

2 Hints from Present-Day Linguistic Theory

Syntax, my lad. It has been restored to the highest place in the republic.

—John Steinbeck on JFK's inauguration speech, 1961

You know as much about this as I do. Argue with me! Teach me something I don't know!

—Morris Halle, cofounder of the MIT Department of Linguistics

The Cognitive Theory of Syntax

We now turn to the development of the cognitive theory of syntax, not before acknowledging that the short introduction in the previous chapter left many general issues untouched. My intention here is not to teach you linguistics but only to introduce the main problems the field is trying to solve, the rationale of the enterprise, and some basic theoretical ideas and tools that are common to virtually all theoretical frameworks that are currently on the market. Finally, I discuss solutions that linguists offer to certain problems, along with implications these solutions may have for philosophy, psychology, and neuroscience. As noted, once matters get a bit technical, the text is boxed and marked gray. You may skip these parts, mostly with impunity (though you'll be missing out on some fun stuff).

A central theme that I leave out is cross-linguistic variability. The legend of the Tower of Babel may have no historical basis, but the fact is that at present, there may be as many as 7,000 languages spoken in the world. These may have different grammars, and TGG as thus far presented obviously cannot be the grammar of all of them. Briefly, the vision (major parts of which have actually materialized) is to have a unified and abstract theory of syntax, with built-in parameters that allow for limited variation across

languages. To take the phrase structure rules as an example, they work for English but not German. In English you say *I have **seen** the girl*, the verb *seen* appearing before the object, whereas in German, it is at the end: *Ich habe das Mädchen **gesehen***. English and German are thus similar in that every sentence has a subject and a verb, potentially an object, and so on—but the grammars also appear to be different; the examples just given show that the position of the main verb is not the same. This difference forces the formulation of rules that are sufficiently abstract and tight but also sufficiently flexible to accommodate such similarities and differences.

While cross-linguistic issues occupy the minds of many, they go beyond the present scope. I will only note (bringing in no evidence) that there is a level of description (sometimes called Universal Grammar) at which the shape of all human grammars, though not the specific rules themselves, is one and the same. Languages and their grammars do differ, but the cognitive principles behind them do not, as far as we know.

Questions and Answers—Critical for Linguistic Ability and Functioning and for Linguistic Theory

As repeatedly noted, you can't do everything in one book. The topics discussed here have been chosen for their relevance, as well as for both historical and illustrative purposes; they are intended to underscore difficult but central empirical problems for the theory of linguistic cognition and to put on display theoretical efforts to understand them and creative moves made to help in that endeavor.

For some philosophers, language may exist in the absence of communication between individual agents. They view it as an internal cognitive tool for thought rather than communication.¹ But even these thinkers agree that subsequent to its evolutionary development, this tool was externalized (or “exapted”)² to serve as a tool for communication or information exchange among agents. A good place to start our discussion would therefore be a paradigmatic type of communication scenario, namely, requests for information and informative responses to these requests—in our case, English questions and their answers:

1. Q: **Who** ate the chicken? A: **John** ate the chicken.
2. Q: **What** did John eat? A: John ate **the chicken**.

These are good Q–A pairs under normal circumstances, given cooperative and truthful interlocutors with a shared body of contextual knowledge (in this case, that there was more than one potential eater; that John is that guy over there; that among the food items around, there was a single chicken). It is of note that such exchanges are not always successful. Schizophrenic patients famously answer questions “tangentially,” responding to them, sharing information, but without answering. For example, Q: “How have you been feeling today?” A: “My flat is just around the corner.” Here, I discuss the grammatical properties of *successful* exchanges via Q–A pairs.

English question words are typically at the beginning of a question sentence (as *who* and *what* are in the questions in 1 and 2), while the center of the answer (the bold element in the answers in 1 and 2) may be in another position. In 1, the question word and the center of the answer are both in a sentence-initial position (this type of question is known as a subject question). In 2, they are not: the question word(s) is sentence-initial as before, but the center of the answer is sentence-final (an object question). In the question in 2, moreover, the verb *ate* breaks into two nonadjacent pieces, the past-carrying *did* and the verb *eat* (reminiscent of the treatment of the verb by the passive transformation; see chapter 1).

Another oddity in 1 and 2 is that while the two questions differ from each other, the shape of the answers is the same—and yet they do not convey the same information. A question carries a request for an unknown piece of information; a verbal answer to the question is expected to convey that information, usually through a declarative sentence. It is but natural to think about this unknown as a variable, whose value will depend on the question’s content and the context in which it is asked. We can reformulate the Q–A pairs in 1 and 2 in a form that is a bit logic-like in flavor:

1’. Q: *What is the identity of x , such that x ate the chicken?*

A: x = **John** (ate the chicken)

2’. Q: *What is the identity of x , such that John ate x ?*

A: (John ate) x = **the chicken**

Speakers are of course aware of these properties of Q–A pairs (even if we may not always be able to articulate them properly). We ask and answer questions daily with little difficulty; the theory of our internal grammar must explain how such exchanges work seamlessly and why deviations from the norm (like those in schizophrenia) are so rapidly detected. So, there are

two goals here. First, to set up a rule such that “deep” structures of Q–A pairs are closely related (similar to active–passive pairs, as discussed in chapter 1). Second, to design a syntax for 1 and 2 that leads to the “meanings” in 1’ and 2’. For simplicity, we will let ourselves be slightly sloppy: for ease of reading, let’s suppress the phrase structure and category notation.

