

*Language Issues
in Literacy and
Bilingual/Multicultural
Education*

Edited by
MASAHIKO MINAMI
BRUCE P. KENNEDY

1971



Reprint Series No. 22
HARVARD EDUCATIONAL REVIEW

Essay Review

Toward a Realistic Theory of Language Acquisition

JAMES PAUL GEE
University of Southern California

LANGUAGE LEARNABILITY AND LANGUAGE DEVELOPMENT
by Steven Pinker.
Cambridge: Harvard University Press, 1984. 435 pp.

Roger Brown, one of the founders of the modern study of child language, refers to Steven Pinker's *Language Learnability and Language Development* as "a landmark of psychological science." Though this remark undoubtedly partakes of some typical dust-jacket hyperbole, there is a sense in which Pinker is in at the beginning of something new. While studies of child language have been for the most part descriptions of what children do, Pinker proposes a theory of language acquisition that explicitly accounts for how children acquire language. That is, he makes a distinction between the study of child language and the study of language acquisition. Although one could argue with almost every proposal Pinker makes, the importance of his book lies not in the degree to which it is right or wrong but in the task he has begun.

Pinker's book is one that anyone interested in child language development ought to know about, though perhaps only a few need read it. Learnability theory represents a new discipline that, if successful, will change the character of the field of child language. It will do this in much the way that Chomsky's work changed the area of language study, even for those who knew little of or disagreed with it, by giving linguistics a formal foundation and by setting new goals and questions for the field. Learnability is in that unique stage of a discipline where anyone could easily read its entire literature, one that generative grammar and the modern study of child language were in not all that long ago. This stage will not last long if "learnability" works, in the sense of giving to the area of language acquisition a new formal foundation and a new set of goals and questions.

While Pinker's book is much less technical than the other major length study in learnability (Wexler & Culicover, 1980), it still requires a fair bit of sophistication in linguistics. Those seriously interested will find it one of the best places to start; others

may want to wait for the more accessible treatments that will undoubtedly follow should learnability theory fulfill its promise.¹ In this review I will concentrate on the historical and intellectual developments leading to the birth of learnability theory. In turn, I hope to place Pinker's book in part of its historical and logical context and then discuss some of its leading claims in terms of this context.

Chomsky's Contribution to Language Acquisition

If Chomsky gave rise to a revolution in linguistics, it was because he made "the logical problem of language acquisition" the key problem in linguistic theory (Chomsky, 1965). He insisted that linguistic theory had to go beyond "descriptive adequacy." It is important, therefore, to recognize the limitations of it. For Chomsky, descriptive adequacy meant that linguistic theory could supply a descriptively correct grammar for every natural language, that is, a grammar that correctly describes the native speaker's intuitions about the linguistic properties of the sentences of their language. The problem is that many different but equally good descriptions are available for each language. To take an example that we will return to later, consider the well-known English Dative Alternation. With many different verbs, such as *give*, *send*, *tell*, *bring*, and others that mean roughly "to transfer something," one can use either of two orderings of the direct object and the indirect object. One can order the direct object (the thing transferred) before the indirect object (the person receiving the object) as in 1a, or one can use the reverse order, placing the person receiving the transferred entity first, as in 1b. Thus, there are two nearly synonymous ways to use these verbs.

- 1a. Mary told the magic words [entity transferred] to John [recipient].
- 1b. Mary told John [recipient] the magic words [entity transferred].

