Problems of Projection

There has been remarkable progress in understanding language in the post-
World War II period, over a very broad range. The questions that students are
raising and investigating today could not have been formulated, even
imagined, when I was a student shortly after the war. A concern among
students then was that although what we are doing is challenging, what will
happen after we have provided a structural analysis for every language by
applying the procedural methods of segmentation and classification that had
been developed, which may creak here and there but need nothing more than
technical repair?

The prevailing attitude more generally was that there are no real questions
about language: language does not exist as an independent object of serious
study. The most influential philosopher concerned with such issues, W.V.
Quine, held that a language is nothing but “a fabric of sentences variously
associated to one another and to nonverbal stimuli by the mechanism of
conditioned response,” hence a rather accidental construct depending on the
course of individual experience; and that if linguists can overcome
methodological “folly,” they will understand that no grammar of a corpus has
any greater validity than any other, much as no axiom system for arithmetic
has a unique claim to validity (Quine, 1960). Other prominent philosophers
and psychologists proposed that “general learning mechanisms” suffice to
account for language acquisition; there is no contribution of some dedicated
“language faculty.” Hence language does not exist as an object of serious
inquiry, beyond organizing data from diverse languages.

In all such proposals there is a question about how “sentences” (verbal
behavior, etc.) become a special subclass of human actions without some
presupposition about what language is.

Among professional linguists, rather similar attitudes were common. Many
adopted the view that languages can “differ from each other without limit and
in unpredictable ways” and therefore the study of each language must be
approached “without any preexistent scheme of what a language must be,”
Martin Joos’s formulation of what he called the “Boasian tradition” (Joos,
1951). The publication that served as the foundation of American structural
linguistics (Harris, 1951) was called Methods because there seemed to be
little to say about language beyond the procedures for reducing linguistic data
to organized form, in some manner depending on the purpose. Nikolai
Trubetzkoy’s classic introduction to phonological analysis (Trubetzkoy,
1935) was similar in general conception. Structuralist inquiries focused on phonology and morphology, the areas in which languages do appear to differ widely and in complex ways, a matter that raises interesting questions to which I will briefly return. It was recognized that there are at least some fixed properties of language, for example, choice of distinctive features, perhaps rooted in articulatory and acoustic properties. But the search for “God’s Truth” going beyond narrow limits was often ridiculed, as the term (Fred Householder’s) indicates.

In general biology, a rather similar picture was familiar, captured by molecular biologist Gunther Stent’s conclusion that the variability of organisms is so free as to constitute “a near infinitude of particulars which have to be sorted out case by case.”¹ That conception has since changed so radically that it is now possible to contemplate seriously the proposal that there might be a “Universal Genome that encodes all major developmental programs essential for various phyla of Metazoa” that emerged at the time of the Cambrian explosion 500 million years ago (Sherman, 2007). From this perspective there is only a single multicellular animal from an appropriately abstract point of view. Observed variety would be superficial, resulting from various arrangements of a “developmental-genetic toolkit” preserved through a long course of evolution.

Although the “variation with few limits” conception has virtually disappeared in general biology, the rather similar “non-existence” thesis remains very much alive with regard to language. It is widely held that “There are well-developed gradualist evolutionary arguments that language is entirely grounded in a constellation of cognitive capacities that each—taken separately—has other functions as well,”² in which case language exists only in the sense that today’s weather exists: a constellation of factors that have independent functions, not an object of serious scientific inquiry in itself. An influential view in language acquisition studies is that it is simply a matter of learning symbols used in communication – words, idioms, this sentence, all essentially on a par, acquired by means available to all primates, and a rather arbitrary collection; also finite apart from some unspecified methods of “induction” or “analogy.” (Tomasello, 2006). Computational cognitive

¹ Quoted by S. Carroll, 2005.
² N.J. Enfield, 2010. A source is cited, but whatever its interest, it barely touches on the question. Nor does any other.
science is dominated by such approaches to language. These studies have the merit of being clear enough to evaluate, and can easily be shown to be dramatic failures when they address questions about the nature of language.³

All of these approaches share the conception of half a century ago that there is no such thing as language in any serious sense, no genetic component to the language faculty – that is, no UG (“universal grammar”) in the technical sense that was proposed in what has sometimes been called “the generative enterprise,” adapting a traditional term to a new context (Huybregts and van Riemsdijk, 1982). This enterprise abandoned the procedural approach of structural linguistics and focused attention on a core property of language that had been largely neglected: each language determines an infinite array of hierarchically structured expressions that are transferred for interpretation to two interfaces: the sensorymotor system SM for externalization, and the conceptual-intentional system CI for thought and planning of action. In this sense, language is “sound with meaning,” in Aristotle’s common sense dictum. The unbounded range is an elementary fact that seems to have received little notice until the 17th century scientific revolution, where it played a significant role.

It follows that each language incorporates a generative procedure GP of some sort that yields the infinite array of internal expressions and provides the appropriate “instructions” for the interfaces, by means of its transfer mechanisms. UG sets conditions on what qualifies as a GP for some human language. We can think of a GP as itself a language in the sense of I-language: language understood as internal, individual, and intensional (the actual procedure of generation, not the class of expressions it generates).

Among the many notions of language, this one is central, in that all others that seek to capture core properties of language presuppose some version of it, at least tacitly.

This internalist approach to language, which began to take shape in the early post-war years, was embedded in the developing “biolinguistic framework” that regards the language faculty as a module of the organism, mostly the

³ Berwick et al (2011). Most of the work has to do with performance. There has of course never been any question about the role of statistical analysis and other cognitive processes in linguistic performance. See Chomsky, 1957, p. 17; Miller and Chomsky (1963). That is a different topic. The same is true of other domains. Study of the digestive system, for example, distinguishes investigation of its structural properties from how it performs under particular conditions – somewhat analogous to the competence-performance distinction in study of language.
brain, a subcomponent of a more complex system with enough internal integrity to be studied independently as part of a broader investigation of its interactions with other such systems in the life of the organism; analogous in this respect to the visual, immune, digestive, and other “physical organs,” and the “mental organs” of motor organization, planning, interpretation of social relations, various kinds of memory, and so on. The approach thus accords with what cognitive neuroscientist C. R. Gallistel describes as the biological norm quite generally: modular systems with special growth/learning mechanisms in different domains and in different species.4

There is every reason to expect human language to keep to the biological norm in this respect. There are in fact crucial features of human language that appear to be isolated in the biological world. They also seem to have emerged very recently in evolutionary time, many millions of years after the separation of modern humans from any other surviving species, among them the reliance on a GP to yield systems of digital infinity with mappings to the two interfaces for language-external interpretation.

