Some Core Contested Concepts


The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Some Core Contested Concepts

Noam Chomsky

Corresponding email address:

chomsky@mit.edu

I’ve been asked to say a few words on core concepts that have been prominent in the study of language in the modern period, most of them seriously contested, often I think on dubious grounds.

There are many questions we can ask about language. The most fundamental one, surely, is: What is language? To the extent that we have some grasp of this, we can proceed to study other significant questions. Lacking that, inquiry is necessarily curtailed. No biologist, for example, would seek to study the development or evolution of the eye without a fairly clear conception of what an eye is, of its essential nature. It would not suffice to say, “It’s used to read”—or of language, that it is used to communicate (among other uses). Those familiar with the field of linguistics will recognize that we have already entered terrain that is highly contested—needlessly, I think.

On the basis of a tentative account of what language is, the more far-reaching the better, we can proceed to ask further questions, among them: How do its concepts, principles, and results inform investigation of acquisition, use, neural representation, historical change, and evolution? Reciprocally, what is learned about these topics can lead to a recasting of the theory
of language (Chomsky, 1986, 1995, 2001). The logical hierarchy of questions does not of course determine order of research into them.

It’s rather curious to see how casually the most fundamental question is treated in classic texts, or in fact today. In this work I will reframe the question as: What conditions should one’s best theory of language satisfy?

Taking some of the classics, for Ferdinand de Saussure, language (in the relevant sense: langue) is a storehouse of word images in the brains of a collectivity of individuals, founded on a “sort of contract.” For Leonard Bloomfield, language is an array of habits to respond to situations with conventional speech sounds, and to respond to these sounds with actions. Alternatively, in his “Postulates for the Science of Language” (Bloomfield, 1926), language is “the totality of utterances made in a speech community”—a variant of William Dwight Whitney’s traditional conception of language as “the sum of words and phrases by which any man expresses his thought” (Whitney, 1873) which however differs in an important way to which I’ll return. Sapir defines language as “a purely human and non-instinctive method of communicating ideas, emotions, and desires by means of a system of voluntarily produced symbols.” An approach that seems to me more useful and suggestive, though left undeveloped, is Otto Jespersen’s: The study of language seeks to discover the “notion of structure” that users of the language have internalized, enabling them to produce and understand “free expressions,” commonly new to speaker and hearer.

Turning to the philosophical literature, the most influential work on the nature of language (not on language use, which is a different question) in the post-war period was that of
W.V. Quine (Quine, 1960). He considered a language to be a kind of theory, where a theory is a fabric of sentences associated with one another and stimuli by the mechanism of conditioned response. Elsewhere, he takes a language to be an infinite set of well-formed sentences, perhaps something like Bloomfield’s “totality of utterances.” This notion, if coherent at all, is derivative, presupposing some mode of generation, which Quine held it would be “folly” to postulate (Quine, 1970).

Within the professional linguistics of the time, a standard view was what Martin Joos called the Boasian tradition that languages can differ arbitrarily, and that each one must be studied without preconceptions. That could not be literally true, of course, but it was held to be a reasonable guideline. The accompanying assumption was that a language is studied by applying analytic procedures to a corpus, basically techniques of segmentation and classification. There is, then, little to say about what language is.

Moving to the present, Max Planck researcher N. J. Enfield, expressing a common view in the cognitive sciences, writes in the journal Science that language may be “a constellation of cognitive capacities” that have other functions (Enfield, 2010). Another influential view is that a language is some rather arbitrary array of “constructions” acquired by general learning procedures and stored in memory (of astronomical size)—perhaps, we read in a recent issue of one of the leading professional journals, just one procedure: statistical analysis of a corpus.

The earlier notions were understandable at the time. But in the past half century an enormous amount has been learned about language in all its aspects, including the discovery of sharp dissociations from other cognitive processes in both directions: high linguistic competence
with defective cognition and empathy, and the converse. Contemporary theses of the kind mentioned merit attention because of their prevalence—and in my view, only for that reason, though perhaps that is a minority view.

By *language* in these remarks I will mean *I-language*, a biological object internal to an individual, and considered *in intension*. That is, we are interested in the actual nature of the system, not just in its external manifestations. Adapting David Marr’s terminology for processing systems, such study of I-language can proceed at the level of computation, algorithm, or mechanism, all of course interrelated.

Another concept that appears in the literature is *E-language*. I have to take responsibility for the term, but not for the ways it is conventionally used, which do not seem to me coherent. Sometimes it seems to be used rather in the sense of the Bloomfield-Quine conception of “totality of utterances,” but as mentioned above, that concept presupposes some finite generative process, neurally encoded, and even then it is by no means clear that any well-defined set is determined, for reasons discussed extensively over half a century ago.