A suitable TGG question–answer rule may be a syntactic transformation that establishes a relation between a *who* or *what* question and its answer by (a) choosing a noun phrase from the answer (*John, the chicken*) and replacing it with the proper question expression containing the variable *x* (the italicized string in 1’ and 2’); (b) copying this expression to the front of the sentence; (c) replacing the original copy with the variable *x*. Take a minute, and try to apply these three operations to the A part of 1 and 2; you should get the Q part of 1’ and 2’, respectively. This question transformation expresses the syntactic relation between an answer and its question. In the spirit of Chomsky’s early work, here are possible transformational rules (with some abbreviations, for ease of reading, e.g. the auxiliary *did* in 2 is suppressed in the notation). On the left side of the double arrow $\Rightarrow$, there is a structural description of the sentences that would undergo the change, described on the arrow’s right side.

1”. Subject question

A: $NP_1 + V + NP_2 \Rightarrow$ Q: *what x*.*x* + V + NP_2

2”. Object question

A: $NP_1 + V + NP_2 \Rightarrow$ Q: *what x*. $NP_1 + V + x$

Note that this rule is formulated backwards (from A to Q) and that it (intentionally) emphasizes the question–answer relation. The representations in 1” and 2” add a semantic perspective: the question operator *what x* (which stands for a request to convey *the identity of x*) and the variable *x* indicate that *who* and *what* questions are requests for information about an unknown. The location of *x* marks the position of this unknown entity in the question sequence. This variable is connected (by a logical relation called binding) to its position in the answer.

This complex representation contains hints about the complexity of the communication process. Yet this complexity does not hinder communication. Q–A exchanges occur frequently and without much reflection—we seem to know what we are doing, or else communication would constantly break down.

How Far Can a Question Go?

What is the difference between the Q part in 1 (Q_1) and the Q part in 2 (Q_2)? Both intuition and experimental evidence suggest that perceptually, Q_2 is more difficult to understand than Q_1 . This may be due to two properties of their respective representations. Perhaps somewhat naïvely, it appears that, as Q_1 is a subject question, all one needs to do in order to convert the answer into a question is replace John with *what x.x*; whereas in Q_2 , an object question, there seems to be an added displacement as the question expression—*what x*, a.k.a. *What is the identity of x*—is copied to the front. Relatedly, the linear distance between the question expression and *x* is larger in Q_2 than in Q_1 .

We will delve into the reasons for the difficulty differential later (in the context of the syntactic brain). At present, we focus on the observation that questions may displace a constituent at various distances. How far can *who/what* travel (a factual issue), and will it affect the question transformation in 1" and 2" (a theoretical issue)?

Facts first. Verbs like *think* may help us track distances between the relevant positions of the words (or phrases) under consideration. Such verbs typically are followed by (or more accurately, embed) a sentence that describes the content of a thought:

3. Q: **Who** did *Mary think* ate the chicken? A: *Mary thought* **John** ate the chicken.
 4. Q: **What** did *Mary think* John ate? A: *Mary thought* John ate **the chicken**.

Applying our quasi-formal rewording routine, we get:

- 3'. Q: *What is the identity of x*, such that *Mary thinks x* ate the chicken?
 A: *x=John* (Mary thinks ate the chicken)
 4'. Q: *What is the identity of x*, such that *Mary thinks* John ate *x*?
 A: *x=the chicken* (Mary thinks John ate)

Embedding made the sentence longer and perhaps more complicated, yet our method allowed us to express the speaker's intention in this way. However, as the reader may verify, the transformational rules in 1" and 2" no longer apply: the current string of words does not match the transformation's structural description (on the left side of the arrow). To fix that, we

could try to add ingredients to the transformation. It would become more cumbersome, but it would fit the new sentences (again, the auxiliary *did* is suppressed for simplicity):

3". **Subject question**_{think}

$NP_1 + V_1 + NP_2 + V_2 + NP_3 \Rightarrow \textit{Who/what } x.NP_1 + V_1 + x + V_2 + NP_2$

4". **Object question**_{think}

$NP_1 + V_1 + NP_2 + V_2 + NP_3 \Rightarrow \textit{Who/what } x.NP_1 + V_1 + NP_2 + V_2 + x$

This move, which adds special rules designed to accommodate this specific embedding under *think*, is not only heavy-handed and ad hoc; it is also hopeless, because language allows the freedom to embed multiple times and still ask questions and give answers, whereas the rules in 3" and 4" do not. Keeping with the rule-adding strategy would force us to formulate a special rule for each embedding, which would only blur our vision, not sharpen it, as it would make the general property of the transformation evaporate into a multitude of specific rules:

5. **Subject question**_{think-believe}

Q: **Who** did Mary think *Bill believes* ate the chicken?

