

1 Grammars, Old and New

*When I pronounce the word Future,
the first syllable already belongs to the past.*

—Wisława Szymborska

The capacity to be puzzled about what looks obvious is a good property to cultivate.

—Noam Chomsky, comment during panel discussion at MIT, 2011

The Beginning of Grammars

In the beginning of grammars, there was Pāṇini. The first linguistic hero, he wrote a grammar of Sanskrit that has been celebrated as the earliest ever. Pāṇini may have belonged to a first generation of Brahmins who could write,¹ and observations on the difference between spoken and written Sanskrit led him to build a system of grammar to describe these differences. The resulting *Aṣṭādhyāyī* (“Eight Chapters,” written ~400–300 BCE) proposed a complete, concise, and consistent analysis of grammatical structure, through thousands of rules and a large inventory of phonological elements and verbal roots, as well as long lists of “exceptions.” Coming from India, it is the earliest known attempt to introduce precision, systematicity, consistency, and conciseness into the study of human language.²

Everyone thinks about language, and everyone has an opinion about its nature. But what is grammar, and what, if anything, does it tell us about ourselves? What exactly did Pāṇini, and a multitude of grammarians and linguists who followed, write? Pāṇini was struck by the richness and regularity of the sounds and sound combinations in Sanskrit, and he constructed a system of rules to capture them. From then on, grammarians

have held that a grammar reflects relations between elements of language (speech sounds, syllables, words, phrases, sentences) as determined by its native speakers (who master the grammatical intricacies of their language without having ever been explicitly taught its grammar).

But what does a rule look like? Let's consider a simple English example from the formation of plural (PL.) nouns: *dog* + PL. → *dogz*; *cat* + PL. → *cats*; *dress* + PL. → *dressez* (or *dresseez* if you're British). These rules are simple indications of how to combine a word (*dog*) with an element of meaning (PL.) and map them ("→") onto another word (*dogz*). But writing a plural rule for each word would not get us very far, as Pāṇini understood. Instead, we should try to understand the generalization behind these specific mappings, and construct a rule that would indicate what is permitted and what isn't. The English plural rule harbors a principle that most readers know: assume that the plural -s is added to a word, but that it can be realized in different ways, depending on the features of the consonant that precedes it. If its pronunciation requires the immediate vibration of the vocal cords (known as "voicing," as in *b*, *d*, *g*, *v*), then the immediately subsequent plural -s inherits this property (hence *dogz*); if the word ends with an "unvoiced" consonant (like *p*, *t*, *k*, *f*, in which vocal cord vibration is delayed), then the plural -s remains as is (*cats*); and if the final consonant of the word is -s or -z, then pluralization would be -ez (*dressez*; alternately -eez, hence *dresseez*). Once formulated thus, we can inquire whether this rule fits into a general mold of English pronunciation: is it in fact related to plurals at all, or does it instead characterize a more general proximity effect, involving two consonants that belong to distinct pieces of meaning, regardless of pluralization? It is clear that the underlying principle has little to do with plurals; rather, the pronunciation of plurals follows a more general rule, one that ties the proximity of an added piece of meaning to the consonant that precedes it, not only in nouns. This is why we say *he hop-s* and *pop-s* but *she rob-z* and *cuss-ez*, which are pronounced in keeping with the same rule, despite their being verbs. The result is more general than the previous one, but it still doesn't cover everything, for what about *he pay-z* or *she fall-z*, in which a final -s is not preceded by a voiced consonant? Are these exceptions, or did we misstate the rules? And maybe we got the rule backwards and its default is -z, not -s? That is, perhaps *dogz* → *dog* + PL. is a better description of this particular instantiation of the pluralization rule? These are serious questions that linguists grapple with, attempting to come up with reasons

to prefer one approach to others, thinking that the best (i.e., most general and precise) formulation is the one that native English speakers tacitly hold in their heads. And we haven't even gotten to question the rule's validity across languages, or to wonder about general laws of pronunciation and their cognitive and neurophysiological bases.

There is much to talk about, but let's turn to what grammatical rules imply. In a way, they define the limits of language. A native English speaker can say *he hop-s* and *pop-s* and *she rob-z* and *cuss-ez*, but not **he hop-z* and *pop-ez* or **she rob-s* and *cuss-s*. The same holds for the possessive *-s*: we say *the dog's* and *the cat's tail* but not **dog's*, **cat's*. The latter are plainly not within the boundaries of English, as judged by its native speakers (hence we mark them with the “*” symbol). The final *-s* rule, then (and grammar in general), codifies regularities in language and draws the line between possible linguistic objects and impossible ones (**dog-s*, **hop-z*, **dog's*). And as we have seen, what began as a rule for plural nouns can be stated at different levels of abstraction.

