Understanding the allure and pitfalls of Chomsky's science.  

Review by Gary Lupyan –  
Department of Psychology  
University of Wisconsin, Madison  
lupyan@wisc.edu

Abstraction, functioning in this way, becomes a means of arrest far more than a means of advance in thought. It mutilates things; it creates difficulties and finds impossibilities … the viciously privative employment of abstract characters and class names is, I am persuaded, one of the great original sins of the rationalistic mind” (James, 1909).

When I was in graduate school, I chanced on a book by Howard Kunstler, The Geography of Nowhere: The rise and fall of America’s Man-Made Landscape (1994). The book is a history and critique of suburbanization of the United States. In Kunstler's words, “the ghastly spectacle of construction and destruction that converted a lovely, verdant, beckoning New World continent into a wilderness of free parking.” What I found most appealing about it was that it helped bring into greater focus my vague dissatisfaction with the suburban landscape. It not only helped answer questions about how the landscape came to be that way, but it put its finger on precisely what it is about the landscape that leads to this dissatisfaction. Chris Knight's Decoding Chomsky (Yale, 2016) does much the same for helping understand how Noam Chomsky became the most highly cited person alive today, analyzing the historical and intellectual landscape that led to Chomsky being compared variously as Galileo, Newton, and Einstein, voted as the world’s top public intellectual in 2005. He was even crowned royalty. Before a 2014 address at the Vatican, Chomsky was introduced as "one of the princes of linguistics". Knight counterposes this acclaim with the strangeness of the ideas Chomsky has championed since the 1950s and makes a compelling case for the scientific vacuousness of these ideas.

It is routine for Chomsky to be hailed as the person who solved how language works. For example, writing for the New Yorker, Anthony Gottlieb singles out David Marr’s work on vision and Noam Chomsky’s work on language as “the most solid … accounts of mental mechanisms” (2012). The culture critic David Golumbia writes that not only Chomsky redefined the entire discipline of linguistics, “but his work has been something close to definitive in psychology, philosophy, cognitive science, and even computer science” (Golumbia, 2009, cited by Knight on p. 2).

And so it comes as a shock to the uninitiated that the version of language that Chomsky is supposed to have solved, bears no relationship to language as understood not only by laypeople, but most practicing language researchers. Beginning in the opening chapters of the book and woven throughout the later ones, Knight lays out the full strangeness of Chomsky’s vision. Is language primarily for communication? Did language evolve? Do children need to be
spoken to (or signed to) in order to become competent language users? Is language a social product? If you answered ‘yes’ to any of these questions, you fundamentally disagree with Chomsky on the definition of language.

Language, on Chomsky’s view is reduced to an innate biological (though not gradually evolved) Universal Grammar (UG). Contra a common misunderstanding, UG is not the set of features that all languages have in common (a search for such linguistic similarities was the goal of Joseph Greenberg’s school of linguistics, which Chomsky rejected). Rather, universal grammar is a set of computational properties that make it possible for people to learn (all) natural languages and to produce infinitely many utterances with a finite brain and a finite amount of experience. Stipulations about what is part of UG have changed radically during the course of Chomsky’s career, but as Knight makes clear, the commitment to UG as the correct way to study language has remained. It is plainly stated in recent reviews of the generative approach, e.g., (Everaert, Huybregts, Chomsky, Berwick, & Bolhuis, 2015). Chomsky’s focus on grammatical competence as the subject of study meant denying the reality of what the rest of language researchers term ‘language’. Drawing on Chomsky’s writings and quoting him heavily, Knight describes how according to Chomsky “the technical term “language” has no relation at all to the pre-theoretical term “language.” Chomsky asserts that while UG is “something real, it is in your head, it is in my head, it is physically represented in some fashion,” “what is now ‘language’ does not need any term at all, because it is a totally useless concept … It does not fit with linguistic theory, it has no existence” (p. 201).

