Theorizing about the Syntax of Human Language: A Radical Alternative to Generative Formalisms

Geoffrey K. Pullum

School of Philosophy, Psychology and Language SciencesUniversity of Edinburgh

ORCID ID: 0000-0002-7748-8847

Abstract (English)
Linguists standardly assume that a grammar is a formal system that 'generates' a set of derivations. But this is not the only way to formalize grammars. I sketch a different basis for syntactic theory: model-theoretic syntax (MTS). It defines grammars as finite sets of statements that are true (or false) in certain kinds of structure (finite labeled graphs such as trees). Such statements provide a direct description of syntactic structure. Generative grammars do not do this; they are strikingly ill-suited to accounting for certain familiar properties of human languages, like the fact that ungrammaticality is a matter of degree. Many aspects of linguistic phenomena look radically different when viewed in MTS terms. I pay special attention to the fact that sentences containing invented nonsense words (items not in the lexicon) are nonetheless perceived as sentences. I also argue that the MTS view dissolves the overblown controversy about whether the set of sentences in a human language is always infinite: many languages (both Brazilian indigenous languages and others) appear not to employ arbitrarily iterative devices for embedding or coordination, but under an MTS description this does not define them as radically distinct in typological terms.

Abstract (Portuguese)
Os lingüistas em geral presupõem que uma gramática seja um sistema formal que 'gera' um conjunto de derivações. Aqui delineo uma base diferente para a teoria da sintaxe: a sintaxe modelo-teórico (SMT). Ela define as gramáticas como conjuntos de declarações que são verdadeiras (ou falsas) em determinados tipos de estruturas (gráficas finitas rotuladas, como árvores). Tais declarações fornecem uma descrição direta da estrutura sintáctica. As gramáticas gerativas não fazem isso; elas são supreendentmente mal-adaptadas para dar conta de certas propriedades familiares das línguas humanas, como o fato de que a agramaticalidade e uma questão de grau. Muitos aspectos de fenômenos lingüísticos parecem radicalmente diferentes quando vistos nos termos STM. Eu presto atenção especial aqui ao fato de que que orações que contem palavras inventadas sem sentido lexical (não se encontram no léxico) são mesmo assim percebidos como orações. Eu também argumento que a STM acaba com a controvérsia exagerada sobre se o conjunto de orações de uma língua humana é sempre infinito: muitas línguas (seja uma língua indígena brasileira ou outra língua qualquer) parecem não empregar dispositivos arbitrários e iterativos para o encaixamento ou a coordinação, mas sob uma descrição STM isso não as definem como radicalmente distintas em termos tipológicas.

INTRODUCTION

This paper addresses an issue concerning how syntactic theories should be formalized. It does not present or defend any specific syntactic theory here; it operates at a higher level of abstraction. The aim is rather to clarify certain differences between ways of formalizing the syntactic theories that linguists develop.

There is much room for confusion and ambiguity when dealing with terms like 'formal', 'formalist', and 'formalization'.[1] I am not concerned here with the loose collection of ontological views that is known as 'formalism' in the philosophy of mathematics. All I mean by formalization is the use of suitable tools from mathematics and logic to make the claims of a theory more explicit.

---

[1] See the papers in McElvenny (2019) for discussion from various perspectives.

Some writers on grammatical theory seem to think that anything described by the word 'formal' must involve insensitivity to the role of meaning or use. Michael Tomasello is an example:

> [G]enerative grammar is a "formal" theory, meaning that it is based on the supposition that natural languages are like formal languages … abstract algebraic rules that are both meaningless themselves and insensitive to the meanings of the elements they algorithmically combine. [Tomasello, 2003:5]

But making use of tools originally devised for characterizing formal languages does not imply making 'the supposition that natural languages are like formal languages.' In some respects, they are alike; but I will argue that those are not the most important features of human language. And it should be obvious that formalization can aid the development of either grammar or meaning: the study of semantics has of course been revolutionized since 1970 by the introduction of techniques from logic.

What I draw attention to here is that there are radically different ways in which formalization of syntax could be carried out. One seldom-adopted idea is that grammars for human languages could be formalized as **transducers**: systems of operations which convert meaning representations into pronunciation representations (or conversely, convert pronunciation representations into meaning representations), thus characterizing what could in principle be a way for thoughts to be expressed by a speaker (or for apprehended utterances to be understood by a hearer). Manaster-Ramer (1993) is a rare instance of a paper advocating transducers as grammars.

Zellig Harris adopted a very different view. His work on syntactic structure was aimed at defining a collection of symmetric relations relating sentences to each other in certain ways. He called those relations 'transformations' — a term his student Noam Chomsky took up later with a different sense.

But the view that has dominated the last six decades of syntactic theory, especially in the USA, is distinct from both of the foregoing approaches. It focuses on set-defining systems known as generative grammars. I begin with some observations about them, and then turn to a fourth alternative.
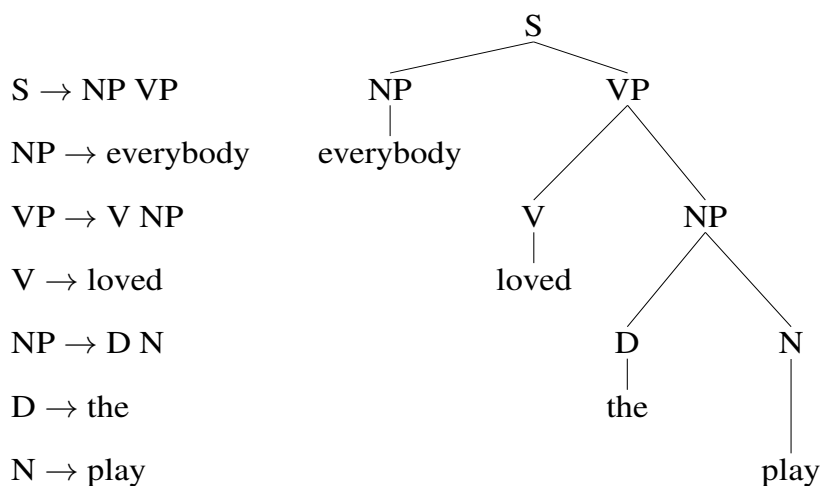
GENERATIVE GRAMMARS

I use the term 'generative grammar' here in the narrow sense developed by Chomsky in the middle 1950s. In this narrow sense, generative grammars are nondeterministic random construction procedures interpreted as definitions of sets of strings or other algebraic objects such as trees. A generative grammar consists of (i) a set of one or more relevant objects that is given initially, and (ii) a set of operations for constructing new objects from those that are already available. A system of this sort defines a certain set of objects if and only if

- **every** object in the set could in principle be constructed from the initially given strings using the operations of the grammar, and
- **only** objects in the set can be constructed from the initially given strings using the operations of the grammar.

There are two strikingly different varieties of generative grammar: **expansion-oriented** and **composition-oriented**. In expansion-oriented grammars, derivations begin with a single symbol, standardly 'S', and the operations expand this to produce longer symbol sequences: S → NP VP, and so on. The grammar generates all and only those sequences of words that can be reached from S. From 1956 to about 1990, the generative theories under discussion by those who followed Chomsky were always expansion-oriented.
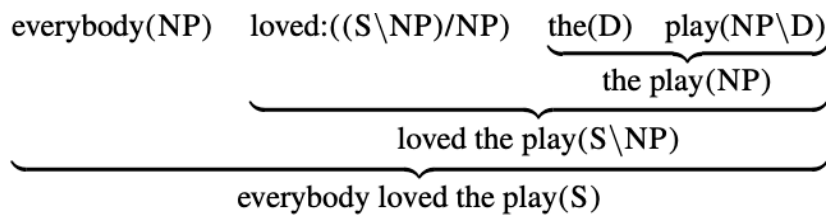
Elsewhere there was a minority community of grammarians, especially those interested in semantics, who were working with composition-oriented generative grammars, specifically **categorial** grammars, in which derivation begins with a collection of categorized words that are combined with each other to former larger units. Yehoshua Bar-Hillel, a friend of Chomsky's in the 1950s, published an expository paper on categorial grammar in *Language* in Bar-Hillel (1953); Montague assumed categorial grammars in his hugely influential work on semantics for natural language (see the papers collected in Montague 1974); and Ades & Steedman (1982) employed them in a paper that founded the sophisticated and influential program now known as combinatory categorial grammar (henceforth CCG; see Steedman 2000 for an exposition). A categorial grammar generates all and only the constructible sequences of words that end up being assigned to the category S (or whatever is used as the category of sentences).