Among the various conceptions of language, the concept I-language is unique in that it presupposes none of the others and is presupposed, if only tacitly, by each of them. Any conception of language will at least assume that there are individual users, and that they use their language by virtue of some internal state of (mostly) their brains—that is, an I-language understood within what has come to be called “the biolinguistic program.” Further inquiries into language are likely to be significant insofar as the concept of I-language they presuppose is adequate to their purposes.
The most elementary property of I-language is that it is a system of discrete infinity. The study of I-language thus falls within the theory of computation (Turing machine theory, theory of recursive functions). An I-language can be taken to be (at least) a computational procedure that yields an unbounded array of hierarchically structured expressions, each assigned an interpretation at two interfaces with other internal systems, sensory-motor (SM) and conceptual-intentional (CI)—roughly sound and thought, though research on sign languages reminds us that sound is only one mode of SM externalization (Neidle et al., 2000).

Computational procedures of this kind are called *generative procedures*. Understanding the term *grammar* to refer to theory of language, the study of I-languages is therefore called *generative grammar*. That concept too is contested, often regarded as wrong or at best as only one of a number of competing approaches to language. Let’s look into these ideas.

To question the validity of generative grammar is to hold that there is no specific sound-meaning correlation, unbounded in range, that differentiates, say, English and Finnish. Since no one believes that, there can be no serious question about the validity of generative grammar, though of course that leaves open the form it might take.

To question the biolinguistic conception of I-language/generative grammar is to hold that what differentiates users of English and Finnish is independent of their internal states, mostly their brains. Again, no one believes that, so there can be no serious question of the validity of this conception either.

Consider the belief that generative grammar within the biolinguistic framework is just one way to study language, competing with others, rather like conflicting theories of phonology,
or of cosmology. That cannot be, if it is true that every approach to language presupposes, at least tacitly, that a language determines a specific sound-meaning correlation and is used by virtue of an internal state. To be sure, one can study language without any interest in sound-meaning correlations and without concern for the internal states of the users of the language. Similarly, one may study bees without being concerned with the internal states that enter into their behavior, though such study will necessarily be limited if it does not take into account what is discovered about these internal states. But there is no competition between these approaches, and no issues arise beyond choice of interests. If these are contested notions, they shouldn’t be.

I-language satisfies a classical concept of language as sound with meaning, though it may be—I think it is—more accurate to say that it satisfies a different classical concept: meaning with sound (or other externalization)—a very different picture, implicit in Whitney’s approach to language, cited above.

A theory of language must at least seek to capture these properties. Correspondingly, an account of acquisition or evolution of language must have at least this much as its target, if only tacitly. It is quite appropriate for a collection of essays on language to have the title Interfaces + Recursion = Language followed by a question mark (Sauerland and Gaertner, 2007), because there may be considerably more; how much more is a central research question of the theory of language. One natural research program can begin by assuming “Nothing more,” then asking how much is left unaccounted for. It will, then, start by assuming some version of what has been called the Strong Minimalist Thesis (SMT) in recent work (Chomsky, 1998).
The core concept of recursion is often confused in the literature with center-embedding, a very different notion. Sometimes recursion is incorrectly assumed to be necessarily infinite. A recursive function may yield a single output (or nothing). If the output is finite, then the generative mechanism is in effect a list—for language, a list of astronomical size even for fairly simple expressions, not a serious proposal. If the list replaces minimal elements by categories, the same hopeless problem remains. If it introduces phrases, then it is presupposing some generative mechanism, and we are back to generative grammar.

It is easy to construct a pseudo-I-language that lacks resources to generate more than a fixed number of expressions—more precisely, of non-deviant expressions, since free use of the mechanisms of the I-language can go on indefinitely, generating deviant expressions of many kinds. For example, we could devise a pseudo-language truncated English (TE) lacking connectives, free adjectival modification, relational expressions like father-of, and other devices that render normal English unbounded. A child learning TE would proceed exactly as in learning English, since she would have no information about these limitations. At some point, however, she would find out that the resources of the language do not permit her to say very much—that is, to express her thoughts beyond very narrow limits.

From the study of child language acquisition, most strikingly the study of children who have acquired language with no input as in Susan Goldin-Meadow’s work (Franklin, Giannakidou, and Goldin-Meadow, 2011; Goldin-Meadow, 2005), we can surmise that children exposed to TE would use their innate capacities to remedy the deficiencies. But if TE or some similarly impoverished system were to persist, it would be a mere curiosity, with no implications for the theory of acquisition, for universal grammar (UG) in the technical sense of work in
generative grammar in the past half-century, or for the study of language generally, contrary to substantial recent literature. That is another contested notion that should be put to rest, I think.

