Problems of projection

Noam Chomsky [*]

MIT – Department of Linguistics and Philosophy, 77 Massachusetts Avenue, Bldg. 32-D808, Cambridge, MA 02139, USA

Abstract

With the crystallization of the ''generative enterprise'' half a century ago, two concepts became salient: the initial state and final states of the language faculty, respectively, UG (the genetic component) and I-languages. Since then inquiry has gained far greater scope and depth. It has also led to sharpening of fundamental principles of language. At first, descriptive adequacy appeared to require rich and complex assumptions about UG. A primary goal has always been to overcome this deficiency. Core properties of concern have included compositionality, order, projection (labeling), and displacement. Early work assigned the first three to phrase structure rules and the last to the transformational component. Simplification of computational procedures suggests that compositionality and displacement (along with the ''copy theory'') fall together while order may be a reflex of sensorimotor externalization, conclusions that have far-reaching consequences. As for labeling, minimal computation restricts options to the few that have considerable empirical support.
© 2012 Elsevier B.V. All rights reserved.

Keywords: Syntax; Minimal computation; Merge; Labeling algorithm; Minimal search

There has been remarkable progress in understanding language in the post-World War II period, over a very broad range, including the general principles that shape this highly special cognitive faculty, dissociated from others in many ways and unique to humans in essentials.[1] One indication is that it was routine and reasonable for prominent linguists in earlier years to write books entitled *Language*. No longer. The task would be radically different today; far too much has been learned. The questions that students are raising and investigating today, empirical and theoretical, could not have been formulated, even imagined, when I was a student shortly after World War II. 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 analysis of texts 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 elsewhere, as an infinite set of well-formed sentences – a notion that if coherent at all, which is dubious, is derivative, presupposing some mode of generation, which Quine held it would be ''folly'' to postulate. 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 inquiry into language reduces to

---

* Corresponding author.
 E-mail address: chomsky@mit.edu.

[1] For review of dissociations, a crucial topic for study of language and other cognitive systems, see Curtiss (2013).

organizing data from diverse languages, refining techniques of analyzing data, and finding some largely valid generalizations – and it even remains unclear why certain actions, not others, belong to language.

Among linguists, rather similar attitudes were common. In the early 20th century classics, Ferdinand de Saussure took language to be a "the sum of word-images in the minds of all individuals. . .of a given community," which is "not complete in any speaker; it exists perfectly only within a collectivity. . .[as] a well-defined object in the heterogeneous mass of speech facts." Leonard Bloomfield defined *language* as "The totality of utterances made in a speech community." For Edward Sapir, language is "a purely human and non-instinctive method of communicating ideas, emotions, and desires by means of a system of voluntarily produced symbols." Evidently, inquiry into other questions about language (or any other system) – in particular its acquisition, use, evolution, neural representation – can proceed only insofar as we have some relatively clear idea about the nature of the object of inquiry. Accordingly, such inquiries were necessarily limited.

By mid-century, 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, 1957). 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. 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, discussed elsewhere. 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 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."[2] 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 half a billion 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 in essentials through a long course of evolution. Deeply conserved properties and factors that constrain variation have taken on a far more significant role in modern evolutionary biology.

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," 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.[3] 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 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.[4]

All of these approaches share the conception of half a century ago that there is no such thing as language in any serious sense. More accurately, while no one can rationally question that there is a genetic component to the language faculty – that is, UG ("universal grammar") in the technical sense of what has been called "the generative enterprise" (Huybregts and van Riemsdijk, 1982), adapting the traditional term "universal grammar" to a new context – nevertheless, under the "non-existence" assumptions, the approaches to UG that have so far led to serious results would have to be reshaped. How, it is idle to speculate without credible proposals.

---

[2] Quoted by Carroll (2005). "Near infinitude" has the standard informal meaning "very large," as when Darwin wrote that "man differs [from animals] solely in his almost infinitely larger power of associating the most diversified sound and ideas." It is likely, however, that he appreciated the observations of his associate and expositor Thomas Huxley on the "predetermined lines of modification" that lead natural selection to "produce varieties of a limited number and kind" for every species.

[3] Enfield (2009). A source is cited, but whatever its interest, it barely touches on Enfield's claim. Nor does any other, to my knowledge. Note that if something like this turned out to be true, it would have no direct bearing on the richness of universal grammar (UG), which would then have to account for the species-specific ways in which these capacities form a "constellation" in humans, dissociated from their other functions (see note 1).

[4] 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, 17) and Miller and Chomsky (1963). But that is plainly a very different topic.

The ''generative enterprise'' abandoned the procedural approach of structural linguistics and sought instead to develop a concept of language concrete enough to guide the study of its essential properties as well as the related inquiries that necessarily rely on such a conception. The first step is to focus attention on a core property of language that had been largely neglected: each language incorporates a mechanism that 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 (broadly understood).[5] 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 characterizes the internal expressions and provides the appropriate ''instructions'' for the interfaces, by means of its transfer mechanisms. UG determines 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 structured objects it generates). Among the many notions of language, this one is central, in that others 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 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 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 generally: modular systems with special growth/learning mechanisms in different domains and in different species (Gallistel, 1998, 1999).

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 behaviorist, procedural, and ''Boasian'' traditions and their more recent descendants. It allows study of Japanese or English to make use of discoveries about the nature of language unearthed in study of Italian or Mohawk, and of inquiries that go beyond linguistic data narrowly construed: the brain sciences, deficit, dissociation, acquisition, genetics, and much more. To reject such options would be like insisting that the study of *Escherichia coli* must in principle ignore what is discovered about other organisms and about general biology, biochemistry, etc. No serious inquiry should be shackled in this way.

UG in the technical sense of the generative enterprise is not to be confused with descriptive generalizations about language such as Joseph Greenberg's universals, a very important topic that has given rise to much valuable inquiry, but a different one.[6]

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 physics Jean Baptiste 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, as a stimulus to further inquiry: Exceptional Case Marking (ECM), Extended Projection Principle (that is, exceptions to the Projection Principle), etc.

---

[5] It is theoretically possible, though unlikely, that some language might be so impoverished in lexical and other resources that only a finite number of non-deviant sentences could be generated by its GP. If so, it would be a minor curiosity, with no bearing on UG, acquisition, or other significant issues, contrary to much media confusion. See Chomsky (forthcoming) for some comment.

