

The Strong Minimalist Thesis is too strong: syntax is more than just Merge*

Deniz Satık, Harvard University
deniz@g.harvard.edu

This paper raises specific puzzles for the Strong Minimalist Thesis (SMT) based on certain crosslinguistic patterns. I do so by pointing out that the SMT entails two undesirable consequences: first, the SMT assumes that the Borer-Chomsky Conjecture is true; in other words, that all syntactic variation across languages is due to lexical differences. Second, it assumes that there can be no ordering restrictions on Merge, because they would imply the existence of an independent linguistically proprietary entity. I first present crosslinguistic evidence from case and agreement that the Borer-Chomsky Conjecture alone is not sufficient to account for syntactic variation. I then present evidence for the existence of ordering restrictions on Merge, based on a cartographic distinction between high and low complementizers. I argue that both of these patterns are purely syntactic, in that they are independent of Merge. I conclude that these independent problems raise puzzles for saltationist theories of language evolution.

Keywords: minimalism, strong minimalist thesis, language evolution, parameters, Merge, cartography

1 Introduction

Determining how language evolved is a notoriously vexing problem for multiple reasons. There is, of course, the dearth of empirical evidence: early human language has not left behind many archaeological traces. Regardless, the little evidence that is available has allowed archaeologists and paleoanthropologists to preliminarily establish the following facts about human language evolution, which ends up making the problem of language evolution even more troubling.¹ First, humanity can trace its origins to a group of anatomically modern *Homo sapiens* living in eastern or southern Africa 150,000–200,000 years ago. These humans had language, or at the very least were linguistically capable: for instance, engravings on red-ochre in South Africa provide evidence for symbolic and abstract thought.²

By contrast, it appears that the Neanderthals did not have such capacity for symbolic thinking, who were present in Europe even as recently as 40,000 years ago. As Pagel (2013) points

*I am indebted to two anonymous reviewers, Susi Wurmbrand and Ljiljana Progovac for comments.

¹My discussion here of the archaeological evidence is based on Pagel (2017).

²See Henshilwood et al. (2002) and Henshilwood & Dubreuil (2009) for further discussion.

out, art, sculptures, musical instruments and specialized tools made by *Homo sapiens* in Europe had become very common at that point. But there is no such evidence of similar creations by Neanderthals—in fact, it appears that they did not even sew their own clothes, instead merely draping themselves with skin. It appears, then, that language must have evolved 150,000-200,000 years ago, together with the first population of *Homo sapiens*.

It is hard to reconcile this with the complexity of language. To see why, let us start with Chomsky (1980)'s Principles and Parameters (P&P) framework, which will soon be presented in further detail. In P&P, differences between languages were captured via parametrizing a finite set of permissible perturbations. For example, it is well known that a language like Turkish is classified as head-final (at least for the most part), whereas a language like English is classified head-initial. One would have the parameter head-initial vs. head-final to account for this difference. Thus, all syntactic possibilities across languages might be accounted for in terms of different parametric values. Syntax, then, ends up looking quite similar to the Periodic Table, like atoms combining into many possible different molecules, as Baker (2002) has suggested.

Now, generative linguists commonly accept that linguistic capacity is due to the *faculty of language*, or *universal grammar* (UG), which is innate to all human beings. In other words, there is an innate system of mechanisms and principles that are unique to humans which is used for language acquisition. Chomsky (2000b) calls this innate system a "language organ": for generative linguists, it is the object of study in the same way that biologists study literal organs like the heart or the lungs.

But the conjunction of the fact that language is innate together with the P&P framework would entail a paradox: namely that it is impossible for so many parameters, potentially in the hundreds, or even thousands, to have evolved. There are two reasons for this: most pressing, such parameters could not have evolved in a mere 150,000-200,000 years, which is a very short amount of time for evolutionary change. Significant change often takes millions of years. And it is also hard to imagine that such parameters could have exerted a strong enough evolutionary pressure to lead to "fruitful sex" in the words of Lightfoot (1991).

By the early 1990s, due to the problems just mentioned above, Chomsky and other linguists had reasonably questioned the amount of computational principles in UG. The most optimal solution would be to assume that UG reduces to the absolute simplest computational principles—perhaps nothing more than the recursive, structure-building operation Merge—which has been called the Strong Minimalist Thesis (SMT). The SMT has been defined in many different ways in the literature.³ But more generally, one could view the SMT as claiming that all of the properties of human language syntax can be derived from three things:

- (1) a. the syntactic operation Merge
- b. interface conditions (involving semantic and phonetic restrictions)
- c. principles of "efficient computation"

Under this way of thinking about UG, Merge is the only linguistically proprietary entity. Chomsky (2020) points out that this radical conclusion seems paradoxical: properties like the linear order of words and copy deletion have nothing to do with language per se. These simply arise from interface conditions and/or principles of efficient computation, both of which are language-independent. The main attraction of the SMT is that it provides an immediate solution of Dar-

³I will focus on a definition provided by Chomsky (2004) in section 2.

win's problem. That is, it makes it conceivable for language to have evolved suddenly, as the result of a single mutation, which endowed the operation Merge onto a single individual 150,000-200,000 years ago. This mutation could have, indeed, led to "fruitful sex," given the great advantages in communication that possessors of the mutation would have had.

The goal of this paper is to show that the SMT is too strong: Merge is not the only linguistically proprietary entity. This is due to the existence of crosslinguistic patterns that appear difficult to account for if Merge is all there is to the faculty of language. As such, Darwin's problem remains just as burdensome. To restate this point more simply, I believe that there is more to syntax, and to language in general, than just Merge.

Although the SMT has been questioned a great deal already in the literature, this paper stands out in that it goes beyond merely questioning it: it provides two specific pieces of evidence where the SMT might go astray. In particular, what both of these patterns have in common is that they involve purely syntactic crosslinguistic generalizations which appear difficult to explain by language-independent interface conditions, or by principles of efficiency. Although I believe that the guiding philosophy of the Minimalist Program is right, a "Weak" Minimalist Thesis may be needed to account for the crosslinguistic patterns that we will see in this paper.

I will first argue for the existence of macroparameters in the sense of Baker (2008a). The SMT, as we will discuss in more detail, entails that all parameters that lead to crosslinguistic variation are attributed to the differences in the features of lexical items. Such parameters are called microparameters. According to Baker, there are parameters within the general principles that shape natural language syntax; in other words, microparameters alone are not sufficient to account for crosslinguistic variation. I will provide Baker's (2008a) crosslinguistic evidence, based on a survey of 108 languages, in favor of the existence of such macroparameters. If Baker is right, I will claim that such macroparameters are also proprietary, in addition to Merge—even if macroparameters can be reduced to microparameters.

I will then provide a second, independent argument in favor of other linguistically proprietary entities. This is based on the cartographic enterprise first developed by Rizzi (1997) and Cinque (1999). Rizzi and Cinque have developed a very finely ordered cartographic blueprint for syntactic structure: Rizzi for the C domain of a clause, and Cinque for the positioning of adverbs within a clause. Although I will grant that much of the blueprint could be reduced to semantic interface conditions following Ernst (2002), I will argue that not all of it can be reduced to interface conditions. To be more specific, I will argue that the positioning of high and low complementizers is linguistically proprietary, because it is a purely syntactic property.