There are, however, interesting things to say about such apparently truncated languages, if the topic is addressed seriously, as it has been. The most illuminating work I know of is by Kenneth Hale, in his very important 1975 paper “Gaps in Grammar and Culture.” Focusing mostly on Australian languages, Hale observes that while they often lack conventionalized linguistic counting systems (numerals), they have various other devices for precise enumeration, and speakers have no problems with counting, which he takes to be a universal human cognitive property. Much the same holds for colors and other “cultural gaps” (Rosch, 1973; Berlin and Kay, 1969). Of particular linguistic-cognitive interest are syntactic and semantic structures that are not instantiated in many of these and other languages, which Hale suggests are also cultural gaps. To account for these, he develops a UG-based approach to a variety of constructions that he suggests may be universal, and apparently lacking in particular languages because some devices made available by UG are not used (as is always the case). These include “relative clauses, the passive, negation with variable sentence-internal scope, topicalization, and others.” Relative clause formation, for example, may be based on universal adjunction and choice of a small number of options for embedding, variously adopted by I-languages, ideas that have been revived in some form in recent work on I-language. In this deeper sense, embedded relatives are universal, Hale proposes, and the same may be true in other cases. The languages appear to be “truncated,” in the sense mentioned earlier, but the appearance is superficial. As Hale showed (Hale, 1973), serious inquiry into alleged truncation can be highly informative, but only if it goes well beyond arrangement of surface forms.
Among many other unbounded recursive devices, English and other I-languages do freely generate center-embedding, though without external aids (time, memory), subjects of course cannot interpret such structures beyond some bound: about seven, as predicted by the late George Miller’s famous magic number seven, plus or minus two (Miller, 1956). Self-embedding, a narrower notion, is much more sharply restricted, a fact of interest for the study of perception/parsing.

Discussion of these topics has often been obscured in the literature by failure to distinguish competence from performance; loosely, to distinguish what we know from what we do. Performance involves many factors, competence being only one. The distinction is contested, but needlessly. It is taken for granted in some form in the investigation of any organic system, cognitive or not, that internal structure has to be distinguished from actual performance of the system under particular circumstances. Inquiries into the nature of the digestive system, for example, may properly abstract away from the effects of stomach flu or what happens after eating a Big Mac, a somewhat analogous distinction.

All of this should, I think, be truism, and it is regarded as such outside the study of language, even in the case of other cognitive systems. Take the simplest of these—arithmetic, an innate capacity that puzzled the founders of modern evolutionary theory since it plainly wasn’t selected. Like Australian Aborigines, each normal person has the capacity to add numbers of arbitrary size, but performance is restricted by available resources—in the limiting case, by what can be computed “in the head,” with no external resources of time and memory. One could construct a theory of calculation in the head, with some elaborate statistical model of how performance degrades, or becomes impossible, as numbers become larger. Fortunately, no one...
undertakes this—I hope. It would be about as informative—at least about fundamental
arithmetical capacity—as a complex list. Arithmetical capacity is captured by internalized rules,
which make it possible to go on indefinitely without changing the internal system, as time and
memory are added. That is the crucial point: *without changing the internal system.*

There is obviously a crucial difference between a device that stores all calculations
internally, and a Turing machine (or standard stored-program computer), which captures our
arithmetical competence with internal rules and has access to external memory. Superficial
experiments might not distinguish the two clearly, but that just tells us that these are the wrong
experiments. The literature on the processing of language by humans, other animals, and
machines is replete with failure to recognize these distinctions. C. R. Gallistel and A. P. King
(Gallistel and King, 2011) make an appropriate observation in their critique of connectionist and
neural net models: “Unfortunately, computer scientists sometimes forget what they know about
the foundations of physically realizable computation when they begin to think about brains.” Or
language, and not just computer scientists. In any event, these are also core notions that are
contested, but I think should not be.

Alongside of knowledge of arithmetic, stored in an internal generative system with access
to unbounded external resources (and possibly derivative from language), we have techniques of
arithmetical calculation, which also have such external access. But that of course does not mean
that the calculating techniques have an independent internal mechanism of recursive generation.
They surely access the system of arithmetical competence. The same holds for externalization of
language (and conceptual-intentional interpretation). The sensory-motor system yields an
unbounded array of utterances in principle, but it cannot be seriously proposed that it has its own
10
generative system. The same holds of parsing and perception. They have their own mechanisms, and can access unbounded external resources, but in doing so they surely access the generative mechanisms of I-languages; otherwise, we could expect to find people who know only Finnish but understand and speak only English.

Other cognitive systems can also carry out computations over an unbounded range of objects. Sometimes they can be arranged in discrete patterns, sometimes continuous or mixed, and varying in numerous other dimensions. This includes the systems of motor organization and vision. Thus both systems can construct infinite arrays of digital patterns, along with or intermingled with continuous ones, but there is no reason to suppose that they have their own digital generative processes, contrary to recent proposals. Rather, they presumably access internal generative systems, possibly those of language.

