RULES OF GRAMMAR

Geoffrey K. Pullum

School of Philosophy, Psychology and Language Sciences
University of Edinburgh

1 'Rules' in the everyday sense

The past six decades—broadly, the period dominated by generative grammar—has seen hundreds of articles and books with titles suggesting that the notion 'rule' is a thoroughly familiar one in linguistics, and that linguists deal with rules all the time. Anyone familiar with post-1960 linguistics will recall seeing titles like these that mention rules:[1]

'On the notion 'rule of grammar' ' (1961)
Some Syntactic Rules in Mohawk (1962)
'Global rules' (1970)
'Evidence for transderivational rules' (1972)
'Evidence that indirect object movement is a structure-preserving rule' (1972)
'Verb agreement as a rule of English' (1972)
'Zero-output rules' (1973)
'The base rules for prepositional phrases' (1973)
Rule Ordering in Syntax (1974)
'Conditions on rules of grammar' (1976)
Rules and Representations (1980)
Extraction Rules in Icelandic (1980)
'A landing site theory of movement rules' (1982)
Words and Rules (1999)
Architectures, Rules, and Preferences (2008)

But strangely, the uses made of the term 'rule' in generative grammar virtually never seem to denote anything like rules in the everyday sense. Rules as we ordinarily conceive of them can be followed or ignored, i.e., respected or flouted. They can be either waived or enforced. This is at root because what they do is to define CORRECTNESS within the terms of some moral, social, legal, conventional, or other such system.

Rules differ crucially from scientific laws. It makes no sense to talk of following, respecting, or waiving the laws of physics. And rules also differ from descriptive generalizations: it makes no sense to talk of complying with a summary of the present state of the traffic on I-95, or waiving the anatomy of the duck-billed platypus, or violating a chemical analysis of a sample of canal water. Human beings can normally conform to rules or violate them, as they wish. It is imaginable that a

[1] I give dates in parentheses merely intended to locate these well-known works by decade; I will not add bibliographical details for each of them in the reference list. A bibliography of modern linguistic works talking about rules would take up much more space than I have here.

person might become so habituated to obeying a rule that they might find it next to impossible to behave in a way that flouts it, but that is not the usual state of affairs, and the compulsion experienced by such a victim of habit does not constitute or define the rule.

Rules as we are familiar with them in everyday life arise in numerous contexts:

- clubs and societies ('Gentlemen must wear ties and jackets in the Cosmopolitan Club members' dining room');
- rights and prohibitions in condominium and office buildings ('Smoking is not permitted anywhere in the building or the courtyard');
- games ('The bishop moves diagonally');
- competitions ('Employees of Kelloggs or its affiliated companies are not eligible');
- table etiquette ('The fork goes on the left of the place setting, the knife on the right');
- parliamentary procedure and order in meetings ('Each member has a vote and each vote is weighted equally'; 'A motion to rescind, repeal or annul or amend something previously adopted requires a two-thirds vote, a majority with previous notice, or a majority of the entire membership'; see Robert 2011);
- driving on public roads ('The speed limit is 25 mph in all residential areas'; 'Stop'; 'Right lane must turn right');
- public behavior ('No Parking'; 'Do not wedge this door open'; 'Wait behind yellow line until called');
- manuscript submission ('Abstracts must fit on one side of letter-size paper in 12-point type with one-inch margins');
- interpretion of legal statutes ('In the absence of a contrary indication, the singular includes the plural, and vice versa'; see Scalia and Garner 2012, 129–131');
- order of arithmetical operations ('Operations are applied working from the left in a formula with mixed division and multiplication, so '$6 \div 2 \times 3$' means '$(6 \div 2) \times 3$' = $3 \times 3 = 9$, not '$6 \div (2 \times 3)$' = $6 \div 6 = 1$').

And grammar, of course, is one of the most obvious domains in which rules are (or used to be) standardly invoked. Every traditional grammar and every prescriptive usage manual deals in rules. A few attested examples:

- 'Assumptive adjective-words precede their head-words' (Sweet 1898, p. 7)
- 'Modern English generally has the normal order subject + verb' (Sweet 1898, p. 13)
- 'Preposition-groups normally follow their head-words … But preposition-groups often take emphatic front-position … *of fuel they had plenty*' (Sweet 1898, p. 26)
- 'The number and person of a Finite verb depend upon the *number* and *person* of its Subject' [Nesfield 1900, §66 p. 55]
- 'Do not join independent sentences with a comma.' [Strunk And White, 2000, p. 5]
- 'Do not use small caps for roman numerals.' [*The Economist Style Guide* 2003, p. 7]
- 'It is a sound rule that *that* should be dispensed with whenever this can be done without loss of clarity or dignity.' [Gowers 2014, p. 216]
- 'When a proper name is used as an adjective, it isn't a possessive and therefore doesn't take an apostrophe. Hence "the Cubs [not *Cubs'*] game is at 1:00 today." [Garner 2022, p. 855]

I have deliberately chosen a diverse array of examples above to illustrate the point that rules expressed in English can be stated in a wide variety of grammatical forms: they may be declarative or imperative in clause type; they may be expressed in either active or passive clauses; they may use the simple present tense or contain modal auxiliaries; they may be in the second person, third person, or even first person ('We do not allow members to bring more than one guest'). But what links them all is the notion of providing guidance about what's right: their intent is to lay down a standard for what is correct or compliant and what is not. That is, they have NORMATIVE force.

2 'Normative' and 'prescriptive'

For my purposes here it is regrettable that there is a century-old practice in linguistics of equating the terms NORMATIVE and PRESCRIPTIVE. References to 'normative grammars' in sources from Bloomfield (1930) to Tieken-Boon van Ostade (2017) almost always mean prescriptive grammars of the authority-dispensing type, which are so often polluted by bossiness or snobbery.

The tradition may have started in the 19th or even the 18th century, but one early use of the term 'normative grammar' was in the work of Hanns Oertel (1901). He seems to regard 'normative' as synonymous with 'didactic,' and contrasts the didactic enterprise with the business of scientific linguistics. 'Normative or didactic grammar sets up a certain standard as correct,' he says (p. 87), and he proceeds from there to discuss the matter of defining a standard variety of a language that can be taught to those who speak non-standard dialects of it. He is clearly referencing 18th and 19th-century efforts to 'improve' the speech of non-standard dialect users, and contrasting normative grammars with scientific ones.

