



Eastern Michigan University

[People & Organizations](#) ♦ [Jobs](#) ♦ [Calls & Conferences](#) ♦ [Publications](#) ♦ [Language Resources](#) ♦ [Text & Computer Tools](#) ♦ [Teaching & Learning](#) ♦ [Mailing Lists](#) ♦ [Search](#)

LINGUIST List 22.4631

Sat Nov 19 2011

Disc: Remarks by Noam Chomsky in London

Editor for this issue: Elyssa Winzeler <elyssa@linguistlist.org>

To post to LINGUIST, use our convenient web form at <http://linguistlist.org/LL/posttolinguist.cfm>.

Directory

1. Geoff Pullum , Remarks by Noam Chomsky in London

Message 1: Remarks by Noam Chomsky in London

Date: 14-Nov-2011

From: Geoff Pullum <gpullum@ling.ed.ac.uk>

Subject: Remarks by Noam Chomsky in London



E-mail this message to a friend

About a month ago (10 October 2011) Noam Chomsky spoke at an invitation-only seminar at University College London (UCL). I attended along with about 90 other British linguists. The announced title was: "On the poverty of the stimulus". The video of both the talk and the question period is available:

(http://www.ucl.ac.uk/psychlangsci/research/linguistics/news-events/latest-news/n_chomsky; henceforth, UCL video)

In what follows I summarize some of the content of Chomsky's London talk and its question session, and explain some of my reactions.

Chomsky's remarks in London were not very different in tone from things he has said elsewhere: the UCL presentation was extremely similar to a lecture given at Carleton University in Canada last April (<http://www.youtube.com/watch?v=XbjVMqOk3uc>), and echoed themes from Chomsky's talk at the symposium on the biology of language at the 2011 Cognitive Science Society conference in Boston last July, and journal articles such as "Language and other cognitive systems" (Chomsky 2011), and particularly the paper "Poverty of the stimulus revisited" (Berwick et al. 2011, henceforth BPYC-2011). These recent talks and papers share a steadfast refusal to engage with anything that might make the debate about the poverty of the stimulus (POS) an empirical one. They issue blanket dismissals of nearly all modern cognitive/linguistic science as worthless, and sweep aside whole genres of work on the basis of what seems to be extremely shallow acquaintance. Claims about parallels in the natural sciences feature prominently, as does a preference for authority over evidence. I will discuss a selection of topics, without attempting to be very systematic.

1. Rocks and kittens

Two aspects of the way Chomsky chose to deal with the topic of stimulus poverty struck me as startling. The first was that he stuck entirely with the version of the argument from POS that the late Barbara Scholz used to call the rocks-and-kittens version.

A child's pet kitten (so the argument goes), exposed to the same primary linguistic data as the child, learns no language at all, and is indistinguishable from a rock in this regard. Since the linguistic inputs are the same, an innate interspecies difference in language readiness and capacity for language acquisition must be involved; therefore linguistic nativism is true. (This is not parody, as I scarcely need to document: Chomsky has happily repeated his views on kittens and the like many times. A Google search on a pattern as specific as {Chomsky granddaughter rock kitten innate} will yield tens of thousands of hits, nearly all relevant ones. See Smith 1999: 169-170, or Stemmer 1999, or Chomsky 2000: 50 for quotable quotes in print.)

At UCL Chomsky didn't really give even this much of an argument: he just noted that humans had a genetic endowment that permitted them to learn language, and stipulated that he would call it Universal Grammar (UG). (Compare, e.g., "The faculty of language then is a special property that enables my granddaughter but not her pet kitten or chimpanzee to attain a specific I-language on exposure to appropriate data..." (<http://hotbookworm.wordpress.com/2010/01/11/noam-chomsky-the-biolinguistic-turn-lecture-notes-part-two/>.)

He even admitted that "intellectually ... there's just nothing to it --- [it's] a truism" (UCL video, 3:42); but he went on to argue that there is "a kind of pathology in the cognitive sciences" (UCL video, 4:24) in that its practitioners obdurately refuse to accept the simple point involved.

The real trouble, of course, is that everyone accepts it --- nobody doubts that there is something special about humans as opposed to kittens and rocks --- but they do not recognize it as a scientific result concerning human beings or their capacities.