Simple context-free phrase structure grammars (CFGs), which are expansion-oriented, and simple categorial theories (CGs), which are composition-oriented, were proved weakly equivalent (in the sense that each can describe exactly the same sets of strings as the other) sixty years ago, by Bar-Hillel, Gaifman, & Shamir (1960). In a CFG, the root category (here S) is expanded into longer and longer strings of categories like NP and VP until a string of words results. Miller (1962) illustrates with the sentence *Everybody loved the play*. It can be represented by a tree diagram, annotated by writing the relevant CFG operation on the left at the point where it is relevant:

```
                                    S
                          ┌─────────┴─────────┐
S → NP VP                 NP                   VP
                          │            ┌───────┴───────┐
NP → everybody         everybody       V              NP
                                       │          ┌────┴────┐
VP → V NP                                          V        NP
                                                   
V → loved                            loved

NP → D N                                       D         N

D → the                                       the

N → play                                               play
```

Categorial grammars have a different-looking theory of categories. A category α/β combines with a following β to form a constituent of the category α, and a category α\β combines with a preceding β to form a constituent of the category α. Thus a transitive verb like *eat* might be categorized as '(S\NP)/NP', meaning 'element that if combined with a following NP yields a

constituent which, if combined with a preceding NP, yields an S'. A categorial derivation of the same sentence analysed above — starting with the four words *everybody*(NP), *loved*(V), *play*(NP\D), and *the*(D) — can be displayed in this way:

$$\text{everybody(NP)} \quad \text{loved:}((S\backslash NP)/NP) \quad \underbrace{\text{the(D)} \quad \text{play(NP\backslash D)}}_{\text{the play(NP)}}$$

$$\underbrace{\qquad\qquad\qquad \text{loved the play}(S\backslash NP) \qquad\qquad\qquad}$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad \text{everybody loved the play}(S) \qquad\qquad\qquad\qquad\qquad}$$

CFGs and simple categorial grammars are weakly equivalent: any set of strings that one can define can also be defined by the other (see Steedman 2000 for detailed discussion of a more sophisticated categorial grammar that has greater expressive power than CFGs). What is important here is that in both cases a grammar provides the means for constructing certain strings or trees but not others, and a grammar is interpreted as defining a specific set: the set of all objects that it would be possible in some way to construct by applying the construction operations of the grammar to whatever is supplied initially.

Before 1990, through all the different versions of transformational grammar, Chomsky always assumed expansion-oriented grammars, but in 1990 he made an unheralded and unexplained switch to composition-oriented grammars as part of his 'minimalist program'. As set out in Chomsky (1993), it took a derivation to begin with what Chomsky called a 'numeration' — formally, a multiset of categorized words.[2] The words are progressively combined by an operation that he calls 'Merge'— it is a rather vague analog of what categorial grammarians call 'application'.

For the purposes of this paper, the differences between CFG, transformational grammar, CCG, and minimalism make remarkably little difference; the issues I am dealing with are at a higher level of abstractness than what separates these varieties of generative grammar.

NATURAL CONSEQUENCES OF GENERATIVE GRAMMAR

I now summarize six inevitable consequences of using generative grammars, of any variety, for describing languages. These are not minor features that could be altered by slight revisions; they are unavoidable side effects of the deepest aspects of the generative type of formalism itself. They should not be controversial: a linguist must either accept them as inevitable consequences or to some extent abandon generative grammars as originally defined. I am not saying they add up to an argument that generative grammars are bad or useless. They are what they are, and they are well suited to some tasks, particularly defining sets of strings (or other algebraically definable objects) within mathematics, logic, and computer science. They are also well integrated with the tools and methods of computational linguistics, which I take to be the most important area in which linguistics finds practical applications (I return later to that topic). My focus in this paper is entirely on characterizig the phenomena of syntax in human languages. This has been the primary application for

---

[2] A multiset is like a set, but with repetitions allowed. It can be treated mathematically as a list of members each of which is paired with a positive integer denoting its number of occurrences.

generative grammars within linguistics since 1957, and I will suggest that there are grounds for a reconsideration.

Generative holism

Generative grammars exhibit an interesting analog of what philosophers of science call 'confirmation holism' (Quine 1951). Just as no datum can confirm or refute any specific prediction or hypothesis of a scientific theory (because other parts of the theory can always be altered to guarantee or counter any specific prediction), likewise no fact about any well-formed expression ever confirms (or refutes) any individual operation that forms part of a generative grammar. A generative grammar's characterization of a language is holistic, in the sense that no individual operation says anything about any expression. It is the whole grammar, taken all together, that provides a simultaneous definition of the entire language. Few linguists seem to have fully appreciated this. It is easy to be tempted to think that the phrase structure operation 'PP → P NP' entails that the generated language has preposition phrases, and that preposition phrases have prepositions as immediate subconstituents, and that prepositions precede their NP complements. It guarantees none of these things.[3] Grammars containing this operation can easily be constructed for languages in which sentences never have PP constituents (because they are always deleted), or PPs never have a P subconstituent (because P is always rewritten as V), or only postpositional phrases occur (because the order is obligatorily reversed). Everything depends on what the rest of the grammar says.

Sharp boundaries

Generative grammars define sharply delineated sets of expressions, in the sense that for any given object (string, tree, or whatever), either it is in the set or is not. There is no gray area: if an object belongs to the generated set it is grammatically perfect, and if not, then its failure to belong is total: the grammar assigns it no grammatical properties.

Fixed cardinalities

The set defined by a generative grammar either has some definite finite size (so there is a maximum sentence length, easily computable from the grammar) or is countably infinite (so its sentences have no upper bound on length and there are as many sentences as there are positive integers).

Holistic acquisition

Acquisition of a language in generative terms has to involve arriving at a complete generative grammar. It cannot involve learning pieces of the grammar, bit by bit, because of generative holism. Gold (1967) did the first mathematical work on language acquisition in generative terms, and formalized learning in terms of instantaneous guessing of complete grammars. In fact the internal workings of grammars play no role at all in Gold's paper: a numbered list of all possible grammars is assumed, and conjecturing a grammar amounts to simply announcing its number. Successfully learning on the basis of an information source is

---

[3] Notice that I do not call it a 'rule'. Rules (like 'Keep to the left') can be followed, or violated. Phrase structure rewriting operations like 'PP → P NP' or 'Merge(A,B)' cannot: they are just steps in a construction process, and do not correspond to anything like a regularity of behavior.

defined as using the information to identify a fully correct generative grammar generating it, and announcing that grammar's number. Gold's proofs depend crucially on the 'sharp boundaries' and 'defined cardinalities' properties. Perhaps his most problematic result is that if you cannot be sure of whether there are infinitely many sentences in the target language or only a finite number, positive information about sentences can never settle the question of whether your current guess at the grammar defines too large a language.

Impossibility of quandaries

Because of the sharp boundaries property it is impossible for there to be grammatical quandaries — cases where the grammar does not have definite implications one way or the other so well-formedness is indeterminate. An expression is either generated (in which case it is linguistically perfect) or not generated (in which case it is not a linguistic entity at all, and the grammar says nothing about it), but it cannot be defined by the grammar as having a dubious or on-the-cusp status. (I will give examples later.)

Lexical dependence

It is part of the definition of a generative grammar that it incorporates a finite lexicon --- a listing of its words and their grammatical (and semantic and phonological) properties. The well-formed expressions defined by a grammar contain only actual words that are in that grammar's lexicon; what the grammar says about grammaticality depends entirely on what lexical items there currently happen to be.

MODEL-THEORETIC SYNTAX

I now want to sketch the radical alternative to generative syntax that is my main topic. Generative grammars are rooted in mathematical machinery originally developed in order to formalize syntactic proofs in logic. The alternative I am suggesting stems from the semantic side of logic: model theory. I will refer to the approach as **model-theoretic syntax** (henceforth MTS). The central idea is that we make the realist assumption that sentences exist, and actually have structure (we do not need to devise a mathematical system to assign it to them), and the purpose of a grammar is to state what that structure is like by giving a set of constraints satisfied by the right structures. The fundamental principles are these:

- A grammar consists of **constraints**, each making a statement that is true or false of any given individual expression.

- A grammar is an **unordered finite set** of such constraints.

- Well-formedness of a structure is determined solely by **satisfaction** of all of the constraints of the grammar, and ill-formedness involves failure to satisfy at least some of the constraints.

Michael Kac's conception of corepresentational syntax (Kac 1978) informally takes an approach that is very close to this, and there are earlier antecedents in the literature as well, but Johnson & Postal's *Arc Pair Grammar* (1980) was the first work adopting this conceptualization of syntax explicitly, arguing for its merits over those of generative grammar, defining a class of candidate structures, and extensively illustrating with scores of proposed constraints (many of them argued to be universal). Postal (2010) is a later work elaborating a modified version of arc pair grammar.

The appropriate mathematical foundations for characterizing structures in a way that can be connected to the description of langugages goes back to work of the American logician Richard Büchi around 1960, and subsequent work by James Thatcher, J. B. Wright, and John Doner in the following decade (for the references, see Pullum 2013 and works cited there). Their papers are highly technical and make difficult reading. They are not cited by Johnson & Postal, who did not know about them. No linguists seem to have known about the relevant work until after it was rediscovered by James Rogers (1994). Knowledge of his work spread slowly in the later 1990s, initially among European computational linguists.

What constraints can be stated on sentence structures depends on the description language chosen. For example, suppose we consider using a first-order logic interpreted on structures that are tree-shaped graphs with labeled nodes but also labeled edges. Node labels will correspond to grammatical categories as usual, but the labels on the edges (the lines linking certain nodes) will signify grammatical relations like 'Head of' or 'Subject of'. I will write 'Head(x, y)' to mean 'x has an immediate subconstituent bearing the Head-of relation to it'. 'P(x)' will mean 'node x bears the label P', and '$\supset$' will stand for material implication. We could then state the constraint (uncontroversially a part of English grammar, and perhaps of universal grammar) that every node labeled PP (preposition phrase) has an immediate subconstituent which is labeled P (preposition) and bears the Head-of relation to it:

$(\forall x)[PP(x) \supset (\exists y)[Head(x, y) \wedge P(y)]]$

'Every PP node has a P node which is its head.'

