

This PDF includes a chapter from the following book:

Born to Parse

How Children Select Their Languages

© 2020 Massachusetts Institute of Technology

License Terms:

Made available under a Creative Commons
Attribution-NonCommercial-NoDerivatives 4.0
International Public License

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

OA Funding Provided By:

The open access edition of this book was made possible by generous funding from Arcadia—a charitable fund of Lisbet Rausing and Peter Baldwin.

The title-level DOI for this work is:

[doi:10.7551/mitpress/12799.001.0001](https://doi.org/10.7551/mitpress/12799.001.0001)

1 Three Visions

1.1 Invariant Principles and Their Successes

The biolinguistic enterprise, seeking to find the biological basis of linguistic structures, reflects the work of many people from many countries analyzing many very different grammars. They have discovered a huge range of interesting, abstract properties, by pursuing a particular, Minimalist vision of what a grammar should look like. This is the first of the three visions of this chapter: the quest for simple, invariant principles.

Acquisition is the process whereby a child selects a grammar conforming to those invariant principles. Invariant principles are universal, restrictive, and appear to be common to the species, serving to explain the similarity of the internal languages of speakers of many historically unrelated languages. A grammar is what we used to call the formal, generative system that characterizes a person's mature language faculty, which is represented in the individual's mind/brain. A grammar is now often referred to as an internal language, individual language, or I-language. Grammars, I-languages—terms used interchangeably throughout the book—are subject to the universal, restrictive principles referred to above.

Rich, invariant principles have emerged, often in response to arguments from *THE POVERTY OF THE STIMULUS*, a notion I will

discuss below. Such principles, defined universally, bridge the gap between information conveyed by a child's typically very limited experience and the rich information codified in mature grammars.

A simple example starts with the fact that in English, *wh*-elements occur at the front of expressions but may be understood in a wide range of positions, the strike-throughs:

- (1) a. Who did you see ~~who~~?
- b. Who did you speak to ~~who~~?
- c. Who did you expect ~~who~~ to win?
- d. Who did you say ~~who~~ left town?
- e. Who did you say Kim visited ~~who~~?

We analyze this as *wh*- phrases being copied from the position in which they are understood to a fronted position where they are pronounced. But there are various positions in which a copied *wh*-item may not be understood, for example, following a complementizer *that*, as in (2a): **Who do you think that has telephoned?* (*** indicates an expression that does not occur in people's speech). Further examples are in (2b–d).

- (2) a. **Who do you think [that ~~who~~ has telephoned]?*
- b. **Whose did she see [~~whose~~ pictures]?*
- c. **Who did she wonder [~~who~~ left]?*
- d. **Who did she meet the woman [who knew ~~who~~]?*

Generally, children are viewed as experiencing, and learning from, simple, robust expressions that they hear. The fact that **Who do you think that has telephoned?* is not said constitutes NEGATIVE DATA, information that something does *not* exist—something that is, in fact, precluded by some principle, which is not learned. It is not learned because it cannot be learned, since it would have to be learned based on negative data available to an analyst but not to a two-year-old child. The two-year-old has no evidence for the restriction. Therefore, what a two-year-old hears, the stimulus, is not rich

enough to fully determine what she comes to know. This is what linguists call the poverty of the stimulus, an important part of the logical problem of language acquisition.

Invariant principles explain negative data like (2), and it is postulated that they are available to children through their biology, that they are attributes of their genetic material, hence not learned, hence the solution to this poverty-of-stimulus problem. The principles explain how simple experiences can trigger rich structures in the biological grammars that constitute some form of Japanese or of Javanese (see Guasti 2016 for good textbook discussion of advances in language acquisition, escorting the reader from basic concepts to areas of current research in a theoretically well-informed fashion). Understanding the invariant principles that have been successfully identified illuminates how we might explain negative data and gain new ways to approach variable properties, properties that occur in some I-languages but not in others, an area where linguists have been conspicuously less successful (as we shall see in the next section).

However, here's an important point showing the need for abstract structures in the parses that children arrive at: the word-for-word translation of **Who do you think that has telephoned?* does exist in a number of languages, as noted by Rizzi 1990: §2.6. This means that those non-English forms must have a different abstract structure than the nonexistent English forms, a structure that is the result of a child's parsing. Superficially similar sentences can derive from structures that differ in crucial and nonobvious ways; that fundamental point is not sufficiently appreciated even by some linguists. Thus, for example, Italian *Chi credi che abbia telefonato?*, literally 'Who do you think that has telephoned?', has an independently motivated (partial) structure *Chi credi che ehi abbia telefonato ehi?*, where the embedded subject DP *chi*, 'who', is first copied to the post-VP position indicated by the second *ehi* and then copied again into the matrix clause. English I-languages do not copy subject DPs

to post-VP positions, but Italian children can learn to parse structures with a post-VP subject DP by hearing and understanding something simple like *Gianni crede che abbia telefonato Maria*, ‘Gianni believes that Maria has telephoned’, or even simpler *Ha telefonato Maria*, ‘Maria has telephoned’. English-speaking children hear no such forms and therefore do not understand or produce expressions like **John believes that has telephoned Mary* or **Has called Mary*. The ambient language does not trigger such inverted expressions, which are therefore not generated by the emerging grammar, unlike what happens for Italian children.

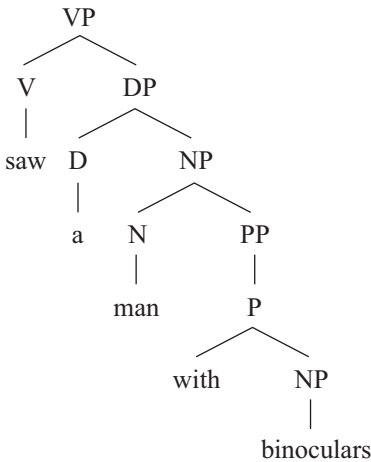
Two universal, invariant properties of all languages that are not learned are recursion and compositionality. All I-languages seem to have three recursive devices, looping functions that allow the repetition of clause types; the existence of these recursive devices means that humans have, in principle, the capacity to generate structures of indefinite length. The three are relative clauses, illustrated in (3a), complement clauses, in (3b), and coordination, in (3c).

- (3) a. This is the cow that kicked the dog that chased the cat that killed the rat that caught the mouse that nibbled the cheese that lay in the house that Jack built.
- b. Ray said that Kay said that Jay thought that Fay said that Gay told ...
- c. Ray and Kay went to the movie and Jay and Fay to the store, while Gay and May worked where Shay and Clay watched.