Let’s now turn to acquisition of language. Like development of any organic system, it involves several factors: (i) external data, (ii) genetic endowment, (iii) more general principles. Factor (ii) includes: (a) language-specific elements (UG), (b) other cognitive systems, (c) constraints imposed by the structure of the brain, about which not enough is known today to be very helpful. Existence of these factors is regarded as uncontroversial, with a single exception: (iia), UG. But that must reflect a misunderstanding. No rational person can believe that there is no genetic basis for the fact that a human infant, but no other organism, instantaneously and effortlessly extricates from the environment language-relevant data, and in a rather comparable way quickly attains rich linguistic competence, again a feat utterly beyond other organisms even in its rudimentary aspects. But that is what is entailed by denying the existence of UG. The same conclusion holds even if one were to adopt the belief that some arrangement of elements of
(iib), other cognitive systems, suffices to account for these achievements; putting aside its implausibility and lack of empirical support or conceptual basis, such a belief still requires that there is some (iia) factor that implements this arrangement. The same would be true if one were to suppose that memory of astronomical scale is what differentiates humans from other organisms—though there is not the slightest evidence for this incredible but not uncommon assumption (and indeed other organisms carry out feats of memory entirely impossible for humans, e.g., jays hiding seeds; Clayton, Emery, and Dickinson, 2006).

Other fashionable approaches are to attribute these achievements to “culture,” a suggestion that gives hand-waving a bad name, or to “theory of mind,” though autistic children highly defective in theory of mind (Baron-Cohen et al., 1985) can acquire rich linguistic competence (and in fact a great deal of language acquisition proceeds before a child shows any sign of having attained theory of mind), and no mechanism is suggested to lead from “theory of mind” to the specific structures of language. Or they may be attributed to “statistics,” leaving virtually everything a total mystery, though no one has doubted, since the early work in generative grammar in the ‘50s, that statistical analysis may well be involved. These comments may seem harsh, but it might be instructive to ask whether they are justified.

Rejection of UG is sometimes based on confusion of UG—the genetic basis for I-language—with linguistic universals, like Greenberg’s very important compilation (Greenberg, 1966). The latter are generalizations, meaning they may well have exceptions. UG does not have exceptions, apart from the margins.
The essentially exceptionless character of UG is not insignificant. One reason is that it tells us *something* at least about the evolution of language—and there is precious little to add of any substance. It tells us that there has been essentially no evolution of the language capacity, at least in the roughly 50 thousand years since our ancestors are assumed to have left Africa. The evidence for this is substantial. Thus infants in Amazonian tribes readily learn Portuguese (Everett, 2005, 2012; for comment see Nevins, Pesetsky, and Rodrigues, 2009), and if brought to New York would speak the local dialects in a way indistinguishable from natives; and conversely. It follows that their language capacity—UG—has not changed. The observation generalizes, with no limits as far as is known. This might be surprising to those who believe—often, it seems, solely on grounds of misunderstanding the modern theory of evolution—that language must have evolved in small steps over a long period. But it is quite consistent with the assumption, much more plausible in my view, that its core properties emerged fairly suddenly (in evolutionary time) by some probably slight rewiring of the brain.

Of course, there is constant historical change, a very different matter. Historical change is not to be confused with evolution. At least in the technical sense of the term *evolution*, languages don’t evolve at all, though they do change over time. Language users evolve, but, with regard to language capacity, not in tens of thousands of years, so it appears, and possibly not since the emergence of cognitively modern Homo sapiens, apparently not too long before they began to leave Africa.

Much of what is published on the “evolution of language” actually consists of speculations about changes that might have taken place long before the trek from Africa. These are speculations about historical linguistics, not evolution. Furthermore, they are speculations of
a kind that are not acceptable in historical linguistics, a discipline that has rigorous standards. It isn’t enough to say, “Maybe this happened, or maybe that.”

For a computational system like language, category (iii) (more general principles) can be expected to include at least principles of computational complexity. As far as I am aware, virtually all of what has been learned about I-language traces back to this factor and to UG (and of course (i), external data). There are many difficult and interesting questions that can be raised about the concept of computational complexity, but they barely arise in this context, if at all, since the only appeal is to principles so elementary that they would be incorporated in any serious analysis of the concept: e.g., less is better than more, minimal search is better than extended search, etc.

