

A refutation of universal grammar

Francis Y. Lin

School of English, Beijing International Studies University, Chaoyang District, Beijing 100024, China

Received 21 March 2016; received in revised form 6 April 2017; accepted 10 April 2017

Available online 23 April 2017



Abstract

This paper offers a refutation of Chomsky's Universal Grammar (UG) from a novel perspective. It comprises a central part, clarifications and comparisons. The central part starts with an examination of Chomsky's research method and then argues that the method is seriously flawed and that it cannot lead to the discovery of any innate universals of language. In the clarifications part, a number of questions that could be raised concerning the central part of the refutation are presented and answered. The answers to these questions help to make clearer why UG is deeply problematic, and thus consolidate the central part of the refutation. The comparisons part discusses some representative critiques of UG by other scholars, shows their inadequacies, and thus further highlights the value of the present critique.

© 2017 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Universal grammar; Refutation; Research method; Scientific theory; Falsifiability

1. Introduction

For several decades Chomsky has been pursuing Universal Grammar (UG) in the attempt to explain the nature of language and the relationship between language and the mind. According to him, there is an innate language faculty, or language organ, whose initial state is UG. Extensive research has been devoted to finding out the content of UG. In the Principles and Parameters (P&P) version of UG, a host of grammatical principles were put forward, which are associated with a set of parameters. To get a glimpse of what UG principles (in P&P) look like, here are two typical ones:

Subadjacency:

Movement cannot cross more than one bounding node, where bounding nodes are IP and NP.

Binding Principle A:

An anaphor must be bound in its governing category, where the governing category for an expression is the minimal domain containing it, its governor, and an accessible subject.¹

Subadjacency was developed to explain why certain movement of words in sentences is allowed (e.g. *What_i did John claim that Peter stole t_i?*) while other movement is not (e.g. **What_i did John make the claim that Peter stole t_i?*). Similarly, Binding

E-mail address: ylin@bisu.edu.cn.

¹ For explanations of the relevant terms in UG mentioned in this paper, see Haegeman (1994), Ouhalla (1999), Cook and Newson (2007). For a general introduction to Chomsky's linguistic thinking, see Smith (2004).

Principle A was devised to account for the grammaticality of sentences such as *John_i shaves himself_i* and the ungrammaticality of sentences such as **John_i ordered Peter to shave himself_i*.

The parameters hypothesized in P&P were used to explain language variation. Suppose that there are 40 binary and independent parameters, this will allow for the possibility of 2^{40} (nearly 1.1 trillion) different grammars, more than enough to cover all the languages that have occurred in human history. Parameters were also employed to account for child language acquisition. When a child is placed in a language environment, the parameters will be set to certain values, giving rise to the grammar of that language. In this way, a child will acquire any human language with ease.

P&P was developed from the late 1970s to the early 1990s (see Chomsky, 1981a, 1981b, 1986). Since then UG has developed into the Minimalist Program (MP) (see Chomsky, 1995a; Boeckx, 2006). In P&P, there are various models, e.g. the θ -theory, the binding theory, the bounding theory, etc., and also various levels of syntactic representation such as S-structure, D-structure and LF (logical form). Are these theoretical posits really necessary? In MP, Chomsky tries to cut UG down to a conceptual minimum. The language faculty is seen as one of the components of the mind/brain, such that it interacts with the rest. The language faculty is capable of generating infinitely many expressions, which are then passed to other systems of the mind/brain for use (these are called interface systems, and they include the sensory-motor and conceptual-intentional systems, and possibly others).

As for how the language faculty generates expressions, Chomsky proposes that the minimum requirement is Merge. For him, Merge is a binary operation, which joins two linguistic entities.² Through reiterative application of Merge, expressions are built up. But there must be constraints on Merge, otherwise all sorts of expressions would be generated, many of which would be of no use. Over the years, various constraints have been put forward, as Chomsky explains:

The “core computational mechanisms of recursion” include the indispensable operation Merge and the principles it satisfies. To discover these principles has been a central research task in generative grammar since 1950s. Among them are, for example, locality conditions that have been investigated intensively. (Chomsky et al., 2005: 2)

Locality Conditions include a host of principles, e.g. Island Conditions, Subjacency, Shortest Move, Greed, etc.³ In addition to these, there are other principles in the literature, such as Last Resort, Minimality, Procrastinate, Inertness, and so on.⁴ All these principles are called economy principles.

Apart from economy principles, some principles of efficient computation have also been hypothesized, such as No-Tampering Condition (Chomsky, 2005: 11–13, 2007b: 5, 2008: 138), Inclusiveness Condition (Chomsky, 1995a: 228, 2007b: 6), Minimal Search and No Backtracking (Hauser et al., 2002: 1578).

It is thus clear that research on UG has been highly vibrant, and the research seems to have been very productive. Yet, UG has been criticized from various disciplines and viewpoints. Of course, Chomsky has not been idle when facing all these criticisms. He and his supporters have produced a barrage of rebuttals, arguing for the rationality and validity of UG.⁵ The debate on UG has lasted for long, and is presently still going on. Suffice it to say that neither side has succeeded in convincing the other that they are wrong.

In this paper I offer a refutation of UG from a novel perspective. I shall argue that a serious problem with UG lies in the method of finding it, a method which is deeply flawed and which makes it impossible to discover any innate linguistic universals.

2. Refutation of Chomsky's UG

In this section and the next, I shall formulate my refutation of Chomsky's UG. For ease of exposition, I shall first present the central part of the refutation. Clarifications will be offered in the next section in order to consolidate it.

In Section 2.1 I shall examine Chomsky's method for finding UG. I shall then in Section 2.2 show that this method is seriously flawed.

2.1. Chomsky's research method

Chomsky claims that he employs the so-called ‘standard method of the sciences’ (Andor, 2004: 97). He also refers to this way of studying the world as the ‘Galilean style’ of inquiry (see Chomsky, 1980a: 8, 218, 2004: 25, 170). Let us now examine his method in some detail.

² See Chomsky (2004: 167–169) for his reasons for taking Merge to be binary rather than n -ary ($n > 2$).

³ See Hornstein et al. (2005: 8) and Gärtner and Michaelis (2005: 114) for explanations of Locality Conditions.

⁴ See Collins (2001) and Boeckx (2006: Chapter 3) for explanations of economy principles.

⁵ In Section 4 below I shall discuss some of the criticisms of UG and Chomsky's rebuttals of them.

Chomsky claims that humans are born with UG, which consists of a small set of highly abstract grammatical principles. A typical example is Subjacency, which says that movement in a sentence cannot exceed a certain range. The development of Subjacency is particularly revealing for our purpose. In the 1960s and early 1970s, some linguists noticed that movement of words/phrases in sentences is constrained, and they formulated some such constraints. For example, the following sentences are ungrammatical:

- (1) *John_i appears [_{CP} it is likely [_{IP} t_i to win]]
- (2) *Which way_i do you wonder [_{CP} why [_{IP} John went t_i]]?
- (3) *What_i did John make [_{NP} the claim [_{CP} that Mary owns t_i]]?

and they were seen to involve movement violating the following constraints respectively:

Specified Subject Condition: An NP cannot move across a subject.

Wh-island Condition: An wh-expression cannot move across a wh-island, which refers to a clause whose [Spec, IP] has been occupied by a wh-expression.

Complex NP Condition: A wh-expression cannot move out of a complex NP.

Apart from these constraints, similar others were also found. Later on, Chomsky (1973) subsumed Wh-island Condition and Complex NP Condition under a more abstract constraint, which was Subjacency (repeated here for convenience):

Subjacency:

Movement cannot cross more than one bounding node, where bounding nodes are IP and NP.

It turned out that Specified Subject Condition could be derived from Subjacency as well. So, Subjacency can be seen as a generalization over Wh-island Condition, Complex NP Condition and Specified Subject Condition.

The development of Subjacency⁶ shows that Chomsky's research method is this: starting from some grammatical data, find some rules which can explain the data and then try to obtain some more abstract principles. Of course, the rules and principles are subjected to revision if recalcitrant data are found.

To make Chomsky's method more evident, let us examine another example: Binding Principle A. Consider the following sentences involving the anaphoras *himself* and *herself*:

- (4) John_i likes himself_i.
- (5) Mary_i admires herself_i
- (6) John_i believes [_{IP} himself_i to be the best]
- (7) *John_i believes [_{CP} that [_{IP} himself_i is the best]]
- (8) *John_i believes [_{NP} Mary's description of himself_i].
- (9) John_i believes [_{NP} any description of himself_i]
- (10) John_i believes [_{CP} that [_{IP} a picture of himself_i will be on show tomorrow]]

Simple sentences like (4) and (5) suggest that an anaphor must be bound in its local clause. But (6) requires that the binding domain be extended beyond the local clause. Yet it cannot be extended too far as to include (7). The difference between (6) and (7) is this: in (6), *believes* governs *himself* (because IP is not a barrier to government), while it is not the case in (7) (because CP is a barrier). In order to allow (6) and disallow (7), the binding domain for an anaphor X must be defined as the clause which contains X and its governor. However, this definition cannot account for (8–10). In order to account for them, the binding domain must be re-defined, yielding Binding Principle A (repeated here for convenience):

⁶ For more discussion on the history of Subjacency, see Roberts (2007), Lasnik (2010), Aboh et al. (2013).

Binding Principle A:

An anaphor must be bound in its governing category, where the governing category for an expression is the minimal domain containing it, its governor, and an accessible subject.

In the above, we have reviewed the development of Subjacency and Binding Principle A, two most illustrative principles in UG. From this review, we can summarize Chomsky's method as this:

Chomsky's Method:

On the basis of certain interesting data, find some general principles that explain them; revise the principles if necessary.⁷

2.2. A serious flaw in Chomsky's method and in UG

Chomsky tries to find laws governing human language. I shall focus on Subjacency (Binding Principle A can be treated in the same way). Subjacency is supposed to be a principle of UG. A human language cannot fail to satisfy it. If a language violates it, then it cannot be a human language.

There is a grave problem with Chomsky's method of research. Subjacency is proposed based on English data, other languages might violate it.⁸ If we consider some other languages and revise Subjacency into, say Subjacency_i, then languages which have not been taken into consideration might still violate the latter. So, it appears that we must test Subjacency against all existing languages, whose total number is between 6000 and 8000 (Evans and Levinson, 2009: 429). Even if we arrive at, say Subjacency_j, which can account for all these languages, it is still far from satisfactory. This is because there have been about 500,000 languages in human history (Evans and Levinson, 2009: 432), most of which have gone extinct, so how could one assure that they will all satisfy Subjacency_j? The 500,000 or so languages arose from certain specific natural and social environments; if human history had been different, there could have been a large number of different languages. The number of possible human languages is impossible to estimate, and it might be infinite. How could one ascertain that all possible human languages must satisfy some version of Subjacency?⁹ No matter how many sentences in how many languages UG theorists have examined, the version of Subjacency posited on the basis of those data cannot be regarded as the law governing movement of words in human language.

Subjacency is supposed to be a principle of UG, and all human languages must satisfy it. If a language violates it, then it cannot be a human language, and a normal human being cannot speak it naturally and cannot acquire it naturally. What the constraint on movement of words in sentences depends ultimately on a person's brain structure. Factors which determine this constraint are likely to include:

memory
attention
information retrieval speed
information processing speed
etc.

Every sentence requires the brain to process it. In other words, for every sentence there are certain *processing requirements* on the brain, e.g. which parts of the brain must be involved and in what way. If the brain can meet the processing requirements of a sentence, then a person can acquire and speak it; otherwise its acquisition would be beyond the person's capability and it could not be a sentence in a human language.¹⁰ So, what languages humans can acquire or

⁷ Chomsky says of his research method: "As for my own methods of investigation, I do not really have any. The only method of investigation is to look hard at a serious problem and try to get some ideas as to what might be the explanation for it, meanwhile keeping an open mind about all sorts of other possibilities. Well, that is not a method. It is just being reasonable, and so far as I know, that is the only way to deal with any problem" (Chomsky, 1988: 190). This is in no conflict with my characterization of his method.