Here I reject the terminological conflation of 'normative' with either 'didactic' or 'prescriptive', because I need to lean on a fundamental terminological distinction between normative statements and prescriptive judgments or motivations. This is not such a radical revision of terminology: I am merely separating two documented senses for the word *normative*. *Webster's Third New International Dictionary* (1961, p. 1540) distinguishes four main senses, sense **1** being 'of, relating to, or dealing with norms' and sense **4b** being 'prescriptive.' I am merely limiting 'normative' solely to sense **1**, as a predicate of statements that defines them as being about what ought to be, not what is—the way analytic philosophers use it. I want to reserve the adjective 'prescriptive' for use only as a predicate of judgments or educational endeavors associated with the presupposition that some people's language is culpably bad—the business of trying to get people to improve their language use.

It seems to me that any grammar worthy of the name has to be normative: if a grammar for a human language does not define correctness for sentences, and define non-sentences as incorrect, it is not doing the job that we have traditionally expected grammars to do—like enabling foreign learners to figure out what they have to learn. Learners of a second language are interested not in learning what sorts of things happen in cases where native speakers make slips of the tongue, but in what it would be right for them to say when using the language themselves.

Notice that we do not get a definition of a language simply by saying that a sentence is grammatical in the language of a speaker A if and only if A reports it as such when successfully accessing his or her intuitions, and that those intuitions are as specified by a generative grammar that is somehow

inaccessibly inscribed in A's brain. The putative systems inscribed in A's mind, whatever they are, have some sort of electrochemical or neurophysiological reality. Claims about them constitute statements of plain empirical fact about the way the world is.

Chomsky has repeatedly asserted that facts about the generative grammars we supposedly have inscribed in our minds are biological facts. And biological realities, whether on a large or small scale, cannot be said to be right or wrong: things simply are the way they are. Neither tectonic plates nor brain states are subject to rules that can provide a basis for judging them to have done something wrong. If speaker A has a certain intuitive feeling when a given sequence of words is presented, that is just part of the way the world is. There is no basis for saying that A is correct—and by the same token, no basis for deeming A to be wrong.

It seems to me that the traditional view under which a grammar has normative force is compelling, and linguists (as distinct from psycholinguists, whose business is the study of what actually goes on in speakers' brains and behaviors) have made a mistake by forgetting or explicitly rejecting this during the past half century.

2.1 Prescriptivism and prejudice

What I have said so far might seem to be bringing grammar in the linguist's sense much closer to prescriptivism than most linguists would ever expect. When I began the study of linguistics as an undergraduate I was rapidly and easily convinced to dismiss prescriptivism as silly prejudiced nonsense, something we should pay no attention to at all, and after encountering the radically anti-prescriptivist rhetoric of works like *Leave Your Language Alone* (Hall 1950), I simply forgot about prescriptivism for at least thirty years of working on syntax and phonology. I was only brought back to considering it seriously by working on the reference grammar Huddleston and Pullum (2002) and, even more so, when I started to write posts about language for a general non-linguist readership on the group blog *Language Log* (see Liberman and Pullum 2006 for a selection of those posts).

One of the reasons linguists dismiss prescriptivism, and even find it distasteful, is that the prescriptive motivation is so often accompanied by DIALECT CHAUVINISM: the touting of one dialect as clearly (almost morally) better than another. Yet this is solely a contingent association: it is not a necessary part of wanting to help others write better that noting a difference between our own variety of English and a different one should immediately inspire us to despise or condemn the latter. Yet suppose someone states that the right way to get a hotdog without onions is to utter (i) not (ii), signaling that judgment in the standard notation of linguistics by putting an asterisk on the latter.

(i) I do not want any onions.

(ii) *I don't want no onions.

(Take the asterisk on (ii) as marking the following sequence of words as not being in the language, but note that, crucially, this is relative to the assumption that we are judging things from within the terms of Standard English, and the word sequence is supposed to indicate that the speaker does NOT want onions.)

Many English teachers have tried to teach children that even the use of the word *don't* is an error—meaning, of course, that they regard 'contractions' as improper in formal written English and want to inculcate in students the habit of not using them; but set that aside for now.

The relevant fact here is that there exist speakers (quite probably half a billion of them) who use (ii) to mean what (i) means in Standard English. When they use *don't* and *no* in the same clause, the two negative markers do not cancel each other out to yield an un-negated clause. Linguists refer to this sort of construction as negative concord, and it is a feature of most nonstandard varieties of English, whether in the British Isles or North America or Australasia.

To say that (ii) is incorrect in Standard English terms is certainly true, but the problem is that teachers who point that out are inevitably interpreted as suggesting not just that (i) is correct English and (ii) is not, but that people who use (i) are in some sense better than people who say (ii), and likewise would say *I never meant to hurt nobody* (for *I never meant to hurt anybody*), and *We ain't got none* (for *We don't have any*), and so on.

So theoretical linguists try to maintain a non-judgmental perspective and say that there are two dialects here, and that ultimately the task of linguistics would be to characterize them both accurately. We should not attribute ignorance or lack of education to the speakers whose native dialect exhibits a negative concord system and who therefore favor (ii). But it is very easy for a hint of social judgment to creep in, or to appear to have crept in.

Dialect chauvinism cannot be just set aside as something that sensible and socially liberal people like us do not have to worry about. The issues have to be faced even in the most aridly mathematical contexts. Surprisingly perhaps, there are dialects in arithmetical notation. An expression like `x^y^z` is not evaluated identically by all calculators and commercial computer software: `4^3^2` may produce 4,096 (in Microsoft Excel and in MATLAB), or 262,144 (in Google, Wolfram Alpha, or the Unix **bc** calculator), or an error message (the LaTeX system used to typeset this book treats `4^3^2` as an illegal expression when used in a mathematical formula).

But a different dialect is not necessarily a contemptible one. If you are using MATLAB, it definitely is WRONG to type `4^3^2` if your intent is to enter an expression that evaluates to 262,144 (you need to introduce some bracketing); but that does not license either the inference that MATLAB is better and more socially acceptable than the Unix **bc** utility, nor the conclusion that it is contemptible and people should be educated out of it.

The issue of dialect chauvinism can be separated off from prescriptivist practice: it may be common to find prescriptivsts taking a snobbish attitude toward non-standard dialects, but that is only a contingent fact about some prescriptivists, not a necessary attribute (I argue this point more fully in Pullum 2023).

2.2 Rule-following, internalized grammars, and culpability

The key worry expressed by Kripke (1982) about linguistics, and in particular the Chomskyan concept of competence, is that no substantive fact about a human being seems capable of giving rise to the normativity he assumes to be there in grammar.