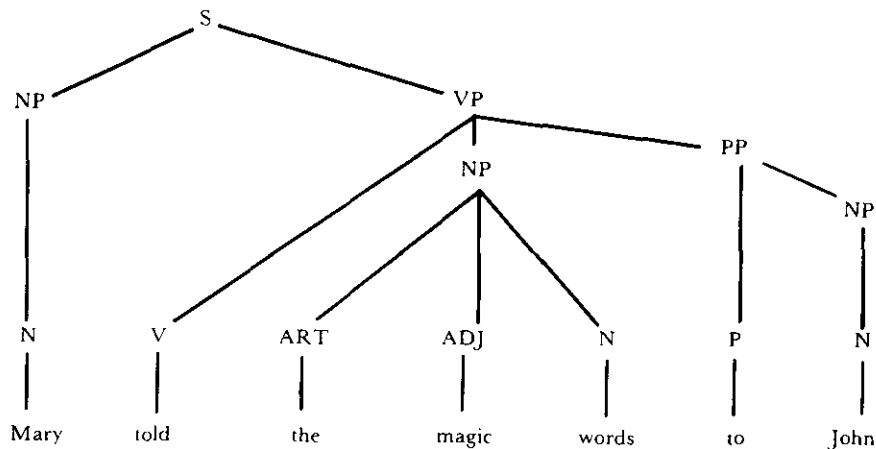
Somewhat paradoxically, however, other verbs that also appear to mean "transfer" will not alternate this way, for instance *donate* or *say*.

- 2a. Mary said the magic words to John.
- 2b. *Mary said John the magic words.²

There are any number of ways to formally describe these facts. One is the classic transformational approach. We set up a deep or underlying structure for the basic, "unmarked," normal sentence type, example 2a in this case. This underlying structure is exemplified in 3a by a "phrase structure tree," which is simply a way to formally represent the syntactic structure of a sentence. This tree is "generated" by a set of phrase structure rules together with a lexicon (dictionary) as in 3b. Phrase structure rules are formal rules that stipulate the structure of possible sentences, or, in other words, parse a sentence (the sort of thing one did in grammar school on the blackboard). For instance, the first rule in 3b says that a sentence in English is made up of a noun phrase (NP)(its subject) and a very phrase (VP)(its predicate). This VP is in turn made up of a verb followed optionally by an object and a prepositional phrase (PP)(for example, "Give the book to the boy"), while an NP can be made up of a

noun optionally preceded by an article (ART) and adjective (ADJ) and optionally followed by a prepositional phrase.

3.a



3b.

S → NP VP
 VP → V (NP) (PP)
 NP → (ART) (ADJ) N (PP)
 PP → P NP

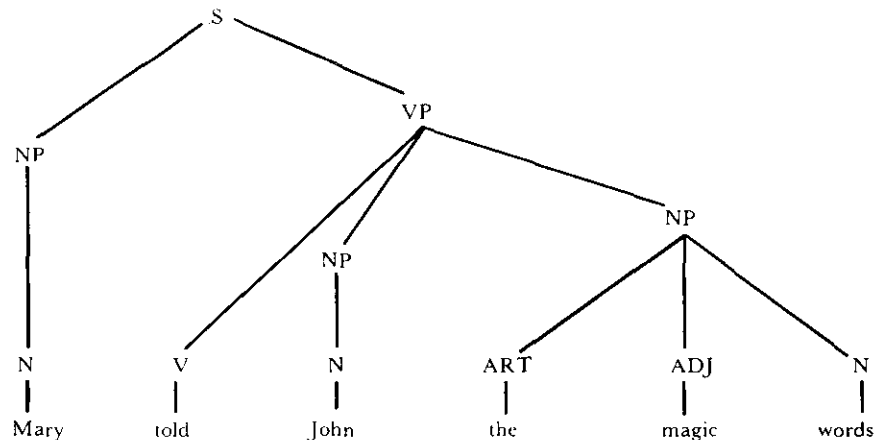
LEXICON: <i>tell</i> =	Verb
	Must be followed by an NP (Direct Object) and can optionally also be followed by a PP headed by <i>to</i>
	Pronunciation: /tel/
	Meaning: "tell"

to = Preposition

We then set up a transformational rule, known as the Dative Transformation, that will change the tree in 3a into an appropriate tree for "Mary told John the magic words" (see 3c). What we are doing, in basic terms, is stipulating that the order represented in 3a, "Mary told the magic words to John," is the basic one, and that the language contains a general rule that stipulates that this basic order can be

changed or transformed into another less basic or more special purpose order, namely, the order represented in 3c.