Let me pause here to note that we are not looking at picayune little things of no consequence. The English plural/third-person/possessive *-s* is small, but like particles in physics, small may be quite important. Indeed, if the *-s* rule is marginal, how come its effects are so far-reaching? Every native speaker of English abides by it: nobody ever utters any of the starred words above. This important point—that grammar marks the boundaries of language and excludes incorrectly structured words we would never utter—is not limited to words; bigger linguistic units also abide by grammatical rules. Aristotle, who lived around Pāṇini's time, famously noted this in his *On Interpretation*, in which he presented a system of logic. While reasoning, not grammar, was his focus, he analyzed sentences into what he thought were their component parts. Indeed, *On Interpretation* is sprinkled with allusions to familiar sentence level concepts, such as *subject* (to Aristotle, the main player in a sentence, the qualities of which the rest of the sentence characterizes), *predicate* (the activity or state that characterizes the subject; usually a verb), *adjective* (a descriptor of a subject or object's qualities, such as *happy*, *tall*, *smart*), and many other terms that indicate the author's interest in grammatical structure (arguably, such an interest is already evident in Plato's *Cratylus*). Grammar and language structure, then, were prominent concepts utilized by thinkers in the ancient world. And we have clear evidence for continued interest in grammar during the Middle Ages, in the

form of French, Arabic, Hebrew, and Latin grammars,³ and the list goes on and on.

To us, the real story begins in the twentieth century, with the birth of modern linguistics, especially in the United States. There, Leonard Bloomfield developed a method to analyze sentences into their *immediate constituents*, a system in which a sentence like *The big dog chased the small cat* is divided into words, then phrases—pairs of “immediately” adjacent words that “go best together”—then larger phrases made out of pairings of phrases with their neighbors, and so on up to the top constituent: the sentence. This is shown in figure 1.1 (where each pairing is marked by two lines that meet).⁴

Constituency implies a hierarchy in relations among elements. The existence of intermediate branches, or nodes, indicates which parts are more closely related to which—in this sentence, the words *small* and *cat* form a constituent (*[small cat]*), and the word *the* forms a constituent with it (*[[the][small cat]]*). Thus, *the* is higher than each of the words *small* and *cat* (in a sense that must be made precise), as it combines with the constituent they form together. This comports with our intuition: what is modified by *the* is neither *small* nor *cat* but their combination. Our linguistic intuition also seems to suggest that *[chased the]* is an arbitrary grouping of words; and indeed, the diagram in the figure (as well as the equivalent bracketing in the text) indicates that this pair of words is not a constituent, because there is no node that contains all these words, and only them. This conclusion also fits with the results of certain tests: for example, you can assert, emphatically, that *it is the small cat that the big dog chased*, but English would not tolerate **it is the cat that the big dog chased small*, which is a plain monster. That is because *the small cat* is a constituent and must appear in one piece.

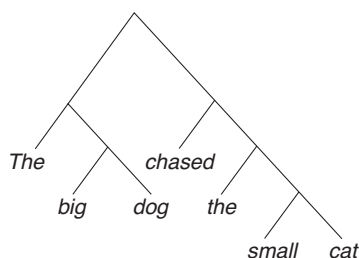


Figure 1.1

A sentence parsed into its immediate constituents.

This system has some transparency, enabling its proponents to justify every one of their moves. For instance, how exactly is the diagram in figure 1 drawn? Why is the division binary, not otherwise? What is the significance of a hierarchy? Linguists have been deeply concerned with these questions. While Bloomfield's insight was that this hierarchy is important for our understanding of how language works, Zellig Harris constructed a formal system of rules that built sentences out of immediate constituents and thus captured many linguistic regularities. Importantly, Harris's formal system captured phenomena but offered no constraints (a critical issue, on which I will elaborate).⁵ The linguistic scene was now ready, as it were, for the main language event of the mid-twentieth century: the appearance of generative grammar, proposed by Noam Chomsky.

The Zeitgeist in the Language Sciences, Early 1950s

We are all captives of the intellectual landscape into which we were born and raised. During the early 1950s, the community of those engaged in the study of language and related behaviors was divided into four prongs: the *structural*, the *statistical*, the *behavioral*, and the *cultural*.

Structuralists, such as Bloomfield and Harris, tried to build a hierarchical grammatical system that would distinguish well-formed words and utterances from monsters and would characterize them precisely, mechanically, and reliably. At the word and sentence level, Harris introduced mathematical methods into the analysis of hierarchical relations in sentence level grammars, but no constraints on grammatical systems were formulated—the linguist could fit rules to a language and build its inventory of “trees” with no limits. As a result, there was no understanding of what a possible syntactic rule of a human language is or of why languages differ from one another.

Parallel developments took place in phonetics. Speech scientists, who investigated the sound shape of language and speech, introduced sophisticated engineering tools for creating, analyzing, and testing different speech sounds. They did so in the hope of identifying the basic building blocks of speech perception and production, and they, too, limited themselves to the characterization of a language's inventory of sounds, which stopped short of being a generalized theory that holds of all human speech.