⁸ It has been reported that Subjacency is violated in Swedish, Italian and Russian (Allwood, 1982; Newmeyer, 2004; van Valin and LaPolla, 1997: 615 ff).

⁹ Newmeyer (2005: Chapter 1) also discusses the difficulties with determining what is possible in human language in terms of UG principles.

¹⁰ It seems that the processing of the sentences in a language by the brain should satisfy some general constraints. One such constraint may be that this processing should facilitate thinking and communication. If an average sentence in a language took the brain hours to process, then we would probably not want to call it a possible human language. What constraints on the brain's processing of language are is an important question, and has important bearings on understanding the precise processing requirements of sentences. These are empirical issues which can be made clearer in the course of inquiry, but for ease of exposition, I shall leave out discussion on these constraints in the rest of this paper.

speak boils down to what sentences the human brain can process, and this depends on factors such as those listed above.¹¹

A couple of clarifications are in order. First, one might object that an infinitely long sentence, e.g.:

(11) John believes that Peter believes that Bob believes . . .

is a sentence in a human language but the brain, being a finite substance, cannot process it. In fact there is no contradiction here. To say that (11) is a sentence in a human language is to say that speakers of that language can speak or understand it. In a strict sense, a human being cannot speak or understand an infinitely long sentence. So, what is going on here is that when saying that (11) is a human sentence we mean something like this: if there were no limitation on memory and other relevant factors, then humans would be able to speak or understand it. In this sense, (11) is a human sentence; and in the same sense, it can be processed by the brain.

Second, one might argue that a sentence, while being able to be understood by people and processed by the brain, might still be ungrammatical (e.g. Sentence (3) above). But again this does not show that there is a contradiction between saying that a sentence is able to be processed by the brain and saying that it is humanly possible. Sentence (3) is ungrammatical in English, but this does not mean that it, or a sentence of the same construction, is not possible in another language. The very fact that people can understand (3) shows that it may be a normal, grammatical sentence in another human language. So, a humanly possible sentence is at the same time a grammatical sentence in a possible human language.

The following point thus becomes clear. If a sentence is grammatical in any one human language, then it is of course humanly possible, and certainly the brain can process it. However, if a sentence is ungrammatical in *N* languages (*N* being any number) that have been investigated, then one still cannot claim that it is humanly impossible, i.e. that the brain cannot process it. So, whether a sentence is humanly possible or not cannot be judged based merely on observed (grammatical and ungrammatical) sentences. To make such a judgment, one must investigate the processing requirements of the sentence and see whether or not the brain can satisfy them.

The following example might help to appreciate the point. Suppose that one wants to discover the limiting time for humans to complete the 100-m race. The person gets hold of some people, let them run the 100 m race, measure their finishing times, and hypothesizes that the shortest one is the human limit for running the 100 m race. It is obvious that this person's method and theory are flawed. A better approach would be to consider the relevant factors (such as the human physiology, the functions of the heart and the lungs, the running posture, the condition of the runway, the quality of the shoes, so on and so forth), make some idealizations, and compute the shortest finishing time under these idealized conditions. Analogously, it is wrong to try to find the limit of movement of words in sentences merely by looking at some sentences in some languages; a better approach would be to consider the relevant factors, such as those discussed in this subsection.¹²

UG contains a set of principles, Subadjacency being a paradigm. Since Chomsky's research method applies to all the UG principles and other putative language universals, they are all seriously flawed. This concludes the central part of my refutation of UG.

3. Clarifications

Section 2 was the central part of my refutation of Chomsky's UG. A natural reaction to it is the following doubt: could this be a *refutation* of UG? To put it differently, the doubt is: can UG, which has been developed by Chomsky and his thousands of followers over the course of several decades, be refuted in such a simple manner? So, in this section I shall try to dispel such doubts, by answering some questions that are likely to be raised concerning Section 2. I shall divide these questions into three groups, concerning respectively the methodology of UG, the minimalist program, and some more arguments for UG. I shall show that the answers to these questions will actually help to make clear why UG is seriously flawed.

3.1. Concerning the methodology of UG

Chomsky appeals to research in several other disciplines to justify his research method. In this subsection, I consider Chomsky's comparison of UG with Marr's vision research, with the thermonuclear theory, and with the theory of evolution in biology.

¹¹ This is similar to the distinction between capacity and ability made in psychology: we may have the capacity to do something, and yet may not have exhibited the ability to do it, because we may not have used our full potential (see for example Kaufman and Kaufman, 2016). From this viewpoint, a human society may use Sentence (3) as a grammatical one in its language, if the members make better use of their language capacity. So this work in psychology can support my argument against UG. But what is language capacity? It seems that in order to answer this question we must know enough about the brain and about the processing requirements of language on the brain.

¹² Lin (2015) illustrates the same problem with UG by contrasting a study of the fall of heavy bodies and one of the fall of live birds.

Q1: Concerning language and the visual system

Question: Chomsky has many times argued that language and the visual system are similar, so they can be studied in the same way. Shouldn't UG be as sensible as Marr's vision theory? In particular, Marr distinguishes three levels: the computational level, the algorithmic/representational level, and the physical level. He proposes to study the computational level, leaving out the details of implementation, such as eye structure, etc. Shouldn't UG be considered a theory at the computational level? If yes, then one can leave out the details of the brain structure in constructing UG.

Answer:

Chomsky compares language with the visual system and points out there are some similarities between them (Chomsky, 1968: 83, 1975: 8, 11, 1980a: 39, 1993: 29). This is one of the reasons why Chomsky calls language 'the language organ'. He proposes to study language in the same way as scientists study the visual system.

Chomsky agrees with Marr's framework of multi-level analysis (Marr and Nishihara, 1978; Marr, 1982),¹³ and thinks that it is consistent with UG. He writes: 'Adopting this framework, we may consider the study of grammar and UG to be at the level of the theory of the computation' (Chomsky, 1980b: 48). In this way, he implicates that in searching for UG, details about the brain need not be considered.¹⁴

I want to point out that while it is fine for Marr to work on the computational level without needing to consider details at lower levels, it is problematic for Chomsky to find UG principles without considering brain details.

An important difference between visual perception and movement of words in sentences is this. Given a cube, one person perceives as a cube, any other person also perceives as a cube. So it is reasonable to assume that people have the same visual system: they run the same computational theory (in Marr's sense).¹⁵ By contrast, given a sentence, a speaker of one language might regard it as ungrammatical, but a sentence with the same grammatical construction might be deemed to be grammatical by a speaker of another language. So, the language organs of different people, if we adopt Chomsky's terminology here, might yield different outcomes for the same sentence (or sentences of the same grammatical construction). So, it seems that people's language organs run different computational theories (in Marr's sense).

So, while the visual systems of different people yield the same outcome given an object, the language organs of different people may produce different outcomes given the same input. The parallel between the visual system and the language organ breaks down here. While it is reasonable to assume that people's visual system run the same computational theory, it does not seem to be reasonable to posit that people's language organs run the same computational theory. So, the foregoing discussion weakens the force of Chomsky's analogy between language and the visual system, and reduces the plausibility of his proposal that language and the visual system be studied in the same way. It does not, however, establish that Chomsky's proposal is unworkable. We must therefore dig deeper (see below).

Why is it plausible to think that the structure of the eye does not affect its perception? That is, why can we neglect the structure of the eye and only focus on the level of the computational-theory? Of course, in an act of perception many, and possibly all, parts of the eye are involved, and each part has a role or function. For instance, in perceiving a cube, one part of the eye may be responsible for color detection, another for line recognition, a still other for depth determination, so on and so forth. The malfunction of any eye-part might affect perception. Here though, we are not dealing with malfunctions: we assume that all the parts of the eye are functioning normally. Then, what can affect, say the perception of a cube? Let us assume that the cube is placed in a fixed location, and the eye's position is also fixed. When a cube is in sight, every part of the eye makes a certain contribution, and in the end a cube image is formed (as described by Marr). Note that each eye-part makes a fixed contribution: if the eye blinks and opens again, each eye-part's contribution will not alter, and the same cube-image will be formed. So, the eye-parts do not affect the perception of the cube. This is why it is reasonable to focus on the computational theory of the eye and not bother about the structure of the eye.

Now, look at the language system. Consider Sentences (12), and (13) (which is (3) repeated here for convenience):

(12) What_i did John claim [_{CP} that Mary owns t_i]?

¹³ Marr and Nishihara (1978) distinguish four levels: the computational, the algorithmic level, the level of particular mechanisms, and the level of basic component and circuit analysis. The last two levels are collapsed into the level of hardware implementation in Marr (1982).

¹⁴ In his later work Chomsky began to distance himself from Marr's theory. He writes that Marr 'was concerned with input-output systems (e.g. the mapping of retinal images to internal representations). Language is not an input-output system. Accordingly, Marr's levels do not apply to the study of language' (Stemmer, 1999: 397; see also Chomsky, 1995b: 12). The reasons behind this move of Chomsky are not entirely clear. But if Chomsky no longer finds Marr's work relevant to UG, then Q1 will cease to be a question for me here. Curiously, Chomsky now appeals to Marr's work again (see Berwick and Chomsky, 2016a: 132–139).

¹⁵ According to Marr (1982), vision proceeds from a two-dimensional visual array (on the retina) to a three-dimensional description of the world as output. Stages of vision include: (1) a primal sketch of the scene, based on features detected from the scene, e.g. edges, regions, etc.; (2) a 2.5D sketch of the scene, which contains information about orientation, contour and depth, etc. of visual surfaces; and (3) a 3D model, where the scene is visualized in a continuous, 3-dimensional map.

(13) *What_i did John make [_{NP} the claim [_{CP} that Mary owns t_i]]?

Since an English speaker does say (12), its processing requirements are obviously met by the brain. For example, (12) might require the use of 30% memory, 50% attention, 80% processing speed, etc., and the brain can certainly meet these requirements. Now, (13) is ungrammatical in English. Is it not possible for a human language to contain a sentence like (13)? If we analyze its requirements, we might find that it requires the use of 40% memory, 60% attention, 90% processing speed, etc. If this is the case, then the brain will certainly be able to support this sentence, and the sentence will be humanly possible.¹⁶ On the other hand, if we find that the sentence requires the use of 120% memory, 150% attention, 200% processing speed, etc., then the human brain will not be able to process it, and it will not be humanly possible. So, even if a sentence is ungrammatical in all the human languages one has examined, one cannot be sure that it is humanly impossible. To make a sure judgment, one must investigate the processing requirements of the sentence on the brain to see whether they can be met by the brain or not. It would not be to find out some constraints on humanly possible sentences, e.g. Subjacency_U, merely by looking at some relevant grammatical and ungrammatical sentences, in however big a number of languages.

To sum up the answer to Q1, language and the visual system share some similarities, but they are fundamentally different. It is reasonable to study the computational theory of the visual system without considering the details of the eye structure, but it is seriously inadequate to find the ultimate version of Subjacency and similar UG principles without considering the processing requirements of sentences and the brain structure.

Q2: Concerning language and the thermonuclear theory

Question: All scientific research is limited by available data. UG theorists only have data consisting of grammatical and ungrammatical sentences (in one or more natural languages), so UG can only be based on such data. So, in this respect is UG not the same as typical scientific theories, such as the thermonuclear theory?

