Language and the Miracle Creed

Noam Chomsky

Abstract

Since Massimo Piattelli-Palmarini’s coining of the term biolinguistics in 1974, biolinguistics research has attracted numerous attentions from the linguistic circle and its neighbouring fields, triggering various ways of pursuing it. The current paper makes clear the origin, significance, and implications of biolinguistics, especially, the reason why the biolinguistics enterprise realizes the miracle creed firstly assumed by Albert Einstein. In this way, the controversies and mis-conceptualization of biolinguistics in current linguistic literature can be clarified, paving the solid foundation for future biolinguistics research.

Keywords: biolinguistics; miracle creed; basic property of language; structural dependence; language acquisition

Affiliation

University of Arizona and MIT, USA
email: chomsky@mit.edu
Now we will begin by just a word about the term biolinguistics, its origin, significance, implications. The term was introduced about 50 years ago at an international conference on cognitive science by Massimo Piattelli-Palmarini, a biologist (Chomsky 2007). He used the term to refer to the generative programs in study of language that had been developing for approximately the previous 20 years. The term is intended to indicate that this approach to language regards language as a biological property of a particular person (Berwick and Chomsky 2016). For example, it’s a property of me that I have knowledge of English and use it in production and perception, but not of Chinese. And this knowledge of English is coded in my brain in some fashion. It’s acquired in the course of growth and development, and it’s used in various ways (Chomsky 2016).

Now, it seems like a truism to describe my language as a property of me coded in my brain. And why even have a term for such an approach? It should be a truism. But in fact, it was an innovation at the time. It still remains at the fringe of the investigation of the study of language and cognition in many ways. The structuralist approach to language was based on Ferdinand de Saussure’s early classical work where he defined language as a social product, a social contract, an agreement among individuals of a community to behave in a certain way (Chomsky 2020). And pretty much same view was taken by Leonard Bloomfield, the leading figure in American Linguistics in the early half of the twentieth century. Among various descriptions, he described a language as a set of utterances that can be used in a community. So it’s a community product. It’s acquired by training. And it becomes a set of habits within the prevailing behavioural framework (Chomsky 2021a). You look at philosophy of language. I won’t run through the details, but the dominant views in the twentieth century that up until today have been that we must consider language as a product of a community, can’t consider it to be an individual phenomenon. So biolinguistics separates itself from that.

Again, in my view, it should be truism. This recognition that having a particular language is a biological property, coded in the brain somehow, does not reject other approaches to language. It simply serves as the foundation for them. You can study communities which can be set up in various ways, and the common or partially shared languages among them. But this is all based on what we know and understand about the individual, the languages of an individual coded in that individual’s brain. We can study acquisition of language, use of language, evolution of language only to the extent that we understand the phenotype, the actual system in an individual, largely shared in many different ways among individuals.

So, we study within this framework, the biolinguistics framework. We study language much the way you study other capacities of the organism, say, the
visual capacity. We want to know how the human, a particular individual, has a particular visual system. We want to know what it is. We want to know the innate structure that determines that you and I have mammalian visual systems, not insect systems, want to study whatever role it has in the growth and development of the system, what you can say about its evolution, so on. But basically, we study vision; we study language much in the same way. And notice that the study of vision is overwhelmingly a study of computational properties. So, for example, one famous result is that successive presentations of a few dots on the screen will give the perception of a rigid body in motion (cf. Ballard and Kimball 1983). The computations involved in that outcome have been studied in some depth and are by now well understood. Nothing is understood, in this case, about the basis for it in the brain. That’s quite a hard problem. And indeed, most of the study of vision is at the computational level, trying to find out how the system functions and operates, whether it’s the mammalian system, the insect system. In fact, other forms of reaction of organisms to light.