Behaviorists, led by psychologist B. F. Skinner (and supported by prominent philosophers such as W. V. O. Quine), sought to study language behavior

but explicitly barred themselves from a scientific inquiry into the cognitive mechanisms that underlie this behavior. Alluding to limitations imposed on “sound” scientific inquiry, Skinner argued that scientists cannot posit abstract cognitive mechanisms as “inner causes” of behavior, because these are not directly observable; therefore, scientists can only make connections between observed environmental factors that the organism experiences (a.k.a. stimulus) and the behavior it displays subsequent to this experience (a.k.a. response).⁶ The rejection of grammatical concepts was quite natural, as they are not directly observable, and hence are devoid of scientific value, according to this highly influential stream in psychology. Grammar, Skinner claimed at the heyday of behaviorist domination of American psychology, should be barred from use in scientific inquiry.⁷

Influenced by behaviorism, students of language turned to human behavior in the language domain, which they sought to study *statistically*, in a manner that excluded grammar. They took a slightly different path, though: attempting to develop tools for quantitative investigations of linguistic behavior, they turned to the recently invented information theory. This approach offered mathematical tools for the measurement of informational content in transmitted messages, quantitatively expressed in terms of the amount of uncertainty that is removed when a new piece of information becomes input.⁸

When applied to human language, this approach was only concerned with the quantification of the degree of uncertainty at a particular serial position in an incoming string. Psychologist George Miller analyzed human language behavior with these tools.⁹ He reasoned that when we listen to a sentence, we typically try to predict the next word from the already-heard context. For example, *After pouring milk into her empty bowl at breakfast, Jeannie realized that an ingredient was missing and ran to the convenience store to get _____*. Given the preceding context, I believe that most Americans would vote for *cereal* (or perhaps *muesli* if your heritage is German or Swiss) as the most likely candidate next word (some might instead say *her*, thinking that the continuation would be *her morning cereal*). Given this expectation, the reduction of uncertainty would be small upon encountering the word *cereal*. If instead the next word were surprising (e.g., *poison*), the reduction of uncertainty would be great. Using this logic, Miller investigated how well people predict the next word, and he tried to model his results with information-theoretic tools. As no assumption of unobservable

“inner causes” was required (especially grammar, hierarchy, parts of speech, or grammatical relations among parts of a string), Miller’s approach gained popularity among behaviorists.

Finally, there were the *culturalists*, whose interests contrasted sharply with those of the other groups. The culturalists viewed language as a window into culture, and they were therefore focused on differences between languages and their connection to cultural variation. Best known among these linguists are Edward Sapir and Benjamin Lee Whorf,¹⁰ who embarked on expeditions to small Indigenous communities whose members spoke languages like Inuit or Hopi, in the hope that documenting these languages would enhance our understanding of the relation between language and culture. Thus, the absence of past tense inflection in Hopi indicated to Whorf that the concept of time in Hopi culture was different from the European concept, a difference he tried to investigate.

However, this (mostly anthropological) linguistic prong did not apply a uniform system of grammatical description. It moreover remained focused on differences between languages at the single word level and on how these differences could serve as windows into the particular cultures in which these languages were used. In contrast to the previous approaches, culturalism did not look at human behavior at the individual level but rather took a broad look at the language–culture relation. Its connection to the present discussion is therefore rather tenuous.

Birth of Generative Grammar

Against this backdrop, Chomsky developed a new method of linguistic analysis, which mostly focused on syntax: the analysis of sentences into their constituent parts. Criticizing the above approaches, Chomsky realized that the simple principles can at times be overly simplistic.¹¹ Take linear proximity, for example: When you ask a question like *Can birds that fly run away?*, a *yes* answer is appropriate if they can run away, but not if they can fly or are flying. Why does the verb *can* relate not to its closest ally, *fly*, but rather to a distant one, *run away*? Chomsky showed how his work—but not others’—offered an answer (that the relevant notion of distance is not linear but structural: it takes hierarchical relations into consideration). Like his teacher Harris and the other structuralists, he derived immediate constituents by rules that annotated the diagrams with the familiar grammatical

categories: Noun (N), Verb (V), Adjective (A) Preposition (P), Sentence (S), and some others. Using these basic building blocks, he formalized a system of *phrase structure rules*. This system used a finite set of rules, symbols, and elements to project into an infinite set of strings. He also added new methods and showed how they were applied in the analysis of phrases and sentences. He also designed empirical tests, to justify the notion of constituent. Let's see some examples, so you can get a flavor of this early rule system.

Chomsky posited grouping rules, whose general format was $X \rightarrow Y + Z$, where X, Y, Z are categories like Sentence, Noun Phrase, Verb Phrase, Noun, Verb, and so on. These rules work through rewriting, designated by an arrow ($\rightarrow$). You rewrite the single category that is to the left of the arrow (X)—starting with the top category, Sentence—as the categories to its right (Y, Z): $\text{Sentence} \rightarrow \text{NP} + \text{VP}$ (or $S \rightarrow \text{NP} + \text{VP}$ for short). You continue with intermediate rules that derives smaller pieces (e.g., $\text{VP} \rightarrow \text{V} + \text{NP}$), and you continue until you get all the way down to the words, which are derived via “category labeling rules,” for example, $\text{N} \rightarrow \textit{you}$ and $\text{V} \rightarrow \textit{flew}$. Thus, to analyze the simple sentence *You flew*, you apply $S \rightarrow \text{NP} + \text{VP}$; $\text{NP} \rightarrow \text{N}$; $\text{VP} \rightarrow \text{V}$; and finally, the category labeling rules.