Answer:

Chomsky thinks that UG is on a par with the thermonuclear theory in terms of evidence. His argument is put clearly and eloquently in the following passages:

Consider the problem of determining the nature of the thermonuclear reactions that take place deep in the interior of the sun. Suppose that available technique permits astronomers to study only the light emitted at the outermost layers of the sun. On the basis of the information thereby attained, they construct a theory of the hidden thermonuclear reactions, postulating that light elements are fused into heavier ones, converting mass into energy, thus producing the sun's heat. ... One might argue that the evidence is inconclusive or that the theory is objectionable on some physical (or, conceivably, methodological) grounds. But it is senseless to ask for some other kind of justification for attributing physical reality to the constructions of the theory, apart from consideration of their adequacy in explaining the evidence and their conformity to the body of natural sciences as currently understood. (Chomsky, 1980a: 189–190)

Our investigation of the apparatus of the language faculty, whether in its initial or final state, bears some similarity to the investigation of the thermonuclear reaction in the solar interior that is limited to evidence provided by light emitted at the periphery. We observe what people say and do, how they react and respond, often in situation contrived so that this behavior will provide some evidence (we hope) concerning the operative mechanisms. We then try, as best we can, to devise a theory of some depth and significance with regard to these mechanisms, testing our theory by its success in providing explanations for selected phenomena. Challenged to show that the constructions postulated in that theory have 'psychological reality', we can do no more than repeat the evidence and the proposed explanations that involve these constructions. (Chomsky, 1980a: 191)

I want now to show that Chomsky's analogy between the thermonuclear theory and UG does not hold.

By the early 20th century, scientists had found, through analyzing the solar spectrum and other means, that the sun is composed largely of hydrogen and helium, with also a small percentage of oxygen, carbon, neon, nitrogen, and so on. In 1929 Robert Atkinson and Fritz Houtermans predicted that large amounts of energy could be released by fusing small nuclei. The laboratory fusion of hydrogen isotopes was first accomplished by Mark Oliphant in 1932. In the late 1930s, Hans Bethe applied the idea of nuclear fusion to explain the energy source that powers the sun, and put forward the thermonuclear theory. According to this theory, under an extremely high temperature four hydrogen nuclei can fuse to make one helium nucleus, releasing a large amount of energy during the process.

When this theory was first put forward, there was no laboratory confirmations (they came later), and nobody could go to the sun to check whether there was nuclear fusion going on there. And yet, few would say that the astronomers were

¹⁶ In fact, the Complex-NP Constraint is violated in some languages, e.g. Swedish (see Allwood, 1982). In these languages, (13) is grammatical.

talking nonsense. Why was the thermonuclear theory reasonable? Well, the astronomers considered the factors that could affect the production of energy of the sun, and tried to build a model of the energy source. The sun is composed mostly of hydrogen and helium, so it was natural to treat hydrogen and helium as the causal factors, neglecting the contributions of other elements such as oxygen and carbon. Then, how could hydrogen and helium cause the sun to have so much energy? Many hypotheses could be formed, for instance, one might think that the burning of hydrogen and helium could do the trick. But, given the prior work by Atkinson, Houtermans, Oliphant and others, the thermonuclear theory seemed more reasonable. So, the evidence for the thermonuclear theory consisted of not only data about the light from the sun, but also evidence coming from laboratory experiments on nuclear fusion. The theory was reasonable, because causal factors for the sun's heat were taken into consideration.

By contrast, Chomsky's search for Subjacency_U does not consider causal factors. Causal factors play no role in his construction of Subjacency.

It is reasonable to say that the evidence for the thermonuclear theory is data about the light from the sun. It is also reasonable to say that Chomsky's Subjacency is based on some sentences in certain languages, after all it is on this basis that Subjacency is devised. In this sense, Chomsky's Subjacency is indeed limited by available data. But if we really want to discover the ultimate Subjacency_U, we must base our research on evidence comprising data concerning factors such as memory and attention, and not merely on data consisting of some sentences in certain languages (these data cannot be evidence for Subjacency_U).

So, although Chomsky's UG and the thermonuclear theory are similar in terms of evidence, they are fundamentally different. The thermonuclear theory considers causal factors but UG does not, consequently the former is unproblematic, whereas the latter is in deep trouble.¹⁷

Q3: Concerning language and biology

Question: According to certain theories in biology, slight changes in the regulatory mechanisms for genes will cause a great of variation in the forms of animals. P&P makes the same kind of hypothesis: slight changes in language environments will cause parameters to be set differently, giving rise to a great number of different grammars. For example, 40 binary and independent parameters will allow for the possibility of 2^{40} (=1,099,511,627,776) different grammars. Doesn't this sound plausible?

Answer:

The molecular biologist François Jacob (1978), speculating on the enormous diversity of organisms, writes:

It was not biochemical innovation that caused diversification of organisms ... What accounts for the difference between a butterfly and a lion, a chicken and a fly, or a worm and a whale is not their chemical components, but varying distributions of these components ... specialization and diversification called only for different utilization of the same structural information. ... It is thanks to the complex regulatory circuits, which either unleash or restrain the various biochemical activities of the organism, that the genetic program is implemented. [In related organisms, mammals for example], the diversification and specialization ... are the result of mutations which altered the organism's regulatory circuits more than its chemical structures. The minor modification of redistributing the structures in time and space is enough to profoundly change the shape, performance, and behavior of the final product. (Quoted in Chomsky, 1980a: 67, square brackets are Chomsky's)

According to Chomsky (2007a: 14):

The P&P approach ... was also suggested by an analogy to new developments in biology, specifically the discovery of regulatory mechanisms by François Jacob and Jacques Monod, and Jacob's speculations about how slight changes in these mechanisms might yield great superficial differences – a butterfly or an elephant, and so on. The model seemed natural for language as well: slight changes in parameter settings might yield superficial variety, through interaction of invariant principles with parameter choices.¹⁸

I want now to show that Jacob's speculation, made in 1978, mentioned in these passages is reasonable, but that the idea of P&P is not.

Prior to 1978, scientists had discovered that there are genes in living organisms and that genes carry information for the production of proteins. They had also discovered that the sequences of chimp and human proteins were nearly identical. Jacob's work on *E. coli* had shown that genes are controlled by some gene switches to produce or not produce certain proteins. If animals are built by similar proteins, proteins are determined by genes, and genes can be regulated by gene switches, then it is natural to think that animals share some common genes, and that they are different because the genes

¹⁷ See Answer to Q4 for more discussion on the issue of evidence.

¹⁸ See also Chomsky (1980a: 67); Berwick et al. (2011: 27–29); Berwick and Chomsky (2016a: 67ff).

have been used differently. Thus, Jacob's speculation was based on the consideration of the causal factors for organism diversity, such as genes, proteins, gene switches, etc. Consequently, it makes a great deal of sense.

Actually, Jacob's speculation has been borne out by the recent theory of evolution and development, or informally Evo-Devo. According to Carroll (2005), animals share a common 'tool kit' of master genes, which have been stable in evolution; the enormous diversity of animal forms is due to the evolutionary tinkering of the gene switches.

P&P looks similar to Evo-Devo in appearance, but they are rather different deep down. The most important difference is in the method. While Evo-Devo considers causal factors, P&P does not. We have seen how Subjacency, a typical UG principle, is obtained by merely looking at certain sentences in one or more languages, and not by considering the relevant causal factors, and that this makes Subjacency deeply problematic. The idea of parameters in P&P does not fare any better, and this is what I shall now show.

Let us examine how parameters are found in P&P. When two languages differ in a certain grammatical phenomenon, the difference is said to be due to a parameter. On the face of it, this way of finding parameters seems innocuous. Yet, when put together with the general idea of P&P, its flaw will be seen. The general idea of P&P is that there are a finite set of UG principles and a finite set of parameters: they together determine the logical space of possible human languages. Now, by examining some existing languages, UG theorists come up with some parameters, which explain the variations in these languages.¹⁹ Will other languages still differ from these languages? Very likely. So, more parameters will have to be posited. But this question can be raised again and again even when more and more languages are taken into consideration. How can UG theorists ascertain that the parameters they have found on the basis of a number of languages they have examined will explain variations in *all* human languages?²⁰ Thus, finding parameters faces the same acute problem as that we saw in the case of finding Subjacency. There is no way to find *the* set of parameters for all human language variations merely by looking at some number of existing languages, no matter how great the number is.

The crucial problem with the idea of parameters in P&P is that parameters are supposed to, together with UG principles, set the limits on possible human languages, but they are to be discovered on the basis of some existing languages. In order to discover the limits of human language variation, one must consider the causal factors relevant to such limits. A language, taken as a set of sentences, has processing requirements on the brain. As long as they can be met by the human brain, the language will be a possible human language. So, to discover the limits of language variation one must investigate processing requirements of various grammatical phenomena. These limits cannot be found by merely examining some human languages, no matter how many of them are inspected.²¹

Q4: Concerning evidence

Question: So far you have been portraying Chomsky as relying only on linguistic evidence, e.g. on grammatical and ungrammatical sentences. Don't Chomsky and his supporters also appeal to other types of evidence in constructing UG?

Answer:

Chomsky indeed talks about making use of all kinds of evidence. He intends UG to be the best theory based on available evidence, which supposedly includes both linguistic and psychological evidence. The following passage expresses his point well:

Approaching the topic as in the sciences, we will look for all sorts of evidence. For example, evidence from Japanese will be used (and commonly is used) for the study of English; quite rationally, on the well-supported empirical assumption that the languages are modifications of the same initial state. Similarly, evidence can be found from studies of language acquisition and perception, aphasia, sign language, electrical activity of the brain, and who knows what else. (Chomsky, 1994: 205)

Here Chomsky talks of evidence from 'studies of language acquisition and perception, aphasia, sign language, electrical activity of the brain'. But the *formulation* of UG principles is not *based* on such non-linguistic evidence, as one can see from the formulation of Subjacency and Binding Principle A (see Section 2.1 above). Chomsky might retort that he does consider some non-linguistic evidence when constructing his theory of UG, for example, in one place he says:

To the extent that notions such as efficient computation plays a role in determining how the language develops in an individual, that ought to be a general biological, or maybe even a general physical, phenomenon. So if you get any evidence for it from some other domain, well and good. That's why when Hauser and Fitch and I were writing

¹⁹ Newmeyer (2005) provides a detailed description of work on parameters in UG. He points out some problems with the idea of parameters.

²⁰ Comrie raises a similar question, and he remarks that 'it would be a serious methodological error to build into our research program aphoristic assumptions about the range of variation' (1989: 6). He argues that more languages need to be investigated in order to talk about language universals. Thus, his argument goes in a rather different direction than the present paper.

²¹ Answer to Q7 below continues to discuss the issue of parameters. The relationship between MP and the evolution of language will be investigated in Section 4.3 below.

(Hauser, Chomsky & Fitch 2002), we mentioned optimal foraging strategies. It's why in recent papers I've mentioned things like Christopher Cherniak's work [on non-biological innateness (2005) and on brain wiring (Cherniak, Mikhtarzada, Rodriguez-Esteban & Changizi 2004)], which is suggestive. (Chomsky, 2012: 60, round and square brackets in original text)²²

The optimal foraging strategies and Cherniak's work on optimal brain-wiring that Chomsky mentions here serve only as inspirations or suggestions. Chomsky thinks that these are 'suggestive' (in his own word) in that they suggest that language might be an optimal system. However, they are not evidence for the thesis that language is an optimal system. More importantly, the formulation of principles and other universals in UG is not based on the optimal foraging strategies or Cherniak's work. Chomsky's talk of such 'evidence' is thus empty.

3.2. Concerning the minimalist program

Q5: Concerning Merge

Question: Isn't Merge a true universal in human language?

Answer:

Any human language must have a means of forming an expression by combining simpler items (e.g. words), this being a defining feature of language. If this combination operation is called merge, then merge is part of the concept of language. The proposition that merge is a universal in language is tautological.

Whether the particular Merge proposed by Chomsky's is universal or not is open to question. Even among MP theorists, there is considerable disagreement about the nature of Merge and its properties, e.g. whether Merge is the basic operation or is the product of two operations which are more basic, whether there are two kinds of Merge (Internal Merge and External Merge) or just External Merge, and so on (see Stroik and Putnam, 2013: 9–10).

The lack of a precise definition of Merge is certainly a problem for UG theorists. The crucial problem, however, lies in the principles of Merge. UG is supposed to provide guidance for possible human languages. Without constraints, Merge would produce all kinds of expressions, many of which may not be humanly possible. So, the important thing is not whether there is Merge or not (any theory of language would need to assume a kind of n -ary merge at a certain level of abstraction, where $n \geq 2$), but what the constraints on Merge are. Chomsky calls the constraints on Merge 'the principles it satisfies' (Chomsky et al., 2005: 2, quoted above). These principles are called economy principles and principles of efficient computation. The crucial problem with these principles is that Chomsky's way of finding them is deeply problematic (see Answer to Question 6 below).²³

Q6: Concerning economy principles and principles of efficient computation

Question: You have been discussing principles in P&P. But the current research framework is MP. Is there anything wrong with economy principles and principles of efficient computation hypothesized in MP?