This paper is structured as follows. Section 2 presents Chomsky's (2004) definition of the SMT, together with a discussion of what it would entail and what would falsify it, in addition to a discussion of the P&P framework which predated the Minimalist Program. Section 3 introduces the reader to Baker's work on macroparameters in syntax, concluding that, at the very least, there must be linguistically proprietary sources of these macroparameters. Section 4 introduces the reader to Rizzi and Cinque's framework, concluding that not all of the cartographic hierarchy can be reduced to the interfaces. Section 5 proposes that syntax ought to return to the P&P framework. The consequence of this is that Darwin's problem remains, given the falsity of the SMT. I provide a tentative discussion of how one might attempt to solve this problem via Progovac's (2004) gradualist approach to language evolution. Section 6 concludes.

2 Background

This section provides a background on the Principles and Parameters framework in 2.1, while introducing the reader to a more formal definition of the SMT in 2.2. 2.3 lays the foundation of how criticism of the SMT might look like, pinpointing particular areas of vulnerability that might arise once we start looking at more language-specific evidence.

2.1 Principles and Parameters

Linguistic variation is ubiquitous. Every aspect of language, including syntax and phonology, seems to vary across languages. Under the terminology of Chomsky (1986), the linguistic variants which are cultural entities—in the sense that anyone who reads this paper is an English speaker, for instance—are E-languages. However, individuals who speak the cultural entity we call *English* each have their own way of internalizing the set of rules and systems that characterize it. For instance, while some speakers may permit weak crossover constructions (*Who_i does her_i father love?*), others may not. Under Chomsky's terminology, each speaker has their own I-language, and variation is found in both E- and I-languages.

Linguistic theory has been driven by the search for language universals—properties which all languages have in common. There have been two paths which linguists have taken in this search. Under the Greenberg (1963) approach to language universals, language typologists catalogue the structural features of languages to find common patterns across them. Greenberg's original sample had 30 languages; the *World Atlas of Language Structures* currently reports data from a total of 2662 languages.

More relevant for our purposes is the second path, inspired by the work of Noam Chomsky, who has consistently argued that humans have a biological predisposition to learning language. The *poverty of stimulus* argument, though controversial, is exceedingly simple. It is difficult to reconcile the fact that languages are extremely complex together with the observation that children pick it up very quickly with little instruction needed.

This indicates the presence of some kind of innate cognitive bias shared by all humans, which *constrains* the hypothesis space in which the learning of languages takes place. One way of characterizing this bias is to constrain the possible grammar that a language can have. Given that, by definition, such constraints must be universal, ought to manifest as structural crosslinguistic universals—this is UG. This paper takes for granted that UG exists, though as we will see, generative linguists might disagree on what UG consists of.

Given that UG raises constraints on the set of possible I-languages, this raises an immediate problem. On one hand, we have very robust constraints on what languages can look like. On the other, we witness a great deal of structural crosslinguistic variation that seems hard to accommodate with the existence of UG. As such, Chomsky (1981) developed the aforementioned P&P framework to reconcile UG together with linguistic variation. First, UG has principles which provide constraints on possible grammars; second, parameters specify the degree to which these possible grammars can vary. Both principles and parameters are innate, admittedly increasing the complexity of the innate capacity for language and raising Darwin's problem for the evolution of language. How could such principles and parameters have evolved?

2.2 The Strong Minimalist Thesis

Indeed, as Berwick & Chomsky (2016) (B&C) point out, any theory of UG, at a bare minimum, has to meet the condition of evolvability. It becomes more and more difficult to meet that condition as we stipulate the presence of additional computational mechanisms like principles and parameters that are innate to all humans. According to Berwick and Chomsky, the only way to meet this burden is via stipulating that syntax itself is simple, and that it evolved as the result of a single mutation. For them, the only serious way to approach the problem of language evolution is to assume that syntax is nothing more than the single and optimal syntactic operation Merge, allowing for recursive sentence structure. This is the simple idea behind the SMT.

Chomsky (2000a) provides a similar definition of the SMT as follows: *language is an optimal solution to legibility conditions*. This follows B&C's assumption that the generative process is optimal from the perspective of efficient computation. Language keeps to Merge, which is the simplest possible recursive operation that is capable of satisfying interface conditions while being efficient. B&C compare language to snowflakes, which is shaped by the laws of nature. By contrast, language is shaped by the interfaces and principles of efficient computation.

Each derivation, at its conclusion, is accessed by the phonological and semantic interfaces for further evaluation. The phonological interface is instantiated by a sensorimotor system for externalization, such as production or parsing. It might be responsible for, among other things, the deletion of Copies in a syntactic derivation. The semantic interface, on the other hand, is instantiated by a conceptual-intentional system for "thought," namely inference, planning and interpretation, among other things. Conditions on representations such as Case theory, binding theory, control theory, θ -Criterion might all be accounted for via this system. These systems are, however, language external, because they are not a part of UG.

We are now ready to present the more formal definition of the SMT by Chomsky (2004). Suppose that the faculty of language has a genetically determined initial state S_0 . S_0 , which is UG, determines all the possible states that a particular language L can be. The goal of the minimalist is to reduce the number of elements present in S_0 . From the perspective of language acquisition, we are initially concerned with the following categories (2a)-(2c):

- (2) a. unexplained elements of S_0
- b. interface conditions (the principled part of S_0)
- c. principles of efficient computation

Chomsky (2004) defines the SMT as the claim that there are no unexplained elements of S_0 : (2a) is empty. Although Merge is the linguistically proprietary operation, it is an explained element rather than an unexplained one, according to Chomsky: it "comes for free" simply because it is the simplest possible operation that accounts for the recursion in human language. Case, agreement, binding theory, the deletion of copies, and all other operations and theories taken to be a part of syntax all should be reduced to either (2b) or (2c), according to Chomsky.

For instance, Chomsky's own theory of the operation Agree holds that a probe P deletes its uninterpretable features by valuing them with the interpretable features of goal G. This seems to be an operation in syntax proper. One's natural inclination is to suppose that, like Merge, it is a part of UG. Why should uninterpretable features, and Agree, exist at all? Chomsky proposes that these are in fact part of the optimal mechanism in order to account for displacement phenomena in syntax, and so can be reduced to (2b)-(2c). The reduction of other phenomena such as control

and binding, among other things, is also supposed to proceed along these lines. As an anonymous reviewer points out, one might take any aspect of language description to be a serious challenge for the SMT, given that it is not very worked out.

At present, however, Chomsky and other Minimalists grant that they are not able to offer formal definitions of computational efficiency. Progovac (2019) argues that there are issues with this line of reasoning. For instance, according to B&C, it is both the case that syntax has to be simple in order to be evolvable and given that syntax must have evolved, it must have been simple, which appears to be a circular line of reasoning. Progovac points out that this makes the SMT unfalsifiable and impossible to analyze from a scientific perspective.

However, for the purposes of this paper, I will put aside issues with the falsifiability of the SMT. Instead, my goal here is to focus on language-specific puzzles that might arise for the SMT, which as far as I am aware is thus far novel in the literature. My hope is that this will lead to further fruitful discussions on the role of the SMT in syntactic theory.

2.3 Puzzles for the SMT

Thus far, it seems there has been little discussion in the literature regarding what the hypothetical truth of the SMT would entail. The first entailment is that the SMT commits its believers to Borer's (1984) conjecture, which is defined as follows:

- (3) The Borer Conjecture
All parameters of variation are attributable to differences in the features of particular items in the lexicon.