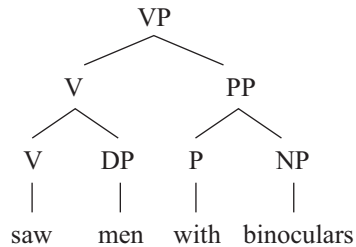
Second, I-languages are compositional; structures are binary branching and consist of units that, in turn, consist of smaller units, which consist of still smaller units. So *saw a man with binoculars* may have the structure of (4a). In that case, since *a man with binoculars* is a DP constituent, the meaning is ‘saw a man who had binoculars’. Another possible meaning of the phrase *saw a man with*

binoculars is ‘saw a man by using binoculars’, in which case the expression has a different structure, the one in (4b). Here *a man with binoculars* is not a single constituent as in (4a); instead, the preposition phrase *with binoculars* is generated as an adjunct to the VP *saw a man*.

(4a)



(4b)



The 1970s saw the formulation of very specific conditions on grammatical operations that disbarred long-distance operations taking place across certain intervening elements. The Subjacency Condition restricted movement to local relationships (accounting for the nonoccurrence of forms like **What did she wonder who bought?*). The Tensed-S Condition ruled out nonoccurring forms like **They expected that each other would win*, and the Specified-Subject Condition eliminated **They expected the women to visit each other* on the reading where *each other* COREFERS with *they* (it may of course corefer with *the women*). These conditions disbarred

operations across certain intervening elements. This was progress, but there were dramatic simplifying effects when scientists reformulated conditions to make analyses more readily learnable. For example, Lasnik 1976 rejected earlier attempts to formulate operations specifying what pronouns could corefer with in preference for specifying what pronouns could *not* corefer with. Lasnik turned things around in a way that led to a far simpler and more readily learnable account of the referential properties of pronouns.

The 1980s opened with Chomsky 1981a, focusing on the central and general role of GOVERNMENT relations and the spectacular simplifications stemming from the BINDING principles. This replaced the complexities of Chomsky 1980, particularly the stipulations of the indexing conventions given in the appendix to that paper. Chomsky built on Lasnik's reformulation and provided elegant accounts of binding relations, which we will examine in detail in §4.1.

Work during this period greatly enriched ideas about what kind of information might be built into our biology, as invariant principles of UG. Those principles enable children to go through that extraordinary third year of life, when, on the basis of rudimentary experience, they develop into recognizable human beings, thinking, understanding, and speaking more or less like an adult over an infinite range of thoughts, or at least like a person who will eventually develop into an adult human being.

My goal here is not to have readers relive the 1970s and learn how various principles were discovered. But readers do need to have a sense of the broad nature of that early work and of what kinds of problems it solved. I am providing some detail but not enough to be comprehensive. That would be a different book, which, having lived through the 1970s, I am not ready to write now.

One example of an early-formulated invariant principle is a condition on DELETION operations: that they can affect only an element that is in a prominent, easily detectable position, namely as (or inside) the COMPLEMENT of an overt head that is adjacent to that

complement (see Lightfoot 2006b). So the complementizer *that* may be deleted in (5) to yield (6) but not in (7) to yield (8). In (7) *that* is not the complement of *nor* in the complement of the adjacent word; in (7a), for instance, [*that Kay wrote*] completes the meaning of the nonadjacent *book* but not of the adjacent *yesterday*.

- (5) a. Jill said [that Jane left].
 b. The book [that Jill wrote] arrived.
 c. It was obvious [that Jill left].
- (6) a. Jill said [Jane left].
 b. The book [Jill wrote] arrived.
 c. It was obvious [Jill left].
- (7) a. The book arrived yesterday [that Kay wrote].
 b. [That Kay left] was obvious to all of us.
- (8) a. *The book arrived yesterday [Kay wrote].
 b. *[Kay left] was obvious to all of us.

This condition on deletion also distinguishes the (a) and (b) examples in (9) and (10): the deleted (empty) VP in (9b) and (10b) fails to meet our condition, since it is not adjacent to the word of which it is the complement, *had*, whose meaning it completes; it is adjacent only in the (a) structures.¹

- (9) a. They denied reading it, although they all had _{VP}e.
 b. *They denied reading it, although they had all _{VP}e.
- (10) a. They denied reading it, although they often had _{VP}e.
 b. *They denied reading it, although they had often _{VP}e.

Over the last two decades or so, under the Minimalist Program, there has been a change of emphasis: linguists have sought to simplify the principles, “minimizing” the information they embody. One form this has taken is to adopt an architecture that subsumes

certain apparently distinct principles. For example, in the 1960s and 1970s linguists wrote very specific top-down phrase-structure rules to capture initial “deep” structures, which included complex structural properties that were, *ex hypothesi*, learned by children. Chomsky’s *Syntactic Structures* (1957: 39) sketches a phrase-structure rule $Aux \rightarrow C (M)$ (*have + en*) (*be + ing*) (*be + en*), which contains much language-specific information, even specific English morphemes. Now things are very different: there is a general procedure for building structures, whereby a single, simple, recursive operation of MERGE creates binary-branching hierarchical structures. The invariant computational operation of (internal and external) Merge builds hierarchical structures bottom up. These structures combine heads with complements and phrasal categories with specifiers and adjuncts; this applies for all languages. This repeatable operation assembles two syntactic elements X and Y into a unit, which may, in turn, be merged with another element to form another phrase and so on. Merge is defined as $Merge(X,Y) = \{X,Y\}$ and thereby derives Third-Factor No Tampering, Inclusiveness, and the restriction to binary-branching structures (for discussion of third-factor elements, see Chomsky 2005). This means that as two elements, X and Y, are merged into a third category, neither of the two merged elements may be changed in any further way as a function of that operation. As Epstein, Obata, and Seeley 2017: 482 puts it, “by definition, neither X nor Y is altered by the operation and no new features are added to X or to Y in the constructed object, nor are any deleted from X or from Y.” Hence apparent properties of UG are derived through the invocation of the simple Merge operation.

Elements are drawn from the lexicon and merged into structures one by one; Merge is the fundamental structure-building operation. To clarify, the verb *visit* may be merged with the noun *London* to yield a VP, $_{VP}[_V visit \text{ } _N London]$, but that also shows the effects of PROJECT, a distinct aspect of structure building, because the verb *visit* projects to a verb phrase, VP. Then the Inflection element *will*

can be merged with that VP to yield an IP, projecting from Infl: $IP_{[Infl\ will\ VP_{[V\ visit\ N\ London]]}]}$. Then the (pro)noun *you* can be merged with that IP to yield another IP: $IP_{[N\ you\ IP_{[Infl\ will\ VP_{[V\ visit\ N\ London]]}]}]}$.