It is unfortunate for linguists that throughout his monograph he sticks to a single rather

unilluminating example: the meaning rule for the term 'plus'. He never considers grammatical rules at all. However, in a couple of footnotes he expresses a tentative worry that the mentalistic interpretation of generative grammar could be in trouble.

He may be right, though deciding whether he is will be extraordinarily difficult: Kripke's book has been the subject of controversy in the philosophy of linguistics for more than forty years now. He challenges us to find an independently specifiable fact about a person that could possibly count as fixing a rule that the person can be said to be following. The property of having a generative grammar inscribed in one's brain does not differ in this respect from any other property. He is asking what specifiable fact F there could be about a specific hairless mammalian biped, possibly about some stable element of the constantly fluctuating electrochemical state of a grapefruit-sized brain, that could guarantee (in his example) the answers to all the future addition sums the person might do or consider, or could determine (in the reapplication I am considering) the well-formedness of some indefinitely large class of never-yet-uttered sentences, and the ill-formedness of the sentences in its complement. The fact F doesn't even have to be discoverable by the likes of us: that's an epistemological issue, and Kripke is concerned only with the metaphysical question. What sort of fact could it be?

Chomsky has never clarified how a stable element of brain states could even be interpreted as an inscription of some unique formal system: we know nothing about how to identify such inscriptions. Still less do we know how having one's brain in a state containing a formal system defining a certain set could actually play a role in the tasks of uttering and understanding sentences, or how having an inscription of some generative grammar in one's brain could impart any conscious knowledge of the properties common to the sentences of the language so that we could supply answers to questions about our native-speaker intuitions.

The basic problem is that Chomsky's 'mentalist' interpretation of the generativist view on grammar seems too lawlike. If inscribing a generative grammar in the brain of some primate causes it to utter and apprehend only sentences structured in a certain sort of way, then the primate is in effect biologically limited to the set of sentences that are structured in that way. Interpreting a sentence in a novel way, or understanding a sentence containing a previously unencountered word, would be impossible for such an organism.

The usual response to anything like this would involve the Chomskyan distinction between competence (the mentally determined definition of the well-formed sentences) and performance (the actual linguistic behavior of the language user). But as far as I can see, it merely posits one hypothesized cognitive module in combat with others.

A purely conjectural internally registered generative grammar is claimed to be less than faithfully subserved by the various ill-understood mechanisms that regulate our breathing, memory, attention, and other physiological functions. But positing these extra layers of bodily and mental activity does not appear to have any bearing on the issue of what can justifiably be called 'right' or 'wrong' in the language. Suppose someone were to utter or type this word sequence, thinking it was grammatical:

> *The Watergate conspiracy was a disaster from which Richard Nixon's second term in the White House was destined never to recover from.*

Suppose further that the generative grammar inscribed in my brain does not generate this string, but

nonetheless other psychological mechanisms in play cause me to apprehend the utterer's intended meaning (that Watergate was a disaster that the Nixon presidency never recovered from) without noticing the mistakenly doubled preposition. If all the brain-internal operations were fully understood, we could understand how it was that the utterer came out with the above utterance; but where can we account for the fact that the sequence of words is not grammatical, but is INCORRECT?

These are the worries that Kripke seems to be hinting at. And Chomsky certainly seems to see them as worries, because he mounts a lengthy challenge in response, devoting most of a long book chapter (Chomsky 1986) to Kripke's objections (I will not go into details here, but see Wright 1989 for a deep and interesting critical response to that chapter).

One point Chomsky makes is a very natural one, concerning the issue of obligations to act. To say that a grammatical constraint has normative force might seem to imply something about WHAT SOMEONE SHOULD DO (or should not do). That is, it might seem that if someone disobeys some condition that is part of a normative grammar, they have thereby done something that they shouldn't have done.

Chomsky sees this implication and objects to it. Characteristically perceptive, he gives exactly the right reason. For some imaginary English speaker Jones, he says (Chomsky 1986, p. 241):

> The rules of Jones's language ... entail nothing about what Jones ought to do (perhaps he should not observe the rules for one reason or another; they would still be his rules).'

This is exactly right: your 'rules'—the system of principles you respect regarding your idiolect—are the same whether you obey them on any given occasion or not. You might have good reasons for not obeying them (you're acting in a play, or you're trying to fool enemy agents by pretending not to be a native speaker, or you're imitating someone else's bad English in a comedy routine, or any number of other reasons).

But how can there be a NORMATIVE rule, defining what is right, without it therefore being the case that you ought to do what's right? Don't philosophers always take the normative notion of what's right to be intimately tied up with concepts such as moral duty, and more generally, what you ought to do?

Alan Millar (2004) elaborates at length on what is effectively this point. He draws the necessary distinction, between (i) defining what is correct and (ii) advising people about what they should do. When a practice is governed by a certain 'rule' (or as I would prefer to put it, when it respects a certain constraint), then 'Participating in the practice makes one subject to that [constraint]' (pp. 168–9); and it is true that if you are subject to it there is a certain sense in which you ought to obey it. But grammatical constraints are not laws of physics: they can be ignored or broken. Millar remarks that you may participate in a practice yet also flout the rules of constraints governing it.

He acknowledges that 'participating in a practice incurs a commitment to following its governing rules and therefore to doing what the rules prescribe': after all if you aim to be thought of as a participant in a practice and you don't respect its defining principles, you can hardly be taken seriously as a participant. But from the fact that a practice is rule-governed, according to Millar, 'It does not follow that one ought to follow the rules' (or respect the constraints, as I would put it). Why not? Why is that not a self-contradiction? Because 'it might be that one ought instead to WITHDRAW

FROM THE PRACTICE' (emphasis added).[2]

Applying this thinking to linguistic prescriptions (though that is not Millar's focus), consider the constraint that defines the so-called 'split infinitive.' Defining it with full rigour is less straightforward than people tend to think (because we need to leave room for possibilities like verb phrase ellipsis, parenthetical interruptions, etc.), but loosely we can say that the constraint forbids adjuncts from being linearly positioned between the infinitival marker *to* and the head verb of the infinitival complement that it introduces. The constraint draws a distinction between phrases that respect it (which are thereby defined as well-formed or 'good') and phrases that don't (which the constraint classifies as erroneous or 'bad'). Yet nothing follows from it about what anyone should do.

A syntactician might formulate the constraint against 'split infinitives' simply as a prerequisite for searching out sentences in literary works that violate it. That is exactly what George Curme (1930) did more than a century ago: his compendious work *Syntax* cites large numbers of attested literary examples (pp. 458–467, esp. pp. 461–465). Curme held the view that 'split infinitives' are natural, frequent, and useful in English writing, and have been attested over many centuries. In his view it is the people who urge avoidance who are making a mistake. But he can't even make that point without formulating the constraint.