Embedding the study of I-language in the biolinguistic framework is entirely natural; an individual’s language capacity is, after all, an internal property. It also unlocks the methodological fetters of the Quine-Joos et al. doctrines and their more recent descendants. It allows study of English to make use of discoveries about the nature of language unearthed in study of Italian or Mohawk; or inquiries that go beyond linguistic data narrowly construed: the brain sciences, language deficit, dissociation of linguistic from other cognitive capacities, language acquisition, genetics, and much more. To reject such options would be like insisting that a theory of E. coli must in principle ignore what is discovered about other organisms and about general biology, biochemistry, etc. No serious inquiry can be shackled in this way.

UG in the technical sense of the biolinguistic framework is not to be confused with descriptive generalizations about language such as Greenberg’s universals, a very important topic that has given rise to much valuable inquiry, but a different one.

4 Gallistel, 1998, 2000. Note that what is modular at one level of analysis (say, David Marr’s algorithmic level) may not be at a more abstract level (Marr’s computational level). Thus different systems of auditory analysis can serve the same general function (escaping predators), imprinting can proceed in many different ways, etc.; facts that have sometimes led to misunderstanding of modularity.
Putting aside genetic variation (an interesting but marginal phenomenon in the case of language) and conceivable but unknown epigenetic effects, the principles of UG, whatever they are, are invariant, and are typically not exhibited directly in observed phenomena, much as in other domains: for example, properties of the perceptual system that enter into identification of persistent objects or that interpret a few scattered stimuli as a rigid object in motion (Shimon Ullman’s rigidity principle). Far more generally, the essential art of science, revealed everywhere, is reduction of “complex visibles to simple invisibles,” as Nobel laureate in chemistry Jacques Perrin put the matter.

In contrast, descriptive generalizations should be expected to have exceptions, because many factors enter into the observed phenomena. Discovery of such exceptions is often a valuable stimulus for scientific research. To mention a classic case, the discovery of perturbations in the orbit of Uranus did not lead to the abandonment of Newton’s principles and Kepler’s laws, or to the broader conclusion that there are no physical laws, but to the postulation – later discovery – of another planet, Neptune. Exceptions to largely valid descriptive generalizations play a similar role quite generally in the sciences. Within the generative enterprise, the exceptional properties that have driven much important inquiry are sometimes purposely identified that way: Exceptional Case Marking (ECM), Extended Projection Principle (that is, exceptions to the Projection Principle), etc.

A sensible approach to the discovery of violation of generally valid principles is captured in Eric Reuland’s remark on early proposals about binding theory 30 years ago: “Too bad to be true, too good to be false” (Reuland, 2011). He and others proceeded to identify and explore various factors that enter into a more complete picture, maintaining much of the spirit of the original.

Binding theory provides useful lessons on taking principles and exceptions seriously. That is illustrated in one of the rare examples of an effort within the non-existence approach to address a significant property of language, binding theory (Chater and Christansen, 2010). Consider the sentence (1):

(1) do they expect to talk to each other?

Here they is the antecedent of each other. Chater and Christiansen propose that the anaphoric relation is simply “an instance of a general cognitive tendency to resolve ambiguities rapidly in linguistic and perceptual input.”
Hence the facts might rely on an innate constraint but not one that is specific to language.

Counterexamples have abounded in the literature for many years, and pursuing them has led to important insights. Thus if (1) is modified to (2), then the quickest way to find the anaphor is again to take “they” to be its antecedent, since John cannot be:

(2) do they expect John to talk to each other?

A variety of such cases show that it is necessary to determine what is a potential antecedent, a matter that quickly becomes complex. Consider for example (3) and (4):

(3) Who do they expect to talk to each other?

(4) (a) They gave instructions to John to talk to each other
    (b) They received instructions from John to talk to each other

The reason why (3) refutes the proposal is intuitively clear: there is an antecedent for each other closer to it than they, namely the unpronounced element in the position of John in (2) and the variable in the interpretation of (3) as (5):

(5) For which persons x, they expect persons x to talk to each other?

Investigation of these properties reaches far into core UG principles. Examples (4) lead to inquiry into the theory of control, and illustrate the ways in which semantic as well as structural properties determine the interpretation of the unpronounced element PRO that serves as the antecedent of the anaphor, in the position of John in (2) and the unpronounced element in (3).

While the Chater-Christiansen principle was known to be untenable as it stands, nevertheless there is an element of truth to it: namely, the role of minimal structural distance, a significant property of UG quite generally. And exploration of the counterexamples yields a very rich harvest and new insights into UG. That is quite often the case.

Early proposals about binding theory assumed that linear order was essential, as the more familiar examples seem to indicate. But work of Tanya Reinhart and others in the 1970s showed persuasively that hierarchy alone was involved in core cases. That suggests a much more far-reaching thesis with
many consequences: that order and other arrangements are a peripheral part of language, related solely to externalization at the SM interface, where of course they are necessary. We return to this question.

Though the non-existence thesis appears to have little or no support, there is something plausible about it, also shared by the generative enterprise since its origins: namely, the concern to reduce UG to the simplest form – to eliminate it entirely, in the non-existence approaches. Within the generative enterprise, the earliest efforts to address the problem of formulating GPs for language postulated rich mechanisms, hence a rich UG. That seemed necessary to attain at least a modicum of descriptive adequacy as soon as even the most elementary properties of language were considered. A main thrust of linguistic theory ever since has been to show that these mechanisms can be significantly reduced. There are several motives. The first is normal rational inquiry: the simpler the assumptions, the deeper the explanatory force; where possible, stipulation should be overcome. A motive more specific to language has to do with the hope for eventual insight into its evolution. The target of an account of language evolution is UG; that is what has evolved. The evidence is quite strong that since the departure of our ancestors from Africa about 50,000 years ago it has undergone no significant change, and not very long before (in evolutionary time) there is no evidence that it existed. The simpler UG, the greater the hope that its evolution – apparently in a narrow time frame -- might some day be at least partially understood.

The two approaches to simplifying UG fall within the general framework for studying any problem of growth and development. We can distinguish three factors in this process: (I) external data; (II) genetic endowment, which determines the general character of growth and development, and in the cognitive domains converts external data to experience; (III) organism-independent factors, including principles of natural law, which play a crucial role in development as in evolution: e.g., the laws of physics that determine that cells divide into spheres rather than cubes, and for computational systems like language, principles of computational efficiency that may well be reducible to laws of nature.