An expression *What did you buy?* is built bottom up in the same way. At a certain point, the IP *you did buy what* has been built: *buy* merges with *what* to yield a VP, then *did* is merged with the VP to yield an IP, and then *you* is merged with the IP to yield another IP, as above. Then what happens is that the previously merged element *did* is copied and merged again, and *what* undergoes the same process. In both cases, the copied element is later deleted in the original position from which it was copied, as indicated by the strike-through: *What did [you ~~did~~ buy ~~what~~]*. Under this approach, there is no primitive operation of movement as such; rather, a copied element may be merged and then subsequently deleted.

Repeated application of Merge is the engine of the computational operations that relate, for any expression, the phonological and semantic forms, which are interpreted at what are commonly called these days the sensorimotor (“articulatory-perceptual” in Chomsky 1995) interface and the conceptual-intentional interface. These interface forms still must meet their own requirements, as we will explore in chapter 4 (for discussion, see Chomsky 1995: 168–170).

Minimalism envisions general principles that are not learned but that limit structures and operations in such a way as to permit learnable operations that express, for example, the former specificities of phrase-structure rules. It is worth emphasizing here that our focus in this section lies in the way that Minimalists have sought to simplify and minimize the information in the invariant principles being attributed to UG and to human biology. Our goal has not been to catalog everything in UG but we do need to recognize that the invariant principles identified have consequences for what we can postulate in individual I-languages, after learning has taken place. For example, consider Chomsky’s Aux rule from *Syntactic*

structures, referred to four paragraphs ago, $Aux \rightarrow C(M)$ (*have + en*) (*be + ing*) (*be + en*). Heads such as tense markers (C), modals (*will*, *may*, *can*, *shall*, etc.), and aspectuals like *have* and *be* merge with their complements sequentially, and the particularities need to be learned by children as they parse their ambient, external language and discover its elements; we will discuss how this happens in §2.4, when we consider how these elements first emerged in the history of English.

Similarly, the conceptual and sensorimotor interfaces have their own properties, which may also capture language-specific, learned, variable properties. Den Dikken 2012 is a comprehensive compendium of developments in generative syntax over the last several decades, beginning with rich, complex transformations that developed into very general operations like Move α , or even Affect α . One sees in children the steady emergence of the simple Minimalist operations, with specificities emerging from interaction across operations. Nowhere is this clearer than in the limits on which DPs may corefer and which must be disjoint in reference. The very complex indexing conventions of the 1960s and 1970s have now given way to the simple and general binding principles to be discussed in chapter 4, under which children must learn which words are anaphors and which are pronouns, apparently a feasible task.

Part of the motivation for minimizing the information in UG is the legacy of William of Occam's simplicity in theorizing, always seeking simpler and therefore more beautiful analyses. Another part is the goal of providing a plausible biological account whereby we might attribute the evolution of the language faculty in the species to a single mutation at some level. Taking Merge to be a universal, invariant property raises the prospect that the possibility of Merge was the mutation that made language and thought possible for *homo sapiens*. Berwick and Chomsky (2016) showed why such a view might be productive and elicited a judicious and informed review from paleoanthropologist Ian Tattersall (2016).

Thinking in terms of simple hierarchical structures resulting from Minimalist computational operations, notably Merge and Project, has also informed remarkable neuroscientific work linking brain activity to the abstract structural units underlying language and thought. For example: as we discussed earlier, repeated application of Merge guarantees that syntactic structures are binary branching and therefore involve a narrow range of hierarchical relations. That has suggested for many years that children instinctively parse expressions in terms of those hierarchical relations and not in terms of purely linear sequences. Now we have neurophysiological evidence that this is so, suggesting that universal aspects of the structure of languages correlate to some degree with a pre-determined brain system.

In an experiment reported in Musso et al. 2003, Andrea Moro and colleagues exposed German speakers who had no previous encounters with Italian to an artificial variety of that language with Italian words but some variable syntactic properties different from those of natural Italian; they did the same with another group of German speakers and an artificial variety of Japanese. So, for example, there were no Italian-style null subjects: people would hear *Io mangio una pizza*, ‘I eat a pizza’, and never the subjectless *Mangio una pizza*, contrary to what occurs in normal, native Italian. The native German speakers were also exposed to “impossible” variable properties of Italian/Japanese, for instance, the negative marker placed after the third word in the unstructured expression, as never occurs in natural languages. So both groups of German speakers were exposed to naturally possible and naturally impossible artificial varieties of each language, Italian and Japanese.

The investigators analyzed their subjects’ behavior and tested the brain activity of those acquiring possible and impossible artificial languages. Subjects learned the real and unreal-but-possible languages with similar accuracy. fMRI results, however, showed significantly different brain activity in Broca’s area and elsewhere:

there was a “correlation between the increase in BOLD [blood-oxygen-level-dependent] signal in the left inferior frontal gyrus and the on-line performance for the real, but not for the unnatural, impossible language learning tasks. This stands as neurophysiological evidence that the acquisition of new linguistic competence in adults involves a brain system that is different from that involved in learning grammar rules that violate UG” and is based entirely on nonhierarchical, linear order (pp. 777–778). The authors go on: “[a]ctivation of Broca’s area is independent of the language (English, Chinese, German, Italian, or Japanese) of subjects,² suggesting a universal syntactic specialization of this area” among natural languages (conforming to principles of UG) (p. 778).³ “Our results indicate that the left inferior frontal gyrus is centrally involved in the acquisition of new linguistic competence, but only when the new language is otherwise based on principles of UG. The anatomical and functional features of Broca’s area allow us to speculate that the differentiation of this area may represent an evolutionary development of great significance, differentiating humans from other primates” (p. 779).

We remain far from understanding the neurophysiology of the language faculty and far from knowing the neural mechanisms for processing hierarchical syntactic structure, but these results do suggest that when the language faculty is “switched on,” producing or understanding language, certain kinds of brain activity are involved that are not involved when dealing with different, nonlanguage events.

As further illumination, David Poeppel and colleagues showed that when people listen to connected speech, cortical activity of different timescales tracks the time course of abstract structures at different hierarchical levels, such as words, phrases, and sentences (Ding et al. 2016). There are some problematic aspects to this study, but results indicate that “a hierarchy of neural processing timescales underlies grammar-based internal construction of hierarchical lin-

guistic structure” (p. 158). Ding et al. found neural activity that directly reflects the abstract structures that linguists have postulated for the infrastructure of language, needed to account for the way that expressions are understood and used. See also Nelson et al. 2017. We always knew that the brain would need a mechanism for encoding the abstract structures of different levels, and now we can begin to figure out some of the brain activity that takes place when that mechanism is operating, a major development.