Here is why. There is no doubt that there is crosslinguistic syntactic variation. If it truly is the case that there is nothing more to syntax proper than Merge, then it alone cannot account for the vast amount of variation that is attested. Nor is it possible to say that different languages have different principles of efficient computation. Therefore, all variation can only be as a result of the different features present in items of the lexicon. As such, Chomsky (1995) incorporates a more specific version of this conjecture into his Minimalist Program, which has been referred to as the Borer-Chomsky Conjecture: given that there are no syntactically proprietary elements apart from Merge, all variation can only be due to the presence of features visible in the two interfaces, namely the conceptual-intentional and sensorimotor systems.

Omer Preminger points out a curious consequence of this entailment that is worth mentioning: it appears to commit the Minimalist to (a non-trivial version of) the Sapir-Whorf hypothesis.⁴ Preminger uses an example of syntactic variation between English and Kaqchikel to illustrate his point: in Kaqchikel, the subject of a transitive clause cannot be targeted for focalization, relativization or *wh*-interrogation, whereas it can in English, as in (4). (4) is ungrammatical in Kaqchikel.

- (4) It was the cat who licked the child.

Recall that all variation ultimately must arise from the two interfaces. Given that speakers of Kaqchikel are able to *pronounce* sentences like (4) before immediately commenting that they are wrong, it appears that the variation does not arise from the sensorimotor interface.

⁴The reader is referred to Preminger (2020) for the blogpost in which Preminger makes this argument.

Of course, as an anonymous reviewer rightly notes, it doesn't follow from the fact that one can say such sentences doesn't mean that the ungrammaticality of them is not explainable at the sensorimotor interface. But, as Preminger mentions, if we do take the SMT for granted, then we have to follow Chomsky in assuming that there is nothing that is linguistically proprietary apart from Merge. In that case, the sensorimotor interface cannot reference any language-specific difference between English and Kaqchikel, because doing so would necessarily imply the existence of something linguistically proprietary that is not Merge.⁵

In addition, Preminger notes that the Minimalist cannot assume that there is a system between syntax and the phonological interface, given that the existence of this system would itself violate the SMT. This leaves the conceptual-intentional interface, which involves the system of thought in the human brain, and is not LF, as Chomsky has stressed.

We are left with two options: either (i) the conceptual-intentional systems of English and Kaqchikel speakers differ, or (ii) sentences like (4) express different interface properties in English compared to Kaqchikel. Though (i) is plainly the Sapir-Whorf hypothesis, (ii) entails it as well, because it implies that English speakers are able to construct sentences with different conceptual-intentional content. Thus, we arrive at the inevitable conclusion that the English and Kaqchikel speakers have different thought processes. One might take being forced to adopt the Sapir-Whorf hypothesis to be an undesirable state of affairs.

Preminger mentions a second and more important entailment, independent of the Borer-Chomsky Conjecture. The SMT also commits one to the following conclusion: if there is any cause for Merge apply or not apply, and this cause is not explainable by reference to the interface conditions (2b) or principles of efficient computation (2c), then it must also be a linguistically proprietary entity, which would violate (2a), given that it would be an unexplained element of S_0 . This will end up forming the basis of the objection based on cartography in section 4.

Preminger lays the foundation for the two independent problems for the SMT I will present. Both of my arguments will involve crosslinguistic patterns which are not amenable to explanation via (2b) or (2c). First, not all syntactic variation can be accounted for by the Borer Conjecture, at least not without positing the existence of unexplained elements. Second, I argue that Merge can apply in a very specific, cartographic order, further indicating the existence of more unexplained elements. I will conclude that it is exceedingly unlikely (2a) to be empty.⁶ If at least one of these arguments is right, then it would raise a serious problem for the SMT.

Before moving on, I want to point out one potential point of contention. As just noted, Minimalists grant that the notions in (2b) and (2c) lack a formal and falsifiable definition. Therefore, how can we know for certain that the upcoming empirical cases I will present cannot be accounted for in terms of interface conditions and/or principles of computational efficiency, without knowing their precise definitions?

This does nothing more than to highlight the problem—the SMT is not falsifiable. At this point, when presented with empirical paradigms, our best option is to take Chomsky and other Minimalists at their word, and rule out explanations involving processes of computation and interface conditions via the process of elimination. Changing the definition of "interface conditions" or

⁵As an anonymous reviewer notes, the very fact that phonological variation exists might be taken to refute the SMT.

⁶Preminger also comes to the same conclusion in Preminger (2011, 2014) regarding the nature of agreement. Preminger argues that agreement is a syntactic phenomenon, and not merely postsyntactic, and it cannot be reduced to interface conditions. I refer the reader to these works for further detail.

"principles of computational efficiency" when presented with novel empirical data has the risk of rendering them trivial and *ad hoc*.

3 The Borer-Chomsky Conjecture is not sufficient

The presence of macroparameters for syntactic variation—whether they exist as a general rule of grammar in the sense of Mark Baker, or as aggregates of microparameters—might raise a problem for the SMT. For my purposes, it is in fact irrelevant whether macroparameters exist independent of microparameters. But first, I will provide some background into the literature on parameters before presenting Baker's (2008a; 2008b) evidence that the Borer-Chomsky Conjecture alone is not sufficient to account for syntactic variation.

But before doing so, let us briefly return to the P&P framework. The P&P approach bridges a gap between the Greenbergian and Chomskyan approaches to language universals mentioned in 2.1. The parameters of UG might in fact be the basis of the variation observed in the Greenbergian tradition, constrained by the principles of UG. The first parameter that was proposed was the Null Subject Parameter by Chomsky (1981) and Rizzi (1982). This parameter was more general in its scope, in that it was meant to account for a cluster of seemingly interrelated properties, such as the overtiness of subjects, subject inversion and complementizer-trace effects. It was proposed to explain differences between the major Romance languages French, Spanish and Italian.

It did not arise as a result of comparing different language families with each other, which would have been broader in scope. Nor did it arise from comparing different dialects of the same language with each other, which would have been narrower in scope.⁷ Indeed, Gilligan (1987) has shown the Null Subject Parameter falls apart if we examine it in either a broader or narrower scope, not matching Chomsky or Rizzi's predictions.

As a result, parameters, at least in the classical sense, have gone out of fashion in recent years: authors such as Culicover (1999), Gallego (2011), Newmeyer (2004, 2005), Boeckx (2011), Richards (2008) among others have rejected the existence of macroparameters like the Null Subject Principle entirely. *Microparameters*, however, have flourished since Kayne (2005). Instead of looking at a cluster of properties, microparameters involve very localized and small differences in the grammars of closely related languages. Works such as Boeckx (2011) claim that macroparameters for syntactic variation can in fact be reduced to aggregates of such microparameters, and one need not posit macroparameters as general rules of grammar after all. This has allowed for the Borer-Chomsky Conjecture to take root in Minimalist syntax.

We are now ready to look at Baker's (2008b) survey of 108 languages regarding the relationship between case and agreement, which provides evidence in favor of the classical parameter theory of syntactic variation.⁸ Here is what the argument is going to look like. Baker's survey indicates that languages can be grouped up into parametric clusters—four to be precise, just as Baker's macroparametric theory would predict. On the other hand, the microparametric theory would predict the distribution of languages with regard to these properties to be in a relatively smooth continuum. This prediction is not borne out, and Baker's theory seems to be on the right

⁷This is one of the reasons why Baker (2008a) calls this the Null Subject Parameter a medioparameter rather than a macroparameter.