For language, we can distinguish three components of the second factor: (a) language-specific endowment, UG; (b) other cognitive processes; (c)

---

5 There are a few suggestions. Thus it has been claimed that likelihood of tone languages has a genetic correlate. Dediu and Ladd, 2007.
conditions imposed on language by the structure of the brain, though too little is known to draw far-reaching conclusions, despite interesting recent progress in neurolinguistics. It may be, as Gallistel and King have argued, that a fundamental reorientation of centuries of study of the brain will be necessary to discover the neural roots of the computational capacities not only of human language, but even of insects, where they are indeed astonishing.

The existence thesis of course does not preclude resort to other cognitive processes, and there are some interesting results integrating them with UG. The earliest step of language acquisition is extricating language-relevant data from the blooming, buzzing confusion, no trivial task, apparently relying on a human-specific component of UG. Infants also quickly acquire knowledge of the general prosodic structure of their language (in part pre-natally), and of the sound system generally, all dependent in large measure on UG, so it appears. Some early steps may also involve other cognitive processes. In my own work on this topic in the 1950’s, I took for granted that extraction of words from running discourse must be based on calculation of transitional probabilities, since there seemed to be no other relevant evidence. But the matter turns out to be more complex and more interesting. Recent work shows that this method fails, though results sharply improve when it interacts with apparent UG principles: six-month-old infants segment into elements with word-like properties when statistical properties are aligned with prosodically organized speech.\(^6\)

The non-existence thesis keeps to (IIb), and I think it is fair to say that it has been a failure in accounting for properties of language (as distinct from language performance; see note 7), beyond the most superficial cases. In contrast, pursuit of the existence thesis has registered many successes in reducing the complexity of UG. These inquiries have dealt with several fundamental features of language. There are, first of all, both contiguous and non-contiguous relations, the latter including the ubiquitous phenomenon of displacement – the fact that some phrase is interpreted both where it appears in surface forms and in some other position – and morphological discontinuity. Another property is linear order. Still another is identification of the category of a phrase (projection, more recently called “labeling”); this differs from the others in that it is a theory-internal notion, not virtually

detectable by direct inspection of expressions. The division of labor in the earliest work assigned contiguous relations, order, and projection to Phrase Structure Grammar PSG, and discontinuous relations to Transformational Grammar TG.

To review briefly some of the major steps in reducing the complexity of UG, the T-markers of TG were eliminated in favor of a recursive PSG, explaining why only some T-markers are found. PSG, with its extreme complexity and many stipulations, was eliminated in favor of X-bar theory. Stipulation of grammatical constructions (interrogative, passive, etc.), with their independent properties, was overcome (though used for expository purposes) by analyzing them into components that function generally, also eliminating redundancies (e.g., raising and passive in ECM constructions). X-bar theory was reduced to bare phrase structure. TG was radically simplified by discovery of general conditions on transformations. And finally the residues of PSG and TG were unified under the simplest combinatorial operation. In several of these moves questionable assumptions and open questions arise, to which I will return.

In recent years, work on these topics has often been called “the minimalist program (MP).” The term has been misunderstood. The program is simply a continuation of the efforts from the origins of the generative enterprise to reduce the postulated richness of UG, to discover its actual nature. The literature contains many criticisms of the MP, including alleged refutations, charges that it is not truly minimalist, and so on. None of this makes any sense. Research programs are useful or not, but they are not true or false. The program might be premature, it might be badly executed, but it is hard to see how it could be fundamentally misguided, since it hardly goes beyond holding that the study of language should keep to standard norms of science.

Within the MP, some new research programs have been proposed and pursued. One asks what an optimal solution would be to the conditions that must be satisfied by GPs for natural language: namely, satisfying interface conditions. Thus we can contemplate a Strong Minimalist Thesis SMT holding that language is a perfect solution to these conditions, then ask to what extent our best picture of actual GPs – I-languages – conforms to SMT; and where it does not, seek to show how the discrepancy can be overcome by

---

7 See, for example, Sauerland and Gärtner, 2007. The question mark in the title directs attention to the vast amount that is not understood.
a deeper analysis of the processes involved. The gap that resists such efforts is UG.\textsuperscript{8}

A separate development has had to do with language acquisition. Early approaches to UG assumed that it provided a general format for grammars, with infinitely many potential instantiations. An evaluation procedure based on “shorter grammar” (under notational transformations intended to capture valid linguistic generalizations) selected one or another grammar, given data (somewhat similar approaches were later developed within computer science). While that in principle provides a “discovery procedure” for grammars, it is completely unfeasible, one of the many problems of abduction. The Principles and Parameters (P&P) approach that took shape 30 years ago seeks to overcome this problem by restricting variety to a finite range. This conceptual shift opened the way to research of unprecedented scope and depth over a very wide typological range, revitalized psycholinguistic inquiry with highly productive exploration of acquisition and use of parametric values, and opened the way to new approaches to historical change.\textsuperscript{9}

Whatever approach to language acquisition one pursues, it will have to face these issues: either there is an infinity of options, in which case hopeless abductive problems arise if the task is taken seriously\textsuperscript{10}; or there is a finite number, and the approach falls within P&P. That leaves open many questions as to how parameters are set, and what role other cognitive processes might play in setting them.\textsuperscript{11}

Returning to the earliest days of the generative enterprise, it was quickly discovered, contrary to prevailing beliefs, that there were puzzles everywhere. Some of these are still very much alive. The period since has been

\textsuperscript{8} As in other domains, fundamental notions like “optimal” (“simple,” etc.) are not entirely clear in advance, though some cases are (e.g., less is better than more, local relations are simpler computationally than global ones). There is always an interplay between applying fundamental notions and sharpening them in terms of empirical consequences.

\textsuperscript{9} On the structure of the system of major parameters (macroparameters), see Baker, 2001. On determination of phylogenetic relations through investigation of parameter matching by methods of bioinformatics, see Bortolussi et al., 2011, and sources cited there. On the failure of efforts within the non-existence approach, see Longobardi and Roberts, 2011. For more general discussion of confusions and failures in the non-existence approach concerning universals, see Longobardi and Roberts, 2010.

\textsuperscript{10} Unless the potential variation is very narrowly restricted. For discussion, see Chomsky, 2010; Berwick and Chomsky, 2011.