Word strings that cannot be analyzed with this given set of rules, for example, **Fly birds*, are deemed ungrammatical, because the tiny system I just sketched features no rule that enables the verb to precede its subject noun. We can therefore evaluate the correctness of a grammar by testing whether or not strings it is capable of analyzing are grammatical by our intuitions—by asking ourselves (or others) whether the line between good and bad sentences that the grammar draws suits our own judgments.

One immediate advantage of this system, which Chomsky was quick to point out, is that it can accommodate structural ambiguity. Certain phrases and sentences can be analyzed in more than one way, with obvious consequences for meaning: the phrase *the first parliament member* may be analyzed as [*the* [*first* [*parliament member*]]] and end up meaning ‘the first among many members of parliament’ or as [*the* [[*first parliament*] *member*]], with a meaning ‘the member of the first parliament (ever).’ Chomsky’s famous example was the ambiguous sentence *They are flying planes*. He used the rules to assign it two different syntactic analyses, giving a formal expression to the observed ambiguity. Here are

the rules, which can be applied in two different ways, to result in two graphical forms (tree structures) in figure 1.2:

Sentence $\rightarrow$ NP + VP

VP $\rightarrow$ V + NP

NP $\rightarrow$ *they*

NP $\rightarrow$ *planes*

V $\rightarrow$ *are*

NP $\rightarrow$ *flying planes*

V $\rightarrow$ *are flying*

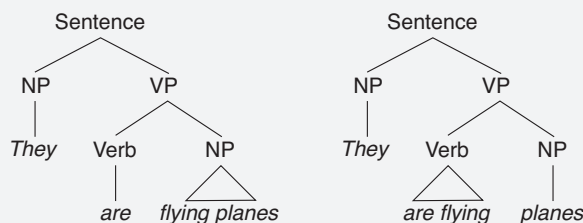


Figure 1.2

Two different labeled trees of a structurally ambiguous sentence (Δ marks a part whose internal structure is suppressed). Each labeled tree gives rise to a different meaning. Left: 'They are the planes that are flying.' Right: 'They are the persons who are flying the planes.' A playful reader may enjoy keeping track of the rules applied at each level of the two trees.

To Chomsky, we humans must possess a cognitive system that can create such trees, or else we would never be able to detect the ambiguity. Still, the example sentences thus far considered have been relatively short and simple, which means that there is more to human linguistic cognition than these examples illustrate—we and our grammar can generate and analyze arbitrarily long sentences. To illustrate this, Chomsky highlighted the grammar's device of *embedding*. This important grammatical tool that we have at our disposal enables us to nest one sentence within another (e.g., *John said that Mary loves cookies*)—and then to further embed yet another sentence (e.g., *John said that Bill thought that Mary loves cookies*) and keep doing so, perhaps ad infinitum. This is a most powerful ability. Indeed, it is widely agreed today that our cognitive syntax has the property of recursiveness.¹² The rule system we build must therefore capture it. To this end, it must be endowed with a new power: it has to be recursive.

Embedding and Recursiveness in Syntax: An Example

We use the grammar's embedding tools many times a day, nesting one sentence within another. Below, I illustrate how a tiny ad hoc rule system ("a toy grammar") generates embedded sentences. The previous rule system has been augmented with one category; let's call it Sentence Glue (SG) for illustrative purposes. This category enables a speaker to glue, or embed, one sentence within another, an operation that may recur any number of times. The tree structure in figure 1.3, associated with the following rules, shows that the top rule (Sentence $\rightarrow$ NP + VP) was applied three times in the sentence *John said that Bill thought that Mary loves cookies*.

Sentence $\rightarrow$ NP + VP

SG $\rightarrow$ *that* + Sentence

VP $\rightarrow$ V + SG

VP $\rightarrow$ V

NP $\rightarrow$ *John, Bill, Mary, cookies*

V $\rightarrow$ *said, thought, loves*

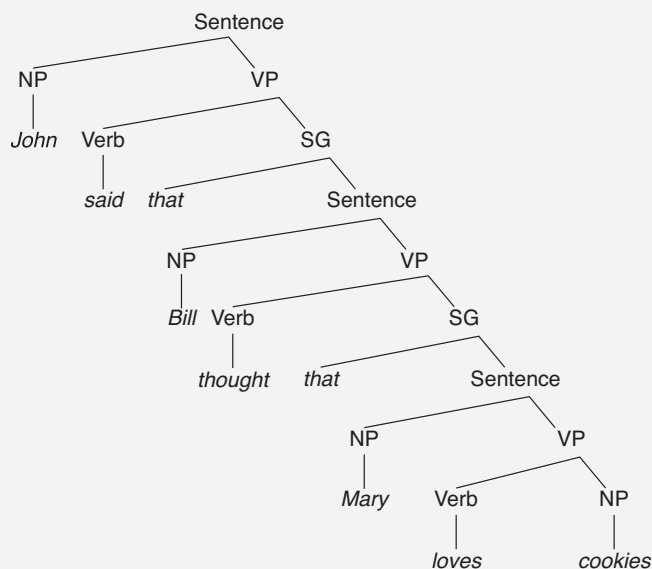


Figure 1.3

A sentence with two embeddings. The category Sentence is called twice by the rule SG $\rightarrow$ *that* + Sentence, which allows for three occurrences of Sentence, each expanded three times into NP + VP.