[6] See Baker (2011), for useful discussion.

A sensible approach to the discovery of violation of generally valid observations 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 tentative 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 Christiansen, 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 this anaphoric relation is simply "an instance of a general cognitive tendency to resolve ambiguities rapidly in linguistic and perceptual input."

Counterexamples abound in the literature, 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?] (*, with *they* as antecedent)

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

The bracketed part of (3) is identical to (1), though the anaphoric relation of (1) is blocked in (3), thus refuting the proposal. The reason is intuitively clear: there is an antecedent for *each other* closer to it than *they*, a "simple invisible" in Perrin's sense: namely the unpronounced element in the position of *John* in (2) and the variable in the interpretation of (3) as (5), present in the mind but not in the external output:

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

Investigation of these properties reaches far into core and invariant UG principles.

Examples (4) lead to inquiry into the theory of control, and illustrate the ways in which semantic and structural properties interweave in determining 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 had long been known to be untenable as it stands, nevertheless there is an important element of truth to it: namely, the role of minimal structural distance, a significant property of UG quite generally, and a puzzling one in many ways. 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 is essential, as simple examples like (1)–(4) seem to indicate. But work of Tanya Reinhart and others in the 1970s showed persuasively that hierarchy alone was involved in core cases, hence structural rather than linear distance. That suggests a much more far-reaching thesis with many consequences:

(T) Order and other arrangements are a peripheral part of language, related solely to externalization at the SM interface, where of course they are necessary.

If (T) holds generally as a principle of UG, then Aristotle's dictum should be modified: language is not sound with meaning, but rather meaning with sound (or some other externalization), a very different concept, reflecting a different traditional idea: that language is primarily an instrument of thought – "audible thinking," "the spoken instrumentality of thought," as William Dwight Whitney expressed the traditional conception. The shift of perspective, to which we return, has many consequences concerning cognitive architecture and its evolution.

Though the "non-existence" thesis appears to have little or no support, and it is not even easy to give a coherent account of it, nevertheless there is something plausible about the guiding intuition, shared by the generative enterprise since its origins: namely, the concern to reduce UG to the simplest form. The earliest efforts to formulate GPs for language postulated rich mechanisms, hence a rich UG. That seemed necessary to attain at least a modicum of descriptive adequacy. 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. It would be idle to study the evolution of the eye without a fairly clear conception of what an eye is, or to study the evolution of language without a fairly clear conception of UG.[7] 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,[8] and not very long before (in evolutionary time) there is no evidence that it existed. The simpler UG, the greater the hope that evolution of language – apparently within a narrow time frame – might some day be at least partially understood.

In the study of any problem of growth and development – language acquisition in particular – we can distinguish three factors: (I) genetic endowment; (II) external data; (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. Genetic endowment determines the general character of growth and development, and in the cognitive domains converts external data to experience.

For language, alongside of UG and whatever other cognitive systems may be relevant, the first factor includes conditions imposed on language by the structure of the brain. About these, too little is known today to draw far-reaching conclusions, despite interesting recent progress in neurolinguistics. There are unique difficulties in proceeding further in the case of language. A great deal is known about the neural basis for the human visual system by invasive experimentation with other species that have very similar systems. But that is impossible for language because of the biological isolation of the human language faculty. There are more subtle reasons having to do with the study of internal cognitive systems, not input systems (such as vision). It also may be, as Gallistel and King (2009) 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 also 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.[9] 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 1950s, 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.[10]

Inquiries into UG have to deal 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 a phrase is interpreted both where it appears in surface forms and in some other position – and morphological discontinuity (including agreement). Another property is linear order. Still another is identification of the category of a phrase (projection, more recently called ''labeling''). The division of labor in the earliest work assigned contiguous relations, order, and projection to Phrase Structure Grammar (PSG), and non-contiguous relations to Transformational Grammar (TG).

Projection/labeling differs from the other notions in that it is not virtually detectable by direct inspection of expressions, but rather is a theory-internal notion, taken over from traditional and structuralist grammar. For that reason it requires even more careful analysis than the others, raising questions that have not been adequately addressed, to my knowledge, despite the near-universal resort to the notion. Its special status becomes clearer as we proceed.

To review briefly some of the major steps in reducing the complexity of UG, 50 years ago T-markers of TG were eliminated in favor of a recursive PSG, and a few years later PSG, with its extreme complexity and many stipulations,

---

[7] Lacking that, speculations about evolution have to do with behavior involving language, say evolution of communication, which is quite different from evolution of language, the internal system that is used for communication and much else. The phrase ''evolution of language'' is shorthand for evolution of the language capacity of language users. Languages change, but they do not evolve, and linguistic change should not be confused with evolution.

[8] There are a few suggestions to the contrary. Thus it has been claimed that likelihood of tone languages has a genetic correlate (Dediu, 2010, 2011). There are, however, questions about the significance of evidence of the kind adduced, including the role of the specific genes allegedly implicated (Nielsen, 2009). A test of linguistic significance would be effects on acquisition, if any.

[9] See Petitto et al. (2012).

[10] Shukla et al. (2011). This extends work of Yang (2004). See Yang (2002), for discussion of the interplay of UG and probabilistic learning theories in language acquisition. On phonotactic evidence for segmentation, see Daland and Pierrehumbert (2011).

was eliminated in favor of X-bar theory, and later bare phrase structure. 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). From the early 1960s, TG was radically simplified by discovery of general conditions on transformations. Postulation of internal levels (D- and S-structure) and the redundant cycles associated with them was later eliminated. And finally the residues of PSG and TG were unified under the simplest combinatorial operation. At each stage, options of variation were sharply reduced, and accordingly explanatory success was improved. In several cases the simplifications involve questionable assumptions and raise open questions, to some of which we 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 (see Freidin and Vergnaud, 2001). 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 such program begins by asking what an optimal solution would be to the conditions that must be satisfied by GPs for natural language: namely, satisfying interface conditions.[11] 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 a deeper analysis of the processes involved. The gap that resists such efforts includes UG.[12]

A parallel and interacting development has had to do with language acquisition. Early approaches to UG assumed that it provided a general format for grammars (I-languages), with infinitely many potential instantiations. An evaluation procedure based on ''shorter grammar'' (under notational transformations designed to capture valid linguistic generalizations) selected one or another grammar, given data (somewhat similar approaches were later developed within computer science). While that approach in principle can provide 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.[13]

Whatever approach to language acquisition one pursues, it will have to face these issues: either there is an infinity of options, in which case challenging and perhaps hopeless abductive problems arise if the task is taken seriously[14]; or there is a finite number, and the approach falls in principle 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.[15]

Returning to the earliest days of the generative enterprise, it was quickly discovered, contrary to prevailing beliefs, that there are puzzles everywhere, some of them still very much alive. The period then and since has been 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 sense beliefs were quickly proven wrong). The capacity to be puzzled is a valuable one to cultivate, as the history of rational inquiry reveals.

