The UCLA Lectures

(April 29 – May 2, 2019)

Noam Chomsky

with an introduction by Robert Freidin

Noam Chomsky's UCLA Lectures: an introductory sketch

Robert Freidin

Princeton University

The revolution in the study of language that began in the 1950s and continues today—what has come to be called *the generative enterprise*, the topic of Noam Chomsky's UCLA Lectures (April 29 – May 2, 2019)—is encapsulated in a single sentence published in 1957, a sentence that is with no exaggeration whatsoever the heart and soul of a research program which, as these lectures demonstrate, remains vibrant and promising well into its third quarter century. The first sentence of the first chapter of *Syntactic Structures* states simply and succinctly: *Syntax is the study of the principles and processes by which sentences are constructed in particular languages*. The radical nature of this definition is easily demonstrated by comparing it to the definitions of *syntax* that came before, all of which focus on linguistic phenomena (see Freidin (2020) for a detailed discussion; see also Freidin (2013)). The comparison highlights a conceptual shift in focus from mere description of data from languages to the principles and processes that underlie the syntactic properties of human languages, a shift that resulted from Chomsky's abandonment of the structuralist goal stating regularities in linguistic data and his adoption instead of what his teacher Zellig Harris characterized as "synthesizing utterances". Harris's *Methods in Structural Linguistics* (1951) mentions both, but opts for the first without any further discussion of the second. (See Freidin (2013) for discussion).

This conceptual shift to principles and processes aligns linguistics with the natural sciences, which since the Scientific Revolution in the 17th century have been concerned with understanding the underlying principles and processes that animate the natural world. (For additional discussion of the conceptual shifts that occurred to initiate the generative enterprise as well as those that have occurred within in it, see Chomsky (1983) and Freidin (1994).)

Chomsky's focus on principles and processes, a constant in his work for more than six decades, is demonstrated once again in these lectures, which address fundamental questions for the study of human language, including the perennial *what is language?* and *what is the nature of the generative enterprise?* as well as a new one about what would constitute genuine explanation in the study of language. As stated at the beginning of the second lecture, the generative enterprise is concerned with *the basic property of language*, which Chomsky characterizes as "namely that each language constructs in the mind an infinite array of structured expressions each of which has a semantic interpretation that expresses a thought, each of which can be externalized in one or another motor system, typically sound, but not necessarily." As

formulated, the basic property of language must be a significant part of the answer to the question *what is language?*. And that answer must crucially involve the principles and processes by which sentences are constructed, for how else are we to account for the infinite array of structured expressions each with an interpretation that expresses a thought?

This formulation also tells us that language as structured expressions in the mind is a mental phenomenon and thus internal to speakers, but with avenues for externalization so that the thoughts these expressions represent can be communicated to other speakers. Therefore the basic property of language is a property of humans, a biological property unique, as it turns out, to the species. For this reason, a program of research with the goal of providing an explanatory account of the basic property interprets the generative enterprise as *biolinguistics*.

As discussed in the first lecture, there is a strand in the history of ideas going back to the Scientific Revolution of the 17th century with Galileo and his contemporaries, who (though their formulations differ in some essentials from Chomsky's (see Chomsky (2017b) for discussion)) recognized the basic property of language and found it both amazing and puzzling. In 1632 Galileo expressed this in the following (cited in Chomsky (2017a), which gives both the reference and the original Italian):

> "But surpassing all stupendous inventions, what sublimity of mind was his who dreamed of finding means to communicate his deepest thoughts to any other person, though distant by mighty intervals of place and time! Of talking with those who are in India; of speaking to those who are not yet born and will not be born for a thousand or ten thousand years; and with what facility, by the different arrangements of twenty characters upon a page!"

In essence, Galileo was expressing amazement for how thoughts could be externalized in an alphabet and thereby transmitted across space and time. Twenty-eight years later, Antoine Arnauld and Claude Lancelot gave the following characterization in their *Grammaire générale et raisonnée de Port-Royal* (see again Chomsky (2017a)):

> It remains for us to consider what is, in fact, one of the great spiritual advantages of human beings compared to other animals, and which is one of the most significant proofs of reason: that is, the method by which we are able to express our thoughts, the marvelous invention by which using twenty five or thirty sounds we can create the infinite variety of words, which having nothing themselves in common with what is passing in our minds nonetheless permit us to express all our secrets, and which allow us to understand what is

not present to consciousness, in effect, everything that we can conceive and the most diverse movements of our soul.

Here we shift from the alphabet to the sounds it represents and from which "an infinite variety of words" can be created, a formulation which links language with infinite creation. What's missing in these formulations is any reference to the infinite array of structured expressions that make the expression of thoughts and feelings possible. Nonetheless, in recognition that Galileo appears to have been the first to recognize what is essentially the basic property of language and to be amazed and puzzled by it, Chomsky has characterized the goal of the generative enterprise to provide a principled explanation for this property *the Galilean challenge.*

What might constitute a principled explanation for some aspects of the basic property is also a focus in these lectures, what Chomsky is now calling *genuine explanation*. In these lectures, a genuine explanation for properties of language, as opposed to descriptions of them, must meet two austere requirements: show that the property is acquired by individuals (learnability) and show that property could have been acquired by the species (evolvability). And for any property under investigation there are three factors to be considered: one, what is genetically determined (innate) and thus universal across the species (Universal Grammar UG); two, the data available for language acquisition (generally impoverished because limited to externalizations of the internal structured expressions); and three, third factor principles not specific to language that may in part help shape the property investigated.

A genuine explanation of properties of language instantiated in linguistic phenomena (for example, structure dependence (discussed in some detail in lecture #3)) must start with the principles and processes with which sentences are constructed, what has come to be called *the computational system* for human language. A computational system consists of the operations for constructing the internal structured expressions of the language including whatever general principles govern the derivations and representations these operations produce in conjunction with a lexicon for a particular language. To the extent that these operations and principles apply to lexicons of all languages, they are part of UG and therefore meet the condition of learnability because they are an intrinsic part of human biology, not something that is learned on the basis of external evidence.

In these lectures Chomsky discusses the state of the generative enterprise: what has been accomplished, what problems remain, and what it might be possible to accomplish in the immediate future. The entire discussion is informed by the Strong Minimalist Thesis (SMT), essentially the hypothesis that the internal computational system for human language, which

generates structured expressions that interface with the Conceptual-Intensional systems of the brain, is a "perfect system" (Chomsky (1995, p. 1). On the same page, Chomsky characterizes what might be considered a perfect system as follows:

> This work is motivated by two related questions: (1) what are the general conditions that the human language faculty should be expected to satisfy? and (2) to what extent is the language faculty determined by these conditions, without special structure that lies beyond them? The first question in turn has two aspects: what conditions are imposed on the language faculty by virtue of (A) its place within the array of cognitive systems of the mind/brain, and (B) general considerations of conceptual naturalness that have some independent plausibility, namely, simplicity, economy, symmetry, nonredundancy, and the like?

and conjectures that "to the extent that the answer to question (2) is positive, language is something like a "perfect system," meeting external constraints as well as can be done, in one of the reasonable ways."

The first part of the discussion of what has been accomplished in the generative enterprise reviews how the problematic early proposals of the 1950s were eventually replaced with a computational system that could serve as a basis for genuine explanations of linguistic properties and phenomena. The earliest proposals of the 1950s employed two distinct kinds of operation (called *rules*): phrase structure rules to capture composition (how the constituent parts of sentences were hierarchically organized) and transformations to capture what later came to be called *dislocation* (or elsewhere (including in these lectures) *displacement*), where a single syntactic object is interpreted as if it occupies multiple syntactic contexts although in phonetic form it only occurs in one. As Chomsky notes, both operations were "much too complex to meet the conditions of learnability or evolvability", and therefore cannot provide a basis for genuine explanation.

The specific problems with phrase structure rules are one, they allow too many possible rules, including those which would never be proposed for obvious reasons—e.g. S➜P+VP, cited in lecture #1, and two, they conflate three distinct aspects of language which, it has become clear in recent years, ought to be handled separately: hierarchical structure, linear order, and projection (which determines what kind of element each syntactic element is).

These problems are partially resolved with the X-bar theory of phrase structure developed in the late 1960s. The first formulation in Chomsky (1970) replaces some of the phrase structure rules of the base component (base rules) with "schema", given as (1) and (2)

(1)   X' ➜ X …

(2)   X" ➜ [Spec, X'] X'

where X is a variable standing for the lexical categories N, V, and A; … represents the complements of X; X is the head of the phrase X'; and the number of primes (or bars above the category symbol in Chomsky (1970)) distinguish the level of the phrase headed by X. To the extent that these schema restrict possible phrase structure configurations, they are part of the solution for too many possible phrase structure rules. While the configurations specified in the schema (1) and (2) are endocentric, the term *endocentric* does not occur in Chomsky (1970) and the possibility that all phrase structure rules are endocentric is only mentioned in Chomsky and Lasnik (1993). However, in practice the endocentricity of phrase structure was extended to clause structures in the form of CP and IP/TP (see Chomsky (1986)) and was determined via projection, where phrasal categories (with one or more primes) were identified by a zero level category they contained as specified in the schema (1) and (2). (It is worth noting that the term *projection* or its corresponding verb do not appear in the original formulation of X-bar theory, the noun appearing for the first time in Chomsky (1976) and the verb in Chomsky (1986).)

As for conflating hierarchical structure with linear order and projection, X-bar still conflates the first with the third. Chomsky says that X-bar did not also conflate linear order, though this was not recognized at the time, citing how English and Japanese manifest mirror image linear orders but have the same X-bar theory, the same hierarchical structure and the same projection. With 2020 hindsight, this suggests that the schema in (1) and (2) might more accurately have been represented instead as (3) and 4)

(3)   X' : { X, … }

(4)   X" : { [Spec X'], X' }

where the rewrite arrows are replaced with a more neutral colon, suggesting that X' names the syntactic object { X, … } and X" names { [Spec X'], X' }. However these schema are represented, they cannot account for exocentric constructions, which these lectures consider to be a major defect of the theory. This wasn't recognized as a problem until recently (see also Chomsky 2019) and wasn't true for the original formulation of X-bar theory (Chomsky (1970)), which included the rule (5) for clause structure (i.e. Subject + Predicate).

(5)   S ➔ N'' V''

Thus clauses like *John proved the theorem* had exocentric structure in contrast to a corresponding derived nominal *John's proof of the theorem*, which was endocentric.

The problems that remain in X-bar theory are resolved with the formulation of Merge in the early 1990s. Instead of too many possible rules to account for compositionality, there is only one operation Merge. And this creates only hierarchical structure, not linear order or projection. And because it doesn't determine projection, it can create both endocentric and exocentric hierarchical structures depending on how projection is determined. Moreover, the operation Merge accounts as well for dislocation, thereby uniting composition and dislocation, which eliminates the need for the complex transformational rules of the 1950s or even an additional transformational operation Move α of the 1980s. When Merge applies to two unconnected syntactic elements, we get composition via what's called external Merge. And when it applies to two syntactic elements, where one is contained within the other, we get dislocation via internal Merge—the same operation, two different cases for its application. Note further than internal Merge is also a compositional operation that creates new hierarchical structure by merging the element contained with the syntactic object that contains it, creating a new syntactic object.

The third lecture turns to some problems with the formulation of Merge that are revealed by considering how the operation interacts with the workspace (WS) that contains the objects constructed by Merge and the atoms from which those objects are constructed (mentioned but not discussed in Chomsky (2013)). By focusing attention on this essential aspect of derivations under Merge, what was basically assumed but not considered, Chomsky is able to formulate a more restrictive concept of Merge, called MERGE, where the operation becomes a mapping between workspaces in which new syntactic objects are constructed. For example, in the case where two lexical atoms a and b are merged, MERGE takes the WS = [a, b] (using square brackets to designate the workspace) and changes WS into WS' = [{ a, b }, ...].

How MERGE is a more restrictive formulation than previous versions of Merge becomes clear when we consider what ... contains. If MERGE is like normal general recursion (e.g. proof theory), then WS' will contain a and b as well as { a, b }. But as Chomsky points out, this makes it possible to generate illegitimate structures that violate well established linguistic constraints. Thus { a, b } could be expanded to any island construction X and then if the WS containing X contains either a or b as a separate element, either could be merged with X creating a chain with a or b contained in X, violating the island constraint. Therefore the formulation of MERGE must exclude this possibility, which demonstrates that recursion for language differs from recursion in general—it's more restrictive.

This difference between recursion for language and recursion in general is ascribed to a restriction on resources available for computation by MERGE, what Chomsky postulates as a condition on resource restriction prohibiting each application of MERGE from generating more than one additional accessible element in a workspace. If WS' = [{ a, b }], then there are three accessible elements ({ a, b } plus a and b), where in contrast WS = [a, b] contained two. But if WS' = [{ a, b }, a, b] then the increase in accessible elements would be five, in violation of the resource restriction.

This raises an apparent problem for internal MERGE where WS = [{ a, b }] is mapped onto WS' = [{ b, { a, b }}]. WS contains three accessible elements, where WS' would contain five if each element b is counted separately. However, as Chomsky points out, given minimal search b in { a, b } would not be accessible; so in this way internal MERGE satisfies the resource restriction, which Chomsky conjectures "probably reduces to a general third factor property of the nature of the brain".

Another potentially problematic issue with internal MERGE also concerns the creation of copies, considered unique to this operation. The existence of copies raises two questions: a) how does the computational system distinguish copies from non-copies (sometimes referred to as "repetitions" especially when two or more apparently identical syntactic objects show up in phonetic form)? and b) how does the computational system determine what is a copy of what? In answer to the first question, the lectures turn to the concept of *the duality of semantics*: argument structure vs. discourse-oriented/information-related and scopal properties. External MERGE generates only argument structure whereas internal MERGE generates only the other properties. Given the duality of semantics, syntactic objects that are interpreted as having an argument function (θ-role) but occur in non-argument positions are copies of identical syntactic objects that occur in argument positions. This assumes that copies are governed by a principle of Stability by which they must be syntactically and interpretatively identical ("a general property of computations"). And conversely, two syntactic objects in argument positions even when they appear to be identical are never copies (thus in *John saw John*, neither *John* can be interpreted as a copy of the other). Chomsky suggests that the answer to the second question comes from "some internal conspiracy about the nature of language"—"there are answers, but it's not trivial" and leaves it as "something to think through".

In these lectures Chomsky proposes an alternative conception for the operation MERGE in which both external and internal MERGE create copies. Thus when external MERGE generates { a, b } from [a, b], it copies a and b and merges them as { a, b }. The copies of a and b are eliminated by the resource restriction, so in effect the element { a, b } replaces a and b in WS'. In

the case of internal MERGE that generates { *b*, { *a*, *b* }}, the element { *a*, *b* } as well as the element *b* are copied, and these are replaced by the element { *b*, { *a*, *b* }}, again a result of the resource restriction. This conception underscores the fact that internal and external MERGE are indeed the same operation, the operation that produces copies.

The resource restriction, which limits the operation MERGE, also rules out various other applications of Merge including Parallel Merge, Sideward Merge, and Late Merge, which in addition are also compatible with derivations that violate island constraints as discussed above. The lectures credit the interesting results that have been obtained by using these applications, but suggest that these results are only useful ways of organizing data, revealing puzzles to be solved, not explanations. The end of the third lecture gives an example of this where Chomsky reanalyzes a simple case of across-the-board deletion (ATB), which has been described using multidominance and Parallel Merge, in terms of copy deletion and straightforward general principles, including Stability.

The lectures make a powerful conceptual argument for MERGE: it is the simplest computational operation, simpler than recursion in general and simpler than previous versions of Merge that have been explored. If it is the simplest computational operation, then the question of learnability is answered: there is nothing to learn because there's no linguistic data that could determine it; it's just part of an innate faculty of language and the null hypothesis for a computational system for language. Similarly, the question of evolvability is answered: the operation couldn't have evolved from something simpler and without such an operation, language wouldn't exist – that is, this is the simplest operation for language to exist at all. If properties of language like structure dependence (as discussed in the second lecture) can be shown to follow from MERGE, then we have achieved the goal of genuine explanation.

Chomsky's account of structure dependence renders the conclusion that the externalization of language involving linear order is not part of the core, a conclusion that is supported by experimental evidence (reviewed in the lecture) from language acquisition, neurolinguistics, and psycholinguistics. If this conclusion extends to externalization of language in general, then what is a perfect system concerns only the internal computation and the structured objects produced that interface with the C-I systems. So I-language is not only internal to the individual, but limited to what is computed internally. Externalization of language is something else, the result of being forced to interface the internal system with sensory-motor systems. The lectures propose, for example, that the head parameter is the result of "a mismatch between the internal system and the sensory-motor system" where hierarchical structure must be forced into a linear order. That linear order just has to be one way or the other, hence the

parameter. Chomsky suggests trying to demonstrate that all parameters have two properties: a) they are exclusively part of externalization, with no role to play in the interface between the internal language and C-I systems, and b) they are simply options left open for externalization that have to be decided one way or another and therefore don't evolve.

The contrast between the internal language and its externalization leads to Chomsky's conjecture that the communicative inefficiency that shows up in the externalized language in the form of parsing, perception, and communication problems (specifically, with filler-gap constructions, structural ambiguities, and garden-path constructions) is strong evidence that language, based on the simplest computational operation MERGE, is designed to be computationally efficient (even when this results in communicative inefficiency).

The computational efficiency of language will of course be determined on the basis of what other operations aside from MERGE the computational system utilizes. The fourth lecture examines a long-standing problem in generative grammar concerning "unbounded unstructured coordination", illustrated in (6).

(6)   I met someone young, happy, eager to go to college, tired of wasting time, …

The problem with these constructions, first discussed in Chomsky & Miller (1963), was characterized as follows:

> In order to generate such strings, a constituent-structure grammar must either impose some arbitrary structure (e.g., using a right-recursive rule), in which case an incorrect structural description is generated, or it must contain an infinite number of rules. Clearly, in the case of true coordination, by the very meaning of this term, no internal structure should be assigned at all within the sequence of coordinate items.

The fourth lecture proposes a solution to the problem that introduces a second computational operation Pair-Merge, distinct from MERGE (which forms sets, hence Set-MERGE). Chomsky considers Pair-Merge to be the next simplest operation after MERGE.

The main general idea is that Pair-Merge is required to handle sequences. Exactly how this is done is left open with the proviso: "There are a number of possible ways to implement these general ideas, but to go into them here would carry us too far afield. They are a topic of current research." However the discussion in lecture #4 provides a few details that are suggestive. The notation for Pair-Merged elements employs angle brackets, the standard notation for sequences, not the curly brackets of Set-MERGE that are the standard notation for sets. So for a sentence like *I met someone young*, the noun and the adjective would be Pair Merged as *<someone, young>* on the assumption that adjuncts are Pair-Merged, not Set-

MERGEd. Furthermore, neither element that is Pair-Merged is accessible to further operations (e.g. internal MERGE) and therefore might provide a principled account of adjunct islands (though, as Chomsky cautions, the facts are complicated). But the syntactic object *<someone, young>* will be merged with the verb *met* to form the set { *met, <someone, young>* }.

Another extremely intriguing suggestion is that because internal structures need not be limited to two dimensions (like a computer screen), adjuncts can occupy other dimensions, as many dimensions as there are adjuncts in an expression, so that each adjunct can directly associate with the element it modifies. Chomsky suggests that this is also true for constructions in which there is only one adjunct (e.g. *someone young*). Lecture #4 extends the discussion to coordinate structure with conjunctions.

The fourth lecture goes on to investigate the possible scope of Pair-Merge beyond coordination that might include puzzles about perception verbs and quasi-causative verbs (*make* and *let*) in conjunction with bare verbs (e.g. *they saw the man walk down the street* vs. **the man was seen walk down the street* & *they let the man walk down the street* vs. **the man was let walk down the street*) and persistent problems with head movement.

The fourth lecture continues with some comments about what the atomic elements used in computations are, a topic that moves the discussion "towards the general domain of semantics … and how this domain relates to the [generative -RF] enterprise". Chomsky points out that classical semantics as developed in the 20th century (from Frege to Quine and others) is based on notions of truth, reference and denotation which relate to the mind-independent world. In contrast, formal semantics in linguistics and philosophy, which Chomsky characterizes as "some of the richest and most exciting work going on in the field in the last couple of decades, as "pure syntax: symbolic manipulations of postulated entities that are not part of the mind-independent world, whatever their real-world motivation." (The same can be said of generative phonology, as the lecture discusses.)

The question that remains is how the mind-independent notions of truth, reference and denotation might connect to language, which would involve getting beyond syntax. The answer would have to involve words in the lexicon, but Chomsky demonstrates that garden variety words like *house*, *river*, and *London* do not refer. So for the case of *house*, he says:

> … a house is something that we construct in our minds, which has a material element, but a crucial part of it is what Aristotle called the form. That's something that's part of our mental operations. When we use the word *house*, we're referring to a mind-independent object which we interpret as a house. Referring is an

action, but the word *house* does not refer.

This is essentially the situation for any other word in the language. From this it would follow that "language simply does not have the concept of truth and reference at all, which means it doesn't have semantics." And this raises the question of where the words of human language came from? A real mystery given that the atoms of symbolic systems other animals employ in communication seem to correlate with identifiable physical events.