Notice that, unlike the phrase structure operation 'PP $\rightarrow$ P NP', this actually does state that every PP in the language has an immediate subconstituent labeled P which is its head (which we might take to be a statement of universal grammar). And we could add another constraint, a parochial one applying to languages like English and Portuguese, saying that when a P node has the same parent as one or more other nodes, the P node precedes — i.e. that the language is strictly prepositional.

For my purposes in this paper, constraints can be paraphrased in ordinary language, as they are in traditional grammars. The informal English version will do for the constraint above. The grammar of English might include other constraints informally paraphrasable as follows:

- All lexical heads are initial in their immediately containing phrases.
- In a tensed clause with a third-person subject the verb has an agreement value matching the number value of the head noun of the subject noun phrase of the clause.
- A VP constituent follows any NP sibling that it has.
- A strictly transitive verb is accompanied by an NP bearing the direct object relation to the immediately containing node.

Such statements should seem very familiar: stated informally like this, they look much like the statements in any traditional grammar book of the last three hundred years (except that those grammars tended not to assume the notion of constituent structure that we assume today). However, there is now a rich mathematical and computational literature about the expressive power of statements framed in different logical languages on models of different

sorts, and it is the subject of some excellent textbooks and survey monographs such as Ebbinghaus & Flum (1999), Immerman (1999), Libkin (2004), and Grädel et al. (2007).

NATURAL CONSEQUENCES OF MTS

In the rest of this paper I survey some of the consequences of formalizing grammars in an MTS way, and relate them to the properties of generative grammars listed earlier. Some of them (notably the 'lexical dependence' and 'fixed cardinalities' properties) turn out to relate to recent controversies in the linguistics literature and will receive extended discussion.

I note first, very briefly, that the 'generative holism' point disappears immediately. Each individual constraint states a specific constraint on actual expressions of the language being described. What that means is that we have what Michael Kac (1992) has called an 'etiology of ungrammaticality': rather than simply not generating some object, and thus saying implicitly that it is not in the language, we can say for anything that is ungrammatical exactly **what makes it ungrammatical** — we can say what its ungrammaticality consists in. Kac stresses that it should be a desideratum for an adequate theory of syntax that it not only characterizes ungrammatical strings as ungrammatical, but specifies a source—a cause or origin—for the ungrammaticality. His work is the only place I know of in the 20th-century literature of linguistics where this point is explicitly made. I believe he is clearly right.

A closely related point is that we no longer have sharp boundaries to languages, if we do not want them — though we can have them if we want them. Given a grammar consisting of the statements $\{\phi_1, \phi_2, \ldots, \phi_k\}$ we can easily define the set of all structures satisfying the conjunction '$\phi_1 \wedge \phi_2 \wedge \ldots \wedge \phi_k$' if we want it. (That set might be useful if one wanted to prove that there is some type of generative grammar that could generate the same set, and thus establish an equivalence of sorts between a variety of MTS and a version of generative grammar.) But although such a sharply defined set of perfectly grammatical sentences can be defined if we need it, the ungrammatical sentences can nonetheless be represented as finely graded according to the ways in which they are ungrammatical. This is because the number of points at which a structure fails to satisfy some constraint, and/or the number of constraints satisfied or not satisfied, can be counted up. Consider this series of word-strings:

a.		*He is the chair of his department.*	
b.	*	*He are the chair of his department.*	[one error]
c.	**	*Him are chair of his department.*	[2 errors]
d.	***	*Him are chair of he's department.*	[3 errors]
e.	****	*Him chair are of he's department.*	[4 errors]
f.	*****	*Him chair are he's department of.*	[5 errors]
⋮	… · · ·		

In each case we could say explicitly which constraints are violated and where. While notions like 'almost generated but not quite' or 'this sentence is even less generated than that one' simply make no sense, the notion 'structure which almost satisfies the constraints but not quite' is fully coherent. A structure could satisfy almost all of the constraints but violate one of them at some point, or violate several of them, perhaps at more than one point. These are

perfectly coherent possibilities. This means it is possible for an MTS description to assign at least some linguistic properties to certain strings of words that are definitely **not** sentences.

Various objections that have often been voiced to this observation; I have encountered them many times in discussions. They are somewhat off-target, but generally coherent, and I will try to deal with some of them very briefly.

One objection is that one could take the sole task of the grammar to be to define the perfectly grammatical sentences, the ones that fully comport with the ideal native speaker's competence, and regard all of the gradience I just alluded to as belonging to the domain of performance. This view is the one defended at book length by Schütze (1996). The only thing I really want to say about this is that my whole point concerns the way MTS brings the statement of a grammar representing the speaker's competence closer to the syntactic phenomena. Whether the phenomena indeed reflect the character of a user's competence or merely a misleading overlay of failures for it to be reflected in performance is of course a theoretical decision, and a subtle and difficult one. But consider, for example, Lasnik & Saito (1984:266–269). They quite clearly distinguish five levels of grammaticality. For perfect sentences they use no prefix, but for ungrammatical ones they use the prefixes '?', '??', '?*', and '*'. For the expression *What$_1$ do you believe the claim* [*that John bought __$_1$*] they use the prefix '?*', but for *Why$_1$ do you believe the claim* [*that John left __$_1$*] they assign '*', and they note that the first violates merely Subjacency but not the ECP, while the second violates both. This is implicitly adopting the MTS perspective (the GB theory of the early 1980s was in fact assuming an overlay of MTS-style constraints on top of a generative grammar). They are not saying merely that the two cases differ in the extent to which performance fails to match competence; they are saying that the competence grammar differentiates them in the degree of their failure to be grammatical.

A second objection is that Chomsky showed in his early work, in several slightly different versions, how to assign differing degrees of ungrammaticality using a purely generative grammar (see Chomsky 1955[1975], esp. pp. 132–147; 1961; 1964; and 1965:148–153). The answer to this is that his approach does not work; a careful critique is given by Pullum & Scholz (2001:27–30).

A third objection that has often been presented to me is that stochastic (probabilistic) grammars can handle the phenomena that I bracket together under the heading of gradient ungrammaticality. About this I only have space to make two brief remarks. First, it is not true that the use of probabilistic grammars will deal with the phenomena, because there will always be fully grammatical sentences (infinitely many under the standard view) that are far less probable than some patently ungrammatical ones; see Pullum & Scholz (2001:31). But second, nothing in what I have claimed about the existence of degrees of ungrammaticality is incompatible with the existence of productive work bringing probabilistic methods to bear on the study of acceptability judgments and their etiology. For an example of a recent piece of work of this sort, see Lau et al. (2016).

Non-holistic acquisition

The MTS account of knowledge of language is that to command a language is to have grasped a set of constraints on utterance structure (and the mappings to meaning and pronunciation) that broadly have the same consequences as those of the other members of the relevant speech community. Under this view, language acquisition is naturally seen as an incremental process of gradually developing a set of constraints whose effects broadly match the effects of those that other speakers have developed. Hitting on **exactly** the same constraints as other learners of the language is by no means a requirement. Your constraints only need to guarantee that your sentences are structured very largely like other speakers' sentences. This is a view that stands a chance of being compatible with fairly uncontroversial facts of this sort:

- Language acquisition quite clearly takes place gradually over a period of years.
- There are signs of imperfect learning at some points during those years.
- Speakers show minor individual differences concerning the constraints they respect.
- Such individual differences often function as seeds of linguistic change.

Generative grammar has not encouraged work on language acquisition that takes this sort of perspective. Indeed, as has occasionally been pointed out, it seems quite incompatible with even the obvious fact that languages are learned gradually, and for the first few years, spoken imperfectly. I have no space to develop this point further, but it strikes me as eminently worth pursuing.

SYNTACTIC QUANDARIES

I referred earlier to a kind of clash in requirements that I call a quandary. An additional interesting property of MTS grammars is that they define quandaries as possible.

The study of quandaries goes back (at least) to a conference held in 1969 where Charles Fillmore noted facts like these (see Fillmore 1972):

> ??*Either Richard or I am usually there.*
> ??*Either Richard or I is usually there.*
> ??*Either Richard or I are usually there.*

Notice that present-tense **be** shows agreement for person and number, but singular disjunctive NPs of different persons have no person specification, hence intuitively there is a clash of requirements. Under MTS this intuitive idea gets a formal expression: for sentences like the ones just exhibited it is impossible to satisfy all the appropriate constraints for standard English simultaneously.

Consider another example. I assume that among for standard English we would want the grammar to include constraints informally paraphrasable as follows:

- All coordinates in a coordination of NPs take the case assigned to the whole coordination (thus coordinates do not differ in case).[4]

---

[4] There is in fact a lot of variation in whether native speakers of English adhere to this constraint. But for now, just assume we are talking about a rather puristic kind of educated standard English such as we find in professionally written nonfiction prose.

- NP determiners take the genitive case.
- Genitive case on multi-word NPs is marked by the suffix *'s*.
- The suffixed *'s* on multi-word NPs is enclitic on the last word, as in *I was amazed at the person I spoke <u>to's</u> rudeness*.
- The *'s* suffix is never encliticized to a pronoun that has its own irregular genitive. **The water damaged a picture of <u>me's</u> frame*.