---

[11] The goal is captured in the title of Sauerland and Gärtner (2007), the question mark directing attention to the vast amount that is not understood.

[12] As in other domains, fundamental notions like ''optimal'' (''simple,'' etc.) are not entirely clear in advance, though some subcases are (e.g., less is better than more, local relations are simpler computationally than global ones). There is commonly an interplay between applying fundamental notions and sharpening them in terms of empirical consequences.

[13] On the structure of the system of major parameters (macroparameters), see Baker (2001). For important discussion of parametrization and comparative grammar, see Kayne (2012), whose remarkable work has provided much of the basis for the flourishing of the field in recent years. 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 similar 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).

[14] Unless the potential variation is very narrowly restricted. For discussion, see Chomsky (2010) and Berwick and Chomsky (2011a,b).

[15] See note 10.

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 sentences (6):

(6) (a) can eagles that fly swim?
     (b) instinctively, eagles that fly swim

In (b), the adverb is associated with *swim*, not *fly.* Similarly, in (a), the question is about ability to swim, not to fly. In both cases, that is clear from the 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. As in the case of the missing interpretations of (6a,b), 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 both (6a,b), the association of the clause-initial element to the verb is remote rather than proximal. Generalizing to other cases, it satisfies the condition of minimal structural rather than minimal linear distance, though the latter 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 these constructions but for all others as well.[16]

Case (6a), 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, to the present. 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. And for such cases as (6b), the proposals (relying on statistical analysis of corpora) are in principle incapable of yielding the result. A general 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 languages 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 (Musso et al., 2003, adopting the paradigm of Smith and Tsimpli, 1996). It was found that in the former case there is 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. 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 and origins of this condition. The best result would be that linear order is never available for computation in I-language, apart from externalization to SM. That is the more far-reaching thesis (T) mentioned earlier in connection with binding theory: the thesis that order and other arrangements are a peripheral part of language, reflexes of SM interface properties (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[17] – 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, we return to a traditional conception of language as ''an instrument of thought,'' as noted earlier. It would follow that particular uses of language that depend on externalization, among them communication, are even more peripheral aspects of language,

---

[16] There are many other cases. The basic facts have been familiar since the earliest investigation of islands in the 1960s, particularly in John Ross's fundamental work (1967, 1986).

[17] See e.g., Kayne (2010).

contrary to belief so widespread as to be virtual dogma, but without serious support.[18] It would also follow that the much of the extensive speculation about language evolution[19] is on the wrong track to begin 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 earlier that contiguous and non-contiguous relations, in earlier years distributed between PSG and TG, can be unified – namely, under the simplest computational operation, embedded in some manner in every relevant computational procedure, call it *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 Merge (the "No Tampering Condition"), and that they appear in Z unordered, the latter a plausible assumption for reasons already discussed. Hence, Merge(X,Y) = {X,Y}. Suppose neither X nor Y is part of the other, as in combining *read* and *books* to form the syntactic object {X,Y} corresponding to "read books." Call that case *External Merge EM*. Suppose that one is part of the other, say Y is part of X. Then the result of 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 *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 Merge.[20]

Clearly it is necessary to distinguish copies from repetitions. Thus in (i) "what hit what?," the two occurrences of *what* are repetitions, but in (ii) "what was hit what?," they are copies. (i) is analogous to the possible answer: "the stone hit the wall." The two occurrences of *what* are independent, just as the two occurrences of *every* are independent in "every boy read every book" (or the two occurrences of "the stone" in "the stone hit the stone"). In contrast, (ii) derives from "was hit what" by IM, raising "what" to surface subject position, externalized as "what was hit?"

The intuitive basis for the distinction is clear, in more complex cases as well: it reduces to independent extraction from the lexicon. There is a simple mechanism to make the distinction if operations apply simultaneously, as in the phase-theoretic systems mentioned below, where IM and transfer both apply at the phase level.

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 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. The same (simplest) assumptions account for example (5): 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 from (9b) by deletion of the lower copy:

(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", as in (9b), 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). In 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]?

The results follow straightforwardly from IM, read off from (9b), as in the simpler examples already discussed.

---

[18] 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 14.

[19] As noted, 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 dialog, to take only the most obvious example – an interesting topic, barely investigated.

[20] EM and IM are the only possible cases if Merge is binary, the simplest assumption. The concepts of multidominance, "late Merge," and some others postulate an extension of Merge.

In a wide range of cases, IM unproblematically 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, as in the examples already mentioned (with an interesting class of exceptions that in fact support the principles involved[21]). That follows from another application of the third factor principle of Minimal Computation: pronounce as little as possible. For GP, that translates to the condition that transfer to SM should be simplified by copy deletion (among other principles). 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,[22] 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'' or ''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, structurally ambiguous, garden path, and island structures result from free functioning of the simplest rules, yielding difficulties for perception. Where there are conflicts between communicative and computational efficiency, it seems that the latter prevails, again supporting the revision of the common sense Aristotelian dictum.

Note that what is relevant here is *conflict*: even if conflict favors computational efficiency, as seems to be true generally, that does not entail that perceptual/parsing considerations play no role in language design. There is good evidence that they do – for example, the differential roles of left/right periphery.