The lectures concludes with some comments about two kinds of mystery, those that could be solved if we had the appropriate evidence (though that evidence may not be available to us) and those for which we have no good ideas even what the appropriate evidence would be let alone where to look for it (and so a deeper kind of mystery). For language, the latter type of mystery concerns questions of the choice of actions— specifically, how an individual chooses a particular linguistic expression from the infinite array of expressions an I-language generates, a choice that is free from stimulus control, and both coherent in and appropriate to the situation in which an expression is uttered—what Chomsky has labeled *the creative aspect of language use* (see Chomsky (1966, 2006)). And it is the recognition of this mystery concerning the basic property of language that puzzled and amazed Galileo and his contemporaries back in the 17th century.

There is one aspect of the basic property of language that is no longer a mystery thanks to the generative enterprise—namely the basis for the infinite array of linguistic expressions available to speakers of a language. The solution is simply and succinctly the principles and processes by which linguistic expressions are constructed in a language (part of a speaker's unconscious knowledge), principles and processes that are finite in number but infinite in scope.

At the very beginning of the generative enterprise—in the first paragraph of the first chapter of *Syntactic Structures*, Chomsky enunciates one of the ultimate goals of the study of syntax as follows: "The ultimate outcome of these investigations should be a theory of linguistic structure in which the descriptive devices utilized in particular grammars are presented and studied abstractly, with no specific reference to particular languages." These lectures show us how far the generative enterprise has progressed towards realizing that goal—from language particular phrase structure rules and transformations to MERGE (and Pair-Merge), where these "descriptive devices utilized in particular grammars are presented and studied abstractly". Nonetheless, as the

discussion in the lectures of unbounded unstructured coordination shows, the study of underlying processes must be informed by an attention to the particular phenomena that are found in languages. The situation is reminiscent of the one in mathematics noted in Courant (1937).

> The point of view of school mathematics tempts one to linger over details and to lose one's grasp of general relationships and systematic methods. On the other hand, in the 'higher' point of view there lurks the opposite danger of getting out of touch with concrete details, so that one is left helpless when faced with the simplest cases of individual difficulty, because in the world of general ideas one has forgotten how to come to grips with the concrete. The reader must find his own way of meeting this dilemma. In this he can only succeed by repeatedly thinking out particular cases for himself and acquiring a firm grasp of the application of general principles in particular cases; herein lies the chief task of anyone who wishes to pursue the study of Science.

And for the science of language as discussed in these lectures, Chomsky demonstrates how this continues to be a process of discovery that results from rethinking over and over the conceptual foundations and empirical basis of the field.

References:

Chomsky, Noam. 1966. *Cartesian Linguistics: a chapter in the history of rationalist thought.* Harper and Row.

Chomsky, Noam. 1970. Remarks on Nominalizations. In R. Jacobs and P. Rosenbaum (eds.), *Readings in English Transformational Grammar.* 184-221. Ginn & Co.

Chomsky, Noam. 1976. Conditions on Rules of Grammar. *Linguistic Analysis*, 2, 303–351.

Chomsky, Noam. 1983. Some Conceptual Shifts in the Study of Language. In L. S. Cauman, I. Levi, C. D. Parsons, and R. Schwartz (eds.), *How Many Questions?: Essays in honor of Sidney Morgenbesser.* 154-169. Hackett Publishing Company, Inc.

Chomsky, Noam. 1986. *Barriers.* M.I.T. Press.

Chomsky, Noam. 1995. *The Minimalist Program.* M.I.T. Press.

Chomsky, Noam. 2006. *Language and Mind.* 3rd edition. Cambridge University Press.

Chomsky, Noam. 2013. Problems of Projection. *Lingua* 130, 33-49.

Chomsky, Noam. 2017a. The Galilean Challenge. *Inference: International Review of Science*, vol. 3, No. 1 (https://inference-review.com/article/the-galilean-challenge).

Chomsky, Noam. 2017b. The Galilean Challenge: Architecture and Evolution of Language.

Journal of Physics: Conf. Series 880 (2017) 012015.

Chomsky, Noam. 2019. Some Puzzling Foundational Issues: The Reading Program. *Catalan Journal of Linguistics* (Special Issue), 263–285.

Chomsky, Noam. and George Miller. 1963. Introduction to the Formal Analysis of Natural Languages. In R. D. Luce, R. Bush, and E. Galanter (eds.), *Handbook of Mathematical Psychology*, vol. 2, 269-322. Wiley.

Courant, Richard. 1937. *Differential and Integral Calculus*. Vol. 1. Blackie and Son.

Freidin, Robert. 1994. Conceptual shifts in the science of grammar 1951 – 1992. In C. P. Otero (ed.) *Noam Chomsky: Critical Assessments*, vol. 1, tome 2, 653–90. Routledge. Reprinted in Robert Freidin. 2007. *Generative Grammar: Theory and its History*. 266–298. Routledge.

Freidin, Robert. 2013. Chomsky's contribution to linguistics: A sketch. In Keith Allan (ed.), *The Oxford Handbook of the History of Linguistics*, 439–67. Oxford University Press.

Freidin, Robert. 2020. *Syntactic Structures*: a radical appreciation. LingBuzz (https://ling.auf.net/lingbuzz/004996).

Harris, Zellig. 1951. *Methods in Structural Linguistics*. University of Chicago Press.

<center>The UCLA Lectures</center>
<center>Noam Chomsky</center>

Preface (October 2020)

These lectures[1] seek to evaluate the state of the generative enterprise within the biolinguistic framework, considering its past, present, and possible future. They present a personal perspective, with only limited attention to other approaches to the study of language and mind.

It may be useful to briefly situate the generative enterprise in the immediate context from which it emerged in mid-20th century. There was, at the time, a general consensus about the achievements of 20th century structural linguistics.[2] Synchronic linguistics had been established as an independent science, a "taxonomic" science, with sophisticated procedures of analysis and some fundamental principles, notably the phonemic principle. There was also a largely shared framework of understanding of the place of language in the larger world: language is an array of habits, extended in usage by analogy, acquired by training and general procedures of habit formation. In many circles the consensus was supplemented with anticipation that the new sciences of communication and radical behaviorism would provide a firm basis for further progress.

The generative enterprise grew from skepticism about the solidity and reach of these achievements. Many of the assumptions seemed dubious at best. The procedural basis seemed too constrained along with inherent distortions. The drive to understand seeks explanation, beyond taxonomy, often yielding very different conceptions of organization of phenomena, a record familiar from the biological sciences.

The initial efforts 75 years ago to construct explanatory theories (generative grammars) quickly revealed that the entities that enter into them lie beyond the reach of taxonomic procedures and often disregard what they yield, though it took a few years for the import of these discoveries to be understood. The procedures themselves were found to have serious flaws, even more so the general framework. Habit, analogy, conditioning and similar concepts cannot begin to account for the nature, acquisition and use of language.

The new mathematical theories of computation made a substantial contribution to the generative enterprise. For linguistics, these discoveries provided the first clear concept of a generative grammar, a

---

[1] The lectures were delivered in the linguistics department at UCLA in April/May 2019 (https://linguistics.ucla.edu/noam-chomsky). They have been carefully transcribed and edited by Robert Freidin, who also contributed valuable suggestions for clarification and improvement, leaving the substance essentially unchanged.
[2] For further discussion see Chomsky (2021b), and from a more personal perspective Chomsky (2021a).

system that provides "a recursive specification of a denumerable set of sentences," ultimately in phonetic form,[3] later sharpened to yielding what is called "the Basic Property" in these lectures and elsewhere.

More generally, the theory of computation enabled the first clear formulation of Aristotle's crucial distinction between "the possession of knowledge and the actual exercise of knowledge" (*de Anima*), between competence and performance, in modern terminology, the former being the essential element. The "Galilean challenge" discussed in the lectures kept to the "exercise of knowledge," production and perception, which access the knowledge that is possessed and of course involve other faculties as well.[4] The same limits hold for the rich tradition of "general and rational grammar" that developed from the Galilean challenge: possession of knowledge was ignored. Humboldt's by now famous aphorism that language involves "infinite use of finite means" keeps to *use of knowledge*, not the more fundamental notion of *possession of knowledge*, which can be accessed and used in various ways. The same holds more generally into the modern period, and in many ways still does.

With these concepts available it becomes possible to move from the "closed world" of taxonomic science to the "infinite universe" of search for explanation, to paraphrase the title of a famous work on history of science. In the closed world of taxonomic science, there remains little to do beyond applying existing tools to more data. In the "infinite universe" of search for explanation, even the rudiments are little understood and each new discovery raises new and exciting challenges. The lectures below discuss some of the ways these challenges have been explored in the generative enterprise.

Lecture #1 (April 29, 2019)

What I would like to do in these lectures, which is actually just one continuous talk broken up into parts, is take them as far as I can up to contemporary work and problems, if we make it. I would like to discuss the state of the generative enterprise as it's been called by some of its leading practitioners: what's been accomplished, what the problems are, and what we can hope to see in the future. From the origins of this initiative, which incidentally revived a tradition that had long been forgotten and that was unknown at that time. But from the origins, the holy grail was genuine explanations of fundamental properties of human language, of the faculty of language. And that's not such a simple matter to capture properly and to the extent you can, it's been an elusive goal. And I think present moment is unusual in the long history of the field, twenty-five hundred years. And that goal, I think, seems perhaps within reach. And if that's the case, it would be a matter of no slight significance not just for linguistics, but beyond. These are the questions I'd like to explore in this extended lecture.

---

[3] Chomsky (1949).

[4] The study of production, furthermore, is crucially limited by the inability – which persists -- to deal with the Cartesian property of "creative use of language."

So to begin with, we have to clarify some basic questions, but highly contested questions about what the field is about. What's the nature of the enterprise? I've personally always found it helpful to rethink these matters over and over. I hope you will too.

So let's begin with what sounds like the simplest question—namely, "what is language?". Well, that question is plainly consequential. The answer to it will determine what we focus on, what kind of work we do, how we proceed, what counts as a result, and critically what counts as an actual explanation, a genuine explanation. There've been many proposed answers over the years. They differ in interesting ways, and–if we think about it a little–the question turns out to be not so simple. So suppose, for example, we asked the question in some other discipline, let's say physics. We ask: what is the physical world? What is energy, what is mass, what is work? Any such question. The answer that we'll get is some technical definition internal to explanatory theory. So we won't get an account of what people intuitively think of as the physical world or think about energy and so on. That's not to the point. We'll find answers within a particular explanatory theory. Suppose we ask biologists "what is life?". There it will be a little bit more ambiguous because the theoretical understanding has not reached the point where it's obvious what the essential conceptual notions are. So it's exploratory. Suppose we ask "what is thinking?". Well, here it gets a little more complicated. Actually as you know, the question was posed by Alan Turing in a famous paper in 1950 which initiated the field of artificial intelligence and papers about whether machines think. And he starts off by saying that the question is too meaningless to deserve discussion.[5] So he's not going to discuss it because the notion thinking is so vague and amorphous that you can't give a response in the manner in which you might in say physics or even biology. He's asked what thinking is, he says it's some kind of buzzing in the head, but nothing much more to say than that. So what he does is something quite different. He proposes a notion which he says, might be somewhere within the range of what people call thinking and maybe it's useful notion. He suggests that it is, and in particular it might stimulate the development of new software, new machines. That's the famous imitation game, the so-called Turing Test.

Let's move on. Notice that when you ask the question what is thinking?, what is language?, what is meaning?, what is belief?, and so on, the answers that you get are really what philosopher Charles Stevenson once called *persuasive definitions*, saying here's what I think is interesting in the general domain of this loose notion. Here's something I think it's worth looking at. Well, you go back to the Turing test, notice that it's not an attempt to explain and understand anything about thinking. It's about an attempt to simulate some of the aspects of thinking. That's a quite crucial difference. It didn't seem so crucial in

---

[5] Turing, Alan. 1950. Computing Machinery and Intelligence. *Mind* 59:236,433–460. Cf. Chomsky, Noam. 2009. Turing on the "Imitation Game." In Epstein, Robert, Gary Robert and Grace Beber (eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, 103–106. Springer.

Turing's day, but now it's highly crucial since a good part of what goes on in the study of language and cognitive science—up-north the Silicon Valley version of this—is basically simulation, not efforts to understand and explain. And in fact, that's the direction that AI and deep learning have taken. There's a lot to say about that, but I'll put it aside unless it comes up later.

When we ask about these things, we're basically told "here's what I think is interesting". Ok, then the next question is "is it interesting?". Is it a sensible choice? And if it is a sensible choice, how can you proceed to place it within the framework of some kind of explanatory theory? And insofar as you can do that, you can discuss the validity of the concept that's proposed. Other than that, there is lively debate about what is language, what is meaning, what is belief. But it's basically "here's my preference". It's not a clear thing that you can give an answer to. You can ask if the preference is sensible one, can we develop it, and so on. But there aren't questions of sort of validity or invalidity. Sometimes it's useful to develop a new technical term to make that clearer; that's what I'll be doing here.

So let's go back to the question "what is language?" and have a look at some of the preferences over the centuries. I think if you look, you can roughly say that they fall into two major categories. One approach to "what is language?" considers the concept that we're focusing on to be something internal to a person. So my language is something that's up here [head/brain]. It's the buzzing that goes on up there in Turing's terms. That's one concept of language. The other concept of language is it's something external to any person, which people make use of somehow. Terminologies are often imprecise, but I think you can roughly see this distinction. And what kind of work we do and how it's evaluated—all of that is going to depend crucially on which of these enterprises is undertaken.

A classic illustration of the first kind, language as an internal object, is, I think, the great linguist Otto Jespersen about a century ago, one of the clearest exponents of this notion. He was the last representative of a long tradition. For Jespersen in his *Philosophy of Grammar*, I'll quote him, a particular language is a system that "come[s] into existence in the mind of a speaker" on the basis of "innumerable sentences heard and understood." This internal system in the mind yields a "notion of their structure which is definite enough to guide him in framing sentences of his own", crucially what "free expressions" that are typically new to the speaker and the hearer. And there's a more general concern of linguistic theory: to discover what is universal to language. Working within the tradition, Jespersen wrote that while "we must not expect to arrive at a 'universal grammar' in the sense of the old philosophical grammarians [based on logic],..What we obtain is the nearest approach to it that modern linguistic science will allow." Here "universal grammar" is understood in the sense of generally valid properties of language (like Greenberg's universals), not UG in the contemporary sense: not generalizations about languages, but the principles that underlie them, again "in the mind of the speaker," indeed humans generally. That's the first approach, which regards language as a property of persons.

The second approach is illustrated by the structuralist, behavioralist approaches to language of the first half of the twentieth century—still, of course, continuing. That took language, the object of study, to be, say, a corpus of materials that a field worker would elicit from an informant, or perhaps a set of sentences, or some other entity that's external to people. So if you look at the actual formulations, for de Saussure, the founder of the structural linguistics, a language is a kind of a social contract in a community, some collection of word images in the minds of the people of the community. Go to the leading American linguist of the early half of the twentieth century, Leonard Bloomfield, one of whose answers to the question "what is language?" is that "language is the totality of utterances that can be made in a speech community"—so something out there. Go to philosophy of language. For W.V.O. Quine, perhaps the leading and most influential philosopher of language in the mid twentieth century, a set of "significant sequences, [which,] being subject to no length limit, are infinite in variety" ("The Problem of Meaning in Linguistics").

David Lewis, another influential philosopher took the same view in his influential article "Languages and Language": language is an infinite set of sentences used by a population. Both Quine and Lewis concluded that while it makes sense to say that a population uses this infinite set, it doesn't make any sense to say that there's a particular way of characterizing the set. To look for that, Quine said, would be "folly." Lewis wrote that he could make no sense of the notion. If that's what language is, what's linguistics? Well, linguistics, naturally, be a way of taking data, however you get it, typically from an informant, by applying various procedures and methods to get an organized form of that data. The most sophisticated version of this was (as Tim Stowell mentioned) Zellig Harris's *Methods in Structural Linguistics*. In Europe, Trubetzkoy's *Principles of Phonology* was constructed on similar grounds. Well, this characterizes almost completely the structuralist-behavioralist approach to language.

There is something kind of paradoxical about it. So what are these entities? What is the set of sentences spoken in a speech community? How can members of population use an infinite set unless they have some way of determining what's in the set or out of the set? In fact, how can we even coherently talk about an infinite set unless we have a method that characterizing it? So it seems to me at least that the approach of leading philosophers and logicians was kind of confused.

It's really the opposite. If you want to talk about infinite sets, you first have to discuss what the internal mechanism for characterizing that set is—what's been called an I-language (I for internal). Well, whatever these ideas are supposed to mean from the structuralist, behavioralist period, which I think is not easy to answer, but whatever it is, there's something external to people, which people have some relation to. Now that by no means has ended; it continues right up to the present. There are strong currents that take very similar views, and I think one can ask the same questions about them, including within, roughly speaking, the generative enterprise.

Suppose instead, we adopt Jespersen's view, then the linguist is studying something that's in the mind of the speaker —namely, the mature state that has been attained, that has, in Jespersen's term "come into existence", and also the innate endowment of the speaker, the faculty of language, which first of all determines the principles that underlie the grammars of all languages and also makes possible the transition from finite data to the state attained, to the I-language in modern terms. As I said, the mature state attained is called the I-language (internal language in technical terms), and the innate principles are nowadays called universal grammar (UG), taking a traditional term and adapting it to a new context. The letter "I" in I-language is convenient. It refers to the fact that the internal language is first of all internal, secondly it's individual, and thirdly it's intensional (with an "s"). We're interested in the actual procedure, the actual algorithm, not the set of things that it enumerates. So for example, if you're studying, say, a person's knowledge of arithmetic, you want to know exactly how that person carries out addition. You're not talking about sets of triples X, Y, Z, such a Z is the sum of X and Y. Here too, we want to understand the generative system in intension.

Well, I should say that with regard to universal grammar, there's a good deal of confusion which exists right up to the present and which is worth dissolving. It's very common to hear that UG has been refuted or that it doesn't exist. What people presumably mean by that is that generalizations about language have exceptions, which is, of course, true. That's true of generalizations, but that's not what UG is about. UG in the contemporary sense is about the innate endowment that enables this transition that Jespersen talked about from finite data to the concept of structure in the mind. The concept of structure is what we call the I-language. So it should be clear that to deny the existence of this is not debatable. That would be senseless. If it doesn't exist, language acquisition is magic.

There is a kind of coherent version of this common claim, see Tomasello and many others.[6] A coherent version would be to claim that there is some general learning mechanism, which has nothing specific to do with language. Or maybe some collection of cognitive capacities which integrates somehow to make it possible to achieve the properties of language, what the faculty of language did.

There are a couple of problems with these proposals. One problem is simply that they reduce to hand waving. Or if they're made at all explicit, they're very quickly refuted. A second problem is that you can expect in advance that they're not going to work for reasons that were discussed by Eric Lenneberg in his classic book on the biology of language fifty years ago, in which he pointed out that there are doubled disassociations between language and other cognitive processes.[7] This work has since been greatly extended. Susan Curtiss is the person who's done the most extensive work on this, and in fact there are

---

[6] Cf. Tomasello, "Universal Grammar is Dead,"*Brain and Behavioral Science* (2009).

[7] Lenneberg, Eric. 1967. *Biological foundations of language*. New York: Wiley.

many examples of cognitive capacities intact, but not language, and conversely. So it's pretty clear in advance that it's not going to work, but it's nevertheless a widely held view.

So I think it doesn't really make sense as far as I can see to claim that there's some problem with UG.

Well, go back to Jespersen, the position that I want to continue with. Jespersen was the last representative of a very interesting tradition, which goes back to the seventeenth century Scientific Revolution, which set the course of modern science in a sharply new direction. The great thinkers of the seventeenth century, Galileo and his contemporaries, simply refused to accept what happens around them as being naturally self-explanatory, not requiring explanation. They recognized that the phenomena of nature were puzzling, mysterious, and demanded explanation—whether it was objects falling to the ground or perception of a triangle or anything else. That willingness to be puzzled about phenomena was actually something pretty new. Though it had happened with the Greeks, there was a kind of a dark ages that followed, but it was revived by the seventeenth century thinkers. And as soon as they began to look around them, they found that everything was really puzzling about things that seemed obvious and therefore didn't require explanation.

Actually, something similar happened in 1950s, if you go back to that period. Linguists generally assumed that everything was more or less understood about language. And that there was nothing general to say about language. A famous characterization by theoretical linguist, Martin Joos—what he called the Boasian principal, named after the great anthropological linguist Franz Boas—is that languages can differ in arbitrary ways and therefore each one has to be studied on its own without preconceptions. There's nothing to say about language except by applying analytic procedures to a corpus , which you could do. So essentially the field was terminal. Actually, I was student at that time and the general mood among students was this is fun, but what would we do when we've applied the procedures to all the languages? Then it's over. However, as soon as we began to look at the phenomena seriously, to try to construct actual generative grammars which would work, you found out that it's not that you understood everything, you understood almost nothing. Everything was a puzzle and it didn't seem as if there'd be any termination to the field, and that's what's happened since. By now the field has just exploded and the kinds of problems that students are looking at today couldn't have even been formulated let alone dealt with not many years ago—that's an enormous change.