\textsuperscript{11} See note 10.
reminiscent of early modern science. For millennia, scientists had been satisfied with quite simple explanations of natural events: a rock falls and steam rises because they seek their natural place, etc. When Galileo and others allowed themselves to be puzzled about these phenomena, and sought real explanations, modern science began (and common intuitive beliefs were quickly proven wrong). The capacity to be puzzled is a valuable one to cultivate, as the history of rational inquiry reveals.

One early puzzle, still alive, has to do with a simple but curious fact, never recognized to be problematic before, though it is. Consider the sentence (6):

(6) can eagles that fly swim?

The question is about ability to swim, not to fly. That is clear from interpretation, and is also shown by morphology, e.g., (7) but not (8):

(7) are eagles that fly swimming?
(8) are eagles that swimming fly"

Sentence (8) does not ask whether it is the case that eagles that are swimming fly. That is a fine thought, but it cannot be expressed by (8). Something about the design of language impedes communication by requiring a more complex circumlocution in this case, instead of allowing this option.

In (6), the association of the auxiliary with a verb satisfies the condition of minimal structural distance rather than minimal linear distance, which is a far simpler computational operation. The puzzle is why this should be so – not just for English but for every language in which the question arises, not just for this construction but for all others as well.

The problem of Aux(iliary)-Inversion has spawned a small industry in computational cognitive science, with many efforts to show how the right result can be attained by statistical analysis of corpora. New articles appear regularly, into 2011. The proposals have interesting features. Each one ignores the obvious explanation in terms of minimal structural distance. Each one that is clear enough to evaluate fails, dramatically (Berwick et al., 2011). Furthermore, even if they succeeded, it would not matter much. The basic problem would remain: why is this the case in all languages and constructions in which the issue arises? The same methods, or very similar ones, would work just as well for languages that rely on minimal linear distance; or that allow both, facilitating communication. And a background question is how
the child even knows what the intended interpretation is in such cases as (6) unless it is already relying on the minimal structural distance principle without any relevant data at all.

There is extensive linguistic evidence showing that the operation of minimal search – presumably a third factor principle (III) – makes use of structural rather than linear distance (apart from virtual adjacency, an unrelated property). There is also some supporting evidence from neuroscience, in studies of brain activity of subjects presented with two types of stimuli: roughly, invented sentences satisfying UG and others not conforming to UG; in the latter case, for example, negating a sentence by placing the negative element after the third word. It was found that in the former case there was normal activation in the language areas, though not when linear order was used. In that case the task was interpreted as a non-linguistic puzzle, so brain activity indicates (Musso, et al., 2003). This is a difficult but potentially rich research area.

The conclusion that core syntax-semantics relies on structural rather than linear distance fits well with what little we know about evolution, and what is known about language more generally. If it is correct, we derive one possible explanation for the fact that the child reflexively knows the right answer in the case of (6): linear order is simply not available to the language learner confronted with such examples, who is guided by a very simple principle of UG that restricts search to minimal structural distance in such cases.

The next question has to do with the generality of this UG condition. The best result would be that it is never available in I-language, apart from externalization to SM. That is the more far-reaching thesis I mentioned before in connection with binding theory: the thesis that order and other arrangements are a peripheral part of language, related solely to externalization at the SM interface, where they are clearly needed (in different ways, depending on the sensory modality used for externalization). The thesis is conceptually very natural, and has considerable empirical support, but also faces empirical problems\(^\text{12}\) -- a standard situation, indicating that as usual further inquiry is needed.

Assuming that the thesis can be upheld, either universally (the best outcome) or with some qualifications, then Aristotle’s dictum should be modified:

\(^{12}\) See e.g., Kayne, 2010.
language is not sound with meaning, but meaning with sound, a very different conception, with many consequences concerning cognitive architecture and its evolution. It would also follow that particular uses of language that depend on externalization, among them communication, are even more peripheral aspects of language, contrary to belief so widespread as to be virtual dogma, but without serious support.\footnote{13} It would also follow that the extensive literature of speculation on language evolution\footnote{14} is on the wrong track to start with. There is independent evidence for that conclusion, in part having to do with evolutionary considerations, but more significantly based on fundamental properties of language design.

I mentioned before that contiguous and non-contiguous relations, in earlier years distributed between phrase structure and transformational components, can be unified under the simplest computational operation, which is embedded in some manner in every relevant computational procedure: call it \textit{Merge}, an operation that takes objects X, Y already constructed and forms a new object Z. The third factor principle of minimal computation dictates that neither X nor Y is modified by \textit{Merge} (the so-called “No Tampering Condition”), and that they appear in Z unordered, the latter a plausible assumption for the reasons discussed. Hence \textit{Merge}(X,Y) = \{X,Y\}. Suppose neither X nor Y is part of the other, as in combining \textit{read} and \textit{books} to form the syntactic object \{X,Y\} corresponding to “read books.” Call that case \textit{External Merge EM}. Suppose that one is part of the other, say Y is part of X. Then the result of \textit{Merge} is again \{X, Y\}, but in this case with two copies of Y, one the original one remaining in X, the other the copy merged with X: call that \textit{Internal Merge IM}. Note that both operations come free: it would require stipulation to bar either of them. Furthermore, there are no operations “form copy” or “remerge,” just simple \textit{Merge}.\footnote{15}

\textit{IM} yields displacement – in fact, in a form appropriate for the CI system. Consider, for example, the sentence “which books did John read.” Here the

\footnote{13} It is worth noting that distinguished evolutionary biologists reject the dogma, among them Nobel laureates Salvador Luria and François Jacob, and neurologist Harry Jerison. See references of note 10.

\footnote{14} The speculations are almost entirely about communication, not language, a different topic: there is communication without language and language without communication (at least if the word “communication” is given some substantive meaning): internal dialogue, to take only the most obvious of many examples.

\footnote{15} \textit{EM} and \textit{IM} are the only possible cases if \textit{Merge} is binary, the simplest assumption. The concepts of multidimensionality, and some others, postulate a richer notion of \textit{Merge}.

13
phrase *which books* has two semantic roles: it receives its role as object of *read*, just as in the case of *read books*. And it is an interrogative operator, binding the variable in the object position, so that the interpretation is something like “for which books x, John read books x.” That is read off directly from the generated structure “which books did John read which books,” the product of IM, with no further comment. Example (5) is the same: it is formed by internal Merge of *who* to the full clause from the position of the unpronounced element that blocks the anaphoric relation.

The same holds in more complex cases. Consider for example the sentence (9a) derived by IM from (9b):

(9) (a) [which of his pictures] did they persuade the museum that [[every painter] likes best]?

