in the adult language, to see whether it might be worth proceeding.

It was quite reasonable to do such pre-checks. Chomsky’s strident claim (Piattelli-Palmarini 1980: 114) was that “you can easily live your whole life without producing” a sentence refuting the first-auxiliary-fronting claim, so we wondered whether such sentences might be entirely absent from *WSJ*. We found that the number of interrogatives in *WSJ* that you have to examine before you hit one of the crucial sort is 15. That seemed interesting enough to encourage some further work.

The same question arises with Baker’s claim that only “an unusual set of circumstances” could reveal that anaphoric *one* could take a multi-word antecedent. How many occurrences of anaphoric *one* in adult English do you have to examine before you find one that clearly shows in context that it has a multi-word antecedent? Despite the fact that most tokens of *one* (as a numeral, etc.) are irrelevant, the answer is that it is easy to find them quite quickly. Browsing through the first two 1987 files of *WSJ* produced good examples within the first 60,000 words. (Recall that a working-class child will hear that many words in as little as two days.) That seemed like a preliminary indication that Baker’s “unusual” should not be taken too seriously, and it might be worthwhile to look a bit further.

However, in every case where accessibility evidence was the issue we followed up our preliminary scans of *WSJ* by collecting examples from various sorts of other texts too: classic children’s novels; transcripts of children’s TV programs; cartoon captions; and utterances from the CHILDES database that were addressed to a toddler.

Nothing supports F&C’s assertion that we were “using the *Wall Street Journal* as a source of information about language input to small children” (p. 106). But in rebuttal of this claim we never made, F&C conduct an unbelievably pointless experiment. They take just *eight* sentences from a random *Wall Street Journal* article, and carry out comparative statistical analysis with similarly

---

a conjecture (due to Barbara Bevington) that children obey the (innate?) maxim “Never learn from anything that rhymes.” This seems unbelievable to us, with its implication that *Jack and Jill went up the hill* cannot be input to the child’s learning; but we will not waste space discussing it further here.

tiny samples of other sorts, solemnly laying out the figures in a table. The reader will not be too astonished to find that 100 % of the 8 sentences of the article examined are declarative, or that the percentage ungrammatical is 0 %. Nor will the reader be surprised to find that mean length of utterance (MLU) is about 5 when talking to a toddler, 10 when talking to a grownup, and 20 in an article about urban excavation access policies for telecom companies pulling fiber-optic.

Is any of this relevant? F&C claim that “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.” But this claim is false unless the frequency of construction types in different genres are both significantly and relevantly different. F&C present no data at all on the frequency of any crucial construction type. And average MLU figures are not relevant. An utterance length of 5 words – indeed, 4 phonological words – is enough to allow a structure that disconfirms the auxiliary-initial clauses generalization:

- (9) a. *Is what’s left mine?* [fronting main clause auxiliary]  
 b. *\*Is what left’s mine?* [fronting first auxiliary]

Our first *WSJ* example of such a construction was 8 words long. If utterances addressed to toddlers are around 5 words long with a standard deviation of 3 or 4 (i.e., if utterances range mostly between 1 and 9 words), there is no reason to think that utterances of the relevant syntactic form might not be amongst them.

We should add that we do not think the category of utterances addressed to toddlers has decisive importance. F&C make repeated references to “child-directed speech”, but our assumption is that children do not learn solely from speech that is addressed to them.<sup>15</sup> We assume they learn from a great deal of what they hear. And they hear all sorts of speech that was never intended for their ears: older children arguing, neighbors complaining, TV comedians joking, popular song lyrics repeated by brothers and sisters, commercials trying to interest them in toys, Uncle Jack telling a dirty joke, the plumber explaining about the toilet leak, and parents making plans. It would be astonishing if they did not learn from material unless it was addressed to them – they cite no evidence supporting this implausible claim. But if children do learn from speech they overhear, then F&C’s list of the characteristics of child-directed speech does not exhaust the properties of the speech that children learn from, and certainly doesn’t show that any of the spot checks of construction types that we did are beside the point.

---

15. Of course, their language learning might well be impeded if virtually *no* speech was addressed to them; but that is a different point.



## 7. Arguments from new linguistic data

A different strategy for responding to EASPA would be to argue that there are other acquirenda that we did not consider, and these do provide support for linguistic nativism. There is hardly any sign of this sort of response in the articles in this volume, except in the case of Crain and Pietrosky. C&P attempt to give arguments for linguistic nativism based on facts not considered in EASPA. But the arguments they construct are very poor ones.