⁸100 of these languages are from the core 100-language sample of the *World Atlas of Language Structures* (WALS) by Haspelmath et al. (2005).

track here. Hence, the Borer-Chomsky Conjecture seems to miss what is really going on in syntactic variation.

Baker's macroparameters are stated as follows in (5)-(6):⁹

- (5) *The Direction of Agreement Parameter*
F agrees with DP/NP only if DP/NP asymmetrically c-commands F.
- (6) *The Case-Dependency of Agreement Parameter*
F agrees with DP/NP only if F values the case feature of DP/NP or vice versa.

These parameters emerged as a result of a detailed comparison between the Niger-Congo (NC) and Indo-European (IE) languages. (5) would be valued Yes in the NC languages, but not IE, while the opposite would be the case for (6). The crucial idea behind these parameters is that NC agreement obeys certain structural configurations that IE does not. Namely, while the IE languages only care about Case matching in regard to agreement, in NC the agreeing head must be in a position lower in the structure than the NP that it agrees with.

Let us see an example of this in action with subject agreement, or in other words, agreement on the finite T head of a sentence.¹⁰ Although in simple clauses both IE and NC languages show agreement between the preverbal subject and the finite verb, we can tease apart the differences in contexts in which something apart from the thematic subject occupies the canonical Spec,TP subject position. This is because in IE, the verb must agree with the nominative NP, regardless of its position, whereas in the NC languages its structure does matter. For instance, in the Bantu language Kinande the finite verb must agree with the fronted object (7a), whereas in Yiddish it must agree with the postverbal subject (7b).

- (7) a. **Olukwi** si-lu-li-seny-a bakali (omo-mbasa)
wood.11 neg-11S-PRES-chop-FV women.2 LOC.18-axe.9
'WOMEN do not chop wood (with an axe).' Kinande
- b. ...az vayn **ken men** makhn fun troybn oykh
that wine can one make from grapes also
'(You should know)... that one can make wine from grapes also.' Yiddish

Here is another example involving locative inversion. In the Kinande example (8) we see agreement on the finite verb with the fronted locative, whereas in English it agrees with the thematic object instead, due to its nominative case:

- (8) **Oko-mesa kw-a-hir-aw-a** ehilanga.
LOC.17-table 17S-T-put-PASS-FV peanuts.19
'On the table were put peanuts.' Kinande

Let us now consider examples in which nothing is moved into Spec,TP. In other words, it is either left empty, or filled with a null expletive. English requires agreement with the postverbal subject in this context, as in (9a). In the Kinande example (9b), the subject agreement slot is filled with a locative prefix, which can either be analyzed as a null expletive *there* or default agreement. The finite verb does not agree with the post-verbal subject.

⁹(5) is simplified. The reader is referred to Baker (2008b) (p. 215) for the complete version.

¹⁰The reader is referred to Baker (2008b) for a much fuller discussion of the differences between the two kinds of languages.

	Agree must be up ((5)) = Yes	Agree can be up or down ((5)) = No
Agreement dependent on case ((6) = Yes)	Turkish, Lango, Greenlandic, Apurina, Chamorro, Mapudungun (n = 6)	IE languages (7), Hausa, Finnish, Abkhaz-Abaza, Kannada, Asmat, Amele, Alamlak, Maung, Mangarrayi, Tiwi, Lavukaleve, Daga, Yimas, Lakhota, Tzotzil, Warao, Barasano, Yagua, Wichi, Choctaw, Hixkaryana, Hebrew, Wari, Chukchi, Makah (n = 32)
Agreement independent of case ((6) = No)	Zulu, Swahili, Kinande, Berber, Arapesh, Tariana, Fijian, Tukang Besi, Slave, Canela-Kraho, Jarawara (n = 11)	Georgian, Arabic, Persian, Warlpiri, Dani, Kewa, Burushaski, Mayali, Halkomelem, Tauya, Ojibwa, Nez Perce, Karok, Otomi, Zoque, Ika, Basque, I. Quechua, Guarani (n = 19)

- (9) a. There arrive three new delegates each day.
b. Mo-**ha**-teta-sat-a mukali (omo-soko).
AFF-there-NEG/PAST-dance-FV woman.1 LOC.18-market
‘No woman danced in the market.’

Kinande

We are now ready to see Baker’s survey. Table 1 provides Baker’s (2008b) survey (p. 221). 68 out of the 108 languages can be categorized into one of the four clusters are predicted to exist by the parameters in (5)-(6). 40 of these languages were unclassified; 29 of these simply have no agreement, whereas the other languages are indeterminate, for which I refer the reader to Baker (2008b) for more details. What is crucial here is that languages are grouped into clusters, rather than a smooth continuum, contrary to what the microparameter approach would predict.

Ledgeway & Roberts (2017) provide Baker’s argument in a diachronic version. They claim that the Modern Romance languages and Latin differ from each other in terms of two (potential) macroparameters: first, a free vs. fixed word order parameter (best described as the availability or lack of scrambling in a language, especially the extraction of adnominal modifiers); second, the head parameter, concerning whether a language is head-final or head-initial. Though both Latin and the modern Romance languages appear to not have polysynthesis and show the same accusative alignment, the modern Romance languages are head-initial with a mostly fixed word order, whereas Latin had many head-final properties with a freer word order. These might be taken to be differences in macroparameter settings between the two.

They note that in addition to macroparametric change, there is a great deal of microvariation between the Modern Romance languages. For instance, Latin did not have clitics at all, while French has the full paradigm of subject clitics, in that there is a clitic for each person and number combination. Italian, on the other hand, is subject to dialectal variation, but no variant of Northern Italian shows the full range of subject clitics like in French. Similarly to Baker, Ledgeway & Roberts note that a purely microparametric view would expect the variation to be greater and not as constrained. Some Romance varieties ought to display unattested clusters of properties which mix features of Latin and Modern Romance syntax. But this is not observed.

They give examples of imaginable Romance varieties in (10a)-(10c) below, each of which have the following property. In (10a), head-final word order and full configurationality is mixed with synthetic passives and articles. In (10b), head-initial word order and non-configurationality is combined with accusative with the infinitive (AcI) and auxiliaries. Finally, in (10c), case, null objects, determiners and discontinuous negation are combined with head-initial word order and full configurationality. None of these are attested in the Romance languages at any period.

- (10) a. Ari nova du pizzaiolu a pizza 'nfornaður.
 at.the nine of.the pizzaiolo the pizza place.in.oven.3SG.PRES.PASS
 'At nine o'clock the pizza is placed in an oven by the pizzaiolo.' *Latalabrese
- b. Trop savons belle la femme avoir dansé
 too.much know.1PL.PRES pretty.F.SG the.F.SG woman have.INF dance.PPT
 'We know that the beautiful lady has been dancing too much.' *Latinias
- c. María visitó pueblom pero yo no conozco paso
 María visited village.ACC but I NEG know step.NEG
 'María has visited the village but I don't know it.' *Latiñol

They thus claim to find a bimodal distribution of macro- and microparameters. Though the Romance languages tend towards being head-initial, non-configurational, accusative and non-polysynthetic, they still have a great deal of microparametric variation. The best explanation of this state of affairs is that the modern Romance languages have had a modest amount of macroparametric change together with a great deal of microparametric variation.