3c.



Transformations are stated in a formal notation (which we can ignore here) that stipulates exactly how one tree can be changed into another. Under this approach, verbs that do not undergo the Dative Transformation, such as *donate* and *tell*, will be marked in the lexicon as negative exceptions to the rule of Dative (as [- Dative]). Thus, the rule will not operate on sentences that contain these verbs and they will appear in only one ordering.

But there is another way to describe the same data. This is the problem with seeking descriptive adequacy only; we have too many ways to describe the same thing without being able to choose between the competing descriptions. We have to list in the lexicon, for each verb, its privileges of occurrence. For instance, we have to say that a transitive verb such as *love* requires an object ("John loves Mary" but not "John loves"), and an intransitive verb like *die* does not allow an object ("John died" but not "Mary died John"). Thus, why not stipulate in the lexical entries for verbs like *tell* or *give* that they allow either a following NP and PP (as in 1a) or a following NP NP (as in 1b). No need for a transformation. One might say, however, that we were missing the generalization that many verbs meaning "transfer," but not all, have these two possibilities. To capture this generalization we set up a lexical redundancy rule. Just as English allows many, but not all, adjectives to add *un* and mean "not adjective" (for example, *happy* and *unhappy*, but *silly*, though not *unsilly* [see 4a]), so we add a word formation rule to the lexicon that says that there is one version of many verbs meaning "transfer" that is followed by NP PP and another version, related to it, that is followed by NP NP (see 4b).

- 4a. Lexical rule: For many adjectives, you can prefix "un" and get an adjective meaning "not adjective."

- 4b. Lexical rule: For many verbs meaning "transfer," which have the possibility of being followed by NP PP, you can add another verb just like it in sound and meaning that can be followed by NP NP.

The actual lexical entries will, of course, have to say whether rule 4a or 4b is true for that word, since they are true of many but not all words they could apply to. Rules 4a and 4b just express tendencies or limited generalizations in the language. The two accounts of the Dative Alternation already provided are both perfectly good descriptions of English as far as they go. They are both descriptively adequate, and it would be a matter of splitting hairs to choose between them. In fact, there are many other such descriptions. The two I have given both use syntactic terms such as NP and PP. Accounts are possible that use semantic terms like *agent* and *patient*, and alternative accounts are available that use discourse terms like *topic*, *new information*, and *old information* (Givon, 1984). As far as description goes, there are no grounds, apart from particular uses we might make of the descriptions, on which to choose one or the other of the alternatives. Similarly, in mathematics there are many different ways to mathematize a problem (such as in terms of different branches or kinds of mathematics), all equally correct, apart from the particular purposes one might have on particular occasions.

Chomsky demanded more of a linguistic theory than descriptive adequacy. He demanded "explanatory adequacy"; in so doing, he construed explanatory adequacy to be the "logical problem of language acquisition" (Baker & McCarthy, 1981; Chomsky, 1965, 1984). To gain explanatory adequacy, linguistic theory must select one and only one descriptively adequate grammar for each human language on the basis of "primary linguistic data," which are the data available to a child in the early stages of language acquisition.