You can always respect the constraint against 'split infinitives' if you want to. That might serve a useful purpose, such as ensuring that your writing is not criticized by those who (wrongly) think the construction is a grammatical error. But the other option you have is to simply withdraw from the practice of writing the kind of English that respects the constraint. Curme thought that was exactly what we hould do; he even asserts (p. 461) that the practice of occasionally violating the constraint 'is more characteristic of our most prominent authors'; split infinitives tend to be shunned more in the work of 'the minor writers, who avoid it as they fear criticism.'

Normative statements of constraints, then, do not intrinsically imply any recommendation about the use of any construction, or about using the relevant language at all. Linguists hypothesize sets of grammatical constraints in order to characterize phenomena accurately. If that is done well, people inclined toward supplying prescriptive advice may be better placed to consider how people should be advised to use the language; but stating the constraints is not to be equated with advising people to respect them. As Millikan might put it, the priciples governing Standard English don't tell you that you can't or shouldn't give up on writing Standard English, but merely define what would constitute continuing to use it.

They provide that definition because they are the governing principles FOR STANDARD ENGLISH, and the only place from which they can possibly draw their authority is from the practice of using Standard English, as exemplified in the speaking and writing of numerous native speakers one meets all the time.

This raises another problem about the normativity of rules which has received a lot of attention from philosophers. Alma Diamond (2024) sees it as intimately bound up with the philosophy of law, and asks: 'how can rules, operating as normative standards, derive from the very actions they are meant

[2]Millikan (2005, p. 148) independently makes a very similar remark: 'The rules of chess don't tell you that you can't quit, but only what would constitute going on.' It is a constraint on the movement of a king within chess that it can only move to an adjacent square, but that does not put you under an obligation to carry on playing chess as normally understood.

to regulate?'

The example appealed to throughout Diamond's paper involves social expectations, community-wide practices, neighbourly duties, explicit rule-setting, and disagreement about justification—important for some aspects of rule-following, but for the most part not even distantly relatable to syntax. Obviously linguistic rules have social roots in some sense: the point of learning American English is to be able to speak and write very much as it used in the social world of English-speaking Americans. But characterizing those social roots is extraordinarily difficult, and linguists who have mostly followed Chomsky's lead have been exposed to decades of training in ignoring the relevant issues. Much work will have to be done in coming decades to make any progress on the crucial question that Tyler Burge (1989) phrases as 'Wherein is language social?', but here I set it aside, because I am pursuing only the issue about the role and interpretation of formal grammatical systems.

3 Generative grammars are not (and cannot be) normative

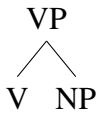
Consider the phrase structure-defining formulas that linguists write in the form 'X $\rightarrow$ Y Z' and typically call 'phrase structure rules'. I will call these PRODUCTIONS here (using the term that the logician Emil Post introduced in 1920–21 for a much more general class of formulas of which 'X $\rightarrow$ Y Z' is a special case—see the clear summary of Post's work by Martin Davis in the preface to Davis 1994, xiii–xv). Linguists have typically called such formulas 'rules', but this (as I have already suggested) is a highly misleading usage. A newcomer to linguistics might imagine that a rule of grammar says something about sentences in English are supposed to be like, but a production such as 'Sentence $\rightarrow$ NP VP' cannot be interpreted in any such normative terms.

- It does not say that every Sentence contains an NP and a VP, or even that every Sentence contains a VP, because the grammar might also contain a formula saying 'Sentence $\rightarrow$ Interjection' (to allow for utterances like *Wow!*).
- It does not say that when an NP and a VP are found in a Sentence they come in that order, because the grammar might also contain a formula saying 'Sentence $\rightarrow$ PP + VP + NP' (to allow for utterances like *After that came a plague of locusts*).
- It does not say that where an NP and a VP are found, they will make a Sentence, because the grammar might also contain a formula saying 'VP $\rightarrow$ V + NP + VP' (to allow for utterances like [$_{\text{VP}}$ [$_{\text{V}}$ *make*] [$_{\text{NP}}$ *the world*] [$_{\text{VP}}$ *go away*]]).
- It does not say that at least some Sentences will be found to contain an NP and a VP, because either NP or VP might be rewritten by other formulæ in the grammar, or completely deleted by rules of the type $\alpha \rightarrow \varepsilon$. That is, a production can erase a symbol completely.

Productions (whether context-free or context-sensitive) are actually **operations**, in the algebraic sense, like set union or numerical addition. A production (and for simplicity I'll give examples of the context-free type, 'X $\rightarrow$ Y Z') is formally equivalent to a subset of $V_N \times (V_N \cup V_T)*$ where V_N is the inventory of categories and V_T is the set of all words or morphemes or other formatives. Just as the operation notated '+' maps a pair of natural numbers i and j to the natural number k that comes j

steps after i in the order defined by the successor relation on the set $\mathbb{N}$ of natural numbers, so, analogously, a standard way of interpreting the phrase structure formula 'VP $\rightarrow$ V NP' is as an operation on $(V_N \cup V_T)*$ that maps any string 'X Sentence Y' to the string 'X NP VP Y' and thus defines one possible pair of adjacent lines in a derivation. A terminal string is defined as well formed if and only if there is a derivation for it in which every pair $\langle \varphi \mathrm{X} \psi, \varphi \mathrm{Y} \psi \rangle$ of adjacent lines (φ and ψ being strings over $(V_N \cup V_T)$) is licensed by a phrase structure operation saying 'X $\rightarrow$ Y.'

There are other interpretations of productions, as an important paper by McCawley (1968) pointed out, but they are not quite as different as some have supposed. One reinterpretation (which McCawley attributed to Richard Stanley) essentially equates 'VP $\rightarrow$ V NP' with the mini-tree that has root label VP and frontier 'V NP', depicted like this:



What McCawley suggests merely provides a different way of characterizing a set—a set of trees rather than strings. (McCawley pointed out that the algorithm for building parse trees from phrase structure derivations did not always lead to a unique result, unless certain rules about building derivations were carefully observed.) The set of trees defined by a collection of McCawley–Stanley productions is the set of all trees in which every local subtree exactly matches one of the mini-trees that make up the grammar.

One can imagine the definition process either in terms of randomly selecting local subtrees in a tree and verifying them (by finding mini-trees that license them) until all of the tree has been verified—this is basically the view that McCawley attributes to Stanley)—or in terms of randomly constructing trees by plugging mini-trees together with local subtrees (matching up frontier nodes to root nodes of mini-trees). Either way, what you have is simply a definition of a set of trees.