### 7.1. Negative polarity and inclusive disjunction

C&P appear to accept that the basic distribution pattern of NPIs in English could be learned from linguistic experience. They then move on to a different point: “In the vast majority of cases, English sentences with the disjunction operator *or* are naturally understood with an ‘exclusive-*or*’ interpretation” (p. 169). Taking the exclusive reading to be basic,<sup>16</sup> they assume that an account is needed of which contexts cause *or* to be interpreted “conjunctively” (i.e., inclusively with both disjuncts taken to be true), and they claim that the statement in (10) describes those contexts:

- (10) The conjunctive interpretation is assigned to disjunctive statements if and only if an NPI can appear in that linguistic environment. (C&P’s example (25), p. 171)

Then they cite (11) and claim that the “(deeper) property of the relevant environment, that of *downward entailment*”, has greater explanatory power than (10), moving us “in the direction of an explanation” of why (10) holds:

- (11) Downward entailing linguistic environments license NPIs *and* constrain the interpretation of disjunctive statements (to conjunctive readings). (C&P’s example (27), p. 171)

While (10) states necessary and sufficient conditions for the conjunctive interpretation of *or* and for the possible distribution of NPI’s, (11) states a sufficient condition for the conjunctive interpretation of disjunctions and the possibility of NPIs. So if “license NPIs” in (11) means merely that these environments permit NPIs to occur, then (10) entails (11), although there is no entailment the other way.

C&P then refer to a general Gricean pragmatic principle as adapted by Chierchia. In this case it is assumed that the inclusive reading of *or* is basic. The

---

16. This is the opposite of what we think is the correct analysis, incidentally; see Gazdar (1979: 78–83) for discussion.

Gricean idea is that *or* corresponds to the inclusive disjunction of classical logic, but in those contexts where the exclusive reading of *or* is more informative, it will be understood as exclusive *or*. C&P don't state this principle, but they think that it is "plausible" (which will apparently be incompatible with (11), since it takes a different reading of *or* as basic). Their summary is:

Regardless of whether either of them is correct, however, we see no reason to doubt [...] that some semantic principle – call it 'downward entailment' – unifies what otherwise seem to be [...] disparate phenomena" (p. 172).

Clearly, nothing said so far supports linguistic nativism. Everyone agrees that the more unifying theory or hypothesis is, *ceteris paribus*, to be preferred. But C&P want to conclude from the above observations that linguistic nativism is true. They need a principle like (12) to get from the phenomena to nativism:

- (12) Any principle (or hypothesis) that unifies disparate phenomena is innate.

Given (12), all maximally unifying theories would be innate, whether they are linguistic theories or not. Superstring theory would be innate by (12), since it unifies quantum mechanics and relativity theory. So (12) is completely implausible.

Yet we are not raising (12) as a straw man. C&P actually seem to tacitly accept it, for they claim that scientific rigor requires us to address the possible existence of a unifying generalization: "simply denying apparent generalizations is just bad science." Nativism must therefore be true: "One can't avoid nativist conclusions by refusing to do linguistics."

One of the most remarkable things about this non sequitur is that there can be no generalization that unifies (11) and the unstated third principle, because they rest on contrary assumptions about the core meaning of *or*.

Section 3 of C&P presents another purported generalization about the distribution of NPIs, namely that every NPI must have a licensing negative element that c-commands it. C&P assume:

- (13) C-command is an abstract structural relationship that cannot be defined in terms of perceptible features of word strings. (p. 174)

The relevant question for the linguistic nativist is whether there is linguistic evidence detectable and taken up by children which either is an instance of the c-command relation or is an observational basis from which children infer such a relation. What is needed is a lacuna specification and some inaccessibility evidence. For if children detect instances of that relation or infer it from what they do detect, there is no lacuna, and so no accessibility evidence, and so no argument for nativism.

In EASPA we set out a specific type of argument that can be used to support some form of linguistic nativism (a plausible one, we supposed). It demands, in the first instance, a statement of what is alleged to be known, the acquirendum specification. Suppose we take C&P's (11) as the acquirendum. What (11) says is first that downward entailing linguistic environments are sufficient to permit the occurrence of NPIs, and second that downward entailing linguistic environments are sufficient for the conjunctive reading of *or*. For example, if the verb *doubt* introduces a clause, then an NPI like *anything* can follow it in the same clause (*I doubt that they'll do anything*), and *or* will be interpreted inclusively if a verb like *doubt* introduces the clause (*I doubt that they'll go to Oxford or Cambridge*). There are positive instances of (11) from which it might be learned.<sup>17</sup> So, ex hypothesi, there is no lacuna – no absence or poverty of the stimulus required for learning (11). Indeed C&P present no poverty of the stimulus argument at all, only a tacit argument for a highly implausible form of nativism based on a highly questionable linguistic analysis.

## 7.2. Crain and Pietroski's "strongest argument"

C&P attempt to develop an another argument for nativism in their final section. But this argument is not a poverty-of-the-stimulus argument under any construal. Rather it purports to be an argument for linguistic nativism based on covariation of children's and adults' speech. They claim that "mismatches between child and adult language may be the strongest argument for Universal Grammar" because there is some evidence that "Child speech can differ from the local adult languages only in ways that adult languages can differ from each other." If we understand them, the relevant comparison is not between child and

---

17. Notice, it would not be confirmable from positive data if they perfected the conditional to a biconditional, and said that the acquirendum was this:

- (P2') NPIs are licensed, and likewise disjunctive statements take their conjunctive readings if and only if the linguistic environment is a downward-entailing one.

