The Minimalist Program proposes to rebuild a theory of the language capacity from absolutely minimal assumptions. While this goal is absolutely legitimate, I will show that the implementation adopted by the Minimalist Program is in many respects empirically and methodologically inadequate. An alternative minimalist approach, based on more robust basic principles and a constructionist view of the relation between lexicon and grammar, offers a more satisfactory starting point on grounds of empirical coverage, learnability, and possibly evolution.

There are views of linguistics afoot outside of the field of linguistics that stem from a vision from the late 1960s or early 1970s. Then, there are some ideas that are floating around that are more recent. Most of this is associated specifically with the work of Chomsky. Although I was a student of Chomsky and I think he is 150% right about some things, over the past 10-15 years I have come to think that he was wrong about some very important things as well. There are alternatives on the market that meet his primary goals better than his own way of doing things. Some of this is in my book, Foundations of Language, and some of it is in a book that is now in press that I have written with Peter Culicover called, Simpler Syntax.

The primary goal of contemporary linguistic theory is an explanation of how a child attains adult competence in language, and I think this is Chomsky’s most lasting contribution to intellectual life in general. Here, we construe competence as an ability to use this very rich combinatorial system in a creative way. A theory of the child’s ability to learn language is under the constraint that it must show how the child arrives at the full complexity of human language, something I call the Descriptive Constraint. The theory must account for the full complexity of adult language. The more complex competence proves to be, the more there is to explain for a theory of acquisition.

One theoretical strategy to explain acquisition is to minimize what the child actually learns. There are two ways to do that that have been tried and true for a long time. One is sort of standard science, to say that underneath all this complexity, there are really much more general principles that interact in such a way as to produce complexity; and everyone is trying to pursue this strategy. A second possibility is to acknowledge that there is a lot of complexity, even when we pare it down to the most general principles, and it is hard to explain how the child learns this, so we can claim that a great deal of this complexity is part of the innate endowment of the child. The child comes knowing that language is going to be a certain way, and the learning process involves picking a way through a rather limited range of options that the heard language presents. This is the strategy of Chomsky’s theories starting around the late 1970s and 1980s, the so-called theory of Principles & Parameters.
This latter strategy of innate capacity is in tension with another constraint on the theory. If humans have an innate ability to learn language, and chimps do not, then there had to be some changes in the potential for language over evolutionary time. These changes might have been chance mutations, or they might have been driven by adaptation—we do not know, but we would like to have a theory that also minimizes the number of changes that occurred between chimps and humans. In particular, we would like to minimize the part that is special for language. There are other things we do that chimps do not—there is some debate, but the consensus seems to be that chimps do not have a theory of mind, they do not imitate and they do not point. Those are things that you need to acquire language, but they are not strictly linguistic.

Anything that serves the broad faculty of language is free and does not have to be explained as part of the theory of language. Whatever is left over is the part that is really special for language, which we might call the narrow faculty of language or Pinker’s language instinct or what linguists sometimes call Universal Grammar. This is what we would like to minimize by saying that there were minimal changes in the course of evolution. This evolutionary constraint gives minimalist inquiry its empirical bite: Linguistic theory should be more than just elegant, there is an empirical question of how much change in the genome is necessary from chimps to humans to get the ability to learn language.

The hypothesis of the Minimalist Program, which is the latest incarnation of Chomsky’s (1995) line of thinking, is that the narrow language faculty is in some sense perfect. It satisfies the Descriptive Constraint, it manages to map between sound and meaning, with an absolute minimum of machinery. The complexity of language arises only by virtue of interactions with independent properties of sound and meaning (that is, Broad Language Faculty). The idea is to get rid of as much of the richness of the Universal Grammar posited by Principles & Parameters, and still derive the same results. I think this ongoing program is not very successful, but some people think they are making progress.

For many people outside of linguistics, and certainly outside of Chomsky’s narrow circle, this latest version has seemed like a recantation of his basic principles because he has spoken for years about a rich innate language capacity, and all of a sudden he is presenting a very stripped down theory. Is this the right move? It is one possibility, but it is discrepant with the actual complexity of language.