Ding et al. discovered that the brain tracks units at each level of hierarchical structure simultaneously. Such tracking requires knowledge of how words and phrases are structurally related. Heidi Getz et al. (2018) also asked how neural tracking emerges as knowledge of phrase structure is acquired in an artificial language. They recorded electrophysiological data (magnetoencephalography) while adults listened to a miniature language with distributional cues to phrase structure or to a control language without the distributional cues. They found that neural tracking of phrases developed rapidly when participants formed mental representations of phrase structure, as measured behaviorally, thereby illuminating the mechanisms through which abstract mental representations are acquired and processed by the brain.

Minimalist ideas about hierarchical structures being formed by multiple applications of Merge not only help us think differently about the evolution of the language faculty and of thought in the species and stimulate new neuroscientific work; they have also facilitated new approaches to the acquisition of language by children. The hierarchical structures formed by multiple applications of Merge constitute the means by which people, including very young children, begin to analyze and parse what they hear—the key component of the acquisition process, as Janet Fodor argued long ago in an important pair of papers, “Learning to parse?” and “Parsing to learn” (Fodor 1998b,c). To parse is to assign linguistic structures to I-language units and their interrelationships. We may now

be at the point where we can dispense with independent parsing procedures, a goal to which Colin Phillips has devoted much of his career; see Phillips 2003a,b. Given the way in which hierarchical structures are built, we might argue that assigning structures to expressions is simply a matter of using the binary-branching structures that UG (in particular the Project and Merge operations) makes available and the structures that children discover as they parse their ambient E-language. Under this view, parsing is not a function of nonlinguistic conditions. Rather, parsing is just a function of the emerging I-language (as we will explore in chapter 3).

Work has shown repeatedly that children rely on the tools provided by their biology and learn much from little experience. Research has examined language acquisition by children exposed only to unusually restricted data, much of the work focusing on the acquisition of signed systems. A striking fact is that 90 percent of deaf children are born to hearing parents, who are normally not experienced in using signed systems and often learn a primitive kind of pidgin to permit rudimentary communication. In such contexts, children surpass their models readily and dramatically and acquire effectively normal mature capacities, despite the limitations of their parents' signing (Newport 1998; Hudson Kam & Newport 2005; Singleton & Newport 2004).

This is not surprising in light of studies of creoles more generally (Aboh 2017) and of new languages beyond creoles, which show that children exposed to very limited experiences go well beyond their models in quickly developing the first instances of rich, new I-languages (Lightfoot 2005, 2006a). Not much is needed for a rich capacity to emerge, as documented by many contributors to Piattelli-Palmarini and Berwick 2013 and now by Belletti 2017. Belletti offers a new kind of poverty-of-stimulus argument, showing that children sometimes overextend certain constructions, using them much more freely and creatively than their adult models.

Extraordinary events have cast new light on these matters: the birth of new languages in Nicaragua and in Bedouin communities in Israel. In Nicaragua the Somoza dictatorship treated the deaf as subhuman and barred them from congregating. Consequently, deaf children were raised mostly at home, had no exposure to fluent signers or to a language community, were isolated from each other, and had access only to home signs and gestures. The Sandinistas took over the government in 1979 and provided a school where the deaf could mingle, soon to have four hundred deaf children enrolled. Initially the goal was to have them learn spoken Spanish through lip reading and finger spelling, but this was not successful. Instead, the schoolyard, streets, and school buses provided good vehicles for communication and the students combined gestures and home signs to create first a pidgin-like system, then a kind of productive creole, and eventually their own language, Nicaraguan Sign Language. The creation of a new language community took place over only a few decades. This may be the first time that linguists have witnessed the birth of a new language *ex nihilo* and they were able to analyze it and its development in detail. Kegl, Senghas, and Coppola 1998 provides a good general account and Senghas, Kita, and Özyürek 2004 examines one striking development, where certain signs proved to be not acquirable by children and were eliminated from the emerging language, an interesting example of where adult learning differs from that of children.

Sandler et al. 2005 discusses the birth of another sign language among Bedouin communities in Israel, which has arisen in ways similar to Nicaraguan Sign Language and was discovered at about the same time. These two discoveries have provided natural laboratories to study the capacity that children exposed to unusually limited linguistic experience have to go far beyond their models and attain mature I-languages that conform to general, known principles.

Now Rachel Mayberry has discovered a third new sign language (Mayberry & Kluender 2018). In March 2016, thirty-five deaf children were brought together in two all-deaf classrooms in Iquitos, Peru. They were identified as still using home signs, before their new, common language began to emerge. The goal of Mayberry's work was to understand the very first stages of language emergence. In contrast, the first work on early speakers of Nicaraguan Sign Language was done ten years after the common language had begun to emerge, so much development had already taken place. Mayberry was able to gather data about deaf children's gestures before they met other deaf children, studying how their gestures changed throughout their first year of interaction with each other. Deaf children across the world often grow up without access to a shared language, so Mayberry's data will provide the first documentation of how children's communicative interactions via gesture become codified into the initial I-language elements of the new language. If it is correct that this is a common experience for deaf sign users, this work will likely lead to an outpouring of comparable data from different contexts, heralding new insights into the emergence of new languages.

If successful language acquisition may be triggered even by exposure to very restricted data, then perhaps children learn only from simple expressions. They only need to hear simple expressions, because there is nothing new to be learned from complex ones. This is degree-zero learnability, which hypothesizes that children need access only to unembedded material (Lightfoot 1989). Such a restriction would explain why many languages manifest computational operations in simple, unembedded clauses but not in embedded clauses (e.g., English subject-inversion sentences like *Has Kim visited Washington?* but not comparable embedded clauses **I wonder whether has Kim visited Washington*), whereas no language manifests the reverse: operations that appear only in embedded clauses and not in matrix clauses.⁴ One explanation for this

striking asymmetry is that children do not learn from embedded domains. Therefore, much that children hear has no consequences for the developing I-language; nothing complex triggers any aspect of I-languages.

In 2014, Norbert Hornstein and Bill Idsardi, working with a pre-publication version of Heinz 2016, proposed that investigating whether children's I-languages are influenced only by simple structures, specifically from unembedded binding domains, be taken as a central "Hilbert problem" for linguistics. That is, a problem or area of investigation where advances would influence the future of the field. Postulating degree-zero learnability has already led to productive reanalyses of phenomena that *seemed* to suggest that children learn from embedded material; there are now significantly better analyses that are based on simple triggers (Lightfoot 2012). That kind of analytical productivity is the hallmark of a Hilbert problem. For related discussion, see also Heinz and Idsardi 2011 and 2013. This Hilbert problem also goes beyond approaches to language acquisition that view it as a process of setting binary parameters.