A formula 'A $\rightarrow$ B C', whether interpreted in Stanley/McCawley mode or not, does not say that the tree contains an A or a B or a C; it does not say that every A in a tree will have a B daughter; it does not say that every B C sequence will have an A parent. It does not say anything about trees at all.

Something very similar could be said about composition-oriented grammars like categorial grammars. The starting point is a lexicon of items with categories of the form R/G (think of G as mnemonic for 'given' and R for 'result'). An item belonging to the category R/G, if given something of the category G, yields a resultant object belonging to the category R. (In Mark Steedman's combinatory categorial grammar, constituent order is handled by having both R/G and $R\backslash G$ categories. An R/G item yields a resultant object of category R if combined with a G that follows it; an $R\backslash G$ item yields an R if combined with a G that precedes it; see Steedman 2019 for an elementary overview.) A sentence is well formed if and only if there exists some way of constructing it using the combinatory operations.[3]

[3]Chomskyan minimalism, with its binary 'Merge' operation, looks to me to like very casual categorial grammar that pays insufficient attention to the details of lexical items. This is not the place for a critique of the 'minimalist program', but see Steedman (in press) for a very sophisticated and detailed account of how minimalist ideas find their best realization in terms of combinatory categorial grammar.

'Transformational rules,' in Chomsky's original sense (now mostly of historical interest), likewise say nothing about sentences. Take a 1957-style transformation abbreviable thus:

X – *wh*-phrase – Y ⇒ *wh*-phrase – X – *t* – Y

It does not say that all *wh*-phrases will appear at left-hand ends of clauses, or even that they CAN appear at left-hand ends of clauses, because there could be another rule that moves them back again, or moves them to the end of the clause, or deletes them in all cases. It says nothing about what clauses will actually look like. It merely defines one operation that may be used in the course of constructing a derivation, which will terminate in the construction of a sentence if it succeeds (that is, if it doesn't dead-end in a situation where non-terminal symbols remain in the string but no further transformational operation can apply to it).

4 What generative grammars really are

The nature of the formal systems known as generative grammars has been much misunderstood by philosophers of linguistics and even by generative linguists themselves. A generative grammar – whether its components are phrase structure formulæ, transformations, composition operations, or whatever – is a formal system of a rather peculiar sort interpretable as a HOLISTIC, NONDETERMINISTIC, RANDOM, CONSTRUCTIVE procedure, intended as a set definition. Let me explain the four modifers I used in that sentence.

- A generative grammar is HOLISTIC because no element of it has any consequences independently of the whole; the entire grammar defines the whole set all at once (it is the smallest set containing everything that can be built using the formulæ), and no proper subpart of the grammar defines any element or subset of the set.
- A generative grammar is NONDETERMINISTIC because at any point in a derivation there may be a choice (or multiple choices) about what could be done next. (In fact usually there are large numbers of such choices.)
- A generative grammar operates in a RANDOM way in the sense that which string will be built on a given run through the operations cannot be foreseen, and some runs may not lead to any string at all. The strings in the set are not produced systematically in a fixed order, such as shortest first, or produced without repetitions. (Indeed, if an infinite set can be enumerated in a strict ascending order, it is decidable, hence has a recognition algorithm: see Janssen, Kok, and Meertens 1977, p. 115. Some sets that can be generated by transformational generative grammars, however, are not decidable.)
- A generative grammar is CONSTRUCTIVE in that its sole function is to supply a method for building algebraic objects to form a set – the set of all objects that could in principle be built from one of the initially given strings using the operations.

Since 1957 people have sometimes referred to generative grammars as algorithms, or have compared them with algorithms. They have nearly always been mistaken in this. They are not algorithms in any normal sense of that term, because they do not compute anything.

Chomsky's writings have not helped to stem misunderstandings about this; he has made many confusing and incorrect statements about generative grammars over the years, sometimes seeming to imply they are computable functions taking arguments, or algorithms for computing the values of such functions. Some examples:

- Chomsky (1959, p. 138) says: 'A grammar of L can be regarded as a function whose range is exactly L. Such devices have been called "sentence-generating grammars." ' But generative grammars (the rewriting systems that Emil Post developed in order to study logical proof, and called 'production systems') are not interpretable as functions at all. Functions take arguments and yield unique values; rewriting systems do nothing of the kind.
- In a footnote attached to the words 'sentence-generating grammars,' correctly crediting Post (1944) as the source for this use of the term 'generate,' Chomsky elaborates on his error. He says: 'We can consider a grammar of L to be a function mapping the integers onto L, order of enumeration being immaterial (and easily specifiable, in many ways) to this purely syntactic study' (p. 138, fn. 1). But since generative grammars are not functions, *a fortiori* they are not functions from integers to sentences. Nobody has ever suggested a variety of generative grammar that takes an integer as argument and gives a string in the language as value.
- He proceeds to say of such a function F from integers to sentences of L that 'The weakest condition that can significantly be placed on grammars is that F be included in the class of general, unrestricted Turing machines.' But Turing machines are not functions any more than grammars are. A Turing machine does include a transition table which specifies what to do next when in a given state and reading a given tape symbol, and that table is indeed a function, but it is not a function from integers to symbol strings. Rather, it maps a pair of a state and a current tape symbol to a symbol to be written and a direction in which to move next on the input. Although it would be possible to write a program for a universal Turing machine that would map integers to the sentences of a language (that is, taking an integer as input it would write out a specific sentence of English as output), that program would not be doing anything like what a generative grammar does.
- A quarter century later we find a strange passage in Chomsky (1986, p. 26) claiming that a generative grammar of English will assign an analysis to any input at all, even a Japanese sentence or the sound of a squeaky door hinge. This might seem to echo the 1959 idea that generative grammars are functions taking inputs, but it is not true: generative grammars of the various kinds Chomsky has proposed have never had the property of taking an input, so they cannot assign an analysis to anything. They simply define strings over a given finite vocabulary, possibly generating some kind of structure for it as well.

Pullum and Scholz (2001) discuss generative grammars under the heading 'generative-enumerative syntax,' but in light of what I have just said, the reference to enumeration might be called misleading: a generative grammar for a language cannot really be said to enumerate the sentences of a language, in the sense of placing them one by one on a growing list without repetitions, so that the first can be called sentence no. 1, the second no. 2, and for each $n > 0$ some unique sentence is number n. This is not what any sort of generative grammar does.