There is another strategy for satisfying Descriptive Constraint besides positing that the child actually has less to learn. Find a way to formulate complexity in adult grammar so that more of it can be learned. Set it up so that acquisition is more graceful than it is in the standard versions of generative grammar. One way to do that is to minimize the elements of syntactic structure for which the child has no evidence. For years we have taken for granted that syntax is full of pronominal elements that have no pronunciation. For example in a sentence like, John tried to leave, how do you know who is the subject of the verb, to leave? You know it is John, but the way that is incorporated in the theory is to say that leave actually does have a subject, an invisible pronoun PRO, which then refers to John. Now, the child has to know that there is a PRO there, even though the child does not hear it. That ability requires more principles somewhere in Universal Grammar.
Another even grosser example is the standard stock and trade of Universal Grammar, the distinction between deep structure and surface structure. Surface structure is the order of words as you hear them, and deep structure is some other syntactic structure that is more canonical, more regular, closer to the meaning in various ways. Then, there is a sequence of operations on deep structure that moves things around and deletes and copies things and so forth, yielding the surface structure. To say that the child learns this implies that the child infers from the signal all of the hidden covert structure. If we can have a theory of linguistic structure that does away with that, then we can simplify the problem of acquisition by having things that are more learnable.

From the point of view of mainstream generative grammar of the 1960s, How is it possible to do without null elements & movement? That was the great advance of Chomsky's linguistic theory in its technical details. Over history, the main motivation for having these covert levels of structure is because we know that the surface structure of language does not conform to the meaning. The active and passive constructions have basically the same meaning with very different surface forms. The main advance in generative grammar was to account for that by saying that they are both derived from the same underlying form, and things moved around. So, there are all these mismatches between form and meaning, and those are now going to be encoded in the relation between this covert level of syntax and the surface structure. The covert syntax will be more-or-less homomorphic to the meaning, and that has been a presumption from the beginning. The strategy whenever you find some semantic distinction that is not expressed directly in surface form is to assume that the distinction is made in underlying syntax and things are moved around to get the surface form. This is a heuristic dating back to Katz and Postal in 1964.

An alternative is to encode these mismatches directly in the relation of meaning to surface form. Back in the 1960s, we did not have a theory of meaning that was stable enough to do that, so it was mostly a matter of speculation. It made more sense to do it in terms of syntax. Now, we have somewhat more robust theories of semantics and perhaps can bring it off. For example, instead of mapping the active and the passive from a single underlying syntactic structure, which corresponds to the meaning, you start with the meaning and say that there are two different ways of mapping it into syntax. The default way is an active, but then there is this other way that you can use if for some reason you want to put the patient in subject position, perhaps to topicalize it.

Is that the same thing as a movement transformation? It is, because you have to encode the same mismatches one way or the other, but you save a component in the grammar. A transformational grammar has a simple mapping between meaning and covert syntactic structure, and a complex mapping between covert syntax and surface form. In a direct mapping theory, you go right from meaning to a complex mapping to surface form. From the point of view of minimality, you should prefer the direct mapping theory, if you can bring it off. Can you bring it off?

There are theories that do it that way, the most prominent is Head-driven Phrase Structure Grammar. It does this technically, formally, and very precisely as a direct mapping theory. The problem with it is that the theoretical framework in which it is usually couched does not say much about the primary goal of explaining learnability. There is not much language acquisition research and no discussion of evolution.
Nevertheless, if one could adapt those techniques of description to a theory that takes the primary goal seriously, then we would be in business.

We could say that direct mapping is a priori superior, but is it empirically superior? As one example, we can look at the passive form. If you think of the passive as a syntactic movement that exchanges the subject and object positions, that commits you to say that you should always find a subject. For example, *The dog chased the cat* becomes *The cat was chased by the dog*. However, there are many passive verb phrase constructions where there is no evident subject:

1. Dick had John followed by the FBI.
2. The man followed by the FBI is my brother.
3. My brother heard insults shouted at him by the cops.
4. Followed day after day by the FBI, John went slowly nuts.

These are all cases where there is no evident source for the passive in a corresponding active sentence.