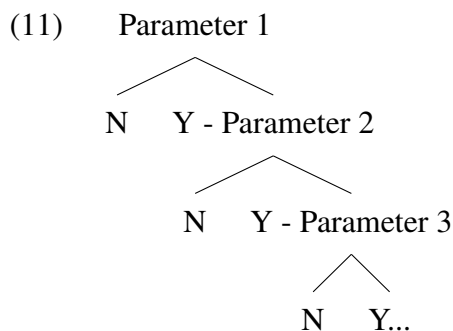
Given the concrete existence of languages clustering towards certain properties, there are now two paths we can take. We can either accept that macroparameters do exist as a general rule of syntax, as Baker does. This would immediately falsify the Borer-Chomsky Conjecture, and provide a potential refutation of the SMT. This is because macroparameters would end up being an unexplained element in the faculty of language, that cannot be reduced to interface conditions or principles of efficiency computation. The problem with this line of reasoning, however, is that Minimalists today assume that macroparameters can in fact be reduced to sets of microparameters. This is able to avoid the objection I have just mentioned while maintaining the Conjecture.

Examples of such approaches are Kayne (2005), Boeckx (2011) and Roberts (2019) among others, all of which claim that macroparameters do exist, but not as a general rule of grammar: they are in fact aggregates of microparameters, so they are in fact one and the same. For this reason, Boeckx goes a step further, claiming that Minimalist theory would benefit from abandoning P&P theory altogether. Though Boeckx acknowledges that parametric variation may exist in the very weak sense, only involving microparameters, he claims that it is not possible for a single parameter to have complex effects across many different phenomena.

But even if macroparameters are reducible, Boeckx doesn't address Baker's argument—it doesn't change the fact that languages cluster around certain *sets* of microparameter values. This seems to be an incontrovertible empirical truth. In other words, even if the Borer-Chomsky Conjecture is correct, it tells us nothing about *why* certain features in the lexicon cluster. There has to be a reason why certain clusters exist, but not others. And this might arise due to unexplained elements in UG.

One possible explanation of how features cluster is as follows. Biberauer et al. (2014), Sheehan (2014) and Roberts (2019) among others all propose a more involved definition of macroparameters in terms of a parameter hierarchy. Though these accounts differ in technical details, the

basic idea behind them is that parameters are dependent, in the sense that there are implicational relationships between parameter settings. Macroparameters again end up being aggregates of microparameters, but they are microparameters that act together on the basis of learning strategies that derive from conservative computational principles.¹¹ In the tree (11) below, we see that a positive setting for Parameter 3 depends on the setting of 2, and the same holds between 2 and 1:



A key advantage of this model is that it drastically reduces the space of variation, while still maintaining descriptive adequacy.¹²

But one has to ask where these hierarchies come from. As Sheehan (2014) notes, it seems implausible that "such a rich system of parameter hierarchies" is a part of the innate endowment, due to Darwin's problem. For example, even if one assumes that Case, EPP and ϕ -features are all a part of UG, it is unclear what regulates the ordering between the parameters—it seems difficult to derive *all* of the parametric hierarchy via interface conditions and principles of efficiency.

Regardless, the fact that apparent macroparametric variation exists still has to be accounted for. Following Holmberg & Roberts (2014) and Roberts (2019), I would like to now propose a specific conception of UG in order to leave room for parametric variation which, in my view, contradicts the SMT. Holmberg and Roberts, contra Boeckx, take UG to create space for parametric variation by leaving certain choices underspecified. These choices involve formal features on certain heads like T and morphological realizations. For instance, the Null Subject Parameter (if there is one) might be taken to be the presence or lack of a D-feature on finite T. Therefore, UG merely specifies whether T may or may not have a D-feature. The job of the language acquirer is to "fill in the gap" via conservative learning strategies.

I take this to be a "weaker" minimalist conception of UG that is necessary in order to capture the empirical landscape, and it is capable of accounting for the macroparametric patterns of variation we have seen in this section. To conclude, it seems that the only way the existence of such microparametric clusters could be explained is by supposing that there are unexplained elements in universal grammar. I take such elements to be, at the very least, a UG with underspecified parameters in the sense of Holmberg & Roberts (2014) and Roberts (2019).

¹¹Computational conservatism is the idea that cognitive computations are costly and there is a general pressure to lessen the cost as much as possible.

¹²I refer the reader to Biberauer et al. (2014), Sheehan (2014) and Roberts (2019) for further details, which would go out of our scope.

4 A syntactic ordering restriction on Merge

We have just seen evidence, in my view, that not all crosslinguistic variation can be accounted for by assuming that there are no unexplained elements in S_0 . I would like to present more evidence in favor of there being unexplained elements, that is independent of the existence of macroparameters of variation. The argument comes from the cartographic enterprise in modern syntax; in particular, I will argue that there are at least some purely syntactic, ordering-based restrictions on Merge that cannot be reduced to interface conditions or principles of efficient computation. I will argue that at least some of Rizzi's (1997) cartographic blueprint of the C domain is purely syntactic, and is therefore an unexplained element as in (2a).

The goal of cartography in modern generative syntax is to draw highly detailed maps of syntactic structure. For Cinque & Rizzi (2009), under this conception, cartography ought to be seen as more of a research topic rather than a theory or hypothesis that attempts to determine the right map for syntactic structure.¹³ There is disagreement as to what the right order of projections is in either Cinque or Rizzi's cartographic frameworks, but Cinque & Rizzi still think that this doesn't alter the fact that cartography is a relevant question for modern syntactic theory. Let's start by looking at Cinque (1999), although I will ultimately conclude that Rizzi's cartographic split C-domain raises the main problem for the SMT.

Cinque seeks to account for a crosslinguistic pattern regarding the ordering of adverbs that can appear in a sentence. If there are multiple adverbs in a sentence, for the most part, they have to obey the ordering in (12).

- (12) frankly > fortunately > allegedly > probably > once/then > perhaps > wisely > usually > already > no longer > always > completely > well

An example of this can be seen in English. Below, we have a sentence with two adverbs: *any longer* and *always*, and both appear before the verb. What we find is that the adverb *any longer* must precede the adverb *always*:¹⁴

- (13) a. John doesn't any longer always win his games.
 b. * John doesn't always any longer win his games.

Cinque tests Norwegian, Bosnian/Serbo-Croatian, Hebrew, Chinese, Albanian and Malagasy in addition to Italian and English. He finds that the ordering in (12) is maintained in each language. Of course, for such fine ordering to be attested in all of these languages would be a remarkable coincidence—it appears that there are some general cognitive constraints from which these patterns derive. Cinque argues in favor of the existence of many and finely ordered functional projections within each clause, into which adverbs can be inserted.¹⁵

¹³Cinque & Rizzi (2009) note that there prima facie may be tension between Minimalism and cartography, if cartographic blueprints truly are innate. This appears to contradict the SMT. But they think that there is no inherent conflict between the two viewpoints: because Minimalism studies the mechanism by which syntactic structure is created—via Merge—whereas cartographers study the ordering in the maps that are created. I will argue that cartography still raises a problem for the SMT, however, in this section.

¹⁴There is one little catch with this data. Notice that the sentence *John doesn't always win his games any longer* is acceptable, in which *always* appears to precede *any longer*. This is also possible in Italian, according to Cinque, but only if *any longer* is emphasized. Without emphasis, it is not possible. As Cinque notes, appearances are deceiving: one could suppose that it involves movement of the adverb from its initial position.

¹⁵The first to argue in favor of this was Alexiadou (1997).

