

## To the Editor of *Language*:

In the paper derived from his 2008 Presidential address, Stephen Anderson (2008) makes two misleading statements about how evidence might bear on language acquisition, one relating to corpus use and the other to typological generalizations. His errors have been made previously in the literature, and should not be repeated in *Language* without correction.

Anderson discusses a familiar claim of Chomsky's: that children do not learn from experience that the formation principles for interrogatives in English are structure-sensitive rather than string-sensitive. Take the facts in (1):

- (1) a. Everything that wasn't eaten will be thrown away.
- b. Will everything that wasn't eaten be thrown away?
- c. \*Wasn't everything that eaten will be thrown away?

Comparing (1b) with its corresponding declarative (1a) reveals that the simple string-sensitive hypothesis in (2) is wrong:

- (2) The first auxiliary in the declarative must be positioned initially in the corresponding interrogative.

This hypothesis would predict (1c). I will call sentences like (1b) telltale sentences, since their testimony brings out the difference between the first auxiliary in the string and the auxiliary that follows the subject of the clause. If language were learned from the evidence provided by experience, then encountering telltale sentences would help the learner by permitting (2) to be eliminated. But Chomsky claimed (Piattelli-Palmarini 1980:40, also 114--115) that 'A person might go through much or all of his life without ever having been exposed to relevant evidence' --- i.e., evidence that would confirm the correct hypothesis over the tempting but incorrect (2). In other words, he claimed that telltale sentences are so rare that you might well never encounter one in your whole life.

Pullum & Scholz 2002 (and before that Pullum 1996) probed that claim a little by looking in a readily available body of text, the *Wall Street Journal* corpus (*WSJ*). Telltale sentences showed up immediately. However, Anderson remarks: 'one might well question the extent to which the *Wall Street Journal* is representative of the input to the child' (804).

People have said such things before: Fodor 2001 (relying on a discussion in Cowie 1999 of the preliminary report in Pullum 1996); Fodor & Crowther 2002:108--114; and Boeckx & Hornstein 2007. People offer their armchair opinion that *WSJ* could not be relevant, but no one goes back and checks *WSJ*. The first telltale sentence in *WSJ*'s 44 million words is the 16th interrogative that occurs, and it is not an instance of financial journalistic prose or an editorial about capitalism. It is a 9-word sentence from spontaneous speech, the penultimate sentence of this passage:

Afterward, one of Mr. Tsongas's partners at the Boston law firm of Foley, Hoag & Eliot, told him: "You've been invited to join your last corporate board."

Mr. Tsongas says he is puzzled by such observations. "Is what I'm doing in the shareholders' best interest? Then what's the problem?"

It is just wrong to assert that *WSJ* cannot provide any evidence that might bear on child language acquisition, because there is no reason to think sentences of the sort Mr. Tsongas uttered in the above passage will be absent from the speech that children hear.

Who knows how many times children hear *Will whoever is making that noise please stop?*, or *Could those who have to leave early sit over here?* I don't (though I know I heard an ordinary person say *Is what you're doing enough?* when speaking spontaneously to a reporter on the BBC World Service). What *WSJ* shows us immediately is that whatever the frequency of telltale sentences in unscripted speech, it is certainly higher than Chomsky asserted.

*WSJ* does not have to be 'representative of the input to the child' to be relevant for the purpose Pullum & Scholz had in mind; it only needs to contain a sample of things real people actually say. (Essentially this point is made in Scholz & Pullum 2002:206-208, which appeared in the same journal issue as the one Anderson cites, but he does not mention it.) In any case, telltale sentences turned out to be present in every corpus Pullum & Scholz looked at.

Anderson then gives a second argument for setting Pullum & Scholz 2002 aside. He says that Legate & Yang 2002 'develop a precise account of the statistical prominence in the input data that seems to be necessary, and show that the level attained by [telltale sentences] is far below this threshold.' But Legate & Yang do no such thing. They supply an estimated frequency of 1.2% for existential clauses in child-directed speech and an estimated frequency of 0.068% for telltale sentences, and note that the latter is 40 times smaller. That is backing up a hunch with a couple of ballpark guesses, not 'a precise account of statistical prominence in the input data.'

Nonetheless, Legate & Yang's argument, while not statistical in nature, is a complex and interesting one. It can be summarized roughly as follows.