The movement theory can be salvaged by positing a null (or deleted NP) that has undergone movement, but that makes the theory less than minimal—adult is obliged to know more about syntax. The theory of the passive in terms of syntactic movement commits you to its being semantically blind. The whole point of syntactic transformations was that they are indifferent to semantics. Unfortunately, there are cases of passives where the lexical semantics plays a role, the less canonical prepositional passives:

5. The bed was slept in/on/*under/*beside by John.
6. The telescope was looked through/*inside by the technician.

The good constructions are about the proper function of the surface subject. There is some semantic dependence between the surface subject and the verb and preposition, which is very hard to characterize in any terms other than semantics. In the classic approach, the transformation can not know about that at all. So, a theory of transformational movement has no way to handle this problem. In a theory of direct mapping, you still have to say that there is a semantic dependence for this kind of passive, but at least you can account for it because you are mapping directly from the semantics, where that information is available.

Another problem with transformational movement, which was recognized very early in the learnability literature, is that learning structural descriptions of ordered transformations is one of the most severe obstacles to language acquisition (Wexler & Culicover, 1980). Learning sequences of operations that can be potentially unbounded presents severe difficulties. Wexler and Culicover (1980) showed that grammar learning could only occur by supplementing the theory of movement with very heavy restrictions on where things could move, which had to be innate. In order to have transformations, you have to have a repertoire of rich constraints on movement that somehow emerge from the genome.

Next, I want to look at the implementation: If there are these tree structures, how does the theory describe their construction? The heuristic from the evolutionary constraint is to have the most minimal mechanisms. Let us see how minimal the Minimalist Program actually is at ending up with complicated recursive structures. The simplest possible way is to take two constituents and stick them together into a
tree and give it the name of one of the elements. Then, keep adding more trees, an
operation called Merge: Take A and B and create either \([A, A B]\) or \([B, A B]\) or \([A, B A]\) or \([B, B A]\).

You can see how applying this over and over again results in big trees, but how do
you get this process started? You start with a numeration: Take a set of elements from
the lexicon, in a bag like Scrabble letters, then pull something out and pull something
else out and build those into a tree, and so on. The basic process is pulling items out of
the lexicon. What is in the lexicon? Minimally, the lexicon comprises words and/or
morphemes, coded nonredundantly. All redundancy in the lexicon is squeezed out into
rules. Chomsky quotes Bloomfield here: “The lexicon is really an appendix of the
grammar, a list of basic irregularities.” Bloomfield says exactly that, but he does not say
it is nonredundant, I think Chomsky reads that in.

That sounds very plausible and minimal, but let us see what that approach
presumes.

i. Organization of syntactic structure is to be characterized in terms of
putting pieces together one after another sequentially. (Derivational,
\textit{a.k.a.}, proof-theoretic)

ii. Binary branching is the optimal and minimal kind to have in syntactic
structure.

iii. The lexicon is nonredundant.

iv. There is a strict division between the lexicon and the grammar: They are
totally different beasts, the grammar is all the regularities and the
lexicon is all the irregularities.

v. Semantics is strictly locally compositional (Fregean): The meaning of a
sentence is constructed word by word, combined according to
structural branching.

There are some alternatives to these assumptions that have become prevalent in
other frameworks. The rules do not construct trees, they license trees. You can check
whether each piece of structure and each relation among pieces of structure online
is licensed by a relevant principle. This allows the possibility that pieces of structure
could be licensed by multiple constraints at the same time. The constraints could
conflict with ways to resolve conflicts. These are called constraint-based, or
representational, or model-theoretic grammars. Is this different from derivation?
Instead of building the structure, you just check the structure at every point.
Chomsky usually says it is a notational variant, and that derivational is right. To a first
approximation, they look like notational variants, but I want to show you that they are
really different.

Here is a simple case of the difference between the two from Paul Postal’s (2004)
book, \textit{Skeptical Linguistic Essays}. He points out that there are sentences that contain
pieces of non-English:

7. The space alien said ‘klaatu barrada nikto’ to Gort.
8. [teenspeak:] And then, I am all like, [gesture of exasperation].
9. The sign @ was invented in 1451.
10. \textit{Skelerf} does not rhyme with \textit{nikto}.
11. \textit{Jean est mangé le pain} is ungrammatical in French.

How do you turn these into sentences? These cannot be described using a
derivation starting from enumeration. One possibility is that these are not English
sentences, which is unlikely. Another possibility is that they are all in the lexicon, too,
but the lexicon is supposed to be knowledge of English. The idea of building structure from enumeration of the lexicon gets into trouble. In a constraint-based model, there can be particular contexts that do not constrain constituents to items of English. It is a weaker theory, but it allows you to describe sentences such as these in a way that you can not in a derivational theory.