Before we turn our attention to variable properties in the next section, let us first consider Charles Yang's Tolerance Principle, a general invariant principle he has recently discovered that governs irregularity. Yang 2016 shows that children make a categorical distinction between rules and their exceptions and offers a computational account that relates patterns of regularity to the number of possible irregular forms. Yang draws on thinking in classical economics about the price of goods reflecting a balance between supply and demand. In this view, children discover a productive rule only if it yields a more efficient organization of language, with the number of irregular forms falling below a well-defined threshold. Yang argues for his principle by showing what it predicts for irregular linguistic phenomena across countless languages. A striking property of the principle is that it makes predictions about where generalizations may also be formulated for nonlinguistic cognitive

domains. There must be a principled distinction between a core I-language and peripheral irregularity. Given Yang's Tolerance Principle, for a linguistic rule to be productive, the number of exceptions must fall below a critical threshold that can be calculated precisely.

Recall that this chapter is entitled "Three Visions." Our first vision is that of Minimalists who have focused on invariant principles. That focus has been remarkably successful in generating understanding of what happens in that astonishing third year of life when a child matures into a full-fledged member of the human species. However, while Minimalists are well aware of the interrelated problems of language variation and language acquisition, they have had strikingly little to say about them.

Obviously, there is much more to say about our first general vision. Indeed, Chomsky's Uniformity Condition, to be discussed in the next section, seems to indicate that there is no variation at the level of I-languages and that all children attain one, invariant language, Human, stated at a level of abstraction that embraces a competence in English, French, Javanese, Japanese, or any other of the apparently diverse languages. We will return to the matter of Human and an abstract skeletal structure again in §1.3 and in chapter 6. We will eventually argue for a form of that vision, by arguing that UG is "open" and that variable properties are learned through the parsing operations. Meanwhile, let us consider the difficulties generativists have had in characterizing variable properties in I-languages and their acquisition, specifically the problems with parameters and how they are set.

1.2 Parameters and Their Problems

Postulating hierarchical linguistic structures formed by a simple Merge operation has yielded new understanding of the invariant properties of language and has generated an immensely fruitful

research program, bringing explanatory depth to a wide range of phenomena (Den Dikken 2012). However, a hallmark of human language, alongside its invariant properties, is its apparent variation. Some properties occur only in certain I-languages, not in others; these are the variable properties. The second vision motivating some linguists focuses on those variable properties. As Charles Yang (2016: 1) puts it,

A theory of language needs to be sufficiently elastic to account for the complex patterns in the world's languages but at the same time sufficiently restrictive so as to guide children toward successful language acquisition in a few short years (Chomsky 1965).

The environmentally induced variation that one finds in language is biologically unusual, not what one sees generally in other species' faculties or in other areas of human cognition, and requires a biologically coherent treatment. Children appear to attain significantly different internal languages, depending on whether they are raised in contexts using some form of Swedish or a kind of Vietnamese. English speakers in seventeenth-century London typically acquired different grammars from those acquired three generations earlier. Furthermore, people speak differently depending on their class background, their geography, their interlocutors, their mood, their alcohol consumption, and other factors.

For a good biological understanding, variation in grammars, I-languages, needs to take its place among other types of variation. This is an area where we have made much less progress than with invariant properties and where there needs to be new thinking. Indeed, syntacticians have very different ideas about what biology provides with respect to variable properties. Although our second vision focuses on issues of variation and acquisition, its implementation has turned out not to be as successful as was hoped.

Chomsky 1981b initiated the "Principles and Parameters" approach, seeking to find a UG with both invariant principles and a set of formal parameters that children were thought to set on exposure

to PRIMARY LINGUISTIC DATA. PLD are the simple data that all children can be expected to hear and that constitute the triggering experience for language acquisition. PLD do not include nonprimary data about what does *not* occur, how things might be translated, what are paraphrase relationships or the scope of quantifiers, or other exotic data that are useful for linguists figuring out the best analysis for a language's syntax but are largely irrelevant to children's acquisition.

For four decades, linguists have been postulating parameters, ideally binary parameters (either structure *a* or structure *b*, for example, head-final or head-initial), but no real, general theory has emerged, and genuinely binary parameters are scarce. The key idea is that children set parameters depending on what they experience (deciding whether their PLD require a head-initial parameter setting, for example). Minimalists set on reducing the complexities of the invariant principles that had emerged by the mid-nineties have not devoted equivalent efforts to minimizing the complexities of UG parameters, the variable properties, nor, more importantly, to giving an account of how parameter settings might be acquired by young children. Scientists following the parameter-based vision have sought a theory of variable properties by defining those properties at the level of UG, which incorporates a set of parametric options alongside the invariant principles. A much cited metaphor equates parameter settings to on–off switches that children flip in response to the PLD that they experience.⁵ The vision was that those switches would have multiple consequences, capturing “harmonic” properties. For example, a language with null subjects, like Italian and Spanish, would also manifest complex verbal morphology, *that*–trace violations, and other related properties.

The term *parameter* was first introduced into the syntactic literature by Luigi Rizzi in footnote 25 of Rizzi 1978, when he “parameterized” bounding nodes for the Subjacency Condition. Different I-languages have different phrasal categories serving as

bounding nodes, and Rizzi addressed acquisition issues by postulating markedness relations between them. Elaborating the Principles and Parameters approach to universal and variable properties, Chomsky 1981b takes principles to embody the universal properties of I-languages and parameters to embody the limits on variation in I-languages. Parameters have been used for quite different forms of variation: sometimes for specific structural properties, such as whether IP might count as a node limiting long-distance movement (Rizzi 1978); sometimes defining language “types,” such as “null-subject languages,” which have null subjects and related harmonic properties (Roberts & Holmberg 2010; and see Haider 2005 for a “typological” approach to differences between Icelandic and German and Cinque 2013 for broader considerations); sometimes distinguishing parameters and microparameters for different scales of variation (Baker 2008; Kayne 1996; Westergaard 2009, 2014, 2017), perhaps keying microvariation to features on functional categories as in Kayne 2005; or even keying *all* variable properties to features on functional categories, as in the Chomsky–Borer Conjecture (Borer 1984; Chomsky 1995). A fundamental observation that motivated the idea of parameters was that variation does not come in random ways but in *clusters*. Parameters, whatever they are exactly, are abstract and designed to capture the clusters of phenomena that make up actual points of variation between I-languages. If we argue against the existence of binary parameters, we will need another way of capturing the harmonic properties that variation manifests, a matter that we return to in chapter 5.

