Metaphysics Handouts
Revised
10 Nov 2026
07:40

Knowledge of language: Chomsky

OUTLINE. Chomsky's thesis is that to know a language is to be in a certain state of one faculty of the mind/brain, which he calls the 'language faculty'. The state that a speaker attains incorporates an 'I-language', an internalized language, so that knowing the language just is having a particular I-language. A grammar is a theory of that I-language, and if the structure it describes is really there in the speaker, then the grammar is true or false in the way that any other scientific theory is. This is why, of two I-languages that generate the same expressions, one may be the speaker's while the other isn't.

Along the way Chomsky argues against two rival views. The first is that knowledge of language is a practical ability to speak and understand. The second is that a language is an externalized object, a set of sentences or behaviours, which any grammar that enumerates it describes equally well. The step that carries the whole case is the argument from the poverty of the stimulus: children come to know facts about reference, and about what a sentence can't mean, which their data don't settle, and so the knowledge must come from something that is already in them.

We ask three questions: how much the structure-dependence example can show, whether knowledge of a language can be prised apart from the ability to use it, and what makes one of two equivalent grammars the speaker's own.

THE EVIDENCE NEVER MET. Chomsky's starting point is that children end up with rich and specific knowledge which their data don't determine. His examples include the facts about what 'them' can refer to in the text's sentences (2) and (3), and the meaning of its sentence (7), all of which speakers know without training or relevant evidence. His most familiar example, however, is the formation of questions, and it's the one that he works through step by step.

Example 1. Imagine a scientist watching a child who has learnt to turn statements into questions, as in these two pairs:

the man is tall	→	is the man tall?
the book is on the table	→	is the book on the table?

The scientist's first hypothesis is that the child has picked up the simple rule H1, and the hypothesis to which the evidence will drive him is the rule H2:

H1: Scan the sentence from the left and move the first 'is' to the front.

H2: Parse the sentence into phrases, and move to the front the 'is' that follows the first noun phrase.

Both rules yield the two questions above. They come apart on a statement in which 'is' occurs twice, and the arrows show what each rule makes of the statement (11a-1):

(11a-1)		the man who is tall is in the room
(11a-2)	$\xrightarrow{\text{H2}}$	is the man who is tall in the room?
(11a-3)	$\xrightarrow{\text{H1}}$	is the man who tall is in the room?

H1 predicts that a child will turn (11a-1) into (11a-3), and yet children who meet (11a-1) for the first time produce (11a-2) and never (11a-3). Their rule must therefore be H2, which Chomsky calls 'structure-dependent' because it applies to the sentence as parsed into phrases, whereas H1 applies to it as a bare string of words. By any reasonable standard H2 is far more complex than H1, and yet nobody trains children to use it. Chomsky adds that a person can go through a good part of life without meeting the relevant evidence, which would be a question like (11a-2), so that everything they've heard fits H1, and that they'll still use H2 without hesitation. He concludes that the preference for structure-dependent rules isn't learnt, and that it belongs instead to 'universal grammar' (UG), the equipment that the language faculty has before any experience.

Here is the difficulty: the argument takes two steps for granted. The first step is the ranking of the two rules by complexity, which Chomsky rests on the point that phrases are abstract, in the sense that nothing in the sound of a sentence marks where a phrase begins or ends.

Example 2. Suppose you had to teach the question rule to someone who already hears (11a-1) as a subject, 'the man who is tall', followed by a predicate. For that learner H2 comes to a single instruction, 'front the "is" after the subject', whereas H1 asks them to flatten a structure they've already built and then to count across it. So for a learner who starts from phrases H2 is the simpler of the two rules, and the scientist's ranking is the ranking of someone who processes a sentence as a bare string of words.

The second step is the claim that the relevant evidence may never arrive, which is a claim about what children hear. The note that Chomsky attaches to it cites only one empirical study, by Crain and Nakayama, and that is a study of what three-to-five-year-olds produce and not of what they hear. What survives is the observation that no child passes through a stage of making errors like (11a-3). For the argument to go through on that basis, it's enough that children avoid the error before the relevant evidence reaches them, even if it does reach them later. But whether they get the rule right before or only after they've heard such questions is a matter of timing, and the text leaves it open.