As in any rational inquiry, we seek the simplest theory of UG, eliminating stipulations and avoidable complexity. From the earliest origins of the modern study of these topics 60 years ago, research has been directed toward this goal, a matter that should be uncontroversial. In recent years this research has sometimes been called the *Minimalist Program* (MP), another notion that has been widely misunderstood and is often pointlessly contested. The MP is nothing more than normal science, and in fact is a seamless continuation of early inquiry. It differs in that it makes use of what has been learned over the years and, in the light of this progress in understanding, it sometimes suggests some new research programs, such as the one I mentioned: adopt the Strong Minimalist Thesis and ask what more is required. That has proven to be a fruitful suggestion, I think, but one can pursue the MP by adopting it or pursuing a different path. Notice that there is no question of the truth or falsity of the MP. Those are not the kinds of questions that one can sensibly ask of a research program. Such a program can be premature or
inappropriate in some other way, perhaps badly formulated, and so on, but it cannot be true or false. Note also that the MP is theory-neutral: whatever one thinks is the right approach to language, one can be concerned with the MP, as in normal scientific inquiry, or not.

A computational mechanism such as I-language includes some operation to construct a new element Z from elements X, Y already constructed, and to continue applying the operation without limit. Call the operation *Merge*. Unless there is evidence to the contrary, we take *Merge* to be as simple as possible: binary, lacking order, and involving no further computation. We can therefore take Merge(X,Y) to be simply \{X,Y\}. The operation meets the minimal requirement of generating an unbounded array of hierarchically structured expressions, to be supplemented by operations mapping these to the sensory-motor and conceptual-intentional interfaces.

It is hard to imagine a theory of I-language that does not at least include *Merge*. Alternatives that have been suggested are more complex, presupposing *Merge* and something additional—for example, resort to associativity to strip away hierarchy. There are notational equivalents—e.g., Fregean ancestral—which we can ignore, unless some authentic empirical issue arises.

Adopting these optimal assumptions at once raises a number of questions. One has to do with word order: where does it come from? It is clearly needed in the sensory-motor system. One possibility, then, is that it is introduced in the course of externalization to the sensory-motor system so as to satisfy the interface condition. If so, it would not be expected to enter into narrow syntax and mapping to the conceptual-intentional system. For core properties of syntax and semantics, this appears to be true. This includes the essentials of binding theory, theta-role
assignment, and other components of language. The thesis is contested, but in this case quite reasonably. There are many empirical consequences that merit investigation, some leading into complex areas, introducing performance and so-called “functional” factors.

The architectural assumption has a certain initial plausibility, and has some interesting consequences beyond accounting for the apparent disregard of word order in core syntax and semantics. One is that it provides a straightforward explanation for the curious phenomenon of structure-preservation of rules, recognized to be a puzzle from the early days of generative grammar because the computational procedure universally found in language in all relevant constructions, and adopted by language learners without evidence, is more complex than the obvious linear alternative (Berwick, Chomsky and Piattelli-Palmarini, 2012).

To take a simple case: given the sentence “Instinctively, eagles that fly swim” in (1), with the structure shown in (2), we understand that the adverb is not connected to the closest (and most plausible) verb, but rather to the more remote one.

(1) Instinctively, eagles that fly swim.

(2) Instinctively, [[eagles that fly] swim].

The facts follow on two assumptions: (a) the association is determined by the third factor principle of minimal computation, in this case minimal search by the adverb for a verb; (b) the postulated architectural assumption—order is not available for computations that yield the conceptual-intentional interface. On these assumptions, it follows that in the structure (2), the
adverb will modify the structurally closest verb \textit{swim}, not the linearly closest verb \textit{fly}. The same argument extends to the familiar cases of Auxiliary-inversion: thus we can ask (3) but not (4).

(3) Are eagles that fly swimming?

(4) *Are eagles that flying swim?

The latter is a fine thought, and much more simply computed than (3), but the thought is expressible only by circumlocution: “Is it the case that eagles that are flying swim?” Again the structurally closest but not linearly closest position is the one selected instantaneously and universally. The same holds for many other constructions in English and other languages, in fact every relevant case that is known.

Auxiliary-inversion is one of the very few cases of a linguistically significant phenomenon that has been addressed by statistical approaches. In fact a small industry has been devoted to the project. Each effort ignores the quite elementary UG explanation based on a well-motivated architectural assumption. And each fails, dramatically (Kam et al., 2007; Berwick et al., 2012).

There is independent neurolinguistic evidence supporting the same conclusion about the architecture of I-language. A research team in Milan, including linguist Andrea Moro (Musso et al., 2003), found that the processing of nonsense expressions observing UG principles engages the normal language areas of the brain, as expected, but that this does not hold for comparable expressions formed or interpreted by reliance on order—say, negating a sentence by fronting the third word—though that is a much simpler computation than the actual ones that are used.
Apparently subjects were treating the examples violating UG as a puzzle, not as language—
which, as is well known, is dissociated from other cognitive capacities. Similar results were
found by Neil Smith and Ianthi-Maria Tsimpli in their studies comparing normal learners with
the autistic savant Christopher, who has remarkable linguistic capacities but very low cognitive
and social competence and has been their subject in much illuminating work. In experiments
with nonsense languages, normal learners treated rules relying on order as puzzles, while
Christopher, who had no problem with nonsense structures satisfying UG, could not deal with
them at all, as expected given his limited cognitive capacities (Smith and Tsimpli, 1995; Smith et
al., 1993).