In earlier years it was assumed generally (by me in particular) that displacement is problematic, an ''imperfection'' of language, to be explained by external factors, perhaps parsing considerations. A residue of that error is the belief that EM is somehow simpler and preferable to IM. There is no basis for that belief. If anything, IM is simpler, since it requires vastly less search than EM (which must access the workspace of already generated objects and the lexicon). But the question doesn't arise: both are freely available, given the simplest combinatorial operation, Merge.

Another residue of the same error is the suggestion, frequently advanced, that copies should be independently generated (by EM or some counterpart). That is multiply wrong. It requires a stipulation that IM is barred. Furthermore, it requires far more computation than copy-formation by Merge, since copies may be arbitrarily complex and would have to be generated separately (as in (9b)), then matched somehow and distinguished from repetitions.

A computational procedure requires a collection of atoms, minimal elements that enter into computation though their parts do not. As a first approximation we may 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 *cyclist*, based on the principle of compositionality in the Fregean sense: interpretation of X is a function of the interpretation of its parts and the way they are combined. While commonly employed in some form in traditional linguistic practice, the concept perhaps received its first clear formulation and application in an early 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.

---

[21] On this matter, see Trinh (2011). I put aside here important questions about covert movement.
[22] See note 12. The issues do not arise here, because the notions involved are scarcely controversial.

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 proposal also coexisted uneasily with PSG, with 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 syntactic objects SO, once generated, will not be modified by further computation, yielding considerable computational saving. These ideas have been developed in *phase theory*,[23] which seeks to identify the SOs that are subject to this restriction in the optimal way. One plausible general principle is that operations are restricted to the phase level – with the exception of EM, which is required to construct the phase in the first place: hence IM, agreement, and the operation of transfer to the interfaces. 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}. It is the interior that is subject to no further modification. Elements of the edge – H and a sister of {H, Z} (and a sister of the resulting SO, etc.) – 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 surface position (as in ''the man who said that Z was elected'').[24] I will put further issues aside here.[25]

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, respectively). Morphological discontinuity presumably falls together with the operation Agree, consisting of matching (simple non-distinctness) under minimal search and valuation of features unvalued in the lexicon. That leaves projection, which has no place in the simplest Merge-based system.

Let us look more closely at these components of language.

EM includes familiar cases of compositionality. Under the simplest conception of Merge, EM(X,Y) yields {X,Y} with no projection or order, and independently of the character of X and Y. X-bar theory introduced a further condition: 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 in this regard between PSG and X-bar theory is illustrated by the first basic rule proposed in PSG: S→ NP VP VP.[26] 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) $[_C \text{ C } [_T \text{ NP } [_T \text{ T VP}]]]$

Here NP is SPEC-T, the second Merge to T. But that is just a stipulation. There is no more reason for NP to be SPEC-TP than for TP to be SPEC-NP. In the simpler Merge system, we derive just (16), much as in early PSG, with no relation between NP and T, the T head of TP:

(16) $[_C \text{ C } [_\alpha \text{NP TP}]]$

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 only if, as now generally assumed, the surface subject is a derived position, raised from object in unaccusatives/passives (''John arrived/was elected'') and from predicate–internal subject in transitives (''John saw Bill'').

---

[23] See among others Chomsky (2008) and Gallego (2012). On possible semantic basis for phases, see Larson (2011).

[24] See Obata (2011).

[25] Among them are identification of phases and their options (if any), and the status of root sentences, where the edge has to be interpreted by some special device. Presumably internal constructions with left dislocation, topicalization, and other root phenomena, typically with special intonation, are treated as root sentences. For further discussion, see references of note 23.

[26] 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. See Chomsky (2007) for some discussion.

The PSG/Merge version is clearly preferable to the X-bar theoretic stipulations, hence to be adopted unless there is convincing evidence for the extra complications. 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 (like order) is a distinct property.[27]

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.[28] 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.'' Such considerations as these suggest that the C–T relation is established before the NP surface subject is introduced into (15)/(16), at which point TP = {T, XP} is labeled T, 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 introduces a verbal phrase[29]:

(17)   T[$_\beta$ (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 structure altogether and selects the closest Aux once EA is raised to surface subject position (and the same with (6b) and many other constructions).

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 most closely approximates SMT. The general course should, I think, proceed along the following lines.

Projection is a theory-internal notion, part of the computational process 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. Under PSG and its offshoots, labeling is part of the process of forming a syntactic object SO. But that is no longer true when the stipulations of these systems are eliminated in the simpler Merge-based conception of UG. We assume, then, that there is a fixed labeling algorithm LA that licenses SOs so that they can be interpreted at the interfaces, operating at the phase level along with other operations.[30] The simplest assumption is that LA is just minimal search, presumably appropriating a third factor principle, as in Agree and other operations. In the best case, the relevant information about SO will be provided by a single designated element within it: a computational atom, to first approximation a lexical item LI, a head. This LI should provide the label found by LA, when the algorithm can apply.

Suppose SO = {H, XP}, H a head and XP not a head. Then LA will select H as the label, and the usual procedures of interpretation at the interfaces can proceed.

The interesting case is SO = {XP,YP}, neither a head (we return to the only other possibility, {H, H}). Here minimal search is ambiguous, locating the heads X, Y of XP, YP, respectively. There are, then, two ways in which SO can be labeled: (A) modify SO so that there is only one visible head, or (B) X and Y are identical in a relevant respect, providing the same label, which can be taken as the label of the SO. These are the two cases that are prominently found.

Consider first (A). SO can be modified for labeling by raising either XP or YP, a reinterpretation of Andrea Moro's principle of ''dynamic antisymmetry'' in terms of labeling.[31] Thus Moro 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.

---

[27] There is a large and instructive literature on problems with Specifiers, but if the reasoning here is correct, they do not exist and the problems are unformulable.

[28] On this matter, see references of note 23.

[29] V may be simply an unspecified root, adopting ideas of Alec Marantz and Hagit Borer. We assume that v has (at least) two ''flavors'': v for unaccusative/passive, v* for transitive/unergative. Note that insertion of EA into surface subject position in (15)/(16) is countercyclic, the complementizer C having already been introduced. For an approach to this problem that maintains Merge unchanged but with a narrow extension of its application, see Chomsky (2008). For other approaches, see Epstein et al. (forthcoming-a) and Gallego (2011). We return to the question.