(b) [which of his pictures] did they persuade the museum that [[every painter] likes [which of his pictures] best]?

In (9a), the phrase “which of his pictures” is understood to be the object of “likes”, analogous to “one of his pictures” in (10):

(10) they persuaded the museum that [[every painter] likes [one of his pictures] best]

Furthermore, the quantifier-variable relationship between *every* and *his* in (9a) is understood to be the same as that in (10): the answer can be “his first one” -- different for every painter, exactly as it is for one of the interpretations of (10). By contrast, no such answer is possible for the structurally similar (11), in which “his pictures” does not fall within the scope of “every painter”:

(11) [which of his pictures] persuaded the museum that [[every painter] likes flowers]?

Quite generally, IM yields appropriate structures for interpretation at CI. However, these are clearly the wrong structures for the SM system: universally in language, only the structurally prominent copy is pronounced (with an interesting class of exceptions that in fact support the principles involved\(^\text{16}\)). That follows from another application of Minimal Computation:

---

\(^{16}\) On this matter, see Trinh, 2011.
pronounce as little as possible. The result is that the articulated sentences have gaps, and as well-known in the study of perception and parsing, that yields difficult problems of interpretation.

The conclusion, then, is that if language is optimally designed, it will provide structures appropriate for semantic-pragmatic interpretation but that yield difficulties for perception (hence communication). There are many other kinds of examples that yield the same conclusion, among them structural ambiguities and garden path sentences. A case of particular interest is islands, as illustrated in the interrogatives corresponding to (12):

(12) they asked if the mechanics fixed the cars

We can ask “how many cars” and “how many mechanics,” (13) and (14) respectively:

(13) how many cars did they ask if the mechanics fixed?

(14) how many mechanics did they ask if fixed the cars?

Clearly the two differ sharply in status: (14), like (8), is a fine thought, but it has to be expressed by some circumlocution. In technical terminology, (14) is an ECP violation, a topic that has been quite extensively studied, including several revealing types of exceptions to the descriptive generalizations. As usual, apparent exceptions do not call for abandoning the generalization as far as it reaches, but for seeking deeper reasons to explain where and why it holds. One important early step was Luigi Rizzi’s discovery that null subject languages, in which counterparts to (14) are unproblematic, have a post-verbal position from which the subject can raise, not violating ECP (Rizzi, 1982).

In general, so it appears, ambiguous, garden path, and island structures result from free functioning of the simplest rules, yielding difficulties for perception (hence communication). Where there are conflicts between communicative and computational efficiency, it seems that the latter prevails, again supporting the revision of the dictum that traces back to Aristotle.

A computational procedure requires a collection of atoms, minimal elements that enter into computation though their parts do not. As a first

\[\text{-------------------------}\]

\(^{17}\) See note 7. The issues do not arise here, because the notions involved are scarcely controversial.
approximation we take the lexicon to be the class of atoms of computation, though this choice remains problematic in ways we see directly.

Another familiar concept in linguistic theory is *cyclicity*, based on the intuition that the properties of larger linguistic units depend on the properties of their parts. While ubiquitous in traditional practice, the concept perhaps received its first clear formulation and application in a 1956 paper on stress contours (Chomsky et al., 1956). At the time, one of the most lively topics in American structural linguistics was the study of stress and pitch levels, developed in its fullest form in a descriptive system that provided a four-stress, four-pitch notation designed to cover all English dialects (Trager and Smith, 1951). The reanalysis proposed that the stress contours are determined by a few simple rules operating cyclicly, in accord with syntactic structure, ideas since elaborated in many ways.

The cyclicity proposal was inconsistent with standard assumptions at the time, in particular the procedural approach to language mentioned earlier, which, for principled reasons, imposed a ban on “mixing of levels”: establishing “higher levels” such as syntactic or even word structure before “lower ones” are fixed (with marginal exceptions, which raised their own problems). That was no longer problematic within the generative enterprise. The cyclicity proposal also coexisted uneasily with PSG, given its “top-down” conception of generation. That tension was resolved by the abandonment of PSG, ultimately in favor of Merge.

Elementary considerations of computational complexity motivate a stronger notion of *strict cyclicity*, requiring that for certain syntactic objects SO, once generated they will not be modified by further computation, yielding considerable computational saving. These ideas have been developed in *phase theory*, which seeks to identify the SOs that are subject to this restriction in the optimal way, also restricting operations to the phase level, including the operation of transfer to the interfaces (the sole exception, by necessity, is EM, required to construct the phase in the first place). To achieve these outcomes it is necessary to distinguish the *interior* of a phase from its *edge*. If H is a phase head with complement Z, then Z is the interior of the phase; the edge is H along with anything merged to {H, Z} (Specifiers, in earlier terminology, now abandoned). It is the interior that is subject to no further modification. Elements of the edge – H and a sister of {H, Z} -- can be modified in the next higher phase; for example, they can raise, as in V-to-T or successive-cyclic A’-movement. While Z is immune from further changes, it does not disappear. If an SO containing Z is raised by IM, then Z will appear in the

---

18 See among others Chomsky, 2008.
surface position (as in “the man who said that Z was elected”). I will put further details aside here.

We have found good reasons to disarticulate the properties that were assigned to PSG in the earliest inquiries into I-language. Order might be a property of externalization induced by SM, leaving contiguous relations and projection. The former fall under Merge along with the non-contiguous property of displacement (EM and IM). There is no projection in the simplest Merge-based system. Let us look more closely at these components of language.

EM includes familiar cases of compositionality, with no projection or order. Under the simplest conception of Merge, EM(X,Y) yields {X,Y} independently of the character of X and Y. X-bar theory introduced a more complex notion: Merge is always to a head, which projects, yielding universal endocentricity and the concepts Specifier (SPEC, second Merge) and multiple SPEC. These notions are not definable under the simplest form of EM; and endocentricity, where it exists, derives from some other source.