Instead of the notion of Merge, the fundamental combinatorial device for these kinds of grammars is unification (sort of like Boolean union on feature structures). If you take two things that share some features, when you unify them, you get something that coincides where they are the same and still has the differences sticking out: Unification of \([V, +past]\) and \([V, +3\; sing]\) = \([V, +past, 3\; sing]\); Unification of \([VP\; V\; NP]\) and \([V, +past]\) = \([VP\; V, +past\; NP]\). Unification is different from Merge, but you can state Merge as a special case of unify: Unification of \(A\) and \([x, y]\) = \([A, y]\); Unification of \(B\) and \([A, y]\) = \([A, B]\). Unify cannot be reduced to a special case of Merge. So Merge is not the conceptually simplest combinatorial operation, as claimed.

What about binary branching? If you have a ten-word sentence, and you are restricted to binary branching, then you will have a pretty high tree with many nodes in it. On the other hand, if you allow ten-way branching, you can have a tree with just one node and ten branches. Multiple branching trees require fewer nodes than binary branching trees. Which is minimal, fewer nodes, or fewer branches per node? You do not know in advance. Multiple branching recursion is present elsewhere in cognition, so it arguably comes for free. Here is an example from visual grouping:

```
xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx
 ooooo  xxxxx  ooooo  xxxxx  ooooo  xxxxx  ooooo  xxxxx  ooooo
 xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx
 xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx
 xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx
 ooooo  xxxxx  ooooo  xxxxx  ooooo  xxxxx  ooooo  xxxxx  ooooo
 xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx
 xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx
 ooooo  xxxxx  ooooo  xxxxx  ooooo  xxxxx  ooooo  xxxxx  ooooo
 xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx  xxxxx  ooooo  xxxxx
```

This is recursive in the standard sense (and could be further embedded in an array of arrays). But there is no justification for binary branching here. The pattern is not composed of two \(x\)s and then another \(x\) and another \(x\) and another \(x\) to make a group of five \(x\)s. There is no empirical reason to suppose binary branching here, although, there is an empirical reason to suppose that there is three-way and five-way branching. What this means is that cognition gives you multiple branching. Saying that binary branching is more economical is denying yourself something you ought to be able to get for free. The alleged simplicity of binary branching is spurious.

There are some claims in the literature that binary branching simplifies binding theory, the way pronouns find their antecedents. It turns out that it simplifies binding theory by not using linear order as a principle for determining the referents of pronouns. However, linear order comes for free, so why throw that out?
There is another claim that binary branching simplifies acquisition. If you hear a
string of three things, and you know that the language has to be left-ward binary
branching, you get a unique structure for it. That is fine, but if you have a string of
three things, and the assumption is that unless there is evidence to the contrary, it is
just a three-way branch, that is just as easy for acquisition. As long as you choose one
as a default, there is no reason binary branching is any cheaper than any other kind.
There is no fair argument for uniform binary branching as the minimal operation in
terms of either the Descriptive or the Evolutionary Constraints.

Although Chomsky often asserts that the lexicon is nonredundant, no one (to my
knowledge) has ever formulated how all redundancy is squeezed out of the lexicon into
rules. The best guess is that redundancy is characteristic of brain processes, it helps
stabilize them and make them more reliable. In terms of simplicity, the general
environment in which language sits favors redundancy.

To return to the main question: What does the child have to learn in addition to the
grammar? The child has to acquire lots and lots of words. In addition, you need a
learning procedure that acquires them, whether general or special or some combination. It is a very difficult task, and it is certainly not done by setting a bunch of
parameters.

What else do you have to learn? Thousands of clichés, titles, etc. A few years ago,
my daughter was devoted to Wheel of Fortune, and just for fun, I asked her to write
down all the answers. After about six months, she had collected about 600 of them.
They are all clichés and idioms, and titles of songs and names. These are all part of an
American English speaker’s knowledge. Otherwise, they would be useless as puzzles on
this show. Wheel of Fortune shows no sign of running out of them after 25 years. What
is interesting about idioms is that their meanings are not compositional, some of them
are discontinuous, and they are redundant in the sense that they use known words to
create a nonredundant meaning. This is a problem for using syntax to arrive at these
idiosyncratic meanings, if words get inserted only at the bottom of the tree (e.g., kick
the bucket means dying). But, idioms fall out very nicely in a constraint-based theory.