A: Mary thought *Bill believes* **John** ate the chicken.

6. **Object question**_{think-believe}

Q: **What** did Mary think *Bill believes* John ate?

A: Mary thought *Bill believes* John ate **the chicken**.

Questions can be asked from the bottom of even longer sentences, with three embedding verbs:

7. **Subject question**_{think-say-believe}

Q: **Who** did Mary think *Sue said* *Bill believes* ate the chicken?

A: Mary thought *Sue said* *Bill believes* **John** ate the chicken.

8. **Object question**_{think-say-believe}

Q: **What** did Mary think *Sue said* *Bill believes* John ate?

A: Mary thought *Sue said* *Bill believes* John ate **the chicken**.

Something more abstract and general is needed, in order to precisely codify the question expressions' apparent freedom to travel large distances, across many nouns and verbs.

A promising possibility leaves the context blank and allows the transformation to create questions everywhere, as long as they position the question expression on the left edge of the representation:

9. Subject question

$\dots NP_n + V_n + NP_{n+1} \Rightarrow \textit{Who/what } x \dots x + V_n + NP_{n+1}$

10. Object question

$\dots NP_n + V_n + NP_{n+1} \Rightarrow \textit{Who/what } x \dots NP_n + V_n + x$

The symbol “...” stands for “any otherwise legitimate sequence of words.” This move would force a new rule type—one that contains a variable that can take any legitimate context as a value. This way, the transformations in 9 and 10 would account for 5–8 (e.g., for 5, just replace “...” of the rule with “*did Mary think Bill believes*” and you will see that the rule works!).

This more relaxed formulation of the transformation seems to work. But new problems arise as the rules are modified from the rigid but unrealistic 3” and 4” to the somewhat looser 9 and 10. The latter harbor incorrect predictions. The Qs in 11 and 12 are in keeping with these rules (for 11, replace “...” of the rule with “*the woman who taught*”; for 12, replace “...” with “*the thug who*”), and the linear distance between *who* and its original position is shorter than in the case of *what* in 8, but the Qs are nonetheless ungrammatical (marked “*” and annotated henceforth as Q*). Importantly, the request for information each of these bad questions conveys makes perfect sense, as seen by the paraphrases Q_✓. Yet, as stated, no Q can be asked.

11. A: Bill married the woman who taught **his sister**.

Q*: **Who* did Bill marry the woman *who* taught ~~*who*~~?



Q_✓: *Who* is the person such that Bill married the woman who taught her?

12. A: The thug who **Bill** saw was scary.

Q*: **Who* did the thug *who who* saw was scary?



Q_✓: *Who* is the person such that he saw the thug who was scary?

I have added two helpful notational innovations: an arrow marks the link between the target and original positions—the sentence-initial copy of *who* and the original *who* (the one that replaced the boldface center of the

answer)—and the latter now appears *struckthrough*. The third, middle *who* is our problem, as I will suggest imminently.

These new discoveries bring us to a crossroads: question expressions can travel quite a distance, but not in every context. We would like to formulate general rules that make every good question possible, no matter how many embeddings the question expression travels through. Yet, at the same time we want this rule to mark questions like 11Q* and 12Q* as bad, because this is how native speakers judge them. Importantly, the reason for the ungrammaticality of the latter seems to be syntactic, not semantic: it is easy to see that the ungrammatical monsters 11Q* and 12Q* do convey coherent meanings (as seen by their paraphrases 11Q_✓ and 12Q_✓). The system we built is not working, and it needs to be repaired. In sum, at least with respect to the puzzle in 11 and 12, the prime suspect for disrupting the Q* may now be identified: it is the second token of *who* that seems to be culpable. Once the arrow crosses it, the sentence is rendered ungrammatical, and both sentence and arrow earn a star.

Once again, these observations hit the core point. If any sequence of syllables constitutes a good word, we don't need a grammar; likewise, if a question can be asked by any sequence of words, a syntax is unnecessary. On the other hand, to rule out sequences of syllables and words such as 11Q* and 12Q*, which are deemed unacceptable by native speakers of English, rules must be invoked. Words and sentences are constrained as part of any speaker's knowledge, and it is our task to understand why this is so and to construct explanations for these phenomena. So, let's carry on.

From “Anything Goes” to Constrained Rules

Ungrammatical sequences are nothing new. Yet the discovery of *systematic* ungrammaticality in the context of transformations, as presented above, was an important landmark in the development of the theory of syntax some 60 years ago.³ It taught the field an important lesson. A theory in which anything goes, that is, where you can write rules at will as long as they do the job, comes with a heavy price tag: If you let transformations apply freely, your grammar produces monsters (this is called the “overgeneration” problem, as the grammar is too permissive and allows, or generates, too many possible structures). One voice may suggest a quest for ways to

tame our rules, in order to block monsters like 11Q* and 12Q*. An opposing voice, however, might retort, “Why bother?” Why waste time on monster blocking, given that monsters hardly ever appear in texts or conversations? After all, at its daily toil, a language processing machine with a loose, unconstrained grammar never encounters monsters like 11Q* and 12Q*; so why are we dwelling on this marginal issue?