This would not be confirmable from positive data, because it implies a negative existential over an unbounded domain: that there is not a single occurrence of an NPI in a grammatical expression of the language where the context is other than downward entailing. The trouble with (P2'), though, is that it is not true, as shown by non-monotone quantifiers like *exactly 3*:

- (i) a. Exactly three people brought me any fruit while I was in the hospital.  
 b. Exactly three people brought me any grapes while I was in the hospital.

The inference from (ia) to (ib) is downward from a set (the set of people who brought me fruit) to a subset (the set of people who brought me grapes); and the inference does not go through: in a world where the only three people who came to the hospital all brought me strawberries, (ia) is true but (ib) is not. Yet the NPI *any* is licensed.

adult grammars, but rather between adult speech and child speech (with performance errors idealized away, of course). So we can restate the claim this way:

- (14) All ways in which normal, considered, grammatical ways of speaking by adults can differ from one speech community to another are ways normal, considered, grammatical ways of speaking by children in a given speech community can differ from normal, considered grammatical ways of speaking by adults in that community.

Suppose (14) is true. Then the variation between child and adult grammatical speech within a speech community does not exceed the variation between kinds of adult speech across all speech communities. Does this support the claim that “UG” exists? That will depend on what “UG” is in this case, and whether it says any more than what (14) says. There are many reasons to think (14) is true, most of them independent of any linguistic nativist claim or view about what UG is. Consider the following:

- (15) All ways normal adult chin shapes can differ from each other across breeding populations are ways normal children’s chin shapes can differ from normal adult chin shapes within a breeding population.

Although (15) is probably true, it doesn’t follow that some *theory* of chin shape is innate. Note also that chin shape is a standard example of a feature that is not an evolutionary adaptation (but merely a by-product of adaptations relating to dentition), and thus cannot really be said to be biologically determined in the relevant way. It is the same with (14): it may be true, but that does not entail that some theory of grammar is innate, or that (14) has any specific biological determination.

Perhaps all C&P intend to say is that (14) supports the claim that the scope of variation in speech is unlearned. We could likewise infer from (15) that chin shape is unlearned. But in neither case does it follow that some tacitly known universal grammar is unlearned, and that is the claim the linguistic nativist needs to support. If this is the “strongest argument for UG”, we fear it still needs quite a bit of work.

## **8. Arguments from statistical frequency**

A further way in which an opponent might respond to EASPA would be to address the actual statistical claims made: are the supposedly inaccessible data rare enough? Sampson and Legate and Yang address this topic. Both focus on the corpus frequency of sentences refuting the front-the-first-auxiliary generalization.

Sampson did some useful corpus work on the basis of his access to the British National Corpus (which was not available to us when EASPA was written; until very recently scholars outside the United Kingdom had no access to it). He reports a surprising finding: that the 4.2m words of the demographically-sampled spontaneous speech part of the corpus appears not to contain a single example of the relevant kind – with an auxiliary preceding a subject that is complex enough to contain an auxiliary.<sup>18</sup> He does, however, note the presence in the corpus of *wh*-interrogative sentences with complex subjects, like *Where's all the parts that fell apart?* (see Sampson's Fn. 9), and confirms that five such examples occur in the 80,500 words of the structurally annotated Christine corpus. That suggests there may be one for every 3,000 utterances. Perhaps open interrogatives are more important as evidence for learning the relevant syntactic point than polar interrogatives. But would one occurrence per 3,000 utterances (if that figure were confirmed) be enough to ensure that no one could ever form the impression that “front the first auxiliary” was the correct rule? Would every child hear at least some such sentences before starting nursery school? These are the kinds of questions that we would assume a linguistic nativist should show some interest in.

Legate and Yang do show some interest. They propose a novel method for determining whether or not some frequency of occurrence of a crucial construction type is high enough to provide a basis for data-driven language learning. We will discuss their interesting contribution in detail.

First we should clear up a small misunderstanding. L&Y describe EASPA's treatment of the data from the Nina corpus in CHILDES as “curiously selective”. We looked in the corpus of dialogue between Nina and her caregivers for examples of interrogatives that had clause-initial auxiliaries that would not be first auxiliary in the related declarative. L&Y affect surprise at finding “counts from only one file, NINA05.CHA, which happens to be the file that has the most number of critical sentences, out of all 56 files” (p. 157). The suggestion is that we found all the relevant examples in the corpus, but then reported only the results for one file, the file that had the largest number.