The same in the case of language. Now language is different from vision in a number of critical respects. For one thing, vision is an input system. Language is not. Language is an internal system. We have knowledge of language, and it’s used in an input system, perception, parsing. But it’s also used in an output system, production – what I’m doing now. That’s not true of vision. Most of the interesting questions about the neurology of vision are so far much too hard to study. Nevertheless, we know something about the neurology of vision, some quite interesting things in fact. The reason we know that is because invasive experiments have been undertaken with other mammals, cats, monkeys by putting an electrode into the striate cortex to see what reactions there are to controlled signals. We can’t do that with humans for ethical reasons (whether it’s ethical for other animals we may debate). Since we have essentially the same visual system as other mammals, we know a lot about the human visual system. Again, in the case of language there is no way to do this. We don’t do the invasive experiments with humans. There are no other species. Language, unlike vision, is species-specific. It’s a property of humans. There’s nothing analogous in other organisms. In fact, in many respects humans are a singular creature, quite unlike others. That raises interesting questions for the theory of evolution, theory of the nature of the brain, and so on. But it’s a fact. Human language is a species property, genuine species property. The faculty of languages is as far as we know common to all languages, all humans, however remote the human you can find, a tribe in New Guinea which hasn’t had other human contact (Hauser et al. 2002). If an infant from the tribe is raised in Beijing it will speak Chinese and can become a linguist or quantum physicist or wherever life takes them. Linguistic and other capacities seem to be common to humans, but in fundamental respects without analogue in other species.
Evolutionary history tells us that it is a very recent development, that humans have been on earth for a couple of hundreds of thousands of years. We know from genomic analysis that humans began to separate a very short time after they appeared, maybe 120,000–150,000 years ago. That’s nothing in evolutionary time. Among the various groups that separated, their descendants have the same language faculty. So the language faculty was in place before the separation. Before the appearance of modern humans, there’s almost no evidence in the archaeological record of any significant symbolic activity. Not long after their appearance (again, in evolutionary time), there is very sophisticated symbolic activity and evidence of complex social organization (cf. Tattersall 2010). So, a plausible conclusion, not demonstrated, but plausible, is that human language developed pretty much along with modern humans.

Come back to the study of the neural basis for language. It is far more difficult than the study of the neural basis for vision. We simply can’t experiment with humans in the obvious ways. There are no other species which can be studied as we can in the case of vision. For vision study of the brain is hard enough. It is multiply hard in the case of language. However, there are some quite interesting and striking results. If there’s time, I’ll go into some of them, mention some of them. That’s biolinguistics.

What are we interested in when we pursue it? The same as in the other sciences. Study of vision, study of chemistry. Whatever we’re investigating, the goal of inquiry is explanation. We want to understand the world. There’s a difference between explanation and description, crucial difference. Description tries to tell us this is the way things are. That’s hard enough, but explanation goes far beyond. It tries to say this is why things are this way and not some other way. That second part, not some other way, is very crucial, often overlooked. That gives a test of whether we actually have explanation. This sounds elementary.

So why discuss it? The reason is it’s too often overlooked. I’ll give two examples, one from the earliest days of the modern cognitive sciences and the generative enterprise. Another from today. There are many others in between, but these will illustrate. So, let’s go back to the 1950s, and earlier, what’s language? As I mentioned, for Leonard Bloomfield, language is a matter of habit and training. It’s the collection of expressions used in a speech community. We go to the leading figure in philosophy of language, W. V. O. Quine of Harvard and it’s pretty much the same: language is a complex of dispositions acquired by conditioning, and the language is the set of sentences used in the community (Quine 1960). Linguistics at the time was described as a taxonomic science. It’s a descriptive science. You give descriptions of language based on procedures, carefully worked out procedures that can be applied to a corpus of materials to provide a structural analysis. These can be applied to any corpus of materials.
There are no explanations. None is required; none is possible even. There’s no ‘why not that’. You can apply the procedures to anything.