Now, if these functional projections truly are innate, then they would be an unexplained element in UG, contradicting the SMT. But this is exceedingly unlikely: Chomsky et al. (2019) notes that taking this theory at face value would be unable to minimally meet the conditions of evolvability and acquirability.¹⁶ How could such fine ordering between adverbs like *any longer* and *always* have evolved? People even rarely use them in the same sentence. Based on such concerns, linguists such as Ernst (2002) have provided purely semantic explanations of Cinque’s hierarchy. This can be fortuitously used by adherents of the SMT to account for (12) in terms of interface conditions. In other words, the ordering in (12) could be explained via semantic or pragmatic reasons that are independent of syntax. I concur, and this does not make a strong argument against the SMT.

But let us move onto the cartography of the C domain. Rizzi (1997) provides empirical evidence for two different kinds of complementizers. In Italian, for example, we see in (14) below that it is impossible to place topics in a position to the left of the high complementizer *che* (which Rizzi calls a finite complementizer), but it is possible to place topics to its right.

- (14) a. Credo che, il tuo libro, loro lo apprezzerebbero molto.
 I.think that[+fin] the your book them it will.appreciate much
 ‘I think that they will appreciate your book very much.’
 b. *Credo, il tuo libro, che loro lo apprezzerebbero molto. Italian

This contrasts with the behavior of the low complementizer *di* (which Rizzi calls a nonfinite complementizer), which only allows one to place topics to its left in (15).

- (15) a. Credo, il tuo libro, di apprezzar-lo molto.
 I.think the your book that[-fin] appreciate-it much
 ‘I think that they will appreciate your book very much.’
 b. *Credo di, il tuo libro, apprezzar-lo molto. Italian

This contrast is visible in English as well, to an extent. *That* is a high complementizer, given it doesn’t allow topics to precede it:

- (16) a. I think that Aspects, Chomsky wrote *t*.
 b. *I think Aspects, that Chomsky wrote *t*.

How do we account for this contrast? If we had just one C projection—CP, as is commonly assumed—it would be impossible for a single projection to be responsible for the contrast between high and low complementizers. Along with other pieces of evidence, Rizzi splits up the C domain as follows.¹⁷ Relevant for the present is that *che* is located in ForceP, because it necessarily precedes all focalized elements and topics. This makes it a high complementizer. *Di* is located in FinP, because it necessarily follows them, making it a low complementizer:

- (17) ForceP > TopicP* > FocusP > TopicP* > FinP

I would like to point out that there is considerable crosslinguistic evidence that the distinction between high vs. low complementizers is not unique to Italian. It is widely attested crosslinguistically, in other Romance languages (such as Spanish, according to Villa-Garcia (2012), in the

¹⁶See also Bobaljik (1999) for problems for Cinque’s hierarchy.

¹⁷The * indicates that TopicP is recursive, in that it can appear before or after any other functional projection between ForceP and FinP, but not before ForceP, or after FinP.

Scandinavian languages (Larsson (2017)), in the Niger-Congo language Lubukusu (Carstens & Diercks (2009)) and even in English (Haegeman (2012)). Regardless of whether one buys the generative enterprise, there really does appear to be two kinds of complementizers—one which necessarily precedes topics, and one which necessarily follows them.

The problem for the SMT is, in fact, exceedingly simple. Thanks to Preminger’s observation, we noted previously that any kind of syntactic functional structure that is innate would contradict the SMT, because it would be an unexplained element in UG. I granted, for the sake of argument, that all of Cinque’s hierarchy could be reduced to semantic/pragmatic explanations, following Ernst (2002). I will even grant that much of Rizzi’s hierarchy could be reduced to semantic/pragmatic explanations. Rizzi (2013) provides a possible explanation of the crosslinguistic asymmetry between the ordering of topic—which can be reiterated in many languages—and left-peripheral focus, which cannot. But not all of Rizzi’s hierarchy can be reduced in such a manner.

Recall Preminger’s point that any restriction on Merge would itself be an unexplained element in UG. Why must *that* or *che* in Italian be Merged *after* all topics? Why must *di* in Italian be Merged before all topics in the left periphery? A complementizer by definition simply marks a clause as the subject or object of a sentence. It has no semantic purpose: it is often treated as semantically vacuous, for instance by the seminal Heim & Kratzer (1998).

To see this in more detail, it would be useful to look at a language like Russian, in which the phonetic form of a high complementizer can vary depending on the meaning of the clause that it marks. The high complementizer *čto* is used to mark indicative embedded clauses, just like *that* in English, while *čtoby* is used to mark subjunctive embedded clauses. An example of *čtoby* in use is given below:

- (18) Ivan xočet čtoby Maša pročitala/čitala [Vojnu i Mir] Antonenko (2008), Russian
 Ivan wants that.SUBJ Maša read.PST.PERF/.PST.IMPERF War and Peace
 ‘Ivan wants for Masha to read War and Peace.’

Like *čto*, *čtoby* is a high complementizer. Topics may not precede it:

- (19) * Ivan xočet [Vojnu i Mir] čtoby Maša pročitala/čitala *t* Russian
 Ivan wants War and Peace that.SUBJ Maša read.PST.PERF/.PST.IMPERF
 (Intended reading) ‘Ivan wants for Masha to read War and Peace.’

Herein lies the problem. The fact that the same position (ForceP under Rizzi’s terminology) can be taken up by both a subjunctive and an indicative complementizer in Russian indicates that the position is not semantically derivable. Indeed, in languages like English, the low complementizer *for* has an irrealis semantics.¹⁸ As Pesetsky (2021) points out, a *for*-infinitive can have an irrealis or generic use, but not a factual one:

- (20) a. For it to rain would be helpful. irrealis
 b. For it to rain is always helpful. generic
 c. For it to rain was helpful last night. factual

Subjunctive clauses are in the irrealis mood. In Russian, the irrealis complementizer is only permitted in ForceP, but in English, it is only permitted in FinP. This is further evidence that such a contrast is not semantically derivable.

¹⁸I follow Adger (2007) in assuming that *for* is a low complementizer in FinP because it does not allow topics to its left.

Might we instead look at the sensorimotor interface? Complementizers do not appear to have any phonological function: to borrow Preminger’s reasoning from section 2, speakers are perfectly capable of pronouncing sentences like (16b), before immediately commenting that they are wrong. And there does not seem to be reason to think that such complementizers play a role in principles of efficient computation. The fact that there are restrictions on *where* different complementizers must appear seems to contradict the SMT, given that they must be strictly ordered. At the very least, there is some purely syntactic truth to Rizzi’s cartographical blueprint.

Satik (2022) proposes a further problem driving this point forward. Note first the seemingly trivial observation that the English infinitive doesn’t allow a complementizer like *that*, though *for* is a possibility, which is often treated as a nonfinite complementizer in English.

- (21) a. I seem (*that) to be happy.
b. Mary is eager (for/*that) John to please.

Satik (2022) provides a survey of 26 languages belonging to several different language families. He notes that although it is common for infinitives to allow complementizers crosslinguistically, just like English, what all of these complementizers have in common is that they are lower in structure—in other words, located in Rizzi’s FinP. It is impossible for a high complementizer to be located within an infinitive, in any kind of infinitive in any language of the survey. This leads him to define finiteness in terms of clause size—in other words, the truncation of ForceP. Given that high complementizers are blocked from being Merged with an infinitive, this is further evidence in favor of ordering restrictions on Merge,

All of the observations thus far have been theory-independent, in that they are purely empirical observations. *Why* should any of this be the case? As Preminger notes, anything that prevents Merge from applying is linguistically proprietary. This seems to imply the presence of unexplained elements in UG that force the position in which topics and focalized elements can be Merged with respect to complementizers. To conclude, there appear to be at least a small set of cartographic generalizations which are not amenable to a Chomsky-style reduction. There may indeed be an innate blueprint that is a part of UG.