Here is my short answer, on which I will elaborate in part III: despite many similarities, language tool building and linguistic theorizing are not the same. The engineering effort is about machines that perform *actual* tasks, whereas science seeks the true explanation of human cognitive structure through inquiries into the full range of humanly *possible* performances. These two approaches may not always invoke the same considerations. In the present case, the contrasts in human judgments that I described—between legitimate long-distance travels of question expressions and a ban on such travels in certain contexts—reflect what humans know and can do (indeed, these examples come from grammatical testimonies of human owners of adult linguistic cognition). As such, they are at the center of scientific inquiry and demand an explanation, even though they may be rare, hence of little interest to machine builders.

Later I will reflect on whether my perspective on the commitments of a scientific linguistic theory is hopelessly romantic. For now, let’s be more concrete: we have an urgent need to block 11Q* and 12Q* and deem them ungrammatical, while preserving all the grammatical questions. The tools we currently have at our disposal cannot do this. Indeed, the discovery of monsters like these led to an important shift in theoretical focus, which began in the late 1960s. Since then, a thorough investigation has attempted to uncover precise *constraints* on the format and mode of application of both syntactic transformations and phrase structure rules. These constraints block monsters, and they also allow rules to be stated more generally. The end result has been a new cognitive architecture for the grammar, which I will now motivate, explain, and then illustrate.

In discussing transformations and embedding (5–8) and attempting to formulate the rules (9 and 10), I considered a transformation that contained an unspecified element that I represented as “. . .” The idea was to allow the transformation to cross “any otherwise legitimate sequence of words.” Yet this led to an overly permissive rule that admitted monsters (11Q* and

12Q*). The next move is to impose a constraint on transformations, one that would selectively bar them from applying in certain contexts: for example, if their application would lead to the crossing of a *who* or a *what*. A way to implement this is to state the contexts in which transformations may apply and to bar them otherwise.

In reflecting on the way theoretical constraints on transformations should be stated, several strategies present themselves. One way to formulate constraints on a syntactic system is to restrict a rule's domain of application; namely, to specify the conditions that must be met for a rule to apply (in its mapping from "deep" onto "surface" representations). An alternative strategy is to allow rules to apply freely but set limits on the format of the resulting representations: that is, to apply a kind of filter that looks at the output of rules and sifts out deficient results. Choosing between these possible methods involves serious considerations with far-reaching consequences. Over the years, syntacticians have provided precise answers.

The introduction of constraints on movement has led to generalization. Specific transformations (for questions, passive, etc.) were subsumed by the more abstract and generalized rule *MOVE*, an operation that copies elements from one position in a representation onto another and that has the capacity to delete the original copy (e.g., *What did John eat what?*). The generalization ranges over different structures: for example, *That steak, John ate ~~that steak~~ yesterday*; *It is my soup that John ate ~~my soup~~ yesterday*; *What did John say that Bill heard that Mary is ~~what~~?* And much more, within a language and across languages.

By the same logic, phrase structure rules were subjected to a severe constraint (which will be revisited in chapter 5) that limited their format. This constraint later helped to formulate a single rule *MERGE*, which takes two distinct pieces in a representation and merges them into a single syntactic unit. It can do so to sublexical units, to single words, and also to phrases. By this theoretical development, many rules that could earlier be part of the human cognitive syntax were now blocked. The resulting theory, which is at the heart of an approach known as Minimalist Syntax, is much more constrained than its previous incarnations.⁴ Other approaches to syntax may use different names and methods, but the idea is one and the same: constrain, simplify, and hence become more explanatory.⁵

How can a list of specific rules be squeezed into two? The answer lies in the architecture of the grammar: it now consists of the two rule types just discussed (though not formally presented), namely MERGE and MOVE, along with some other ingredients that will be laid out shortly. Critically, each sentence is associated with several highly constrained representations, mapped onto one another by these rules. The constraints set clear bounds on the notion “possible grammatical rule” and, in doing so, allow a restricted set of rules to be formulated. When stated correctly, they block monsters like 11Q* and 12Q*. Constraints on representations are key to the success of the grammatical system in assigning a correct structural description to any grammatically acceptable sentence and rejecting all ungrammatical strings.

An Example Constraint: Relativized Minimality

The ungrammatical questions (Q*) in 11 and 12 were said to be bad because the distance between the question word and its original position was too large. But how to define the notion “too large a distance” precisely? It must be done in such a way as to allow the acceptable travel of a question word out of multiple embeddings (5–8) but block the bad questions.