Well, let's go back to the seventeenth century. Among the many phenomena that intrigued and puzzled Galileo and his contemporaries, one actually was language.[8] So they expressed awe and amazement

---

[8] Chomsky. 1966. *Cartesian Linguistics*. Harper & Row. See also (https://inference-review.com/article/the-galilean-challenge)

at a quite remarkable fact that, as they put it, with just a few symbols, a couple dozen symbols, it's possible to express an infinite number of thoughts and to convey to others, who have no access to our minds, all of the workings of our minds. And they asked how that magical accomplishment could be made. I like to quote their own words, which puts it evocatively, that they were awed by the method by which we are able to express our thoughts, the "marvelous invention by which using twenty-five or thirty sounds, we can create the infinite variety of expressions which, having nothing in themselves in common with what is passing in our minds, nonetheless permit us to express all our secrets, allow us to understand what is not present to consciousness [crucial point—NC], in effect everything we can conceive and the most diverse movements of our soul."

And if you stop to be willing to be puzzled, it is a pretty amazing fact. It by no means seems natural, we do it all the time, but it is quite amazing. Furthermore, there's nothing similar to it in organic world—which they recognized—which raises very crucial questions. How did this unique human achievement come about, and how can it be understood and explained? Well, for Galileo the alphabet was, he said, the most stupendous of all human inventions, comparable to the achievements of Michaelangelo and Titian—nothing like it. The reason was first that it captured this amazing property and secondly because it allowed us to express all the wisdom of the ages and beyond that it included the answers to any question that we could pose, all in this small collection of symbols. It was kind of like what we would call these days a universal Turing machine. The Port Royal grammar and logic which followed shortly after gave many serious insights into logic and linguistics, became the basic logic text for many centuries. It initiated a tradition in linguistics of what was called rational and universal grammar. Rational because it was seeking explanations, not descriptions. Universal because it was trying to find the principles that underlie all languages, Jespersen's great principles.

Well, the traditional formulations are not precise, but I think it's fair to interpret them as recognizing that the capacity for language as well as individual languages are internal properties of persons, as Jespersen says quite explicitly. And it was also generally assumed (without much evidence, but as we know now quite reasonably) that this capacity, whatever it is, is a human characteristic, shared among all human groups–there are no known group differences. And it is furthermore unique to humans in all the central respects; there's nothing analogous in the organic world, so it's a true species property and as they recognized, it's the foundation of human culture and human creativity.

Well, these ideas actually had a very substantial impact on philosophy and general intellectual culture mainly through the influence of Descartes, who adopt similar views at roughly the same time. Descartes's famous dualistic approach, the idea that in addition to the material world there is also mental world, was based very substantially on the recognition that this unique ability to create an infinite number of thoughts is somehow unique to humans. And it cannot be captured by machines. Machines for early

modern science (Galileo through Newton and beyond) meant the kinds of artifacts that were being created by skilled artisans and were proliferating all over Europe—very complicated artifacts that could do all sorts of intricate things. And their approach to science is: well, that's what everything is. It's called the mechanical philosophy (philosophy, of course, meant science). So mechanical science that's real explanation. The criterion for intelligibility for Galileo and his contemporaries was constructing or at least devising in principle a machine that could account for something. If you could do that, you had an intelligible theory. But Descartes recognized quite correctly that this amazing capacity of language couldn't be captured in those terms, so postulated a second substance *res cogitans* (thinking substance) which would capture this capacity somehow and be linked to the material world. That's Cartesian dualism.

This is all a scientific program perfectly sound science based on correct observations about the limits of mechanical objects. The Cartesian scientist took the natural next step, especially Jacques de Cordemoy, a leading Cartesian philosopher-scientist. De Cordemoy designed a series of experiments to determine whether some other creature could exhibit the capacities that a human can exhibit. This sounds sort of like the Turing Test, but with a crucial difference. Turing was trying to find something that would simulate aspects of human behavior. De Cordemoy was pursuing a scientific project. It's kind like a litmus test for acidity. Does some other entity, some other object or organism have a particular property. That's a similar looking project, but very different in character. This is a real science.

Well, what happened to these developments? Their fate is commonly misinterpreted. So often believed that science as it developed got rid of what Gilbert Ryle called "the ghost in the machine", the second substance. But what happened actually is the exact opposite. What happened is that Isaac Newton, much to his dismay, exorcized the machine, but he left the ghost intact. Newton showed that there are no machines, that the material world simply cannot be captured in mechanical terms because of interaction without contact, which is inconsistent with the mechanical philosophy. Newton himself regarded this result as a complete absurdity which no one with any scientific understanding could contemplate; he agreed with the other great scientist of his day, Leibniz and Huygens and others, that this was utterly absurd, but couldn't seem to find a way out of it. So the end result is we have theories like Newton's that we can understand, but no intelligible world. What they describe is simply unintelligible.

That was understood and science just changed: it stopped seeking intelligible accounts of an intelligible world and just moved to the weaker objective of finding intelligible theories of the world, which is quite different and quite unacceptable to early modern science. It's a major shift in intellectual history and it was understood, very well understood at the time. So shortly after, David Hume, who regarded Newton the greatest genius in history, wrote in his history of England, while "Newton seemed to draw the veil from some of the mysteries of nature, he showed at the same time the imperfections of the mechanical philosophy and thereby restored nature's ultimate secrets to that obscurity in which they every did and ever

will remain"—and they do in fact remain in that obscurity. Science just stopped looking for them after some period. John Locke, shortly after Newton's great treatise *Principia* appeared, carried the conclusion further in a highly consequential way. He expressed it within the theological framework of the day, but we can change the terms (the point is correct); he argued that the incomparable Mr. Newton as he called him had demonstrated that God had added to matter properties that are inconceivable to us, specifically interaction without contact, and so perhaps God had superadded to matter the capacity of thought, a property of certain kinds of organized matter. That's thought. That idea was pursued extensively through the 18th century into the early 19th century. Darwin mentions it in his notebooks. It was then forgotten completely and it's been revived in recent years as what's called a radical new idea in philosophy of mind. It's now a commonplace of the cognitive and brain sciences, picking up a forgotten tradition that followed directly from Newton's demonstration that there are no machines. That's a crucial part of intellectual history, not too well understood. And it's worth remembering that as Hume and Locke correctly recognized, Newton had in fact left issues in mysteries and obscurity—in which they remain. That's quite interesting question.

Well, let's put that aside and go back to the tradition of rational and universal grammar culminating in Jespersen. All of that was swept aside completely by the 20th century behavioralist-structuralist currents, which typically, in fact, I think, almost universally adopted the second approach I mentioned, taking language to be something external to people which people somehow grasped. The whole tradition was totally forgotten, still unknown, which is unfortunate, I think. There are lot of wealth and richness there. It was so forgotten that even Jespersen, a famous linguist in the early 20th century, was gone. There's an interesting article by a historian of linguistics Julia Falk, who reviews this and points out that even a major linguist like Bloomfield and others just knew nothing about him, essentially nothing.

Well, that general program that culminates say roughly from Galileo to Jespersen falls within the natural sciences. It was revived with the generative enterprise in the early 1950s. It's called the biolinguistics program, but it should be understood that this is only one current within the generative enterprise. Much of the ongoing work within the generative enterprise does not accept this internalist view, but that's the one I'll keep to.

Well the early efforts in the tradition ran into plenty of difficulties: empirical difficulties and conceptual difficulties. The empirical difficulties were there just wasn't enough evidence. What was understanding of language was pretty thin. The conceptual problem was that there was no way of really understanding this notion of concept of structure in the mind that enables this achievement of expressing an infinite number of thoughts to be captured. Well, it was somehow recognized that there is what we may call the basic property of language—reformulating it in our terms, a concept of structure in the mind is capable of generating an infinite array of structured expressions, each of which captures a thought (to the extent that we understand the notion thought) and each of which can be externalized in some sensory motor

modality, typically sound (but as we know now very well could be some other modality, could be signs (sign is virtually identical to speech), even with some reservations could be touch. So modality is basically irrelevant, a matter of some significance that I'll come back to.

Well, by the mid 20th century, the conceptual problems had been overcome. That's why the generative enterprise was able to take off and revive the tradition; Turing, Emil Post, Gödel of course, and other great mathematicians had given a precise and clear understanding of what became the theory of computability, which allows us to understand very clearly how it can be that a finite object (like the brain or your laptop for that matter) can capture within it the basic property. Now that's well understood, which means you can proceed with the enterprise that had lapsed with Jespersen. Well, you can deal with what sometimes I've called the Galilean challenge, the original formulation of what the field, I think, ought to be about. It's a persuasive definition again; you can decide whether you like it or not.

Well, if you want to meet the Galilean challenge, there are several tasks. In logical order, the first task is to try to discover the I-languages for languages of the widest possible typological variety. It's a huge task, of course. Second, to the extent you can do that, you can turn to the next problems, theoretical problems. The first one is to determine how a speaker of a language when he's producing a sentence selects a particular expression from the infinite set that's generated by I-language. The next question is how is that expression externalized in some sensory-motor system. A third question is the inverse of that for the hearer, how is the expression processed mapped from something, say, sound to an expression of the I-language. Well, the second and third tasks are input-output problems, kind of problems we know how to handle and in fact a great deal has been learned particularly about processing, also about the externalization of the internal object generated.

How about the first task? How does the speaker select a sentence, an expression from the infinite array generated by the I-language? That's another total mystery. There's nothing to say about it. In fact that's true of voluntary action generally, of which this is an instance. There is basically nothing to say about it. This fact is captured kind of fancifully (as they put it) by two of the leading neuroscientists who deal with voluntary action, Emilio Bizzi and Robert Ajemian, who have a recent review of the state of the art in the field of voluntary action.[9] What they say is we're beginning to understand the puppet and the strings, but we have absolutely nothing to say about the puppeteer. We can't say anything about why one or another action is selected, in particular that holds for the first task. So there's another mystery that so far is beyond, it doesn't even have bad ideas. There's nothing to say about it. Well, the I-language is clearly a property of

---

[9] Cf. Bizzi, Emilio and Robert Ajemian. 2015. A hard scientific Quest: Understanding Voluntary Movements. *Daedalus* 144:1, 83–95.

the individual, by definition, and the same is true of the faculty of language, although it's a shared property of humans with insignificant variations. It is in fact a property of each individual.

And the faculty of language faces two empirical conditions, two crucial conditions to be faced by any theory of the faculty of language for this internal system. One is the problem of learnability. Second is the problem of evolvability. So the faculty of language has to be rich enough so that it can account for the properties of all languages, and even more strikingly, for the remarkable property of this enormous leap from finite data to the internal system, which is carried out by the faculty of language which has to be rich enough to overcome the acute problem that's called the problem of poverty of stimulus, often unappreciated. So one demand on the faculty of language is that it has to be rich enough to achieve these goals. But it also has to be simple enough, so that it could have evolved, meeting the evolvability condition, and very more specifically to have evolved under the very specific conditions of evolution of language. These two goals are, at least on surface, antithetical. By enriching the theory you make the problem of evolvability harder, and conversely. And a lot of the field over the last years has been involved in some kind of effort to overcome this apparent conflict. So let me stop for a second and talk about the specific conditions under which language evolved which make the problem much harder and more striking.

In general, very little is known about the evolution of cognition. It's a very hard topic to study. One of the leading evolutionary biologists, Richard Lewontin, has a famous article in the four volume MIT *Invitation to Cognitive Science*.[10] He wrote the article on evolution of cognition and his basic conclusion is, in effect, "I'm sorry you guys, you're never gonna learn anything about it." It just can't be handled by the techniques available to current science. Notice it's not that it's mystery in the sense of the other things I imagine. It's just that it's beyond the possibilities of research. If you had, say, tape recordings from a hundred thousand years ago, maybe you can learn something, but we don't. Now we're never going to get them. So his conclusion was there is roughly nothing to say about it. There's a lot to what he said and that's worth thinking about. But I think it's little too pessimistic.

There is some evidence that has come to light and it's kind of suggestive. We know from plenty of fossil evidence that modern humans appear roughly two or three hundred thousand years ago, essentially in that range. It's now known that human groups, which were very small at that time, began to separate roughly 150 thousand or more years ago. And the groups that separated have the same faculty of language as far as we know. San people in Africa, the first group that separated. Well, that tells us that the faculty of language was already established not very long after modern humans appeared. So it seems essentially a property of modern humans as they appear. Notice that these periods of time are extremely small from the

[10] (eds.). *Method, Models, and Conceptual Issues: An Invitation to Cognitive Science*. Vol. 4, 107–132 Lewontin, Richard C. 1998. The evolution of cognition: Questions we will never answer. In Don Scarborough and Saul Sternberg. MIT Press.

perspective of evolutionary time, which doesn't deal with notions like tens of thousands of years. So essentially, we can say that the species characteristic appeared essentially along with modern humans. A second fact known from the archaeological record is prior to the appearance of modern humans, there doesn't seem to be any serious evidence of any kind of symbolic activity in the archaeological record. And not long after the appearance of humans, you start getting quite a rich record of complex symbolic activity. The Blombos Cave in south Africa is the most famous example. So there's further work on this by a very fine linguist, Riny Huybregts, most of you know, who pointed out that the earliest separation, roughly maybe a hundred fifty thousand to two hundred thousand years ago of the San people in Africa, although they have, as far as we know, the same faculty of language, they have a somewhat different form of externalization.[11] As he showed, these are all and only the languages that have complex click systems. So few exceptions, but I think he actually managed to show that they're meaningless. So what that suggests, Riny points out in his article, is that the faculty of language developed prior to the separation, but externalization took place after the separation in slightly different ways. There's some minor physiological adaptation about click languages like change in the structure of palate, but not very much.

Well, that's all very suggestive. If you put all this together, what it strongly suggests is that whatever emerged along with modern humans and yielded the faculty of language must have been very simple, simple such as something that nature would hit upon immediately as soon as some small rewiring of the brain made this task of satisfying the Galilean challenge possible. That converges with developments that have been taking place within the generative enterprise in quite a suggestive and important way. Well, in order to fix it, suppose that in fact something simple did develop along with modern humans yielding the faculty of language. We would expect it to be very simple in structure and with very simple modes of computation, which would satisfy the evolvability condition as a genuine explanation.

Then what remains to fix a language? Well, the individual has to fix a language on the basis of data, with very simple evidence. The reason is that we know now from psycholinguistic studies that acquisition of the essentials of language has already been carried out very early, in fact about as early as you can test. Two or three year olds have enormous understanding of the fundamental principles of language—and I'll come back to some examples of that— while the evidence available to them is very limited: maybe they've heard few million sentences, but that gives them extremely little evidence. That's been shown very well in careful statistical work, especially by Charles Yang, who showed that when you look at the effect of what's called Zipf's law, the rank frequency distribution of words, it turns out that almost all the evidence that children are getting is just repetitions of very few things. Even bigrams barely are repeated in millions of

---

[11] Huybregts, Riny. 2017. Phonemic clicks and the mapping asymmetry: How language emerged and speech developed. *Neuroscience and Biobehavioral* review 81:B, 279–294.

sentences; trigrams almost never or very rarely. So the evidence is really very slim. The knowledge that's acquired is very rich. We conclude in general what we expect to find is a very simple faculty of language, and the actual acquisition of language should be based on some kind of capacity to pick out what's significant and important from quite impoverished data. That's what you'd anticipate.

And more generally then, we of course always would —in any field—be looking for the simplest theory. That's simply a general fact about explanation. It's clear that as the foundations of a theory becomes simpler, its explanatory depth increases. So if science is interested in explanation, not just simulation, it'll be, of course, looking for simplest theory. There's a second reason for looking for the simplest theory which is a kind of a precept going back to Galileo again, who urged that we accept the idea that nature is simple and it's the task of the scientist to show this. That's true from study of failing bodies to the flight of eagles to whatever you look at. That's of course, what's called a regulative principle, a precept, but it's one that's been spectacularly successful in the sciences. So it's simply taken for granted in the sciences and there's every reason for us to take it for granted too. And thirdly for linguistics, there's third reason to expect a very simple theory of the faculty of language, namely the specific conditions under which this faculty appears to have evolved.

Well notice that it's often argued that evolution violates Galileo's precept. Evolution is what François Jacob called tinkering, a bricolage, which tries lots of different things and ends up with very complex objects. Whatever one thinks of that, it doesn't seem to apply to the special case of the acquisition of language simply because of the specific conditions under which it seems that language evolved.

Well, considerations like these arise very clearly in the development of the minimalist program to which I'll return; but the point I want to emphasize here is that learnability and evolvability provide the conditions for genuine explanation. That's the holy grail. These are the conditions for meeting the Galilean challenge. So genuine explanation will, of course, be at the level of UG and it will be in a form that meets the demands of learnability and evolvability. That's a very austere requirement. Anything short of that is not a genuine explanation. Anything short of that is a partial account, maybe a useful one, but not an explanation. It's a way of presenting materials as a problem to be solved. That's a very important endeavor. It's much better to have some organized presentation of some carefully structured problem. That's a great advance over just chaos, of course. So it's by no means denigrating those achievements, but we should not confuse them with genuine explanations.

Well, any device that's introduced in linguistic description—any device—to deal with some problem—whatever the problem is—has to be measured against these two conditions. Is it learnable? Is it evolvable? I think we're finally, maybe, in a position today to take the Galilean challenge seriously—which if true, is quite important. That's a new step.

So just to illustrate with a concrete example to which I'll return if there'll be time, there is very interesting paper by very good linguist Zeljko Boškovic, most of you know, on the coordinate structure and the adjunct island constraints, and in the paper, he points out each of these, coordinate structure and adjunct island, pose many problems, many mysteries. But his paper attempts and, I think, in a way succeeds in showing that these two collections of mysteries actually are the same mystery. What he tries is to reduce the adjunct island constraint and the coordinate structure constraint to a single mystery, relying on a neo-Davidsonian, event semantics, which in fact treats adjuncts as coordinates. So based on that idea, you can take two collections of mysterious phenomena, putting them together into one collection of mysterious phenomena, which is a significant advance. That leaves the mysteries, but they are now more susceptible to successful inquiry. And I think that virtually every achievement in the field is pretty much like that, that manages to reduce some collection of mysteries to a simpler more manageable collection, which is a major achievement, but it's not a genuine explanation. So we're still searching for the holy grail. At least, all of this is the way things look within the biolinguistic program. If you're pursuing a different enterprise there are different considerations.

The first proposals back to in the early 1950s were basically dual. There were two different problems that had to be faced. One was the problem of compositionality, how do you put structures together. The other was the very puzzling property of dislocation: expressions are heard in one position, but they are interpreted both there and somewhere else. So in "What did John see," you interpret the *wh*-phrase "what" as a quantifier ranging over the whole clause, but you also interpret it as the object of "see", where you don't pronounce it. That's a ubiquitous property of language. Very complex cases have been studied over the years.

Well, the proposals back in the 50s were for two different kinds of mechanisms: phrase structure grammar for basic compositionality, transformational grammar for dislocation. You look back at the proposals, each of these systems was much too complex to meet either the conditions of learnability or evolvability—that is, to provide genuine explanations. That was understood, but it was very unclear what to do about it. It was generally assumed at the time that compositionality is something natural—that we can kind of handle. Dislocation seems very strange. You don't build dislocation into formal systems, for example. It is just something that seems unique to language and a very weird property of language—one that was considered what's called an imperfection of language that somehow adds, for odds reasons, this complex notion. That's still widely believed, but I think it's exactly the opposite of the truth. As research has progressed, it turns out, first of all, these two apparently different properties can be unified and that the more primitive of them in fact is dislocation. I'll come back to that, but it seems that the most primitive operation is dislocation, whereas compositionality is considerably more complex—although they can be unified into a single operation, something I'll want to come back to.

A couple of comments. Within a particular current of the generative enterprise by the 1960s, it was quickly recognized that phrase structure grammars are completely unacceptable. They're way too complex. Phrase structure grammar for one thing allows totally impossible rules. So there's nothing in a phrase structure grammar that says you can't have a rule, say, sentence becomes preposition + verb phrase (S➜P+VP), or anything else you can imagine. So it just allows a huge number of rules that are completely impossible. So there has be something completely wrong with it. Also I think in retrospect; we can now see that phrase structure grammar conflated three quite different notions. One is the notion of just hierarchical structure, a second is the notion of linear order; a third is the notion of what was called projection: how to decide whether some unit you form is a such and such. And over time, it's been recognized that these are quite different properties.

Well, a step was taken by the late 1960s to overcome some of these problems with the development of what was called X'-theory (X-bar theory). I won't discuss it—I assume you know what is it—but X'-theory did have a number of consequences, which interestingly were not really understood very clearly at the time. (Tim will remember this.) For one thing, X'-theory keeps to structure but has no linear order. So you have the same X'-theory in effect for, say, English and Japanese, which are close to mirror images. Well, the significance of that wasn't really entirely grasped. What it tells you in the first place is that you have to have a principles and parameters approach. There can't be a rule system, at least for compositionality, which took some years for that to settle in. But it's immediately obvious once you look at X'-theory. So take English and Japanese—essentially the same X'-theory, but there has to be something distinguishing them, something that says the linear order is one way, in one language, the other way in another language. But that's a principles and parameters approach. Furthermore, if you look at that parameter, you see that it doesn't affect the meaning of the sentence. So whether you have verb object or object verb language, the meanings are exactly the same. The theta structure, the argument structure is the same. That at once suggests that the parametric difference, the linear order, simply doesn't have anything to do with the core of language, namely the construction of an infinite number of thoughts. To put it in more technical terms, it doesn't feed the conceptual-intentional level, doesn't yield semantic interpretation. That's an observation which has many consequences if you think it through. It's elaborated in later work in other ways, but it's already a hint that somehow properties like linear order and other aspects of externalization don't strictly speaking belong to the I-language. A lot of consequences when you think it through. I'll return to it.