What I had imagined would be under discussion in this seminar is the specific view about the character of human first language acquisition that is known as linguistic nativism. This is a substantive thesis asserting that language acquisition is largely guided by an intricate, complex, human-specific, internal mechanism that is (crucially) independent of general cognitive developmental capacities. This assertion seems to me worthy of serious and lengthy discussion. The rocks-and-kittens claim is surely not. We all agree that kittens and rocks can't acquire language, and that it's not because they don't get sufficient exposure. But that hardly amounts to support for linguistic nativism over general nativism (Scholz & Pullum 2002: 189).

It's not that Chomsky doesn't recognize the distinction between linguistic nativism and general nativism. He says (Chomsky 2000: 50, reproduced at (<http://www.chomsky.info/books/architecture01.htm>):

"Now a question that could be asked is whether whatever is innate about language is specific to the language faculty or whether it is just some combination of the other aspects of the mind. That is an empirical question and there is no reason to be dogmatic about it; you look and you see. What we seem to find is that it is specific."

But to say that you simply look and see, when the question is as subtle and difficult as this one and concerns mechanisms inaccessible to the tools we currently have, is surely not a responsible characterization of what science involves.

2. Stimulus poverty without the stimulus

The second striking choice Chomsky made was to address the poverty of the stimulus without ever mentioning the stimulus at all. This was POS without the S. One would expect that when someone claims that the child's input is too poverty-stricken to support language acquisition through ordinary learning from experience, they would treat empirical observations about the nature of that input as potentially relevant. It would give a POS argument some empirical bite if one could specify ways in which the child's input was demonstrably too thin to support learning of particular features of language from experience of language use. That would seem worthy of attention. The rocks-and-kittens version does not. I was very surprised that Chomsky stuck to it so firmly (though that does explain his lack of interest in the child's input: the rocks-and-kittens argument doesn't need anything to be true or false of the input).

The POS issue is going to take a long time to resolve if we can't even focus on roughly similar versions of the purported argument. Yet Chomsky regards it as crucial that it be resolved. He began his talk, in fact, with some alarmist remarks about the prospects for linguistics ("the future of the field depends on resolving it": UCL video, 4:38). If we do not settle this question of stimulus poverty, he claimed, we are doomed to seeing our subject shut down. So he portrays current skepticism among cognitive scientists about linguistic nativism as not just obtuse, but actively harmful, a threat to our whole discipline.

This is an interesting (if rather risky) new way of stoking enthusiasm for linguistic nativism: appeal to linguists' self-interest and desire for security (you don't want to be shut down, do you?). But it's hard to take seriously. Linguistics is not going to die just because a fair number of its practitioners now have at least some interest in machine learning, evolutionary considerations, computational models of acquisition, and properties of the child's input, and are becoming acquainted with probability theory, corpus use, computer simulation, and psychological experimentation --- as opposed to waving all such techniques contemptuously aside.

3. The lesson of Bayes' Theorem

Chomsky went on to remind us all of the linguists and psychologists in the 1950s who (allegedly) stuck so rigidly to corpus data that they regarded experiments going beyond the corpus data as almost a betrayal of science. And he stressed that the work of people today who work on Bayesian learning of patterns or regularities from raw data has no value at all ("zero results"). He compared their modeling of phenomena to physicists making statistical models to predict the movements of medium-sized physical objects seen outside in the street (UCL video, 36:41).

I think such a blanket dismissal overlooks a crucial conceptual contribution that Bayesian thinking makes to theoretical linguists, one that has nothing to do with the statistical modeling on which Chomsky pours such scorn. Many linguists have given the impression that they think it is impossible to learn from positive data that something is not grammatical. Lightfoot (1998: 585) suggests, for example, that although you can perhaps learn from experience that auxiliary reduction is optional in the interior of a clause, you cannot possibly learn that it is forbidden at the end of a clause; hence linguistic nativism has to be true. This reasoning is flawed, and Bayes' Theorem teaches us why.

The lesson is that probability of a generalization G being correct given a body of evidence E is not dependent merely on whether E contains crucial evidence confirming G over its rivals. The probability of G is proportional to the product of the antecedent probability of G's being true with something else: the probability that the evidence would look like E if G were true. That means that what is absent from experience can be crucial evidence concerning what the grammar has to account for. For example, all the thousands of times you've heard clause-final auxiliary verbs uncontracted strengthen the probability that they're not allowed to contract.