We noticed that in 11Q* and 12Q*, the link that the transformation formed between the question word *what* and its ~~struck through~~ original position crosses another *who*. An influential proposal by University of Siena’s linguist Luigi Rizzi (here simplified) turned this observation into a general constraint on representations containing a transformational relation; this constraint, stated in 13, would derive the facts.⁶

13. Relativized Minimality

In the configuration . . . X . . . Z . . . Y . . .



Y cannot be related to X by MOVE if Z *intervenes* and Z is of the *same type* as X.

Z will be annotated IV (for *intervenor*).

This formulation prohibits the presence of an intervenor between two MOVE-related positions, or else the resulting representation would be ungrammatical. Above, Z is present between X and Y. But to count as *intervenor* (IV), Z must be relativized to X, Y (i.e., X, Y, Z must all be of the same type, and that is a complicated matter). In 14–16, I illustrate how the *intervenor* concept works. In each instance, I added a paraphrase, Q_r, showing that the ungrammatical output of the transformation, or MOVE, nonetheless conveys a coherent meaning.

(continued)

14. A: I wonder who could solve the problem **in this manner**.

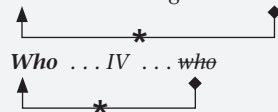
Q*: *How do you wonder **who** could solve this problem **how**?



Q_r: **What** is the manner by which someone, about *whose* identity you were wondering, could solve this problem?

15. A: The dog which **Mary** saw was scary.

Q*: *Who did the dog **which who** saw was scary?



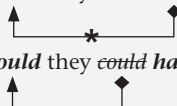
Q_r: **Who** is the person such that the dog *which* she saw was scary?

Relativized Minimality is violated by 14Q*, because **how** (= X) is related to **how** (= Y) by MOVE (a rule), where *who* (= Z) is an intervenor; X, Y, and Z are question words, hence of the same type. The string is thereby rendered ungrammatical. Similar considerations hold for 15.

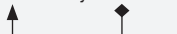
Relativized Minimality covers a set of facts that seem unrelated and yet fall under the same generalization:

16. A: They could have left.

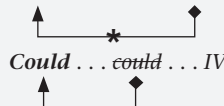
Q*: *Have they **could have** left?



Q_r: **Could** they *could have* left?



Have ... IV ... have



Could ... could ... IV

In 16Q*, X, Y=have and Z=could. The link between X and Y crosses the intervenor Z, and ungrammaticality follows. Yet in 16Q_r, X, Y=could, the relevant category Z=have is not an intervenor, and the result is a grammatical question. Surprisingly, then, Relativized Minimality covers such questions, too, and reveals itself as a constraint on representations with broad empirical coverage.

More Dependency Relations: *A Student Is . . . or A Student Are?*

The relation between positions in a sentence that is captured by MOVE turned out to be a central one in syntax, with many implications for the schema of the new cognitive syntax. It has also had ramifications for how humans process sentences in real time—ramifications that have been thoroughly

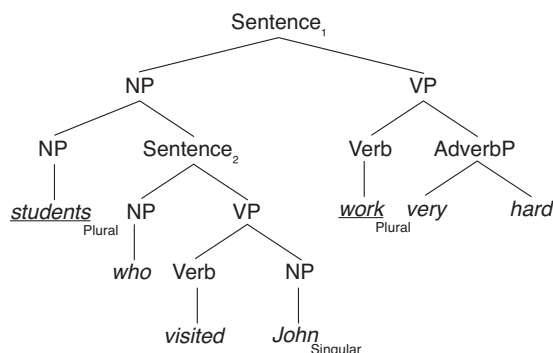
investigated. But MOVE is not the only dependency relation one encounters in sentences. These relations are important, especially in the context of LLMs, so I will elaborate a bit. First among them is the familiar person–gender–number agreement between a subject and a verb in English (and between a verb and its object in languages like Italian). Person, gender, and number are taken to be features that typically cluster together (hence the reference to “person–gender–number” agreement), and they manifest in many of the world’s languages. Even in a poorly inflected language like English, we say *This student is smart* but never **This student are smart* (here, I assume that if a student chooses *they* as their personal pronoun, then this sentence takes the form *These students are smart*). But what makes the starred one unacceptable? What rule or constraint does this sequence of words violate? Needed is a constraint that prohibits a lack of number congruence (**This student_{Singular} . . . + . . . are_{Plural}*). While this appears to be a simple formulation, we must be clear about what “. . .” can possibly be. That is, can any material that intervenes between a subject and a verb in English substitute for “. . .”? The answer is in the negative. Here is a small demonstration. It appears that “. . .” cannot always stand for *nothing*. That is, mere linear proximity between a noun and a verb does not ensure that they must agree in person, gender, and number. This can be seen in the contrasts below (where agreement is annotated with dashed arrows, to underscore that the relevant dependency relation may be a relation other than MOVE).