[30] Note that this is a modification of earlier proposals (of mine in particular) that labeling is a prerequisite for entering into computation. That cannot be, as Samuel Epstein points out, or it would block many cases of EM, e.g., Merge (Z, {XP, YP}) = {Z, {XP, YP}}, as in (18), below.

[31] As suggested in Chomsky (2008). See Moro (2000).

If, say, XP raises, then the result will be the structure (18), with two copies of XP:

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

The intuitive idea is that the lower XP copy is invisible to LA, since it is part of a discontinuous element, so therefore β will receive the label of YP. The phenomenon is similar to intervention effects, where the head of a chain (more appropriately, the whole chain, the discontinuous element) induces an intervention effect, but not the lower copy.[32]

The standard convention has been to take each of the copies to be an independent element. Adopting that convention, it may seem to be a stipulation to take the whole chain to be the relevant SO. But the convention has no particular merit. It is quite reasonable to take α to be "in the domain D" if and only if every occurrence of α is a term of D.[33] That yields the intended results for intervention and for the labeling interpretation of dynamic antisymmetry.

Moro's principle has a number of applications. One case is the "split topic" construction in German and many other languages, as in (19):

(19)   Nagetiere hat Peter nur zwei Eichhőrnchen gesehen ("rodents has Peter only two squirrels seen")

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

Another case is the predicate–internal subject construction (17), repeated here, taking v to be v*, with V (perhaps root-based) raising to v*:

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

Since β 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 here and below just for exposition – then β will be labeled v*, as required, since EA is part of a discontinuous element, hence invisible to LA. Therefore EPP is forced in this particular case (though other EPP problems remain). Suppose EA remains in situ and IA raises. 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*.[34] Alexiadou and Anagnostopoulou (2001) have proposed that in structures of the form β, either EA or A must raise.[35] If that thesis can be sustained, then the conclusion could simply follow from labeling.

The most general case of lack of label is successive-cyclic movement. The intermediate steps are of the form {$_\alpha$ XP, YP}, where XP can be for example a *wh*-phrase with YP a CP. The syntactic object α cannot be labeled, but it must be interpreted, if only for theta-marking. If XP raises, then α will be labeled Y, as required. Therefore XP must raise, and successive cyclic movement is forced.[36] 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 in X-bar theory and its descendants.

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 JFK was assassinated in which Texas city?

---

[32] On intervention effects and their subtleties, see Sigurðsson and Holmberg (2008).

[33] Following a suggestion of Hisatsugu Kitahara, using the notions "term" and "occurrence" in the technical senses defined elsewhere. For improvements, see Collins and Stabler (forthcoming), who also reinterpret discontinuity in these terms. See Krapova and Cinque (2008), for the same interpretation of intervention effects.

[34] Technically, what is visible to LA is {EA, {v*}}, v* the complex element formed by head-raising of V (or, perhaps, {v, R}, along lines discussed earlier), and the internal argument part of a discontinuous element, hence invisible to LA. The labeling algorithm has to be designed so that search into a singleton set is minimal, and the nature of head-adjunction has to be clarified.

[35] 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 as a unit (see Wurmbrand, 2006). That leaves open the question how German (and perhaps Japanese) differs from English-like languages, perhaps because of some difference in the internal structure of the verb phrase, perhaps because of options in fixing LA, perhaps in some other way. As noted, Ott shows that the Moro-based analysis does hold for German in other constructions.

[36] One may ask why YP doesn't raise; similarly, why v*P doesn't raise stranding EA. That opens some interesting questions that I will put aside here.

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

(21)   they thought [$_\alpha$in which Texas city [C, [JFK was assassinated]]]?

The fact that α has no label bars (21), as required – and also eliminates recourse to a variety of devices that have been proposed over the years to account for successive-cyclic raising with the interesting residues it leaves.

Let's turn next to (B), {XP, YP} in situ without raising. One illustration is indirect questions, as in (22), as contrasted with (21):

(22)   they wondered [$_\alpha$in which Texas city [$_\beta$C [JFK was assassinated]]]

In this case too α is of the 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 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), a condition that may relate to questions of non-phasal movement more generally. 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 familiar examples as ''PRO to seem to be intelligent is not as easy as you might think.'' This appears to be a straightforward matter of CI-interpretation of the subject–predicate phrase α. Surface subject has other special properties (what Luigi Rizzi calls an ''aboutness'' property), illustrated in various ways; in null subject languages by the fact that silent *pro* can refer to the surface subject (but not the object) of the preceding sentence.[37] Such considerations suggest that α receives its own interpretation and should be labeled – in accord with the general principle that all SOs that reach the interfaces must be labeled. Perhaps that can be achieved by the device suggested for embedded interrogatives. NP and TP share prominent features, namely φ-features – so-called ''SPEC-Head agreement,'' known to have a variety of interesting properties.[38]

Mere matching of most prominent features does not suffice (Marcel den Dikken, pc), as shown, for example, by the copular constructions that Moro studied. What is required is not just matching but actual agreement, a stronger relation, which holds in the indirect question and subject–predicate examples but not small clauses. Sharpening this condition requires a closer analysis of Agree, which would carry us too far afield.

For this approach to be tenable, it must be that LA 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 natural, though the implications remain to be explored.

It seems further that only certain features can serve as labels. Coordination provides motivation for this conclusion. First, we have to distinguish structured from unstructured coordination, the latter as in ''John is tall, happy, hungry, bored with TV, etc.'' 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 the generative enterprise, since they seem to require infinitely many rules to avoid unwanted structure. A natural 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.[39] It can be applied indefinitely often, adding individual predications without further structure, thus avoiding the problems that arise under PSG and TG.

---

[37] Rizzi (2010), part of his theory of criterial positions. For similar ideas on such positions, see Epstein (1992). Depending on how they are analyzed, anaphora, quantificational structures, negative polarity items, and other phenomena might also provide evidence for independent interpretation of subject–predicate structures, as also commonly assumed in traditional grammar.

[38] The same may be true of possessives ({John, {Poss, book}}), assuming hidden agreement as manifested in Turkish. On subject–predicate structures with non-nominal SPEC, see in particular Miyagawa (2010).