5 Resurrecting Darwin’s problem

In the two preceding sections, I have argued that there are in fact unexplained elements within the language faculty—purely syntactic properties that cannot be reduced to interface conditions or explained in terms of principles of efficiency. My solutions to the two presented problems have been different. For the first, I have argued in favor of a return to a watered-down and more Minimalist version of the P&P framework. For the second, I have proposed that it is possible for UG to carry a rudimentary blueprint of the C domain, specifying the position of high and low complementizers. If I am right, then I must confess that we are facing a fiendishly difficult problem. How could such linguistically proprietary elements have evolved, in addition to Merge? It is unlikely that all of these syntactic constraints evolved saltationally—that is, a large and sudden mutational change from one generation to the next. We need more innate building blocks, as Haspelmath (2020) would suggest.

In order to maintain our hypothesis that language is innate—UG does exist—it is necessary to consider an alternative to BC’s saltationist approach to language evolution. Many researchers

have proposed gradualist accounts of language evolution, even in syntax.¹⁹ To see how syntactic constraints on deriving movement constraints in the generative grammar framework, let us consider how Progovac (2009) derives islandhood under a gradualist account of language evolution.

The seminal dissertation Ross (1967) notes the existence of *islands*: these are defined as syntactic environments which do not allow movement out of them. Note that there is a clear difference in acceptability between (22a)-(22b) below. A classical example of an island is the coordination structure in (22b):

- (22) a. What did Mary eat ham with <what>?
b. * What did Mary eat ham and <what>?

The existence of islands is puzzling from an evolutionary perspective. How could constraints on movement have led to "fruitful sex," in the words of Lightfoot (1991)? Why would a grammar with island constraints be selected over a grammar without islands? Of course, concerns such as these were the original kind of justification for B&C's saltationist approach.

Progovac (2009) suggests that islandhood constraints could have evolved gradually. Taking movement itself as an exceptional operation, she argues that islandhood is in fact the default state of syntax. Progovac makes the observation that movement itself is only available out of a subset of complements, forming a natural class. But the set of islands do not form one, because islands are things like conjuncts and adjuncts, among other things. According to Progovac, movement evolved from a proto-syntax that only had small clauses and one-word utterances. Subordination and movement evolved due to the need to embed multiple viewpoints within each other. Adjunction and coordination were not sufficient enough for this purpose, as the example in (23c) illustrates. Only (23c) allows a person's knowledge about someone else to be reported:

- (23) a. [As you know], [as Mary knows], he is a linguist.
b. He is a linguist, [and you know it,] [and Mary knows it].
c. You know [that Mary knows [that he is a linguist]].

We now have a gradualist account of islandhood—the need to be able to embed multiple viewpoints does seem important, after all. But can such a gradualist account of islandhood be extended to macroparameters, and certain parts of the cartographic framework? This seems to be difficult, to say the least. It is difficult to see how the evolution of a macroparameter for agreement and case, for instance, could have ever led to fruitful sex. For the purposes of my paper, I have only sought to merely show that there are unexplained elements of the language faculty—I am leaving open how these unexplained elements came into being for future research.

But here is my tentative attempt at explaining how at least Baker's macroparameters on case and agreement could have evolved. Suppose that agreement and case evolved as a result of gradual evolution—perhaps as the result of a "feedback loop" between adaptive cultural and biological changes. Perhaps this was driven by the cultural need for speakers to disambiguate their utterances by marking dependent nouns and verbs with the relationship that they bear to each other.

¹⁹Apart from Progovac, some examples are Givón (1979, 2002, 2009), Pinker & Bloom (1990), Newmeyer (2005), Jackendoff (1999, 2002), Culicover & Jackendoff (2005), Tallerman (2014), Heine & Kuteva (2007), Hurford (2007, 2012), Gil (2017) and Progovac (2009, 2016, 2019) among others. For a helpful survey of the field of language evolution, the reader is referred to Progovac (2019).

One possibility is that there is some truth to Chomsky's notion of computational efficiency, in that principles of computational mechanism require for languages to have either upward or downward agreement, and for case to be dependent on agreement or not. UG, in the "weak minimalist" sense we have seen in section 3, is underspecified for one of these options. It is up to the learner to determine whether the language has upward or downward agreement and specify this parameter in their language faculty. This leaves open many questions, of course, but my goal here has been to provide a sketch of how such an approach could play out.

6 Concluding Remarks

The fundamental goal of this paper has been to present specific puzzles Strong Minimalist Thesis from linguistic evidence. These puzzles imply that Merge is not the only linguistically proprietary element present in the language faculty of humans. My strategy has been to first present the two consequences assuming the SMT would have. First, that all syntactic variation would be due to the Borer-Chomsky Conjecture, and second, that there can be no purely syntactic cause that makes Merge apply in certain orders. I have argued that both of these consequences lead to independent contradictions, raising problems for the SMT. Linguists interested in the origins of Universal Grammar might perhaps consider assuming a gradualist approach to its evolution.

Admittedly, this does end up opening more problems than it solves. How could unexplained elements of the language faculty, in Chomsky's words, have evolved, in addition to the structure generating and recursive operation Merge? At the very least, my hope is to have helped to rule out a specific saltationist account of language evolution by Berwick & Chomsky (2016). This is a vexing problem, and it will likely require assuming that language evolved gradually in multiple evolutionary steps. Indeed, it would not be wise to discard the philosophy driving the Minimalist Program in its entirety, given that it is driven by reasonable evolutionary concerns. I believe that a "Weak Minimalist Thesis" is the right way forward for syntactic theory.

References

- Adger, D. (2007). Three domains of finiteness: a minimalist perspective. In I. Nikolaeva (Ed.), *Finiteness: empirical and theoretical foundations*. Oxford: Oxford University Press.
- Alexiadou, A. (1997). *Adverb placement: A case study in antisymmetric syntax*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Antonenko, A. (2008). Puzzles of Russian Subjunctives. In *UPenn Working Papers in Linguistics* (Vol. 14.2). Philadelphia: Penn Graduate Linguistics Society.
- Baker, M. C. (2002). *The atoms of language*. Oxford: Oxford University Press.
- Baker, M. C. (2008a). The macroparameter in a microparametric world. In T. Biberauer (Ed.), *The limits of syntactic variation* (pp. 351–373). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Baker, M. C. (2008b). *The Syntax of Agreement and Concord*. Cambridge: Cambridge University Press.
- Berwick, R., & Chomsky, N. (2016). *Why only us: language and evolution*. MIT Press.