These statements are not mutually contradictory, but taken together they entail that no fully grammatical coordinate determiner NP can end in a personal pronoun.

The result is that every attempt to use a genitive NP determiner to say 'the book written by him and me' fails to be fully grammatical:

?*he and I's book*	?*he and me's book*
?*he and my book*	?*he and my's book*
?*him and I's book*	?*him and me's book*
?*him and my book*	?*him and my's book*
?*his and I book*	?*his and I's book*
?*his and me book*	?*his and me's book*
?*his and my book*	?*his and my's book*

Many speakers would be inclined to say that the best of this bad bunch is ?*his and my book*, and if so, that is interesting, because of course that one only violates the third constraint (that genitive case on multi-word NPs is marked by the suffix *'s*), but satisfies all of the others. It is less ungrammatical than any of its rivals.

The same is true (perhaps worse) if the coordination of NPs is in the **independent genitive** form (as in *mine*, *yours*, *hers*, *ours*, *theirs*), and its last coordinate is a personal pronoun with an independent genitive form distinct from its dependent genitive:

?*This book is him and my*.
?*This book is him and mine*.
?*This book is him and me's*.
?*This book is his and me's*.
?*This book is his and mine*.

There is really no convincingly well-formed way, using this construction, to say what these sentences are clearly attempting to say; we need to recast the sentence (e.g., as in *This book was co-authored by him and me*).

The point established by examples of this sort is that they suggest **grammars are collections of independent constraints** which can, on occasion, be in conflict in a way that makes a certain construction unusable or unacceptable. A generative grammar simply defines a set of perfectly well-formed strings and leaves nothing unclear. But the facts seem to be that human languages are not well-defined sets of this sort. Rather, they are ways of structuring sentences, expressed by sets of constraints on syntactic form that, in certain marginal circumstances, are not always capable of being simultaneously satisfied.

Now let us return to the crucial fact that generative grammars base their account of syntactic phenomena entirely on the contents and properties of the lexicon. Several writers on philosophy and semantics have noted, in effect, that this is a fundamental error. The most famous, and also the earliest, was Lewis Carroll (the pseudonym of a Cambridge logician named Charles Dodgson), who hinted at it in his famous children's book *Alice's Adventures in Wonderland* (1865), through a nonsense poem, *Jabberwocky*. It begins:

> *'Twas brillig, and the slithy toves*
> *Did gyre and gimble in the wabe*
> *All mimsy were the borogoves*
> *And the mome raths outgrabe.*

All the nouns, verbs, and adjectives here are complete nonsense, invented by Carroll for the amusement of children. But remarkably, a native speaker of English will not only recognize this stanza as English, but will perceive that it is somewhat archaic and poetic English, in rhyming verse with an ABAB pattern. But how could it possibly be perceived as linguistic material at all, given that by design none of its major-category words are in the English lexicon, and thus these strings of words would not be generated by a correct generative grammar for English?

The first person to make a serious theoretical point about what such phenomena mean seems to have been Andrew Ingraham (1841–1905), a Massachusetts schoolmaster, in remarks in a book called *Swain School Lectures* (1903:154), quoted at length in a famous philosophy book, *The Meaning of Meaning* by Ogden & Richards (1923:46). Ingraham points out that 'there are results of thinking which could not be obtained at all without language or symbols of sight or sound analogous to language'. He presents this thought experiment:

> Suppose someone to assert: *The gostak distims the doshes*. You do not know what this means; nor do I. But if we assume that it is English, we know that *the doshes are distimmed by the gostak*. We know too that *one distimmer of doshes is a gostak*. If, moreover, the doshes are galloons, we know that *some galloons are distimmed by the gostak*. And so we may go on, and so we often do go on.
> [https://archive.org/details/swainschoollectu00ingruoft/page/154]

Ingraham's point here is that we can actually derive items of knowledge from given assumptions purely through our command of the syntax of our language. Without any knowledge of what a gostak might be, we can come to understand, simply from being told that the gostak distims the doshes, that (for example) doshes are capable of being distimmed. Of course, knowing what the properties of gostaks are, or what distimming is like, is quite another matter; but in principle—if there were gostaks, and if there were such a process as distimming, that would be a matter of finding things out about the world.

Rudolf Carnap, in *The Logical Syntax of Language* (1937:2), possibly prompted by Ogden & Richards, makes a comparable point, inventing the sentence *Pirots karulize elatically*, which is visibly English despite containing no English words at all. He writes:

> [G]iven an appropriate rule, it can be proved that the word-series "Pirots karulize elatically" is a sentence, provided only that "Pirots" is known to be a substantive (in the plural), "karulize" a verb (in the third person plural), and "elatically" an adverb; all of which, of course, in a well-constructed language—as, for example, in Esperanto—could be gathered from the form of the words alone. The meaning of the words is quite inessential to the purpose, and need not be known. Further, given an appropriate rule, the sentence "A karulizes elatically" can be deduced from the original sentence and the sentence "A is a pirot"—again, provided that the type to which the individual words belong is known. Here also, neither the meaning of the words nor the sense of the three sentences need be known.

Carnap notes that we would have to be sure *pirot* is a noun with a regular plural, and that *karulize* is a regular verb, and *elatically* is a typical *-ly* adverb. He believes that any 'well-constructed language' would signal the category of each of its words (or terms or signs) overtly in its form.[5] But the main thing he wants to stress is that drawing inferences from a sentence need not require knowing anything about the meaning of the constituent words: we do not know what pirots are (because there is no such thing), but that is not necessary in order for us to be able to infer that if some specified object *A* were a pirot, and the statement *Pirots karulize elatically* were true, then it would be true that *A* karulizes elatically. What I am pointing out is something different, and it does not involve recognition or reasoning. It is simply about the linguistic phenomena. It seems that we have to classify sentences like Ingraham's and Carnap's not as gibberish, but as English. They refer to things and processes that do not exist, but they do so in English. This is simply not compatible with saying that English is what English speakers know about their language, and knowledge of English is expressed as an internalized generative grammar with a finite lexicon.

Charles C. Fries, in *The Structure of English* (1952:111) makes the point that 'we can make nonsense words and produce utterances in which the structural meanings are perfectly clear', using his own invented example:

The vapy koobs dasaked the citar molently.

This has been quoted several times in the English language teaching literature, or replaced by other such examples with the same general point (Benjamin & Oliva 2007:63 use the example *I found a flindering fleek on the floot*; they cite a blog by Kristin Denham using *The dorbling groobies frandled a bonkled slank*; and so on). The point of all these nonsense sentences for teachers of English grammar and language use is to stress that students can be taught to spot the nouns and the verbs even in sentences with invented nonsense words. And there are sound points to be made thereby, it seems to me. One of them is that the traditional definition of a noun as a word naming a person, place, or thing is exposed as hopelessly

---

[5] Logical calculi are designed so as to meet this condition, of course, employing *p* and *q* to denote propositions, *x* and *y* for individual variables, etc. Carnap mentions Esperanto because it too was designed by its inventor, Ludoviko Zamenhof, to meet the condition: all common nouns end in *-o*, adjectives end in *-a*, adverbs end in *-e*, etc.

inadequate: there are no people, places, or things named *grooby*, *slank*, *floot*, or *fleek*, yet these are clearly nouns in the examples just cited.

Chomsky (1957:104) cites Carnap's example, without attribution, in the course of arguing against what he says is a common argument for 'structural meaning'. But he seems not to have appreciated the fundamental point, and it has been completely missed by subsequent generative grammarians as well: if our knowledge of English is theoretically modeled by a generative grammar, it is inexplicable that any strings with novel words in them could be construed as English, or as any kind of linguistic material. Since they are not generated by the grammar that constitutes a speaker's internalized linguistic knowledge, they are defined as not being English at all.

It is remarkable that generative grammarians have not realized what a devastating argument this is against the claim that our linguistic competence is accounted for simply by assuming an internalized generative grammar ('I-language') defining the set of strings that belong to our language (our 'E-language'). Our behavior when faced with sentences containing novel lexical items does not suggest that at all; in fact it seems flatly incompatible with it.

Under an MTS account, things can be different. If lexical statements are conditions on phonological forms specifying what grammatical and semantic properties are associated with them, a sentence like *Pirots karulize elatically* can be characterized as not just similar to English in its structure but as fully well-formed English. And not just well-formed but also meaningful. Because Carnap is right in what he says: an English speaker who hears it knows that it describes something called a pirot as capable of karulization, and it entails that karulization can be elatic, and that it is the elatic kind of karulization that pirots undergo, and so on. What they don't know is what sort of thing a pirot is, or what happens to something when it karulizes, or how elatic processes differ from non-elatic ones, and so on.

What makes MTS relevantly different is that it represents the syntactic structure — the NP-VP form of the clause, the Adverb label of *elatically*, the plural NP node over *pirots*, etc. — quite independently of the lexical details. There are lexical constraints that say *the* must be under a D (determinative) label in an NP (either singular or plural, it doesn't matter), and *this* must be labeled D and also singular, but none of the lexical constraints say anything about *pirot* or *karulize*, so those words cannot possibly violate any such constraints.