They are wrong about this. When we remarked, “It did not take long to find three examples in the file NINA05.CHA”, we did not intend to imply that a systematic count of the whole corpus had been undertaken (we explicitly cautioned that the needed corpus work “has not yet been undertaken, and we do not claim to have undertaken it”). We quoted these three sentences from NINA05.CHA simply because they were the first ones we found. Our method was simply to read through the files in the Nina directory until some relevant cases were found, and then stop. Unfortunately we missed a few. L&Y worked

---

18. Recall that a sentence as short as 5 words could in principle display this complexity, as in our (9) above.

more systematically, and provide a useful correction. What we found suggested that one needed to go through about 30,000 words of conversation with toddlers before encountering sentences of the relevant sort. But L&Y's investigation reveals that the first case comes much earlier: after about 8,200 words of dialogue (at the 600th question in the Nina database). If this were typical, then from the results of Hart and Risley (1995), it would follow that a child of professional parents should hear one of these sentences every day on average, and even a child of parents on welfare would hear one every three days or so.<sup>19</sup>

L&Y undertake the work of counting the crucial example types in the Nina corpus for a specific reason. They propose a new method for assessing whether a construction's corpus frequency is sufficient to be deemed accessible to infant learners. The idea is to compare the frequencies of two constructions that are crucial data for the acquisition of two different acquirenda that are typically acquired at the same age.

To put this in more precise terms, first we take two decision problems that the learner has to solve; then we verify that children typically solve these problems at approximately the same age; and finally we determine the frequencies of the two relevant kinds of crucial evidence in the typical child's experience, and compare them.

Call the problems  $P_1$  and  $P_2$ , and let the frequencies of occurrence of utterance tokens instancing the two relevant constructions be  $F_1$  and  $F_2$ . L&Y reason:

- (16) "if  $F_1$  and  $F_2$  turn out significantly different, then  $P_1$  and  $P_2$  must represent qualitatively different learning problems." (L&Y, p. 155)

We suspect that they mean that some deep qualitative difference is revealed about the ways in which  $P_1$  and  $P_2$  are solved, viz., that the solution to one of them is preinstalled as part of innate universal grammar while the solution to the other is not. (We will return below to another conclusion that could be drawn.)

The specific problems they choose for an application of their method are these:

$P_1$ : Is the generalization about auxiliary-initial clauses structure-dependent (stated in terms of fronting the auxiliary of the matrix clause)?

---

19. We cannot be sure, because it is certainly possible that a child of a welfare mother hears fewer subordinate clauses per thousand words. We know too little about what the input to welfare children is like. But of course, we also know too little about whether their syntactic development is delayed.

$P_2$ : Is English a null subject language (one in which nonemphatic pronominal subjects may be freely dropped)?

The correct answers to the two decision problems are of course yes to  $P_1$  and no to  $P_2$ . L&Y cite works by Crain and Nakayama on  $P_1$  and Valian on  $P_2$  to support the claim that both are normally acquired by the age of 36 to 38 months. The null hypothesis is that  $P_1$  is solved by attending to utterances of one type that have frequency  $F_1$ , and  $P_2$  is solved by attending to utterances of another type that have frequency  $F_2$ . Under the null hypothesis, the frequency of the crucial evidence in each case should be similar. L&Y's empirical contribution is to make a preliminary assessment of  $F_1$  and  $F_2$  and to show they are quite divergent.

In identifying the indispensability evidence, L&Y follow us, as we followed Chomsky and all the other literature, in assuming that the evidence crucial for solving  $P_1$  is the occurrence of sentences like (17a), which are in the lacuna for the auxiliary-initial sentence generalization.

- (17) a. *Will those who are interested stand?*  
b. *Those who are interested will stand.*

It is the second of the underlined auxiliaries in (17b) that appears initially in (17a). But L&Y also accept our point that *wh*-interrogative clauses are evidence for  $P_1$ , as are polar interrogatives like (17a). For example, the occurrence of (18a) is evidential support for  $P_1$  because the auxiliary that is positioned after the fronted nonsubject *wh*-phrase in (18a) is not the first of the underlined auxiliaries in the corresponding declarative sentence (18b):

- (18) a. *Where's the other dolly that was in here?*  
b. *The other dolly that was in here is there.*

Regarding the evidence that the child needs to solve  $P_2$ , L&Y make a very important decision without providing much supportive reasoning (we return to discuss this point extensively below). They say:

Following the generalization that the use of *there*-type expletives correlates with obligatory subjects, *there*-type expletives have been argued to be the evidence disconfirming an optional subject grammar (Hyams 1986; Jaeggli and Safir 1989; among many others). Thus, we need only to count the frequency of *there* expletive sentences to get  $F_2$ . (p. 156)

Assuming with no further ado that the frequency of *there*-insertion sentences is the critical factor in learning that English is not a pro-drop language, L&Y report that expletive *there* sentences are 40 times more frequent than sentences like (17a) or (18a) in the 20,651 sentences in the Nina corpus. It now follows

straightforwardly by modus ponens, from their own statement (16) above, that the two acquirenda are qualitatively dissimilar.<sup>20</sup> Inexplicably, however, L&Y draw instead the conclusion that linguistic nativism is true:

[...] the conclusion then seems to be Chomsky's (1975: 33): "the child's mind [...] contains the instruction: Construct a structure-dependent rule, ignoring all structure-independent rules. The principle of structure-dependence is not learned, but forms part of the conditions for language learning." (p. 159)

This conclusion does not follow. The logic of what *does* follow is rather involved, but we think it is interesting, since L&Y are on the threshold of an argument against general nativism that actually might work, so we will examine it in detail. We adopt the following temporary abbreviations:

- G General nativism: there is innate knowledge, but none of it is linguistic.  
 Q Qualitative similarity:  $P_1$  and  $P_2$  are qualitatively similar problems.  
 T Temporal coincidence:  $P_1$  and  $P_2$  are solved by learners at approximately the same age.  
 F Frequency similarity:  $F_1$  and  $F_2$  are approximately the same in relevant corpora.