Well, these are some of the consequences at once from looking at the X'-theory. It should have been recognized instantly, but only gradually came to be realized later on. X'-theory does have problems. I'll mention these and then put the rest off until later. There's fundamental inadequacy of X'-theory which was not recognized. It still conflates projection and compositionality. It does separate order, but it still

conflates those of two. And that runs aground as soon as you look at exocentric constructions, which can't be generated in X'-theory. So you can't have say subject-predicate or you can't have any movement because all movement yields exocentric constructions. So if you have *wh*-movement, it gives you a construction, a *wh*-phrase CP. Two, just two, structures. Neither one is dominant. Subject-predicate, if you accept, say, the predicate internal subject hypothesis (Sportiche and Koopman)[12] so you have a noun phrase and a verb phrase, but they're just two parallel phrases. Well a lot of artificial devices were constructed within X'-theory to try to get around this and to give you what you intuitively know is the thing you're after, but that's not allowed, that's trickery. So there was a fundamental problem with X'-theory that finally was resolved a couple of years ago with the development of labeling theory, which finally separates totally this problem of projection from compositionality, separates all three; tells you when some operation of dislocation must take place, when it may take place, when it need not take place. So that finally breaks up the conflation of three notions that were mixed up in phrase structure grammar—lot of interesting results, plenty of interesting problems. Well, that brings out us up until about the 1990s. So why don't I stop there and go on next time.

Lecture #2 (April 30, 2019)

The recapitulate briefly, of the many different ways one can approach language, we've focused on one particular one, what's been called the generative enterprise, which is concerned with what's been called *the basic property of language*, namely that each language constructs in the mind an infinite array of structured expressions each of which has a semantic interpretation that expresses a thought, each of which can be externalized in one or another motor system, typically sound, but not necessarily. As I mentioned, this approach revives a traditional conception—goes back to the 17th century—went as far as the early 20th century, the last representative was Otto Jespersen, then forgotten, then revived without knowledge of the tradition.

Within the generative enterprise we also have settled on a particular approach, which is not universal, not required—namely, what's called the biolinguistic program, that is, regarding the language that a person has as a property/trait of the person, like the person's immune system or the person's visual system. Like other such systems, it is a particular instance of something more general. There is a general mammalian visual system which is innately determined. There's a particular visual system that is the result of interaction with experience, other factors that govern growth. And the same is true of the internal

---

[12] Koopman, H and D, Sportiche. 1991. The position of subjects. *Lingua* 85:211–258.

language (clarification: call it I-language, to make it clear that there's a particular technical notion involved). "I" fortuitously stands for internal, individual, and also intensional (with an *s*)—namely we're interested for the purposes of science in the actual method by which this property is expressed, not just any method that happens to yield the same external consequences. So that's the approach.

Within this approach there is a notion at once of what constitutes genuine explanation. A genuine explanation will have to meet two conditions: the conditions of learnability (the system has to be able to be acquired by an individual) and also evolvability (the innate internal system, the faculty of language, has to have evolved). So if a device is used for description that can't meet those two conditions, we don't have a genuine explanation, we have a description. Descriptions can be very valuable. It's much better to have organized descriptions than chaos, but it's short of genuine explanation. Now that's a very austere requirement, and until recently it hasn't even been possible to consider it. But it is the goal to be aimed at by (certainly) any approach within the sciences or rational inquiry generally.

Well the early work from the 1950s distinguished two different aspects of I-language (the term came later, but it doesn't matter). One of them was compositionality; the other was dislocation. For compositionality the early approach that was adopted was Phrase Structure Grammar (PSG); for dislocation, transformational rules. It was assumed at the time (and in fact for a long time) to the present (and still is widely believed) that dislocation is a strange property of language. You wouldn't expect it. It's kind of an imperfection as it was called—have to kind of explain it away. Compositionality was taken for granted. As I mentioned, and will talk about today, it seems like the opposite is true, which is kind of an interesting result.

The PSG had two fundamental problems, which were already recognized by the 1960s. One is, it simply allows way too many rules—that alone disqualifies it, can't have a system that permits all sorts of impossible rules. The second problem, which was more subtle, is that it conflates three different aspects of internal language which really should be treated separately. That's become clearer over the years. One is simply compositionality, having structure—the core elements of the basic property. The second is linear order, which is not part of the basic property. The third is projection—that is, determining what each structured element is (what kind of element it is). The PSG tries to handle all three. It's become clearer over the years that's not the right way to proceed.

By the late 60s all of this was becoming partially clear. This approach turned to what was called "X-bar theory" to overcome the problems of PSG. Well X-bar theory takes care of some of them. It eliminates the main problem of far too many rules. In fact, there's basically no rules in an X-bar theoretic approach. So we don't have all these impossible rules hanging around. It also distinguishes linear order from hierarchical structure. So there's no order in an X-bar theory. There is still the conflation of projection and compositionality, which only recently (last couple of years) was recognized to be a basic problem—

primarily because of the existence of exocentric constructions, which are excluded by X-bar theory, although they are all over the place. As I mentioned, in order to get around this annoyance, special tricks were invented to pick out the right answer, intuitively what is the right answer but without any particular basis. So there's nothing in X-bar theory that tells you that if you have an NP + VP structure, you should regard it as essentially verbal rather than essentially nominal. It's anything, or is it an inflectional element—totally arbitrary. And that of course is a very serious problem. It's overcome, I think, by labeling theory, which I'll simply assume you are familiar with (I don't have any time to go into it, but it's an approach which overcomes this problem and has, furthermore, the advantage of accounting for when dislocation operations (movement operations) may take place or must take place, and when they have to terminate; does it automatically). But that I'll put aside.

The X-bar theory also (although this was not recognized at the time) forces you to have a Principles & Parameters approach. You have the principle of X-bar theory, then you have to somehow get the order that's in particular languages. So English and Japanese are sort of mirror images at the core. You have to somehow state that separately. And that led right off within a couple of years to general Principles & Parameters approach, which raises interesting questions on its own. Where did the parameters come from? Do they have to evolve? Well, if they have to evolve, then we're very far from genuine explanations. And in fact the simple case that X-bar theory exposes points to what should be the answer. Say, when you have V–O or O–V order, that has to be stated for each particular language. But the answer to it has no effect on the semantic interpretation, so it doesn't feed the conceptual-intentional level—which is suggestive. It suggests that whatever the parameters—and if you think about the other parameters that have been proposed (say, null subject or others), they have the property that they do not enter into determining semantic interpretation. So if you pick say Richie Kayne's proposals about basic SVO order vs. say a basic SOV order, it doesn't make any difference to the semantic interpretation. All of that is very suggestive. If the parameters don't enter into semantic interpretation, we have to ask why they are there altogether. After all, the core part of language is just expressing thoughts basically, having semantic interpretations.

So it begins to look right off as if the parameters should be external to the core part of language. (We'll come back to that.) It looks more and more plausible to me that that's the case. What about their evolution? Well, think again about the head parameter. It didn't evolve; it's just forced by the requirement of having to externalize in a sensory motor system that requires linear order. So it doesn't evolve at all. There's a mismatch between the internal system and the sensory-motor system. And that mismatch just has to be answered one way or another. But there is no evolution of the parameter. And in fact the goal ought to be to try to show that all parameters have this property. There's very interesting work on this by Epstein-Seely-Obata-Baptysta studying a particular array of dialects of Cape Verde Creole. They try to show that

the differences between the dialects simply reduce to an unstated property of the set of principles: it allows alternative orders of rules. It turns out that several dialects just pick different answers to this.

If this kind of approach can work for all parameters, then the problem of evolution is overcome and we go back to having a hope getting genuine explanations. This is all kind of work in progress—in fact, not even in progress—it's an open problem to be investigated to try to show that for all parameters—all the options that distinguish languages—the goal would be to have two properties: 1) to show that they are simply part of externalization (they don't feed the conceptual-intentional level (semantic interpretation)) and 2) they don't evolve at all; they are simply options left open by core grammar, which have to be settled one way or another. So each language picks its way of settling them. That would be the perfect answer if you can find it.

Well, X-bar theory partially dealt with these problems. That still left the problem of conflating projection and composition, and it had the huge problem of excluding exocentric constructions, which, as I say, abound, so that's no good.

The Principles and Parameters approach which developed mainly through the 1980s led to a huge explosion of empirical work in a very wide variety of typologically varied languages. I think more was learned about language in the 80s than in the preceding couple of thousand years. Also, there were conceptual steps forward. Still there was the problem of not being able to reach genuine explanation.

Well, by the early 1990s it appeared to a number of people, myself included, that enough had been learned so that it might be possible to tackle the whole general problem of language on more principled grounds—that is, to actually seek genuine explanations, to formulate what was called the "Strong Minimalist Thesis", which says that language is basically perfect, what's called "approaching Universal Grammar from below", starting by assuming (let's see how far we can get) a perfect answer that would be a genuine explanation and how much can we explain this way. That's what's been called the Minimalist Program. Now the way to proceed with it is to start with the simplest computational operation and just let's see how far you can go. Where do you run aground?

Well, what you have to do is/you have to explain the actual phenomena of language on the basis of the simplest computational operation—that would be the genetic element plus other factors that of course enter into acquisition of language. And there are several such factors. One of course is the data with which the child is presented. So, we don't all speak Tagalog. It has got to be something in the data that picks out which language we end up with. And it should be the case that the correct theory when we converge on it will show that very limited data suffices to fix the I-language. There's an empirical reason for that, that is, as psycholinguistic evidence/ child studies have increasingly shown, the basic properties of language are really known quite early with very little evidence available.

There are other factors, sometimes called third factor properties, just general principles of growth and development, which are independent of language—maybe natural laws. For a computational system like language, the natural one to look at is computational efficiency, just assuming, for the reasons I mentioned last time, that nature seeks the simplest answer: so computational efficiency plus the simplest computational operation plus whatever contribution data makes, that should yield the I-language attained. That's the goal. This program has been called the Minimalist Program.

So where this can be achieved, you do have genuine explanation. If you can explain something simply on the basis of the first and the third factors, namely UG (the simplest computational operation) along with, say, computational efficiency, if anything can be explained in just these two terms, you do have a genuine explanation. You've solved the learnability problem because there is no learning. That's the optimal solution to the learnability problem. You've also addressed the evolvability problem in the best possible way (if you think it through). The basic property is simply a fact. It's a fact about the faculty of language that it satisfied the basic property. It therefore follows that there has to be some computational operation that yields the basic property. And whatever computational operation it is, it's going to incorporate the simples computational operation and optimally nothing else (or not much else). So if you can reduce something to the simplest computational operation, you have solved the problem of evolvability, even if you don't know the exact answer as to how the rewiring of the brain took place. You know this must have happened. Okay, so you're free and fair.

So whenever you can account for something on the basis of just the first and third factors (UG and third factor properties like computational efficiency), if you can do that, you certainly have a genuine explanation. You could have more complex genuine explanations if you bring in the effect of whatever the right theory of determining the I-language from data is. When that factor is introduced you can have richer kinds of genuine explanations. But the simplest kind will simply reduce to the simplest computational operation plus computational efficiency.

Well, notice, of course, that the innate properties, including the simplest computational operation, may be triggered by data, by evidence; but that's normal for innate properties. You commonly have to have triggering (evidence) to get the system to start functioning. So in the case of the visual system, for example, it's known for the mammalian visual system including humans, unless you have the right kind of stimulation in very early infancy (the first couple of weeks), patterned stimulation of certain kinds, then the visual system just doesn't function. It's not learning how to function. It knows how to function. Something's got to set off the system in operation. That's true for innate processes altogether, so presumably true for language as well. So for example, if a child is raised in isolation (no noises), the faculty of language would collapse, presumably. There's in fact even some evidence for that for odd cases, unfortunate cases, that

have been discovered. But apart from that, it seems that we can at least begin by asking how far can we get to genuine explanation by just looking at first and third factors.

So what's the simplest computation possible? Well, the simplest computation which is found buried in any computational system just takes two objects that have somewhere already been constructed and creates a third object from them. In the optimal case, the operation will not modify either of the elements that is operated on. It will leave them unchanged and it won't add any additional structure. So it just takes 2 objects, call them X and Y, and forms a new object Z, which doesn't affect X, doesn't affect Y. That's basically binary set formation. So the simplest operation, which has been called Merge in recent years, is simply binary set formation. And the question we can ask is (if we can account for something in terms of Merge and third factor properties like computational efficiency) have we in fact reached genuine explanation for the first time?

Well if you think about the operation Merge (binary set formation), it has two possible cases. So we have X and Y, we're forming the set {X, Y}: one case is that X and Y are distinct, the second case is that one of them, say X, is an element of Y (it's part of Y)—the technical term is *a term of Y*. We define the notion *term* to mean X is a term of Y if X is a member of Y or it's a member of a term of Y (normal inductive definition, irreflexive). So if one possibility is that X or Y (doesn't matter which) is a term of the other, the other possibility is that they're distinct. There is a third imaginable case—namely, that they overlap. But if you think about the process of construction of binary sets, you can't have that case; so that case we can rule out. So we have basically two options: the names that are used for them are external Merge for two distinct elements and internal Merge where one is part of the other. The linguistic analogues would be, for example, forming *read books* from *read* and *books*, that's external Merge. Internal Merge would be forming *what did you see* from *you saw what*. Notice what you actually form is *what you saw what*. You take *what*, merge it to *you saw what*. You get *what you saw what*. That's critically important. That's internal Merge. Internal Merge is dislocation. But "dislocation" is not a strange word because you are actually keeping the original element. Okay, that's internal Merge. External Merge just takes two separate syntactic elements and brings them together.

Notice that these two are one operation, Merge. They are not two different operations. These are just the two possible cases of the single operation. There is a lot of confusion about this in the literature, so you really want to be careful about it.

Just to be clear, the Minimalist Program is a program of research; it's not a theory. You read in the literature discussion of whether the Minimalist Program is true or has been refuted—that's meaningless. It's a program of research that asks how far can we get towards genuine explanation. Okay, that's the program.

As I mentioned last time, by now it has some (in very recent years) independent motivation from what's very recently been discovered about the conditions of evolution of the language faculty. To repeat quickly, it's been established by genomic evidence that the language faculty was established roughly at the time of the appearance of anatomically modern humans. There's no evidence for symbolic activity before that. So we have an indication, not a proof, but an indication that probably what happened is very simple. And it apparently hasn't changed since, since the language faculty is shared first of all among the earliest separating groups, but in fact as far as we know among all human groups. Se we have independent reason to believe that the program is a plausible program. It's likely that the right answer to what the language faculty is will in fact be something like the simplest computational operation.

Well, how far can we get on this? It seems to me that in the last couple of years there have been pretty substantial achievements which give us an indication of how far we might be able to go. One achievement is cases where you have reached genuine explanation. One is simply the unification of the two fundamental properties of language that were recognized back in the 50s, composition and dislocation. If they can both be unified in Merge (I think the first person to point this out was Hisa Kitahara), if you can bring these together, it's a pretty substantial achievement. These were always thought to be two totally separate parts of language. But maybe they are just the same operation, in fact the simplest operation. So maybe the null hypothesis, the perfect case, is that you have both dislocation and composition from the same operation.

Actually you can push that a little farther. If you think about the two cases of Merge, internal and external, which one is more primitive, the simplest, the one that involves least computation? Actually. Internal Merge. External Merge requires search—you have to search all the available possibilities. But what are these? First of all it's the entire lexicon, say maybe 50,000 items for a normal knowledge, plus anything that's already been constructed. This is a recursive procedure, so you're constructing new syntactic elements along the way and any of them can be accessed for further operations. So external Merge involves a huge search procedure. Internal Merge, in contrast, involves almost no search at all. You're just looking at a particular structure, taking a look at its terms. That's it. So internal Merge is by far the more primitive operation; external Merge is much more complex—which is an interesting conclusion because the opposite has always been assumed. It's always been assumed that compositionality is the obvious case; we don't have to explain that. Dislocation is the complicated case, the imperfection, the strange property of language. It turns out (when you think it through) the opposite is the case. Internal Merge (dislocation) is the simplest case. What you have to ask is why have external Merge at all? Why not just keep to the simplest case? (Well, you think about that for a minute. I'll come back to it in a second.) But that's what you discover when you begin to think about the nature of operations.

Incidentally, there are some independent reasons for this. Not proofs, but suggestive. If you think about internal Merge alone—so imagine the simplest case where you have a lexicon of one element and we have the operation internal Merge. (I could write this down, but it's simple enough so you can figure it out in your head). You have one element: let's just give it the name *zero* (0). We internally merge zero with itself. That gives us the set {0, 0}, which is just the set *zero*. Okay, we've now constructed a new element, the set zero, which we call *one*. Then we apply internal Merge again; we take the term of the set *zero*, we merge it with the set {0}, we now have a new element which consists of two elements, zero and the set containing zero {0, {0}}. We call that *two*. And we just keep going. We've created the successor function. And in fact (if you think about it a little more), we've also created addition, because if you take one of the longer terms and you merge that, you're essentially getting addition. Multiplication is just repeated addition, unbounded repeated addition. So we basically have the foundations of arithmetic, just from internal Merge.

That raises an interesting question. There is a question that very much troubled Charles Darwin and the co-founder of evolutionary theory, Alfred Russel Wallace (who were in contact). They were both concerned with what appeared to be a serious contradiction to the theory of natural selection—namely, the fact that all human beings understand arithmetic. If you think about it, arithmetic has never been used in human evolution except in a very tiny period tacked on at the very end. But for almost the entire history of human beings, nobody bothered with arithmetic. You might have small numbers, you might have what's called *numerosity* (knowing that 80 things are more than 60 things), but that's different from arithmetic (knowing that numbers can go on forever, knowing what the computations are of say addition, multiplication). So where did this come from? Darwin and Wallace simply assumed that all humans have it. By now there's pretty good evidence that they're right, that you find indigenous tribes that have no number words beyond maybe 'one', 'two', and 'three' and do handle numerosity. Turns out that as soon as they're inserted into, say, market societies, they immediately compute freely; they have the entire knowledge. And in fact on conceptual grounds, we know that this must be the case, because there's no way to learn it. There's no possible way to learn that the number go on forever. You can't have evidence for that, but every child just knows it. So the problem is: where did it come from, since it was not selected and it's universal. It looks like it contradicts the fundamental ideas of the theory of evolution.

Well, a possible answer to this is that it's just an offshoot of the language faculty. The language faculty did emerge. It's here, we're using it; so it's around. The language faculty at the very least in its simplest case give you arithmetic. So it's quite possible the reason everyone knows arithmetic is just that we've got the language faculty. So therefore we have the simplest case of it.

Now on the surface it looks as if there's counter-evidence, which has been brought up quite often— namely, dissociations. The kind of cognitive phenomena Susan Curtiss discusses. There are dissociations between arithmetic competence and linguistic competence. So there are people who have a perfect language

faculty and can't do arithmetic—and conversely. However, it's not clear that's a counter-argument for reasons that were pointed out a couple of years ago by Luigi Rizzi (a great linguist who most of us know). He pointed out that the tests for dissociation are testing performance, they're testing what people use, not the internal knowledge. So as, for example, there are also dissociations between reading and language competence. There are people who can have normal language competence but can't read, and conversely. Now it can't be that there's a separate reading faculty; that's impossible. So in fact what it shows is just that the tests are studying the utilization of the competence; they're not testing the competence itself. So it may very well turn out that this is the answer to the Darwin/Wallace problem, a deep problem of the evolutionary theory.

Going beyond, there's other very suggestive evidence. Marvin Minsky (one of the founders of Artificial Intelligence (who was a very good mathematician)) about a couple of decades ago actually in the 60s, he and one of his students experimented with the simplest Turing machine. A Turing machine is just an abstract computer basically, the ones that have the smallest number of states and the smallest number of symbols. Just ask the question what happens if you just let these machines run free. And they got an interesting result. Most of them crash: they either get into infinite loops or they just terminate. Of the ones that didn't crash, they gave the successor function, which suggests strongly—and thus Marvin went on to draw the plausible conclusion; he said probably nature in the course of evolution will pick the simplest thing, so it will pick the successor function. So he was thinking of things like extraterrestrial intelligence. He said that we can expect that if we ever find anything like that, it will be based on the successor function because that's the one that's the simplest one. Well that's internal Merge. So from quite a number of points of view, we have a conclusion that dislocation is probably the primitive element of language. Composition is the one that has to be explained.

So why do we have external Merge altogether? Well, suppose we didn't have external Merge, what would language look like? Actually, it would be nothing but the successor function, because that is all that internal Merge can produce: the successor function and addition. So if you had say 10 elements in the lexicon instead of one and used only internal Merge, what you would get is ten instances of the successor function, nothing else.