Sometimes researchers invoke what is called the Chomsky–Borer Conjecture as a theory of parameters, saying that parameters constitute features on functional categories. Restricting variation to elements of morphology (lexical features of functional categories) avoids any direct, clear role for structures and leaves much unsaid; it cannot be a restrictive theory of parameters until we have a substantive theory of lexical features, and it pushes matters back to

asking what are the lexical features and what theoretical ideas might unify them.⁶

The vision behind this conception of on–off parameter switches is that points of variation fall into a narrow class, with perhaps just thirty or forty parameters defined at UG. Under this second of our three visions, variation would be expected to be narrowly defined and limited. However, when one looks at the kind of variation attested in I-languages, the vision of a limited number of points of structural variation does not look plausible; one would expect the variation to be less varied than it in fact appears to be. For example, the parade case of a parameter, the null-subject parameter, takes many different forms, and one sees languages developing very narrow and specific new variable properties (chapters 2 and 3). Even so, the parameter-based vision presents an important challenge for Minimalist thinking, since it attributes a great deal of information to the linguistic genotype, the kind of rich information that has proven dispensable in the UG-defined invariant principles. However, there are other problems with this conception of grammatical variation, which suggest that it may be time to try another approach (cf. also Boeckx 2014; Epstein, Obata, & Seeley 2017).

Meanwhile, there is no coherent definition of parameters and the term is used to refer to quite different things, as just noted. And separately, there has been very little discussion of how children set the parameters that have been proposed. Indeed, variation in parameter settings does not have the explanatory depth that we see in invariant principles of UG, and I am not aware of arguments for any particular parameter-based form of poverty-of-stimulus reasoning. But Crain 2012 does argue that one needs parameters to understand the scope properties that children demonstrate when acquiring English and Mandarin quantifiers; I have not yet worked on these cases to see how a parsing approach might fare. Parameters used to be the only way of approaching issues of variability in mainstream generative work, but now we have parsing-based

approaches. For example, Getz 2018a tackles an old problem treated in terms of parameters and, developing an analysis based on parsing, shows how children can learn the relevant distinctions. Newmeyer 2004 holds that parameters have not illuminated the nature of linguistic variation and that linguists have been unprincipled in their use of them (but for critical discussion see Biberauer 2008, including the editor's introduction, and Holmberg 2010 and Roberts & Holmberg 2010).

Linguists study variation in silos: syntacticians studying parameters have little contact with sociolinguists studying variable rules, and proponents of variable rules do not interact much with Optimality theorists studying constraint reranking, who, in turn, do not collaborate with cartographers seeking the variable properties associated with different functional heads. Indeed, Minimalists have devoted little attention to variation and acquisition (the two go together: variable properties must be acquired by children during development, whereas invariant properties may be provided in advance by UG and need not be triggered by primary data). The idea of UG parameters (macro- and micro-) grossly violates Minimalist aspirations to minimize information at UG. Indeed, Chomsky 2001 invoked a Uniformity Principle, abstracting away from variation: "in the absence of compelling evidence to the contrary, assume languages to be uniform, *with variety restricted to easily detected properties of utterances*" (my emphasis). This opens the logical possibility that all children attain the same I-language, which might be called Human, whether they are raised in a Javanese- or Japanese-speaking environment. Human would be stated at a level of abstraction that would embrace the specificities of Javanese or Japanese. Chomsky has never argued for such a possibility, as far as I know, but it would comport with his Uniformity Principle, and we will in §1.3 suggest a way of implementing this idea in terms of UG being open to information that is learned through parsing.

Hornstein's 2009 effort to transform the Minimalist Program into a Minimalist Theory has essentially no discussion of variation or acquisition, apart from a brief, four-page discussion of the history of parameters (pp. 164–168). Boeckx 2015 goes further and seeks to eliminate lexically determined features on the remarkable grounds that “they are obstacles in any interdisciplinary investigation concerning the nature of language,” as if linguists should deal only with analytical machinery invoked by present-day biologists.⁷

Difficulties with parameters are aggravated by the absence of an adequate account of which PLD set which parameters and how. It is often supposed that children evaluate candidate grammars by checking their generative capacity against a global corpus of data experienced, converging on the grammar that best generates all the data stored in the memory bank. But that entails elaborate calculations by children and huge feasibility problems (Lightfoot 2006a: §4.1).

The best-worked-out evaluation systems are Robin Clark's FITNESS METRIC (1992) and Gibson and Wexler's TRIGGERING LEARNING ALGORITHM (1994). Both systems involve the *global* evaluation of grammars as wholes, where the grammar *as a whole* is graded for its efficiency (Yang 2002). Children evaluate postulated grammars as wholes against the whole corpus of PLD that they encounter, checking which grammars generate which data. This is “input matching” (Lightfoot 1999) and raises questions about how memory can store and make available to children at one time everything that has been heard over a period of a few years.

Clark's genetic algorithm employs his Fitness Metric to assign a precise number to each grammar based on what it can generate and how many “violations” there are, that is, how many unacceptable structures the grammar generates wrongly. The key idea is that certain grammars yield an understanding of certain sentences and not others; put differently, they generate certain sentences and not oth-

ers. The Fitness Metric quantifies the failure of grammars to parse sentences by again counting the “violations,” in this case, sentences experienced that cannot be generated by the grammar being evaluated. There are two other factors involved in his Fitness equation, a superset penalty and an elegance measure, but those factors are subject to a scaling condition and play a minor role, which I ignore here. The Fitness Metric remains the most sophisticated and fully worked-out evaluation measure that I know. It is a global and very precise measure of success, assigning specific indices to whole, fully formed grammars against a whole corpus of sentences, rating their success exactly according to the extent that their output matches the child’s input.⁸

Gibson and Wexler take a different approach, but in their view as well, children effectively evaluate whole grammars against whole sets of sentences, although they react to local particularities. They are “error-driven,” and they determine whether a certain grammar will generate everything that has been heard and whether the grammar matches the input. Gibson and Wexler’s children acquire a mature grammar eventually by using a hypothetical grammar (a collection of parameter settings) and revising it when they encounter a sentence that their current grammar cannot generate (an “error” in Gibson & Wexler’s terminology, corresponding to Clark’s “violation”). In that event, children follow Gibson and Wexler’s Triggering Learning Algorithm to pick another parameter setting, and they continue until they converge on a grammar for which there are no unparsable PLD and no errors. Gibson and Wexler 1994 uses a toy system of three binary parameters that define eight possible grammars, each of which generates a set of sentence types. The child essentially tests the generative capacity of grammars in light of the complete set of sentence types experienced. Both Clark’s child and Gibson and Wexler’s child *calculate* the degree to which the generative capacity of the grammar under study conforms to what they have heard.