Readers would be justified in thinking that this sounds like doing science by decree. Is Knight exaggerating? It would be a mistake to generalize the critiques of Chomsky to the field of generative linguistics at large, which contains many linguists who have distanced themselves from Chomsky (particularly in the wake of the Minimalist program), but it is Chomsky rather than generative linguistics at large that is at the center of Decoding Chomsky, and Knight’s characterization of Chomsky’s modus operandi seems valid.

Consider, for example the program of University of Edinburgh’s Language Evolution and Computation Unit (e.g., Kirby, Cornish, & Smith, 2008; Kirby, Dowman, & Griffiths, 2007; Kirby, Griffiths, & Smith, 2014; Thompson, Kirby, & Smith, 2016). This work turned the dogma that the “stimulus is poor therefore children must rely on an innate language acquisition device” on its head by showing that it is because children have limited memory and are exposed to only a subset of utterances they will need to produce, that led to the emergence of grammar in the first place. Basic design features of language such as compositionality are, on this view, cultural adaptations. Like the related work of Christiansen and Chater (2008), this approach attempts to explain the evolution of linguistic structures as languages adapting to the human brain rather than the brain evolving—or on Chomsky’s view, forming in a single mutation—Universal Grammar. Berwick and Chomsky dismiss the Edinburgh approach with a casual, but telling remark: “In brief, [this work] does not appear to deal with the nature of the language faculty as we construe it here, and hence has nothing to say about the evolution of language” (Berwick & Chomsky, 2016). By the “language faculty
as we construe it here”, Berwick and Chomsky are referring to UG. The idea that UG “can serve as the object (and the sole object) of a truly scientific study of language” is, according to Knight, “the foundational error at the root of all Chomsky’s other intellectual contradictions and difficulties” (p. 237).

**How did it begin? The original allure of Chomsky’s vision.**

The work that propelled Chomsky to stardom was the 1957 publication of *Syntactic Structures*. It was, Knight writes, “the snowball which began the avalanche of the modern “cognitive revolution”. .... ‘In the beginning was *Syntactic Structures*’ (p. 14). Knight points out that both *Syntactic Structures* and the similarly influential *Aspects of the Theory of Syntax* were funded in part by grants from the military. Knight asks two questions: First, why did Chomsky (“an outspoken anarchist and anti-militarist”) take the money? A second more interesting and pertinent question is, “what did the military think they were buying?” (p. 16). Quoting some of the original Air Force backers of the work, Knight argues that the military “sponsored linguistic research in order to learn how to build command and control systems that could understand English queries directly” (p. 17). It’s not that the military thought Chomsky would deliver a product that would enable some kind of thought to machine-code translator. Rather, Chomsky’s vision was attractive because it promised to “reduce the amount of knowledge needed to understand the field.” (p. 18). Rather than having to bother with details of specific languages and cultures, language could be reduced to pure, culture-free computation. By establishing a division between competence and performance, all “imperfections” of language (that is, aspects that were not well fit to the theory) could be ascribed to performance, with *competence* remaining an object of a purely naturalistic science.

Several chapters of the book are devoted to describing the intellectual climate that made this vision of language so appealing. In one of my favorite passages of the book, Knight quotes the mathematician Warren Weaver, envisioning—in 1955—a kind of Babylonian antitower. Weaver imagines people living in a series of tall closed towers and attempts to communicate between the towers can only achieved with great difficulty. “But, when an individual goes down his tower, he finds himself in a great open basement, common to all the towers. Here he establishes easy and useful communication with the persons who have also descended from their towers ... the way to translate from Chinese to Arabic, or from Russian to Portuguese, is not to attempt the direct route, shouting from tower to tower. Perhaps the way is to descend, from each language, down to the common base of human communication – the real but as yet undiscovered universal language” (p. 55). Chomsky never strove for the development of a universal language and did not share Weaver’s enthusiasm for machine translation. But enough people in the 1950s had this dream to make Chomsky’s research program seem like the perfect fit for turning it into reality.