Trying to visualize a machine or algorithm that closely corresponds to what a generative grammar really does is a bit strange. In effect it starts with a blank output tape, writes a string on it (bounded

by left and right end markers), then moves on and writes another string, and so on, forever. In a sense this embodies a set of precise step-by-step instructions for doing something, which is what we ordinarily think of algorithms as doing,but it does not meet one key condition we normally assume for an algorithm: IT DOES NOT HALT.

Ryan Nefdt suggests an original conception of grammars that does not equate them with accepting or enumerating machine programs; rather, it relates them to COMPRESSION algorithms (Nefdt 2023, pp. 70–79). He is reminding us here of an important insight: by capturing generalizations, grammars permit information about syntactic distribution and morphological structure to be expressed vastly more compactly (infinitely more, if the language is infinite), and we a smaller grammar that covers the same set of strings is taken to be a better grammar. Kolmogorov complexity (the theory of 'minimum description length') is thus a fundamentally important branch of mathematics for formal linguistics if we want to understand how grammars could be evaluated as better or worse than others and what criteria might be relevant in learning them.

But Nefdt's suggestion that 'grammars can be viewed as compression algorithms' should not be taken to mean that grammars literally take languages as input and produce compressed representations of their structure as output. Nefdt (p. 71) quotes Millhouse (2021) as saying that 'C is a compressed representation of D' should be taken to mean that 'C is considerably smaller than D and that there is some practical method for generating C given D and *vice versa*.' But although it is right that D (the grammatical sentences) can be generated from C (the grammar), it is a very different matter to generate a grammar from a language, or from the unbounded set of intuitions a speaker of it has about sentences. It was in a sense the goal of the methods of analysis that American descriptive linguists (often known as the 'structuralists') were trying to develop in the decades before 1957, and success in any such project would be firstly a realization of the dream of a mechanical discovery procedure for grammars (always dismissed by Chomsky as too ambitious), and secondly, a gigantic advance in developmental psycholinguistics.

So we can accept Nefdt's suggestion that any explicit and complete grammar for a language must offer a compressed account of the structure shared by a typically vast array of linguistic expressions. But grammars are not algorithms, and (a fortiori) are not compression algorithms analogous to the **gzip** program for computer files. We have no algorithmic procedure for taking a set of sentences as input and producing a grammar as output.

There is an interesting question about whether the Large Language Models (LLMs) of recent 'artificial intelligence' (AI) research embody such procedures (and I am grateful to Brett Reynolds for bringing up this point). I cannot explore the topic exhaustively here, but I will point to a few specific issues. The use that has been most commonly made of LLMs is to create 'chatbots' like the infamous ChatGPT. These are programs which take an arbitrary-length string or 'prompt' as input and produce as a response another string. The prompt might be a question, for example, and the response would be intended to be understood by the interlocutor to be an answer to it. But such chatbots are nothing like either generative grammars or algorithms for building grammars. They might perhaps be said to have grammars implicitly present in their structure, but no one has been clear about how such a grammar might be extracted and examined.

The 'training' process for an LLM has some similarities to a compression algorithm. The process involves taking the content of a corpus of training data (usually of truly vast proportions — gigabytes of text) and building from it a gigantic database of statistical facts that can be used to set parameters

for a chatbot implementation to use. In the chatbot implementation, features of the prompt are used to set a huge number of parameters that connect it to various aspects of the training data. The chatbot then responds with a string in which the character-to-character transition probabilities are controlled by these parameters.

The key thing that has lent LLMs importance in the contemporary linguistic sciences is that the responses of chatbot implementations are delivered in flawless Standard English. LLMs hardly ever violate the constraints of English spelling, punctuation, morphology, or syntax. And Piantadosi (2024) has made the very important point that this is a devastating blow to the claims Chomsky has made famous under the slogan 'poverty of the stimulus.'

It has been repeatedly asserted by Chomsky and his followers that learning English from mere exposure to a random assortment of English sentences would be not just difficult but actually impossible: the input that a human infant receives in terms of utterance of well-formed sentences just does not provide enough information. Negative information, giving clues as to what is not syntactically allowed, would be needed as well. LLMs are trained solely on positive data — strings that are in the training corpus. They are not warned that infinitely many other strings are ruled out by the grammar and must not be used in responses. Piantadosi argues that this leaves no avenue of escape for Chomskyan assertions about the impossibility of language acquisition from text. LLM training, and the grammatical flawlessness of LLM chatbot responses, seems to demonstrate that conclusively.

Returning to our main theme, however, it would be wrong to think that a generative grammar (or any kind of grammar) can be seen as a procedure for compressing a corpus by extracting its regularities and expressing them more compactly, but it is true that grammars in effect supply such a compact representation. What LLMs do is much less clear; settling the question would need a great deal more discussion than there is room for here.

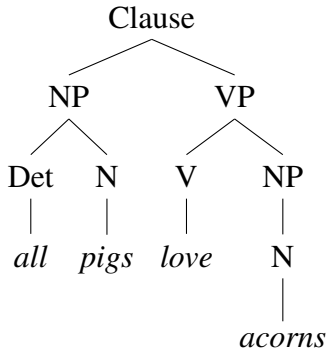
5 A model-theoretic alternative

One way in which grammatical statements can be formalized makes it much more plausible to see them as normative—like rules in the everyday sense. A grammar can be expressed as a set of independent constraints on linguistic structures, rather than a collection of operations forming part of a generative procedure for building the entire (potentially infinite) set of linguistic structures.

What I mean by a constraint from this point on is simply a statement that can be evaluated as true or not true within a certain kind of abstract object. (The mathematical notion of a graph should be general enough to capture the class of suitable objects.) A structured object will count as well-formed according to a constraint if it evaluates as true when evaluated in the structure of that object.

If a grammar is merely a finite set of constraints in this sense – statements about the structure of phrases, clauses, and sentences – it will be much more similar to a grammar in a sense that a traditional grammarian would recognize. And there is a kind of mathematical machinery tailor-made for theorizing about the relation of constraints to objects: the machinery of model theory.

For simplicity and concreteness, assume that phrases, clauses, and sentences have structures represented adequately by the phrase structure trees that linguists use. A sentence like *All pigs love acorns* might be diagrammed like this:



As a graph, this is directed, ordered, singly-rooted, and acyclic, with nodes (vertices) labeled by either one of the six category names (Clause, Det, N, NP, V, VP) or by one of the four word forms (*acorns*, *all*, *love*, *pigs*). The edges in the graph are indicated by the lines, their direction is represented by the top-to-bottomw dimension, and the ordering is shown informally by horizontal positioning on the page. For a logician, such a tree can be seen as a relational structure: a set of 12 elements (the nodes or vertices), with ten unary predicates (the category names Clause, NP, etc.) and two binary relations: **parent of** (the lines or edges) and **precedes**. Constraints on such tree representations for Standard English might include statements that could be expressed very informally like this:

(C1) If a node is parent of both an NP node and a VP node, the NP node precedes.