In fact these nonsense words do not violate any constraints at all. The constraint that (in all but a limited class of nouns) a plural noun must end in *-s* is satisfied by *pirots* (if we take it to be a noun). The constraint saying that a present-tense verb must not have a final *-s* if its subject is plural is satisfied by *karulize* if we take it to be a verb. The constraint saying that a transitive verb must have an object is satisfied (vacuously) if we take *karulize* to be an intransitive verb. No constraints mention *pirot* or *karulize* or *elatically* specifically, precisely because those words do not exist.

A RELATED POINT ABOUT PARSING AND LEARNING

When touching on the point about the lexical independence of MTS in lectures, I have nearly always faced questioners suggesting that a generative grammar could be modified to accommodate the facts in some way. In effect they ignore the issue I addressed (namely, what

a grammar should say about the linguistic facts), they shift attention to to how parsing and acquisition work. Neeleman (2013) is one person who articulated this response in print. He says:

> The problem dissolves if parser and grammar are taken to be descriptions of the language faculty at different levels. The robustness of the parser makes it likely that the parsing process will not terminate if a terminal is identified whose phonology does not match an existing word. . . In case the input contains a phonological form that is not part of a speaker's permanent lexicon, a good strategy might be to store the relevant form in a 'temporary lexicon', and to try and identify a meaning for it (presumably, this is how new words are learned).

This is a change of subject to something different from my point, but the new subject is important, and deserves some discussion. My concern so far has been with how the syntactic facts of English could be described in a way that does not turn it into a mystery that there can be sentences of English containing words that are not in the English lexicon. This is purely about the language, and how the statement of the grammar might best be framed. The new topic is quite different, and brings together two psycholinguistic issues: (i) how people process utterances that they encounter, and (ii) how they learn novel words.

These are important questions because we constantly read or hear sentences containing words we happen not to know. The day before I first drafted this paragraph I was reading a newspaper science article and came across the word *solenodon*. I had never heard of solenodons before (perhaps you haven't either: they are small, highly endangered mammals with long, hairless snouts, sharp teeth, and poisonous saliva, and they live in the forests of Haiti and the Dominican Republic). How did I even recognize that I was reading an English sentence, and that *solenodons* was a plural noun? A minute later I had the experience again when I came to a sentence containing the word *kallikreins*. (A kallikrein is a type of enzyme that a solenodon has in its venomous saliva.) Later the same day in the science and technology section of *The Economist* (December 14th, 2019) I encountered the words *therianthrope* and *therianthropy* for the first time in my life. We barely notice how often we have such experiences.

It is not just about nouns: we encounter novel verbs and adjectives as well, almost as often. The article mentioning therianthropy used the adjective *Sulawesian*, which I had never seen (it means 'from the Indonesian island of Sulawesi', and I could guess that), and the adjective *buccal* occurred on the same page ('*Baleen whales suck in mouthfuls of water and extract small organisms such as krill, using fibrous buccal filters*') — there was no indication that buccal filters are in or near the cheeks.

And as for verbs, an article in *The Economist* about America's abandonment of the agreement with Iran on production of fissile nuclear material (28 March 2018) said:

> *Iran might resile from the deal, further roiling an unstable region at risk of tit-for-tat nuclear proliferation.*

Resile is a verb that I have never used and do not recall encountering. You might well not know it, just as you might well not know the verb *roil*.

Even novel prepositions show up occasionally. The preposition *outwith* appears to have spread from the English of Scotland to the English of the rest of the United Kingdom over the past 40 years.

An account of knowledge of language that says your knowledge of your language is physically realized by a generative grammar with a defined lexicon inscribed in your brain, it is mysterious how such a person could recognize a sentence with a novel noun or unfamiliar verb as being linguistic material at all, since their grammar would not assign the string any derivation, and thus would not say anything about it.

I have stressed is that if syntactic constraints can be imposed on structure independently of what nouns or verbs there happen to be in the language, they could be held unchanged in the grammar no matter what items were dropped from the lexicon or added to it. I take this to be a powerful argument in favor of looking at syntax in MTS terms. But notice that I put matters entirely in terms of the theory of syntax: I have been talking about how to describe syntactic constraints independently of any predetermined lexicon. This should not be confused with the psycholinguistic question of what happens in real time when a user of a language encounters an utterance with an unknown word in it.

The confusion is common: generativists with whom I have discussed the issue of lexical independence nearly always respond like Neeleman, suggesting that humans could possess some kind of special cognitive module that responds to novel substrings by triggering some kind of operation that builds a new lexical item for them on the fly, conjecturing a suitable category for the item to see if a parse results. On its own, that is just an unimplemented speculation about a sort of imaginary device that would make the problem go away, and does not even bear on my main point about representing the structure of sentences containing nonexistent words in a way that makes lexicon-independent syntactic structure theoretically intelligible.

However, lexical acquisition is a matter of fundamental importance for psycholinguistics. Learning what words there are, and what their syntactic categories and meanings are — one aspect of language acquisition that must be for the most part parochial, i.e. largely independent of universal grammar — is going to constitute the bulk of the acquisition task for an infant. Under some highly lexicalized post-1980 theories of grammar it constitutes essentially the whole of the task. But how is such lexical acquisition accomplished? In the past decade there have been significant developments within computational psycholinguistics that bear on that question. Modern machine learning techniques have made it possible to actually implement a relevant acquisition process.

The most interesting and sophisticated work of this sort, without any doubt, is that associated with Mark Steedman's research program. Steedman and his collaborators have developed working computational implementations of CCG, in conjunction with a probabilistic model to reduce indeterminacy caused by ambiguity. These implementations handle not just syntax but also logical form. And based on them, the elements of language acquisition have also been modeled. Kwiatkowski et al. (2010) describes an algorithm that simulates language acquisition via 'semantic bootstrapping', where the input consists of unanalysed sentences paired with logical forms (adopting the plausible assumption made by many computational language acquisition researchers that children guess meanings for

sentences they hear uttered in context). The input to the algorithm is sentence/meaning pairs, and the algorithm works out how best to break the sentences into words and hypothesize syntactic categories and meanings for those words.

A more recent paper, Abend et al. (2017), reports success not only in the matter of breaking utterances into words and hypothesizing lexical information for the words but also in one-trial learning of categories for invented nonsense words: their model was able to guess the categories of an invented verb *dax* in the *Jacky daxed Jacob*, even though all three words were previously unseen, given the logical form *daxed*(*Jacky*, *Jacob*). It was also able to diagnose the categories of invented nouns, verbs, and even prepositions.

Abend et al. are in effect showing a way in which a composition-oriented generative grammar can be reconciled with the apparently insuperable stumbling-block of being unable to generate any sentence if its words are not known in advance. It is true that a derivation of a sentence is impossible if the words in the sentence, complete with their syntactic categories and semantic contributions, are supplied at the start. But it is also true that a CCG-based parser given unanalysed sentences paired with logical forms can develop its own lexicon using modern machine learning techniques. My proposal is that we should view the syntactic facts as indicating that syntactic constraints are independent of accidental lexical facts. That is not in any sort of conflict with the goals of the productive and important program of Steedman and his colleagues, which spans the boundaries of computational linguistics and computational psycholinguistics. Grammars constructed within the purely generative CCG framework play an important role in their work, and so do other systems, such as logical form representations, probabilistic modeling, and a sophisticated machine-learning algorithm.

IRRATIONAL CONTROVERSY OVER INFINITUDE

I made the point earlier, under the heading 'fixed cardinalities', that once we regard languages as sets of expressions, they have to have fixed sizes (cardinal numbers, or cardinalities). That defines a fundamental distinction between finite and infinite languages. An interesting feature of MTS is that it completely dissolves certain disputes rooted in this distinction.

Some linguists seem to have looked at these questions from entirely the wrong perspective, apparently assuming that we could first determine that a language has infinitely many sentences and then start developing a theory to account for this remarkable property. For example, Epstein & Hornstein (2005) say:

> This property of discrete infinity characterizes human language; none consists of a finite set of sentences. The unchanged central goal of linguistic theory over the last fifty years has been and remains to give a precise, formal characterization of this property and then to explain how humans develop (or grow) and use discretely infinite linguistic systems.

Likewise, Lasnik (2000:3) states:

> Infinity is one of the most fundamental properties of human languages, maybe the most fundamental one. People debate what the true universals of language are, but indisputably, infinity is central.

These writers have it backwards. Infinite size is not a discovered property of human languages: no evidence of any kind could either confirm it or refute it. Epstein & Hornstein's phrase 'discretely infinite linguistic systems', incidentally, is a mistake: it is not the system of rewriting operations that is infinite — nobody ever suggested that what Chomsky calls 'I-languages' are infinite, since under Chomsky's view they are physically realized in finite brains; Epstein & Hornstein must mean something like 'how humans develop (or grow) and use linguistic systems that are capable of specifying the syntactic properties of discretely infinite sets of sentences.' But that is also wrong: how could the contingent existence of some electrochemical regularity in a mammalian brain possibly determine that some set of sentences is infinite?[6]

The apparent necessity of an infinite number of sentences arises because we adopt the practice of using generative grammars as descriptions, and a generative grammar, as pointed out earlier, can define just two kinds of set. The first possibility is that it defines a finite set, so that the length of sentences is limited to some fixed number of words. The other possibility is one that can result if there is some reachable category with non-zero yield that can, non-trivially and without limit, be embedded within larger instances of itself.