But language has other properties, for example it has argument structure, θ-theory (it has semantic functions like agent, patient, and so on). You can't get that from internal Merge. You're required to have external Merge develop the structures which will yield argument structure, θ-structure. So an empirical property of language, namely that it has argument structure forces you to use this more complex operation of external Merge. Another reason for it is, again, the existence of exocentric constructions like, say, subject–predicate. You can't get those by internal Merge of course. So there's just empirical reasons that force language to include the more complex operations of external Merge. Notice that this turns everything

on its head. That's exactly the opposite of what has been assumed. The basic operation is dislocation, internal Merge (arithmetic essentially). Then because of special properties of language (argument structure, exocentric constructions), you are just compelled to use also the more complex case of external Merge.

But again notice that these are both the same operation, just two cases of the same operation, one more primitive than the other. You are forced to use also the less primitive one, compositionality, because of properties of language.

Well, notice again in the case of internal Merge, you not only get dislocation, but you have several copies of the element that has been dislocated. Those of you who know how this works, you can have a great many copies. You can have successive cyclic dislocation, which leaves a copy in every position with consequences for externalization and interpretation.

Well that's very important because of a phenomenon known in the linguistics literature as "reconstruction"—namely, a dislocated element is actually interpreted in the position from which it came (even intermediate positions). So standard examples are sentences like *it is his mother that every boy admires* with the interpretation 'every boy admires his own mother'. *It is his mother that admires every boy*, meaning 'his mother admires every boy' doesn't work the same way: the quantifier *every* does not range over the variable *his*. And the reason (if you think about it) is that the unpronounced copy, what's reaching the mind, actually has the phrase *his mother* in the position in which quantification works. *Every boy admires his mother* is reaching the mind, although not the ear. And there are many more complex intricate examples of the phenomenon of reconstruction.

Now in a lot of the modern literature there is an operation of reconstruction that somehow takes the kind of structurally highest element and reinterprets it back in the initial position. But that's unnecessary because that comes automatically; it's just there, so what reaches the mind is what's called *the chain of elements*, all the elements that have been dislocated, but it doesn't reach the ear. What reaches the ear just cuts out all but one of them. (Actually, that's (interestingly) not entirely correct; I'll come back to it in a minute. Put that aside for a minute, I'll return to it in a second.)

And notice that when you look at these cases, you have a strengthening of the conclusion that was already suggested by X-bar theory—namely, that the core language, the system that reaches the mind to give semantic interpretations, pays no attention to the way expressions appear in phonetic form. It just asks what happens on the internal operations. What we hear has linear order and is missing copies, but the mind doesn't care because it's only paying attention to the internal structure.

And that raises the question: why does linear order even exist? Why do you have it? Well, the trivial answer to that is the sensory-motor system requires it—partially, if you use a different sensory-motor system for externalization like sign language, which doesn't have strictly linear order because visual space provides options that acoustic production does not have. But if you have visual space you can use positions

in it to convey the meaning. It's in fact done in sign language. So anaphora in sign language works by picking a point in space and referring back to it, which you can't do in spoken language. You can also have simultaneous operations in sign language like raising your eyebrows while you're signing, which turns something into a question. Can't do that in spoken language. In general, the nature of the sensory-motor system is determining properties of language in a broad sense, which are not part of core language at all. They just reflect the characteristics of the sensory-motor system.

Well, now there's something that came up in the discussion last time; I'll repeat it. The sensory-motor systems have nothing at all to do with language. These were in place long before (in fact millions of years before) the language ever evolved. The time language evolved may be roughly a couple of hundred thousand years ago. The sensory-motor systems were in place independently of language. They haven't changed with the usage of language. They're still what they were at the origins of homo sapiens, with the most extremely minor variations. There's no indication of any adaptation of sensory-motor systems to language. The early homo sapiens who developed this internal system (by whatever means they did (some rewiring of the brain which gave the simplest computational operation)) had the task of getting this stuff out somehow and they had to use the sensory-motor systems that are around, as every infant does. And the sensory-motor system imposes conditions; but those are not linguistic conditions (they are not strictly speaking part of language). From which we can conclude—it's a contentious claim (as Hilda Koopman pointed out last time), but it looks from this kind of argument as though properties like linear order or elimination of copies just don't have anything to do with language. They're conditions imposed by the mode of externalization, which is language-independent. And (as I mentioned last time) this has the paradoxical conclusion that just about everything that's been studied in linguistics for the last 2500 years is not language. It's the study of some amalgam of language and sensory-motor systems. But if we want to really think through the internal language, the entity that's computing away and giving you semantic interpretations, that doesn't have any of these properties. It doesn't have operations of reconstruction, doesn't have linear order; just operates along producing thoughts in your mind which you can then try to externalize somehow.

Well, in fact if you think about it, that's even true of what's called *inner speech*. So almost all the time, say 24 hours a day, you're basically talking to yourself; you can't stop. It takes a tremendous act of will to keep from talking to yourself. But if you think about inner speech, it's externalized speech. For example, it has linear order. Or you can think about two sentences you just produced in your mind and you can ask do they rhyme, are they the same length, do they have copies, and so on. (The answer is they don't.) So what we call inner speech is very superficial. It's not what's going on in the mind. All of this has a lot of implications for the study of consciousness and preconscious mental operations. Almost everything that's

involved in language seems to be inaccessible to consciousness. It's prior to whatever reaches consciousness, even in inner speech. A lot of implications to this that are worth thinking through.

A conclusion that kind of begins to appear on the horizon is that core language (the internal computations building linguistically articulated thoughts) could be universal. It's possible that it's just common to the species. It's part of the faculty of language, which would solve the learnability problem because nothing would be learned if that's true. And it solves the evolvability problem in so far as you can reduce operations to the simplest computations. So that's a kind of goal that you can see formulating itself on the horizon, something to strive for in research. You would also anticipate, it this is correct, that the variety of languages is just a surface phenomenon. It's not really a fact about language.

I mentioned last time that in the structuralist-behavioralist period, it was assumed almost as a dogma that languages can differ from one another pretty much arbitrarily and therefore each one has to be studied without any bias, any preconceptions. That was what was called the *Boasian doctrine*. If what I have been saying is correct, then that's just completely false. There's only maybe just one language and a lot of variation that comes from mapping it into externalization systems. You would expect the same to be true of the complexity of language because relating two independent systems is a complex affair. So you would naturally expect the complexity of language to reside there. Also just the fact that languages change very fast. In fact, in every generation there's language change. It could be that this is just externalization. Anyway, those are goals to be thinking about—kind of coming into view when you just think about the nature of the problem.

Well, under externalization (as mentioned before) the ear hears only one case, whereas the mind perceives every case. So when you say *it is his mother that every student admires* you don't hear *it is his mother that every student admires his mother*. That's hitting the mind so you know how to interpret it, but it's not hitting the ear. And that's in general true.

Actually there are some very interesting qualifications to this. It's known that in some languages the initial position that you don't hear in dislocation actually has some kind of a phonetic mark, some sort of a sound that indicates 'I was here'. There's a study by a Vietnamese linguist Tue Trinh that goes into this in some detail, relating it to prosodic properties of the language which pretty well determine when this is the case. More intricate in successive cyclic movement (for those of you that know this stuff), where you have successive cyclic dislocation to higher and higher positions, there are in many languages residues, strange residues, along the way. That's true at both of what are called *phases*: the CP-level and the v*P-level. In some languages you get phonetic traces indicating that something went through here. So there is some indication in some languages in the externalization that there was something there in the internal computation; but overwhelmingly you just don't hear anything. You hear one case, typically the structurally highest case. Although in what are called *in situ languages* like say Chinese or Japanese, you can hear it in

base-generated position, the place where it started from. Actually, that's true in languages like English too. So you have *in situ* constructions in English and similar languages in very special circumstances (which are interesting), but overwhelmingly languages vary this way.

So the question that arises is: why don't you pronounce everything, why bother deleting a lot of copies? Well here we turn to a third factor property: computational efficiency. Suppose you were to pronounce all of them? I gave a simple example, *what did you eat*, but instead of *what* it could be of arbitrary complexity, so *which book that appeared in 1690 and then was destroyed did you read?* let's say. If you repeat that as *which book that appeared in 1690 and then was destroyed did you read which book that appeared in 1690 and then was destroyed?*, you've got a lot of computation going on, mental computation to set the stage for phonetic articulation and then motor control to articulate the whole thing. So you have massive reduction in complexity if you just don't pronounce extra copies. So computational efficiency compels the deletion of the copies. You can't delete them all of course or else there's no indication that the operation ever took place. So you have to have one at least, typically the structurally highest one—then the rest you delete.

Well that's fine from the point of view of meeting the condition of computational efficiency, but it has a consequence for language use: it causes problems for perception, serious problems in fact. Those of you who've worked on mechanical parsing programs know that one of the toughest problems is what's called *filler-gap* problems. You hear a sentence which begins with *which book*, then comes the whole sentences and you have to find the place where *which book* is interpreted. There's nothing there in the sentence to indicate where that place is. That's called a filler-gap problem. These can be pretty tricky. It's simple in the case that I mentioned, but if you think through complex cases, it can become quite intricate. So we have the following strange situation: computational efficiency is forcing communicative inefficiency; computational efficiency is forcing parsing problems, perception problems, communication problems.

What does that suggest? It suggests that the way language evolved and is designed, it just doesn't care about efficiency of use. It's of no interest. It just cares about efficiency of computation. Now there a lot of evidence for this, many linguistic phenomena that are quite familiar to us that nobody bothered thinking about. Take for instance structural ambiguity, examples like *flying planes can be dangerous* which could mean either 'the act of flying planes can be dangerous' or 'planes that fly can be dangerous'. There's a lot of structural ambiguity in language. Where does it come from? Well if you look at the cases, it comes from just allowing the rules to run freely and not caring whether it causes perception problems. But all structural ambiguity problems are perception problems. You have to figure them out.

Take what are called *garden path sentences* (discovered by Tom Bever about 50 years ago), sentences like *the horse raced past the barn fell*. When you present this sentence to people, they are

confused at first—what's *fell* doing there extending the sentence *the horse raced past the barn*? It's called a garden path sentence; it leads you down the garden path to the wrong conclusion. But if you think about it further, it's a perfectly grammatical sentence with the interpretation 'the horse that somebody raced past the barn fell'. These are all over the place.

Incidentally, if you run the Google parser on garden path sentences, the one you can pick off the internet, it crashes all the time on these even though they are normal parts of language. And again, these are just cases where the rules run free and they don't care whether it causes perceptual problems or not.

Actually the most interesting case, too complicated to go into here (but many of you are familiar with it) is islands, syntactic structures out of which dislocation is not possible. You can think them, but you can't produce them. The cases of islands that are understood, there are lot of problems with them. All of them turn out to be cases where you have computational efficiency forcing the island, compelling a paraphrase, often a complex paraphrase to somehow get around it. Take say the so-called "wh-island condition, which blocks "which book did he met the man who read." There's a good argument that the explanation traces back to a phase-theoretic/strict cyclicity argument with a basis in computational efficiency, hence requiring a paraphrase.

Well, the general conclusion seems to be that, as far as I know, whenever there is a conflict between computational efficiency and communicative efficiency, computational efficiency wins. Language just doesn't care about communicative efficiency. That also has a consequence. There is a very widely held doctrine, almost a dogma, that somehow the function of language, how it evolved, its basic design and so on is as a system of communication. And apparently that's not the case. Apparently, language just doesn't care about communication. Of course it's used for communication—very useful, but that just seems to be something on the side. Probably the background for this doctrine is, first of all, that we do use language for communication all the time—so on the surface, that's what you see—and also a kind of belief going back to early Darwinian notions that anything that evolved had to proceed by small steps from what came earlier. But it seems to be just entirely false. Language seems to have come from something totally different having nothing to do with communicative systems. (Well, a lot to say about this, but time is running out I think. Let me just go on a little bit because there is a lot I want to get to.)

There are other genuine explanations, some of them pretty striking, for nontrivial properties. The most interesting one of all, I think, is what's called *structure dependence*. It's a strange property of language, a puzzling property noticed years ago, that the operations of language simply don't seem to pay attention to linear order. They only pay attention to structural relations. So a standard example in the literature is a sentence like *the man who is tall is happy*. If you turn that into a question it becomes *is the man who is tall happy?* not **is the man who tall is happy?*. That's obvious; every English speaker knows that.

Well, as soon as generative grammar began, the question arose: why? Why don't you use the simple computational operation of picking the first occurrence of the copula and putting it in front? It's a much simpler operation than the one that's used. The one that's used in fact picks the structurally closest one, which in the example above is the second *is*. So if you draw the tree, the structure of this, you see that the second *is* is structurally closer though linearly more remote. The language faculty just doesn't look at the linear order. Or take a sentence like *can animals that fly swim?*. We understand that to mean 'can they swim?' not 'can they fly?'. So it doesn't mean 'is it the case that animals that can fly also swim?'. Why doesn't it mean that? It's a perfectly fine meaning. So why doesn't the sentence mean that? That's what it would mean if you picked the first occurrence of *can* and moved it or interpreted it. It's not a mistake that any English speaker can possibly make, that any child makes. In fact, it's by now known that by about 30 months old infants already are operating with structure dependence. That's something that's just automatic; you have essentially no evidence for it. And this is true for all constructions in all languages, as far as anybody knows.

Well, all of these things we take for granted, but the usual question arises: why? Why, it's particularly puzzling for two reasons. For one thing, we avoid what looks like the computationally simple operation (based on linear order): pick the first element, that's trivial. We avoid that; we use what looks like a more complex operation. The second reason that it's puzzling is we ignore everything we hear. What we hear is linear order, we don't hear structure. That's something the mind constructs internally. So we ignore everything we hear; we ignore what looks like the simplest operation; and we do it universally.

Examples like *is the man who is tall happy?* have been a little misleading. They have led cognitive scientists to believe that maybe you could learn structure dependence because you have data for it. And there's a huge literature trying to show one way or another that you might acquire this, all of which collapse on analysis. But there's plenty of cases where the evidence is just zero, totally zero. So take for example the sentence *the guy who fixed the car carefully packed his tools*. It's ambiguous: he can fix the car carefully or carefully pack his tools. If *carefully* appears in the front as in *carefully the guy who fixed the car packed his tools*, it's unambiguous. You unambiguously pick the more remote verb for the adverb to modify, not the closest verb. Evidence for that is zero. There's no way you can learn that by deep learning analyses of the *Wall Street Journal* corpus or something even more voluminous. And this is universal in language.

So there is a puzzle. And there is in fact an answer. The answer is that language is based on the simplest computational operation—namely Merge, which yields no linear order. So the option of using linear order just isn't there in the internal core of language. So here's a situation where there is no learning necessary because that's just innate. There's no evolvability problem because it's the simplest computational operation. So you have a genuine explanation of quite a deep property of language, a strange property of language: structure dependence, no linear order.

There happens to be confirming evidence from other sources in this case. There are neurolinguistic studies initiated by a linguist who you know, Andrea Moro, who devised the study, working with a group of brain scientists in Milan. They took two groups of subjects each of which was presented with an invented nonsense language. One of these languages was modeled on a real language that the subjects didn't know; the second used rules that violate UG like those based on linear order. So for example in the second language, the invented one, negation would be say the third word in a sentence, which is very trivial to compute. In the one modeled on a real language, negation was whatever negation is, which is pretty complicated when you look at it. Well what they found is that the subjects who were presented with the invented language modeled on a real language handled it with activation in the normal language areas of the brain (mostly Broca's area). In the case of the subjects presented with the language, say in which negation was the third element, they didn't get Broca's area activation. They got diffuse activation over the brain, which indicates that it's just being treated as a puzzle, not a language problem.

Similar evidence had been presented by Neil Smith and Ianthi Tsimpli, who work with a particular subject who has very low cognitive capacities, extremely low, but has remarkable language capacities (picks up languages like a sponge basically). When they presented him (Chris, they call him) with the invented language based on a real language, he learned it very quickly. When they presented him with the one that has, say, negation in the third position, he couldn't handle it at all. It's a puzzle; he can't deal with puzzles.

So here's a very striking case, a rare case, where we have the optimal explanation for something, perfect explanation, genuine explanation, for a fundamental property of language. We have confirming evidence from all the relevant sources: language acquisition (it's known right away), neurolinguistics (you've got the right kind of brain activation), and psycholinguistics (you get dissociations). You couldn't ask for a better result.

Now here's a very curious fact about the field of cognitive science altogether. There's a huge literature on this. In fact a major topic in computational cognitive science is trying to show how this could be learned somehow from data, how it could be learned without what's called *inductive bias* for structure. If you look at this literature, first of all the methods that are proposed—if they are clear enough to investigate—they all of course fail. But a more important fact is it wouldn't matter if they succeeded. So let's imagine (though it's an incredible stretch) that you could, say, take the *Wall Street Journal* corpus and use some deep learning method and figure out from it that speakers of English use the structural rule. The significance of that would be absolutely zero because it's asking the wrong question. The question is: why is this the case? In fact why is it the case for every construction in every language, and without evidence. So even if, inconceivably, you could show how it could be learned, it wouldn't matter. It's totally irrelevant because it's the wrong question.

What about the inductive bias? There is no inductive bias; it's just the simplest explanation. So it's as if this entire effort (which is a large part of computational cognitive science) is an effort to try to disprove the null hypothesis. Now if you think about it, that just doesn't happen in the sciences. It's just not the kind of research project you undertake. Like if there's a perfect solution for something, you don't try to find some complicated way in which you could have gotten the result some other way – more accurately, if you look at the actual work, some tendency to approach the result from a distance on the basis of huge amounts of data and extensive statistical analysis. It doesn't make any sense. This raises serious questions about the nature of the field. It's a strange departure from the sciences.

There are other successes, but starting next time I want to turn to something else: problems with Merge, the places where it really doesn't work and what we can do about this.

Lecture #3 (May 31, 2019)

Yesterday, I talked about some of the ideas that work out, some of the successes in finding genuine explanations, a term that I had discussed. That's actually something really new in the history of the subject. Until recently, even the concept of genuine explanation wasn't well understood and there were no examples of it. One of the striking developments over the past two years is that there are some genuine explanations (in the sense that I discussed) for some fundamental properties of human languages; and there are some rather surprising consequences which overturn long-held beliefs. Well, that was the successes.

Today, I want to talk about some of the failures, some of the things that don't work, and what they imply, and some of the ways around them. And we'll see that those two have some curious and unexpected and rather interesting consequences.

So let's start with X'-theory. X'-theory had some successes, some failures. One of the successes was it eliminated the major problems of phrase structure grammar. The fact that phrase structure grammar permitted vastly many rules that are totally unacceptable, that couldn't be tolerated. Also phrase structure grammar conflated three distinct properties of language: compositionality, linear order, and identification of the categories (projection). X'-theory separated linear order from the other two, which has lots of consequences that I discussed last time. There were problems. The problems were two. One, it still conflated two distinct concepts, joining elements together and identifying their category. And secondly, and even more fundamentally, it ruled out "exocentric constructions," which abound in language. They are all over the place. And that required all sorts of artifices to figure out how to get what you knew was more or less the reasonable answer, but without any justification for it. So we want to get rid of that. And the simplest way to get rid of that in the most principled way is to assume the strong minimalist thesis (SMT) and see how far you can go from there. That means you take the simplest

compositional operation, the one that has to be embedded in any computational procedure, and see how much can be explained in those terms. And I discussed some of the successes last time.

However, there is an immediate problem: what about exocentric constructions? So suppose you're constructing a subject-predicate construction. They have to be done in parallel. You have to construct the verb phrase and the noun phrase before you can merge the two together. That means you have to have a workspace in which syntactic objects are constructed. And the workspace will consist of, first of all, the atoms of computation (call them lexical items), and also anything constructed by the computational operations. So the workspace is just the set including the atoms and anything constructed by Merge.

Let's call this "capital merge" (i.e., MERGE) because it's different from the concept "Merge" that has been used in recent years. So MERGE actually is going to take two syntactic elements (call them P and Q) in the workspace WS and put them together as {P, Q}, thereby forming the new workspace WS' containing the new element created.

$$\text{MERGE}(P, Q, WS) = [\{P, Q\}, \ldots] = WS'$$

So it is an operation on the workspace which is going to change it. It's a much broader operation. Note that the workspace is a set, the set which includes everything that's been constructed; but it's a different kind of set from the sets that formed by MERGE. It's not accessible to computation. You don't merge the workspace with elements within it. So to just indicate that I'll use a different notation. I'll use square brackets [ ] for the set that is the workspace: the result of merging P and Q, and then other objects that will constitute the resulting workspace.

And that raises the question: what's here? So what else is in the workspace? Well, let's take the simplest case, the case where P and Q are the only elements in the workspace. So the simplest case is the workspace that consists of P and Q.

$$WS = [P,Q]$$

And then you ask, what does MERGE yield? It yields the set {P, Q}—and what else?

$$WS = [P, Q] \rightarrow [\ \{P, Q\}, \ldots]$$

Well, if this was normal recursion, the kind of recursion that applies generally, the rest of the workspace would include P and Q.

$$WS = [P, Q] \rightarrow [\{P, Q\}, P, Q]$$

So for example, if you're doing, say, proof theory, general recursion, you have axioms and you have rules of inference. If you apply the rules to either the axioms or some theorem you've already constructed, you get a new theorem, and what is accessible to further operations is everything that's

already been produced: the axioms and every previous theorem. So that is essentially saying you include the set {P, Q} you've just formed, and P and Q.