There is a further difficulty even when the example works, because it then shows at most that learners represent sentences hierarchically, as phrases inside phrases. Chomsky's own test for a distinct language faculty is that the properties of its initial state be specific to language, either one by one or in the way they're organized and combined. Hierarchy on its

own isn't obviously specific to language, since vision and the planning of action have it too. So the example has to show something about how hierarchy is combined with the other properties.

Nor will hierarchy alone account for what speakers know, because no hierarchy of phrases predicts that the text's sentence (13) has a meaning while its sentence (14) is gibberish. The upshot is that, if the example is to go Chomsky's way, he has to hold that the child's bias is a content and not merely a format: it isn't just the hierarchical form in which sentences are represented, but a principle that mentions linguistic categories. That is something which the formation of questions can't show.

KNOWLEDGE THAT OUTRUNS ABILITY. Chomsky denies that knowledge of a language is a practical ability to speak and understand, and he gives two arguments. The first argument is from ordinary usage, which he thinks gets the matter right: a speaker's ability can improve, decline, be lost through injury and be recovered, all without any change in what the speaker knows. 9

Example 3. Suppose that Jones takes a course in public speaking and comes out speaking and understanding better, although his knowledge of English hasn't changed. The ability theorist can't put it that way, and has to say instead that Jones has improved his ability_1 to use his ability_2, where ability_1 waxes and wanes while ability_2 stays put, even when we can't tell what it entails. Chomsky's complaint is that the two uses of the word are hardly more than homonyms, and that ability_2 has simply been handed every property of knowledge. 10

Chomsky grants that we do speak of 'abilities' that we can't exercise, as when we say that a swimmer whose hands are tied 'is able' to swim, but he insists that the cases he is discussing aren't of this sort.

Against Chomsky, however, the speaker whose injury recedes and whose lost ability returns is exactly a case of this sort. The reason is that an ordinary ability is already a standing capacity, distinct from its exercise, and so ability_2 is no neologism but simply ability in the ordinary sense. 9

Chomsky's second argument is better, and it turns on the text's sentence (14). We're able to give that sentence either of the readings that analogy with other sentences suggests, and yet we recognize that our language provides neither of them. What we have, then, is knowledge of what the sentence can't mean, and not a lack of skill. 12

In a later passage Chomsky asks how far such knowledge reaches. Take John who follows the rule R = 'Adverbs don't separate a verb from its object.' According to Chomsky, John knows that R, although he doesn't know that R holds or that it's a rule of his language (English). Suppose now that R follows from a deeper rule R' = 'Case assignment in English requires strict adjacency.' Then, Chomsky says, John knows that R' too, and our reluctance to say so comes only from our own unfamiliarity with the notion of Case, which is irrelevant to the state that John is in. If the word 'know' resists this use, Chomsky is content with 'cognize' instead, and he adds that the question isn't ultimately an important one. 266 267

Notice that these two passages are in tension with each other. The paragraph about Jones convicts the ability theorist of departing from ordinary usage, whereas the passage about John declares such a departure harmless when it's Chomsky's own. Moreover, ability_2 was never a worse coinage than 'cognize'.

Chomsky has a reply, because his real charge against the ability theorist is a different one. The reduction of knowledge to ability was meant to dissolve the puzzles about knowledge by restating them in dispositional terms, which are terms closer to behaviour. His charge is that it dissolves nothing, since what the knowledge is, where it comes from and how it's used are as hard to say under the new names as they were under the old. But the same question can then be turned on 'cognize': what does it explain to call the speaker's state knowledge? 10 11

The best answer open to Chomsky is that the state has the structure of a system of rules and assigns a status to every sound that a speaker meets, which is something no disposition does. The cost of this answer is that 'knowledge' now names whatever the best theory attributes to the state, so that John knows that R' although he has no route to the concept of Case. 26

Remark 1. Russell's *KAKD* militates against exactly this possibility, that someone should know a proposition while lacking a concept that occurs in it. Russell's principle is that every proposition we understand is composed wholly of constituents with which we're acquainted, and his reason is that nobody judges without knowing what they're judging about. Universals are among the constituents, and a universal of which we're aware is a concept, so for John to judge that R' he must have Case assignment and adjacency before his mind, just as he has verb and object before his mind when he judges that R. But he doesn't have them before his mind, for two reasons: nothing has brought the universal Case assignment before him in the way that enough yellow things bring yellow before us, and he can't describe it by means of concepts he already has, since he has never heard of it. It follows that John can't even understand the proposition that R', let alone judge it or know it. KAKD 154 KAKD 150