Answer: The crucial fault with these principles lies in the way they are to be found. They are hypothesized after looking at some sentences in one or more languages. The method is the same as that used in finding Subadjacency, only these principles are more general and more abstract. The procedure, which was used to argue against Subadjacency, can be employed to show that these principles are also seriously in trouble. There is no need to repeat the procedure here.

Q7: Concerning parameters in MP

Question: In MP, many attempts are being made to reduce parametric variations to the lexicon or the interface between the faculty of language and the sensory-motor system so as to make the narrow syntax parameter-free. Doesn't this sound plausible?

Answer:

The idea of using a *finite* set of parameters to set the limits of language variation is deeply problematic, as has been shown earlier. This being so, there is no point in talking about where the parameters should be placed in the theory, whether to be associated with UG principles or to be dumped into the lexicon (Chomsky, 1995c: 54; Kayne, 2000; Rizzi, 2014) or the externalization process (Berwick and Chomsky, 2016a: 83, 92, 108).

What will happen if in MP the set of parameters is assumed to be *infinite*? Chomsky does contemplate this possibility:

I guess that most of the parameters, maybe all, have to do with mapping [to the sensory-motor interface]. It might even turn out that there isn't a finite number of parameters, if there are lots of ways of solving this mapping problem. (Chomsky, 2012: 53–54, square brackets in the original text)

²² Cherniak's work on non-biological innateness and on brain wiring is mentioned in Piattelli-Palmarini et al. (2009: 16) and Chomsky (2005: 6) respectively.

²³ Section 4.3 below will continue to discuss Merge in relation to the evolution of language.

Yet if infinitely many parameters are allowed, then it will mean that languages can differ in unlimited and unpredictable ways. Many years ago Joos (1957: 96) remarked that ‘languages [can] differ from each other without limit and in unpredictable ways’, and Chomsky has rejected this remark repeatedly in the past few decades. So, Chomsky will have to take back, at least water down, his long-standing criticism of Joos’s remark. Of course, this will not prove that Chomsky’s UG is wrong. It is nonetheless important to ask what UG will be like when parameters are infinitely many. Since UG principles cannot be associated with an infinite number of parameters, these parameters will have to be placed elsewhere, for example in the lexicon or in the externalization process. Dissociating UG principles from infinitely many parameters may be fine, but what about UG principles themselves? They are as problematic as before; postulating an infinite number of parameters cannot change this fact.

3.3. Concerning more arguments for UG

Q8: Concerning the ‘argument from the poverty of stimulus’

Question: A very important argument for UG is the ‘argument from the poverty of stimulus’. Isn’t this argument persuasive?

Answer:

The term ‘argument from poverty of the stimulus’ (APS) first appeared in Chomsky (1980a: 34), though one can find traces of this argument in his writings in the 1950s and 1960s (see Thomas, 2002). In fact, there is no single version of APS, rather many scholars have proposed different versions of it. Pullum and Scholz (2002) identify no less than 13 separate arguments with different premises. Almost all of the premises involved in APS have been questioned, and the reasoning involved in APS has been criticized. The literature on the debate on APS is huge,²⁴ and to explain adequately the debate on APS would require a book-length treatment. Two things though are clear enough: one is that APS is an argument for the existence of UG, and other being that neither side of the debate has managed to convince the other side.²⁵

In the light of the preceding discussions in Sections 2 and 3, there is no need to go into the debate on APS. APS is used to argue that UG exists. Now, does UG exist? As argued earlier, there is no way to find the innate language universals (e.g. Subjacency_U) hypothesized in UG. So, APS can no longer support UG.

Of course, how the child acquires language is puzzling and needs explaining. Evidently there must be innate constraints on the child’s acquisition of language; but they cannot be those postulated in UG. What the innate constraints on language are must be investigated in other ways.

Q9: Concerning the ‘best theory’ argument

Question: Chomsky says that in science normally theoretical constructs in the best theory are taken to be psychologically/physically real. Doesn’t that sound correct?

Answer:

Chomsky’s ‘best theory’ argument is well developed in Chomsky (2000: 108 ff) (but also see Chomsky, 1980c: 12; 1986: 243ff, 2000: 142, 145; 2002: 50, 53, 68, 2009: 173). He reviews the history of the advanced sciences, points out that many things posited in the past (e.g. gravity, fields of force, atoms, and chemical bonds) were initially regarded to be ridiculous and non-existent but were later accepted to be real, and argues that this is because the theories which hypothesized these entities turned out to be the best ones that explain the relevant phenomena. Chomsky holds that UG is the best theory that explains a person’s knowledge of language (2000: 60), and thus implicating that the hypothesized UG principles may well turn out to be psychologically/physically real.

Now, what Chomsky says about best theories in science seems fine. However, his own theory is deeply flawed. The ‘best theory’ argument is therefore has little force in upholding UG.

Q10: Concerning the ‘Martian scientist’ argument and linguistic naturalism

Question: Chomsky says that if a Martian scientist comes to the earth, he will find it natural to study language in the same way as he studies typical natural objects, and might conclude that there is only one language on the earth. Isn’t his argument persuasive?

Answer:

Chomsky never provides a complete version of the ‘Martian scientist’ argument, rather he presents fragments of it in several of his writings (Chomsky, 1988: 41ff., 2000: 3, 7, 97, 118, 141–142; Hauser et al., 2002: 1569; Antony and Hornstein, 2003: 291; Berwick and Chomsky, 2016a: 61, 78). The motivation behind the argument is clear: it is used to justify Chomsky’s own thinking on UG. The gist of the argument is as follows. Suppose that a Martian scientist comes to

²⁴ See, for example, Goodman (1967), Putnam (1967), Cowie (1999), Laurence and Margolis (2001), Sampson (2005), Clark and Lappin (2011), Berwick et al. (2011), and papers in the special issue on APS in *The Linguistic Review* 19(1–2), 2002.

²⁵ See also Pullum and Scholz (2002: 12) on this point.

the Earth. He observes many phenomena, not only those such as the blowing of the wind, the flowing of rivers, the fall of apples, and the growth of plants, but also the phenomenon of language. If the Martian scientist reasons as Chomsky does, i.e. by following APS and the 'best theory' argument, and by making analogies between on the one hand language and on the other hand the visual system, the thermonuclear theory and Evo-Devo, etc., then he may well think that there is an innate language faculty – a natural object, whose initial state is UG, and that UG consists of some abstract principles, which are psychologically real and are to be discovered by using the standard method of the sciences.

Indeed, if the Martian scientist reasons as Chomsky does, he may well come to the same conclusions about language as those reached by Chomsky, and he may well pursue UG in the same way as Chomsky does. However, if the Martian scientist is more careful, and reasons as what is described here in Sections 2 and 3, he will likely find that UG is deeply problematic, and will likely try to explain the phenomenon of language in other ways.

This is not to say that no aspects of language should be treated as a natural object. Certain parts of the brain and certain innate mechanisms are involved in language processing, and these are natural objects. The problem is how to identify them. They must be investigated by taking approaches other than the UG approach.

3.4. Summary

In this section some questions that are likely to be raised concerning the central part of the refutation of UG (presented in Section 2 above) were answered. These answers help to make more evident that UG is seriously flawed. Of course, more similar questions could be asked. For example, in recent years Chomsky has made a lot of discussion on optional/perfect design of language, the distinction between language in the broad sense (FLB) and the narrow sense (FLN), three factors in language design, biolinguistics, the evolution of language, and so on (Chomsky, 1995a, 2002, 2005, 2007a; Hauser et al., 2002; Fitch et al., 2005; Berwick and Chomsky, 2016a, 2016b). Are these discussions not plausible? The answer to questions like this is this: these discussions may, or do, make a lot of sense, but Chomsky's goal is to find UG; since the research method employed in UG is fundamentally flawed, and it is therefore impossible to find the real, innate UG (if it exists) using this method.²⁶

4. Comparisons with other critiques of UG

As said at the end of Section 1, UG has been criticized by many scholars, from a number of disciplines. It is impossible to analyze all these critiques in this paper, but it would be useful to discuss some representative ones, to show their inadequacies and to highlight the value of my critique.

4.1. The critique by Ibbotson and Tomasello

Ibbotson and Tomasello (2016) present a high-profile critique of UG. Their paper is short, but it contains several arguments against UG, which have been often heard in the field. In this subsection I would like to scrutinize these attacks, to see how effective they are against UG.

First, Ibbotson and Tomasello claim that linguistic evidence shows that the universals hypothesized in UG are false. Of the subject-drop parameter, they write: 'This notion of universal principles fits many European languages reasonably well. But data from non-European languages turned out not to fit the revised version of Chomsky's theory'. Of the idea that recursion is a universal property of human language, they state: 'Some languages – the Amazonian Pirahã, for instance – seem to get by without Chomskyan recursion'. I shall call this argument the 'false-universals argument'. In fact, many scholars have also argued that the posits in UG are falsified by research in specific languages. Sampson (2005: 157–8) cites evidence showing that the Island Constraint is violated in English. He also objects to a number other universals proposed by Chomsky and his followers. Evans and Levinson (2009) claim that Subadjacency, Binding Principles, syntactic constituency, and recursion etc., are not found in all languages. Newmeyer (2008) points out some other problems with some UG constructs. See also Tomasello (2009) for similar points.

Second, Ibbotson and Tomasello advance the 'unfalsifiability argument', which says that Chomsky's proposals are 'difficult to test in practice' and 'pretty much unfalsifiable'. Again, this argument has been rehearsed before, see Itkonen (1996), Seuren (2004: 28–29), Sampson (2005: 162–164), Tomasello (2007), Evans and Levinson (2009: 429), and Dąbrowska (2015). Since (un)falsification is related to (un)scientificity, some researchers have gone as far as to advance the 'unscientificity argument', claiming that UG is unscientific (see Lappin et al., 2000; Stokhof and van Lambalgen, 2011).

²⁶ Section 4.3 below will discuss UG in relation to the evolution of language.

Third, Ibbotson and Tomasello present the ‘linking-problem argument’. The linking problem is how the innate UG is connected with the grammar of individual languages. Ibbotson and Tomasello argue that ‘the linking problem – which should be the central problem in applying universal grammar to language learning – has never been solved or even seriously confronted’. The linking-problem criticism has been offered by other scholars too. For example, [Ambridge et al. \(2014, 2015\)](#) claim that UG suffers from the linking problem, as well as the problem of inadequate data coverage and that of redundancy.

Fourth, Ibbotson and Tomasello produce the ‘alternative-theory argument’. They propose a usage-based theory (e.g. [Tomasello, 2003](#)). It suggests that ‘children learn language through social interaction and gain practice using sentence constructions that have been created by linguistic communities over time’, that these constructions, at least some of them, are ‘evolve[d] over long periods’ because of ‘various cognitive-processing mechanisms – with names such as schematization, habituation, decontextualization and automatization’, that they are learned by children, and that analogy plays an important role in understanding communicative intentions while there are ‘constraining mechanisms’ which limit the power of analogy. In the mind of Ibbotson and Tomasello, the usage-based theory is an alternative theory of language and is better than UG, as evidenced in their concluding remark: ‘Universal grammar appears to have reached a final impasse. In its place, research on usage-based linguistics can provide a path forward for empirical studies of learning, use and historical development of the world’s 6000 languages’. Again, this alternative-theory argument has been seen in the literature before. For example, [Ambridge et al. \(2014, 2015\)](#) try to replace some key principles in UG (e.g. Subadjacency and Binding Principles) with certain discourse-pragmatic principles, and Hofmeister and colleagues attempt to account for island-effects in terms of processing factors ([Hofmeister and Sag, 2010; Hofmeister et al., 2012a, 2012b](#)). These researchers all argue, explicitly or implicitly, that their alternative theories are better than UG.