A recursive function is one that can call itself (whether directly or indirectly). For the example just given, this system must have rules that allow it to analyze a sentence within another sentence. To equip our phrase structure rule system with this property, we need to modify it a bit. Thus far, the category to the left of the rule's arrow was rewritten as two others to the right. Take in particular $S \rightarrow NP + VP$. The rules $NP \rightarrow N$ and $VP \rightarrow V + NP$ have no way of calling S again. But to embed a sentence within a sentence, we need to allow rules that call S . In other words, S should be allowed to be on both the left and right sides of the arrow (e.g., $S \rightarrow NP + S$); we can also achieve the same effect with rules that allow us to get from S to another S in two steps (e.g., $S \rightarrow NP + VP$; $NP \rightarrow N + S$). Note that we can now keep expanding our rules on end: $S \rightarrow NP + VP$; $NP \rightarrow N + S$; $S \rightarrow NP + VP$; $NP \rightarrow N + S \dots$). This gives our rule system a lot of power: we can embed time and again.

Early TGG and Its Properties

Chomsky then moved on to other important relations between sentences, and he set up rules to account for them. He highlighted the intimate relationship between the active sentence *The man ate the food* and the sentence *The food was eaten by the man*, which contains roughly the same words and has roughly the same meaning but is in passive voice. This, he argued, is an important relation that the grammar should not miss. He then showed that his phrase structure rules, which assumed linear adjacency between constituents, were incapable of capturing the relation between these two sentences: to get the passive from the active, you need to switch the positions of *the man* and *the food* in the sequence, then change the verb *ate* to the participle *eaten*, and finally, insert the auxiliary verb *was* and the preposition *by*. With a grammar that only has phrase structure rules, you cannot carry out such operations.

Chomsky's new grammatical system was designed to solve such problems. First presented in detail in his 1957 monograph *Syntactic Structures*,¹³ it contained two components. First was a system akin to Harris's phrase structure rules, which analyzed, or generated, sentences that were annotated for their immediate constituents. Then, to fix the active-passive problem described above, this rule system was augmented with a system of *grammatical transformations*, rules that have a different shape: given the sequence of words in the sentence (a concatenation), as generated by the

phrase structure rules, they may switch the sequential order, substitute some words, and add or delete others.

I suggest taking a minute to see the fine mechanics of this 1957 system through the active–passive sentence pair *The man ate the food* and *The food was eaten by the man*. The sequence of operations takes the active sentence *The man ate the food* and first applies the rule $S \rightarrow NP + VP$. This operation yields $NP_1 = [the\ man]$; $VP = [ate\ the\ food]$. The next step parses the pieces into yet smaller ones: $VP \rightarrow V + NP$ results in $V = ate$; $NP_2 = [the\ food]$. Finally, the verb is broken into a tense marker (categorized as an Auxiliary, or Aux for short) and a verbal stem: $V \rightarrow Aux + V$, results in $ate = [past][eat]$. The sequence of labels of the resulting string is $NP_1 + Aux + V + NP_2$.

This string is then input to the passive transformation: it is on the left side of the double arrow symbol $\Rightarrow$ below. On the right side is the output of the transformation: the sequence of categories and the special words of the passive sentence.

$$NP_1 + Aux + V + NP_2 \Rightarrow NP_2 + Aux + V + en\ by\ NP_1$$

Expanding the categories to their terminal nodes, we obtain *The man ate the food* $\Rightarrow$ *The food was eaten by the man*.

This setup has several implications: the grammar is now more complex, because it is divided into two rule types: the first type features generative “structure-building” phrase structure rules; the second is a new rule type—a grammatical transformation, which relates pairs of sentence types (such as active–passive, question–answer, and much more) to one another. Thus, phrase structure rules generate basic structures, and grammatical transformations operate on the basic structures that are the output of these rules/transformations and thus account for regular relations among structures. The result: transformational generative grammar (TGG), as it became known. Transformations enrich a relatively limited inventory of basic (or “kernel”) sentences of the language: they enable derivation of passive sentences, questions, and other complex sentences, which are just one or two transformations away from their active declarative counterparts. Transformations constitute a purely syntactic operation, one that (supposedly) does not affect meaning. The grammar becomes more cumbersome, but the payoff is an explanation for a new set of phenomena: we get an answer to the question why languages feature pairs of sentences that have the same words and the

same meaning, despite the fact that they differ in their syntactic form. The answer should be clear by now: one member of the pair is derived by a syntactic transformation from the other. Note that words in this grammar are syntactic objects, an integral part of the phrase structure rule system (e.g., there may be many rules of the shape $V \rightarrow eat$). This would later change: a reformulated grammar would accommodate a separate store of words, or mental lexicon, with its own properties.