The advantage of an MTS description is simply that it is **silent** on questions of language size and expression size, which is the best stance for a syntactic theory. A constraint on expression structure, or a finite set of such constraints, does not entail that infinitely many objects with that structure exist, or that they don't. And this is welcome: no syntactician should have to pretend there are serious answers to questions about how many sentences exist, or how long they can get. On the assumption that there are infinitely many sentences, all of the extremely long ones are surely unacceptable to every speaker (one can hardly call a sentence acceptable if it is not even humanly utterable, e.g. if uttering it or listening to it would take longer than any human can remain awake). But the extremely long sentences are the overwhelming majority. In other words, most of any infinite set of grammatical sentences will be totally unacceptable. It would be preferable if we did not have to make (or deny) that assumption about the language. An MTS grammar provides exactly that kind of neutrality, because MTS constraints make statements about **what expressions are like**, structurally, with no entailments about how many there are, or how big they are allowed to be.

Fifteen years ago the question of whether there are human languages in which the stock of sentences went from being never considered to being hotly controversial when Daniel Everett (2005) called specific attention to a Brazilian indigenous language, Pirahã, in which the main features taken to invalidate any finite upper bound on the number of sentences seem to be absent.

---

[6] This should look like an allusion to the problem discussed by Kripke (1982). The allusion is intentional: Kripke himself saw that the problem he was studying bore directly on questions about the mental reality of competence grammars.

This seemed to contradict a hypothesis expressed by Hauser et al. (2002) about the exclusively human and specifically linguistic aspects of the human capacity for language, the 'faculty of language narrow sense' (FLN). They claim (p. 1571) that 'a core property of FLN is recursion', which 'yields discrete infinity'. The latter phrase recalls the remarks in Chomsky (1991:50) about language being 'at its core, a system that is both digital and infinite' — that is, it exhibits structure calling for a discrete algebraic representation rather than the mathematics of continuous functions. This core property 'is intuitively familiar to every language user' in that people are aware that 'There is no longest sentence' , i.e., 'there is no non-arbitrary upper bound to sentence length' (p. 1571). Clarifying in a later paper, the same authors say that they are focusing on 'a known property of human language that provides its most powerful and unusual signature: discrete infinity'; they repeat: 'Whatever else might be necessary for human language, the mechanisms underlying discrete infinity are a critical capability' of humans (Fitch et al. 2005:182). The infinity they speak of resides not, of course, in the size of the grammar, which must be finite and representable in a brain of finite capacity; it is the number of definable sentences that is infinite. There is no doubt, then, about what Chomsky, Fitch, and Hauser are saying: they believe it to be a 'signature' property of human language that it involves 'at its core' a system defining an infinite set of sentences. On that point, Everett directly challenges them: if he is right about Pirahã, it apparently does not exhibit this 'discrete infinity'.

Some of the reactions to Everett (2005) were clearly driven by emotional rather than scientific impulses. When Everett was invited to speak at MIT's Department of Brain and Cognitive Sciences in December 2006, one Chomsky defender broadcast to two MIT email lists a furiously abusive message urging a boycott of Everett's lecture and accusing him of lying about the Pirahã — of wanting to 'enjoy the spotlight of mass media'. You or anyone can do the same, he urged sarcastically: 'Just find a remote tribe and exploit them for your own fame by making claims nobody will bother to check!'.[7] And Chomsky himself showed unusual signs of irrational anger concerning Everett's claims, a couple of years later, when Brazil's biggest-selling daily newspaper asked him about Everett. Chomsky said (as translated into Portuguese, in *Folha de S. Paulo*, 1 February 2009):

> Ele virou um charlatão puro, embora costumava ser um bom linguista descritivo. É por isso que, até onde eu sei, todos os linguistos sérios que trabalham com linguas brasileiras ignoram-no.

> ['He became a pure charlatan, although he used to be a good descriptive linguist. That is why, as far as I know, all the serious linguists who work on Brazilian languages ignore him.']

---

[7] See http://itre.cis.upenn.edu/~myl/languagelog/archives/003837.html for a brief account on Language Log. In 2013 the same person teamed up with two Brazilian colleagues to produce a video (https://www.youtube.com/watch?v=J3jWI4cPRMg&t=391s) purporting to show a Pirahã man alleging various ethically dubious acts by Everett during his early years as a missionary. But having no knowledge of the language, the makers of the video could not tell that they were being duped by their interpreter, who has hardly any knowledge of Pirahã at all and simply made things up (see Everett and Gibson 2019, p. 781, fn. 3, for some details).

That is an extraordinary piece of personal defamation, attacking Everett's integrity and reputation rather than addressing linguistic facts. And the part about how 'serious linguists ... ignore him' is outright dishonesty: Chomsky has never worked on Brazilian indigenous languages and never discusses the work of those who have, and he has no grounds for estimating Everett's standing among Amazonianists. Everett's expertise is not questioned by the SIL missionary linguists among whom he used to work, or by any of the roughly twenty researchers who have spent time with him among the Pirahã, or by any of the linguists who have actually made progress on learning the Pirahã language (Arlo Heinrichs, Keren Madora, Jeanette Sakel, Steven Sheldon, Eugénie Stapert). Chomsky then continues with an unverifiable claim about Everett's thought processes:

> Everett espera que os leitores não entendam a deferença entre a GU no sentido técnico (a teoria do componente genético da linguagem humana) e no sentido informal, que dis respeito às propriedades comuns a todas as línguas.

> ['Everett hopes that the readers do not understand the difference between UG in the technical sense (the theory of the genetic component of human language) and the informal sense, which concerns properties common to all languages.']

The tactic here is to suggest that the remarks about 'recursion' in Hauser et al. (2002) never intended any claim about languages, but only meant to say that the genetically transmitted ability to acquire languages with such features as hypotaxis was innate in all human beings. This claim, Everett never addressed; it is empty, as he notes (Everett 2009:439): it seems to say merely that 'only humans speak because only humans are humans.' Anyway, the quotations from Hauser et al. given earlier ('There is no longest sentence', etc.) clearly refer to 'properties common to all languages'.

The replacement of scientific objectivity by personal abuse is accompanied by a remarkable inattention to syntactic typology on the part of Everett's critics. The claim that natural languages could lack features like hypotaxis was not novel or unprecedented. As Fred Karlsson helpfully reminds us (2010:46–49), there are several distinct syntactic devices that, if applied unboundedly, could cause the set of all sentences to be infinite. These include syndetic coordination (*You need celery, apples, walnuts, and grapes*); asyndetic coordination (*You need celery, apples, walnuts, grapes*); stacking (*a [nice [cosy [little [cottage]; [my [brother's [wife's]]] cousin*); apposition (*London, the capital, a major financial center*); repetition (*She wept and wept and wept*; *It's very, very, very rare*); and hypotaxis, or syntactic subordination (*Knowing that [the doctors thought [it was incurable]] didn't help*). If it is not guaranteed that every human language will employ all of them (with no bound on iteration), then in principle there could be a language that exhibited none of them (or had limitations on all of them), and thus had a finite number of sentences. Everett's claim is limited and specific: he finds that Pirahã has no coordination, stacking, or hypotaxis. This should not provoke accusations of charlatanism.

Coordination is completely lacking in other languages; for example, Dyirbal seems to be one of the many languages that use simple juxtaposition of sentences to express conjunction: Dixon (1972) gives no indication any word meaning 'and'; and he states explicitly (p. 361)

that Dyirbal has no exact correspondents of English *or*, *if*, or *because*. It is also uncontroversial that there are languages in which stacking is disallowed; for example, Nevins et al. (2009:368) point out that German does not allow stacked genitive determiners. The keenest focus of attention for syntacticians, though, has been on hypotaxis: Pirahã seems to have no subordinate clauses, only simple main clauses. Nevins et al. (2009) seem desperate to reinterpret Everett's data in ways that are compatible with some sort of nonfinite subordination. But what no one seems to have noted is that linguists had long noted the apparent absence of hypotaxis in a variety of human languages. I offer ten examples in chronological order, from the half century before Everett (2005) was written:

- Paavo Ravila, in a chapter contributed to Collinder (1960), makes the following remark about Proto-Uralic (PU) and many of its modern descendants (mainly languages of northern Russia such as Erzya, Moksha, Mari, Udmurt, and Komi):

    > In PU there was no hypotaxis in the strict sense of the word. The sentences were connected paratactically, and in the absence of conjunctions, the content determined the mutual relations of the sentences. In PU, as in the Uralic languages spoken today, the subordinate clauses of the Indo-European languages had as counterparts various constructions with verb nouns…

- Hale (1976) drew attention to the fact that relative clauses in Warlpiri are not embedded within the NP but rather loosely adjoined to the clause; in Australan languages quite generally the distinction between relative and complement clauses is often unclear, and so is the question of whether there is any true syntactic embedding of finite clauses (Nordlinger 2006 discusses another language of this sort).

