Born to Parse
How Children Select Their Languages

David W. Lightfoot

The MIT Press
Cambridge, Massachusetts
London, England

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

ARCADIA
A charitable fund of Lisbet Rausing and Peter Baldwin

Library of Congress Cataloging-in-Publication Data

Names: Lightfoot, David, 1945– author.
Title: Born to parse : how children select their languages / David W. Lightfoot.
Description: Cambridge : The MIT Press, 2020. | Includes bibliographical references and index.
Identifiers: LCCN 2019044359 | ISBN 9780262044097 (hardcover)
Subjects: LCSH: Language awareness in children. | Grammar, Comparative and general—Parsing. | Grammar, Comparative and general—Syntax. | Second language acquisition—Study and teaching (Elementary)
Classification: LCC P118.3 .L54 2020 | DDC 401/.93—dc23
LC record available at https://lccn.loc.gov/2019044359:

d_r0

Contents

Preface

This book is about the acquisition of language by young children. For all children, acquisition begins very early, arguably in utero, when they often respond differently to the sounds their mother makes in her language than to sounds from other sources. Initially, for example, babbling happens with a wide range of sounds, drawn from all the languages of the world and possibly even beyond, and steadily narrows to the smaller number of sounds that the child will use for the rest of her life, selecting them for what we will be calling her internal language. There is nothing voluntary about this, no more than children decide voluntarily that it would be good to see in three dimensions. They just get on with it, growing their language system as their biology demands, developing the sounds they will use and the structures. Deaf children develop a gestural system and express their thoughts by that means. Much interesting work has been done on these early stages of acquisition, but this book will focus on what happens a bit later, in the extraordinary third year of life. In that year, syntax emerges rapidly and children develop into more or less full-fledged human beings. This holds regardless of whether the language users surrounding the child are speakers or signers, who express thoughts with equal richness.

Every parent has witnessed this apparent miracle. In that third year, children come to express and understand a wide range of thoughts in a language that they discover, select, and make their own, perhaps an individual, private form of what we call, say, Japanese or Javanese. Children experience and produce a finite number of expressions but they have the capacity to understand and use an infinite range. Once they can say something simple like *Heidi made that*, they might also understand similar structures like *Kirsten saw the movie there* and *Eric ate the cake when Alex drank OJ*. And so on, literally ad infinitum, in principle. Sometimes it is hard to make out the words but it gets easier as children fine-tune the sounds of their language and fine-tune the syntactic system in ways I will elaborate.

There is much for scientists to discover, many things that vary from one language user to another, but a central theme of this book is that, just as some birds are born to chirp, ants to follow a chemical trail back home, so humans are born to parse, born to assign fine-grained linguistic structures to what they experience. Everybody does it, normally at about the same age. These linguistic structures are drawn from a range of possibilities defined by invariant principles that have been discovered over recent decades and that are required to interpret particularities of the surrounding external language. Parsing is key.

There have been significant conceptual and technical shifts over past decades in our understanding of the abstract, INVARIANT principles of language, which are a function of what linguists call UNIVERSAL GRAMMAR (UG) and do not need to be learned (I use small caps for the first instance of major technical terms). Those shifts lead now to new approaches to parsing and to new analyses of VARIABLE properties. Variable properties show up in some but not all languages and do need to be learned; they differ from language to language, unlike invariant properties. We can take a new approach to language acquisition and there are good reasons to do so, because there are difficulties with our current notions of language variation and acquisition.

Children DISCOVER the structural CONTRASTS manifested by the variable properties in the ambient language, beginning with the contrasts among sounds or gestures on which their language system will be built. A little later, as they experience their ambient language, they begin to SELECT the first elements of their language system, which they need in order to understand what is said. Children parse with their emerging grammar. In parsing, they identify nouns, verbs, and other categories that may vary somewhat from one language user to another. They assign linguistic structures to the external language they hear, which TRIGGERS specific internal elements required for specific aspects of the parse. Children use what UG makes available, notably the binary-branching structures that emerge from recursive, bottom-up procedures called Project and Merge, plus what their emerging system, already partly formed, affords them. We used to call that system a grammar, but more recently we call it an I-LANGUAGE, *I* for internal and individual, emphasizing the fact that the internal system holds for an individual's brain and not for groups of people, like the group of all English speakers.

External language (E-LANGUAGE), on the other hand, is a very different kind of entity, language out there. It is what a child hears, and it is not structured, not discrete, nor represented in people's brains. E-language has no inherent structure but has structure assigned to it through parsing, after an initial I-language has begun to be triggered. If E-language shifts, for whatever reason, children may assign different structures, parse differently from earlier generations of language users, and thus attain a new grammar, a new I-language. What those new parses are and how they are selected by the innovative children provides information about acquisition and explains new variable properties emerging through new parses. Both E-language and I-languages play crucial, interacting roles: unstructured, amorphous E-language is parsed and an I-language system results.