However, that fails for human language. And it falls in a very straightforward way. For example, if you were to do this, then you could take {P, Q} and turn into something of arbitrary complexity violating every imaginable island or any other property—call it X, and then you MERGE P (a member of the WS) to X. You'll get {P, X}. But this P in {P, X} would be a copy of that P in {P, Q}.

$$\text{MERGE } (X, P) = WS_i = [\{P, X\}, X, \{P, Q\}, P, Q, \dots],$$

where {P, Q} is deeply embedded in X. You will now have violated every imaginable combination of islands and other linguistic properties. All islands would be violated no matter how radical they are, and this would be done by simple legitimate operations.

So that tells us that there is something different about recursion in human language versus general recursion. We have to bar some elements that were already there. They can't remain in the workspace. That raises interesting questions.

Actually, I should say that the original definition of Merge back in 1995 actually incorporated this property, but without recognizing it. It was really (if you look back at it) an operation that said "replace" P and Q by the set {P, Q}. And the formalizations of this, if you look at them, did add an operation (an operation Remove) to get rid of elements that you've already had. That, we don't want. We don't want to have extra new operations. It should be that the right concept of MERGE just has this property. But that raises the question why do we have that property. Why should language have the property that it deviates from normal recursion by getting rid of the things that had been generated and you are not allowed to access them anymore?

Well, that raises an interesting question. We would, of course, like to reduce this to some general principle, some essentially third factor property. Remember that in the study of growth and development, acquisition of language in particular, there are three factors you have to look at. One is what is genetically determined—that is, UG. Second is whatever the data happened to be. The third is principles that hold independently of the operation in question. That would include laws of nature and special things about the system that is carrying out the operation. So some properties of the brain in our case, which make it language ready. That possibility has been noted for long time, but nothing was ever said about relevant properties about the brain—about which very little is known in fact. But what we would look for in this case is some property of the brain that forces this. That would be a third factor principle. That would be the optimal conclusion.

Before coming to that, notice that there are already embedded in linguistic theory as its

developed over the past years, some examples of this. One case is that Merge in the old sense never increases the workspace. Merge itself doesn't increase the workspace unless it happens to be bringing in a new lexical item. So the old definition just didn't allow arbitrary expansion of the workspace. More interestingly, conditions like the Phase Impenetrability Condition (PIC), a fundamental condition with many important consequences, reduces the set of objects accessible to MERGE. It says there are elements you've produced that you can't see anymore. That restricts resources available for computation. Another case is simply c-command. If you think about successive movement of something, we take it for granted that if you're doing successive cyclic A'-movement (let's say), the next move will pick the top element, not the lower elements. Well what that means is that you're using essentially minimal search. You're using a computational principle, a third factor principle, that says look for the closest copy and don't look any further. Well again, that reduces accessibility.

But we need something beyond. We already have a number of proposals to reduce accessibility and hence limit the richness of computation; but we want something beyond that. And actually there is something highly suggestive, a property of neural computation which has been noticed elsewhere. A very striking property of neural computation is that the brain is very slow: it's a very slow computer. And that shows up in pretty dramatic ways. I'll quote from Sandiway Fong, a cognitive neuroscientist who's a colleague of mine at Arizona:

> "The marvel that we call the human brain is actually the weak link in our cognitive apparatus. Our sensory apparatus far outstrips the brain's capacity to process the high-resolution input to the sensory organs. The eye responds at the single photon level, the minimal possible detection, the eardrum vibrations respond at a level smaller than the diameter of hydrogen atom."

Not for everybody, incidentally (I'm an exception). But in general, it's pretty remarkable. And it's the same for every sensory system. They're essentially perfect, physically perfect. The brain can't handle that. So almost all of the computation that the brain is doing is throwing out information. The brain isn't capable of handling the amazing capacity of the sensory systems to be physically optimal detectors, better than anything that can be constructed. It does raise a strange question. Why did mother nature bother to produce these astonishing systems if the next step is to get the brain to throw out all the information that they produce? That's not true just for humans, it's true for all organisms, so it seems. It's a question. You can ask why that happened? Probably there's some physical law that explains that if you are going to have sensory systems, you have to make them perfect. But, what the reason for that is unknown; it just is a fact.

Anyway, the brain is essentially throwing out almost everything that comes in. And it is very tempting to think that the resource limitation that's part of language computation (and

presumably part of the computation that's done by any organism), that it's just a special case of the fact that the brain is so slow and incompetent, it wants to get rid of tons of information. That's a plausible candidate for accounting for the fact that one of the properties of linguistic recursion, as distinct from recursion in general, is simply that it's got to work through the brain. Remember that whatever the properties of the brain are that implement the linguistic computation is some third factor condition which is just presupposed by language—presumably by any organic computation.

Well, if you think back a step, we've been trying all along in pursuit of genuine explanation to restrict the amount of computation that takes place, to find the simplest theory, the one with the fewest mechanisms, least computation. And there were good reasons for that, which I discussed. One, just the general fact that simplicity of theory corresponds to the depth of explanation, which is what you're trying to achieve in any serious scientific inquiry. Secondly is this Galilean precept, which has been spectacularly successful about nature being simple. And third, in the special case of language, there are the empirical conditions on language evolution, which we don't know for certain, but seem to show, as I discussed, that it was a very rapid development which hasn't changed since (and of course, very recent in evolutionary time)—all of which converges to suggest that the computation should be restricted to the minimum.

Here we have a second point. Not only should computation be reduced to a minimum, but the resources available to computation (the set of elements accessible to the operations), that too should be reduced to a minimum. So we seem to have a very broad principle of reduction of computation (broader than normal explanation in the sciences) which says that not just are the means restricted to the minimum, but also the resources accessible to the means have to be restricted to the minimum. That has a lot of consequences. One consequence that I already mentioned tells us that the workspace is going to be as small as possible.

Well, let me just make a comment on notation. We want to say that [X], the workspace which is a set containing X is distinct from X.

$$[X] \neq X$$

We don't want to identify a singleton set with its member. If we did, the workspace itself would be accessible to MERGE. However, in the case of the elements produced by MERGE, we want to say the opposite.

$$\{X\} = X$$

We want to identify singleton sets with their members.

This is actually something goes back to phrase structure grammar. In phrase structure grammar you didn't allow rules that look like V→V. Nobody said why, but it's just the kind of

phrase structure rule that you didn't formulate. It's not allowed. For one thing it would get you into infinite loops. You don't want that, so it's just kind of barred. Now there was a trick used in phrase structure grammar, an illegitimate trick, which introduced VP—so if you have a verb phrase that was only a V, it would be described like this (I'm using the tree notation because it's easier to describe). You have NP VP, this VP would become V: VP→V. That's barred by X'-theory. In X'-theory, you don't have any VP if it's only a V. That was taken for granted in X'-theory.

Actually it has consequences. For one thing, for Richard Kayne's Linear Correspondence Axiom (LCA) it was necessary to introduce VP if you want to get the asymmetry between a subject and a verb. But that's illegitimate, so it's a problem for the LCA. And there are other consequences. Carrying that over into the Merge framework, what it comes down to is this ({X} = X). So crucially in the case of the workspace, we do not identify a singleton set with its member. In the case of something constructed by MERGE, yes we do identify a singleton set with its member. That has quite a number of consequences.

Well the simplest case that I just mentioned actually provides a criterion for determining what are the proper and improper forms of Merge. We have to ask for every notion of Merge that's developed, every case that's proposed, whether it yields illegitimate derivations. If it does, we have to throw it out. So take something that I think nobody has ever proposed, but let's just imagine it. Suppose we construct a syntactic object X with terms a and b unrelated to one another (on separate branches in the usual tree notation). And then we want to merge *a* and *b*. So that's going to give us ([{a, b}, X]).

There is nothing in the original definition of Merge that blocks that. It's just something nobody ever wanted to do, because it doesn't make any sense. But suppose you do do it, then what happens? Well, same story as before. X can become arbitrarily complex, call it Y. *a* can join to Y. And we've again violated all conditions. So this is ruled out because it allowed too much accessibility. We want a definition of MERGE that restricts accessibility so you won't be able to generate structures like that. Well, that was something that has never been proposed. But take things that have been proposed and are very widely used in linguistic description—like, for example, what's called Parallel Merge. Parallel Merge is usually described like this.
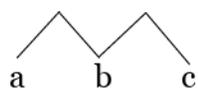
What you do is construct a tree notation in which the element b is part of {b, c} and also part of {a, b}. That's called Parallel Merge.

Well, I should make a comment at this point. The tree notations are kind of convenient, but they're very misleading and you should really pay no attention to them. For one thing, a tree

notation kind of leads you to suggest that there has to be something at the root of the tree. But that's conflating compositionality with projection. And in fact you often don't have anything at the root of the tree—for example, every exocentric construction. That's what labeling theory takes care of, which eliminates that conflation. Another reason is that when you draw trees, it looks easy to do lots of things that don't make any sense. So especially with contemporary computer graphics, you can draw funny lines from one point to another and connect those two elements. Or you can pack something in lower in the tree and call it Late-Merge, and so on. And it seems to sort of work out by playing around with trees, but if you try to translate it into MERGE, it doesn't make any sense.

So let's take Parallel Merge. What this really means is you started with a set {a.b}; this is the workspace now, and it had c in it.

$$WS = [\{a, b\}, c]$$

And then what you did was merge b with c, which gives you a new workspace WS', which consists of the original set ({a, b}) and the new one ({b, c}).

$$WS' = [\{a, b\}, \{b, c\}]$$

And here we have the old problem. The problem is that we can turn {a, b} into something as wild as we want, call it X, and then we can merge X with the b in {b, c} forming Y and again the two b's in Y are a chain that violates every possible condition. So Parallel Merge looks if it's easy to draw, but it can't be a legitimate operation because again it violates this paradigmatic problem—which tells you that you're proposing an operation which is going to yield illegitimate objects, objects that violate every imaginable syntactic condition. So that has to be ruled out.

Actually, those of you familiar with the literature know that Parallel Merge has been used to give lots of interesting results. It's the basis for trying to construct what's called "multi-dominance" constructions. That's been used for across-the-board deletion, for lots of other syntactic phenomena, to give accounts of them. Now those accounts in the literature are called explanations. But they are actually not. They are descriptions. They are interesting descriptions which have the property that I discussed earlier: they are illegitimate ways to impose some kind of organization on chaotic data—which is very valuable, often. It's more organized than before so we can kind of look and see if there is a possible way of accounting for it. But it's not an explanation. It's a waystation on the road to conceivable explanation.

So there's got to be something about the definition of MERGE that's telling you can't create Parallel Merge structures. And if you think about it for a second, you can see what it is. Too many accessible elements have been produced. The right definition of MERGE, when we get to it, should allow only one new accessible object—namely the one you're constructing.

When you put P and Q together, you're constructing a new object, the set {P, Q}. That set is accessible to computation, but nothing else should be. That should be the right definition of MERGE. (It might seem as though this excludes internal MERGE, but as we will see directly, it does not). Parallel Merge is adding two new accessible objects. It's adding the new set {b, c} but also the new occurrence of b in {b, c}. And that's too many. The right definition of the computation should permit only the minimal number—namely one new accessible object. And again, I suspect that that goes back to the general property of the brain that says it's just very slow. It has to throw out information.

Actually, before going on, I might mention that this is kind of familiar in other domains of language acquisition. So when you study language acquisition of phonology in infants, what you find is that of course every infant is capable of all the phonemic distinctions in any possible language. But by about a year old, or maybe even earlier, most of them have just been lost. What the early stage of phonological development is doing is saying throw out all the possibilities, and just leave us with these. When you study critical periods in language acquisition, it's pretty much the same. What you're saying is that at these apparently several critical periods, lots of stuff is just thrown out that you are not allowed to look at anymore. That's what makes a critical period.

Take a theory of language acquisition like, say, Charles Yang's, which assumes that the child just starts with all the possible I-languages and as data comes along, the probability distribution over the set of I-languages changes. So some data comes along and says, okay, I'll lower probability of this guy and raise the probability of that one. Ultimately, in the course of language acquisition, it need not converge in a single point, but it gives you a skewed probability distribution in which many languages are way down there, you're not going to bother with them. And there is one or maybe a couple of languages that have high probability, so you hang on to those and that's your I-language. You could again describe that in the same terms. It's saying what the acquisition system is doing is simply throwing out lots and lots of data. What we know the brain does massively because of the extraordinary sensitivity of the sensory organs.

Well anyhow, Parallel Merge goes and with it all of the interesting, very interesting, empirical consequences that have been produced that are based on Parallel Merge. Now those consequences remain, but as puzzles not as things that have been explained, as kind of organized data which now we want to look at.

Take a look at say, what's called Sideward Merge. Same problem. I won't run through it, but if you formulate Sideward Merge, not just as a tree with a line going from here to here, but in terms of actual MERGE, it has exactly the same problem. So this basic problem that we started with, the very simple problem, turns out to be a very good diagnostic to tell you when some proposed concept of Merge is legitimate or not. It's very simple, a very simple idea, and it cuts

very deeply.

Let's take Late-Merge. Well, Late-Merge has the same problem, but an additional problem too (which was pointed out by Epstein, Kitahara, Seely)—namely in the case of Late-Merge you're attaching something within the syntactic object – at a lower point in the tree in the (misleading) tree notation. One, it has the problem that I already mentioned. But secondly, it requires a substitution operation—namely, this new element constructed has to substitute into where one of its terms came from. How does that happen?

Well, you know, when you draw a tree, you can just draw it there, but that's not enough. You need some kind of operation that says that the new element you've constructed has to go right in the position of what you've constructed it from. It's a non-trivial operation; it's way beyond the bounds of SMT. So Late Merge, which has lots of consequences (it's used very widely), also can't be correct. And in fact I think most of the Late Merge literature (I won't have time to talk about this) probably reduces to Ellipsis in a way that I've discussed elsewhere, improved by Dennis Ott. But I'll put that aside here, just a problem to work on.

Anyway, all of these operations are out. All the consequences that follow from them (there's lots of stuff in the literature, very interesting consequences), they remain as just puzzles. Well, if you think about these proposals, they do fall within the original loose and vague characterization of Merge, so it's natural that people should have proposed them. They seemed to fall within the original (I won't say definition, because it was not clear enough to be a definition)—but the original characterization of the idea. And it's led to interesting results, but misleading interpretations of them. They are not results; they are simply presentations of interesting data, organized in a form that poses a problem. We have to figure out a genuine explanation for these phenomena, one that conforms to SMT, or as close as possible to it as we can come.

Well that suggests a research program. The research program is, first of all, to take all of the kinds of operations that have been proposed for Merge and ask yourself which ones are legitimate, and which ones are not legitimate. And then the second task is to formulate a new definition which captures just the legitimate ones and leaves out the illegitimate ones. And then the third part of the research program is to explain why. So those are the three steps.

I'll skip the first step. There is a lot of material in the literature showing which ones are legitimate, which one aren't. It runs right along essentially the lines I've just outlined. It turns out that the only ones that are legitimate are the simple ones that we had in mind when it was first developed: the narrow version, external and internal Merge. All of the rest: Parallel Merge, Sideward Merge, Late-Merge, and other similar proposals that have been made, they are all illegitimate, illegitimate on the very same grounds. They yield legitimate operations which

produce illegitimate results, hence have to be excluded. And they all have the property that they are increasing accessibility too much. So that's the first part.

We then want to formulate Merge so that it will block these operations. We want a definition looks like this:

$$\text{MERGE (P, Q, WS)} = [\{P, Q\}, \dots] = \text{WS}',$$

with two conditions, which are easy to formalize – I'll skip that. Just intuitively, the first condition that has to be met by this new definition is that nothing that was accessible in the workspace can be lost. More specifically, if something was accessible to MERGE in WS it has to still be accessible in WS' (which raises interesting issues that I'll put aside for lack of time). That's the first condition. The second condition is that the new definition of MERGE has to be minimal, the optimal definition. It has to restrict accessibility as fully as possible. Now MERGE itself is creating one new accessible object. That's the point of merging, you're creating a new object. But it can't increase accessibility beyond that. That's the second thing. And the third is just we don't want any arbitrary junk around. So not a lot of other stuff that had nothing to do with the operation. Those are the basically the conditions on the proper definition of MERGE.

That leads us to the next step: what's the explanation for it? And now we basically have the explanation. The first point, which says you can't throw things out, can be regarded as a special case of the No Tampering Condition (NTC), the general SMT condition of minimal computation which says you can't modify the elements that are entering into the computation. Well, the most extreme form of modifying something is to throw it out, okay. So you can't do that. So that's the first condition. The second condition is the resource constraint, a constraint which, I think, probably reduces to a general third factor property of the nature of the brain, the slow element in the cognitive system that's hampering cognition: the brain. So it probably just goes back to that. So we have the resource constraint which limits accessibility. The third thing that says don't throw in any extra junk is actually a consequence of restricting accessibility. If you throw in anything else, you're increasing accessibility. So we therefore have an optimal definition of MERGE which meets the all conditions we want. And we have explanation for all of them. So that's the set of flaws, failures, and a possible solution for all of them, a plausible one.

There is another consideration that we have to keep in mind, let's call it "stability". It's a general property of computations which we don't usually bother to think about, but we have to be explicit about it for reasons that show up when we try to explain things. So for example, if you're forming, say, topicalization, the mouth produces (i) but it's (ii) that is in the mind of the speaker:

(i)   Mary's book, I want to read

(ii)   Mary's book, I want to read Mary's book

The optimal internal MERGE operation doesn't delete anything. When we pronounce it, we drop the lower copy; but that's again for reasons of computational efficiency. But we are now talking about just what is in the mind. We're forgetting about what reaches the ear, the product of externalization, which doesn't really belong fully to language, strictly speaking.

Something that's taken for granted about this, but that we have to be clear about, is that the two occurrences of *Mary's book* have to be absolutely identical in every respect. So it can't be one "Mary" who is the topic and another "Mary" whose book I'm reading. It can't be that the topic is the book that Mary owns and the lower copy, the thematic object, is the book that Mary wrote, let's say. That's simply taken for granted. The interpretive system has to know when elements are precisely absolutely identical. So let's call that condition "stability"; there turn out to be consequences.

Now that immediately begins to raise some interesting questions. How are copies distinct from pairs that are identical in syntactic and phonetic form but interpreted independently. You can say

John saw John

That's ok, but those are two different *John*s, they have nothing to do with one another. Those are repetitions. On the other hand, here when you've done internal Merge, you have copies. How does the interpretive system know what's a copy and what's a repetition? They look alike, but it has to know them. That turns out to be a non-trivial question. There is an interesting paper by Erich Groat and Chris Collins (you can find it on LingBuzz**)** in which they go through a lot of problems with the various efforts to explain this. But you have to be able to explain it.

The interpretive systems have got to know what's a copy and what's a repetition. Now assuming phase theory (for good reasons), that means that precisely at the phase level, the system has to know the distinction. Well, of course at the phase level, you do have the information about what was produced by internal MERGE and what was produced by external MERGE. So you have some information, but it's not complete because there are cases of internal MERGE internal to the phase, like for example raising the predicate-internal subject to subject position inside the CP, internal to the phase. But at the phase level, you don't see that it's just happened here. So how can the interpretive system solve this problem? That raises quite interesting questions.

I think the basic answer to it is given by a general property of language, which is sometimes called "duality of semantics". If you look quite generally at the interpretation of expressions, it falls into two categories. There is one category which yields argument structure (theta-roles and the interpretation of complements of functional elements. There is another

category which is involved in displacement, which has kind of discourse-oriented or information-related properties or scopal properties and so on, but not argument properties. That's duality of semantics. If you think about it a little further, you see that the first type, argument structure, is invariably given by external MERGE. The second type, non-argument structure (other factors) is always given by internal MERGE.

Now there appear to be some exceptions to this. As usual when you produce a generalization, you look at details, and you may find maybe something doesn't seem to work. If it's a strong generalization, the proper approach, just as in standard science, is to put the exceptions aside temporarily, supposing that they may be misunderstood, and try to make the strong generalization work always. I think that's a good rule of thumb; we should use it here.

So I think duality of interpretation is probably a very strong principle. The question is where does it come from? Well, probably it's just a property of the nature of thought. We don't know a lot about thought. In fact, one of the main ways in which we come to understand thought is to ask how it's linguistically articulated. That's one of the very rare avenues into what's the nature of thinking. And if it's correct (as I've been suggesting throughout) that the internal language really doesn't care about use or communication but that it cares about expression of thought, then we expect the design of language to capture these fundamental aspects of thinking. That's what I'm proposing.

Now if you think about duality of semantics, you have a technique right away to determine what's a copy and what's a repetition. If something is in a theta position, it's not a copy (unless it's been raised, in which cases it's a copy of what's been raised). If it's in a non-theta position, it's a copy. And at the phase level, the system simply has to take a look and say what's in a theta position, what isn't in a theta position. That tells us what's a copy and what's a repetition. We can simplify that by taking the operation MERGE to produce copies always, including both cases of MERGE (external and internal MERGE). That produces copies, no other linguistic operation does. So a copy is just whatever is created by MERGE. In the case of internal MERGE, it creates two copies. In the case of external MERGE, it also creates two copies: the original occurrence and the one that's in the set you form. But the operation MERGE, which is in effect a replace operation, gets rid of one copy in the case of external MERGE (and by NTC it has to be the "loose one," the one not within the new item formed). So you never see it. So MERGE always produces two copies, but in the case of external MERGE, just by the nature of the operation, the minimal operation, only one of them remains. That's the resource restriction part.