The difference between PSG and X-bar theory is illustrated by the first basic rule proposed in PSG: S → NP VP. Projection as S is stipulated, and there is no hierarchic relation holding between NP and VP. In the X-bar-theoretic counterpart, this is revised in terms of Merge to head, as in the standard version (15), with complementizer C and inflectional element T:

(15) [C C [T NP [T T VP]]]

Here NP is SPEC-T, the second Merge to T. But that is just a stipulation. Within X-bar theory there is no more reason for NP to be SPEC-TP than for TP to be SPEC-NP; and simple Merge yields just (16), with no relation between NP and T, the T head of TP:

(16) [C C [a NP TP]]

In traditional grammar the priority of V in subject-predicate (NP-VP) structures is sometimes assumed, based on the fact that the semantic roles of subject and object are determined by V. A residue of that observation remains under simple Merge, but the reasoning does not apply to subject-

---

19 We put aside here the internal structure of Subject and Predicate, taking NP to be some nominal structure, perhaps DP or perhaps nP, where n is a functional element determining category.
predicate sentences if, as now generally assumed, the subject is a derived position, raised from object in unaccusatives/passives (“John arrived/was elected”) and from predicate-internal subject in unergatives (“John saw Bill”).

The PSG version of Merge is simpler, hence to be preferred for that reason alone unless there is convincing evidence to the contrary. It also has empirical support, as we see below. I will therefore assume that Merge applies in the simplest form: therefore there is no concept SPEC, structures need not be endocentric, and projection is a distinct property.

In (16), no relation is established between C and elements within α. However, there clearly is a C-T relation: that is shown directly by Aux-Inversion, but also by more complex theory-internal phenomena, such as inheritance of features of C by T. Since there is no notion of Specifier, minimal structural distance does not distinguish NP from TP. If we reduce Aux-inversion to the simpler principle that inversion depends on locality independent of category, then inversion could just as well yield “eagles [young are flying]” rather than “are [young eagles flying]” as the interrogative counterpart to “young eagles are flying.” These considerations suggest that the C-T relation must be established before the NP subject is introduced into (13), at which point LA will yield the label T for TP = {T XP}, establishing the C-T relation and permitting Aux-raising, feature-inheritance, and any other C-T relations.

We therefore have an independent argument for the Predicate-internal Subject hypothesis, which takes the basic structure of TP to be (17), with EA the external argument (missing for unaccusatives), IA the internal argument, and v a functional element that determines verbal category:

(17) T [β (EA) [v [V IA]]]

Note that it is still necessary to maintain the barrier against accessing linear order, so as to bar the elementary computational principle that disregards

---

20 On this matter, see among others Chomsky, 2008.

21 V may be simply an unspecified root, adopting ideas of Alec Marantz and Hagit Borer. We assume that v has two “flavors”: v for unaccusative/passive, v* for unergative. Note that insertion of EA into surface subject position in (13) is countercyclic, the complementizer C having already been introduced. For approaches to this problem see Chomsky, 2008 and Kitahara, 2011a (extending Epstein et al., in press; also talks since 2008).
structure altogether and selects the closest Aux once EA is raised to surface subject position.

Let us turn next to projection (labeling). There are many ways to describe how it might work, but we are interested in finding the most principled answer, the solution that accords best with SMT. The general course should, I think, proceed along the following lines.

Projection is a theory-internal notion, part of the computational process GP. We assume, then, that there is a fixed labeling algorithm LA, available for every GP. For a syntactic object SO to be interpreted, some information is necessary about it: what kind of object is it? Labeling is the process of providing that information. It follows that every SO that is to receive an interpretation must be labeled when it is transferred to the interfaces at the phase level. If an SO is to be interpreted but is not labelable by LA, then it must be modified so that it can be labeled, a fact with important consequences as we see directly. The idea that every SO must be labeled when it is formed is a residue of X-bar theory and its stipulations, and falls away under the simpler theory of compositionality. In the best case, the relevant information about SO will be provided by a single designated element within it; to first approximation a lexical item LI, a head. This LI should provide the label found by LA, when the algorithm can apply.

There are two ways in which the designated LI can provide the label: LI could itself be the label, or the most prominent features in LI could be the label detected by LA. The former is the conventional assumption, and I will keep to it for the moment, turning to the latter – which I suspect will prove correct – later on.

The optimal LA should conform to the overriding principle of minimal computation, and should therefore be nothing more than minimal search. Suppose Z = \{H, XP\}, where H (a head – containing no internal SOs) and XP does contain internal SOs. Then minimal search will uniquely detect H, which will be the label of Z. In any other case, minimal search in the simplest form will yield no result. Suppose we limit LA to this characterization, with no further complication.

Labeling of \{H, XP\} requires that H not be of the form \{X, Y\}, constructed by Merge. If, as we have so far assumed, the LIs of the lexicon are the atoms of computation, then they must not have this form. At this point questions arise about the nature of the lexicon: the criteria that determine what
constitutes an LI, and how these objects are constructed. Can the lexicon satisfy the condition that its elements are not constructed by Merge, yielding LI = \{X, Y\}? If, as sometimes proposed, φ-features can move independently of the complex in which they appear in an LI, then they are atoms of computation but not LIs. There are a variety of questions of this nature. For present purposes I will assume that the atoms are the LIs, begging many interesting questions, some of which arise below.

There are many other questions that arise about LA in the simplest form given above. Clearly a pronoun X can appear in a structure \{X, YP\}, as in S = “he left.” But it cannot be a head, or it would label S incorrectly. Therefore a pronoun must be a more complex structure, perhaps D-pro, as has been suggested. The same holds for nouns, as in “John left.” Therefore the noun must also be a complex structure, perhaps n-root, where the root is ambiguous as to category and n, v and other functional elements determine category.\(^22\) Similarly the determiner that (and others) must be morphologically complex, though not necessarily the definite article, raising questions about internal structure of nominal phrases. Similar questions arise about free relatives, as in “I like \[α what you wrote\],” where α is a nominal phrase (but not *“I like what book you wrote”). On the basis of these and other facts, Caterina Donati suggests that what and its counterparts are heads that can ambiguously project or be interpreted as Specifiers of the clause (Donati, 2006). In other languages, however, the head of the free relative is complex – e.g., Spanish lo quel or French ce qui, which may suggest that English what is similar, morphologically something like it-that, with it a reduced determiner that can project. If English who lacks such an analysis, that would account for the fact that it cannot head a free relative (*“I like who you invited”). Counterparts to who do however appear quite freely as heads of free relatives in other languages, sometimes with a complex form, as in Hebrew “kol mi” (literally “all of whom”). The complex form whatever can also head free relatives. In general, many questions about this construction remain open, and with them, questions about labeling.\(^23\)

Let us turn now to constructions \{XP, YP\} where neither is a head. These do not receive labels under LA. Andrea Moro argues that such structures require raising of one or the other term, his principle of “dynamic antisymmetry.”

\(^{22}\) See preceding note.