There is no EVALUATION of I-languages and no PARAMETERS defined at UG. That, I will argue, has been an ill-chosen wild-goose chase. Rather, it is time to pursue a new vision of what variable properties are and how some are selected by young children, with UG open in ways I will describe. There is no miracle here; all children go through a similar kind of development, and we can achieve substantial scientific understanding of it.

We can have a more productive research paradigm if we abandon the search for parameters defined at UG, discard any evaluation metric, and dispense with positing an independent element of cognition known as a PARSER. That will represent a major simplification, reducing the machinery required for our theories and minimizing the information being attributed to our biology, along the lines sketched by proponents of the Minimalist Program. In eliminating UG-defined parameters, evaluation of grammars, and a distinct parser, I aim to make a substantial contribution to Minimalist ambitions.

Rather than seeing parsing as a processing "approximation" or "add-on" to the grammar, we will connect parsing more tightly to the grammar. This emphasis on integrated parsing echoes work by Bob Berwick, Janet Fodor, Heidi Getz, Marit Westergaard, Virginia Valian, Colin Phillips, and others. Phillips (2003a,b) works under the motto of "the grammar is the parser" and like him, I will dispense with an independent parser. Instead, I will focus on how children might use their internal language system to assign linguistic structure to what they hear, that is, to parse their ambient, external language. That enables them to select structures in their internal, private language. As a result, I adopt a new vision, at least within generative perspectives on acquisition: parsing is central to that vision, but it is implemented by the emerging language system, not by an independent parser.

There are some good case studies that point to a productive new paradigm, but we will also examine some difficult cases that show where new work is needed and how that new work should be conducted. We will focus on these difficult cases, on what is needed to improve analyses as our research paradigm is recast. Thinking in terms of parameters has led to much interesting work on variable properties, and that will need to be reconstrued in this new paradigm. Thinking through difficult cases where things remain to be discovered will be helpful for making the transition from one paradigm to another. I will be keen to show that under the new paradigm, we can have good empirical coverage as we make our radical simplifications.

Part of the pleasure of writing a book like this is to show people in related disciplines how linguists have come to analyze language acquisition and variation, in the hope that this will inform analyses of other areas of human cognition: memory, spatial cognition, emotion, and beyond. This book is written for philosophers, psychologists, neuroscientists, and linguists who see themselves as addressing questions of cognitive science broadly. A good case can be made that vision and language are two areas of human cognition where successful theories have been developed. Linguists have sometimes followed the lead of vision scientists, and understanding how languages may vary, how they are parsed by young children, may now cast light on the acquisition of other cognitive capacities and lead to deeper understanding. I shall include enough technical information to give the ideas real substance but shall use as little jargon as possible to make ideas accessible across disciplinary boundaries.

In putting together books of this kind, authors try out ideas in papers and lectures. Lectures include those given in classes at our home universities and some on the road. Readers familiar with my earlier work will recognize that the ideas here continue the commitment to examining "the logical problem of language acquisition" that I first formulated in *The Language Lottery* in 1982 and to "cue-based acquisition" from *The Development of Language* in 1999. Since then, ideas have changed significantly, as reflected in Lightfoot 2017b, as a result of the emphasis on parsing in a particular, grammar-based way. Those changes were first explored in lectures I gave at the Beijing Language and Culture University (BLCU) in 2015, where I learned from the lively linguistic community led by Fuzhen Susan Si. In fact, China has played a significant role, thanks to two visits generously organized by Ping Li, which also allowed me to try out ideas at the 2015 Brain Science meeting in Shenzhen and in 2018 at Jiangsu Normal University in Xuzhou, at Shanghai Jiao Tong University, and again at BLCU. I also presented these ideas in lectures at Newcastle University in the UK, including one as part of their Insights public-lecture series, at the Universities of Connecticut and Pennsylvania, and at the nineteenth Diachronic Generative Syntax meeting (DiGS 19) in Stellenbosch, South Africa. I am immensely grateful to those audiences and to the individuals who followed up afterwards.

There is nothing quite like being able to sustain a thorough investigation over the period of a fifteen-week course, and I am indebted to students in my Georgetown Diachronic Syntax class in spring 2019, who worked with a preliminary draft of this book. I am grateful to the three outstanding referees who advised MIT Press under the editorship of Marc Lowenthal; they all knew my work over a long period and understood how and why it had changed over recent decades, giving me much helpful advice about how to knit my story together and make it fit more coherently. John Whitman and Waltraud Paul helped me with the analysis of Chinese, and Heidi Getz proved to be an invigorating coauthor when we put our work together and came to understand the similarities in our analyses and the differences (Getz & Lightfoot to appear). Heidi and I share a view that linguists' reliance on parameters leads them to underestimate the richness of learning in the acquisition of language. Enriching the learning involved by basing our analysis of it on parsing offers a good alternative to analysis through parameters.