**Where are the data?**

Readers may naturally ask: Surely, Chomsky and his collaborators have offered empirical support for the reality of Universal Grammar? After all,
Chomsky frequently reiterates that this is the only truly scientific approach to the study of language. Science requires data. What are the data that have been offered in support of Chomsky’s theories? This is one area in which I wish *Decoding Chomsky* offered additional details as it would help strengthen Knight’s argument that Chomsky’s ideas lack empirical support. Some recent articles address some of the empirical shortcomings in greater detail (Edelman, in press; Evans & Levinson, 2009; LaPolla, 2015; Lin, in press; see Ibbotson & Tomasello, 2016 for a discussion aimed at a more general audience). Chomsky’s supporters tend to dismiss such critiques on grounds that the authors simply do not understand generative linguistics (see, e.g., the commentaries to Evans and Levinson, 2009).

As someone studying language outside the generativist tradition, what has always struck me about the generativist approach to data is that (1) the only data offered seem to be in the form of sentence $X$ is grammatical while sentence $X'$ is not, and (2) it is the job of alternate approaches to show how they could address the theoretical constructs of the generativist approach. Here is an example from a recent review paper by Everaert et al. (2015). Their central argument is that approaches to language that make use of ordered strings of words/morphemes can never succeed. Why not? Because, argue the authors, it is only by analyzing language using the generativist approach that one can understand phenomena like “parasitic gaps.” A parasitic gap (PG) is defined in the article’s glossary as: “A gap (a null variable) that depends on the existence of another gap RG [real gap], sharing with it the same operator that locally binds both variables. PG must conform to a binding condition asserting that PG cannot be c-commanded by RG.” (p. 732). It is this phenomenon that is supposed to explain why the sentence “Guess which politician your interest in Jane clearly appeals to” is grammatical while the sentence “Guess which politician your interest in clearly appeals to Jane” is not. (If the grammatical sentence strikes you as no more comprehensible than the ungrammatical one, worry not, you are simply the victim of processing constraints).

There are two key problems with such data. First, the methods used to collect and analyze grammaticality judgments are characterized by a “deplorable … lack of rigor” (Schütze, 2016; see also Grandy, 1980). Typically, there is no systematic collection of grammaticality judgments and no statistical analysis. In other words, there is no attempt to do serious data collection in the one area that is supposed to provide empirical support for the theories. In an attempt to find out why this is the case, Schütze reached out to Chomsky, who replied that research practices in linguistics ought to follow the natural sciences where “almost no one devotes attention to ‘methodology’” (Schütze, 2016). I have a hunch that ‘natural’ scientists would disagree.

The second problem (which helps to explain the first) is that even if we take grammaticality judgments as the behavior to be explained (which is rather strange in itself), *behavior* should be the target of explanation, not theoretical constructs like parasitic gaps. Suppose that an alternative to explaining the pattern of grammaticality judgments is offered based on this or that domain-general psychological principle or an analysis of language statistics or
differences in learnability of one kind of construction or another. The response by Chomskyan linguists to such demonstrations tends to be “but this does not explain parasitic gaps.” This is precisely the argument of Everaert et al.: “Applying analytical or statistical tools to huge corpora of data in an effort to elucidate the intriguing properties of parasitic gaps will not work.” (Everaert et al., 2015, p. 735). Why should the goal be to account for the theoretical construct that is a parasitic gap? In the absence of independent evidence that a parasitic gap is something real, theories of language are not obliged to explain it!

Such missing evidence concerns far more than esoteric constructs like parasitic gaps. What is the independent evidence for the reality of empty categories, C-command, even the notion of movement itself? If it becomes possible for a machine to parse natural language without the use of these constructs, as it is increasingly the case, does it not show the superfluousness of these constructs (see Norvig, 2011 for a discussion relevant to this point)? Knight cites the linguist Frederick Newmeyer, as saying that the proof of Chomsky’s success lies not in any evidence that his theories actually worked but in the fact that “anyone who hopes to win general acceptance for a new theory of language is obligated to show how the theory is better than Chomsky’s” (p. 180). It is an unhealthy state of affairs if the test of alternate theories is to see how well they can explain Chomsky’s constructs rather than how well they can address empirical phenomena.