With a system of rules at hand, Chomsky proceeded to argue that syntax is blind to meaning (let alone connotation): a well-known example he gave contrasted two nonsensical word strings, *Colorless green ideas sleep furiously* and *Furiously green sleep ideas colorless*. Both, he observed, make no sense, and yet only the former is grammatically well-formed. Syntax, he concluded, is therefore about form and is independent of meaning and context. He thus embarked on the development of methods for the study of syntactic structure that is independent of content.

How does this affect alternative approaches to grammar? The cultural approach immediately becomes unrelated to Chomsky's object of inquiry. As for the statistical approach, he criticized it by showing that an information-theoretic system, incapable of analyzing sentences into a hierarchy of parts and subparts, misses important linguistic generalizations. An example he gave was the complex sentence *If [John sleeps late] then [he will miss the train]*. Its logical and syntactic structure, he argued, resembles *If [I don't eat the carrot], then [I can't lift the weight]*: while apparently unrelated, both sentences conform to the conditional structure *If Sentence₁ then Sentence₂*. But if we look at the neighborhood relations of the words *if* and *then* in these two sentences, we get a different perspective: the first *if* has a sleeping context, whereas the second, an eating one; the first *then* has a train-missing context, whereas the second has a strength one. Chomsky wrote that if we start just by looking at neighborhood relations and count how many times *if* and *then* appear in particular contexts, we would be led to treating these two sentences as unrelated, and we would miss an important generalization: that the two sentences have common structural properties and participate in the logical relation *if . . . then . . .* To Chomsky, this was among many profound and critical regularities in human language that a purely statistical perspective misses, which indicated that it was deeply lacking.

This is how TGG began. Importantly, in making his then-novel proposals and contrasting them with extant alternatives, Chomsky followed a methodological strategy, one philosophically known as "inference to the

best explanation.” The account he proposed was not only supported by evidence; it was also accompanied with a demonstration that it was superior to alternative accounts (structural, statistical, cultural), having solved problems that they had not.

Much has been written on the skepticism, to say the least, with which this new approach was greeted. I will only note that traditional linguists argued that it only focused on English, ignoring other languages; philosophers accused it of assuming an obscure version of mentalism, even dualism (by which minds are not causally related to bodies); psychologists chided its bold acceptance of inner, unobservable, cognitive processes. But even its fiercest critics today would admit that it heralded the cognitive revolution, which placed its focus on the mental, and then neural, mechanisms that govern human behavior, rather than on behavior itself. Some of these issues will play a role in chapters to come.

With this short history in our pocket, we can fast-forward to today. Whereas the legacy of this 70-year-old system does live on in certain respects, its descendants look rather different. While there are at present multiple versions of linguistic theory, which debate one another, they all seem to be more generalized, more constrained, more precise, and more abstract.¹⁴ Possibly, many if not most of them are even more explanatory than their 1950s predecessor, and their empirical scope is certainly much broader. Let’s look into this in some detail.

What the Linguist Is Talking About—The Concepts “Grammar” and “Grammaticality”

Grammatical transformations, as a mere technical innovation, do not a scientific revolution make. Indeed, the new linguistic theory in which Chomsky couched his technical ideas had larger scientific goals. In a nutshell, he looked beyond the descriptive horizon of his predecessors, viewing linguistics as being not about linguistic description but rather about the *human knowledge* codified in this description. In other words, TGG, if accurate, properly concise, and comprehensive, is not a mere grammar book for foreign language learners.

So, what is the nature of this construct, TGG, that we are said to possess deep inside our minds (and brains)? It is a theoretical account of what human language users know about their language—knowledge by virtue of which they are capable of using language the way they do. Importantly,

this theory is not just about what humans do when they use language; it is about what they *can* do with language, as owners of a human language faculty. The point pertains to all grammar rules, from phrase structure rules ($S \rightarrow NP + VP$; $NP \rightarrow N + S$; $VP \rightarrow V + NP$, . . .) to the complex transformational rules ($NP_1 + Aux + V + NP_2 \Rightarrow NP_2 + Aux + V + en$ by NP_1) down to that little *-s, -z, -ez* rule that every native speaker of English knows. The last rule, in fact, is followed strictly, excluding even wrong plurals of invented words, as a famous experiment showed long ago. Children were asked to turn a singular noun into its plural form. They were shown an unfamiliar object and told, “This is a *wug*!” Then, they were shown two such objects and asked, “What are these?” Remarkably, even preschoolers returned *wug-z*, not **wug-s* or **wug-ez*.¹⁵ Speakers seem to have clear intuitions about grammatical well-formedness at the word, phrase, sentence, and text levels. Notably, these intuitions are rather uniform across native speakers of the same language (there are some sources of noise and distortion, which we linguists have learned how to control and filter out).