The argument we take L&Y to be approaching runs as follows:

- |      |    |                                       |   |
|------|----|---------------------------------------|---|
| (19) | a. | $(G \wedge Q \wedge T) \rightarrow F$ | Under general nativism two qualitatively similar problems $P_1$ and $P_2$ , solved at similar ages, should have similar $F_1$ and $F_2$ . |
|      | b. | $\neg F$                              | But the frequencies $F_1$ and $F_2$ are not the same.   |
|      | c. | $\neg(G \wedge Q \wedge T)$           | So the conjunction of general nativism with the claims of qualitative similarity and temporal similarity does not hold.                   |
|      | d. | $\neg G \vee \neg Q \vee \neg T$      | I.e., either general nativism or qualitative similarity or temporal similarity does not hold.   |
|      | e. | T                                     | But $P_1$ and $P_2$ are solved at similar ages.   |
|      | f. | $\neg G \vee \neg Q$                  | So either general nativism or qualitative similarity does not hold.   |
|      | g. | Q                                     | But $P_1$ and $P_2$ are qualitatively similar.  |
|      | h. | $\neg G$                              | Therefore general nativism is false.  |

- 
20. (i) "if  $F_1$  and  $F_2$  turn out significantly different, then  $P_1$  and  $P_2$  must represent qualitatively different learning problems" (L&Y, p. 155)  
 (ii) " $F_1$  [...] is 40 times lower than 1.2% [=  $F_2$ ]" (L&Y, p. 158)  
 (iii) therefore  $P_1$  and  $P_2$  represent qualitatively different problems (by modus ponens)



For the sake of argument, let us suppose that the conditional claim (19a) is true (though doubtless there is more to be said on that). The argument needs empirical support at three critical points: (19b), (19e), and (19g). For the nonce, we assume that  $P_1$  and  $P_2$  are solved at similar ages, so (19e) will not be questioned here. With (19g), that  $P_1$  and  $P_2$  are qualitatively similar problems, we believe L&Y will face trouble. They need the truth of this premise because, otherwise, all they can conclude is that *either* general nativism is false *or* the two problems are qualitatively different. But  $P_1$  and  $P_2$  do not seem to be qualitatively similar at all.

$P_1$  is a special case of what nearly all syntacticians take to be a universal generalization (whether innate or not) about natural language syntax: that it involves sequences of atoms paired with a categorized hierarchical structure.<sup>21</sup>  $P_2$ , on the other hand, is entirely parochial, and definitely settled by access to some sort of English data (since in some languages the data are different and  $P_2$  has to be settled in the opposite way).

But it is with (19b) that the most serious trouble arises. Showing that  $F_1 \approx F_2$  depends on more than just showing that two numbers are not significantly different. It must also be the case that the right classes of utterance tokens have been selected for determining  $F_1$  and  $F_2$ . And we think it is clear that L&Y cannot support their claim that the right choice of  $F_2$  is the frequency of sentences with expletive *there*. L&Y say that expletive *there* sentences “have been argued to be the evidence” for solving  $P_2$ , citing Hyams (1986) and Jaeggli and Safir (1989) for support, and conclude that “we need only to count the frequency of *there* expletive sentences to get  $F_2$ ” (p. 156). But the works cited do not provide support for the claim.<sup>22</sup>

A purported typological generalization (originally from Perlmutter 1971) lies at the basis of Hyams’s case: in certain western European languages we find (i) subjects obligatory in nearly all finite clauses, and (ii) expletive NPs (usually with locative etymology) as subjects of existential clauses. Moreover, in a somewhat larger class of European languages both (i) and (ii) are false: subjects are not required in finite clauses except for emphasis, and no overt expletive NPs appear as subjects in existential clauses. But Hyams exaggerates this typological correlation when she suggests that “lexical expletives are *always* ‘avoided’ in a pro-drop language” (1986: 74, emphasis ours). Elaborating on this claim, she says (1986: 92):

---

21. For a suggestion (admittedly brief and speculative) concerning how sensitivity to hierarchical structure might not be specific to language, see the thought-provoking discussion in Sampson (1999: 122 ff).

22. We will ignore Jaeggli and Safir, in fact, since they simply mention Hyams’ work briefly in the course of an introductory overview. Hyams’ work is the sole source of the idea under discussion.