Kripke's 'sceptical paradox' presses Chomsky from the other side. On Kripke's view competence is normative and not descriptive, which is to say that it settles what a speaker should say and not merely what he does say. A brain state, by contrast, settles only what John does say, and the trouble is compounded when the rules are tacit ones that the scientist has to reconstruct. The upshot is that the position Chomsky needs is that 'knowledge' is a theoretical term whose application is read off the theory of the speaker's state, just as 'mass' is read off physics, and if he takes that position he has to give up the argument from usage. 28

TWO GRAMMARS, ONE LANGUAGE. A grammar is a theory of the I-language that may be true or false. Its statements are about structures of the brain, described at some level of abstraction from the mechanisms, in the way that a nineteenth-century theory of valence described the elements in abstraction from the mechanisms that account for their properties. This is why, of two I-languages L and L' that generate the same expressions, L may be the speaker's while L' isn't. Quine denies this and Lewis doubts it. Their thought is that where two grammars are 'extensionally equivalent', that is, where they generate the same set of sentences, there is nothing to make one of them correct and the other not. 22 23 20

Chomsky takes it that their model is a formal system such as arithmetic, where the well-formed formulas are given in advance and the rules that generate them can be picked at will. His reply is that Quine and Lewis have the story backwards. What's given to the child is a finite array of data, from which the child's mind builds an I-language, whereas the externalized set of sentences, the 'E-language', is a derivative artefact.

The difficulty with this reply is that the claim that a grammar is true or false was conditional on such a notion of structure existing, and that is just what Quine's indeterminacy and Kripke's sceptic deny, namely that there is a fact in the speaker which makes one of two equivalent rule systems his.

Remark 2. Chomsky doesn't think Quine showed any novel indeterminacy beyond the normal underdetermination of theory by evidence, and he finds Quine's formulations inconsistent. Hmmm. These pages don't answer Kripke's sceptic.

Consider how the matter looks from inside Chomsky's view, on which L is an abstract entity to which a speaker H stands in a relation R, the relation of knowing or having a language. A grammar's statements, however, are about structures of the brain, which are specific things in the world. So if L and L' are two abstract objects, only H's brain could settle which of them H is related to. The trouble is that the mechanisms of the brain aren't known, and on Chomsky's own account linguistics joins the natural sciences only to the extent that they're discovered. So the fact that makes L the speaker's language is, for now, a promissory note on neuroscience.

Example 4. Suppose you're shown two grammars of English that agree on every sentence and every meaning, together with a speaker who sees immediately that the text's sentence (14) is gibberish but needs a moment to work out its sentence (13). Chomsky wants the 'parser', the device that assigns a structure to each sentence a hearer meets, to mirror difficulties like this one, so the speaker's hesitation might seem to tell us which grammar is his. But the parser adds an organization of memory and some heuristics to the rules of the I-language, and it's meant to depart from the I-language in just such cases. It follows that the hesitation at (13) favours one grammar over the other only if the rest of the parser is held fixed. The sentences alone, on which the two grammars agree, can't tell them apart.

The cost of letting the brain settle the matter is that the abstract object L drops out in favour of the 'steady state' of the language faculty from which it was abstracted, so that realism about a grammar becomes realism about a brain state under a description.

The Quinean can now come back. Evidence that goes beyond the sentences themselves, such as the speaker's hesitation, is finite too, and grammars that fit all of it will still diverge on cases that nobody has tested. If that is only the normal underdetermination of theory by evidence, as Chomsky holds, then a grammar of an E-language is only normally underdetermined too. Chomsky's asymmetry between the two then shrinks to a difference in where the facts are placed, in the speaker or in the set of sentences. The Quinean, for his part, must then say why physics, which its evidence underdetermines in the same way, gets to keep its facts.

Kripke's sceptic presses a sharper objection, since he asks what fact about a speaker could make it the case that the speaker follows one rule and not another. The speaker's dispositions are the one candidate for such a fact that the text considers, and Chomsky rules them out for reasons of his own. What remains is a state of the brain, but a brain state is finite and it sometimes malfunctions. To say which of its outputs are errors, we'd have to appeal to the grammar that it embodies, and which grammar it embodies is the very question at issue. The upshot is that, for Chomsky to win here, he has to hold that the brain has exactly one system of rules in it, so that the question 'which grammar?' is like the question 'which valence?', and is settled whether or not we ever find the mechanisms.