The argument from absence of stimulus is pretty much demolished by this Bayesian insight: the argument form simply is not valid. And for people who use the phrase "the logical problem of language acquisition" (as linguistic nativists

have been doing since 1981), that ought to mean something. It certainly seems to me sufficient to justify including at least a brief introduction to Bayesian statistical reasoning in the education of every theoretical linguist.

Suppose, though, that it ultimately turns out that the current fashion for constructing Bayesian computational models of learning is something of a dead end. It still doesn't follow that it is deleterious. Much can be learned by watching models ultimately fail. There is no threat to the discipline here: linguistics is not so fragile that it will collapse just because one possibly false trail was followed.

The people interested in Bayesian modeling and similar computational lines of research are smart enough to eventually perceive its inadequacy (if indeed it is inadequate), and will move to something that looks more interesting. People get bored in dead-end ventures. I talked to Roger Brown in 1968 and he told me that the reason he had abandoned Skinnerian behaviorism ten years before had nothing to do with any revolutionary new ideas in scientific thinking about cognition or the impact of Chomsky's famous review of Skinner: he was just bored with the work that behaviorism demanded, and wanted to try something more interesting. Intellectually agile people want to move on.

4. Bias at the NSF

About half-way through his talk, Chomsky made some claims about the probability of success with proposals to the NSF to fund research projects on Universal Grammar (UG). He said: "If you want a grant from the National Science Foundation, you better not include that [the phrase "UG"] in your proposal; it will be knocked out before it even reaches the review board" (UCL video, 30:35).

He warmed to this theme: "If you want to get a grant approved, you have to have the phrase 'sophisticated Bayesian' in it, and you also have to ask for an fMRI, especially if you have nothing whatever to do with it" (he chuckled here and there was general laughter) "... if you meet those two conditions, you might make it through the granting procedures" (UCL video, 31:02).

Then he returned to the claim that "UG" will doom your proposal: "But if you use a dirty word like UG, and you say there's something special about humans and we've got to find out what it is, that pretty much rules it out" (UCL video, 31:18). And then, with no chuckling, he added: "I'm not joking; I have concrete cases in mind ... of good work that just can't get funded, because it doesn't meet these conditions... Right at MIT in fact" (UCL video, 31:28).

Since award details are public information, it is trivial to find out whether the NSF is making awards for purely theoretical study of UG in a Chomskyan perspective. And it is. Željko Bošković's grant "On the Traditional Noun Phrase: Comparing Languages With and Without Articles" (BCS-0920888) is an example. And MIT is not left out. For example, David Pesetsky obtained Doctoral Dissertation Research grant no. BCS-1122426 for a project "Argument licensing and agreement"; the abstract begins: "Which properties of human language are universal, and which may vary across languages? Answering these questions will help us understand the unique human capacity for language, through which we hope to gain insight into the overall architecture of the human mind." And Chomsky must know that his co-author Robert Berwick received grant BCS-0951620 for a "Workshop on Rich Grammars from Poor Inputs" at MIT in 2009.

Naturally, many NSF proposals mentioning UG will go unfunded --- the majority, given that across the board less than 25% of grant proposals get funded. But (of course) proposals are sent out for peer review whether they mention UG or not, and whether they mention Bayes or not.

It seems a strange strategy to make claims of this sort to an audience of linguistics professionals in a foreign country who would have little knowledge of the NSF, and send out the message to young investigators internationally that following Chomsky's theoretical line will blight their careers by dooming their chances of NSF funding. Even if this were true, it would give the impression of a fractious field that has bad relations with its most important Federal funding

agency. But it is much stranger to make such statements when they are easily discovered to be false.

5. An uncomprehended question about machine learning

In the question period there was an extremely unfortunate interaction when the computational learning experimenter Alexander Clark tried to ask a question. Chomsky interrupted and began his answer before Clark had managed to make his point. The question Clark wanted to put was roughly the following (I knew enough to see where he was going, and he has confirmed to me that this was what he meant).

A paper Clark had published with Eyraud (2007) on learning some kinds of context-free grammars (CFGs) from positive data is dismissed in BPYC-2011 as useless. Chomsky repeated that dismissal in his talk. But Clark's more recent work has focused on languages in the much larger context-sensitive family that are generated by minimalist grammars as formalized by Edward Stabler. These are strongly equivalent to the Multiple Context-Free Grammars (MCFGs) that were invented by Seki & Fujii (1991), as Clark tried to begin to explain. He was not attempting to say anything about CFGs, but to raise the issue of learning the languages of minimalist grammars, or equivalently MCFGs. This is a wildly different class, vastly larger than the class of CFGs. It corresponds to the infinite union, for all natural numbers N , of a hierarchy of classes of languages (each definable in several ways) in which the first few steps are these:

$N = 0$ finite languages

$N = 1$ regular (finite-state) languages

$N = 2$ context-free languages

$N = 3$ tree adjoining languages

$N = 4$...