Given that lexical expletives appear only in non-pro-drop languages [footnote omitted], these elements may serve to trigger a restructuring of [the child's early mistaken analysis] ... We assume that the child ... knows that a lexical pronoun is avoided where a null pronominal is possible ... [And] he knows that because [expletives] are void of semantic content they do not signal emphasis, contrast, change of discourse topic, etc. In other words, they are not present for pragmatic reasons. The alternative is that they have a strictly grammatical function. At this point, the child may deduce that the expletives are necessary because a null expletive is impossible in [subject position] ... If this hypothesis is correct, we expect that at the point at which the child begins using expletive *it* and *there* (and hence, we know that he is aware of these elements) he will no longer produce subjectless sentences.

There are many weaknesses in Hyams's argument (in many cases she notes them herself, so they are not really in dispute). First, three of her typological claims about languages are false.

A. It is not true that lexical expletives, in those languages that have them, are just subject slot-fillers, teleologically connected to the property of subject obligatoriness as a device for complying with subject requirements. Expletive pronouns appear in nonsubject positions in many different constructions. When Hyams (1986) was being written, this may not have been widely appreciated, but it is very fully supported in Postal and Pullum (1988). The relevance is that the nonsubject positions in question (direct object, object of preposition) are not obligatorily filled in English. This undercuts the whole motivation for viewing expletives as inherently tied to subject obligatoriness, or obligatory filling of any positions.

B. It is also not true that "lexical expletives appear only in non-pro-drop languages". Hyams herself observes in a footnote (1986: 102–103, Fn. 10) that Modern Hebrew is a pro-drop language that uses the third person pronoun *ze* as an extraposition expletive, and the 13th-century Italian of Boccaccio and Petrarcha had free subject pronoun drop but used *egli* as an existential expletive. Hyams dismisses these facts with highly unconvincing and unsupported suggestions about the Hebrew usage possibly being "scholastic" and the Italian being "literary". There is no evidence for either claim. In fact, Modern Italian still uses an existential expletive, *ci*, and far from being literary, this is a spoken form, the literary language having *vi* instead.<sup>23</sup>

C. Finally, it is not true that subjects are uniformly obligatory in English. The learner who adopted that view would find data from colloquial English baffling. The facts are quite complex, and we cannot describe them in detail here, but anyone interested will have no trouble in documenting the fact that subject

---

23. We are indebted to Pieter Seuren for this observation.

pronouns are dropped in English. There can hardly be many English-speaking children who never hear anything like these examples:

- (20)
- a. *Had a talk with the boss this morning.*
  - b. *Beats the hell out of me.*
  - c. *Sorry I'm late; missed my bus.*
  - d. *Makes me wonder why I bother.*
  - e. *Doesn't make any difference to me.*
  - f. *Makes a lot of sense, that does.*
  - g. *Really knows his stuff, that guy.*
  - h. *No problem; won't take a minute.*
  - i. *Can't have people running around naked.*
  - j. *Drives me crazy.*
  - k. *Woke up, fell out of bed, dragged a comb across my head; found my way downstairs and drank a cup, looking up I noticed I was late ... Found my coat, and grabbed my hat, made the bus in seconds flat ... (Lennon and McCartney, "A Day In The Life")*

Interrogative examples (missing an auxiliary verb in addition to the subject) are also relevant:

- (21)
- a. *Got milk? (TV commercial)*
  - b. *Ever seen one of these before? (John McEnroe, menacingly holding a tennis ball up in front of an umpire's face; TV commercial)*
  - c. *Wanna make something of it?*
  - d. *Seen any good movies lately?*
  - e. *Like anything you see?*
  - f. *Have a good time last night?*

In short, there are subjectless independent clauses all over the place in conversational English; all learners will encounter them. This could considerably slow down the learning of the relevant acquirendum here (that English is indeed, to a substantial degree, a language in which finite clauses need overt subjects). Universal grammar cannot help in analyzing the implications of the various colloquial constructions just cited. Since the grammars of other languages differ on such points, the child has to learn from linguistic experience what to infer from the above sorts of examples, what the register and style connections are, and so on. With a fair amount to figure out about where subjects can legitimately be dropped and where they cannot, it is perhaps not too surprising if the generalization seems to be learned rather late.

Three of Hyams's assertions about infant learners and the course of the acquisition process cast further doubt on her thesis:

A. The claim that children never use expletive subject pronouns until they stop dropping subjects is false: Hyams herself notes (1986: 108, Fn. 35) that child

production corpora do show occasional uses of expletives in the period before subjects become obligatory. She dismisses these “precocious expletives” as anomalous.

B. Hyams does not show what the direction of implication is: even if her typological generalization were sound, and it were true that children discovered expletives and stopped dropping subjects at around the same time, she would have no more evidence for her generalization than for the one obtained by reversing the direction of implication. That is, it could be the case that the child figures out from the data that one is expected to have a subject in nearly every kind of finite clause, and *then* notices overt existential expletives, and starts using them, the better to comply with that generalization.

C. Finally, innate knowledge on the part of the child is again being assumed in an argument for innate knowledge on the part of the child: L&Y follow Hyams in assuming that the occurrence of expletive *there* will be relevant to the child’s learning process because “the child knows that a lexical pronoun is avoided where a null pronominal is possible” (Hyams 1986: 92). But this assumes the very thing that is at issue in this debate, namely whether the infant innately knows some language-specialized contingent facts.