Now there was a rationale for that. It wasn’t overlooked. The rationale was, and it was explicitly stated, that languages can differ from one another virtually without limit, and that each language has to be studied without presupposition. This was formulated famously by Martin Joos in the mid-1950s, a leading theoretician of language, a physicist in his background. If that’s the case, then there are no impossible languages. And it’s fine if the procedures can apply to any corpus.

I should mention that similar ideas were held at the time about organisms. It was held very commonly among biologists that each organism has to be studied on its own. Organisms can vary in virtually any possible way. There’s little general to say about them. In biology, this has all collapsed, totally. By now, it’s even considered plausible to speculate that there might be a universal genome, single genome for all of life (cf. Sherman 2007; Mao et al. 2020). Minor variations give the superficially different organisms. There was very interesting classical work by D’Arcy Thompson, by Alan Turing, which have entered into modern biology, providing laws of form, general principles by which organisms can grow and develop and that determine their nature.

I think something similar is happening in linguistics. And we will talk about it as we proceed.

Well, that’s an example from the early days of modern cognitive science and the generative enterprise. None of this was sustainable. The procedures of analysis cannot give you the elements that enter into explanatory theories of the language or of the faculty of language, and training has almost nothing to do with acquisition and the nature of language. So the whole story collapses pretty much as it did in biology in the same years.

At the time, it looked as if basically everything was known. No problems. The procedures can apply to any corpus. As soon as the effort to develop explanatory systems was undertaken, to develop generative grammars – which are a weak kind of explanatory theories – it turned out that we know almost nothing. Every sentence is a problem. That’s the beginning of modern cognitive science, the generative enterprise. That was then. Let’s come to today.

A lot of you, I’m sure, are familiar with the extensively reported work on language modelling, like the GPT (Generative Pre-Trained Transformer) theories. There’s lot of excited discussion in the media and the science journals about how these new modelling systems based on deep learning can produce language pretty much the way humans do (cf. Floridi and Chiriatti 2020). They may surpass us. Maybe they’re sentient. Maybe they’ll develop further to be kind of super intelligences. You see this all over the place. I’m sure you’ve seen them.
There is a good deal of work on the failures of these models. They can't get this kind of sentence. They make mistakes on some other kind of sentence. These are efforts to show that the models are too weak. The response to that is standard. We'll just add another trillion parameters and will double the number of terabytes of data scanned. Bring in a couple of supercomputers. Maybe we can do better. Well, that enterprise can go on forever, but there's a much simpler objection. The systems are much too strong. They do not answer the question: Why not that? As a result, the methods, if you look at them, work just as well for impossible languages – we now have a clear conception of that (Chomsky and Moro 2022) – just as well as they do for actual languages. So essentially, they're telling us nothing about language, because they do not explain 'why not that'.

A lot of you might ask why to undertake the enterprise at all, since it tells us nothing about language. That's clear, also nothing about cognition. Does it produce anything useful? Actually not. There is work with deep learning approaches that does provide things that are useful. I happen to be hard of hearing. So in a conference like this, I use live transcription, which is achieved by brute force. It tells you nothing about translation, transcription, or anything else, but it's useful. The modelling systems like GPT aren't even useful. Their only function is to waste a lot of California's energy. That's going to be true no matter how many trillion parameters they add and how many times they double the number of terabytes of data. This is something that you should pay attention to. Basically useless activities instead of trying to answer the questions 'Why this and why not that? What's the nature of cognition, language, in particular?' And the public is very seriously misled. So a good deal of caution is necessary in that connection.

Let's return to the distinction between description and explanation, which should be obvious and uncontroversial even though it isn't in the actual world. For science, for efforts to understand the world, careful description is needed. But as a preliminary to the search for explanation.

What would be a real explanation for language? Let's take a crucial property of language a very surprising one, a fundamental feature, the property that's called technically structure-dependence. I'll just give a couple of examples to illustrate.