There are also noncanonical utterance types to be learned, which are not predicted
by X-bar theory. These could be prepositional phrase with noun phrase constructions,
like Off with his head! or Into the trunk with you! Here are some more examples:

12. How about X? (How about a cup of coffee? How about we have a little
talk?)
13. NP+acc Pred? (What, me worry? Him in an accident? John drunk?)
14. NP and S (One more beer and I am leaving. One more step and I shoot.
15. Scores (The Red Sox 4, the Yankees 3)
16. The more ... the more (The more I read, the less I understand).

These have to be stored as exceptional pieces of syntax, complete with some sort of
special interpretation. It was these kinds of things that led to the theory of
Construction Grammar. Syntax alone can have meaning attached to it, without
particular words being involved. The point is that you can not derive these pieces of
syntax by deletion and movement in any account of standard phrase structure.

Finally there are other noncanonical pieces of syntax. Here are some examples for
names of geographical features: the Atlantic Ocean, the Hudson River, the
Mediterranean Sea; the Bay of Biscay, the Gulf of Aqaba, the Sea of Azov; Arrowhead
Lake, Wissahickon Creek, Laurel Hill, Loon Mountain; Lake Michigan, Mount Washington. These are idiosyncratic productive systems, which follow rules for feature and name positions. The grammar of numbers is another idiosyncratic element: three hundred fifty-five billion, fourteen million, one hundred twenty-five thousand, six hundred thirteen. Another system is focus reduplication:

17. You make the tuna salad, and I’ll make the SALAD-salad.
18. Would you like some wine? Would you like a DRINK-drink?
19. Do you LIKE-her like-her?
20. Are you guys, um, LIVING-together living-together?

These indicate a generic member of a category or a special marked member. Unlike most standard phonological reduplication rules, this one can copy phrases as well as words. Another one is the N-P-N construction: dollar for dollar, face to face, house by house, month after month, volume (up)on volume of phonology texts. These are productive in some respects, but they are riddled with special features. How is meaning related to the meaning of the preposition, if at all? All these little patterns have to be learned. They are rules. But they do not follow from any standard notion of Universal Grammar. Presumably every language has lots of these sorts of things.

There is a possible objection from the minimalists to consider, that these are merely peripheral aspects of grammar. The problem of language acquisition and the goal of perfection apply only to core grammar (that is, argument structure, passive, raising, long-distance dependencies, basic cases of binding). So the sorts of phenomena I have been describing are irrelevant to these accounts. Such an approach explicitly abandons the Descriptive Constraint: The theory is no longer responsible for the structure of the adult grammar (or, postpones it indefinitely). Moreover, if you have a learning procedure that can acquire words and all these peripheral grammatical phenomena, can the same learning procedure not acquire core grammar as well? Without an account of the learning of the periphery, you can not tell.

The research strategy of Minimalist Program is to idealize away from periphery, not to mention its acquisition, so it will never investigate this question. However, similar peculiarities are found in indisputably core areas of grammar. For some verb phrase constructions, where the verb normally determines all the arguments, in English these are often infested with parasites. If you look at something like, He sang/drank/slept/laughed HIS HEAD OFF, you can insert many verbs, but they can not be transitive. You can not say, He drank scotch his head off. Why? Because his head has taken over the object position. The idiom, his head off, takes over the head and particle positions and you can put a verb there, but there is no room for anything else, and it means something like excessively. There are many of these examples in which the verb phrase is parasitized by peculiar constructions to demote the verb to a kind of modifier.

One of the big victories of core grammar is to unify long-distance dependencies like relative clauses and topicalization, etc., by saying there is only one principle, move wb/alpha, that moves to the front and satisfies the same constraints every time. Early on, we were originally concerned about getting the front of these long-distance dependencies right. So that the relative clause had the right stuff at the front, and it was really hard to move something and then make sure it came out the right way. The ones that were particularly vexing were the infinitival relative clauses, like the man to
whom to speak and not the man who to speak to, or the man for you to hire and not the man with whom for you to talk. The interesting thing is that nobody looks at these any more. Once the movement was unified, everyone stopped worrying about them.