In sum, Hyams’s claims, both from facts about languages and from facts about children’s learning, are very shaky. We cannot assume that learners of English simply note that existential clauses contain an overt expletive *there* and set the Null Subject Parameter on that basis. Children’s success in learning where to use clausal subjects by shortly after the age of three may well have nothing to do with the occurrence of existential *there* clauses, so there might be no implication to draw regarding the parallel with the age of acquiring the generalization about auxiliary-initial clauses. That means it is not true that “we need only to count the frequency of *there* expletive sentences” to get the evidence we need (L&Y, p. 156).

L&Y have made a useful contribution by pointing out the possibility of broadening the relevant database through comparison of learning times and data frequencies for different constructions. But the particular application they choose is problematic in a number of ways. More work remains to be done on this idea.

We conclude with one interesting point that might be relevant to that further research. In Scandinavian languages like Norwegian and Danish, the frequency of occurrence of the existential expletive pronoun appears to be about double the frequency of expletive *there* in English (Ebeling 2000; Mikkelsen 2001). L&Y claim that frequency of triggering data determines time of learning (see their Footnote 5, where they insist that if more abundant data on some acquirendum is available, it will be acquired earlier). They therefore predict a dramatically earlier setting of the Null Subject Parameter to its negative value among Scandinavian learners. We do not know what the facts are about acqui-

sition age. It would be embarrassing for the hypothesis of L&Y if Scandinavian children only learned to make subjects obligatory at around 3;2, but it would be interesting support for them if those children seemed to acquire the generalization much earlier, markedly sooner than children acquiring English.

## 9. Conclusion

Let us summarize. In Section 2 we distinguished linguistic nativism (the view that there is innate knowledge about language) from four rival views, at least two being plausible competitors to linguistic nativism: *general nativism* and *anti-nativism*. Section 3 addressed Thomas's literature review on the history of the notion 'poverty of the stimulus'. Taken together, Section 2 and Section 3 show that there is no unified position advanced in the replies to EASPA. There is no unitary concept of linguistic nativism, because of the polysemy of the word 'innate', which is used variously to mean a priori, genetically inherited, unlearned, or biologically determined, and an argument that supports one of the resultant versions of linguistic nativism will not necessarily support another; and there is no unitary concept 'poverty of the stimulus' either (in fact we suggested that the phrase is just a shibboleth with social function).

In Section 4 we reminded the reader that, contrary to what some linguists seem to believe, lack of negative evidence does not make language learning impossible in any demonstrable way. Section 5 then identified three different purported arguments for linguistic nativism based on underdetermination of theories by evidence – arguments distinct from the APS that we considered in EASPA. We showed that none of the arguments based on the underdetermination of theory by evidence supports linguistic nativism.

By the end of Section 5 we had dealt with all of the content of Lasnik and Uriagereka, and also with the arguments from underdetermination presented by Fodor and Crowther and Legate and Yang.

Certain critiques of our use of corpora by F&C were addressed in Section 6; we rebutted two new putative arguments from empirical data by Crain and Pietroski in Section 7; and then in Section 8 we turned (at last) to the main proposal of EASPA, the idea that frequency statistics might bear on assessment of the APS. We examined an extended argument against EASPA by L&Y. Although we argued that it fails as presented, we drew from it an argument form that we think might be fruitfully deployed against general nativism.

It is an interesting fact that one of the respondents to EASPA, Sampson, thinks we should become anti-nativists like him – and the other five think we already are. Let us try once again to be clear: *we are not advocating anti-nativism*. EASPA set out a form of argument, dubbed "the APS", that would be capable of providing *support* for linguistic nativism, and explored four cases

that attempt to support its empirical premise. In each of those cases our (tentative and admittedly incomplete) probing of corpora available to us suggested that the empirical support was not there. But claiming that four instances of an argument for linguistic nativism appear to lack support does not amount to presenting an argument for anti-nativism, or for general nativism. Indeed, in Section 8 of the present article we have attempted to develop, on the basis of ideas in L&Y's interesting work, a new kind of argument that looks as if it might be capable of refuting certain forms of both anti-nativism and general nativism, if the necessary empirical support were forthcoming.

This, it seems to us, is the kind of work that needs to be done. Different versions of linguistic nativism will have different forms of general nativism as plausible competitors. There are varieties of linguistic nativism (for example, those based more within ethology and evolutionary biology) that we have not addressed in this article, and they will have general-nativist competitors too. There may ultimately be ways to decide between the competing views. It is clear to us that this work has, as yet, hardly begun. But, it is to be hoped that the contributions to this volume, taken together, make it considerably clearer where the field stands, what sorts of issues remain open, and what kinds of evidence and argument are, or are not, relevant to settling those issues.