Even considering a small number of parameters, the problems become clear. If parameters are independent of each other, forty binary parameters entail over a trillion possible grammars, each generating an infinite number of structures. Parameters, of course, are not always independent of each other, and therefore the number of grammars to be evaluated might be somewhat smaller. On the other hand, Longobardi et al. 2013 postulates fifty-six binary parameters just for the analysis of noun phrases in Indo-European languages, which would suggest much larger numbers. On any account, the relevant numbers are astronomical.

It is worth bearing in mind that all I-language variable properties result from change (assuming that all languages have a common, single, monophyletic origin). Alongside addressing feasibility issues, one's theory of variation must accommodate grammatical change from one generation to another. Seeking to explain changes through language acquisition requires an approach to language acquisition different from input matching. If children acquire grammars by evaluating them against a corpus of data, then for an innovative grammar, they need to be confronted with the data generated by this grammar in order to select the grammar that generates them. This introduces problems of circularity: what comes first, the new grammar to generate the new data or new data that require the child to select the new grammar? Given that change seems to be ubiquitous and is the source of all variation, this circularity is a reason to discount the input-matching approach to the acquisition of variable properties.

There are other conceptual problems with viewing children as evaluating the generative capacity of numerous grammars, calculating how best to match the input experienced, and setting binary parameters accordingly (see also Boeckx 2006, 2014; Lightfoot 2006a; Newmeyer 2017). Certainly, given the abstractness of grammars, it will not do to suppose that triggers for elements of I-languages are sentences that those elements serve to generate, as

seems to be assumed by Crain and Thornton (1998); that would introduce other problems of circularity.⁹

To illustrate the lack of explanatory power in parameters, consider a formal parameter, postulated to be part of UG, according to which a head either precedes or follows its complement: {head, complement}. An alternative is to say that a child parses head-first expressions like [_v *sing* _{DP} [*three songs*]] and [_N *books* _{PP} [*about cities in California*]] as such. The “ordering parameter” solves no poverty-of-stimulus problem and does not reduce the information needed for a child to converge on the right structure. Similarly, it is not helpful to say that languages like Dutch have a parameter set to yield verb-second structures like _{DP} [*drie studenten*] _{CP} [_C *bezoeken* _{VP} *Utrecht*] ‘three students visit Utrecht’, as opposed to saying that children parse such structures and consequently have I-languages that generate such structures. It would also not be explanatory to say that English I-languages have a parameter set to allow complex DPs like *the man from Utrecht’s daughter with long hair* as opposed to saying that children parse such expressions with ‘s analyzed as a Determiner clitic licensing the complex DP *the man from Utrecht*: _{DP} [*the man from Utrecht*] _{Det} ‘s _{NP} [*daughter with long hair*]. Rather, UG is open to such structures being postulated when children are exposed to relevant ambient language that requires or EXPRESSES such structures.

In summary, combining our first two visions so that I-languages are viewed as consisting of invariant properties and a set of formal parameter settings, with both principles and parameters being defined at UG, fails to account for language acquisition and language variation. In particular, the input-matching approach has not been successful and faces apparently insuperable difficulties. Parameters stated as part of UG have no useful role to play, and it may be opportune to consider other approaches, based on a different vision, where parsing plays a more central role.

1.3 Principled Parsing for Parameters

For our third vision, we build on earlier work that treats children as predisposed to assign linguistic structures to what they hear, as born to parse. Rather than calculating and ranking the generative capacity of grammars and setting formal parameters, children are endowed with the tools provided by a restrictive and Minimalist UG that has no specific parsing procedures (Lightfoot 2017b). They assign structures to what they hear and use the structures necessary to interpret what is heard. Children thereby discover and select the structures, in principle one by one, as those structures are required by what they are hearing, as they are demanded by new parses.¹⁰ Those structures are part of the child's emerging I-language and, in aggregate, constitute the mature I-language. Such an approach enables us to understand how children develop their internal system and how those systems may change from one generation to another, as revealed by work on historical change in syntactic systems. After all, *all* variable properties of I-languages must originate in change (Lightfoot 2018).

So internal languages consist of invariant properties over a certain domain *plus* supplementary structures that are not invariant but are required in order to parse what children hear, consistent with the invariant properties. Children select what they need to accommodate new parses, drawing on what UG makes available, notably structures built by Merge and Project as sketched in §1.1. Put differently, children parse the external language they hear, assigning to expressions structures provided by the bottom-up procedures of Project and Merge. They discover and select specific I-language elements required for certain aspects of the parse, and they do so by using what UG makes available and what they have in their emerging I-language. I-languages grow over time through parsing, since parsing enables the child to discover new contrasts in the ambient E-language, hence variable properties in the emerging I-language.

The aggregation of those parsed elements constitutes the complete I-language beyond what is given by UG.

When E-language shifts, children may parse differently, and thus a new I-language emerges. Children discover variable properties of their I-languages through parsing with the available hierarchical structures; there is no evaluation of I-languages, there are no binary parameters provided by UG, and there are no special parsing procedures. That constitutes a major minimization of current theories of grammar, declaring substantial aspects of them redundant and not needed.

There is an interplay between E-language, which is parsed, and I-languages, which result from parsing. Parsing is a direct function of an emerging I-language, not of independent parsing procedures designed to produce the structures required (more on this in chapter 3). Children discover the structural contrasts in their ambient, external language and select the structures of their emerging internal language to accommodate the contrasts.

This approach grows out of earlier work on “cue-based acquisition,” which also took parsing to be a key element of acquisition (Dresher 1999; Fodor 1998a; Lightfoot 1999; Lightfoot & Westergaard 2007; Sakas & Fodor 2001). However, that early work postulated that the cues that children might discover were specified in UG, as in a restaurant menu, thereby opening itself to the objections to richly specified parameters discussed in §1.2 and predicting that new variable properties fall into a narrow class, depending on the number of parameters. Ideas that involve rich and specific information about variable properties stipulated at UG risk being biologically implausible and at variance with the aspirations of the Minimalist Program. In the current vision, variable structures, those showing up in some but not all I-languages, are not stipulated at UG but are discovered as children parse the ambient E-language, hence learned. The approach allows good understanding of why

English has adopted so many structural innovations, apparently quite idiosyncratic and unprincipled and not shared by its closest historical relatives. A researcher following this vision would expect a greater range of variation in the world's languages and would seek new, more biologically plausible ideas about variation (see chapters 5 and 6).

Two well-studied English examples very briefly; for details see Lightfoot 2017b and sections §2.4 and §3.3 here.