So suppose I say, 'the man who fixed the car carefully packed his tools'. Notice that the sentence is ambiguous. Could mean 'fixed the car carefully' or 'carefully packed his tools'. A description of the language, the generative grammar, will have to state that fact. Now let's take the word carefully and put it in the front of the sentence. 'Carefully, the man who fixed the car packed his tools'. Now it becomes unambiguous. It means he carefully packed his tools. That's surprising. The word 'carefully' has to find some verb phrase to modify, but it doesn't use the simplest procedure, namely, find the closest verb phrase. Finding the
closest is an easy cognitive operation. Infants can do it. We do it all the time, but you don’t use it here. You find the most remote verb phrase ‘packed his tools’, which happens if you think about the structure of the sentence to be the structurally closest verb phrase.

What we do is to ignore the linear order of words. We’re relying solely on operations that apply to abstract structures. Think what that means. It means we ignore 100% of what we hear: words in linear order. We attend only to what our mind constructs. Furthermore, this is known by infants as early as they can be tested, less than two years old (cf. Shi et al. 2020).

Give you just one more example to illustrate. Take the sentences, ‘The bombing of the cities is a crime’, ‘The bombings of the city are a crime.’ There is something surprising there too. Agreement of the copula does not depend on the closest noun phrase. We ignore adjacency, the simplest rule. We ignore adjacency in determining the inflection of the copula. What we do is carry out quite a complex operation, which, first of all, identifies the subject of the sentence, ‘the bombing of the cities’, and then selects the core phrase within it, which happens to be ‘bombing’ and not ‘cities’. Two fairly complicated operations carried out on what we never hear but what the mind constructs, which we do it reflexively, ignoring simple operations like adjacency. And furthermore, these properties hold of all constructions in all languages. I’ll give some interesting qualifications about that in a moment. So here we have a remarkable property of language, a very surprising fact about learning and acquisition from infancy on. We ignore everything we hear, and attend only to what we never hear and what the mind constructs. And this is done reflexively without training, without evidence. Just part of our nature.

That raises questions ‘Why is this the case? Why do we not use linear order?’ It turns out we have an answer to that. If the language is a computational system, it’s based on recursive enumeration of an infinite array of entities. That’s very clear. The simplest computational operation is binary set formation, called Merge in contemporary discussion (Chomsky 2021b). So if language uses the simplest computational operation Merge, it turns out it’s very easy to show that you get structure-dependence and no resort to linear order. So assuming that the mind of humans is constructed so that it uses the simplest possible operation, then we have an immediate explanation for this quite surprising and far-reaching fact.

Now comes the qualification. What I’ve just been talking about holds for part of language, the internal part that is used to construct thoughts: the thoughts ‘The bombing of the cities is a crime’, ‘Carefully, the man fixed the car packed his tools.’ When we construct those thoughts, we use Merge, but when we externalize those thoughts into a sensorymotor medium, usually sound but could be sign, then we abandon this. We use linear order all the time. Lots of rules of
externalization use linear order. We conclude that language has two quite distinct parts. There’s an internal part, syntax and semantics, basically a thought creation system. There’s an external part which maps these internal objects into one or another sensory motor system (Chomsky 2017). It’s the first part that is keeps to the simplest possible computational operations.

Let’s go on. Why does it use the simplest possible operation? Here, there’s a thesis called the Strong Minimalist Thesis, which proposes that the internal system is basically perfect (Chomsky 2000). It’s designed optimally using the simplest computational operation. Why should we think of accepting it? The reason is what’s been discovered in science in the last couple of hundred years, what Albert Einstein called the miracle creed. Quoting Einstein, the true scientist believes that ‘the totality of sensory experience can be “comprehended” on the basis of a conceptual system built on premises of great simplicity. The skeptic will say that this is a “miracle creed”. Admittedly so, but it is a miracle creed which has been borne out to an amazing extent by the development of science’ (Einstein 1950: 13).