17. Subject–verb agreement ($N_{\text{Singular}} \dots N_{\text{Plural}} \dots V_{\text{Singular}}$)

a. **A student_{Singular} who visited us_{Plural} work_{Plural} very hard.*

b. *Students_{Plural} who visited John_{Singular} work_{Plural} very hard.*

Why aren’t the nouns closest to the verb agreeing with the verb in number? The answer lies in the hierarchy seen in figure 2.1. The embedded sentence, Sentence₂, is within the main sentence, Sentence₁, which describes the subject *student(s)*—it refers to those students who visited John and to no other student. The underlined verb should therefore agree with its (underlined) subject, which is linearly distant but structurally on a par. The adjacency between the noun *John* and the verb *work* is not a determinant in the subject–verb agreement rule.

**Figure 2.1**

Embedded nouns may not agree in number with the closest verb. While the noun *John* is adjacent to the verb *work*, the latter agrees with the distant *students*, because they are on a par hierarchically (parts of immediate constituents of Sentence₁).

Do note that while the agreement relation is subtle, it may not be related to MOVE, which governs the relation between a displaced element and some other position. By what we have seen thus far, agreement is a relation between two stationary words or phrases from different categories, hence potentially independent of MOVE.

In sum, we have seen the complexity of the agreement dependency, its critical dependence on hierarchical structure, and its distinctness from MOVE. People occasionally make agreement errors when they produce such sentences, but for the most part, these errors are immediately recognized as such. The rule is clearly followed by native speakers, who can readily acknowledge the ungrammaticality of 17a. Upon careful examination, the agreement relation seems to not be subject to the Relativized Minimality constraint, which leads to the conclusion that it is not MOVE but constitutes, rather, a distinct dependency relation that requires separate regulation.

I know that keeping track of all the notation I introduced is not easy. But syntax—that piece of knowledge that our cognitive system possesses—is complex, and setting constraints on its rules is even harder. We shall soon see how this recognized complexity has led to design changes in the model for our cognitive syntax.

Even More: Did *She* Look at *Herself* or at *Her*?

A knight and three princesses (and no one else) are standing in front of a large mirror, and the knight says “Look at your own image, please!” Sentence 18a is an accurate report of what happened, indicating that *herself* picks up its reference from its most closely neighboring noun phrase *every princess*. Curiously, this cannot be said of 18b, which may describe a scenario in which the knight’s request was that every princess should look at some other unmentioned female (though there are none in the image), not at herself. Again, the dashed arrows stand for a dependency relation that is potentially different from MOVE.

18. a. The knight asked *every princess* to look at *herself* in the mirror.

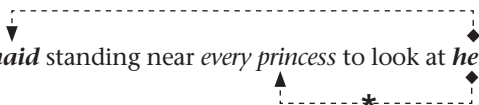


b. The knight asked *every princess* to look at *her* in the mirror.

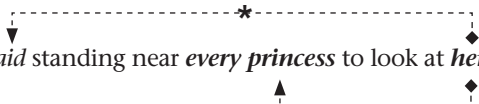


Now imagine that a (distinct) maid is standing near every princess, and the knight is telling each maid to look at her princess’s image in the mirror. The event cannot be described by 19a (as *herself* can only refer to *the maid*). However, 19b is an accurate report of what happened.

19. a. The knight asked *the maid* standing near *every princess* to look at *herself* in the mirror.



b. The knight asked *the maid* standing near *every princess* to look at *her* in the mirror.



Now this is really peculiar: each pair of sentences conveys roughly the same message, but the way native English speakers interpret the dependency between *herself/her* and neighboring nouns turns out to be subtle: in the pair 18, *herself* must refer to its nearest noun, whereas *her* cannot (hence the starred arrow), but it can refer to a more distant one; in the pair 19, the pattern is different. Curiously, the linear distance between *herself/her* and *every princess* is identical in all four cases: they are separated by the three words *to look at*. The possible and impossible readings, marked by the arrows and

stars, present a subtle, highly restricted pattern, which we cannot attribute to mere linear distance.

I will not burden you with a detailed syntactic explanation of this pattern (it can be found in any decent syntax primer) and will just say this: whatever the correct explanation of the peculiar distribution of pronouns and reflexives may end up being, it must rely on an analysis that assumes a structural hierarchy between parts of a sentence and highly constrained principles of shared reference between (reflexive and nonreflexive) pronouns like *her*, *herself*, and other nouns around them.

Interestingly, this set of restrictions on the referential dependency of pronouns interacts with other parts of syntax. This will become very important soon, when we come to the functions of “attention heads” in LLMs; they are entrusted with the task of linking positions within a sentence. Can they provide an account for the subtleties presented? We shall soon find out.