These particulars are not predictable from a general rule that says to move things to the front. They have to be learned. It is not easy to write rules that come up with these configurations after fronting. The attempt was abandoned with the onset of Principles and Parameters (i.e. in disregard of Descriptive Constraint). In a constraint-based (non-movement) theory, these can be learned as idiosyncratic configurations associated with surface forms – i.e. syntactic idioms with particular constructional meanings. Generalizations about long-distance dependencies are not a consequence of movement, but a consequence of relating the signature to a gap within the clause.

There is an unbreakable continuity between core and peripheral phenomena, and between core generalizations and complete lexical idiosyncrasy. There is probably a multi-dimensional space from the most general rules of verb phrases all the way to individual verbs. A theory that posits a principled difference between them is missing a deep and important fact about language (not to mention abandoning the Descriptive Constraint). A derivational movement-based theory does not lend itself to expressing this insight. A constraint-based theory does. Therefore, derivational and constraint-based theories are not notational variants, and constraint-based theories are more adequate for expressing insights about the texture of linguistic structure. Virtually all the basic properties of implementation of phrase structure in the Minimalist Program are either formally non-minimal, empirically inadequate, or methodologically unsound.

With respect to learning and innateness, the Minimalists could object that a constraint-based theory requires a proliferation of rules in order to meet the Descriptive Constraint. How does this approach address acquisition, so as to comply with Evolutionary Constraint (that is, to reduce the volume of innate components of a Narrow Language Faculty)? Let us look at the difference between a word and a rule: Both are pieces of structure stored in memory, but a rule has variables as part of its structure, which must be satisfied by unification with something else. To examine this smooth transition from idiosyncrasy to maximal generality, we start out with a fully specified piece of structure, like kick the bucket, we also have idioms with variables, like take ___ to task, and we can go on up the line to more general frames with more variables, until we see things that start to look like the head parameter or X-bar theory. The core principles of phrase structure are general schemata; idiosyncratic rules and fully specified items are specializations. There can also be idiosyncratic rules that are not specializations of more general schemata (e.g. N-P-N).

What does this formulation of rules say about learning? This presents a possibility of a learning procedure that people as different as Tomasello (2003), Culicover, and Nowak (2003) and also Martin Braine in the 1970s have proposed. A child learns particular constructions holistically at first. When multiple items share a part, create a new item (that is, a rule) that consists of the constant part plus a variable corresponding to parts that differ from item to item. More and more general schemata arise by recursive application of this process. This is much easier in a theory without movement.

What is the role of Universal Grammar in this: How is this different from plain analogical learning? It is different from analogical learning in that there is an
extraction of variables. That is really important, it is the thing that has yet to be added
to connectionist learning. The way I see Universal Grammar playing a role is as sort of
attractors for generalizations—you are aiming in a particular direction with your
generalizations, if you can get there.

What kinds of generalizations would the child be looking for?

Some aspects of Universal Grammar:
I. Basic organization of conceptual structure, growing directly out of
primate cognition (hence part of Broad Faculty of Language).
II. The notion of words being used symbolically to communicate
intentionally about perceived world – the evolutionary breakthrough
(Deacon 1997). The rest is refinement.
III. Use of Unification plus variables in stored structures to permit
productivity and recursion.
IV. Basic principles of phrase structure:
   A. X-bar theory.
   B. Other common alternatives such as conjunction schema.
V. Basic default principles of syntax-semantics interface:
   A. Semantic heads map to syntactic heads, semantic arguments to
      syntactic arguments, semantic modifiers to syntactic adjuncts.
   B. Agent First order preference.
   C. Topic First, Focus Last.
VI. Basic principles of morphological agreement and case-marking.
VII. Basic principles of long-distance dependencies.

Not to mention Universal Grammar aspects of phonology and morphology.

This is not a perfect system by any means, but it appears relatively minimal, given
the need to satisfy the Descriptive Constraint. Unlike the Minimalist Program, this
conception of grammar allows for proliferation of learned rules, under a potentially
realistic learning regimen. Learning rules is mostly an extension of learning words. The
narrow language faculty is sort of a toolkit that results in tendencies toward language
universals.

If any approach to language is eventually going to satisfy the primary goal of
linguistic theory, to satisfy the Description and Evolutionary Constraints, and to make
meaningful contact with cognitive neuroscience and evolutionary biology and
psychology, it will be an approach growing out of constraint- and construction-based
minimalism, not out of the Minimalist Program.