Once upon a time, people thought that the burning of substances released phlogiston. Phlogiston was used to explain why some substances became lighter when burned and what made some metals rust. In time, our understanding of oxidation reactions made phlogiston (a theoretical construct) unnecessary for explaining observable phenomena like burning and rusting. Chomsky’s rejection of non-generative approaches to studying language because these fail to explain constructs like parasitic gaps is akin to rejecting modern chemistry because it has failed to isolate phlogiston.

Politics and Science

Although Decoding Chomsky is primarily focused on Chomsky’s science, Chomsky’s role as a public intellectual is linked to his political activism. When asked, Chomsky denies there is any connection between the scientific and political persona, remarking sometimes that the linguistics takes away time from what really matters (Horgan, 2016). The distinction between Chomsky the scientist and Chomsky the activist is a stark one. Chomsky the scientist believes that Language (scientifically understood) is devoid of communicative intent, social meaning and “anything else which the rest of us would associate with language” (p. 136): “While the scientist says language is not for communication at all, the ordinary human Chomsky uses language precisely to communicate — to denounce his own state, his own government, his own employers, his own institutional milieu… opposing just about everything which he embodies in his alternative role” (p. 136). In order to understand the peculiarities of the science”, writes Knight, “we must understand the political commitments against which it
has always been counterposed" (p. 130).

Knight believes that Chomsky's politics and activism are indeed kept separate by what he refers to as a “firewall” erected by Chomsky and “designed to separate ‘science’ from any kind of social or political activism” (p. 193). In perhaps the most provocatively titled chapter in the book — *Mindless activism, tongue-tied science* — Knight presents a compelling argument that Chomsky’s activism, barred from drawing on the scientific method becomes, by design, mindless and scientifically illiterate (which according to Knight would “prove a disaster for the global revolutionary left”, p. 200). At the same time, in an effort to be “naturalistic”, the science expunges all aspects of culture and socialization as outside its purview and is consequently *tongue-tied*, having nothing to say about politics. On this model, “you are either a scientist or an activist; you cannot play both roles at the same time. … A climate scientist, for example, will be respected for reporting worrying findings, but condemned for resorting to direct action to avert the consequences. Those who do confuse roles in this way risk being accused of betraying their vocation” (p. 197). That this may appear entirely normal to current scientists is precisely Knight's point. Knight suggests that the current separation between science and activism is one of Chomsky’s legacies.

Politics aside, there is a second sense in which Chomsky’s science is tongue-tied. By defining language as an idealized grammatical competence which cannot be studied using normal scientific methods, the science becomes dedicated to solving problems of its own making, having nothing to say about the kinds of scientific questions that everyone else cares about. As Robin Tolmach Lakoff argues in *The Language War*, accepting the generative approach to studying language means “accepting the impossibility of saying almost everything that might be interesting, anything normal people might want or need to know about language” (2000, p. 7).

The road ahead: Chomskyan linguistics versus modern language research

Reading *Decoding Chomsky* may give the impression that the state of modern language research is decidedly poor: that linguistics and the language sciences are dominated by a powerful figure whose intuitions of what a theory ought to look like “[lead] an army of people to go out and reanalyze everything to conform to that intuition” (Pinker, as cited by Kenneally, 2008 p. 271; Knight, p. 179) (Of course given that the data are largely introspective judgments about grammaticality, reanalysis can simply involve adjustments of one’s grammatical intuitions).

I am saddened by the brilliant minds who have dedicated themselves to trying to resolve the specific problems posed by Chomskyan linguistics (of the why is X grammatical and *X is not variety), given that so many of these problems appear to be the field’s own making. Chomskyan generative linguistics seems to be an abject case of what William James called “vicious abstractionism”. It is what happens when we single out “some salient or important feature [of a phenomenon] and instead of adding to its previous characters all the positive consequences which the new way of conceiving may bring, we proceed to use
our concept privatively … treating it as a case of ‘nothing but’ that concept, and acting as if all the other characters from out of which the concept is abstracted were expunged.” This, according to William James, is one of the “great original sins of the rationalist mind” (James, 1909/2008). It is difficult to see a future for a scientific study of language as a grammatical competence which did not evolve and does not lend itself to empirical investigation aside from casual reliance on grammatical intuitions of linguists (Schütze, 2016; see also Massaro, 2017).