Psychologist Betty Tuller has become a partner in so many ways that it was natural for her to sharpen the ideas of the book and make it more readable and accessible to people outside of linguistics. Linguist Terje Lohndal started working on this approach to language analysis as a teenager and first wrote to me then, beginning a long-term, fruitful correspondence; both his work and the correspondence have helped to keep me honest in this book. However, these days publishers use advanced, digital facilities for their editorial work, which can be a challenge for those of a certain age. My standard response is to look for somebody under the age of thirty five. The young person who helped me navigate those challenges was Kate Kelso, who was masterful. I am enormously grateful to the wide-ranging intellects of Elan Dresher, Bill Idsardi, and Barbara Lust; they read the whole manuscript and gave rich commentary and much helpful advice.

It was a particular pleasure to be invited to sketch the ideas of this book at BLCU in 2018 at the inauguration of China's first department of linguistics, a major development for Chinese linguists, who build on hundreds of years of thinking about language. There I argued that studying language can be a good vehicle for teaching undergraduates about scientific investigation quite generally, as we did in the early days of the Department of Linguistics at the University of Maryland, which I helped to found in the early 1980s. Given the way our field has developed, researchers can lead students to generate productive, innovative findings early in their careers, without expensive equipment and without having to develop full command of a discipline like history, biology, or nuclear physics. Instead, students may take advantage of the laboratory in their heads, their knowledge of forms of Chinese in Beijing and of English in Maryland. The dynamism of Chinese academic life and the recent history of work

on Chinese syntax will bring exciting developments, and I would like to think that this book might assist those efforts by exploring new ways of thinking about language acquisition and variation.

Work on the book turned into a family affair: I was helped by my sister-in-law, Sue Lightfoot, who prepared the index, and my daughter, Heidi Lightfoot, Founding Director of Together Design, who worked with MIT Press to produce the cover design.

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).

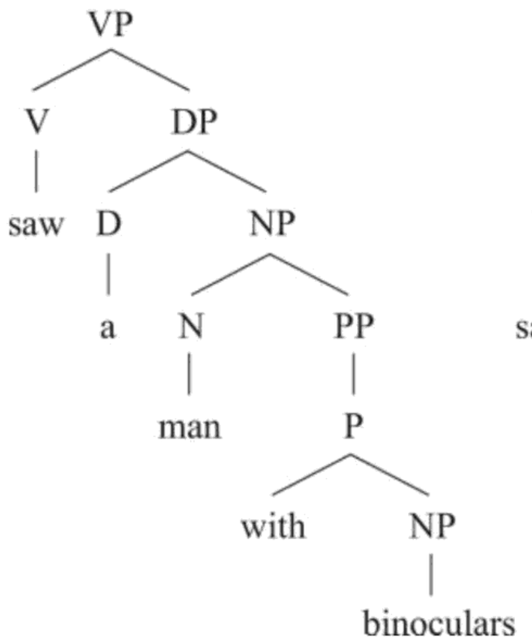
However, here's an important point showing the need for abstract structures in the parses that children arrive at: the word-for-word translation of **Who do you think that has telephoned?* does exist in a number of languages, as noted by Rizzi 1990: §2.6. This means that those non-English forms must 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 ~~chi~~ abbia telefonato ~~chi~~?*, where the embedded subject DP *chi*, 'who', is first copied to the post-VP position indicated by the second ~~*chi*~~ 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).

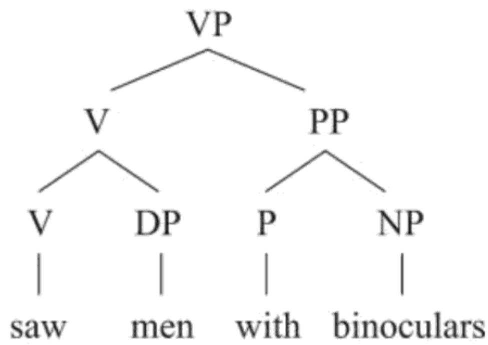
(3) a. This is the cow that kicked the dog that chased the cat that killed the rat that caught the mouse that nibbled the cheese that lay in the house that Jack built.
b. Ray said that Kay said that Jay thought that Fay said that Gay told ...
c. Ray and Kay went to the movie and Jay and Fay to the store, 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 $_{\text{VP}}e$.
b. *They denied reading it, although they had all $_{\text{VP}}e$.

(10) a. They denied reading it, although they often had $_{\text{VP}}e$.
b. *They denied reading it, although they had often $_{\text{VP}}e$.

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 → 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, $_{\text{VP}}[_{\text{V}}$*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: $_{\text{IP}}[_{\text{Infl}}$*will* $_{\text{VP}}[_{\text{V}}$*visit* $_{\text{N}}$*London*]]. Then the (pro)noun *you* can be merged with that IP to yield another IP: $_{\text{IP}}[_{\text{N}}$*you* $_{\text{IP}}[_{\text{Infl}}$*will* $_{\text{VP}}[_{\text{V}}$*visit* $_{\text{N}}$*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 → 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 predetermined 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 linguistic 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