I have described four main arguments contained in the short paper of [Ibbotson and Tomasello \(2016\)](#), and have made clear that these arguments are shared by many opponents of UG. The unscientificity argument is not made by Ibbotson and Tomasello, but it is closely related to the unfalsifiability argument and has been made by others. It is time to examine the effectiveness of all these arguments against UG. I shall do this in two steps. First, let us see how Chomsky and his followers have reacted, or could react, to these arguments. Second, I shall make my own assessment of this debate.

The above arguments are all linked, directly or indirectly, to the conception of scientific research. So it is necessary to have a good understanding of this conception. Three most influential theories of scientific research in this regard are [Popper \(1959\)](#), [Kuhn \(1970\)](#) and [Lakatos \(1970\)](#). Popper regards falsifiability as a criterion of demarcation between what is and what is not genuinely scientific: a theory should be considered scientific only if it is falsifiable. According to Kuhn, science in general is broken up into three successive stages: pre-science, normal science and revolutionary science. Pre-science lacks a central ‘paradigm’. In normal science scientists attempt to enlarge the central paradigm by puzzle-solving. When many anomalous results build up, science reaches a crisis, a new paradigm then occurs which replaces the old one. This ‘paradigm-shift’ is then revolutionary science. While Kuhn views progression in science in terms of paradigm-shifts, Lakatos sees it through the concept of program-change. A ‘research program’ has a ‘hard core’, which contains its basic assumptions, and a ‘protective belt’ of auxiliary hypotheses. When refuting evidence is encountered, scientists will not abandon the program straightaway, rather they will try to alter the hypotheses in the protective belt such that the hard core of the program can be retained unscathed. When a program does not predict new phenomena, and/or none of those new predictions are confirmed by observation, the program is degenerating and will be superseded by a better, more progressive, program.

We are now in a position to examine how Chomsky and his followers have been arguing that UG is scientific according to the above three theories. First, Chomsky argues that UG is falsifiable, therefore implying that UG is scientific according to Popper’s theory:

An innate hypothesis is a refutable hypothesis. Any hypothesis which says that such and such a property of language is genetically determined is subject to the most immediate refutation of the strongest kind. Such hypotheses have been refuted over and over again in the past by just looking at the next phenomenon in the same language or the next language. That is why it has been so hard to formulate specific hypotheses about genetically determined structures. ([Chomsky, 1980c: 80](#))

Second, Chomsky and his supporters try to make UG consistent with Kuhn’s account of scientific research. In an interview around 1980, [Chomsky \(2004: 67\)](#) says that his work (on P&P) ‘may lead to the possibility of a conceptual revolution’ in the Kuhnian sense. Ten Hacken, a follower of UG, develops a notion of research program that is ‘strongly based on Kuhn’s concept of paradigm’ ([2007: 29](#)), and argues that UG is a revolution in the Kuhnian sense ‘because it is based on a different research program from Post-Bloomfieldian linguistics and it gradually replaced the latter’ ([2007: 179](#)).

Third, Lakatos’s theory has been used to support the claim that UG is scientific. [Chomsky \(1980a: 257, n. 15\)](#) cites [Lakatos \(1970\)](#) when discussing the Galilean style of inquiry. [Boeckx \(2006\)](#) applies Lakatos’s characterization of

research programs to Chomsky's Minimalist Program. He highlights 'parallelisms between minimalism and the methodological foundations in better-developed sciences such as physics' (Boeckx, 2006: 8), effectively arguing that the former is scientific in the Lakatosian sense.²⁷

Thus, Chomsky and his followers have developed a strong argument that UG is scientific according to theories of Popper, Kuhn, and Lakatos. In this way, they have defused the unfalsifiability argument and the unscientificity argument. This move plays an important role in defending UG against other criticisms adumbrated in the preceding text. The move creates a scientific outlook for UG, which has been a shield for Chomsky to block other attacks on UG. Let us now look at how this is done.

To rebut the false-universals argument, Chomsky contends that counterexamples are dealt with in UG in the same way as they are handled in the natural sciences. In the first place, data that seem to falsify some theoretical hypotheses may be neglected, as long as 'we [can] show that certain fairly far-reaching principles interact to provide an explanation for crucial facts – the crucial nature of these facts deriving from their relation to proposed explanatory theories' (Chomsky, 1980d: 2). In other words:

Apparent counterexamples and unexplained phenomena should be carefully noted, but it is often rational to put them aside pending further study when principles of a certain degree of explanatory power are at stake. (Chomsky, 1980d: 2)

In the second place, falsification in science is seldom naïve falsification, as Lakatos has made clear. According to Smolensky and Dupoux (2009), who are UG supporters, language universals can be divided into descriptive universals and cognitive universals. The former are summarized from some linguistic data and are therefore liable to be confronted with counterexamples. The latter are meant to be about how the brain functions, and they are what UG theorists are striving to discover. Smolensky and Dupoux (2009: 468) argue:

Counterexamples to des-universals are not counterexamples to cog-universals. . . . a hypothesized cog-universal can only be falsified by engaging the full apparatus of the formal theory.

Here falsification presumably means sophisticated falsification discussed by Lakatos. So, seemingly falsifying data may, on a close look, turn out not to be so.

We have just seen how Chomsky and his followers have rebutted the unfalsifiability argument and the false-universals argument. These rebuttals might not convince the opponents, such as Ibbotson and Tomasello. What Chomsky and his supporters say on these issues might not be correct. But the problem for the opponents is that they cannot prove Chomsky wrong. The crux of the matter is that there is no precise definition of (sophisticated) falsification and no precise definition of what a scientific theory is. So the opponents of UG may still grapple with Chomsky on the issue of falsification or scientificity, but it is certain that nothing conclusive can come from such fights, at least in the short run.

So the unfalsifiability argument and the false-universals argument are not really effective against UG. What about the linking-problem argument? Ibbotson and Tomasello do not give a detailed analysis of this problem, only claiming that 'the linking problem – which should be the central problem in applying universal grammar to language learning – has never been solved or even seriously confronted' (quoted above). So their claim lacks evidence and argument. By contrast, Ambridge et al. (2014, 2015) have done better. Through examining a number of cases, they argue that UG suffers from the problem of linking, as well as the problems of data coverage and redundancy. But these researchers have at best shown that the current version of UG suffers from the linking problem (and data coverage and redundancy), and they themselves admit this:

Of course, we claim only to have shown that none of the categories, learning procedures, principles, and parameters proposed under *current* UG-based theories aid learning; we have not shown that such innate knowledge could not be useful in principle. It remains entirely possible that there are components of innate linguistic knowledge – yet to be proposed – that would demonstrably aid learning. (Ambridge et al., 2014: e82; emphasis added)

The above remark reveals a grave problem with the argument by Ambridge et al. against UG. If it is 'entirely possible' that a future version of UG can 'demonstrably aid learning', then why is UG wrong? Why not try to find a better version of UG in which the linking problem will be solved? Neither Ibbotson and Tomasello nor Ambridge et al. have shown that this route is impossible. So the linking-problem argument against UG is not successful either. The deep reason for this unsuccess is still connected with the notion of scientificity. Chomsky could insist that UG is a scientific theory, and argue that the linking problem is a scientific problem to be solved. In this way he could argue that the existence of the linking problem in no way invalidates UG.

²⁷ For more accounts of 'scientific revolutions' in linguistics, see Kertész (2010a, 2010b).

We are left with only the alternative-theory argument to examine. How effective is it against UG? Ibbotson and Tomasello seem to know well the theories of scientific research by Popper, Kuhn and Lakatos, as they remark that ‘Of course, scientists never give up on their favorite theory, even in the face of contradictory evidence, until a reasonable alternative appears’. In the mind of Ibbotson and Tomasello, such a reasonable alternative has already appeared, which is the usage-based theory. And they also write that ‘The paradigm shift is certainly not complete’, implying that there has been a paradigm shift from UG to the usage-based theory. But it seems too early to make such an assessment. UG and the usage-based theory can be regarded as competing paradigms, but it would be a complex matter to tell whether there is a paradigm shift going on, and it would be even more difficult to ascertain that the usage-based theory is winning the battle (as is implied in Ibbotson and Tomasello’s words). As Ibbotson and Tomasello admit, the usage-based theory is ‘far from offering a complete account of how language works’, and ‘There is much work to be done by usage-based theorists to explain how these forces interact in childhood in a way that exactly explains the path of language development’. Indeed, there are many important problems for usage-based theorists to solve, for example, how constructions are exactly learned, what the constraining mechanisms on analogy exactly are and how they exactly work, and how the usage-based theory can explain island-effects.²⁸ Since the usage-based theory still faces these problems, none of which is trivial, it would simply be premature for its proponents to claim victory over UG now, or to say that they will win the competition. This is also true of other alternative theories, such as those proposed by Ambridge et al. and Hofmeister and colleagues. So, Chomsky could stick to the claim that UG is a scientific theory, and argue that the existence of alternative theories (paradigms, or programs) in no way refutes UG.

To sum up, the four arguments against UG I identified in Ibbotson and Tomasello (2016) are shared by many other researchers. Ibbotson and Tomasello do not advance the unscientificity argument, but it is implied in their words and has been made explicit by other scholars. But all these arguments are not strong enough to refute UG, neither alone nor in combination. Chomsky has rebutted, or could rebut, all these criticisms by insisting that UG is scientific, and it is impossible at present to prove that UG is not.²⁹

4.2. *The critique by Everett*

Everett has been arguing that UG is wrong through a series of publications (e.g. Everett, 2005, 2009, 2012). His evidence is from the language Pirahã, spoken by a community of a few hundred speakers living along the Maici river of Amazonas, Brazil. He claims, among others, that Pirahã lacks numbers, color terms, and in particular recursion. It is the claim that Pirahã lacks recursion that has attracted the most attention in the field.³⁰ Since Chomsky asserts that recursion is the most important component of UG (see especially Hauser et al., 2002), Everett takes that the case of Pirahã poses a serious challenge to UG. His own proposal for explaining language is that culture provides a casual explanation to language. In particular, he hypothesizes a so-called ‘Immediacy of Experience Principle’ (IEP), to account for the grammatical phenomena in Pirahã. According to Everett, the Pirahã people only express immediate experiences, and this explains why numbers, color terms and recursion are absent in their language.

There are a number of problems with Everett argument against UG. First, his claims about the Pirahã language have been strongly disputed. For example, Nevins et al. (2009a, 2009b), through analyzing the same data as used by Everett, find no evidence supporting the claim that Pirahã lacks recursion. So, whether recursion is absent in Pirahã is not yet decided. It is thus premature to cite Pirahã as a counter example to UG.

Second, for the sake of argument let us suppose that Pirahã really lacks recursion. Would this falsify UG? As Everett (2009: 438–439) himself realizes, falsifying UG is a complicated matter. Essentially, he distinguishes between naïve and sophisticated falsification (although he does not use these terms). If UG says that all human languages must exhibit recursion (this is the case of UG1 in Everett, 2009), then in Everett’s eye this is falsified by Pirahã. But if UG says that recursion is a human biological property (this is the case of UG2 in Everett, 2009), then Pirahã does not falsify it. But Everett complains that in this latter case UG would be unfalsifiable. In the light of the preceding subsection, it is clear that this unfalsifiability argument cannot refute UG.

In fact Everett’s unfalsifiability argument concerning recursion has a more serious problem. While many scholars complain that Subadjacency is unfalsifiable and hence doubt that it is an innate principle, it seems that few would deny that

²⁸ For description and analysis of Chomsky’s criticisms of learning theories, including those based on analogy, generalization and distribution, see Lin (1999, 2000, 2002). See also Berwick et al. (2011) and Clark and Lappin (2011) for more recent debate on learning theories.

²⁹ There are other arguments against UG in the literature, which concern data selection, idealization, the functions of language, the nature of rules, the nature of explanation, and psychological reality, etc. Chomsky has fended off all these attacks using the scientificity shield. Limited by space, I shall not go into this matter here.