The set of rules that the linguist posits is therefore expected to cover all the well-formed sentences in a language that a native speaker may ever utter, and only these. The actual status of sentences (distinguishing between, e.g., the well-formed *The man ate the food* and the ill-formed *The man ate a food* and *The food ate the man* and certainly *Food the ate man the*) is determined solely by the testimony, or judgment, of native speakers of the language—evidence obtained by behavioral experiments.

If TGG is capable of distinguishing between the well- and ill-formed sets of word strings and if it is the best way to do so that is found on the market of ideas, then we must conclude that *it is* the very knowledge that the speaker possesses. Such a bold claim is again in keeping with the well-known philosophical principle “inference to the best explanation” (also known as abduction), which seems to be followed in most sciences: if you have several accounts of a set of natural phenomena and if you also have criteria that help you decide which account is best, then it follows that the best explanation is the *true* theory, until further notice.¹⁶

Pursuing this line of scientific reasoning, Chomsky’s research program aimed to characterize the human grammar, which should lead to an understanding of the human communicative ability—our exceptional ability (perhaps unique within the animal kingdom) to produce and comprehend any possible utterance in our native language, regardless of whether we have previously heard or read it.

Are We Born with a Grammar in Our Head? The Innateness Hypothesis

If speakers of English know its grammar (and that includes young preschoolers, who say *bee-z* and not *bee-s* just like everyone else), how did they get to know it? Did the rules get into their heads by rote learning? Was there a teacher? Moreover, if English speakers know their grammar, why can't they just write down its rules?

An influential idea for an answer was articulated by Plato some two thousand years ago, in his dialogue *Meno*. The issue debated there is whether knowledge is acquired through the recall of preexisting ideas or through learning by exposure to information (a.k.a. the nature vs. nurture debate). Through a Q&A experiment with a slave boy, Plato argues for the former position. That is, he claims that much of the knowledge we possess is actually ours from birth and only needs triggering via an interaction with the environment. Some knowledge, then, is innate. Chomsky, relying mostly on Descartes's later endorsement of a version of this idea, argued for innate knowledge from a linguistic angle in his celebrated *poverty of the stimulus* argument. We have an ability to produce and comprehend grammatically well-formed sentences that we have never seen, he said. This ability, to generalize from examples to a rule, is not learned. Hence, aspects of our grammatical knowledge transcend our past experience and must be innate.

Chomsky's fresh approach and sharp arguments, as well as his personal charisma, won him young supporters early on. As philosopher John Searle put it, "Chomsky did not convince the established leaders of the field, but he did something more important, he convinced their graduate students."¹⁷ George Miller, the mathematical psychologist who had started out with statistical models for next word prediction, was among the first to change his views, and he became an early collaborator. Together, Chomsky and Miller wrote a program for the new language science, which consisted of three questions:

- A. What is the precise nature of linguistic ability?
- B. How does it arise in the individual?
- C. How is it put to use?¹⁸

At the time, little was known about the biological bases of linguistic ability. Later, two additional foundational questions were added to the list:

- D. How does the human brain support this ability?
- E. Is it uniquely human and how did it evolve?

It appears, then, that the new science of linguistics has touched on perennial psychological and philosophical questions from a linguistic angle; later, issues concerning its neural bases also arose. It is for this reason that this science has had far-reaching implications; it is also one of the reasons why it ignited so many debates and controversies and has acquired a considerable number of critics and enemies.

The Precise Nature of Our Linguistic Ability

Seventy years past these early efforts, the current theory is obviously different. Generations of linguists have improved it vastly, broadened its empirical coverage, and turned it into a highly formal and constrained theoretical apparatus. Still, the ultimate goals of this theory have largely remained similar to those Chomsky originally formulated: to account for human linguistic ability.

What, then, is the nature of this ability and how does science characterize it? If linguistic ability were only characterized by the sentences humans produce in speaking and writing (perhaps with special emphasis on those that are most frequent, like *The man ate the cake*), the linguist's life would be relatively simple. Almost anyone who has taken a basic computer science course could write a grammar for such a language, which would likely implement some basic phrase structure rules, as was done above. That should not be too difficult for a programmer—after all, Chomsky's early work laid the foundations of the theory of programming languages and their complexity.

The matter becomes more serious once properties of language that are less visible are taken into account. We saw how the earliest version of TGG rationalized the need for transformations by observing abstract properties of the language: the theory created a typology of sentences (active–passive, question–answer, and more) and argued that these types needed to be related to one another, which was the function of the transformational rules. The relation between such sentence pairs is an abstract property of language, which must be accounted for. Transformational rules, however, complicate the grammar. Indeed, the 1950s syntax featured two distinct rule types, with a clear ordering imposed on their application. Phrase structure rules were used to build structure, and the output of this process was then fed into transformational rules. Take active and passive sentences. The thought was that, while

they may sound different on the surface, deep down their meanings are identical. By the 1960s, this thought was made explicit: pretransformational representations were called “deep” structures, whereas their posttransformational derivations were called “surface” structures. Active and passive sentences thus had the same “deep” structure but differed in their “surface” structures—the external shape of the sentence as we finally utter or hear it.