In fact, by now, it’s simply taken for granted by scientists. Your goal is to show that the miracle creed has been achieved. If it’s good enough for all of science, it’s good enough for us. Maybe language falls together with everything else in the world and satisfies the miracle creed, the internal system at least (and probably, in different ways, externalization too). If that’s true, we have the ultimate in explanation for this fundamental property of language.

That raises another question. How is it acquired? It’s obviously unlearnable. Therefore, it must somehow derive from our innate biological endowment called Universal Grammar (UG) in contemporary usage – adapting a traditional term to a new context: the innate, genetic basis for the language faculty, common to all humans. If it’s part of our innate basis, it evolved. That sounds paradoxical because it’s highly dysfunctional. Why should evolution produce something so dysfunctional? We’re using complex mental operations on what we never hear instead of simple, trivial operations on what we also always hear, highly dysfunctional. Why should evolution produce something perfectly simple, but highly dysfunctional? There’s an answer to that. Mother nature doesn’t care whether what evolves is dysfunctional. Contrary to many beliefs. Let’s spell that out more carefully.

How does evolution work? Think about it. There are basically three stages in evolution (cf. Chomsky 2021b). The first stage begins with some kind of structure, say organisms in some state. Then some disruption takes place. The first-stage disruption could be mutation, could be gene transfer. Lots of other possibilities. Might be a bacterium swallowing another microorganism. That actually happened once which led to complex cells, eukaryotic cells, ultimately,
humans. That’s why we’re not all bacteria. Some disruption takes place at the first stage.

Second stage, Mother nature comes along and observes the miracle creed, tries to find the simplest possible way of dealing with this new system that emerged. No concern for function. Doesn’t know anything about function.

Then comes the third stage, winnowing, natural selection. The organisms that are best fitted to the environment and reproduce more tend to stay around. That’s Darwinian evolution. When Darwin (1859) said in *Origin of Species* that all changes must be very slight, that’s the way evolution must work, he was talking about the third stage. For the second stage, that’s not true at all. There can be radical changes. We know many examples right now. But the crucial point is that at the second stage, nature cannot know about functions and doesn’t pay any attention to them.

So it’s perfectly normal for language to have emerged to satisfy the miracle creed but to be dysfunctional for (communicative) use. One illustration is structure-dependence. A dramatic illustration of the phenomenon. There are others. Don’t have a lot of time, but I’ll give one more case.

Merge systems plus economy considerations combine to delete unnecessary elements in externalization. That saves an enormous amount of computation. Take a simple sentence like ‘What did John eat?’ We know that ‘what’ is the object of ‘eat’, but we don’t hear that. Actually, the mind does hear that. That’s what’s produced by the simplest system of the internal grammar, the thought producing system. It produces ‘What did John eat what?’ That’s what reaches the mind, but it’s not what reaches the ear. What reaches the ear is ‘What did John eat?’ Actually, that’s dysfunctional, causes problems with perception and parsing. Those of you who work in natural language parsing will recognize this as the filler-gap problem. You hear the phrase, like a *wh*-phrase at the beginning, and have to find the place where the gap is. Hard problem. In fact, the hardest problem in parsing. And it can get very hard.

Let us give another example. Take the sentence ‘The girls expect the boys to like each other.’ ‘each other’, an anaphor, doesn’t refer. It has to find something to determine what it refers to. It looks for the closest thing, structurally closest, which is ‘the boys’. ‘The girls expect the boys to like each other [the other boys].’

Suppose we question the boys, we get ‘which boys did the girls expect to like each other?’ Notice that here, ‘each other’ does not look for the closest thing we hear, ‘the girls’. It looks for the more remote thing, ‘the boys’. The same process as before. What reaches the mind is ‘which boys, the girls expect those boys to like each other’. Then ‘each other’ finds the closest thing that gives the thought. That’s what reaches the mind, not what reaches the ear. What reaches the ear has a gap. You have to find it, and finding that gap can get quite hard. Take the sentence ‘The girls told us to expect the boys to like each other’, not too hard to
understand. How about ‘which boys did the girls tell us to expect to like each other’? Here there are two gaps. The subject of ‘expect’ is not pronounced. The subject of ‘like’ is not pronounced. It’s already a pretty complicated computation. And it goes on. It gets much more complex as you proceed.