[39] For further discussion, see Chomsky (2004). On the problems such constructions pose for PSG and TG, see Lasnik (2011).

Structured coordination is different. Possibly the underlying structure of [Z and W] is (23a), capturing the semantic symmetry of coordination (abstracting from pragmatic properties). To label β, one of Z or W must raise, say Z, yielding (b):

(23)   (a) $[_\alpha \text{Conj} [_\beta Z\ W]]$
        (b) $[_\gamma Z [_\alpha \text{Conj} [_\beta Z\ W]]]$

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 but rather the label of Z, typically shared 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 available as a label, so that γ receives the label of Z.[40]

Resort to Pair-Merge may well be appropriate in other circumstances. Recall the Alexiadou–Anagnastopolou 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 the verbal phrase, 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.[41] 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. The approach could be extended to double objects, capturing the direct–indirect asymmetry familiar from traditional grammar. If the approach proves viable, it would be another case where labeling forces an appropriate analysis.

To summarize the proposals about labeling: We assume that a label is required for interpretation at the interfaces, and that labels are assigned by a minimal search algorithm LA applying to an SO (like other operations, at the phase level). There are two basic cases: {H, XP} and {XP, YP}. The first is straightforward: LA selects the label H and the usual interpretive procedures apply. The second includes three types: (i) successive-cyclic movement, (ii) criterial positions, and (iii) other symmetric structures.

Let us turn to at least a brief mention of some of the many remaining problems.

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.[42] Clearly LIs are complex objects with many properties that do not derive from operations of narrow syntax, and there are many questions about where these complexities enter into a derivation (e.g., are phonological properties introduced in externalization to SM, as proposed in distributed morphology?). If φ-features can move independently of the complex in which they appear in an LI, then they are atoms of computation but are not LIs. There are a variety of questions of this nature, some already mentioned.

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 it must be a more complex structure, perhaps D-pro (see Uriagereka, 1988). The same holds for nouns, as in ''John left'' or ''sugar is sweet.'' 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.[43] 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] 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 the head of the free relative is complex – e.g., Spanish *lo que* 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''[44]). Counterparts to such constructions do however appear as heads of free relatives in other languages,

---

[40] The assumption under consideration is that although C is not a possible label, it must still be visible for determining the structure. Otherwise, as a reviewer points out, Z and W would be equally prominent in (b).

[41] For a carefully worked out version in more current terms, see Larson (1988). For an approach to the puzzles that arise without resort to Pair-Merge, see Kitahara (2012). For discussion and analysis of issues that arise in these and related constructions, see particularly Pesetsky (1996).

[42] Including the question of ordering within the lexicon and the phonology, interesting questions but not relevant to the discussion of order in the computational system, the topic here.

[43] See note 29. It might be even more complex. Thus proper nouns could be of the form {D, {n, R}}, D a determiner, spelled out in some languages. Assuming this analysis, the complement of an LI is actually the complement of the more complex form. E.g., the object of a verb is actually the object of {v, R}, R a root. Selectional properties would have to be interpreted accordingly.

[44] For my dialect at least, cases range from impossible, like this one, to occasionally marginal.

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.[45]

Another long-standing problem has to do with head-head constructions, the first step in a derivation. If the Marantz-Borer conception is adopted, these will be of the form *f-root*, where *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.

Interesting questions arise with regard to other complex constructions, such as ECM:

(24)   they consider [$_\alpha$John to be intelligent]

Work tracing back to 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* α] (Specifier, in conventional analyses), followed by raising of *consider* to v*, which restores the base order but with different constituency, a rather surprising puzzle, variously addressed.[46] Raising of *John* to sister-of-[*consider* α] can be attributed to the now familiar labeling problem, permitting α 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, though loose ends remain.

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 (2007), there is good reason to suppose that the φ-features of T are in fact inherited from C; and though his argument does not extend to this case, the tense feature as well. The system is simplified if features of an LI cannot move independently of the feature bundle to which they belong. That would entail that all the features of C should be inherited by T, including not only tense (as is overt) but also Q. If so, then in (12), spelled out more fully here, the Q feature of C should also appear in T:

(25)   they asked [if-Q [$_\alpha$[how many mechanics] [T-Q fix the cars]]]

The surface *wh*-phrase subject is therefore already in its criterial position as sister-of-QP. Its Q-feature agrees with the head of the sister phrase QP, so it therefore need not – in fact cannot – raise any further (and α is labeled Q). That leaves an escape hatch for a *wh*-phrase object, permitting (13) while blocking (14). 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.[47]

As is well known, the paradigm of (13)–(14) appears to extend 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 (Rizzi, 1982), 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 English[48] know these facts without evidence.

One possible approach, suggested by Luigi Rizzi (pc), is that *criterial freezing* in his 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 agree with 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

---

[45] For much information about these structures, see Caponigro (2003). For an interesting approach to free relatives on assumptions related (but not identical) to those here, see Ott (2011b).

[46] For discussion and sources, see. Lasnik (2003). These approaches assume raising to SPEC-AGR-O, not SPEC-V, but it is not clear that the extra assumptions are necessary. On a possible generalization with EPP, see Chomsky (2008).

[47] The approach is reminiscent of the Vacuous Movement Hypothesis of Chomsky (1986), but avoids the special stipulations required there, and the fact that vacuous movement was defined in terms of ordering, hence is not part of narrow syntax if the discussion here is correct. 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. For discussion of these options and consequences, see Ouali (2008, 2011).

[48] 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). For a different approach, see Epstein et al. (forthcoming-b).

freezing. Deletion of *that* in (14′) might leave only a weakened form of C (perhaps FIN of Rizzi's left periphery), lacking F. Under these (unfortunately still stipulative) assumptions, the unification and variations would follow.

The cases discussed leave many questions open, and there remain a substantial array of others, well beyond the hints here, 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, as well as accounts that seem straightforward but remain unexplained in principled ways, familiar features of the sciences, particularly since their modern origins.