(C2) If a node is parent of both a V node and an NP node, the V node precedes.

These are particular facts about the structure of English sentences; in Malagasy, the opposite of (C1) holds, and in Turkish the opposite of (C2) holds. There might also be constraints that hold universally, like these:

(C3) Every VP node is parent of a node labeled V.

(C4) Every NP node is parent of a node labeled N.

We could add a statement that the Det node has to precede the N node that has the same parent, and that if the word form under the Det is *all* then word form under the N node with the same parent must be plural. We could connect the fact that *love* appears in its plural form (not in its 3rd-person singular form as *loves*) to the fact that *pigs* is in its plural form (the traditional way of stating this is that the verb agrees in number with its subject noun).

Such statements can be made increasingly general as the range of sentence types is expanded, and they can be formulated with as much precision as desired, using a description language with a model-theoretic semantics, such as first-order logic, or monadic second-order logic. A great deal is

known mathematically about the expressive power of systems of constraints on structures. There is a mathematical literature devoted to proving theorems about them: it falls within a subdiscipline known as finite model theory. Different types of generative grammar are provably equivalent in certain respects to properties of trees defined using description languages of different expressive power. Libkin (2004) provides a general introduction to finite model theory; Rogers (1998) develops a sophisticated application of the theory to linguistics, introducing the term 'model-theoretic syntax.'

Pullum (2013) argues that stating grammars as sets of constraints has genuine conceptual advantages over stating them as holistic nondeterministic random sentence construction systems. For example, under model-theoretic as opposed to generative assumptions, we are not forced to assume a sharp boolean distinction between sentences that are generated and nonlinguistic material that is not. Instead you a finely graded theory of degrees of ungrammaticality becomes available (because more constraints or fewer may be violated, and at more or fewer points in the structure).

Intuitively, this makes it somewhat easier to see how the halting utterances of a toddler or a foreign learner, or a speech blunder by a native speaker, can in practice be understood: although *Is our children learning?* is not grammatical, it comes close enough that we can see what question the blunder-prone President George W. Bush was intending to ask. The sentence he uttered satisfies all the constraints of English other than agreement of the verb with the subject noun.

The acquisition problem also becomes a little easier to conceptualize in a plausible way. Generativist views of first language acquisition have tended to hypothesize that human infants have a mysterious innately-driven and species-limited ability to rapidly identify a generative grammar that generates the entire language perfectly, after exposure to a relatively small number of its shorter sentences. (Notice, grasping only part of a generative grammar would be utterly useless.) But if we take the model-theoretic view, acquisition can be envisioned as a gradual process (perhaps lifelong) of coming to grasp constraints on expressions that make the childs own utterances, and interpretations of other speakers' utterances, more similar in structure to those of parents or peer groups.

We can also avoid the intuitively false prediction made by a generative grammar to the effect that a nonsense sentence like *The gostak distims the doshes* will not be recognizable by an English speaker as linguistic material at all: if the lexicon is treated as a set of constraints on phonological or orthographical material requiring certain forms to be associated with certain grammatical and semantic properties, then no constraint is violated by *gostak* or *dosh* being treated as nouns (simply because there is no constraint saying that they aren't), or by *distim* being treated as a regular verb. The sentence can be depicted, correctly, as the active counterpart of *The doshes are distimmed by the gostak*, and so on.[4] It is true that the addressee of a sentence like *The gostak distims the doshes* will be puzzled, but only about what gostaks and doshes are, and what distimming involves—not about whether the sentence says something in English.

Finally, what I have called syntactic quandaries—cases of apparent clashes of grammatical requirements where no possibility comes out fully well-formed—are predicted to be possible under the model-theoretic view, whereas generative grammars make them impossible to represent (since any given sentence is either generated, meaning that it is perfectly grammatical, or not generated, in which case it has no linguistic properties at all). And the fact is that quandaries do arise in human

[4]This important point was, in effect, made 120 years ago in Ingraham (1903, p. 154), which was quoted at length by Ogden and Richards (1923, p. 46). Essentially the same insight, together with its relevance to the model-theoretic view of syntax, was independently recognized much later in Chapter 14 of Johnson and Postal 1980.

languages. A simple example in English is noted by Fillmore (1972). He points out that the choice between alternatives like these seems vexingly impossible to make:

??*Either I or my brother am responsible.*

??*Either I or my brother is responsible.*

??*Either I or my brother are responsible.*

The problem is that *am* doesn't go with *my brother*; *is* doesn't go with *I*; and *are* doesn't go with either. Intuitively, a constraint on the verb paradigm is demanding a correct choice between *am*, *is*, and *are* according to person and number, but the disjunctive singular subject NP is non-committal as to the choice between 1st person and 3rd person. Fully satisfying all of the constraints is not possible, and that forces a quandary. A generative grammar for English is in the unfortunate position of having to get off the fence: it must generate either three, two, one, or none of them. But none of those four ways of setting up the grammar would yield a theoretically-based recognition that there is a grammatical quandary here; a generative grammar cannot provide it. A model-theoretic account can. Informally again, the grammar will say that (i) *am* must have a 1st-person singular subject NP, (ii) *is* must have a 3rd-person singular subject NP, (iii) *are* must have a subject NP that is neither 1st-person singular nor 3rd-person singular, (iv) *I* is 1st-person singular, and (v) *my brother* is 3rd-person singular—leaving the person specification of a disjunctive NP like *either I or my brother* undefined. Each of the three sentences will satisfy nearly all the constraints of English, but not quite all.

6 Conclusion

Any philosopher of linguistics will encounter the notion of a 'grammatical rule' repeatedly, and it would be easy to make the mistake of imagining that it is an unproblematic notion that we understand from everyday life, one that the scientific work done in the field known as generative grammar has tried to explicate over the past seven decades. But that is not so.

There is an ordinary, everyday sense of the term 'rule' under which rules can be obeyed, disregarded, violated, waived, enforced, etc., and their whole point is to define certain things as correct or permitted and others as incorrect, inappropriate, or forbidden; such rules have NORMATIVE force. But linguists abandoned the ordinary sense of 'rule' after 1957, and started using the term to denote such abstract entities as phrase structure formulas (A $\rightarrow$ B C) and transformations (X $\Rightarrow$ Y). Such expressions do not have normative interpretation at all.