\(^{23}\) For much information about these structures, see Caponigro, 2003.
Thus he takes copular structures to be of the form [copula-small clause], where the small clause is of the form [XP-YP] (e.g., “[be [lightning, the cause of the fire]].” One of the terms of the small clause must raise, with interpretive consequences that he discusses.

This account can be reinterpreted in terms of labeling (see Chomsky, 2008). If, say, XP raises, then the result will be the structure (18):

(18) XP copula \{\beta XP, YP\}

Here the lower XP copy is invisible to LA, since it is part of a discontinuous element (however this notion is formally expressed). Therefore \(\beta\) will receive the label of YP. The phenomenon is similar to intervention effects, where the head of a chain (equivalently, the whole chain, the discontinuous element) induces an intervention effect, but not the lower copy.\(^{24}\)

Moro’s principle has many applications. One case is the “split topic” construction in German and many other languages, as in (19):

(19) Nagetiere hat Peter nur zwei Eichhörnchen gesehen (“rodents has Peter only two squirrels seen”)

Adopting the labeling reinterpretation of dynamic antisymmetry, Dennis Ott argues persuasively that (19) is derived by raising of “Nagetiere” from the small clause SC = \{zwei Eichhörnchen, Nagetiere\} (DP, NP), forced by the need to label SC along the lines just discussed for copular constructions; and shows that many complex properties of these constructions follow directly from this analysis (Ott, 2011).

Another case is the Predicate-internal subject construction (17), repeated here, identifying \(v\) as \(v^*\), with \(V\) (perhaps simply a root) raising to \(v^*\):

(17) T \[\beta (EA) [v^* [V IA]]\]

Since \(\beta\) is of the form \{XP, YP\}, it is not labeled by LA. As in the cases just mentioned, if EA raises to surface subject – SPEC-T, using the term SPEC now just for exposition – then \(\beta\) will be labeled \(v^*\), as required, since EA is part of a discontinuous element, hence invisible to LA.\(^{25}\) Therefore EPP is

\(^{24}\) Moro, 2000. On intervention effects and their subtleties, see Sigurdsson and Holmberg, 2008.

\(^{25}\) Note that this observation is non-trivial. It depends on the exact formulation of LA, in detail beyond what is given here.
forced in this case. Suppose EA remains in situ and IA raises to SPEC-T. Then the part of the structure visible to LA is EA-v*, with EA the “complement” of v*, and the structure is again labeled v*. Alexiadou and Anagnostopolou (2001) have proposed that in structures of the form β, either EA or IA must raise. If that thesis can be sustained, then the conclusion could simply follow from labeling.

Another case of lack of label is successive-cyclic movement. The intermediate steps are of the form \{α XP, CP\}, where XP can be for example a wh-phrase. The syntactic object α cannot be labeled, but it must be interpreted, if only to be theta-marked by the LI that selects it. If XP raises, then α will be labeled C, as required. Therefore XP must raise, and successive cyclic movement is forced. We see here again the empirical advantages of dropping the requirement that every syntactic object must be labeled, as in the more complex head-oriented concept of Merge.

Note that apart from the labeling problem it is not obvious why XP must raise in this case. English allows wh-in-situ constructions, as, for example, in asking questions in a quiz show, where the construction is preferred:

(20) they thought that JFK was assassinated in which Texas city?

That raises the question why (21) is impossible, with an analogous interpretation:

(21) they thought that [α in which Texas city [JFK was assassinated]]?

The fact that α has no label bars (21), as required.

Let’s turn next to indirect questions, as in (22):

(22) they wondered [α which books [β C [John read]]]

In this case too α is of the unlabelable form \{XP, YP\}, but XP does not raise. How then can it be labeled? The solution may lie in the fact that the most prominent feature of α and of β is shared, namely the interrogative feature Q, a feature of C and the head of α if we adopt a plausible analysis of

________________________

26 There are a number of empirical problems. One, noted by Dennis Ott (pc), is that in German β can remain unchanged, and can even raise. That leaves open the question how German (and perhaps Japanese) differ from English-like languages, perhaps because of some difference in the internal structure of the verb phrase, perhaps because of options in fixing LA (see preceding note), perhaps in some other way.. As noted, Ott shows that the Moro-based analysis does hold for German in other constructions.
interrogative *wh*-phrases that takes Q to be the most prominent element (Cable, 2007, 2010; Narita, 2011). Searching \{XP, YP\}, then, LA finds the same most prominent element -- Q -- in both terms, and can take that to be the label of α.

With this in mind, let’s return to the (not so simple) Subject-Predicate construction (16), now understanding the subject NP to be raised from EA or IA of the verbal phrase:

(16) \[ c \ C \ [\alpha \ NP \ TP] \]

The phrase α has certain restrictions. Thus it does not raise (apart from the context of Richard Kayne’s influential LCA framework), but it is not clear that this property relates to labeling. Does it receive a semantic interpretation? That seems to be the case. Thus the surface subject has a kind of secondary agency, as we see in such sentences as “PRO to seem to be intelligent is not as easy as you might think.” It also bears what Luigi Rizzi calls an “aboutness” property, illustrated in null subject languages by the fact that silent *pro* can refer to the surface subject (but not the object) of the preceding sentence (Rizzi, 2010). Such considerations suggest that it receives its own interpretation and should be labeled. Perhaps that can be achieved by the same device suggested for embedded interrogatives. NP and TP share prominent features, namely φ-features – so-called “SPEC-Head agreement.”

For this approach to be tenable, it must be that labeling seeks features, not only LIs – or perhaps seeks only features, in which case it would be similar to probe-goal relations generally, specifically Agree. That seems quite natural, though the implications remain to be explored. It must also be that only certain features can serve as labels; thus we do not want a small clause \{DP, DP\} to be labeled D. But there is independent evidence that this is the case.

Coordination provides motivation for this conclusion. We have to distinguish structured from unstructured coordination, the latter as in “John is tall, happy, hungry, bored with TV,….” In this case each AP is predicated individually of the subject; the APs are not predicated as a unit. Such structures as these have been a problem since the early days of generative grammar, since they seem to require infinitely many rules to avoid unwanted structure. A natural

\[\text{\footnotesize Rizzi, 2010. For similar ideas, see Epstein, 1992.}\]

23
approach is to invoke Pair-Merge, differing from simple Merge in that it forms a pair, not a set, inducing the asymmetry characteristic of adjuncts. It can be applied indefinitely often, adding individual predications, and avoiding the problems of incorporation into PSG or TG.