These observations suggest a broader thesis: The fundamental “design” of language
satisfies conceptual-intentional interface conditions; externalization in one or other modality is
ancillary. In traditional terms, language is primarily an instrument of thought. Language is
“audible thinking,” as Whitney expressed the traditional conception; it is “the spoken
instrumentality of thought.”

If so, then a fortiori, this is true of the various uses of externalized language. In particular,
it holds for communication, often taken to be the essential property of language, on dubious
grounds. The shift of perspective appears to be well grounded from several points of view
beyond language design, and is of considerable broader significance, matters I will not explore
further here.

Returning to the generative process, the simplest form of Merge yields other
consequences. Suppose Y is merged with X. Then, still assuming binary Merge, there are two
possibilities: (1) Y is distinct from X, (2) Y is contained in X—in technical terms, a term of X, that is, an object formed by an earlier step in the computation. The two possibilities are External Merge (EM) and Internal Merge (IM), respectively.

Without further comment, IM yields the ubiquitous phenomenon of dislocation, and also the so-called “copy theory of movement,” since Merge leaves X and Y unmodified. Note that these conclusions follow directly. To bar either EM or IM-with-the-copy-theory requires stipulation, and indeed multiple stipulations when other devices are introduced to achieve the effects of the displacement and “copying” that “come free” with Merge. These are important consequences, insufficiently appreciated.

By automatically yielding copies, the simplest operation provides the right forms for the conceptual-intentional system in crucial cases, as we would expect if language is “designed” to satisfy the CI interface. Consider for example the sentences (5)–(11), with structures as shown.

(5) We thought [they expected to see each other].
(6) We wondered which boys [they expected to see each other].
(7) We wondered which boys [they expected which boys to see each other].
(8) [Which of his pictures] did they persuade the curator that every artist likes best?
(9) They persuaded the curator that every artist likes [his first picture] best.
(10) [Which of his pictures] did they persuade the curator that every artist likes [which of his pictures] best?
(11) [Which of his pictures] persuaded the museum that every artist paints flowers?
In (5), the anaphor *each other* is dependent on the closest possible antecedent, *they*, in accord with the natural (third factor) principle of minimal computation (minimal search). But in (6), with the identical embedded (bracketed) phrase, *each other* is dependent on the more remote antecedent *which boys*. The result follows directly from the fact that Merge yields not (6) but rather (7), with displacement leaving a copy under Internal Merge, and it is (7) that is transferred to the conceptual-intentional interface, not (6), which derives from (7) by externalization.

In (8), as in (9), the quantifier *every* can bind *his*; hence, (9) is an appropriate answer to (8), which is derived by Internal Merge from the underlying form (10) that is transferred to the conceptual-intentional interface. We know that *every* cannot bind *his* without Internal Merge, as we see in (11), structurally similar to (8) in relevant respects. Throughout, the simplest version of Merge yields the appropriate forms for the conceptual-intentional interface, though plainly not for the sensory-motor interface; we will return to that.

Note that we have to distinguish copies from repetitions. Thus in “What hit what?” the two occurrences of *what* are repetitions, which have nothing to do with one another, just as the two occurrences of *the* have nothing to do with one another in “The boy likes the dog.” But in “What did they see what?” the two are copies, receiving the required dual interpretation at the conceptual-intentional interface (and yielding the surface form “What did they see?”); similarly in (7) and (10). The intuitive basis for the repetition-versus-copy distinction is straightforward. Thus if a repeated lexical item is drawn from the lexicon twice, the two are repetitions; if only once, they are copies, an intuition that extends at once to more complex cases like (7) and (10). There is a simple mechanism to make the right distinction if the relevant operations apply simultaneously, as in phase theory, but I will put that aside here. (See Chomsky, 2013.)
In earlier generative grammar, it was assumed (by me in particular; see Chomsky, 1965) that composition of elements and dislocation are two distinct processes, and that dislocation is more complex and hence problematic, an apparent “imperfection” of language, a departure from optimal design that has to be explained somehow. But that was incorrect. Keeping to the simplest theory, we see that composition falls under External Merge and dislocation under Internal Merge, which are the two possible subcases of the simplest compositional operation Merge. If either were lacking, that would have to be explained. It is often felt that Internal Merge is more complex than External Merge and should somehow be avoided. This is a residue of the earlier errors. If anything, Internal Merge is simpler, since External Merge requires search of the entire workspace and lexicon, while Internal Merge keeps to a single object. But the question does not arise in the best theory: both are available, freely, when Merge is in its simplest form.