A Final Dependency: Ellipsis, at the Meeting Point Between Form and Meaning

Our last dependency relation displays fascinating interactions with the previous ones. At issue are incomplete sentences (a.k.a. ellipsis), in which the missing words—call this “a gap”—are not heard but nevertheless can be completed in the speaker’s mind, as determined by a previous sentence. What is special here is that this completion is highly constrained—not just any completion will do:⁷

20. a. John devoured a cake, *but* Bill didn’t devour a cake.
(No ellipsis → Bill devoured no cake)
- b. John devoured a cake, *but* Bill didn’t ____.
(Ellipsis → Bill devoured no cake)
- c. John devoured ice cream, *and* Bill ____ cake.
(Ellipsis → John and Bill devoured sweets)

Importantly, despite the missing pieces, no ambiguity is noted in any of these sentences. Although Bill’s negated action in 20b is not mentioned, and yet the sentence *only* has the meaning implied in parentheses. It cannot mean that Bill did not bake a cake or that he did not devour a steak. The same meaning uniqueness holds of 20c—John and Bill ate different foodstuffs, but the reconstructed meaning can place no verb other than

devoured in the blank, which is easily completed by the hearer. Whatever the process of meaning reconstruction in speakers' heads may be, its results in these cases appear to be strict. As with questions, we utter and hear such sentences daily and are able to complete the missing part easily and without error or hesitation. How can we formulate the meaning reconstruction algorithm correctly to ensure that the meaning of the missing piece is always unique? Representing it in a somewhat richer way, we replace the gap with a whole struck through Verb Phrase in 21b, (without its past tense marker, which remains on *didn't*), and a Verb in 21c:

21. a. John devoured a cake, *but* Bill didn't devour a cake.
 b. John devoured a cake, *but* Bill didn't ~~devour a cake~~.
 c. John devoured ice cream, *and* Bill ~~devoured~~ cake.

I can see a reader nodding and thinking: what's the fuss about? A simple rule that says *Complete whatever is missing according to the preceding clause!* would do. But wait: how exactly is a grammatical representation obtained? For 21b, the process must begin with the identification of the missing material in the second clause, which would somehow lead to the conclusion that the VP is missing; then, it would search for its parallel in the first clause ([_{VP} *devoured a cake*]), strip it off its tense, add an auxiliary with that tense (*didn't*); finally, copy the resulting material (*devour a cake*) into the abstract representation of the whole sentence. For 21c, the completion would require the identification of the verb as missing in the second clause, its copying from the first clause ([_V *devoured*]), and inserting it in the right place in the second clause.

This complex process is the minimum required for successful language use of such sentences, and it is a sequence of operations that speakers perform routinely and rapidly with novel ellipsis sentences. Moreover, writing a program that does all that reliably with any selection of words may not be easy at all, mostly because providing a general yet precise definition of the piece that must be copied into the ellipsis is difficult. The above description is incomplete, as well. To see that, consider the ellipsis sentence in 22. Unlike 20, this ellipsis contains an ambiguity: the pronoun in the ellipsis may connect to more than one antecedent. But most importantly, the ambiguity is highly constrained:

22. John walked his dog, *and* Mary and Bill did, too.

As any native speaker of English would agree, 22 may mean:

- Mary and Bill each walked his/her respective dog.
- Mary and Bill both walked John's dog.

But not just *any* antecedent for the pronoun will do—22 may *not* mean:

- Bill walked John's dog, and Mary walked Bill's dog.
- Bill walked John's dog, and Mary walked her own dog.
- Bill walked Mary's dog, and Mary walked her own dog.
- Bill walked Mary's dog, and Mary walked John's dog.
- Bill walked Mary's dog, and Mary walked Bill's dog.

Thus, on the one hand, ellipsis requires that the missing material would be filled by an exact copy that comes from the first clause; on the other hand, we see that sometimes, a constrained ambiguity is found. A theory of human linguistic ability must account for these highly systematic peculiarities, which we are all aware of automatically, as it were. I will stop short of providing an explanation, as it requires advanced syntax and semantics, but will only point out that, unlike simple cases of MERGE and MOVE, one must invoke interactions between the pieces in order to explain such odd phenomena.

The Cognitive Architecture of TGG—Take Two

Let's summarize this more technical chapter on linguistics. We have seen two generalized and highly constrained grammatical rules, MERGE and MOVE. The former builds hierarchical structure, and the latter establishes dependencies between positions in that hierarchical (tree) structure. We have also seen small hints of the need for additional operations, which may not involve displacement but do represent dependencies between positions in a tree (agreement, pronoun–antecedent dependencies, and the interactions seen in ellipsis). Reformulations of rules and constraints, with ever-increasing precision and generality, led over the years to predictions that could be tested, as further important linguistic phenomena were being uncovered, which motivated further theoretical modifications and extensions. In this way, two interfaces with syntax were established: one that makes contact with the logical and reasoning parts of our mental life, the other, with the physical qualities of speech. In some sense, this schema determines the flow

of information from an idea to an utterance. Each of the interfaces is implemented as a linguistic level of representation, with rules and constraints of its own. To better understand how this computational machinery works, what constrains it, what empirical discoveries motivate it, and what accomplishments it can be credited with, we need to display the context of the computational medium within which the rules operate. The changes I just described and motivated lead to a new cognitive architecture, depicted in figure 2.2.