**APPLAUSE**

**Questions**

**Professor Gary Marcus:** It seems like the point you were making with the Martian
language example is equally a problem for any theory; what you need is a use-mention
distinction. I do not understand why this distinction is more of a problem for the
Minimalist Program than it is for any other account of grammar. These are mentioned
instances of language, rather than using the language as such.

**Professor Jackendoff:** Unless you have something in the lexicon that says mention,
and there is an invisible item in the syntax that allows you to put anything in it, this
suggestion will not work. In some cases, for direct quotes in your own language, those
get semantic interpretations, and the truth conditions matter.
**Professor Boris Gasparov:** Your discussion of the proliferation of clichés and the impossibility of separating grammar from the lexicon strikes me as being rather close to the ideas of Charles Fillmore.

**Professor Jackendoff:** Absolutely, he is one of the originators of construction grammar, and that is one of the tenets, that words are one kind of construction and there are other chunks that come in all sizes.

**Steven Frisson:** When you are attempting to build more abstracted rules, and you want to distinguish things that are idioms, which can not accept variables, from things that can, this is going to be difficult. Could it come out of the semantic-syntactic interface?

**Professor Jackendoff:** It seems like a lot of these sort of grow out of one canonical case that is of much higher frequency than any other instance of the construction. There has to be some sort of frequency and variety sensitivity. In general, I do not know the answer to that.

**Professor Gary Marcus:** I am trying to figure out why I am not quite satisfied with the argument, even though I agree with so much of what you are saying. Learners could easily be seduced by certain kinds of distributional generalizations. The field of acquisition has been down the distributional path before. I do not think they should search the entire search-space, I think there are critical parts of the space they did not search. Like, what if you did not just learn the strings of syntax, but you learned them with the semantics. There is a way in which what you are proposing is at least reminiscent of these old distributional approaches. The difference between you and Tomasello, and I think you are on the right side, is to say that Universal Grammar is integrated in this; you are not just making any generalization. What is somehow funny about the structure of your talk is that it makes arguments for the construction side, but it does not really give the classic examples of why that is not enough by itself. One needs more of an integration of Universal Grammar to figure out why it is that this approach not going to fall into the same trap.

**Professor Jackendoff:** I think that is exactly right, which is why I am saying that if this works out, it is at least believable. One of the early points of Wexler and Culicover is that you can not get to first base without having a meaning to correlate with it. That presumes that the system of meaning is in place for the child, and that Universal Grammar is at least telling you a default way to say something. At this point, I do not think we have any hope of a learning theory for derivational grammar, but it remains to be seen whether we can work it out in this case.

**Professor Robert Remez:** I think this is possibly the same question that has been proposed a few times in slightly different form, but it is a psychologist’s question again about learning. There are two steps in your proposal that have elements that are completely incommensurate. In one, the infant is described as learning something holistically, and in the next, the infant is said to make various pairings over sets of elements, and the elements can not be given in a holistic description. So, where do the elements derive from?

**Professor Jackendoff:** I think the set of primitives have to be given, you have to know that there are syntactic categories. I think the notion of a syntactic category is not discoverable, the notion of a phonological decomposition of a syllable is not
discoverable. Maybe the notion of a syllable is discoverable. I think the basic bedrock has to be given.

Professor Robert Remez: So, if you tried to describe the initial state, it would be a capacity to elaborate a sample according to a set of elements that are given a priori.

Professor Jackendoff: Yes, and then the question is how to do the segmentation and classification.

Professor Robert Remez: But, you can do it every which way—it is inherently multi-stable.

Professor Jackendoff: Yes, it is going include statistical properties of the environment plus the attraction of Universal Grammar plus whatever you can bring to bear from the meaning. So, you are coming at it from a number of different angles. The hope is that maybe this might work. I think there are serious questions about which things sort of go regular because you are just hearing samples of both. How does the learning extract that regularity from a few examples? I think that is a fundamental problem.

Mr. Martin Jansche: Everything that you have sketched out is in HPSG and some other formulations, and there is some acquisition literature within HPSG, by Georgia Green and some of her students. I am hoping that your approach will garner some respect for these endeavors.