- Derbyshire (1979) provides a detailed grammar of Hixkaryána, a Cariban language of northern Amazonia, in which he notes that it has no finite complement clauses, no indirect speech reports, and no clause-taking verbs of propositional attitude. Derbyshire says: 'Subordination is restricted to nonfinite verbal forms, specifically derived nominals' or 'pseudo-nominals that function as adverbials' (compare Pavila's reference to 'verb nouns', i.e. nominalizations); and 'There is no special form for indirect statements such as "he said that he is going"…'. Hixkaryána does have a verb with the root *ka-* meaning 'say' that can introduce directly quoted speech, but that does not involve subordination (the quoted material in a direct speech report can be an independent sentence, a sequence of sentences, or some speech in a different language). Derbyshire also notes that Hixkaryana lacks 'formal means … for expressing coordination at either the sentence or the phrase level' — that is, there are 'no simple equivalents of "and", "but" and "or".'

- Talmy Givón (1979) relates absence of hypotaxis to both diachronic and phylogenetic language evolution. He claims: 'diachronically, in extant human language, subordination *always* develops out of earlier loose coordination', and adds that the evidence suggests it 'must have also been a phylogenetic process, correlated with the increase in both cognitive capacity and sociocultural complexity'. He further observes:

    > [T]here are some languages extant to this day — all in preindustrial, illiterate societies with relatively small, homogeneous social units —

> where one could demonstrate that subordination does not really exist, and that the complexity of discourse-narrative is still achieved via "chaining" or coordination, albeit with an evolved discourse-function morphology…

- Dixon's (1981) grammar of Wargamay (Pama-Nyungan) has a section headed 'complement clauses', but his examples there are clearly not complements in the usual sense; they are untensed adjunct phrases of result or purpose. No tensed subordinate clauses are exhibited.

- Moore (1989) discusses the way the Tupian language Gavião (Rondônia, Brazil) uses nominalization as a substitute for what in English might be expressed by relative or complement subordinate clauses, very much as Derbyshire (1979) reports for Hixkaryána. This nominalization, according to Moore, forms a complex noun. I find no indications in his work of any possibility if nominalizing further clauses end embedding the results inside such nouns. The same is true for a longer and earlier study, Moore (1984). There are clear signs of paratactic sequences of VPs in a manner reminiscent of apposition, but I see no clear examples of clausal hypotaxis.

- Evans (1995) notes that an obligatory marker of oblique case is added to each constituent in a relative clause in Kayardild, and this makes it impossible to embed further relative clauses inside it, so embedding is limited to a depth of one.

- Deutscher (2000) argues at length that when Akkadian was first written it did not have finite complement clauses (though later in its history it developed them).

- Engelbretson (2003) finds no clear evidence for any subordination in conversational spoken Indonesian, a conclusion that Gil (2009) firmly endorses for the spoken Indonesian of the Riau province.

- Comrie & Kuteva (2005), citing John Roberts, mention Amele (Papua New Guinea) as a language in which the semantic role of relative clauses is played by clauses that show no signs of subordination (as opposed to mere juxtaposition). In such languages (as others have also noted) it is not at all clear that one can distinguish relative clauses or complement clauses from independent clauses following them.

And so on, and so on. For half a century, linguists have been describing languages that lack freely iterable hypotaxis — or at the very least, show no unambiguous signs of such properties in attested sentences. It is irrational and ethnocentric to express horror at the idea of a language not having the same sorts of syntactic devices for sentence-lengthening that we find in the languages of Western Europe, with their long history of literacy.

But the relevant point I want to add is that in addition to being empirically unjustified, the ill-tempered debate over Everett's observations about Pirahã is also mathematically misguided. The infinitude issue is not an empirical one; it is an artifact of misguided reifying of generative theory. Claiming that human languages have infinitely many sentences (the error that Epstein & Hornstein make) is not the right way to put it, even for languages like English; rather, what is true is that when languages with certain syntactic and lexical

properties are modeled in the most perspicuous way with generative mechanisms, infinite sets play a role in the resultant theoretical reconstruction.

Just as a thought experiment, consider a language like English which respects the constraint (mentioned above) that lexical heads are initial in their minimal phrases. The constraint entails that any internal complement selected by a lexical head will follow that head. English happens to have verbs, nouns, adjectives, and prepositions that select clause complements. The constraint is very general, entailing not only that direct objects of transitive verbs will follow the verb (*devoured _the food_*), that PP complements will follow adjectives (*proud _of his accomplishments_*), and that clause complements will follow verbs (*believe _he died_*), nouns (*belief _that he died_*), adjectives (*certain _that he died_*), and prepositions (*after _he died_*). But English exhibits evidence confirming the position of clause complements only because English has lexical items that select them. That is a matter of lexical accident. If there happened not to be any lexical items that selected clause complements, the constraint could stand unaltered, and still a valid part of English syntax, but there would be no evidence from clause complements supporting it. The syntax would not forbid clause complements; it would actually stipulate their positioning (they would follow any lexical head that selected them), but given a complete lack of lexical items selecting clause complements, it would be satisfied vacuously in that regard.

Whether clause complements are found in all languages should thus be seen not as a deep syntactic fact, but as a lexical one, both for MTS frameworks and modern highly lexicalized generative frameworks. The syntactic constraints of an MTS grammar can stipulate where clause complements would go, but it is the lexicon that controls whether there will be any. It could turn out that there aren't. And that lexical accident would decide whether set of all possible sentences was infinite or merely a large finite set.

More generally, no MTS description will either entail that there are infinitely many well-formed clauses or entail that there aren't. The same set of constraints can be compatible with either situation. When a constraint in English grammar says that lexical heads precede their complements, it doesn't have any implications for how many sentences there are or how long they can get. If for some reason there are no sentences beyond a certain finite length, or we decide for some theoretical or practical purpose to ignore sentences greater than that length, the constraint will have the same effect. For most purposes, that will be all a syntactician is interested in: the size of the set of all sentences, or the maximum sentence length if there is one, will be irrelevant to the grammarian's concern.

The question of whether all human languages have verbs taking subordinate clause complements should be investigated carefully and objectively. At present the answer seems to be no (recall the relevant citations above). But those involved in the controversy over Everett's finding about Pirahã who have implied that he is demeaning or insulting its speakers by drawing his conclusion are mistaken. His more than 40 years of close study of the Pirahã language, his respectful attention to the culture of its speakers during many years of residing with them, and his professed friendship with them all make such charges implausible. His conclusion certainly says nothing about their thought processes, because it is obvious that a person can think a thought that cannot be expressed directly in a single

sentence of their language. English syntax does not permit *wh*-extraction from clausal subjects:

Which subject did that you did well in surprise your teachers the most?

Yet obviously an English speaker can think the relevant thought. You can surprise your teachers by doing well in several subjects at school; there can be one that strikes them as the most surprising; and I can ask which subject that was. It's just that I can't ask the question directly by means of a sentence like the one above. That does not mean that I find the thought unthinkable, or that English speakers should feel insulted by the (true) allegation that their language does not permit its direct expression.

As Everett has frequently affirmed, Pirahã speakers are clearly capable of evincing beliefs that imply embedding of propositions within propositions. See also Levinson (2013) for evidence from discourse and pragmatics with patterns of embedding that English would not allow in its syntax. In this domain, restrictions on language simply do not entail restrictions on thought.

CONCLUSION

Generative grammar arose in mathematical logic and early theoretical computer science as a way of defining sets of strings of symbols for various purposes, one of them being the study of the structural characteristics of recursively enumerable sets of integers. For six decades, syntactic theorists have been applying it to human languages, as Chomsky proposed.

A different mathematical apparatus is available: syntax can be formalized in model-theoretic terms. The relevant mathematics has been developing since 1960, and has spawned a couple of research subareas within logic (finite model theory) and theoretical computer science (descriptive complexity theory). I have argued that various linguistic phenomena, particularly facts about the gradience of ungrammaticality and the fact that grammatical and meaningful sentences can contain nonsense words, suggest that an MTS mode of formalization might be best for theorizing about the kinds of syntactic systems humans use.

Among other points in its favor, MTS dissolves away the irrationally hostile and pointless debate about whether or not Pirahã has 'recursion' in a sense that entails it has infinitely many sentences. Some linguists seem to treat this as an issue of great ideological importance, as if it were essential for the dignity of human beings that they should be in possession of an infinite fund of sentences. But the question is not a serious one, and should not engage the attention of syntactic theorists. The job of syntactic theory is to provide descriptions of, and theories about, the internal structural properties of phrases, clauses, and sentences. It not the job of syntactic theory to count them or measure them.

queries. Their help was crucial in permitting me to develop and improve the paper to its present (doubtless still flawed) state. They should not be assumed to agree with what I say; in some cases tbey definitely do not. And the remaining errors and inadequacies are of course solely my fault.

REFERENCES

Abend, Omri; Tom Kwiatkowski; Nathaniel Smith; Sharon Goldwater; and Mark Steedman. 2017. Bootstrapping language acquisition. *Cognition* **164**, 116–143.

Ades, Anthony E. and Mark J. Steedman. 1982. On the order of words. *Linguistics and Philosophy* **4**(4), 517–558.

Bar-Hillel, Yehoshua. 1953. A quasi-arithmetical notation for syntactic description. *Language* **29**, 47–58. Reprinted with revisions in Bar-Hillel (1964), 61–74.

Bar-Hillel, Yehoshua. 1964. *Language and Information: Selected Essays on Their Theory and Application*. Reading, MA: Addison-Wesley.