That’s another example where language is designed to give the simplest system to the mind, to understanding, to thinking, but it is dysfunctional in (communicative) use. And that’s true in case after case. I won’t give other cases. The point is nature doesn’t care. Wherever we have a conflict between computational efficiency and communicative efficiency, communication is sacrificed. Language doesn’t care, just uses computational efficiency. That, incidentally, is one of many reasons to conclude that language is not fundamentally an instrument of communication, contrary to twentieth-century dogma. It was believed for millennia back to classical Greece and India that language is fundamentally a system of thought (cf. Hinzen 2017). In the twentieth century, Behaviourist, Structuralist pressures, misunderstanding of Darwinism, led to the conclusion that language is fundamentally a system of communication evolved somehow from animal communication. Soon as we study the actual system, we see that’s not true. It’s a system of thought, as was understood for millennia, the modern dogma is just wrong. And when we look at the way evolution actually works, that’s not in the least a surprise. In fact, it’s what we expect.

Well, I see the time is running out, so I’ll skip some more discussion, simply say that very much the same thing is true just of acquisition of words. So it’s been shown by careful experimentation with children, mainly the works of the late Lila Gleitman and her associates, that words are acquired on the very few presentations, two or three presentations. And what’s acquired is extremely rich. The semantic content of even the simplest words is quite complex when you look into it.

Identifying the sound of the word – its phonetic structure – is also non-trivial. For a chimpanzee or a dog, which have roughly the same auditory system, it’s just noise. But for an infant, it’s highly structured by the fundamental structures of phonology. Here, incidentally, neurological brain studies have given considerable insight (cf. Mampe et al. 2009 among others). Could go into this if you want. But that tells us that even what we regard is the simplest case of so-called learning actually involves complex, innate structure. There’s very little learning involved.

If you go on, you find that you raise serious questions about whether the notion learning is even a coherent concept. It happens to be a belief, a dogma, of contemporary philosophy and cognitive science, most of linguistics that learning is the uncontroversial notion, the null hypothesis. That is absolutely untrue. The evidence is all to the contrary. There’s been no success in finding generalized learning mechanisms. No results. We shouldn’t expect any results. That
Noam Chomsky raises the question whether learning is even a significantly identifiable component of the growth of an organism from conception to maturity. That raises a lot of questions which there is no time to go into.

Well, returning to the study of language. The major question is to see how far we can extend genuine explanation (Chomsky in press). How far does language satisfy the magical creed? How far does language fall within what we increasingly are coming to see in the essential nature of the world? That seems to be a very exciting prospect that we’re entering into in the new phase of the generative enterprise, the biolinguistics programme.

Note

1. The contents of this paper come from the invited keynote speech given by Professor Chomsky at 2022 First Conference on Biolinguistics and Language Acquisition, which was held in Beijing Foreign Studies University on 29 October 2022 (initiated by Professor Baocheng Han and Tiaoyuan Mao at the Institute of Linguistics of Beijing Foreign Studies University). The contents were transcribed by Tiaoyuan Mao and Xiangyu Chang and proofread by Professor Chomsky.

About the Author

Noam Chomsky, a laureate professor of linguistics at the University of Arizona and an institute professor emeritus at the Massachusetts Institute of Technology (MIT), is credited with revolutionizing the linguistics field with many groundbreaking books as Syntactic Structures, Language and Mind, and The Minimalist Program. His work has strongly influenced the fields of cognitive science, philosophy, psychology, computer science, mathematics, childhood education and anthropology.

References


Chomsky, N. (2021b) Minimalism: where are we now, and where can we hope to go. *Gengo Kenkyu* 160: 1–41.