It is sometimes proposed that copies should be formed by External Merge, with Internal Merge barred. This is another needlessly contested issue. Forming copies by External Merge is far more complex. It requires independent generation of the copies, which can be of arbitrary complexity. It requires some more complex method to distinguish copies from repetitions. And it has to be stipulated that Internal Merge is unavailable—another extra complication, like barring any other freely available option.

Given two freely available options for Merge, it is natural to ask whether they have different interface properties. At the sensory-motor interface, the distinction is obvious: dislocation versus adjacency. At the conceptual-intentional interface, the distinction appears to correspond to the “duality of semantics”: argument structure (External Merge), and other
properties (Internal Merge: discourse-related, information structures, and scopal properties). The correlation is close. Whether it is perfect is contested, and interesting questions arise (as in the case of control; Polinsky, 2013). But here it is a serious empirical issue.

Quite generally, the simplest form of Merge yields correct conceptual-intentional interpretations. But what reaches the sensory-motor system is of course quite different: Only the hierarchically most prominent copy is externalized in cases of the kind discussed here; others are deleted. That again follows from elementary considerations of computational complexity: compute and externalize as little as possible.

Accordingly, while third-factor considerations of efficient computation (Hauser et al., 2002) provide the correct semantic interpretations, they yield perceptual problems. A well-known difficulty for parsing programs is “missing elements”: filler-gap problems. The sensory-motor correlates of (7) and (10), i.e., (6) and (8) respectively, have gaps that have to be filled to interpret the sentences properly. The perceptual difficulty also derivatively impedes communication. We therefore have a conflict between computational efficiency and perceptual/communicative efficiency. Universally, computational efficiency prevails.

These are interesting situations, not previously investigated, but worth exploring. There are many other examples: structural ambiguity, garden path sentences, and extraction islands are relevant cases. Insofar as these are understood, the attested facts follow from optimal computation, again yielding fine thoughts that can only be expressed in complex ways. Consider for example violations of the trace-licensing condition ECP (Empty Category Principle), illustrated in (12)-(14):

22
(12) He wondered if the mechanics fixed the cars.

(13) How many cars did he wonder if the mechanics fixed?

(14) *How many mechanics did he wonder if fixed the cars?

The thoughts expressed by (13) and (14) are both fine, and both are formed by straightforward application of IM followed by minimizing computation in externalization. But only (13) is acceptable. The interrogative (14) requires a circumlocution, such as: “For how many mechanics is it the case that he wondered if they fixed the cars?” The matter is intricate, but it appears that (14) may be barred by considerations of computational complexity within a phase-theoretic analysis of generation (Chomsky, 2008) that keeps closely to optimal computation. If so, that would be another case of conflict between computational and communicative efficiency, again resolved in favor of the former (Chomsky, 2013).

As far as I know, the conclusion generalizes to all conflicts between computational and perceptual/communicative efficiency that are at all understood, and in some cases understood quite well, as illustrated by (5)–(11). These results again support the conclusion that externalization (hence a fortiori communication) is an ancillary aspect of language, peripheral to its core nature.

Note that I am referring here to conflicts: even if these favor 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 the left and the right peripheries. These crucial distinctions undermine some recent experimental
work that seeks to refute the conclusion that language design favors computational over perceptual/communicative complexity, as appears to be true in every known case. Proceeding along these lines, we can consider the research program I mentioned earlier: asking how closely the basic design of language approximates a “perfect” system, one that includes the core computational operation Merge and conforms as closely as possible to general efficiency conditions: the Strong Minimalist Thesis. The residue would have to be accounted for by complication of innate endowment: UG or, in principle, other elements of innate endowment, if their relevance could be shown. Or the residue might be attributable to interface conditions.

We should expect to find that externalization violates conditions of minimal computation. It is a complex task because the sensory-motor system is unrelated to core language design, even in evolutionary origins, it appears. And in fact, most of the apparent complexity of language lies in the externalization. In learning English, it is necessary to acquire the details of phonetics, morphology, order and arrangement, and so on, but no one learns (or of course is taught) such facts as those I have just mentioned, which abound in all languages. The former are also the aspects of language readily susceptible to historical change, as one would expect if they are ancillary.

These conclusions have extensive empirical support in many languages, and also conform to the little we can plausibly surmise about evolution. The emergence of language required at least the appearance of a computational mechanism to yield an infinite hierarchy of expressions linked to the conceptual-intentional system—hence Merge, in the simplest case. That would have required some rewiring of the brain, presumably by a slight mutation—in an individual of course, not a group. That individual would have had the capacity to think: to plan, to interpret,
and so on. Over generations, the capacity might have proliferated through a small hunter-gatherer group, so that there would be some point in devising modes of externalization—a complex task, which can be solved in many ways over time, yielding what appears superficially to be a wide variety of languages, but as is increasingly discovered for language, and in biology more generally, with a core design that has changed little if at all.