Structured coordination is different. Very likely the underlying structure of [Z and W] is (23a). To label β, one of Z or W must raise, say Z, yielding (b):

$$(23)(a) \left[ \alpha \text{Conj} [\beta Z W] \right]$$

$$(b) \left[ \gamma Z \left[ \alpha \text{Conj} [\beta Z W] \right] \right]$$

Now β receives the label of W, but what about γ? It is, again, an \{XP, YP\} structure, hence unlabelable. But it needs a label. We know what the right answer is: the label is not Conj, the most prominent head, but rather the label of Z, shared typically with W; if the coordinated expressions are APs, then γ is an AP, etc. It follows that Conj and the construction α that it heads are not visible to LA, so that γ receives the label of Z.

Resort to Pair-Merge may well be appropriate in other circumstances. Recall the Alexiadou-Anagnostopolou thesis that implies that if IA raises, EA can remain in situ. That will not follow from the suggested labeling analysis if a residue remains in VP, as for example, in such constructions as “put the book on the table.” A solution might be to adopt some version of proposals that go back to the 1950s, which interpret “put-on the table” as a complex verb with “the book” as its object. The complex verb itself might be understood as the Pair-Merged SO <put, on-the-table>, consisting of a verb and an adjunct (with possibly multiple unstructured adjuncts, as in “put on the table in a box…”). In that case the normal labeling procedure for [v [V, IA]] will apply.

Another long-standing problem has to do with head-head constructions, the first step in a derivation. If the Marantz-Borer conception is adopted, as suggested earlier, these will be of the form \textit{f-root}, where \textit{f} is one of the functional elements determining category. Suppose that root, like
conjunction, does not qualify as a label. In that case these constructions will be labeled \( f \), as intended, because no other element is visible to LA.

Which features qualify as potential labels? It seems that the set should at least include those that function as probes: \( Q \) and phi-features. And as we have just seen, the functional elements that determine lexical category. More extensive study is required to sharpen and motivate the class of potential labels and to explore the implications of feature-based LA generally.

Interesting questions arise with regard to other complex constructions, such as ECM, as in (24):

(24) they consider \([\alpha\text{John to be intelligent}]\)

Work going back Paul Postal’s investigations of “raising to object” 40 years ago, reformulated in a contemporary framework by Howard Lasnik and Mamoru Saito (based on ideas of Masatoshi Koizumi), has shown convincingly that \( John \) is actually in the higher clause in surface structure. In current analyses, that is assumed to be the result of raising of \( John \) to sister-of-[\( consider \alpha \)] (Specifier, in early analyses), followed by raising of \( consider \) to \( v^* \), which restores the base order but with different constituency, a rather surprising puzzle, variously addressed.\(^30\) Raising of \( John \) to sister-of-[\( consider \alpha \)] can be attributed to the now familiar labeling problem, permitting \( \alpha \) to be interpreted. The resulting structure is still unlabelable, but the effect of obligatory raising of \( consider \) to \( v^* \) perhaps overcomes the problem since the resulting structure is labeled \( v^* \), and for interpretability that might suffice.

Another property of sister-of-T (SPEC-T, informally), mentioned earlier, is the island condition ECP, illustrated in (12)-(14). As shown by Marc Richards (Richards, 2007), there is good reason to suppose that the \( \phi \)-features of \( T \) are in fact inherited from \( C \); and though his argument doesn’t extend to tense, the tense feature as well. Earlier we noted that the system is simplified if features of an LI cannot themselves be atoms of computation. That would entail that all the features of \( C \) should be inherited by \( T \), including \( Q \). If so, then in (12), spelled out more fully here with the \( wh \)-phrase subject, the \( Q \) feature of \( C \) should appear in \( T \):

\(^30\) For discussion and sources, see. Lasnik, 2003. These approaches assume raising to AGR-O, not \( V \), but it is not clear that the extra assumptions are necessary.
(25) they asked [if-Q [[how many mechanics] [T fix the cars]]]

The surface wh-phrase subject is therefore already in its criterial position, in Rizzi’s sense (see note 27), sister-of-QP. Its Q-feature is matched by the sister phrase QP, and it therefore need not – in fact cannot – raise any further. That leaves an escape hatch for a wh-phrase object, permitting (14) while blocking (13). This suggests in outline what might be a principled account of the core cases of ECP, explaining why the thought is fine but cannot be expressed, as discussed earlier.31

As is well known, the paradigm of (13)-(14) extends to (13’)-(14’), the “that-trace” effect:

(13’) How many cars did they say that the mechanics fixed ___?
(14’) * How many mechanics did they say that ___ fixed the cars?

In a null subject language like Italian, all of these cases are fine,32 strongly suggesting that the phenomena should be unified – an outcome that has long been sought. Furthermore, if that is dropped in (14’), the sentence is fine. And it is of course necessary to explain how speakers of English33 know these facts without evidence.

The preceding comments suggest some possibilities. Suppose that criterial freezing in Rizzi’s sense, blocking IM, is induced by a feature of T inherited from C: perhaps a force feature F, subsuming Q as a special case. If F is understood to match both the Q-feature of the subject of (14) and the φ-features of the subject of (14’), then both cases will be subject to criterial freezing. Deletion of that in (14’) might leave only a weakened form of C (perhaps FIN of Rizzi’s left periphery),

31 The approach is reminiscent of the Vacuous Movement Hypothesis of Chomsky, 1986, but avoids the special stipulations required there. Note that inheritance has to be understood as copying, leaving Q in its original position for selection and labeling. The basic argument for inheritance due to Richards is that if valued uninterpretable features remain at the phase head position, the derivation will crash at the next phase. That is not a problem here, because Q is interpretable. For φ-features it may mean that they are deleted or given a phonetic form (as in West Flemish), hence invisible at the next phase.

32 Rizzi, 1982. The following comments are based on suggestions of Luigi Rizzi’s, pc.

33 There are dialects in which (14’) is acceptable, presumably because learners have positive evidence indicating that the operative principle does not hold. Learners might then adopt a subject-extraction strategy discussed by Rizzi and Shlonsky (2007).
lacking F. Under these (unfortunately still somewhat stipulative) assumptions, the unification and variations would follow.

The cases discussed leave many questions open, and there remain a substantial array of others that require analysis and explanation. But the material reviewed should indicate the value of seeking principled explanations in accord with the minimalist program and its antecedents since the 1950s. And also the value of taking puzzles seriously, including apparent exceptions to generally valid descriptive generalizations, and accounts that seem intuitively obvious but remain unexplained in principled ways, familiar features of the sciences particularly since their modern origins.