The Cognitive Architecture of TGG—Take One

Each sentence had two representations: “deep” and “surface.” For an active sentence, they contained the same word sequences, but for a passive sentence, different ones: the deep structure of the passive fed into the transformational rule $NP_1 + Aux + V + NP_2 \Rightarrow NP_2 + Aux + V + en + by + NP_1$, making its overall structure more complex than its active counterpart. To simplify matters, we can choose the total number of symbols of the rule as our complexity metric (repetitions allowed); there are seven symbols to the left of the arrow ($NP_1, +, Aux, +, V, +, NP_2$) and 11 to the arrow's right ($NP_2, +, Aux, +, V, +, en, +, by, +, NP_1$). The grammatical system of TGG was shaped as in figure 1.4.

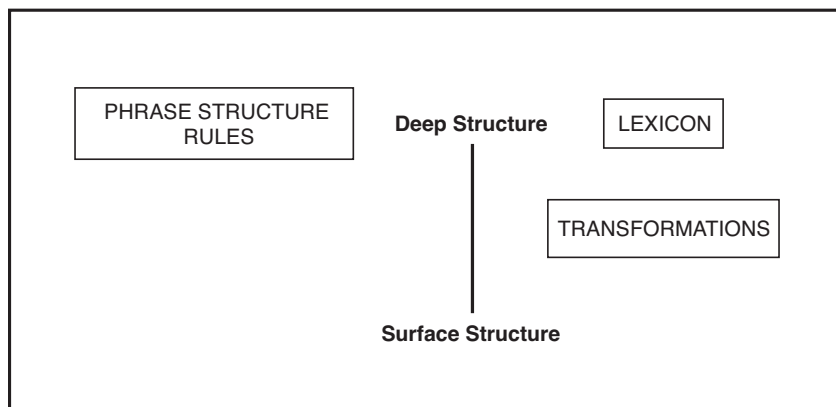


Figure 1.4

The structure of the grammar, Take One. Each sentence is associated with two representations. Deep Structure is fed by phrase structure rules that combine with words pulled out of the lexicon. The result is mapped onto Surface Structure by transformational rules (which may or may not apply).

On Take One, TGG posited syntax as a specialized cognitive system for syntax that interfaces with the semantic system and with general cognition via Deep Structure. Deep structures are stored in every speaker's brain. At its other end, Take One had Surface Structure, interfacing with the sound system of language. Transformations mapped Deep Structure onto Surface Structure and vice versa. The path from meaning to sound and back was assumed to be through the syntax, which, upon investigation, was becoming increasingly complex.

The metaphors “deep” and “surface” inspired many thinkers and aggravated some, leading to multiple attacks on the theory, although the representations were neither deep nor surface—these names merely denoted abstract objects, the grammatical descriptions of sentence structure.¹⁹

This is how the nature of grammar was viewed. It was further seen as a key to our mental makeup, as it enabled rich and flexible communication. In the next chapter, we dig a bit deeper into its later development, which I dub Take Two.

Taking Stock

In this chapter we saw:

- Snippets of historical grammars, from Pāṇini's Sanskrit grammar
- Three twentieth century approaches to linguistics
- Reflections on what grammar might be
- Structural ambiguities providing motivation for immediate constituents
- Chomsky's phrase structure rules as a basic way to generate structure. Their recursive property enables embedding (in fact, as many embeddings as you wish)
- Chomsky's transformational rules—part of TGG—accounting for regular relations among sentence types, such as active–passive and question–answer
- The broader significance that TGG has for psychology, philosophy, and neuroscience
- Chomsky and Miller's program for the language sciences
- Innate ideas then and now (from Plato to Chomsky)
- Take One of the architecture of our cognitive syntax: Deep and Surface Structure

This is a section of [doi:10.7551/mitpress/15932.001.0001](https://doi.org/10.7551/mitpress/15932.001.0001)

How Deeply Human Is Language?

Chomsky, the Brain, and the AI Fantasy

By: Yosef Grodzinsky

Citation:

How Deeply Human Is Language?: Chomsky, the Brain, and the AI Fantasy

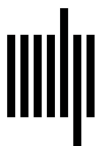
By: Yosef Grodzinsky

DOI: 10.7551/mitpress/15932.001.0001

ISBN (electronic): 9780262052023

Publisher: The MIT Press

Published: 2026



The MIT Press

The MIT Press
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139
mitpress.mit.edu

© 2026 Massachusetts Institute of Technology

All rights reserved. No part of this book may be used to train artificial intelligence systems or reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data is available.

ISBN: 978-0-262-05200-9

EU Authorised Representative: Easy Access System Europe, Mustamäe tee 50, 10621 Tallinn, Estonia | Email: gpsr.requests@easproject.com