In contrast, the state of modern language research—at least from where I stand—looks very different. There is a vast chasm between the self-referential program of Chomskyian linguistics and modern research on just about every aspect of language that is happening outside the Chomskyan fold. Research in linguistic typology is being standardized and unified (e.g., http://wals.info; http://glottobank.org), and the ability to look at the full diversity of human languages is enabling us to draw richer inferences about the human language capacity (Dunn, Greenhill, Levinson, & Gray, 2011; Lupyán & Dale, 2016), the study of language history is being made more rigorous by the application of quantitative phylogenetics (Gray & Atkinson, 2003; Gray, Drummond, & Greenhill, 2009). Combined psycholinguistic data and computational models are helping to show how more abstract grammatical knowledge emerges from experience with specific utterances (Chang, Dell, & Bock, 2006). Theories of language comprehension and production are being integrated with theories of memory and motor control (MacDonald, 2013). Modern approaches to distributional semantics make the idea of learning word meanings through word use increasingly plausible (e.g., Bruni, Tran, & Baroni, 2014; Mikolov, Chen, Corrado, & Dean, 2013). There is also a growing excitement about comparative and computational approaches to the study of cultural evolution, and understanding the relationships between the evolution of cooperation and language (e.g., Henrich, 2015; Kirby et al., 2014; Oller & Griebel, 2004; Smith, 2010; Tomasello, 2008).

Predictably, Chomsky believes none of this work has any relevance for understanding Merge (the latest of many formulations of UG), making it irrelevant to our understanding of language (Berwick & Chomsky, 2015; Everaert et al., 2015)—an opinion most practicing language researchers fortunately ignore. Applying the scientific method to questions decreed by Chomsky as irrelevant and unscientific is paying dividends. For example, Chomsky’s repeated assertion that the input children receive does not matter because language is not something children learn, it is something that happens to them “like puberty” (e.g., Chomsky, 1987) led researchers to ignore, for many decades, the relationship between language input and language outcomes (see e.g., Bates et al., 1994 for an important exception). Of course children’s language comprehension and production are enormously affected by input (Hart & Risley, 1995; Hoff, 2003; Fernald, Marchman, & Weisleder, 2013)—an issue of significant public importance. Anyone insisting that it is linguistic competence that is independent of language input needs to explain why competence in the absence of performance matters and provide empirical evidence for truly preserved competence in the face of truly compromised input.
Conclusion

When questioned about the progress of generative linguistics, Chomsky has often remarked that linguistics and cognitive science are in a pre-Galilean state, with thinkers beginning to formulate the questions in the right way and that “someday someone is going to come along and say ‘Look, you guys, you’re on the right track, but you went wrong here. It should have been done this way. Well, that will be it. Suddenly things will fall into place.’” (p. 178). This quote is taken from an interview conducted in 1983 (Chomsky, 2003). A nearly identical statement appears in an interview with The Atlantic in 2012 (Katz, 2012), and in a lecture at Princeton where he attempted to sum up 60 years of generative linguistics (Chomsky, 2014). In Knight’s words, “With each new disappointment, [Chomsky] turns with undimmed optimism toward the future — to a moment of revelation, when, quite suddenly things will fall into place.” (p. 174). In an especially vivid assessment of Chomsky’s many versions of UG, Knight is “reminded of a man on the doorstep fumbling with is key in the half light. He … turns it this way and that. Despite all his fumblings, the lock just will not yield. To those watching, the most likely explanation is that he’s got the wrong key.” (p. 178).

The future of linguistics and cognitive science may indeed look very different from its present. Our intellectual descendants may see the present period as primitive. Pre-Galilean even. But what are the chances that future scientists will confirm that the key to understanding language lies in formulating language in a way that strips away all that makes it language? Is it believable that a view that according to Chomsky is “obvious to any thinking person” (Chomsky, 2014) is true and yet has failed to produce empirical evidence outside of an ever-shifting theoretical framework of its own making? I wouldn’t bet on it.

References