References

Alexiadou, A., Anagnostopolou, E., 2001. The subject-in-situ generalization and the role of case in driving computations. Linguistic Inquiry 32, 193–231.
Baker, M., 2001. Atoms of Language. Basic Books, New York.
Baker, M., 2011. The interplay between universal grammar, universals, and language specificity. Linguistic Typology 15, 473–482.
Berwick, R., Chomsky, N., 2011a. In: Di Sciullo, A.M., Boeckx, C. (Eds.), The Biolinguistic Program: The Current State of its Evolution.
Berwick, R., Chomsky, N., 2011b. In: Di Sciullo, A.M., Boeckx, C. (Eds.), The Biolinguistic Program: The Current State of its Evolution and Development.
Berwick, R., Pietroski, P., Yankama, B., Chomsky, N., 2011. Poverty of the stimulus revisited. Cognitive Science 35, 7.
Bortolussi, L., Sgarro, A., Longobardi, G., Guardiano, C., 2011. How many possible languages are there? In: Bel-Enguix, G., Dahl, V., Jiménez-López, M.D. (Eds.), Biology, Computation and Linguistics. IOS Press, Amsterdam, pp. 168–179.
Cable, S., 2007. The Grammar of Q. Ph.D. Dissertation, MIT.
Cable, S., 2010. The Grammar of Q: Q-Particles, wh-movement and pied-piping. Oxford University press, Oxford.
Caponigro, I., 2003. Free Not to Ask: On the Semantics of Free Relatives and Wh-Words Cross-Linguistically. Ph.D. Dissertation, UCLA.
Carroll, S., 2005. Endless Forms Most Beautiful. Norton, New York.
Chater, N., Christiansen, M.H., 2010. Language acquisition meets language evolution. Cognitive Science 34, 1131–1157.
Chomsky, N., 1957. Syntactic Structures. Mouton, The Hague.
Chomsky, N., 1986. Barriers. MIT, Cambridge, MA.
Chomsky, N., 2004. Beyond explanatory adequacy. In: Belletti, A. (Ed.), Structures and Beyond. Oxford University Press, Oxford, NY.
Chomsky, N., 2007. In: Sauerland, U., Gärtner, H.-M. (Eds.), Approaching UG from Below.
Chomsky, N., 2008. On phases. In: Freidin, R., Otero, C.P., Zubizarreta, M.L. (Eds.), Foundational Issues in Linguistic Theory. Essays in Honor of Jean-Roger Vergnaud, vol. 291–321. MIT, Cambridge, MA.
Chomsky, N., 2010. Some simple evo-devo theses: how might they be true for language? In: Larson, R., Déprez, V., Yamakido, H. (Eds.), The Evolution of Human Language: Biolinguistic Perspectives. Cambridge University Press.
Chomsky, N. Minimal recursion: exploring the prospects. In: Speas, M., Roeper, T. (Eds.), Proceedings of the Conference on Recursion: Structural Complexity in Language and Cognition. U. Mass, Amherst, forthcoming.
Chomsky, N., Halle, M., Lukoff, F., 1956. On Accent and Juncture in English, for Roman Jakobson. Mouton, The Hague.
Collins, C., Stabler, E. A Formalization of Minimalist Syntax, forthcoming.
Curtis, S., 2013. Revisiting modularity: using language a window to the mind. In: Piattelli-Palmarini, M., Berwick, R.C. (Eds.), Rich Languages from Poor Inputs. Oxford University Press, Oxford.
Daland, R., Pierrehumbert, J.B., 2011. Learning diphone-based segmentation. Cognitive Science 35, 119–155.
Dediu, D., 2010. Linguistic and genetic diversity – how and why are they related? In: Brüne, M., Salter, F., McGrew, W. (Eds.), Building Bridges Between Anthropology, Medicine and Human Ethology. Europäischer Universitätsverlag, Bochum, pp. 169–178.
Dediu, D., 2011. A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. Procedings of the Royal Society 278 (1704), 474–479.
Di Sciullo, A.M., Boeckx, C. (Eds.), 2011. The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty. Oxford U., Oxford.
Donati, C., 2006. On wh-head movement. In: Cheng, L.L.-S., Corver, N. (Eds.), Wh-Movement: Moving on. MIT, Cambridge, MA, pp. 21–46.
Enfield, N., 2009. Without social context? Science 239 (September).
Epstein, S., 1992. Derivational constraints on A′ [A-bar]-chain formation. Linguistic Inquiry 23 (2.), 235–259.
Epstein, S., H. Kitahara, T. Seely. Structure Building That Can't Be. In: Uribe-Etxebarria, M., Valmala, V. (Eds.), Oxford U, Oxford, forthcoming-a.
Epstein, S., H. Kitahara, T. Seely. Set Intersection in Bare Phrase Structure is Generable by Simplest Merge: Implications for Complementizer ''Trace'' Explanation, forthcoming-b.
Freidin, R., Vergnaud, J.-R., 2001. Exquisite connections: some remarks on the evolution of linguistic theory. Lingua 111, 639–666.
Gallego, A.J., 2011. Deriving Feature Inheritance from the Copy Theory of Movement: Non-Phase Heads as (Copies of) Phase Heads. UAB, MS.
Gallego, A.J. (Ed.), 2012. Phases: Developing the Framework. Mouton de Gruyter, Berlin.
Gallistel, C.R., 1998. Symbolic processes in the insect brain. In: Scarborough, D., Sternberg, S. (Eds.), Methods, Models and Conceptual Issues: An Invitation to Cognitive Science, vol. 4. MIT, Cambridge.
Gallistel, C.R., 1999. The replacement of general-purpose learning models with adaptively specialized learning modules. In: Gazzaniga, M.S. (Ed.), The Cognitive Neurosciences. 2nd edition. MIT press, Cambridge, MA.
Gallistel, C.R., King, A.P., 2009. Memory and the Computational Bralin: Why Cognitive Science will Transform Neuroscience. Wiley-Blackwell, Malden, MA.
Harris, Z.S., 1951. Methods in Structural Linguistics. U. of Chicago, Chicago.
Huybregts, R., van Riemsdijk, H., 1982. Noam Chomsky on the Generative Enterprise. Foris, Dordrecht.

Joos, M., 1957. Readings in Linguistics. American Council of Learned Societies, Washington.

Kayne, R., 2010. Why are There No Directionality Parameters? NYU, MS.

Kayne, R., 2012. Comparative Syntax. NYU, MS.