*Stevenson College, UCSC*

## References

- Bateson, Patrick and Paul Martin (2000). *Design for a Life: How Behavior and Personality Develop*. New York: Simon and Schuster.
- Chomsky, Noam (1957). *Syntactic Structures*. The Hague: Mouton.
- (1975). *Reflections on Language*. New York: Pantheon.
- (1980). *Rules and Representations*. Oxford: Basil Blackwell.
- (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.
- Crain, Stephen (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 14: 597–612.
- Crain, Stephen and Mineharu Nakayama (1987). Structure dependence in grammar formation. *Language* 63: 522–543.
- Crain, Stephen and Paul Pietroski (2002). Why language acquisition is a snap. *The Linguistic Review* 19: 163–183.
- Ebeling, Jarle (2000). Presentative constructions in English and Norwegian. A corpus-based contrastive study. PhD. dissertation, University of Oslo.
- Epstein, Samuel David, Erich M. Groat, Ruriko Kawashima, and Hisatsugo Kitahara (1998). *A Derivational Approach to Syntactic Relations*. New York, NY: Oxford University Press.
- Feldman, Jerome (1972). Some decidability results on grammatical inference and complexity. *Information and Control* 20: 244–262.
- Fodor, Jerry A. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.
- Fodor, Janet and Carrie Crowther (2002). Understanding stimulus poverty arguments. *The Linguistic Review* 19: 105–145.

- Gazdar, Gerald (1979). *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.
- George, Alexander (1989). How not to become confused about linguistics. In *Reflections on Chomsky*, Alexander George (ed.), 90–110. Oxford: Basil Blackwell.
- Gold, E. M. (1967). Language identification in the limit. *Information and Control* 10: 447–474.
- Gould, Stephen Jay (1981). *The Mismeasure of Man*. New York: W. W. Norton.
- Griffiths, Paul E. (2002). What is innateness? *The Monist* 85: 70–85.
- Haegeman, Liliane (1994). *Introduction to Government and Binding Theory*. 2nd edition. Oxford: Basil Blackwell.
- Hart, Betty and Todd R. Risley (1995). *Meaningful Differences in the Everyday Experiences of Young Children*. Baltimore: Paul H. Brookes.
- Hyams, Nina (1986). *Language Acquisition and the Theory of Parameters*. Dordrecht: D. Reidel.
- Jaeggli, Osvaldo, and Kenneth Safir (1989). The Null Subject Parameter and parametric theory. In *The Null Subject Parameter*, Osvaldo Jaeggli and Kenneth Safir (eds.), 1–44. Dordrecht: Kluwer Academic.
- Jain, Sangay, Daniel Osherson, James S. Royer, and Arun Sharma (1999). *Systems That Learn*. 2nd edition. Cambridge, MA: MIT Press.
- Lasnik, Howard and Juan Uriagereka (2002). On the poverty of the challenge. *The Linguistic Review* 19: 147–150.
- Legate, Julie and Charles Yang (2002). Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review* 19: 151–162.
- Mikkelsen, Line (2001). Reanalyzing the definiteness effect: evidence from Danish. Unpublished ms., Department of Linguistics, University of California, Santa Cruz.
- Perlmutter, David M. (1971). *Deep and Surface Structure Constraints in Syntax*. New York: Holt Rinehart and Winston.
- Peters, Stanley (1972). The projection problem: how is a grammar to be selected? In *Goals of Linguistic Theory*, Stanley Peters (ed.), 171–188. Englewood Cliffs, NJ: Prentice-Hall.
- Piattelli-Palmarini, Massimo (ed.) (1980). *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.
- Pinker, Steven (1994). *The Language Instinct: How the Mind Creates Language*. New York: William Morrow.
- Postal, Paul M. (1974). *On Raising: One Rule of Grammar and its Theoretical Implications*. Cambridge, MA: MIT Press.
- Postal, Paul M. and Geoffrey K. Pullum (1988). Expletive noun phrases in subcategorized positions. *Linguistic Inquiry* 19: 635–670.
- Pullum, Geoffrey K. and Barbara C. Scholz (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19: 8–50.
- Russell, Bertrand (1948). *Human Knowledge: Its Scope and Limits*. New York: Simon and Schuster.
- Saffran, Jenny, Richard N. Aslin, and Elissa L. Newport (1996). Statistical learning by 8-month-old infants. *Science* 274 (13 December): 1926–1928.
- Sampson, Geoffrey (1989). Language acquisition: growth or learning? *Philosophical Papers* 18: 203–240.
- (1999). *Educating Eve*. London: Cassell.
- (2002). Exploring the richness of the stimulus. *The Linguistic Review* 19: 73–104.
- Shinohara, Takeshi (1994). Rich classes inferable from positive data: length-bounded elementary formal systems. *Information and Computation* 108: 175–186.
- Thomas, Margaret (2002). Development of the concept of ‘the poverty of the stimulus’. *The Linguistic Review* 19: 51–71.
- Wright, Crispin (1989). Wittgenstein’s rule-following considerations and the central project of theoretical linguistics. In *Reflections on Chomsky*, Alexander George (ed.), 233–264. Oxford: Basil Blackwell.