There has been much relevant mathematical work on these matters between 1984 and the present by people like Gerald Gazdar, Henk Harkema, Aravind Joshi, Greg Koble, Jens Michaelis, Carl Pollard, Kelly Roach, James Rogers, Edward Stabler, K. Vijay-Shanker, and David Weir (it is easily findable; I will not try to give even a brief bibliography here.) If Stabler has accurately captured the intent of the hints in the "minimalist program" about Merge and feature-checking, then minimalism embraces an enormous proper superset of the context-free languages. (I say "if" because Chomsky declines to refer to any of Stabler's work, so we don't know whether the formalization is acceptable as a precise reconstruction of the minimalist program as he conceives of it.)

Clark was trying to get Chomsky's reaction to recent results (see e.g. Clark 2010) exhibiting efficient algorithms for learning various subclasses of the MCFGs, including some fairly large classes going well beyond CFGs.

Chomsky interrupted the question and began to talk about CFGs. But he misspoke, and talked about having proved in 1959 that CFGs are equivalent to linear bounded automata (they aren't; LBAs are equivalent to context-sensitive grammars). Even if CFGs had been equivalent to LBAs, and even if Chomsky had been responsible for results on LBAs in 1959 (he wasn't, it was Kuroda five years later), CFGs had nothing to do with the observation Clark was trying to make about MCFGs. And Chomsky had in any case never proved any theorems about learnability, which was what Clark was trying to ask about. Clark's question not only was never answered, it was not even heard, hence of course not understood.

6. Languages evolving

After Clark's question, there were only a few more. I was lucky enough to be allocated time to ask two brief questions before the session ended. Chomsky had condemned language evolution work wholesale ("a burgeoning literature, most of which in my view is total nonsense": UCL video, 27:08), and I asked him to speak more directly about Simon Kirby's research on iterated learning of initially randomly structured finite languages, which he has shown leads to the

rapid evolution of morphological regularity.

Chomsky's answer was that it is not at all interesting if successive generations of learners regularize the language they are trying to learn: the regularity emerges only because human intelligence and linguistic competence is utilized in the task, and if you gave the same task to computers the same evolution would not happen.

Kirby's group has in fact addressed both those points, and both claims appear to be false. It seems to be the cognitive bottleneck of memory limitation that forces the emergence of regularity (decrease in Kolmogorov complexity) in the language over learning generations, not human linguistic capacity or intelligence (note the remark of Kirby, Cornish, & Smith 2008: 10685, that "if participants were merely stamping their own linguistic knowledge onto the data that they were seeing, there would be no reason we would find rampant structured underspecification in the first experiment and a system of morphological concatenation in the second"). And the effect of weak learning bias being amplified by cultural transmission through iterated learning does indeed turn up when the learner is simulated on a computer (see e.g. Kirby, Dowman, and Griffiths 2007).

There is an opportunity for substantive discussion here. And since both Chomsky and Kirby are invited speakers at the upcoming EvoLang conference in Kyoto (<http://kyoto.evolang.org/>), there will be a forum where it could happen. I hope it will. But maybe I'm too optimistic: I see the current integration of computationally-assisted cognitive science with careful syntactic description and theorizing as precisely what should inspire confidence that the language sciences in the 21st century has a bright future rather than spelling doom to linguistics.

7. Genetic fixity

The other topic I was able to ask about was the scientific plausibility of a view that has a remarkable genetic quirk arising between 50,000 and 200,000 years ago, giving a single developing hominid species an unprecedented innate UG that permits articulate linguistic capacities, and then remaining absolutely fixed in all of its details until the present.

A very few linguists (they include James McCawley, Geoffrey Sampson, and Philip Lieberman) have pointed out this prediction of genetically determined variation in UG between widely separated human groups. Lieberman notes that dramatic evolutionary developments like disappearance of lactose intolerance or radical alteration in the ability to survive in high-altitude low-oxygen environments can take place in under 3000 years; yet (as Chomsky stresses) the evidence that any human being can learn any human languages is strong, suggesting that UG shows no genetic variation at all.