Now that still raises interesting questions. If you start thinking of the complexity of constructions: double object constructions, topicalization, small clauses, and so on, it looks as if

there are ambiguities in searching for what is a copy of what. And what we have to show is that language has a conspiracy that determines the answer in each case. So for example, in a case of double objects, Vergnaud's abstract case theory may uniquely determine the answer. In a case of topicalization, when you have topic up there, you want to find out which of many elements that look identical it comes from. Luigi Rizzi's left periphery theory plus the theory of labeling which determines criterial positions will solve that problem. The case of small clauses is extremely interesting; but there is a way of resolving that ambiguity.

But here is a problem that I posed: there has to be a conspiracy in the structure of language that automatically resolves ambiguities about what's a copy of what. We have two problems about copies. One, what's a copy and what's a repetition That's solved by defining copy as just anything produced by MERGE (and no other rule) and by duality of semantics. Then the second question what's a copy of what? That has to come from some internal conspiracy about the nature of language. I think there are answers but it's not trivial. (I leave it as something to think through.)

I should point out in addition (you can think about this, I won't go into it) that none of this works for the illegitimate operations. So if you do Sideward Merge or Parallel Merge and so on, none of this is going to work out. There you do have the problems of copy/repetition and identification. That's an independent reason indicating why the illegitimate cases are illegitimate. The main argument is what I discussed: the yield illegitimate outputs. But it also turns out that the concept of copy/repetition and the procedure for unambiguously finding copies fail for the illegitimate cases. (You can try it out and see how it works for yourself.)

Well, notice that you can now rephrase the whole process of constructing MERGE and its application in different terms -- the normal way in which recursive processes are described. So suppose you want to define the set of integers. How do you define the set of integers? Well, what you say is the set of integers is the smallest set, the least set, containing 1 and containing the successor of any integer. That's the set of integers. That's transitive closure, the Fregean ancestral, the classic way. So what would we say here? Well, what we say is that for a given I-language, the set of workspaces—the set notice, not the least set—is the set containing the lexicon and containing MERGE (P, Q, WS) for any P,Q and WS that has already been generated. It's the same as the definition (the transitive closure) of the set of integers with one exception. We don't have to say "the least set." Simpler here. We don't have to say it because resource restriction already forces it to be the least set. So essentially we are in the right ballpark. We have a standard recursive inductive definition under transitive closure of the set of workspaces, and it's exactly where we want to be.

Well, I've kind of a skipped a lot of steps here, there's a lot that has to be filled in, but I

think maybe this is enough of a picture so you can see how it can be filled in.

Well, to finish for today (I want to go to another set of topics tomorrow), let's just ask what we can say about the things left unexplained. So let's take say ATB, across-the-board deletion. A simple case of ATB would be the following example.

What book did John read and Mary buy?

This is nicely handed in terms of multi-dominance and Parallel Merge. But, unfortunately, that's illegitimate. So we have to ask is there legitimate way of doing it? Well, the simplest approach would be to say that it just derives by deletion from

[What book]$_1$ did John **read** [what book]$_2$ and [what book]$_3$ did Mary **buy** [what book]$_4$,

occurrences numbered for later reference. So it's the conjunction of these two questions.

There is standard objection to this. The standard objection is that nothing tells you it's the same book in both cases. So that looks like a problem. However notice that we already have a solution to that problem—namely, let's take a look at the copies. 1 and 2 are copies, and 3 and 4 are copies, so 2 and 4 delete by the usual principle of copy deletion in externalization. What about 1 and 3? They are independent. No principle tells us whether they are copies or repetitions. So one option is that they are copies. In that case, 3 will delete by the usual externalization principle of deletion of all copies except the first (since deletion is part of externalization, linear order is available).

So we automatically get the right phonetic form, but what about the right interpretation? How about the fact that the two books might be different? Well, here the notion Stability enters. Deletion in general has the property quite apart from this (as in the topicalization case and Ellipsis, anything else)—that you have to have absolute identity; otherwise you can't delete. So if you want to delete under, say, VP ellipsis, or topicalization, or whatever it may be, there has to be absolute identity. That's the Stability property. So in fact, it automatically turns out—without any comment at all—that you get ATB. Just follow the principles already established.

Now there is something crucially important here—namely, you don't have c-command between 1 and 3. Suppose you did have c-command. Then it wouldn't work. So suppose you have the sentence

[Which boy]$_1$ did John ask [which boy]$_2$ [which boy]$_3$ Bill met [which boy]$_4$

("which boy did John ask which boy Bill met")

Here 1 c-commands 3. That means that at the phase level, 1 c-commands both 2 and 4, each of which has a theta role, a violation of duality of semantics. So 1 and 3 must be repetitions, not copies. And we do not have an ATB-type interpretation.

So crucially, the lack of c-command in coordination yields ATB, but the comparable interpretation is unavailable at the phase level when c-command holds.

[An aside for those familiar with phase theory. The problem actually arises at a position I've omitted, the *ask* phase. But the point is the same.]

Pretty much the same is true of the basic cases of parasitic gaps. Those of you familiar with the topic can try it your own. So the standard examples that they are used for multi-dominance and Parallel Merge fall into the place without any comment. They just come from following simple processes of minimal computation, resource restriction, and stability, the general principle of deletion. Well, the task is: show this for everything. Not a small task. I'll leave it there turn to some other problems tomorrow.

Lecture #4 (May 2, 2019)

Well this has been pretty sketchy, especially yesterday. There are lots of loose ends that ought to be tied up. I think it would be best maybe to delay that to the discussion period tomorrow, so I hope you guys will come in with lots of questions and objections, and we'll be able to see if we can fix up what wasn't done properly, of which there is a lot. There is actually one question that came up in the question period yesterday about an important matter, and I didn't actually give the right answer to it, so let me go through it here.

So the paradigmatic example that I was giving about what distinguishes the legitimate from the illegitimate operations can be illustrated with Parallel Merge, what is represented in tree notation as (A):

(A)
        a   b   c
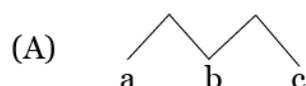
Let's now ask how this can be constructed by Merge. We begin with WS = [a, b, c]. We then have to merge two of the members, say a and b, yielding WS' = [{a, b}, c]. Then we merge b in {a, b} to c, by Parallel Merge, which yields the new set {b, c} and the workspace

$$WS'' = [\{a, b\}, \{b, c\}]$$

WS'' is the Merge-based structure represented as (A) in tree notation – and the comparison with (A) demonstrates how tree representations of syntactic structure can be misleading.

In WS'' there are two copies of b, two distinct occurrences. We put aside here formalization of these notions, which raises no problems relevant here.

WS'' is illegitimate for reasons already discussed. You can expand {a, b} to something as

47

complex as you like (call it X). Since b in{b, c}) is accessible, you can merge the occurrence of b in {b, c} to X forming the workspace WS'''.

$$WS''' = [\{b, X\}, \{b, c\}]$$

Now {b, X} has a chain connecting its two occurrences of b, which can violate all conceivable properties of language. That's illegitimate. The source of the illegitimacy is that the Merge-based operation assumed in Parallel Merge adds two accessible terms — namely, b in {b, c} and the set {b, c} itself. The operation violates resource restriction, which allows adding only one accessible term.

Note that Parallel Merge raises another problem. Merging b from {b, c} with X creates two b-chains, each headed by b in {b, X}, one terminating with the occurrence of b in X and the other with the occurrence of b in {b, c}. The ensuing computation and interpretation has no way to discriminate between them. There are good reasons, discussed elsewhere, why such indeterminacy should not be tolerated in the generative process, another reason to reject Parallel Merge.

A question raised yesterday was why the same reasoning doesn't block ordinary internal Merge/MERGE. So suppose you have {a, b} and merge b to {a, b}, yielding $X = \{b_2, \{a, b_1\}\}$ (subscripts for convenience), ordinary internal Merge. Why doesn't that violate the resource restriction condition? The operation creates two new accessible items: X itself and b_2. But recall that there are independent principles that restrict accessibility. One is PIC, which prevents further access to material in an earlier phase. Another is the third factor principle Minimal Search MS, a process that terminates once it reaches the head of a chain (a principle that has good empirical motivation, which I'll have to put aside here). MS therefore renders b_1 inaccessible to Merge/MERGE. There are two new accessible items, b_2 and X, but since b_1 is no longer accessible by virtue of MS, total accessibility has increased by 1, satisfying resource restriction.

These examples clarify just what the resource restriction asserts: it states that in mapping WS_1 to WS_2, the number of accessible items can increase only by one. It doesn't say anything about which items are accessible. Any application of internal Merge/MERGE applies to an accessible item X that is a term of Y, forming {X, Y}. It renders X inaccessible (by MS) and adds a new accessible copy of X as well as the newly accessible {X, Y}. Thus it increases accessibility by one, which conforms to the resource restriction. But the particular occurrences that are accessible change.

Well, there are a lot of other things that need to be clarified, but instead of that, I would like to touch on a few other topics briefly, just to bring them up, I won't be able to go into them with any care, but will just to identify them as areas of further inquiry.

One point is that, in addition to the symmetrical operation MERGE, there is quite good

evidence that we also need another operation, an asymmetric operation in addition to MERGE. This is clear intuitively from simple adjuncts. So, for example, if you have a phrase like *young man*, there's an asymmetry between the two words, that's clear. The element that's formed is a noun phrase, not an adjective phrase. So *young* is an adjunct that's not changing the category.

Notice this is quite different from symmetrical MERGE. In symmetrical MERGE, if you happen to have a head and an XP, then the head will provide the label – in earlier versions, what projects. But that's a case of MS (like most topics, not without some questions when we think about it carefully). If it's an XP-YP structure, it becomes an issue for labeling theory – partially understood I think, with interesting results, but with serious gaps if we look more closely. That doesn't work for adjunct asymmetry. So if you have *old portrait of John*, *portrait of John* is not a head, and could be arbitrarily complex, but it still provides the label of the whole unit. So there is a clear asymmetry.

Beyond that, there is a problem that's been lingering for sixty years, a serious problem. It's the problem of unbounded unstructured coordination. So we can have something like (I):

(I)   *I met someone young, happy, eager to go to college, tired of wasting his time, …*

with an unbounded number of possible adjective phrases and no structure among them. They're strung together like beads on a string, and unbounded. That's a real problem. For one thing you can't create this with any kind of phrase structure grammar because you need an infinite number of rules. That's even true, interestingly, of the universal case: unrestricted rewriting systems, the ones that allow you to do anything. You would still need an infinite number of rules.

Notice that unrestricted rewriting systems (as was proven by the great mathematician Emil Post long ago) are universal. They are equivalent to universal Turing machines, which means you can code the right result by some devious means, but you don't get it by just writing the rules and looking at the structure generated. So that's out totally within phrase structure grammar. About sixty years ago, when George Miller and I were working on mathematical linguistics, we thought we had an answer to this in terms of generalized transformations, but Howard Lasnik pointed out that that doesn't work either, so there is no method around, no device around, that allows this.

We need a new device and since we're trying to get the simplest possible operations, the simplest operation after set-formation is pair-formation. At some point in generation of an interpretable structure, each of the adjective phrases in (I) will be predicated of *someone* with an asymmetric relation similar to *young man*, presumably pair-formation, yielding an n-dimensional structure that is merged into the ongoing derivation.

So it seems we need an operation Pair-Merge, which will also apply to the simple adjunct case like *young man*. *Young* will be adjoined to *man*, but you don't see it in the labeling because

it's off in some other dimension. And the unbounded unstructured cases like (I) show you in effect that there are unboundedly many dimensions as adjective phrases are independently adjoined to the host. It's not two dimensional like a blackboard, but there's no reason to suppose that mental representations are restricted this way. You can add any number of adjuncts at any point. So we want to work this out. That means intuitively that every member of this coordinated construction, every individual member of it is individually predicated of what it links to.

Now we know that the order of the unstructured unbounded elements matters. It matters because of reasons that were pointed out by Jim McCawley over half a century ago, namely, notions like *respectively*. So if you say *John and Bill saw Tom and Mary, respectively*, the order in which they appear affects the semantic interpretation. So it's not just a set of paired elements, it's a sequence of paired elements. Same with adjuncts: *John and Bill are young and tall, respectively*. Sequences of this kind are found for all major categories and with other constructions, e.g., *John picked Tom and Bill in that order*. The problem is quite general. By raising it we are moving into largely unexplored territory, so a good deal of experimentation is in order.

More evidence that it is a sequence is that adjuncts can repeat. So you can say *the guy is young, tall, happy, young, eager to go to Harvard,* and so on and so forth—you can repeat them as often as you want. So what we have is a situation where in order to generate these objects, you generate syntactic objects in WS, select a finite set of these, and from that set form a sequence S (which could be any sequence of the elements of the set selected from WS). and that sequence S is the syntactic object that you're then going to merge into the construction to proceed with the derivation and interpretation.

The operation of picking a particular element out of a set is straightforward. There are formal ways of doing it which are familiar. (Those of you who know some logic will recognize that one is David Hilbert's epsilon operator, which picks a single element out of a set. It was part of his work on foundations of mathematics, a basic operation). So it's a straightforward operation, but it does have the property that it is indeterminate. So that's a part of the nature of generation (and secondarily production), as I discussed last time.

Well, there are clearly (at least) two kinds of coordination: conjunction and disjunction. So what we are doing is forming objects that look like this. We're forming a sequence which begins with some conjunction, and then contains elements, each of which is predicated of something. So we have a sequence of elements that looks like this, with links L_i.

$$< CONJ, < S_1, L_1 >, \ldots, < S_n, L_n > >$$

And in fact all the links have to be identical. One of the aspects of coordination is that you're

adjoining everything to same point and the coordination elements have to be similar in a way that should be determined by the common link. So we have an object like that. That's the basic object that gives you unbounded coordination. When you get down to the limiting case, when n = 1, that's just plain adjunction. So *the young man* and so on.

There are a number of possible ways to implement these general ideas, but to go into them here would carry us too far afield. They are a topic of current research.

Now, notice that each element of these pairs is inaccessible. So if you have the phrase *old man*, you can't extract *man* and leave *old*; you can't extract *old* and leave *man*. So the elements of the pairs are inaccessible. If that were all there was to it, this would yield both the coordinate structure and the adjunct island constraints. You have the coordinate structure constraint because every term is inaccessible. You have the adjunct island constraint because you can't pull the elements out of adjuncts. However, it's not quite that simple in fact, as I mentioned earlier, referring to a paper by Željko Bošković which reviews lots of different kinds of complicated cases. There are languages in which you can't extract anything inside the adjunct. There are others where you can extract something that's inside, but not the adjunct itself—and a couple of other cases. So there is more work to be done. And what it in fact shows is that the concept of adjunct is just not sufficiently refined. There are a number of different kinds of adjuncts which behave quite differently and this is in fact kind of an unexplored domain. Here is a major research project: to ask how we can deal with the class of cases that Bošković left unexplained in his bringing together the two kinds of problems.

Well, the next question is what is L? What do things link to? So let's take the simplest case, say, noun phrase and verb phrase coordination, for example, *John, Bill, Tom, the young man,… read the book, walked to the store….* What's the linking in those cases? Well, the assumption that comes to mind right away is that the link is the categorizer: n and v for NP and VP coordination.

[Before turning to that, some asides. I'm going to assume here that nominal phrases are actually NPs. The DP hypothesis, which is widely accepted, was very fruitful, leading to a lot of interesting work; but I've never really been convinced by it. I think these structures are fundamentally nominal phrases. Actually, there is very good paper by Masa Oishi that spells out how this could work. This could suggest that in a nominal phrase, definite articles are actually features of the nominal phrase, not elements merged into it, very much like Semitic, where the definite article appears in every element of nominal phrase, including the determiner. That seems to me probably the way it works. As far as determiners are concerned, like say *that*, I suspect that they are adjuncts. So I'll be assuming that the core system is basically nominal. As I said Masa's paper spells a lot of this out.]

So one suggestion you might entertain is that L for nominal sequences is just n, the categorizer of each of the coordinated phrases. But that runs into an immediate question. If you accept (as I am doing here) the Hagit Borer-Alec Marantz theory of root-categorization (which I think is pretty strongly motivated), the roots in the lexicon are independent of category. The sequence S is actually a Root Phrase (RP) sequence which is categorized by linking to n, v, a, in the basic cases. Note that this differs from the usual analysis, in which R is raised to the categorizer and adjoined to it, an operation of head-raising that raises many problems and is in fact unformulable in any available system that is at all clear. I'll return to a few comments on it. The problems are avoided if there is no head-raising, but rather independent linking to the categorizer. An independent argument for admitting Pair-Merge (or MERGE—no distinction in this context).

The standard examples in the literature, say *books fell*, are revised as the limiting case: a one-membered sequence <RP>, with the RP *books* linked to n. Many consequences follow, which I'll leave dangling. They are currently being investigated in ongoing research.

In fact, I think elements like n and v have been somewhat misinterpreted. They're not just nominal and verbal categorizers; they're basically phase markers. I think a natural way of capturing this is to go back to a classical distinction actually goes back to Classical Greece, and has been used sometimes in generative grammar) and to assume that we really have two fundamental notions. In classical terms they are the notions of *substantive* and *predicative*, which gives you four categories:

[+S, –P] = noun
[–S, +P] = verb
[+S, +P] = adjective
[–S, –P] = preposition

If you look at the classical Greek grammars up to late into the Roman period, these are the basic notions. The basic cases are pure substantive and pure predicative, noun and verb. There are no adjectives. Adjective we can add, but as a non-primitive category, both substantive and predicative. And then there is the fourth one, which is neither. The pure cases have the categorizers n and v, which mark phases (along with C). They are the links, what the RPs link to, assigning category. That's one possibility.

Notice that there is a well-known phenomenon about extraction that holds of both nominal verbal phrases. Only some nominal and verbal phrases resist extraction. For these, we can only extract to the edge, but not out of the phrase. We'll call those strong. In the case of verb phrases, it's the transitive verb phrases, the ones that we usually mark v*. Those constitute strong phases that you can't extract from, you've got to first go to the edge, then you can go on. In the weak

phases, unaccusatives and passives, you can extract without first moving to the edge of the phase. And there is a similar distinction in nominal phrases. So with regard to the Complex Noun Phrase Constraint, it's been known for a long time that it applies strongly for definites (specifics), but not for indefinites (non-specifics). There is a sharp contrast between *what did he believe the claims about* and *what did he believe claims about.* There are lexical and other effects, but this seems to be an overarching distinction

So it looks as though both the nominal and verbal phases subdivide into the categories strong and weak, and what we have been calling the two flavors of v (v and v*) are probably strong and weak phase markers—perhaps framed in the system just described. There has been a lot of discussion over years about whether noun phrases are really phases like verb phrases. One problem about it was that although they have a lot of similarities, there's one notable difference: you don't have an escape hatch in noun phrases. There is nothing that allows you to go to the edge and then go on. But fortunately, as I learned a couple of weeks ago from a student at MIT, there is an African language called Buli, which does have an escape hatch for noun phrases, a marker that says, OK, if I'm here we can move to the edge of the noun phrase and then go on. So if we simply assume that the basis for all languages is proto-Semitic and Buli, then everything works out— we're off and running, we just have to find the right languages. Other languages like English are defective in this respect: they don't have the extra morphological element that allows extraction out of noun phrases. That's one of the nice things about comparative work, you can often fill in the blanks that ought to be there.

Well if you play with coordination you'll notice that it can get pretty complicated. So for example, it could be that in a conjunction, one of the conjuncts is a disjunction. And that disjunction could be an unbounded unstructured element. And if you play with it, it gets very complex. I'll put that aside. The principles for dealing with it seem pretty straightforward. The actual working of it out becomes pretty hairy. But this at least seems not to be a fundamental problem, just a technical problem.

Now there are also fundamental problems, as was made very clear by recent work of Barry Schein's. Barry has a great book that we ought to study carefully. It is, I think, the longest book with the shortest title in the entire literature. The book is entitled *And*, and runs over 1000 pages with details about the extremely intriguing complexities that you get in the semantic interpretation of coordinate structures. It's all done in a kind of neo-Davidsonian event calculus. So here is the task, assuming that we have the right kind of syntactic formalism: how can you go from this to those semantic objects? I'll come back later to the nature of these semantic objects, but I think Barry gives strong evidence that that's the kind of target that the mapping from syntax to formal semantics should be trying to achieve. So here is a huge unsolved problem for those

of you who are looking for dissertation topics, one of many.

I should say incidentally that this is a dramatic difference from the 1950s. At that time it looked like there were almost no topics beyond application of fairly well established ideas to as yet unanalyzed phenomena . It looked like the whole field was essentially finished; I think I mentioned this earlier. One of the really striking things that's happened during my lifetime is that the field went from one that seemed terminal to one that is endless and open. Everywhere you look there are more problems. That's a good sign that at least something is on the right track, the way things ought to be. People who are just coming into the field now probably aren't struck by the dramatic difference, but if you look back it really is dramatic.