³⁰ For instance, both Evans and Levinson (2009) and Ibbotson and Tomasello (2016) make use of the claim that Pirahã lacks recursion to argue against UG.

humans have the innate ability to generate recursive expressions. Everett claims that Pirahã lacks numbers, but at the same time he writes:

If Pirahã culture radically changed, however, and suddenly there was an acute need for ordinal numbers, then presumably the Pirahãs would respond just as all cultures without ordinal numbers have responded to that need throughout human history – they would invent them or borrow them from another language that already had them. (Everett, 2012: 267–268)

So even Everett thinks that Pirahã could have numbers. If so, couldn't Pirahã have recursion too? It seems that Everett would have to answer this question positively. In that case, Everett would no longer be justified to say that Pirahã is a counter example to UG.

Everett's own culture-based theory has some serious problems too. The biggest problem is that his notion 'immediate experience' is imprecise and is difficult to remedy. As Kay (2005: 636) points out, experience of color is about as direct as experience gets. If the Pirahãs obey the IEP, why don't they use color terms to express the immediate color experiences then? Everett's (2005: 642) explanation is that color is an immediate sensation but the naming of it is not, for a property name is an abstraction. But this explanation does not go through. Consider the perception and naming of an object, say a banana. According to Locke (1690), the mind initially forms a set of simple ideas, such as yellow, hardness and impenetrability, then on the basis of this the mind builds up the complex idea of a banana. The Pirahãs do have a name for bananas (see Everett, 2005: 625). If Locke is correct, then the name *banana* must be more abstract than the name *yellow*; and since the Pirahãs have the former, they could have the latter too. In that case, Everett's own data from Pirahã will contradict the very IEP he formulates. Consider another example. According to Everett, in Pirahã one can say something like *John's son* but not *John's son's daughter*. Everett does not have a satisfactory way of explaining this using the IEP (see Nevins et al., 2009b: 676, n.11). In fact, all the explanations of the syntactic phenomena of Pirahã by Everett in terms of the IEP are problematic, precisely because of the troublesome notion of immediate experience.

Another serious problem with Everett's theory concerns falsifiability. To falsify this theory, one would need to find another culture, which satisfies the IEP and yet the language does have numbers, or color terms, or recursion, or the like. Since Everett claims that Pirahã culture is unique, it is nearly impossible to find another IEP-culture. It is also empirically impossible to construct an IEP-culture and see what kind of language the members will come up with. So it seems that Everett's theory also suffers from the problem of unfalsifiability, a very problem he identifies with UG.

Everett of course hopes that his theory is better than UG. Either explicitly or implicitly, he is employing the alternative-theory argument. But from the preceding discussion, the hope of using this argument to triumph over UG seems rather remote. So I conclude that Everett's critique of UG is not successful, and that it does not refute UG.

4.3. The critiques by Progovac and Anogianakis

Like many of Chomsky's books, the new book he co-authored with Robert C. Berwick (Berwick and Chomsky, 2016a), is also provocative. In this subsection I shall consider two critical reviews of this book, Progovac (2016) and Anogianakis (2017), the aim being to see how much effect they have against UG.

But first we must be clear about the main ideas presented in Berwick and Chomsky (2016a). As the title makes clear, the book aims to explain how language has evolved. This breaks into a number of questions, which Berwick and Chomsky (2016a: 110) summarize into the following list:

- 'What': what has evolved?
- 'Who': who has language?
- 'Where' and 'When': where and when did language emerge?
- 'How': how is language implemented in the brain?
- 'Why': why do humans have language?

But they omit a question which is perhaps more important than all these and which they have discussed in great length in the book, it being how language evolved. I shall call this the 'How-evo' question and add it into the above list; and I shall rename Berwick and Chomsky's 'How' as 'How-imp'. With this preliminary note, let us look at Berwick and Chomsky's answers to these questions.

First, the 'What' question. For many years Chomsky has been saying that there is a faculty of language, which interacts with two other mental organs: a sensory-motor system (SM, informally called 'sound') for externalization as production or parsing, and a conceptual-intentional system (CI, informally called 'thought') for inference, interpretation, planning, and the organization of action.³¹ The core of the language faculty is Merge (Internal Merge and External Merge), the reiterative

³¹ Berwick and Chomsky (2016a) rename the language faculty as 'the computational system' (CS), and say that CS interfaces with SM and CI.

application of which generates an unlimited number of expressions with hierarchical structure and recursion. An internally generated expression is passed to SM through the process of externalization, and to CI through the process of internalization (2016a: 111). In Berwick and Chomsky's opinion, externalization has not evolved (2016a: 83); and they do not discuss the evolution of CI or of internalization. For them the evolution of language is reduced to the problem of how Merge came into existence.

Second, 'Who'. Berwick and Chomsky argue that birdsongs do not have hierarchical structure, that even the famous chimp Nim Chimpsky could not manage to produce embedded, hierarchically structured sentences. They explain the difference between humans and nonhumans in terms of Merge: the former have it and the latter do not. So only we, humans, have language.

Third, 'Where' and 'When'. According to Berwick and Chomsky (2016a: 156–157), anatomically modern humans appeared (in southern Africa) about 200,000 years ago, but did not become behaviorally modern until roughly 80,000 years ago, and the African exodus was about 60,000 years ago. Modern human behavior was associated with 'rapid changes in both tools and the appearance of the first unambiguously symbolic artifacts, such as shell ornaments, pigment use, and particularly the geometric engravings' (2016a: 38). From this Berwick and Chomsky estimate that language emerged 60,000 years ago, and likely before 80,000 years ago, and that thus there was about 130,000 years of evolutionary time for language (2016a: 157), which is barely a flick of an eye in evolutionary time. So language seems to have developed rapidly, and appears to be a biological leap.

Fourth, 'How-evo'. To explain this problem, Berwick and Chomsky turn to theories of evolution for help. They distinguish between two theories: one is conventional, Darwinian, adaptionist, gradualist ('by creeps'), and the other being modern, stochastic, nongradualist ('by jerks'). They favor the nongradualist theory and try to apply it to language evolution. The reason is that they seem to think that adaptive evolution cannot scale 'fitness peaks' based on the following points, as summarized by Anogianakis:

- (a) the "genetic drift," (b) the immense amount of time required for large scale adaptation . . . , (c) the concept of "fitness" does not apply to small populations, such as the presumed early *Homo sapiens* communities and, finally,
- (d) that the mathematics of Population Genetics (Gillespie, 2004) support their contention that adaptive evolution cannot possibly scale "fitness peaks". (Anogianakis, 2017)

According to Berwick and Chomsky's nongradualist theory, Merge was the result of a slight rewiring of the brain, only a small number of individuals initially had this beneficial innovative trait, which then spread through the population (2016a: 79–80).³²

Fifth, 'How-imp'. Berwick and Chomsky speculatively posit that the word-like elements are stored in the middle temporal cortex as the 'lexicon' (2016a: 159), and that Brodmann areas 44 and 45 are where syntactic processing takes place. For Berwick and Chomsky, syntactic processing is based on Merge.

Last, 'Why'. Berwick and Chomsky list eleven major alternative theories, from language as gossip to language as 'internal mental tool' (2016a: 80–81). They adopt the last alternative. In particular, they argue that the primary function of language is not for external communication (as assumed in the other ten alternative theories) but for internal thought, through an illustration of how an expression is built up internally and how it gets externalized (2016a: 72–74). Suppose that through reiterative application of Merge, the expression *John is eating what* has been built. Now, the expression is further merged with *what*, yielding *what John is eating what*. This internal expression is then externalized for pronunciation. Pronouncing two copies of *what* would be too computationally costly, so only the first copy is retained, indicating that a question is being asked and at the same time achieving computational efficiency. The expression that gets pronounced is *what John is eating* (or better *what is John eating*), but this imposes a burden on interpretation (communication), because the hearer has to know that the object of *eating* is *what*. There is thus a conflict between computational efficiency and interpretive-communicative efficiency, and computational efficiency wins. Since language favors computational efficiency over interpretive-communicative efficiency, the primary function of language is then for internal thought and not for external communication. In the words of Berwick and Chomsky, "language evolved as an instrument of internal thought, with externalization a secondary process" (2016a: 74).

Having made clear the main questions and the answers presented in Berwick and Chomsky (2016a), we are now in the position to discuss the critical reviews by Progovic and Anogianakis. Among the criticisms by Progovic, three stand out. One concerns Berwick and Chomsky's argument that language is for thought rather than for communication. Progovic (2016) questions why pronouncing both copies of *what* enormously increases computational cost, why the first *what* should be kept to indicate that the expression is an interrogative except for the purpose of communication, and why pronouncing only one *what* shows that language evolved for thought and not for communication (2016: 994). The second major criticism by Progovic is directed against Berwick and Chomsky's idea that language evolution must be explained

³² Chomsky (2005: 12) hypothesizes that Merge occurred in a single individual.

using the nongradualist theory, by pointing out that there is an alternative, gradualist approach, which [Progovac \(2015\)](#) himself has developed (2016: 995). Progovac's third major criticism is leveled at Berwick and Chomsky's research method. He blames their work for lack of falsifiability and implies that it is unscientific (2016: 995–996).

Anogianakis's review focuses on Berwick and Chomsky's argument that that adaptive evolution cannot scale fitness peaks based on the points (a)–(d) (see above). He gives a point-by-point analysis and argues that Berwick and Chomsky are wrong, and that adaptive evolution can scale fitness peaks. He thus implies that language evolution can be explained using the gradualist theory. Anogianakis also questions Berwick and Chomsky's treatment of the relationship between language and thought. According to Berwick and Chomsky, language and thought are separate, with language evolved later than thought. But according to Anogianakis, 'it is evident that thought must have developed either before language, or, at least, in parallel with the development of language. In fact, it is obvious—as a conjecture but not proven in any way – that the very beginnings of language (and possibly the genes associated with its later development) are associated with the normal social activities of the species'. So Anogianakis thinks that, *pace* Berwick and Chomsky, language evolved for communication.

So Progovac and Anogianakis both argue against Berwick and Chomsky's proposal to explain the evolution of language using the nongradualist theory, and both take issue with Berwick and Chomsky's idea that language evolved for thought and not for communication. But they have not shown that Berwick and Chomsky are wrong. First, even if adaptive evolution can scale fitness peaks, it does not mean that it can explain the evolution of all organisms. For example, Darwin's two-cell prototype eye does not seem to be the result of some laborious trial-and-error selection process, rather it seems to be caused by two stochastic and abrupt events ([Berwick and Chomsky, 2016a: 31–32](#)). If the evolution of the eye is to be explained by the nongradualist theory, then why not the evolution of language too? Progovac and Anogianakis have only argued that the gradualist theory is a viable alternative, but they have not shown that the nongradualist alternative is fundamentally wrong. Similarly, both reviewers point out problems with Berwick and Chomsky's argument that language evolved for thought and not for communication, but they have not proven them to be wrong. As for Progovac's unfalsifiability and unscientificity argument, it cannot really do much damage to Berwick and Chomsky's theory, as can be seen from Sections 4.1 and 4.2 above.

Among the six questions above, Progovac and Anogianakis have only dealt with 'How-evo' and 'Why'. About Berwick and Chomsky's answers to the other four questions, these reviewers have said little, if not nothing. On the whole, Progovac and Anogianakis have raised some problems, but these problems are not life-threatening to Berwick and Chomsky's work on the evolution of language.

What about UG? Progovac complains that some notions Berwick and Chomsky employ are imprecise, e.g. Merge, perfect system and computational cost ([Progovac, 2016: 993–994](#)). But this is a very weak criticism. Anogianakis focuses exclusively on the theory of evolution, and hence says little about where UG itself is right or wrong. So both reviews, though very critical of [Berwick and Chomsky \(2016a\)](#), have little force against UG.