Yet grammars must have normative force, or they are not serving their traditional purpose of specifying what is correctly phrased and what is not in the language. They could be of no possible use to learners of the language if they did not specify what differentiates correct usage from incorrect or mistaken usage.

Normativity (in my terms) is not at all the same thing as prescriptivism. And the fact that some grammatical statement is interpreted as normative does not in any way imply that if you violate it you have done something you ought not to have done: Alan Millar (2004) carefully explicates that point.

The nature of the formal systems known as generative grammars has been much misunderstood by philosophers of linguistics and even by generative linguists themselves: many claims about them are widely believed but are not true.

I have very briefly pointed out in section 5 that there is a coherent alternative to generative grammar, formalizable using the mathematical machinery of model theory, under which the statements making up a grammar are simply assertions about correctness requirements for linguistic structure. That alternative is associated with various conceptual advantages that generative grammars do not enjoy, in addition to permitting grammars to be seen as having normative force.

References

Anonymous. 2003. *The Economist Style Guide*, 8th edition. London: Profile Books.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Burge, Tyler. 1989. Wherein is language social? In George (ed.) 1989, 175–191.

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

Chomsky, Noam. 1986. *Knowledge of Language: Its Origins, Nature, and Use*. New York: Praeger.

Curme, George O. 1930. *Syntax*. (Vol. III of *A Grammar of the English Language*.) Boston: D.C. Heath.

Davis, Martin (ed). 1965. *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. New York, NY: Raven Press.

Davis, Martin (ed). 1994. *Solvability, Provability, Definability: The Collected Works of Emil L. Post*. Boston, MA: Birkhäuser.

Diamond, Alma. 2024. Moves & rules: Addressing the puzzle of social rule-following. *Law and Philosophy* 43(5), 549–589.

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

Fowler, Henry & Frank Fowler. 1906. *The King's English*. Oxford, UK: Clarendon Press.

Garner, Bryan A. 2022. *Garner's Modern English Usage* (5th ed.). Oxford, UK: Oxford Oxford: University Press.

George, Alexander (ed). 1989. *Reflections on Chomsky*. Oxford, UK: Basil Blackwell.

Gowers, Rebecca. 2014. *Plain Words: A Guide to the Use of English*, by Sir Ernest Gowers, revised and updated. London: Particular Books.

Hall, Robert A., Jr. 1950. *Leave Your Language Alone!* Ithaca, NY: Linguistica.

Huddleston, Rodney & Pullum, Geoffrey K. et al. 2002. *The Cambridge Grammar of the English Language*. Cambridge, UK: Cambridge University Press.

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

Janssen, Theo, Gerard Kok, and Lambert Meertens. 1977. On restrictions on transformational grammars reducing the generative power. *Linguistics and Philosophy* 1(1), 111–118.

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

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

Liberman, Mark and Geoffrey K. Pullum. 2006. *Far From the Madding Gerund and Other Dispatches from Language Log*. Wilsonville, OR: William, James & Co.

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

McCawley, James D. 1968. Concerning the base component of a transformational grammar. *Foundations of Language* 4, 243–269. Reprinted in *Grammar and Meaning*, 35–58 (New York: Academic Press; Tokyo: Taishukan; 1973).

Millar, Alan. 2004. *Understanding People: Normativity and Rationalizing Explanation*. Oxford, UK: Oxford University Press.

Millikan, Ruth Garrett. 2005. *Language: A Biological Model*. Oxford: Clarendon Press.

Millhouse, Tyler. 2021. Compressibility and the reality of patterns. *Philosophy of Science* 88, 22–43.

Nefdt, Ryan M. 2023. *Language, Science, and Structure: A Journey into the Philosophy of Linguistics*. New York, NY: Oxford University Press.

Nesfield, J. C. 1900. *Outline of English Grammar in Five Parts*. London: Macmillan.

Oertel, Hanns. 1901. *Lectures on the Study of Language*. Scribner. Facsimile archived online at https://archive.org/details/lecturesonstudy01oertgoog/page/n4/mode/2up

Ogden, C. K. and I. A. Richards. 1923. *The Meaning of Meaning*. London: Routledge & Kegan Paul.

Piantadosi, Steven. 2024. Modern language models refute Chomsky's approach to language. In Edward Gibson and Moshe Poliak, eds., *From Fieldwork to Linguistic Theory: A Tribute to Dan Everett*, 353–414. Berlin: Language Science Press.

Post, Emil L. 1944. Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society* 50(5), 284–316. Reproduced in Davis 1965, 305–337 and in Davis 1994b, 461–494.

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

Pullum, Geoffrey K. 2023. Why grammars have to be normative—and prescriptivists have to be scientific. Joan Beal, Morana Lucač, and Robin Straaijer (eds.), *The Routledge Handbook of Linguistic Prescriptivism*, 177–196 Abingdon, UK: Routledge.

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* (Lecture Notes in Artificial Intelligence, 2099), ed. by Philippe de Groote, Glyn Morrill, and Christian Retoré, 17–43. Berlin: Springer Verlag.

Robert, Henry M. et al. 2011. *Robert's Rules of Order Newly Revised*, 11th ed. Philadelphia, PA: Da Capo Press.

Scalia, Antonin and Bryan A. Garner. 2012. *Reading Law: The Interpretation of Legal Texts*. St. Paul, MN: Thomson/West.

Steedman, Mark. 2019. Combinatory categorial grammar. András Kerteész, Edith Moravcsik, and Csilla Rákosi (eds.), *Current Approaches to Syntax—A Comparative Handbook*, 389-420. Berlin: De Gruyter Mouton.

Steedman, Mark. 2025. On internal merge. *Linguistic Inquiry* 56 (in press).

Strunk, W. & White, E. B. 2000. *The Elements of Style by William Strunk, Jr., With Revisions, an Introduction, and a Chapter on Writing by E. B. White* (4th edition). Needham Heights, MA: Allyn and Bacon.

Sweet, Henry. 1898. *A New English Grammar, Logical and Historical;, Part II—Syntax*. Oxford, UK: Clarendon Press.

Tieken-Boon van Ostade, Ingrid, ed. 2017. *English Usage Guides: History, Advice, Attitudes*. Oxford, UK: Oxford University Press.

Wright, Crispin. 1989. Wittgenstein's rule-following considerations and the Central Project of theoretical linguistics. In George (ed.) 1989, 233–264.