Bar-Hillel, Yehoshua; Chaim Gaifman; and Eliyahu Shamir. 1960. On categorial and phrase structure grammars. *Bulletin of the Research Council of Israel* 9F:1–16. Reprinted with revisions in Bar-Hillel (1964), 99–115.

Benjamin, Amy and Tom Oliva. 2007. *Engaging Grammar: Practical Advice for Real Classrooms*. Urbana, IL: National Council for Teachers of English.

Berko, Jean. 1958. The child's learning of English morphology. *Word* **14**(2–3), 150–177.

Carnap, Rudolf. 1937. *The Logical Syntax of Language*. New York: Harcourt Brace.

Chomsky, Noam. 1955[1975]. *The Logical Structure of Linguistic Theory*. Unpublished typescript. Page reference is to the version published by Plenum Press, New York, 1975.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1961. Some methodological remarks on generative grammar. *Word* **17**, 219–239.

Chomsky, Noam. 1964. Degrees of grammaticalness. In Jerry A. Fodor and Jerrold J. Katz (eds.), *The Structure of Language: Readings in the Philosophy of Language*, 384–389. Englewood Cliffs, NJ: Prentice-Hall.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1991. Linguistics and cognitive science: Problems and mysteries. In Asa Kasher, ed., *The Chomskyan Turn*. Cambridge, MA: Basil Blackwell.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In Kenneth Hale and Samuel Jay Keyser, eds., *The View from Building 20*, 1–52. Cambridge, Massachusetts: MIT Press.

Collinder, Björn. 1960. *Comparative Grammar of the Uralic Languages*. Stockholm: Almqvist & Wiksell.

Comrie, Bernard and Tania Kuteva. 2005. Relativization strategies. In Martin Haspelmath, Matthew S. Dryer, and Bernard Comrie (eds.), *The World Atlas of Language Structures*, 494–497. Oxford: Oxford University Press.

Derbyshire, Desmond C. 1979. *Hixkaryana*. (Lingua Descriptive Series, 1.) Amsterdam: North-Holland.

Deutscher, Guy. 2000. *Syntactic Change in Akkadian: The Evolution of Sentential Complementation*. Oxford: Oxford University Press.

Dixon, Robert M. W. 1972. *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Dixon, Robert M. W. 1981. Wargamay. In R. M. W. Dixon and Barry J. Blake, eds., *Handbook of Australian Languages, Volume 2*, 1–144. Canberra: Australian National University Press.

Ebbinghaus, Heinz-Dieter and Jorg Flum. 1999. *Finite Model Theory*. Berlin and New York: Springer, second ed.

Englebretson, Robert. 2003. *The Problem of Complementation in Colloquial Indonesian Conversation*. (Studies in Discourse and Grammar, 13.) Amsterdam: John Benjamins.

Epstein, Samuel and Norbert Hornstein. 2005. Letter on 'The Future of Language'. *Language* **81**, 3–6.

Evans, Nicholas. 1995. *A Grammar of Kayardild*. Berlin: Mouton de Gruyter.

Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* **46**, 621–646.

Everett, Daniel L. 2009. Pirahã culture and grammar: A response to some criticisms. *Language* **85**(2), 405–442.

Everett, Daniel L. and Edward Gibson. 2019. Review article on Amaral et al. (eds.), *Recursion Across Domains*. *Language* **95**(4), 777–790.

Fillmore, Charles W. 1972. On generativity. In P. Stanley Peters, ed., *Goals of Linguistic Theory*, 1–20. Englewood Cliffs, NJ: Prentice-Hall.

Fitch, W. Tecumseh; Marc D. Hauser; and Noam Chomsky. 2005. *Cognition* **97**, 179-210.

Fries, Charles C. 1952. *The Structure of English*. New York: Harcourt Brace.

Gil, David. 2009. How much grammar does it take to sail a boat? In Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*, 19–33. Oxford: Oxford University Press.

Givon, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.

Gold, E. Mark. 1967. Language identification in the limit. *Information and Control* **10**, 447–474.

Grädel, Erich; Phokion G. Kolaitis; Leonid Libkin; Maarten Marx; Joel Spencer; Moshe Y. Vardi; Yde Venema; and Scott Weinstein. 2007. *Finite Model Theory and Its Applications*. Berlin: Springer.

Hale, Kenneth. 1976. The adjoined relative clause in Australia. *Grammatical Categories in Australian Languages*, ed. by R. M. W. Dixon, 78–105. Canberra: Institute of Aboriginal Studies; Atlantic Highlands, NJ: Humanities Press.

Hauser, Mark; Noam Chomsky; and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* **298**, 1569–1579.

Haviland, John. 1979. Guugu Yimidhirr: Sketch grammar. *Handbook of Australian Languages, vol. I*, ed. by R. M. W. Dixon and Barry Blake, 26–180. Canberra: Australian National University Press.

Immerman, Neil. 1999. *Descriptive Complexity*. New York: Springer.

Ingraham, Alexander. 1903. *Swain School Lectures*. Chicago, IL: Open Court. Online at: https://archive.org/details/swainschoollectu00ingruoft/

Johnson, David E. and Paul M. Postal. 1980. *Arc Pair Grammar*. Princeton, NJ: Princeton University Press.

Kac, Michael B. 1978. *Corepresentation of Grammatical Structure*. London: Croom Helm.

Kac, Michael B. 1992. *Grammars and Grammaticality*. Amsterdam: John Benjamins.

Karlsson, Fred. 2010. Syntactic recursion and iteration. In Harry van der Hulst, ed., *Recursion and Human Language*, 43–67. Berlin: De Gruyter Mouton.

Kripke, Saul. 1982. *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Cambridge, MA: Harvard University Press.

Kwiatkowski, Tom; Luke Zettlemoyer; Sharon Goldwater; and Mark Steedman. 2010. Inducing probabilistic CCG grammars from logical form with higher-order unification. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1223–1233. Cambridge, MA: Association for Computational Linguistics.

Lasnik, Howard. 2000. Syntactic Structures *Revisited: Contemporary Lectures on Classic Transformational Theory*. Cambridge, MA: MIT Press.

Lasnik, Howard and Mamoru Saito. 1984. On the nature of proper government. *Linguistic Inquiry* **15**, 235–289.

Lau, Jey Han; Alexander Clark; and Shalom Lappin. 2016. Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science* **41**(5), 1202–1241. Online at: https://onlinelibrary.wiley.com/doi/full/10.1111/cogs.12414

Levinson, Stephen C. 2013. Recursion in pragmatics. *Language* **89**(1), 149–162.

Libkin, Leonid. 2004. *Elements of Finite Model Theory*. Berlin: Springer.

Manaster Ramer, Alexis. 1993. Towards transductive linguistics. In Karen Jensen, George E. Heidorn, and S. D. Richardson, eds., *Natural Language Processing: The PLNLP Approach*, 13–27. Dordrecht: Kluwer Academic.

McElvenny, James, ed. 2019. *Form and Formalism in Linguistics*. Berlin: Language Science Press.

Miller, George A. 1962. Some psychological studies of grammar. *American Psychologist* **17**(11), 748–762.

Montague, Richard. 1974. *Formal Philosophy*. Edited and with an introduction by Richmond H. Thomason. New Haven, CT: Yale University Press.

Moore, Denny [Dennis Albert]. 1984. *Syntax of the Language of the Gavião Indians of Rondônia, Brazil*. PhD dissertation, Department of Anthropology, Graduate School of the City University of New York. Available online at: http://etnolinguistica.wdfiles.com/local--files/tese%3Amoore-1984/moore_1984_gaviao.pdf

Moore, Denny. 1989. Gavião nominalizations as relative clause and sentential complement equivalents. *International Journal of American Linguistics* **55**(3), 309-325.

Neeleman, Ad. 2013. Comments on Pullum. *Mind and Language* **28** (4), 522–531.

Nevins, Andrew; David Pesetsky; and Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* **85**(2), 355–404.

Nordlinger, Rachel. 2006. Spearing the emu drinking: Subordination and the adjoined relative clause in Wambaya. Australian Journal of Linguistics **26**(1), 5-29.

Ogden, C. K. and I. A. Richards. 1923. *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. New York: Harcourt Brace. Page reference is to the third edition, 1930.

Postal, Paul M. 2010. *Edge-Based Clausal Syntax: A Study of (Mostly) Object Structure*. Cambridge, MA: MIT Press.

Pullum, Geoffrey K. and Barbara C. Scholz. 2001. On the distinction between model-theoretic and generative-enumerative syntactic frameworks. In *Logical Aspects of Computational Linguistics: 4th International Conference*, ed. by Philippe de Groote, Glyn Morrill, and Christian Retoré, 17–43. Berlin: Springer.

Pullum, Geoffrey K. (2013): The central question in comparative syntactic metatheory. *Mind and Language* **28** (4), 492–521.

W. V. O. Quine. 1951. Two dogmas of empiricism.' *The Philosophical Review* **60**, 20–43.

Rogers, James. 1994. *Studies in the Logic of Trees with Applications to Grammar Formalisms*. PhD dissertation, University of Delaware.

Schütze, Carson. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago, IL: University of Chicago Press.

Steedman, Mark. 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Tomasello, Michael. 2003. *Constructing a Language*. Cambridge, MA: Harvard University Press.