In this paper my focus is UG, so what I want to do now is to examine it in connection with Berwick and Chomsky's new book. The question to be answered is this: in Sections 2 and 3 above it was argued that UG is seriously flawed and deeply problematic, but now Berwick and Chomsky has written what seems to be an important book on the evolution of language, doesn't this show that UG is important after all? This is a challenging question, and I would like to deal with it by examining the relationship between Berwick and Chomsky's answers to the six questions on the one hand and UG on the other. The six questions, 'What', 'Who', 'When', 'How-evo', 'How-imp', and 'Why', are all interesting,³³ and few would deny their importance. But must we resort to UG in order to answer them?

Consider 'What' first. Since the topic is the evolution of language, what has evolved is evidently language. So what is language? What counts as language? Any system, if it counts as a language, must be able, internally, to express thought and, externally, to show itself through symbols, sounds, gestures or the like. So it is innocuous to say, as Berwick and Chomsky do, that language interfaces with CI and SM. What is more important is the question of what language is. This is a matter of definition, which affects also the answer to the 'Who' question. If language is defined simply as a set of expressions which can express meaning, then many animals too have language. For the sake of argument, let us adopt the definition of language offered by Berwick and Chomsky: language must have such features as combinationality (word-like elements are combined to form expressions), hierarchical structure, embedding, etc. Let us call these the 'defining features' of language. It is by these external features that we judge whether a system is a language or not. It is not necessary at all to invoke Chomsky's technical notions of Internal Merge and External Merge for the determination of language. In fact, it is only after having determined a genuine instance of language using those defining features, be it a public language or a particular individual's language, can one proceed to characterize it, in terms of a theory of grammar (UG, with Merge as a part, being only one alternative). In sum, UG is not necessary for answering the 'What' question.

Next, consider 'Who' and 'When'. As said above, what language is a matter of definition. Let us assume that language must have those defining features such as hierarchical structure and embedding. Then whether animals have language

³³ I omit the 'Where' question here and in below, for it seems not as important as the 'When' question.

depends on whether their communication systems exhibit those features. This can be done independently of the hypothetical notion of Merge. Similarly, when trying to determine the time line of language evolution, e.g. whether language emerged 50,000 years ago (Chomsky, 2005; see also Berwick and Chomsky, 2011), or 60,000–80,000 years ago (Berwick and Chomsky, 2016a), or 70,000–100,000 years ago (Aitchison, 1996), or 157,000 years ago (Berwick and Chomsky, 2016b), or 400,000–500,000 years ago (Dediu and Levinson, 2013), we rely on direct or indirect evidence showing when the relevant systems began to have hierarchical structure, embedding, etc., and we do not do this in terms of the hypothetical Merge.

Now consider ‘How-evo’. Did language evolve by creeps or by jerks? This is an interesting question. What we need to do is investigate how humans came to have the ability to combine elements into expressions with hierarchical structure and embedding. Maybe this ability was achieved adaptively and gradually (as Progovac and Anogianakis suggest), or maybe it was caused nongradually by sudden mutations and rapid increase of the brain size (as Berwick and Chomsky postulate). There are many ways to produce expressions that are combinational, hierarchical and with embedding, as shown in various grammatical theories such as head-driven phrase structure grammar, lexical-functional grammar, tree adjoining grammar, and combinatory categorial grammar. Each of these theories provides an explanation of what the combinational ability is. Even Berwick and Chomsky (2016a: 113–114) admit that these theories are similar to their Merge-based theory. The details of the combination operation are all different in these theories (including UG), but the evolution of this operation can be investigated in the same way, i.e. at certain time in the history the human brain began to be able to perform this operation. So the answer to ‘How-evo’ does not need to be framed exclusively in UG terms.

Let us proceed to consider ‘How-imp’. It is of course valid to investigate where the word-like elements are stored in the brain and where syntactic processing takes place. But it is not necessary to conduct such research in terms of Internal Merge and External Merge.

Finally, consider ‘Why’. Progovac and Anogianakis have pointed out some problems with Berwick and Chomsky’s specific answer. I want now to consider the ‘Why’ question itself: did language evolve for thought or for communication? To answer this question, we must know what thought and communication are, at least know certain things about them. Take the expression *John tickled Mary* for example. In order for this expression to be someone’s thought, with the meaning that John did the tickling and Mary got tickled (and not the other way round), the person himself must know that the same expression on different days retains the same meaning. If the person one day interpreted the expression in the order of SVO, and another day understood it in the order of OVS,³⁴ then he would not have clear thought, and thought would not be possible for him. Similarly, in order for the person to communicate to others the meaning that John tickled Mary, he must not utter *John tickled Mary* one day and *Mary tickled John* another day. This shows that thought and communication are fundamentally similar: to have thought is to be able to communicate with oneself (at different times). So the ‘Why’ question, if put as ‘Did language evolve for thought or for communication?’, turns out to be misleading, for it suggests that thought and communication are very different. An appropriate answer to the ‘Why’ question would be something like: language evolved both for thought and for communication. A more adequate answer will require a more adequate conceptual analysis of the notions of language, thought and communication, but this can be conducted independently of any specific theories of grammar, UG included.

To sum up, the six questions concerning the evolution of language are all intriguing, but the search for their answers can be carried out without considering UG at all. Berwick and Chomsky plant UG in the discussion of all the six questions, giving the appearance that UG is important and indispensable in investigating the evolution of language. This subsection has however shown that it is nothing other than an appearance.

4.4. Summary

In this section I analyzed various arguments against UG, especially those made by Ibbotson and Tomasello, Everett, Progovac and Anogianakis. I showed that these arguments all fail to refute UG. The deep reason for this failure is that there is no precise definition of falsification and scientificity. Chomsky insists that UG is scientific and uses (or could use) this as a shield to fend off all the arguments discussed here, and it is impossible at present to prove that UG is not scientific.

In comparison, my critique of UG in Sections 2 and 3 took a rather different route. I completely avoided talking about whether UG is unfalsifiable or unscientific; rather, I focused on whether Chomsky’s research method can lead to the discovery of any innate language universals. Since my critique does not appeal to any controversial concepts or principles, I believe that it can succeed in refuting UG.

In this section I also examined Chomsky’s recent work on the evolution of language. What Chomsky does, intentionally or unintentionally, is to hide UG under the smoke of the evolution issue. I showed that when the smoke is dispersed, the fundamental flaw in UG is still there as before. Talking about the evolution of language cannot save UG.

³⁴ There are indeed languages with the order OVS, such as Apalaí and Hixkaryana.

5. Conclusion

Since ancient times humans have found language perplexing and have made continuous effort to understand its nature. Chomsky's UG is one important attempt in uncovering the mystery of language. The motivation of UG is highly understandable, yet the method used in finding it is seriously flawed. UG theorists try to discover UG principles, parameters and so on, by examining certain grammatical and ungrammatical sentences in some languages. It is impossible to find the intended innate language universals using this method. Talking about issues such as the evolution of language by Chomsky might have the effect of diverting the critics' attention away from UG's problems, but it cannot remove them.

Many things have prevented people (including the proponents of UG) from seeing that UG is deeply flawed. The analogies between language and scientific theories, such as the thermonuclear theory, Evo-Devo and Marr's theory of vision, have fostered the belief that UG is a scientific theory. Various considerations, e.g. the argument from the poverty of the stimulus, the best-theory argument and the Martian-scientist argument, have contributed to the strengthening of this belief. A bewildering array of research results that have kept appearing in the forms laws (such as Subjacency and Binding Principles) and entities (such as parameters, functional and lexical categories, and lexical features) have given the impression that UG is making steady progress toward the hidden nature of language. But on careful analysis, the analogies turn out to be misleading, the belief unjustified, and the impression of progress false.

There might be innate constraints on the child's acquisition of language, and there might be other innate mechanisms responsible for language; but these cannot be found with the method employed in the theory of UG. If we want to find any innate language universals (e.g. principles, categories, etc.) at all, we must investigate the processing requirements of language on the brain, and get to know more about the brain structure.

References

- Aboh, E.O., Guasti, M.T., Roberts, I. (Eds.), 2013. *Locality* (Oxford Studies in Comparative Syntax). Oxford University Press, Oxford.
- Aitchison, J., 1996. *The Seeds of Speech: Language Origin and Evolution*. Cambridge University Press, Cambridge.
- Allwood, J., 1982. The complex NP-constraint as a non-universal rule. In: Engdahl, E., Ejerhed, E. (Eds.), *Readings on Unbounded Dependencies in Scandinavian Languages*. Acta Universitatis Umensis and Almqvist & Wiksell International, Stockholm, pp. 15–32.
- Ambridge, B., Pine, J.M., Lieven, E.V.M., 2014. Child language acquisition: why universal grammar doesn't help. *Language* 90 (3), e53–e90.
- Ambridge, B., Pine, J.M., Lieven, E.V.M., 2015. Explanatory adequacy is not enough: response to commentators on 'Child language acquisition: why universal grammar doesn't help'. *Language* 91 (3), e116–e126.
- Andor, J., 2004. The master and his performances: an interview with Noam Chomsky. *Intercult. Pragmat.* 1, 93–111.
- Anogianakis, G., 2017. Reinterpreting Darwin to explain evolution of language: a review of Why Only Us: Language and Evolution, R.C. Berwick, N. Chomsky. *Lingua*. <http://dx.doi.org/10.1016/j.lingua.2017.01.001>
- Antony, L.M., Hornstein, N., 2003. *Chomsky and His Critics*. Blackwell, Oxford.
- Berwick, R.C., Chomsky, N., 2011. The biolinguistic program: the current state of its development. In: Di Sciullo, A.M., Boeckx, C. (Eds.), *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. Cambridge University Press, Cambridge, pp. 19–41.
- Berwick, R.C., Chomsky, N., 2016a. *Why Only Us: Language and Evolution*. MIT Press, Cambridge, MA.
- Berwick, R.C., Chomsky, N., 2016b. Why only us: recent questions and answers. *J. Neurolinguist.* <http://dx.doi.org/10.1016/j.jneuroling.2016.12.002>
- Berwick, R., Pietroski, P., Yankama, B., Chomsky, N., 2011. Poverty of the stimulus revisited. *Cognit. Sci.* 35, 1207–1242.
- Boeckx, R.C., 2006. *Linguistic Minimalism: Origins, Concepts, Methods and Aims*. Oxford University Press, Oxford.
- Carroll, S.B., 2005. *Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom*. W.W. Norton & Company, New York.
- Cherniak, C., 2005. Innateness and brain-wiring optimization: non-genomic nativism. In: Zilhao, A. (Ed.), *Cognition, Evolution, and Rationality*. Routledge, London, pp. 103–112.
- Cherniak, C., Mokhtazada, Z., Rodriguez-Esteban, R., Changizi, K., 2004. Global optimization of cerebral cortex layout. *Proc. Natl. Acad. Sci.* 101, 1081–1086.
- Chomsky, N., 1968. *Language and Mind*. Harcourt, Brace and World, New York.
- Chomsky, N., 1973. Conditions on transformation. In: Anderson, S.R., Kiparsky, P. (Eds.), *A Festschrift for Morris Halle*. Rinehart & Winston, New York, pp. 232–286.
- Chomsky, N., 1975. *Reflections on Language*. Pantheon Book, New York.
- Chomsky, N., 1980a. *Rules and Representation*. Blackwell, Oxford.
- Chomsky, N., 1980b. Rules and representations. *Behav. Brain Sci.* 3, 1–61.
- Chomsky, N., 1980c. Contributions. In: Piattelli-Palmarini, M. (Ed.), *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*. Routledge & Kegan Paul, London.
- Chomsky, N., 1980d. On binding. *Linguist. Inq.* 11 (1), 1–46.
- Chomsky, N., 1981a. *Lectures on Government and Binding*. Foris, Dordrecht.
- Chomsky, N., 1981b. Principles and parameters in syntactic theory. In: Hornstein, N., Lightfoot, D. (Eds.), *Explanation in Linguistics: The Logical Problem of Language Acquisition*. Longman, London, pp. 32–75.
- Chomsky, N., 1986. *Knowledge of Language: Its Nature, Origin, and Use*. Praeger, London.