- Biberauer, T., Holmberg, A., & Roberts, I. (2014). A syntactic universal and its consequences. *Linguistic Inquiry*, 45, 169–225.
- Bobaljik, J. (1999). Adverbs: The hierarchy paradox. *Glott International*, 4.9/10, 27-28.
- Boeckx, C. (2011). Approaching parameters from below. In *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*. Oxford: Oxford University Press.
- Borer, H. (1984). *Parametric syntax: case studies in Semitic and Romance languages*. Dordrecht, Holland: Foris Publications.
- Carstens, V., & Diercks, M. (2009). Parameterizing case and activity: Hyper-raising in Bantu. In *Proceedings of North Eastern Linguistic Society 40*. Amherst, MA: GLSA Publications.
- Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, The Netherlands: Foris Publications.
- Chomsky, N. (1986). *Knowledge of language*. New York, New York: Praeger Publishers.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Massachusetts: MIT Press.
- Chomsky, N. (2000a). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–156). Cambridge, MA: MIT Press.
- Chomsky, N. (2000b). *New Horizons in the Study of Language and Minds*. Cambridge: Cambridge University Press.
- Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures* (Vol. 3, pp. 104–131). Oxford: Oxford University Press.
- Chomsky, N. (2020). *The UCLA Lectures*. Retrieved from <https://ling.auf.net/lingbuzz/005485>
- Chomsky, N., Gallego, Á. J., & Ott, D. (2019). Generative grammar and the faculty of language: Insights, questions, and challenges. *Catalan Journal of Linguistics*, 229.
- Cinque, G. (1999). *Adverbs and functional heads: a cross-linguistic perspective*. New York: Oxford University Press.
- Cinque, G., & Rizzi, L. (2009). The Cartography of Syntactic Structures. *Manuscript*.
- Culicover, P. W. (1999). *Syntactic nuts: hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press.
- Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford: Oxford University Press.
- Ernst, T. (2002). *The syntax of adverbs*. Cambridge: Cambridge University Press.
- Gallego, A. J. (2011). Parameters. In C. Boeckx (Ed.), *Oxford handbook of linguistic minimalism*. Oxford: Oxford University Press.
- Gil, D. (2017). Isolating-Monocategorial-Associational Language. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (2nd ed., p. 471-510). San Diego: Elsevier.
- Gilligan, G. (1987). *A crosslinguistic approach to the pro-drop parameter* (Unpublished doctoral dissertation). University of Southern California.
- Givón, T. (1979). *On understanding grammar*. New York: Academic Press.
- Givón, T. (2002). *Bio-linguistics: the Santa Barbara lectures*. Amsterdam, Netherlands: John Benjamins Publishing Company.

- Givón, T. (2009). *The genesis of syntactic complexity: Diachrony, ontogeny, neuro-cognition, evolution*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language*. Cambridge, Massachusetts: MIT Press.
- Haegeman, L. (2012). *Adverbial clauses, main clause phenomena, and the composition of the left periphery*. Oxford: Oxford University Press.
- Haspelmath, M. (2020). Human Linguisticity and the Building Blocks of Languages. *Frontiers in Psychology, 10*.
- Haspelmath, M., Dryer, M., Gil, D., & Comrie, B. (2005). *World Atlas of Language Structures*. Oxford: Oxford University Press.
- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Malden, MA: Blackwell.
- Heine, B., & Kuteva, T. (2007). *The genesis of grammar. a reconstruction*. Oxford: Oxford University Press.
- Henshilwood, C. S., d'Errico, F., Yates, R., Jacobs, Z., Tribolo, C., Duller, G. A. T., . . . Wintle, A. G. (2002). Emergence of Modern Human Behavior: Middle Stone Age Engravings from South Africa. *Science, 295*(5558), 1278–1280.
- Henshilwood, C. S., & Dubreuil, B. (2009). Reading the artefacts: gleanings language skills from the Middle Stone Age in southern Africa. In *The Cradle of Language* (pp. 61–92). Oxford: Oxford University Press.
- Holmberg, A., & Roberts, I. (2014). Parameters and the three factors of language design. In *Linguistic variation in the minimalist framework* (pp. 61–81). Oxford: Oxford University Press.
- Hurford, J. R. (2007). *The origins of meaning: Language in the light of evolution*. Oxford: Oxford University Press.
- Hurford, J. R. (2012). *The origins of meaning: Language in the light of evolution II*. Oxford: Oxford University Press.
- Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences, 3*(7), 272–279.
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Kayne, R. S. (2005). Some notes on comparative syntax: With special reference to English and French. In G. Cinque & R. S. Kayne (Eds.), *The Oxford Handbook of Comparative Syntax* (pp. 3–69). Oxford: Oxford University Press.
- Larsson, I. (2017, 08). Double complementizers. *Nordic Atlas of Language Structures Journal, 1*. doi: 10.5617/nals.5413
- Ledgeway, A., & Roberts, I. (2017). Principles and parameters. In *The Cambridge Handbook of Historical Syntax* (pp. 581–626). Cambridge: Cambridge University Press.
- Lightfoot, D. (1991). Subjacency and sex. *Language & Communication, 11*, 67–69.
- Newmeyer, F. J. (2004). Against a parameter-setting approach to language variation. *Linguistic Variation Yearbook, 181–234*.
- Newmeyer, F. J. (2005). A reply to the critiques of ‘grammar is grammar and usage is usage’. *Language, 81*(1), 229–236.
- Pagel, M. (2013). *Wired for Culture: Origins of the Human Social Mind*. New York, NY: WW Norton Company.

- Pagel, M. (2017). Q&A: What is human language, when did it evolve and why should we care? *BMC Biology*, 15(1).
- Pesetsky, D. (2021). *Exfoliation: towards a derivational theory of clause size*. Retrieved from <https://ling.auf.net/lingbuzz/004440>
- Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 707–784.
- Preminger, O. (2011). *Agreement as a fallible operation* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
- Preminger, O. (2014). *Agreement and its failures* (No. 68). Cambridge, MA: MIT Press.
- Preminger, O. (2020). *Noam Chomsky and Benjamin Lee Whorf walk into a bar*. Retrieved from <http://web.archive.org/web/20220419182803/https://omer.lingsite.org/blogpost-chomsky-and-whorf-walk-into-a-bar/>
- Progovac, L. (2009). Sex and Syntax: Subjacency Revisited. *Biolinguistics*, 3(2-3), 305-336.
- Progovac, L. (2016). A gradualist scenario for language evolution: Precise linguistic reconstruction of early human (and Neandertal) grammars. *Frontiers in Psychology*, 7.
- Progovac, L. (2019). *A Critical Introduction to Language Evolution: Current Controversies and Future Prospects*. Cham: Springer.
- Richards, M. (2008). Two kinds of variation in minimalist syntax. In F. Heck, G. Müller, & J. Trommer (Eds.), *Varieties of competition* (pp. 133–162). University of Leipzig.
- Rizzi, L. (1982). *Issues in Italian syntax*. Dordrecht, Holland: Foris Publications.
- Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (pp. 281–337). Dordrecht: Kluwer Academic Publishers.
- Rizzi, L. (2013). Notes on cartography and further explanation. *Probus*, 25(1), 197–226.
- Roberts, I. (2019). *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press.
- Ross, J. (1967). *Constraints on variables in syntax* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
- Satk, D. (2022). The fine structure of the left periphery of infinitives. In *Proceedings of the North East Linguistic Society 52*. Amherst, MA: GLSA.
- Sheehan, M. (2014). Towards a parameter hierarchy for alignment. In R. E. Santana-LaBarge (Ed.), *Proceedings of the 31st West Coast Conference on Formal Linguistics* (p. 399-408). Somerville, MA: Cascadilla Proceedings Project.
- Tallerman, M. (2014). No syntax saltation in language evolution. *Language Sciences*, 207–219.
- Villa-Garcia, J. (2012). *The Spanish complementizer system: Consequences for the syntax of dislocations and subjects, locality of movement, and clausal structure* (Unpublished doctoral dissertation). University of Connecticut.