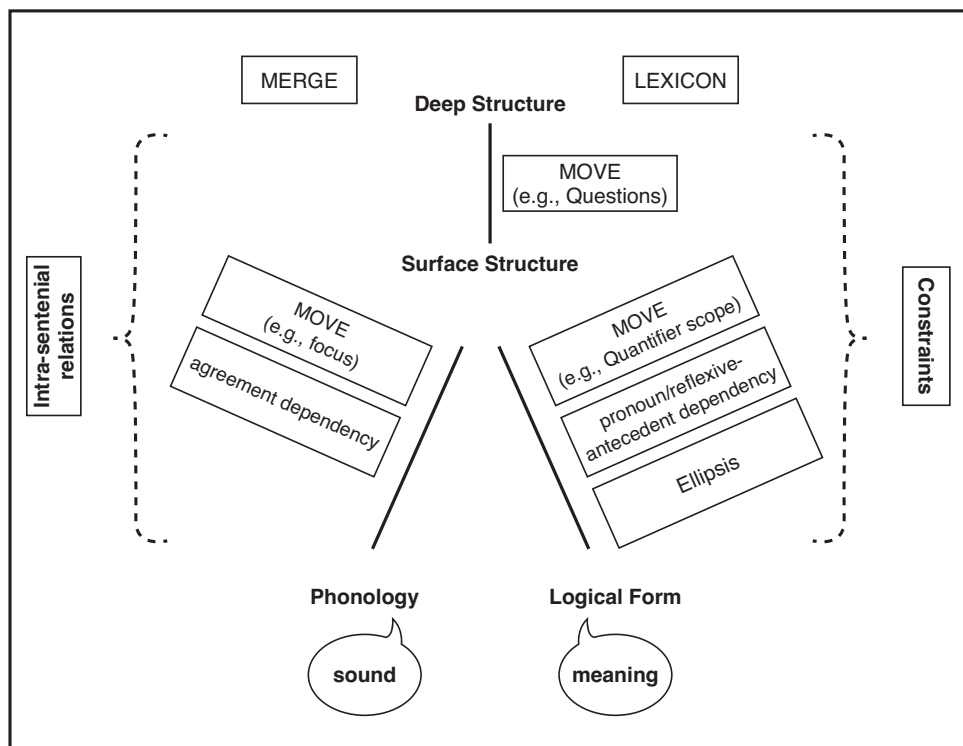


Figure 2.2

The Take Two model contains four levels of representation with which each sentence is said to be associated. Slanted boxes mark the rules that map from one level to another. MERGE applies to lexical entries drawn out of the lexicon. MOVE is an available mapping throughout; other rules are restricted to a particular mapping. The model is richer than its predecessor, yet it is much more constrained and has broader empirical coverage.

Figure 2.2 is a rich but, nonetheless, simplified and hence incomplete representation of the mental structure of syntax in its final, steady-state, adult form. Our mind and brain have representational media to accommodate four levels. Deep Structure is where a hierarchical representation is created when phrase structure rules (now formulated in a more abstract and generalized form as the rule MERGE) apply to the abstract representation of words, each projected from the mental lexicon. The output of this operation, a hierarchical “tree” representation with lexical entries as its “leaves,” is passed on to the Surface Structure level. This process may or may not apply a transformational rule (MOVE) to the representation. From this level, the representation is mapped onto the meaning-related level of Logical Form (a term borrowed from Bertrand Russell) and, at the same time, to a sound-related form, Phonological Form. In each mapping, certain operations (in boxes) may apply. Critically, their application is highly constrained and theoretically motivated. Each sentence is associated with the four complex representations. But these are just names. In fact, other linguistic frameworks differ in their nomenclature, but as far as this book is concerned, there appears to be general agreement on the central ideas, including an architecture with levels of representation, recursive rule systems that build up hierarchies out of labeled elements, formal mapping rules, and constraints. There also seems to be a consensus on the nature of the explananda, namely the linguistic knowledge and behavior of native speakers. At the core of the discussion are therefore the tight constraints on each representation and each mapping (meaning that not just *any* representation or transformational mapping is possible) and the phenomena they are designed to explain. We will soon see how this architecture plays a role in psychological and neurological studies of language.

If you got this far, you should now have at least a bit of a flavor of what the theory of syntax is about, and why the study of its intricacies is important if we want to understand human language abilities. This conclusion holds even before other important topics are discussed: left untouched in this brief review are relations that are at the syntax–semantics interface; semantics (the study of how single meanings compose into more complex ones); and pragmatics (the study of that which is unsaid but implied). Some of these issues will come up in later chapters.

Taking Stock

In this chapter we saw some details about the architecture of our cognitive syntax:

- Abstract theoretical concepts that a speaker allegedly possesses: not only transformations but also constraints on transformations
- The move to a highly abstract theory with a minimal set of rules that are highly constrained
- An example: Relativized Minimality—a constraint on *MOVE*
- Other dependency relations: agreement and pronoun–antecedent relations
- Complex puzzles that ellipsis structures reveal
- The new cognitive architecture of syntax—Take Two—Deep and Surface Structure, Logical Form, Phonological Form, and the rules *MERGE* and *MOVE* and their varied instantiations—and some facts these elements are designed to explain

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