- Chomsky, N., 1988. *Language and Problems of Knowledge: The Managua Lectures*. MIT Press, Cambridge, MA.
- Chomsky, N., 1993. *Language and Thought*. Moyer Bell, Wakefield, RI.
- Chomsky, N., 1994. Naturalism and dualism in the study of language and mind. *Int. J. Philos. Stud.* 2, 181–209.
- Chomsky, N., 1995a. *The Minimalist Program*. MIT Press, London.
- Chomsky, N., 1995b. Language and nature. *Mind* 104, 1–61.
- Chomsky, N., 1995c. Bare phrase structure. In: Campos, H., Kempchinsky, P. (Eds.), *Evolution and Revolution in Linguistic Theory*. Georgetown University Press, Washington, pp. 51–109.
- Chomsky, N., 2000. *New Horizons in the Studies of Language and Mind*. Cambridge University Press, Cambridge.
- Chomsky, N., 2002. *On Nature and Language*. Cambridge University Press, Cambridge.
- Chomsky, N., 2004. *The Generative Enterprise Revisited*. Mouton de Gruyter, New York.
- Chomsky, N., 2005. Three factors in language design. *Linguist. Inq.* 36, 1–22.
- Chomsky, N., 2007a. Bilingualistic explorations: design, development, evolution. *Int. J. Philos. Stud.* 15, 1–21.
- Chomsky, N., 2007b. Approaching UG from below. In: Sauerland, U., Hans-Martin, G. (Eds.), *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*. Mouton de Gruyter, New York, pp. 1–29.
- Chomsky, N., 2008. On phases. In: Freidin, R., Otero, C.P., Zubizarreta, M.L. (Eds.), *Foundational Issues in Linguistic Theory*. MIT Press, Cambridge, MA, pp. 133–166.
- Chomsky, N., 2009. The mysteries of nature: how deeply hidden? *J. Philos.* 106, 167–200.
- Chomsky, N., 2012. *The Science of Language: Interviews with James McGilvray*. Cambridge University Press, Cambridge.
- Chomsky, N., Hauser, M.D., Fitch, W.T., 2005. Appendix. The Minimalist Program. <http://arti.vub.ac.be/cursus/2005-2006/mwo/03b-mp.pdf> (accessed 20.03.17)
- Clark, A., Lappin, S., 2011. *Linguistic Nativism and the Poverty of the Stimulus*. Wiley-Blackwell, Oxford.
- Collins, C., 2001. Economy conditions in syntax. In: Baltin, M., Collins, C. (Eds.), *The Handbook of Contemporary Syntactic Theory*. Blackwell, Oxford, pp. 45–61.
- Comrie, B., 1989. *Language Universals and Linguistic Typology*, 2nd ed. University of Chicago Press, Chicago.
- Cook, V.J., Newson, M., 2007. *Chomsky's Universal Grammar: An Introduction*, 3rd ed. Blackwell, Oxford.
- Cowie, F., 1999. *What's Within? Nativism Reconsidered*. Oxford University Press, Oxford.
- Dąbrowska, E., 2015. What exactly is Universal Grammar, and has anyone seen it? *Front. Psychol.* 6, 852. <http://dx.doi.org/10.3389/fpsyg.2015.00852>
- Dediu, D., Levinson, S.C., 2013. On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. *Front. Psychol.* 4, 397. <http://dx.doi.org/10.3389/fpsyg.2013.00397>
- Evans, N., Levinson, S.C., 2009. The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–492.
- Everett, D.L., 2005. Cultural constraints on grammar and cognition in Pirahã. *Curr. Anthropol.* 46, 621–646.
- Everett, D.L., 2009. Pirahã culture and grammar: a response to some criticisms. *Language* 85, 405–442.
- Everett, D.L., 2012. *Language: The Cultural Tool*. Pantheon Books, New York.
- Fitch, W.T., Hauser, M.D., Chomsky, N., 2005. The evolution of the language faculty: clarifications and implications. *Cognition* 97, 179–210.
- Gärtner, H.-M., Michaelis, J., 2005. A note on the complexity of constraint interaction: locality conditions and minimalist grammars. In: Blache, P., Stabler, E., Busquets, J., Moot, R. (Eds.), *Logical Aspects of Computational Linguistics*. Springer-Verlag, Berlin, pp. 114–130.
- Gillespie, J., 2004. *Population Genetics: A Concise Guide*. Johns Hopkins University Press, Baltimore.
- Goodman, N., 1967. The epistemological argument. *Synthese* 17, 23–28.
- Haegeman, L., 1994. *Introduction to Government and Binding Theory*, 2nd ed. Blackwell, Oxford.
- Hauser, M.D., Chomsky, N., Fitch, W.T., 2002. The faculty of language: what is it, who has it and how did it evolve? *Science* 298, 1569–1579.
- Hofmeister, P., Sag, I.A., 2010. Cognitive constraints and island effects. *Language* 86, 366–415.
- Hofmeister, P., Casasanto, L.S., Sag, I.A., 2012a. How do individual cognitive differences relate to acceptability judgments? A reply to Sprouse, Wagers, and Phillips. *Language* 88, 390–400.
- Hofmeister, P., Casasanto, L.S., Sag, I.A., 2012b. Misapplying working memory tests: a reductio ad absurdum. *Language* 88, 408–409.
- Hornstein, N., Nunes, J., Grohmann, K.K., 2005. *Understanding Minimalism*. Cambridge University Press, Cambridge.
- Ibbotson, P., Tomasello, M., 2016. Evidence rebuts Chomsky's theory of language learning. *Sci. Am.* (September 7), 2016.
- Itkonen, E., 1996. Concerning the generative paradigm. *J. Pragmat.* 25, 471–501.
- Jacob, F., 1978. *Darwinism reconsidered*. Atlas World Press Rev. (January).
- Joos, M. (Ed.), 1957. *Readings in Linguistics*. Chicago University Press, Chicago.
- Kaufman, J.C., Kaufman, A.B., 2016. Capacity, potential, and ability: integrating different approaches to studying animal vs human creative processes. *RUDN J. Psychol. Pedagog.* 4, 29–36.
- Kay, P., 2005. Commentary on Everett (2005). *Curr. Anthropol.* 46, 636–637.
- Kayne, R., 2000. *Parameters and Universals*. Oxford University Press, Oxford.
- Kertész, A., 2010a. From 'scientific revolution' to 'unscientific revolution': an analysis of approaches to the history of generative linguistics. *Lang. Sci.* 32, 507–527.
- Kertész, A., 2010b. Two notions of 'research program' and the historiography of generative linguistics. *Historiogr. Linguist.* 37, 165–191.
- Kuhn, T.S., 1970. *The Structure of Scientific Revolutions*, 2nd ed. University of Chicago Press, Chicago.
- Lakatos, I., 1970. Falsification and the methodology of scientific research programmes. In: Lakatos, I., Musgrave, A. (Eds.), *Criticism and the Growth of Knowledge*. Cambridge University Press, Cambridge, pp. 91–195.
- Lappin, S., Levine, R.D., Johnson, D.E., 2000. The structure of unscientific revolutions. *Nat. Lang. Linguist. Theory* 18, 665–671.
- Lasnik, H., 2010. Brief Overview of Subadjacency/Islands. <http://ling.umd.edu/~lasnik/LING610%202010/Brief%20Subadjacency%20overview.pdf> (accessed 20.03.17)
- Laurence, S., Margolis, E., 2001. The poverty of the stimulus argument. *Br. J. Philos. Sci.* 52, 217–276.
- Lin, F.Y., 1999. Chomsky on the 'ordinary language' view of language. *Synthese* 120, 151–192.

- Lin, F.Y., 2000. The transformations of transformations. *Lang. Commun.* 20, 197–253.
- Lin, F.Y., 2002. On discovery procedures. In: Nevin, B. (Ed.), *The Legacy of Zellig Harris: Language and Information into the 21st Century. Philosophy of Science, Syntax, and Semantics*, vol. 1. John Benjamins, Amsterdam, pp. 69–86.
- Lin, F.Y., 2015. What is really wrong with universal grammar (Commentary on Behme). *Language* 91 (2), e27–e30.
- Locke, J., 1690. *An Essay Concerning Human Understanding*. Thomas Bassett, London.
- Marr, D., 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company, New York.
- Marr, D., Nishirara, H.K., 1978. Visual information processing: artificial intelligence and the sensorium of sight. *Tech. Rev.* 81, 2–23.
- Nevins, A., Pesetsky, D., Rodrigues, C., 2009a. Pirahã exceptionality: a reassessment. *Language* 85, 355–404.
- Nevins, A., Pesetsky, D., Rodrigues, C., 2009b. Evidence and argumentation: a reply to Everett (2009). *Language* 85, 671–681.
- Newmeyer, F.J., 2004. Typological evidence and universal grammar. *Stud. Lang.* 28, 527–548.
- Newmeyer, F.J., 2005. *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. Oxford University Press, Oxford.
- Newmeyer, F.J., 2008. Universals in syntax. *Linguist. Rev.* 25, 35–82.
- Ouhalla, J., 1999. *Introducing Transformational Grammar: From Rules to Principles and Parameters*, 2nd ed. Edward Arnold, London.
- Piatelli-Palmarini, M., Uriagereka, J., Salaburu, P. (Eds.), 2009. *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. Oxford University Press, Oxford.
- Popper, K.R., 1959. *The Logic of Scientific Discovery*. Hutchinson, London.
- Progovac, L., 2015. *Evolutionary Syntax*. Oxford University Press, Oxford.
- Progovac, L., 2016. Review of *Why Only Us: Language and Evolution* by Robert C. Berwick and Noam Chomsky. *Language* 92, 992–996.
- Pullum, G.K., Scholz, B.C., 2002. Empirical assessment of stimulus poverty arguments. *Linguist. Rev.* 19, 9–50.
- Putnam, H., 1967. The 'innateness hypothesis' and explanatory models in linguistics. *Synthese* 17, 12–22.
- Rizzi, L., 2014. On the elements of syntactic variation. In: Picallo, C. (Ed.), *Linguistic Variation in the Minimalist Framework*. Oxford University Press, Oxford, pp. 13–35.
- Roberts, I. (Ed.), 2007. *Comparative Grammar: Critical Concepts in Linguistics. Wh-movement*, vol. 4. Routledge, London.
- Sampson, G., 2005. *The 'Language Instinct' Debate: Revised Edition*. Continuum, London.
- Seuren, P.A.M., 2004. *Chomsky's Minimalism*. Oxford University Press, Oxford.
- Smith, N.V., 2004. *Chomsky: Ideas and Ideals*, 2nd ed. Cambridge University Press, Cambridge.
- Smolensky, P., Dupoux, P., 2009. Universals in cognitive theories of language. *Behav. Brain Sci.* 32, 468–469.
- Stemmer, B., 1999. An on-line interview with Noam Chomsky: on the nature of pragmatics and related issues. *Brain Lang.* 68, 393–401.
- Stokhof, M., van Lambalgen, M., 2011. Abstractions and idealisations: the construction of modern linguistics. *Theor. Linguist.* 37, 1–26.
- Stroik, T.S., Putnam, M.T., 2013. *The Structural Design of Language*. Cambridge University Press, Cambridge.
- ten Hacken, P., 2007. *Chomskyan Linguistics and Its Competitors*. Equinox, London.
- Thomas, M., 2002. Development of the concept of 'the poverty of the stimulus'. *Linguist. Rev.* 19, 51–71.
- Tomasello, M., 2003. *Constructing A Language: A Usage-based Theory of Child Language Acquisition*. Harvard University Press, Cambridge, MA.
- Tomasello, M., 2007. What kind of evidence could refute the UG hypothesis? In: Penke, M., Rosenbach, A. (Eds.), *What Counts as Evidence in Linguistics: The Case of Innateness*. John Benjamins, Amsterdam, pp. 175–178.
- Tomasello, M., 2009. Universal grammar is dead. *Behav. Brain Sci.* 32, 470–471.
- van Valin, R.D., LaPolla, R.J., 1997. *Syntax: Structure, Meaning and Function*. Cambridge University Press, Cambridge.