These seem to me about the only conclusions that can be drawn with any credibility about the evolution of the human language faculty, though of course all kinds of tales can be spun about what might have happened.

The earliest work in generative grammar was concerned with several properties of language, crucially composition, order, projection, and displacement. The division of labor in the earliest efforts assigned composition, order, and projection to phrase structure grammar and displacement to transformational grammar. Subsequent work provided good reasons, some already mentioned, to revise this picture: composition and displacement fall under Merge, and order may be part of externalization, a reflex of SM conditions.

That leaves projection, more recently called labeling. Labeling is different from the other elementary properties in that it isn’t virtually detectable in the immediate data available; rather, it is a theoretical notion, and therefore raises interesting questions that I can only hint at here, cutting many corners (Chomsky, 2013).

Suppose we adopt the standard and plausible convention that labeling is required for interface interpretation. An object of the form [V-XP] (hence labeled V, a verb phrase) will be interpreted differently from one of the form [C-XP] (hence labeled C, a clause). Hence every
syntactic object transferred to the interface has to be labeled, optimally by a minimal search algorithm (like Agree). This is straightforward for [H-XP] structures, where H is a head, the most prominent element. Therefore H is the label, and standard interpretive procedures can be carried out. The interesting cases are XP-YP structures (neither XP nor YP a head). In X-bar theory and its various descendants, such structures are barred by a stipulation requiring endocentricity. That’s clearly wrong. Apart from the extra stipulation, which is already problematic, every application of Internal Merge yields an XP-YP structure (putting aside head-movement here, very likely a different operation). The stipulation does capture something important, however, which is that at the interface level, standard interpretive procedures do typically apply to [H-XP] structures, labeled H.

Therefore one task for the theory of labeling is to find ways to convert [XP-YP] structures to endocentric ones, or to provide some other mode of interpretation.

One way to render [XP-YP] structures labelable is to raise XP or YP, leaving only the other visible to the labeling algorithm (there is a good reason why the copy left behind is not visible)— basically a restatement of Andrea Moro’s dynamic anti-symmetry thesis in terms of labeling (Moro, 2000, 2013). The major illustration is successive-cyclic movement, which is forced in order to provide labeling at the interface. Thus if Internal Merge generates “You say which book John saw” with the structure shown in (15) as a step leading to (16), then one of the two sisters in the brackets in (15) has to raise, and for independent reasons it has to be the wh-phrase, so that the residue is labeled as a clause.

(15) you say [which book - John saw]
(16) Which book did you say John saw?

That leaves two basic cases. One is the final stage of Internal Merge, as in “I wonder [which book – John saw].” Another is the simple subject-predicate construction. In the literature, it is stipulated that one of the two elements is peripheral, the specifier of the other; thus the subject-predicate construction is taken to be \([XP-[T-YP]]\), with \(T\) the head and \(XP\) the specifier. But that is completely arbitrary. The theory of composition provides no reason for this decision, rather than calling \([TP]\) the specifier of \(XP\). In a pure Merge theory, the notion \textit{specifier} is not definable: there are no notions first merge, second merge, etc., as there are in X-bar theory, which adds stipulations beyond Merge.

In both the terminal stage of IM and the subject-predicate construction, \(XP\) and \(YP\) share their most prominent features: \(Q\) in the wh-phrase case, phi-features in the subject-predicate case. Hence the labeling algorithm will find the same most prominent feature in \(XP\) and \(YP\), and can label accordingly. This is the case that Luigi Rizzi calls “criterial freezing,” which again may be reducible to labeling (Rizzi, 2010, 2014).

It seems, then, that \([XP-YP]\) structures, freely generated, yield two basic cases: raising (which may furthermore be forced to be successive-cyclic) and criterial freezing, the two attested constructions. There is a good deal more to say, but this seems to be a promising approach to dealing with the network of issues relating to projection/labeling.

There are other crucial issues. To take one, even the simplest words of language—\textit{tree, river, person}, etc.—have properties with no counterpart in animal communication, and this is even more evident when we proceed to more complex ones. All of this poses major problems for
evolution and acquisition. It is sometimes proposed that these and other properties of language compel us to postulate a prior Language of Thought (Fodor, 1975), but whatever the merits of that conception, it doesn’t help here, since it just sets up an infinite regress.

Naturally, these remarks only skim the surface, but they seem to me to be a fair if rather superficial picture of the state of the art as it has developed over recent years, at least from a personal perspective.

References