Professor Jackendoff: One of the things that has bemused me for some time is that there are all these little schools of grammar who define themselves in opposition to Chomsky, and do not spend much time talking among themselves about what they have in common. I actually tried to get a number of these people together, and they did not know how to talk to each other. What I am trying to do through my work in part is to say there are elements that all these things have in common; let us sort through the additional presumptions in each framework and see what is right and not right. I have some problems with HPSG’s emphasis on the sign, in that I think it is more heterogeneous—not everything is a matching of syntax, meaning, and phonology. Also, there are problems with the emphasis on heads. Some people in HPSG have been moving in the direction of construction grammar. This dialogue is a move in the right direction.

Professor Robert Remez: My corner of the field is bedeviled by this notion of parsimony, or what really, in fact, is prescriptive parsimony. The way we actually learned about this in the olden days is that a pre-requisite for application of Occam’s Razor is functional equivalence. Unless you have two functionally equivalent accounts, an assessment of parsimony is premature. You can not aim to be more parsimonious, you can only make a more parsimonious choice. I would like all claims about parsimony to be postponed until functional equivalence is established. This is a cranky comment.

Professor Jackendoff: I see what you mean. I think in the book with Peter, we take the position that a lot of times people invoke Occam’s Razor and they are not looking at the true price. They are invoking it locally. We can take this condition out of binding, but what you do not see is that you need that thing anyway for seven other things, or you have to add more things elsewhere.

Professor Gary Marcus: One of the puzzles that you raise is how you can have a lot of things that seem law-like and seem to follow Universal Grammar, and then you have
all these idiosyncrasies, and there is almost a continuum between them. I wonder whether this notion of analogy constrained by Universal Grammar escapes from that problem. Do you suffer the same problem if you bring Universal Grammar back into the analogical system. It is nice that you can now represent the constructions given the notation, but there is still this continuum, and you still have to wrestle over things like the gradation from law-like to idiosyncratic.

Professor Jackendoff: Suppose the input does not let you make a generalization in that direction, so you get stuck with something like, *how about some lunch*, which does not follow from anything particular in Universal Grammar. So the generalization goes as far as it can and it gets stuck in an ecological *cul-de-sac*. If there is some way that it can join the main stream, and get up to X-bar theory, it will do so, but if it can not, well, Universal Grammar does not have to tell you what the class of possible grammars is. It is more like the theory of markedness: grammars are going to be easier to learn or more robust if they have certain properties.

Professor Bill Benzon: I want to get back to semantics—how do you incorporate the notion that we inherit so many cognitive abilities with meaning in the real world? I notice that you have a forthcoming paper with Steve Pinker, and I wonder how you reconcile the evolutionary framework. Some propose that languages evolved to be different so that group members could identify each other.

Professor Jackendoff: I have been pushing the view for 20-some years that the semantic system is not restricted to language. If language is to be used for anything useful, it has to connect with the way you understand the world. At what point does language stop and this other stuff begin? There have been these huge arguments back and forth in semantics as to whether there is a specifically linguistic semantics and then world knowledge, or whether to go straight to world knowledge. I tried to argue that it goes straight to world knowledge, perhaps with a particular slant put on it by the fact that it is expressed in language. Not every attribute of language can be cast as an adaptation. Evolution can not code a whole language on the genome. It got as far as it needed to so that we could learn the rest. The fact that we use language for identification is just a side-effect of the fact that we use anything we can for identification.

Pinker and I wrote that article in response to an article in *Science* by Hauser, Chomsky, and Fitch. They claimed that all you need to add to the chimp cognitive repertoire might be just recursion. What happened to words, phonology, morphology, grammatical functions and stuff like that? They posed the right question—what do you need to add to the chimp repertoire—but I think it is the wrong answer. In fact, I think you get recursion for free. It is the words and the phonology and so on that you need.

Professor Remez: Let us thank Professor Jackendoff and adjourn.

APPLAUSE

Place: Kellogg Center, Room 1512
School of International and Public Affairs
420 West 118th Street

**Time:** 4:00 PM

**Chair:** Prof. Robert E. Remez, Barnard College, Columbia University

**Rapporteur:** Jennifer Pardo


**Questions pertaining to this transcript should be sent to the rapporteur via email:**

Jennifer Pardo
jsp2003@columbia.edu