First, in the early sixteenth century a change was completed whereby, after the loss of very rich verbal morphology, a set of preterite-present verbs, verbs with past-tense forms but present-time meaning, came to be parsed as Inflection elements with a very different syntax from verbs, unlike in any other European language. This change has been studied by many diachronic syntacticians and is now well understood. After the widespread simplification of inflectional morphology in Middle English, only two verbal inflections survived, the third-person singular marker in *-eth* or *-s* and the second-person marker in *-st*. The preterite-presents never had the third-person marker. That absence was just one of many facts about morphology, but after the great simplification, the absence of those forms came to make these verbs categorially distinct and unlike any others. In addition, the past-tense forms of these so-called Modals like *could*, *would*, *should*, and *might*, only rarely conveyed past time, unlike *opened*, *thought*, and so on. Rather than expressing past time, they expressed a modal “subjunctive” meaning. The evidence is that these forms were parsed as a distinct, non-verb category of Inflection, which entailed a different syntax and the simultaneous obsolescence of formerly robust forms like *He has could see stars*, *She wanted to can see stars*, *She will can see stars*, and so on.

Because of these morphological changes, Middle English children began to discover new contrasts: the verbs *can*, *may*, *must* now ceased to look like verbs like *open*, *read*, *think*, because they lacked

the third-person marker. Establishing such contrasts is the basis of parsing, and Elan Dresher's work becomes important here, seeking a general approach to contrast hierarchies in phonology and revising Roman Jakobson's work on the topic (Dresher 2009, 2019, and Cowper & Hall 2019, have extended these ideas to syntax).

Second, with the loss of case morphology on nouns, which therefore no longer occurs in the ambient E-language, a set of forty or so verbs indicating psychological states underwent a kind of reversal of meaning and a new syntax, *like* changing from 'please' to 'enjoy', *repent* changing from 'cause sorrow' to 'feel sorrow', and a theme subject changing to a patient with both verbs. With the loss of morphology, children came to parse expressions differently, assigning different structures: *Gode ne licode na heora geleafleast* 'to-God did not like their faithlessness' was no longer parsed with *Gode* 'God' as a dative case, *heora geleafleast* 'their faithlessness' as a nominative subject, and *licode* (*liked*) meaning 'pleased' or 'caused pleasure'. Instead, based on these new parses, children selected new I-languages that entailed the new structures and new semantics of these psych verbs ('God' as a nominative, 'faithlessness' as a theme, and *like* with a new meaning, 'enjoy' or 'experience pleasure'). It is hard to see how an explanation for these changes could be provided if children were evaluating multiple grammars and setting formal parameters, and it is equally hard to see what binary structural parameters might be implicated; we return to this in §3.3.

This is too brief an account, but it is time for an alternative to parameter setting as an approach to variable properties; we need something that matches our success with invariant principles. I suggest that, rather than being parameterized, UG is OPEN, consistent with a variety of structures, depending on what contrasts the parsing process reveals. Given what they experience in their ambient external language, children may or may not select nominal case endings or verbal suffixes in their I-languages; that is determined

by the interplay between E-language and I-languages, and UG does not include particular parameters.

Under this approach we do not overtheorize variation by trying to specify particular options in a very rich UG. Rather, UG is unspecified in certain ways, open to demands that emerge from parsing E-language, which are accommodated by the categories, structures, and operations made available by UG and by the elements in a child's emerging I-language. Variation is inherent to the language faculty by virtue of its openness. This approach recalls the 1970s distinction between core and peripheral properties: UG provides skeletal structures that are enhanced by the results of parsing. Extending the metaphor, UG provides a skeletal structure, fleshed out by the consequences of children parsing the E-language they are exposed to. Under the vision sketched in this section, children discover variable properties by parsing the ambient E-language with the tools provided initially by UG. Through parsing, they also postulate structures in their emerging I-language that are required to understand the language(s) around them. Those variable properties are not stipulated in UG in the form of parameters to be set; that would be too much information for the stripped-down, minimized UG furnished by Minimalist thinking. This vision allows for a greater range of variable properties than thirty or forty UG-defined parameters would permit and requires a larger role for learning.

It remains to be seen whether this line of thinking yields a productive approach to variation beyond what we have viewed in the past as syntactic parameters. Hopefully, as linguists focus on the nature and acquisition of variable properties in all domains of language (see Lightfoot & Havenhill 2019), they will facilitate new approaches to variation, leading to an understanding of why human language, unlike other aspects of cognition, encompasses so much variation.

1.4 Reflections

Most discussion of variable properties concerns either properties that vary across languages or properties that vary within the PLD that an individual learner is exposed to. The former are mostly considered by generative grammarians; the latter are the focus of much work in sociolinguistic, Labovian traditions. Each of these silos could benefit from new thinking about the sources of variation, and perhaps there will be some coming together, as people see interactions between the two paradigms. After all, variation within external language triggers different individual, internal languages.

Another theme in discussion of variable properties concerns the role of statistics with respect to I-languages: statistics can complement theories of language, not replace them, if we see statistical variation within a language as something to be explained, not the explanation. Again Yang's work (Yang 2002) treats statistical information as an aid to children acquiring I-languages; statistics *reflect* properties of I-languages (see Lidz 2010; Lidz & Gagliardi 2015). And Getz 2019 shows that the age at which children learn categorical rules may restrict their nature. One can hope that different types of variation may be understood in a shared perspective and silos may become more porous.

It is interesting to reflect on how thinking has changed about the invariant principles of UG and the variable parameter settings since the early days of the Principles and Parameters approach, the early 1980s (Chomsky 1981b). In those early days the general goal was to find principles and parameters of UG that could solve the poverty-of-stimulus problems that were identified. Since researchers have sought to get "beyond explanatory adequacy," the strategy has been to reduce UG to a minimum. This does not mean that we should retreat from solving poverty-of-stimulus problems and rely on biologists to solve our analytical problems (contra Boeckx 2015; see

Lightfoot 2017b for discussion); it does mean that we need richer notions of learning beyond simply invoking properties of UG.

Two examples where this has happened: first, Getz (2018a,b) has shown that we can get better understandings of some well-known poverty-of-stimulus solutions with a richer understanding of the learning involved and how it may come from parsing ambient language. Second, Dresher, Cowper, and Hall's contrastive theory (Dresher 2009, 2019; Cowper & Hall 2019) does not assume innate features, unlike in earlier generative phonology. They can give up innate features because they postulate something else innate, contrastive hierarchies and the concept of features.

I hope that this book will help the field get to a better, more unified understanding of variable properties, in particular by recognizing that understanding their acquisition better will yield a better understanding of their nature.

Here I have explored three visions, placing bets on the third. Let us now explore that vision more fully and see what its mechanisms might offer people seeking to understand the third year of a child's life.¹¹ I will now sketch three new parses, three instances of new structures being discovered and selected. Meanwhile I will discard the second vision and develop the first and third.