Why would UG remain so astonishingly resistant to minor mutations for so many tens of thousands of years? There is no selection pressure that would make it disadvantageous for Australian aborigines to have different innate constraints on movement or thematic role assignment from European or African populations; yet not a hint of any such genetic diversity in innate linguistic capacities has ever been identified, at least in grammar. Why not?

Chomsky's response is basically that it just happened. He robustly insists that this kind of thing happens all the time in genetics: all sorts of developments in evolution occur once and then remain absolutely fixed, like the architecture of our visual perception mechanism. Human beings, he told me solemnly, are not going to develop an insect visual system over the coming 50,000 years.

This was his final point before his schedule required him to leave, and I had to agree with him (so let's not have any loose talk about kneejerk disagreement, OK?) --- we're not going to develop insect eyes. But I couldn't help thinking that this hardly answered the question. There are parts of our genome that remain identical for hundreds of millions of years, like HOX genes; but generally they cause catastrophic effects on the organism if incorrectly expressed. Even with the visual system, arbitrary changes could put an organism in real trouble. For widely separated populations of humans to have different constraints on remnant movement wouldn't do any damage at all, and it would offer dramatic

support for the view that there is a genetically inherited syntax module (though the "U" of UG would now not be so appropriate).

So it was just as with the rocks-and-kittens POS argument: I agree with the starting observations, as everyone must; but the broader conclusions that Chomsky defends, and more generally his extremely negative attitude to computer simulation work, human-subject experimentation, evolutionary investigations, and data-intensive research don't seem to follow.

I am not pessimistic enough to believe that contemporary experimental research in the cognitive and linguistic sciences --- Bayesian and connectionist work included --- will prove to be some kind of toxic threat to our discipline. I think it represents an encouragingly lively and stimulating contribution. I think we have a responsibility as academics to acknowledge such work and do our best to appreciate its methods and results. It won't do anything clarify our understanding of language if we simply condemn it all out of hand.

Geoff Pullum
University of Edinburgh

REFERENCES

Berwick, Robert; Paul Pietroski; Yankama; and Noam Chomsky (2011). [BPYC-2011] Poverty of the stimulus revisited. *Cognitive Science* 35: 1207–1242.

Chomsky, Noam (2000). *The Architecture of Language*. New Delhi: Oxford University Press.

Chomsky, Noam (2011). Language and other cognitive systems: What is special about language? *Language Learning and Development* 7 (4): 263-278. <http://dx.doi.org/10.1080/15475441.2011.584041>

Clark, Alexander (2010). Efficient, correct, unsupervised learning of context-sensitive languages. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, 28-37. Uppsala, Sweden: Association for Computational Linguistics. <http://www.cs.rhul.ac.uk/home/alexc/papers/conll2010.pdf>

Clark, Alexander, and Remi Eyraud (2007). Polynomial time identification in the limit of substitutable context-free languages. *Journal of Machine Learning Research*, 8, 1725–1745.

Kirby, Simon; Michael Dowman; and Thomas Griffiths (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104 (12): 5241-5245.

Kirby, Simon; Hannah Cornish; and Kenny Smith (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105 (31): 10681-10686.

Lightfoot, David (1998). Promises, promises: general learning algorithms. *Mind and Language* 13: 582-587.

Seki, Matsumura and Kasami Fujii (1991). On multiple context-free grammars. *Theoretical Computer Science* 88: 191-229.

Smith, Neilson Voyne (1999). *Chomsky: Ideas and Ideals*. Cambridge: Cambridge University Press.

Stabler, Edward (1997). Derivational minimalism. Christian Retore, *Logical Aspects of Computational Linguistics (Lecture Notes in Artificial Intelligence, 1328)*, 68-95. Berlin: Springer Verlag.

Stemmer, Brigitte (1999). An on-line interview with Noam Chomsky: On the nature of pragmatics and related issues. *Brain and Language* 68 (3): 393-401.

Linguistic Field(s): Cognitive Science
Computational Linguistics
Discipline of Linguistics

[Read more issues](#)[|LINGUIST home page](#)[|Top of issue](#)

Page Updated: 19-Nov-2011



[About LINGUIST](#) | [Contact Us](#)



While the LINGUIST List makes every effort to ensure the linguistic relevance of sites listed on its pages, it cannot vouch for their contents.