Kitahara, H., 2012. External merge of internal argument DP to VP and its theoretical implications. Reports of the Keio Institute of Cultural and Linguistic Studies 43, 241–248.

Krapova, I., Cinque, G., 2008. On the Order of wh-Phrases in Bulgarian Multiple wh-Fronting. In: Zybatow, G., et al. (Eds.), Formal Description of Slavic Languages: The Fifth Conference, Leipzig 2003. Frankfurt am Main, Peter Lang, pp. 318–336.

Larson, R., 1988. On the double object construction. Linguistic Inquiry 19, 335–391.

Larson, R., 2011. In: Di Sciullo, A.M., Boeckx, C. (Eds.), Clauses, Propositions, and Phases.

Lasnik, H., 2003. The position of the accusative subject in the accusative-infinitive construction. In: Bhaskararao, P., Venkata, S.K. (Eds.), Non-nominative Subjects, vol. 1. John Benjamins, pp. 269–282.

Lasnik, H., 2011. In: Di Sciullo, A.M., Boeckx, C. (Eds.), What Kind of Computing Device is the Human Language Faculty?

Longobardi, G., Roberts, I., 2010. Universals, diversity and change in the science of language. Lingua 120, 2699–2703.

Longobardi, G., Roberts, I., 2011. Non-arguments about non-universals. Linguistic Typology (July).

Miller, G.A., Chomsky, 1963. Finitary models of language users. In: Luce, D.R., Bush, R.R., Galanter, E. (Eds.), Handbook of Mathematical Psychology, vol. 2. Wiley, New York.

Miyagawa, S., 2010. Why Agree? Why Move? Unifying Agreement-Based and Discourse-Configurational Languages. MIT, Cambridge, MA.

Moro, A., 2000. Dynamic Antisymmetry. MIT, Cambridge, MA.

Musso, M., Moro, A., Gluache, V., Rijntjes, M., Bücheli, J.C., Weiller, C., 2003. Nature Neuroscience 6, 774–781.

Narita, H., 2011. Phasing in Full Interpretation. Ph.D. Dissertation, Harvard.

Nielsen, R., 2009. Adaptationism – 30 years after Gould and Lewontin. Evolution 10–63.

Obata, M., 2011. Root, Successive-Cyclic and Feature-Splitting Internal Merge: Implications for Feature-Inheritance and Transfer. Ph.D. Dissertation, U. Michigan.

Ott, D., 2011a. Local Instability: The Syntax of Split Topics. Ph.D. Dissertation, Harvard.

Ott, D., 2011b. A note on free relative clauses in the theory of phases. Linguistic Inquiry 183–192.

Ouali, H., 2008. On C-to-TΦ-feature transfer: the nature of agreement and anti-agreement in Berber. In: D'Alessandro, R., Hrafnbjargarson, G.H., Fischer, S. (Eds.), Agreement Restrictions. Mouton de Gruyter, Berlin, pp. 159–180.

Ouali, H., 2011. Agreement, Pronominal Clitics and Negation in Tamazight Berber: A Unified Analysis. Continuum, New York.

Pesetsky, D., 1996. Zero Syntax: Experiencers and Cascades. MIT, Cambridge, MA.

Petitto, L.A., Berens, M.S., Kovelman, I., Dubins, M.H., Jasinska, K., Shalinksy, M., 2012. The ''Perceptual Wedge Hypothesis'' as the basis for bilingual babies phonetic processing advantage: new insights from fNIRS brain imaging. Brain and Language 121 (2), 142–155.

Reuland, E., 2011. Anaphora and Language Design. MIT, Cambridge, MA.

Richards, M., 2007. On feature-inheritance: an argument from the Phase Impenetrability Condition. Linguistic Inquiry 38.3, 563–572.

Rizzi, L., 1982. Issues in Italian Syntax. Foris, Dordrecht.

Rizzi, L., 2010. On Some properties of criterial freezing. In: Panagiotidis, E.P. (Ed.), The Complementizer Phrase. Oxford U., Oxford, pp. 17–32.

Rizzi, L., Shlonsky, U., 2007. Strategies of Subject Extraction. In: Sauerland, U., Gärtner, H.-M. (Eds.), pp. 115–160.

Ross, J.R., 1967. Constraints on Variables in Syntax. Ph.D. Dissertation, MIT.

Ross, J.R., 1986. Infinite Syntax! Norwood, NJ, Ablex.

Sauerland, U., Gärtner, H.-M. (Eds.), 2007. Interfaces + Recursion = Language? Mouton, New York.

Sherman, M., 2007. Universal genome in the origin of metazoa: thoughts about evolution. Cell Cycle 6 (15), 1873–1877.

Shukla, M., White, K.S., Aslin, R.N., 2011. Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. Proceedings of the National Academic Science of the United States of America 108. (15).

Sigurdsson, H., Holmberg, A., 2008. Icelandic dative intervention: number and person are separate probes. MS. http://www.uni-leipzig.de/heck/agree08/sigurdssonholmberg.pdf.

Smith, N., Tsimpli, I., 1996. The Mind of a Savant. Blackwell, Oxford.

Tomasello, M., 2006. Acquiring Linguistic Constructions. In: Damon, W., Lerner, R., Kuhn, D., Siegler, R. (Eds.), Handbook of Child Psychology. 2. Cognition, Perception, and Language. Wiley, New York, pp. 255–298.

Trager, G.L., Smith, H.L., 1951. An Outline of English Structure. Battenberg, Norman, OK. Corrected edition, 1956, American Council of Learned Societies, Washington, DC.

Trinh, T., 2011. Edges and Linearization. Ph.D. Dissertation, MIT.

Trubetzkoy, N., 1969. Principles of Phonology, trans. C. Baltaxe from 1939 German Original. U. of California, Berkeley.

Uriagereka, J., 1988. On Government. Ph.D. Dissertation, U. of Connecticut.

Wurmbrand, S., 2006. Licensing case. Journal of Germanic Linguistics 18 (3), 175–236.

Yang, C., 2002. Knowledge and Learning in Natural Language. Oxford University Press, Oxford.

Yang, C., 2004. Universal grammar, statistics, or both? Trends in Cognitive Sciences 8 (10).