- (i) Assume that languages fall into two types, optional-subject ('pro-drop') and obligatory-subject.
- (ii) Assume that noticing that existential clauses with *there*-type expletives, and nothing else, triggers learning that a language is of the obligatory-subject type.
- (iii) Assume that noticing telltale sentences, and nothing else, triggers learning that (2) is an incorrect generalization.
- (iv) If the trigger for learning about obligatory subjects is 40 times more frequent than the trigger for learning (2), then children should learn the former generalization long before the latter.
- (v) But this is not so: children seem to learn both at around the same time (by about a month or two after their third birthday).
- (vi) Therefore there are too few telltale clauses to trigger learning from experience.
- (vii) Yet children do learn that (2) is wrong.
- (viii) Therefore not all learning is from experience, so the thesis of linguistic nativism is true.

This argument certainly deserves attention. And it has received it, though Anderson does not note that. It is answered in detail in Scholz & Pullum 2002:217-221. The core problem is that Legate & Yang attribute the claim about the crucial status of clauses with *there*-type expletives to Hyams 1986. But Hyams made at least half a dozen clearly false or highly dubious assumptions, which Scholz & Pullum list (218-220).

Perhaps the most important point is that partitioning the languages of the world into two types, 'optional-subject' and 'obligatory-subject', is a hopeless oversimplification. When considered in the context of the complex and subtle conditions governing subject omission in 'partial pro-drop languages' like colloquial Finnish or Estonian, it looks absurdly simplistic (see e.g. Duvallon & Chalvin 2004).

Perhaps in due course arguments like the one Legate & Yang attempt might enable us to find out whether or not young children hear enough telltale sentences in everyday speech to permit experience-based learning of the principles of interrogative formation. It is worth trying to find out. Scholz & Pullum 2002:220-221 point out an

interesting datum that might play a role in this enterprise (the frequency of *there*-type expletives in Danish is double the English frequency, which under a non-nativist view might suggest that Danish children should learn certain aspects of language at a faster rate). But essentially all of the work remains to be done.

It is not enough simply to wrinkle one's nose at the *Wall Street Journal*, wave an arm in the direction of the 'pro-drop parameter', and move on. The question of whether telltale sentences might be relevant to the learning of auxiliary placement remains open as far as I can see. Anderson is only the latest of quite a few writers who treats the issue too lightly and dismisses non-nativist positions too swiftly. The important question of whether the typical child's linguistic experience is just too meager to support language learning deserves to be taken more seriously than that.

## References

- Anderson, Stephen R. 2008. The logical structure of linguistic theory. *Language* 84.795-814.
- Boeckx, Cedric and Norbert Hornstein. 2007. Les différents objectifs de la linguistique théorique [The varying aims of linguistic theory]. In *Cahier Chomsky*, ed. Julie Franck and Jean Bricmont, 61-77. Paris: L'Herne.
- Cowie, Fiona. 1999. *What's Within*. Oxford University Press.
- Duvallon, Outi & Antoine Chalvin. 2004. La réalisation zéro du pronom sujet de première et de deuxième personne du singulier en finnois et en estonien parlés. *Linguistica Uralica* 4.270-286.
- Fodor, Jerry. 2001. Doing without what's within: Fiona Cowie's critique of nativism. *Mind* 110.99-148.
- Fodor, Janet D. & Carrie Crowther. 2002. Understanding stimulus poverty arguments. *The Linguistic Review* 19.105-145.
- Hyams, Nina. 1986. *Language acquisition and the theory of parameters*. Dordrecht: D. Reidel.
- Legate, Julie & Charles Yang. 2002. Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review* 19.151-162.
- Piattelli-Palmarini, Massimo, ed. 1980. *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*. Cambridge: Harvard University Press.
- Pullum, Geoffrey K. & Barbara C. Scholz. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19.9-50.
- Pullum, Geoffrey K. 1996. Learnability, hyperlearning, and the poverty of the stimulus. Jan Johnson, Matthew L. Juge, and Jeri L. Moxley (eds.), *Proceedings of the 22nd Annual Meeting: General Session and Parasession on the Role of Learnability in Grammatical Theory*, 498-513. Berkeley Linguistics Society, Berkeley, California.
- Scholz, Barbara C. & Geoffrey K. Pullum. 2002. Searching for arguments to support linguistic nativism. *The Linguistic Review* 19.185-223.

GEOFFREY K. PULLUM  
University of Edinburgh