Among the many open questions are those I've barely touched on in these remarks, such questions as how to recast the generative enterprise in terms of the recognition that what has been intensively studied is really the limiting case of one-membered sequences, and related questions about how much of the construction of thought lies within the bounds of the operations that have so far been subjected to close examination – all questions of considerable obscurity but surely large import. It's far too much to undertake here, even to the extent that we now have some grasp of what is involved.

Let's keep to the narrower question of the possible scope of Pair-Merge. There are candidates apart from the very far-reaching topic of unbounded unstructured coordination, including some that may help resolve open problems. So for example, consider the standard examples

(i)   Mary saw the man walking down the street.

(ii)   Mary saw the man walk down the street.

Running through that familiar paradigm, you notice that there is a strange gap in it with bare verbs, as in (ii). You can't passivize with the bare verb, so you can say (iii) but not (iv):

(iii)   The man was seen walking down the street.

(iv)   *The man was seen walk down the street.

This holds for two categories of verbs: perception verbs and quasi-causative verbs (like *make* or *let*)—so *they made the guy walk down the street* and *they let the guy walk down the street* but not *the guy was made walk down the street* or *the guy was let walk down the street* Those two categories.

This has been a problem for a long time, with no good explanation. There is an interesting proposal by Norvin Richards in terms of his contiguity theory, in terms of the output of passivization. But that does not work because the same problem arises with *in-situ* passives, structures that are a little awkward in English but normal in very similar languages, sentences

like

(v)   There were seen walking down the street three men in dark overcoats.

That's more or less OK, but in contrast the parallel sentence with bare *walk* is sharply deviant:

(vi)   *There were seen walk down the street three men with dark overcoats.

So there is a sharp distinction between passives and non-passives whether or not you move the object.

So what could be involved? Well, what could be involved is suggested by the fact that this holds for the quasi causatives *let* and *make*. One might imagine that these elements are the English-style spelling out of causative morphemes, which normally attach to a verb, and are found all over the place. English tends to replace inflection with separate words – modals instead of verbal inflection for example, or *of*-constructions instead of genitive case for verbal and adjectival complements. So it wouldn't be too surprising to find causative inflection affixed by Pair-Merge in the underlying structure (*<let, walk>*, *<make, walk>*) and then spelled out with independent morphemes. Then comes an interesting question: why should perception verbs have the same property? But putting that on the shelf, if they all have this property, what you might expect is that all of these are pair-merged to the bare verb, and the pair-merged element is just immune to the passive operation, whatever you think passive is (e.g., dropping case). That would fill in the gap, making use of the Pair-Merge device.

There are other interesting proposals. One of them is Hisa Kitahara's suggestion about what has always been a pretty serious problem: head raising. Head raising has none of the right properties. It violates the Extension Condition. It's always described incorrectly. If a verb raises to inflection, say to T, it's always described as if the T-V complex becomes a T; but it's not, it's a V—the outcome of the adjunction is really verbal, not inflectional. And the further move of V to C is kind of V-second phenomenon, not a T-second phenomenon. So there are all kinds of problems with head movement.

Another kind of problem with it is that head raising has properties that cross syntax and phonology. So it's almost entirely like phonological processes in that it doesn't have semantic consequences. Consider Jean-Yves Pollock's analysis of French and English verb-raising, obligatory in French in cases where it is barred in English. Whether the verb raises or not, the semantic interpretation is the same. If you have a V-second language, it has the same semantic properties as a non-V-second language. In fact, you have to look very hard, way out at the fringes, to find some kind of semantic consequences to head raising. That makes it look phonological. But it often seems to be successive cyclic: root movement to categorizer to form X, X to T

forming Y, Y to complementizer yielding V-second (and none of these operations formulable, as just noted). So there are many problems, extending to head-raising quite generally.

Some might be obviated by the fairly radical shift to a pair Merge approach involving sequences and linking, as discussed briefly above; an independent argument for that approach, if it can be developed properly.

But keeping to more conventional approaches, Hisa's proposal is this. Essentially (I'll take T-to-C case, it's the same for the others) you generate {T, VP}. C is in the workspace, which we're taking to include the lexicon . Then you pair-merge C and T, yielding <C, T>. The workspace now contains <C, T> and {T, VP}. Notice by pair-merging C and T, you haven't increased the number of accessible elements because neither of these is accessible. That is just like *old* and *man* in *old man*, neither is accessible on its own. And once you've done this, you now merge <C,T> to {T, VP}, yielding {<C, T> {T, VP}}. That gives you the right answer. If you think about it, you could do the all head raising that way.

So that's a possible way of approaching the entire problem of head raising which has been a tricky and serious problem for a long time. I'll leave it in your hands as something else to explore, and to see how far this kind of analysis might reach in handling head raising without departing from the very spare framework that we've been assuming.

Well, there are plenty of questions like this. There are two major tasks. The first one is to account for the cases that have been described usefully by illegitimate operations, as discussed last time. And then there is the much broader task of trying to make sense of what hasn't yet been examined in anything like these terms. We now have a criterion for what 'makes sense': provide a genuine explanation, one that meets the crucial conditions of learnability and evolvability. If you can reduce the first factor to MERGE, you've solved that problem because there is no question of learnability and of evolvability, as I discussed earlier.

If you can reduce explanation to MERGE and Pair-MERGE, you've come pretty close. You still have to ask how Pair-MERGE could have evolved, but at least that's easier than lots of other things you can imagine. We would then like to try to show that other devices that seem critical for explanation, like say phase theory, can be reduced to third factor properties of minimal computation. They do in fact reduce computation, so if you can show that phase theory follows from that, rather like the resource restriction, you'll have a genuine explanation.

That's the general project, but notice again it should be clear (going back to the first lecture) that this project only arises if you accept a particular version of what the whole enterprise of linguistics is – only if you pick a particular kind of answer to the initial question: What is language?, What are we investigating? The enterprise attempts to address what I called the Galilean challenge, the challenge that was posed at the beginning of the scientific revolution,

which translates in our terms into at least capturing the basic property of language, the property of the faculty of language that permits the generation of an infinite number of structured expressions with semantic interpretations and the option of an ancillary operation of mapping them to one or another sensory-motor system.

If that project is accepted, the next question is whether to go on to accept a particular variant of it (which not all people who do generative grammar accept): taking the enterprise to be part of natural science. That means, in other words, adopting the biolinguistic program, which takes language to be a property of human beings (not some object out in the external world which somehow human beings connect with). If you go this far, then you have to meet the conditions of evolvability and learnability. If that is achieved, we have genuine explanation. And we can then turn to the kinds of tasks that are open—and challenging.

Notice again that not accepting this enterprise is perfectly legitimate. One can keep to just describing and systematically organizing linguistic phenomena. That's valuable in itself, and a preliminary to proceeding on to genuine explanation. It's also useful for engineering projects, the kind of Silicon Valley linguistics that is now prominent: just constructs something that more or less works, like a Google translator, however it's done. It's unlikely to provide any insight into language or cognition, but that doesn't undermine its usefulness.

There is another topic that I ought to at least mention. A computational system has operations and atomic elements, the smallest elements for the purpose of the computation. I haven't said anything about those. So what are the atoms of computation?

And at this point, we're moving towards the general domain of semantics and we want to ask how this domain relates to the enterprise? Well, here we have to be a little careful. You have to ask what we mean by 'semantics'. There is a classical interpretation of semantics that has roots far back in history, and in the modern period has been developed extensively by Gottlob Frege, Charles Sanders Peirce, Albert Tarski, Rudolf Carnap, W.V.O. Quine, and many others. This tradition is founded on the notions of truth/reference/denotation and studies such topics as entailment.

Let's take a look at those notions. First of all, entailment is syntax. It's sometimes called 'logical syntax'. Entailment is matter of the formal relations among the expressions. The world doesn't even have to exist. Semantics in the traditional sense is the field that tries to relate symbolic systems – within the biolinguistic framework, internal mental computations -- to the world, the mind-independent real world out there. That's semantics, based on truth and reference.

If you look at what's called formal semantics, some of the richest and most exciting work going on in the field in a last couple of decades, notice that it's pure syntax: symbolic manipulations of postulated entities that are not part of the mind-independent world, whatever their real-world motivation. It's sort of analogous to phonology. If you think about the

externalization process, what do we have? The syntax generates set of objects, we then want to externalize them. The process of externalization has several steps. The first is traditionally called morphophonemics, or phonology in the sense in which Morris Halle and I used the term, a mapping of the syntactic structure to some sort of phonetic form. The phonetic form is a syntactic object. And the mapping is symbolic manipulation. So it's again syntax in a general sense.

I should say that this project of generative phonology, which maps syntactic structure to phonetic form, has pretty much been abandoned in contemporary phonology, mostly optimality theory, which is concerned with other questions.

Generative phonology sets the stage for another task: determining how the generated phonetic form relates to non-linguistic objects like articulators and sound waves – or if you're using sign, gestures and facial expressions.

Formal semantics is rather like generative phonology. It consists of an array of syntactic operations setting the stage, one anticipates, for eventual relation of the symbolic system to the outside world.

Within the framework we have been discussing, the study of language is basically syntax. Narrow syntax constructs the syntactic objects. That's what I've been talking about the last couple of days. Another part of syntax is generative phonology, which sets the stage for construction and interpretation of extra-linguistic objects, like sound and gesture. And another part of syntax is formal semantics. All of this is going on inside the head. It doesn't get to the outside world. It's neither phonetics nor semantics.

Formal semantics differs from the other two syntactic systems in the way it has been pursued. It has not sought to find genuine explanations, in the sense of the term we've been discussing, addressing the problems of learnability and evolvability. There is little effort to ask what's the simplest way to proceed. The goal has been to find some means to accomplish the task at hand. This is not a criticism. The tasks are hard and in large measure newly discovered. And there may be principled reasons why genuine explanation is not an objective. I'll return to this in a moment. But the questions should not be overlooked.

In the far more circumscribed case of generative phonology, there has been a lot of work from the outset of the generative enterprise 70 years ago to find optimal solutions, the simplest and most fully explanatory. In narrow syntax, facing new and more intricate problems, the effort has proceeded along the lines that we've been talking about. In formal semantics, the questions have been peripheral.

Some of the questions may not arise for principled reasons, in particular questions of learnability, which have always been driving considerations in narrow syntax because of the extreme poverty of stimulus issues. Those questions may not arise at all for formal semantics. It is kind of tacitly assumed

to be invariant. There is no serious proposal, to my knowledge, that languages have parametric differences in this domain. That's a very reasonable assumption. For the phenomena that are being investigated, there's virtually no empirical evidence available to the child for language acquisition. So it does make sense to suppose that this part of syntax is invariant.

I've talked a little about whether narrow syntax might be close to invariant. That's an open question. It certainly doesn't look like it on the surface, but a lot of the work that's been done has tended to restrict the assumed variability of narrow syntax to the point where you might imagine that the core of narrow syntax has little or no variation, that parameter setting is restricted to externalization and the lexicon (and in the latter case, to which I'll turn in a moment, much less than is often assumed). These are further open questions.

That's the kind of picture that seems to be emerging, I think. It leaves open the question of how truth/reference/denotation fit in—true semantics in the traditional sense. How do we relate what's going on inside the head to the outside world? It turns out not to be such a simple question. The proposals in formal semantics sound as if they are relating to the outside world, but when you look at them closely, they are actually not.

Take say model theoretic approaches. One tends to treat the individuals and predicates postulated in the model as if they are actually out there in the mind-external world. But in fact, whatever their external world motivation, they're mental objects. There's nothing in the outside world that corresponds to those postulated objects of the model. Model theoretic semantics works very well. You can do all kinds of things with it. But it's pure syntax. It's not getting to the outside world.

What about neo-Davidsonian event semantics, a very rich and productive field. What are events? Does the world come packaged in events? Actually not. Events are our mental constructions imposed on whatever is going on. So events are really internally constructed. We know that to try to count the number of events in some mind-independent way is kind of meaningless. So for example, thanks to Zeno, we know that if you walk across the room, there is a continuous number of events. Same with anything else you're looking at. There are the events your mind imposes. So event semantics is again another form of syntax.

How then do we get beyond syntax? Well, we have to look at the notion of reference. Truth will depend on reference. To investigate 'reference', we start naturally by looking at the words that are used to refer. In the literature on reference, mostly philosophical literature, *London* is taken to be a proto-typical referential word. Lot of problems arise (say Kripke's puzzles of belief when you try to ask questions about London). But is London a thing in the outside world? Suppose for example, I tell you that I visited London before it was burned down and then rebuilt twenty miles up the Thames out of entirely new materials. Is there an object in the real world that

is physical because it can be burned down, but then can be rebuilt somewhere else with different physical objects and looking different? Can there be such an object in the outside material world? We could decide to rebuild Carthage, let's say. It was gone a couple thousand years ago, but we could rebuild it somewhere else where it would look different but it could still be Carthage. So whatever such words are, they're not names of things in the mind-independent world.

What about other words? We might raise the question whether there any words in language that refer to anything in the outside mind-independent world. I think the answer is that there aren't, for normal non-technical language. Some of the reasons for this were given as far back as classical Greece. So, for example, Aristotle asked the question what is a house? His answer was that a house consists of the amalgam of two different kinds of elements. In his metaphysics, they are matter and form. So the matter of the house is the timbers, the bricks, the stuff it was made of. The form of the house is the design, what it's for, what the architect had in mind, how it's used, and so on. So something might look physically just like a house, but in fact be something totally different. It could be a library, let's say. It could be a garage, a paperweight for a giant, all kinds of things. It depends on what the architect had in mind and how it's used, and what it's for basically. But the properties of form are not in the physical world. They're in the mental world. Now when Aristotle talked about this, it was metaphysics. It's what a house really *is*. In fact, it's in his book *Metaphysics*.

When we move on to the 17th century again, there was a kind of a cognitive revolution and a lot of these ideas were reinterpreted in terms of modes of cognition to which experience conforms, which I think is the right way to look at them. So reinterpreting it in those terms, a house is something that we construct in our minds, which has a material element, but a crucial part of it is what Aristotle called the form. That's something that's part of our mental operations. When we use the word *house*, we're referring to a mind-independent object which we interpret as a house. Referring is an action, but the word *house* does not refer.

If you think about the simplest words in the language, it comes out the same way. To take an example earlier than Aristotle, the pre-Socratic example philosopher Heraclitus famously argued that you can't cross the same river twice because when you cross it the second time it's a completely different physical object and you're a different physical object. So it's impossible to cross the same river twice. Of course, we do cross the same river twice, which simply tells us that our notion *river* does not refer to a mind-independent entity in the material world. We use the term to refer to a construction of our minds, and if you start playing around with how you individuate these objects, it becomes very tricky.

So, for example, take the Charles River, which I often crossed on the way to Boston. Suppose you reverse the course of the water, so it goes in the opposite direction. You've still got

the Charles River. Suppose you divert it into a different direction because you don't want it to go into the bay, but somewhere else; it's still the Charles River. In fact, you can make all sorts of massive changes in the physical object and it will still be the Charles River. On the other hand, there are trivial changes that will prevent it from being a river at all. So suppose you put barriers along the side and you start using it for, say, commercial tankers going up and down. Then it's a canal, it's not a river anymore. Suppose you even make a more minuscule change. You make what in physics is called a phase change from the liquid to the glassy state (tiny change, almost undetectable without instruments). But now it has a hard surface and you paint a line down the middle and you start using it to commute to Boston. Now it's a highway. You've made an almost undetectable change. If you play around with it some more you can make huge changes in the physical object and it stays the Charles River; you can make virtually undetectable changes, it's not a river at all.

Take a look at any other word in the language; you find pretty much the same.

All of this must mean that there is very little learning involved in acquisition of a word. No one has empirical evidence for these properties of words. Certainly children do not. Work on child acquisition, particularly Lila Gleitman's, reveals that children acquire words on very few presentations, not by teaching but from the normal environment, which provides little indication of what words are used to refer to, if anything. The rich and complex meanings of words must somehow be available independently of experience. What is learned seems to be primarily sound-symbol association, Saussurean arbitrariness. The rest is more mystery.

It may be, then, that language simply does not have the concept of truth and reference at all, which means it doesn't have semantics. There's syntax and there's modes of using the symbolic objects your mind has constructed, roughly what's called pragmatics, including speech act theory. My guess is that's probably what we're going to find when we pursue the topic further. That raises many very interesting questions. So, for example, take these terms like *house* or *river*, and *London*, whatever you like. Where do they come from?

If you look at animal symbolic systems, it seems that the atoms of these systems, the elements within them, do in fact pick out identifiable physical events. So take the bee that has my name attached to it (so I recently learned). If it's a honeybee then it performs the famous waggle dance: it flies out to a flower and comes back to the hive, waggles, and the symbolic actions that it's performing are one-to-one correlated with particular physical phenomena extraneous to the bee: the distance to the flower, the height of the flower, the orientation in which other bees have to fly, the quality of the flower and so on. Every animal system we know is like that. Monkey calls, for example. There's something we call a warning call, meaning that the leaves are fluttering in a certain way. The monkey reflexively emits some noise, other monkeys run away,

maybe an eagle is coming, something like that. Animal systems seem to be like this completely. The human systems are not like this at all. That raises a very interesting question. Another one of those mysteries, where did this come from?

If you look at the array of mysterious things, there are plenty of them and they fall into two categories. One is the category of mysteries where we can't get the right kind of evidence because it's empirically impossible for us to do it. We can't go back, say two hundred thousand years, and hear what people were saying. In principle it's a theoretically possible task, but you just can't do it. That's one kind of mystery. But then there are the deeper kinds of mysteries which have to do with things like selecting the sequence when you're trying to form a coordination, or more generally just picking what sentence you're going to produce. Questions of choice of action. These kinds of mysteries are the kind where we don't even have bad ideas. There's no idea as to how to proceed.

We can then ask what that category of mysteries tell us. It might involve some restriction on human cognitive capacities. There's a lot of debate about this. Most scientists vigorously reject the idea that there could be a limit on cognitive capacities of humans (unlike other animals, where it's uncontroversially true). That seems to me a strange position. If humans are part of the organic world, not angels, we ought to have the same general properties as other kinds of organisms. Every organism we know of has both scope and limits of cognitive capacity. And in fact the two are kind of logically related: whatever is yielding the scope is also imposing certain limits. And you might ask why humans should be exempt from this.

I'll leave that with you as another problem to contemplate.

Selected bibliography

Bošković, Željko. 2019. On Unifying the coordinate structure constraint and the adjunct condition. In A. Bárány, T. Biberauer, J. Douglas, and S. Vikner (eds.) *Rethinking Grammar*: *a Festschrift for Ian Roberts*. Language Sciences Press.

Chomsky, 1949. *Morphophonemics of Modern Hebrew.* Undergraduate honors thesis, U. Pennsylvania. Updated 1951 version published in 1979, Garland.

Chomsky, Noam. 1995. *The Minimalist Program*. M.I.T. Press.

Chomsky, Noam. 2000. Minimalist Inquiries: The framework. In Roger Martin, David Michaels, and Juan Uriagereka (eds.). *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89-155. M.I.T. Press.

Chomsky, Noam. 2001. Derivation by phase. In M. Kenstowicz. *Ken Hale: A life in language*. 1-52. M.I.T. Press.

Chomsky, Noam. 2005. Three Factors in Language Design. *Linguistic Inquiry* 36:1, 1-22.

Chomsky, Noam. 2019. "Some Puzzling Foundational Issues: The Reading Program". In Á. J. Gallego & D. Ott (eds.), *Catalan Journal of Linguistics. Special Issue: Generative Syntax. Questions, Crossroads, and Challenges*, 263-285. https://revistes.uab.cat/catJL/article/view/sp2019-chomsky/287-pdf-en.

Chomsky, 2021a. Linguistics Then and Now: Some Personal Reflections. *Annual Review of Linguistics.*

Chomsky, 2021b. Reflections. In Nicholas Allot, Terje Lohndal and Georges Rey, eds. *A Companion to Chomsky.* Wiley-Blackwell.

Chomsky, N. and G. Miller. 1963. Introduction to the Formal Analysis of Natural Languages. In R. D. Luce, R. Bush, and E. Galanter (eds.), *Handbook of Mathematical Psychology*, vol. 2, 269-322. Wiley.

Epstein, Samuel David, Hisatsugu Kitahara, and T. Daniel Seely. 2012. Structure building that can't be! In Myriam UribeEtxebarria and Vidal Valmala (eds.), *Ways of Structure Building*, 253-270. Oxford and New York: Oxford University Press.

Groat, Eric & Chris Collins 2018. Copies and Repetitions. LingBuzz (https://ling.auf.net/lingbuzz/003809).

Kayne, Richard. 1994. *Antisymmetry of Syntax.* M.I.T. Press.

McCawley, James D. 1968. The Role of Semantics in a Grammar. In E. Bach and R. Harms (eds.), *Universals in Linguistic Theory*, 124-169. New York, NY: Holt, Rinehart & Winston.

Oishi, Masayuki. 2015. The Hunt for a Label. In (eds), Egashira et al. In *Untiring Pursuit of Better Alternatives*, 322-334. Kaitakusha.

Pollock, J.-Y. 1989. Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20:365–424.

Richards, Norvin. 2016. *Contiguity Theory*. M.I.T. Press.

Schein, Barry. 2017. *And: conjunction reduction redux*. M.I.T. Press.

Sulemana, Abdul-Razak. 2017. "Q-particles and the nature of Covert movement: evidence from Bùlì," ms.