Let me say first what it means to call this a "logical problem." Chomsky ignored the real-world fact that children acquire language over time, perhaps going through certain stages (much as Galileo ignored the fact of friction in formulating a law of falling bodies). He conceived of the problem of language acquisition in abstract and logical terms. A child is born with some initial state of tacit or unconscious knowledge about language; perhaps this state is zero (as a rigid empiricist might argue), perhaps it is relatively rich (as a linguistic rationalist might argue). Call this state S_{INITIAL} . The child is exposed to certain finite data about the language, the "primary linguistic data." On the basis of these data and whatever is known at state S_{INITIAL} , the child induces a grammar of the language to which she is exposed; where this grammar represents a final state of knowledge of the language, call it S_{FINAL} .³ (This is what the adult knows about the language if we ignore the fact that people continue to learn about language throughout their lives, and concentrate on the fact that even older children have an immense mastery of language that far exceeds what they had as younger children.) The problem for linguistic theory, then, is to supply not just a descriptively adequate grammar but the unique grammar that the child arrives at when exposed to the primary data— S_{FINAL} —given whatever knowledge is in S_{INITIAL} .

If children just memorized the data and S_{INITIAL} was empty, then S_{FINAL} could be

a memorized list of the sentences of the language, and the problem would be trivial. But, of course, they cannot do this given the fact that there are an infinite number of possible sentences. At a slightly less trivial level, we might assume that the child memorizes not sentences but patterns, and draws analogies. Thus, S_{INITIAL} would be empty, save perhaps for some very general knowledge about similarity and difference. This doesn't work either, of course. For example, from the occurrence of both *sleepy* and *sleeping* before nouns ("the sleepy girl" and "the sleeping girl"), the child should draw the correct conclusion that they are both adjectives; then from the equally correct observation that *sleepy* occurs after the verb *seems* ("The girl seems sleepy"), the child should draw the conclusion, by analogy, that "The child seems sleeping" is grammatical. But it isn't. From the correct observation that *happen* and *seems* occur in similar syntactic environments ("The girl seems to be sleepy." "The girl happens to be sleepy"), and the observation that one can say "The girl seems sleepy," the child by analogy should conclude that "The child happens sleepy" is grammatical. But it isn't. Analogies based on recognized patterns are a poor basis on which to acquire a language.

One of Chomsky's most important contributions to linguistics was to show that the logical problem of language acquisition is nontrivial (in fact, solving it may be beyond human science-forming abilities). In order even to approximate a solution, one must make fairly rich assumptions about the initial state S_{INITIAL} which amounts to making substantive assumptions about the human biological capacity for language or, put in other terms, about the innate knowledge children bring to the language acquisition task.

Let me give an example which contains the sorts of assumptions about S_{INITIAL} we might have to make and which is also indicative of one type of research methodology in theoretical linguistics. Everything in the primary linguistic data that any child is exposed to would indicate that the Content Question Rule below is true of English (Ross, 1967), though it turns out in fact to be false. That is, if all that children paid attention to were the data they are exposed to in the course of language acquisition, they ought to grow up using this rule. They do not, however, because they "know" (in some sense) that it is wrong. Then the question is, Where did this knowledge come from?

CONTENT QUESTION RULE: One can form a grammatical content question from any declarative sentence by changing any one of its NPs into a question word or phrase and placing it at the front of the sentence (plus inverting its first helping verb, or adding a form of "do" if no helping verb is present, when one has questioned anything but the subject). For example:

- 5a. John likes *Mary*. → *Who(m)* does John like?
- 5b. *John* likes *Mary*. → *Who* likes *Mary*?
- 5c. John thinks *Bill* likes *Mary*. → *Who(m)* does John think *Bill* likes?
- 5d. John thinks *Bill* likes *Mary*. → *Who* does John think *likes* *Mary*?
- 5e. *John* thinks *Bill* likes *Mary*. → *Who* thinks *Bill* likes *Mary*?

Nothing in the primary data (nothing, in fact, that any English speaker—child or adult—has ever said or heard) contradicts this statement. Nonetheless, as 6a and 6b show, it is false, since not just any NP can be questioned.

- 6a. John recognized the truck that hit *Mary*. →
6b. **Who(m)* did John recognize the truck that hit?

Since nothing in the primary data can explain why 6b is ungrammatical, that is, why adults in S_{FINAL} know that it is ungrammatical, then we must attribute some knowledge to the $S_{INITIAL}$ to account for this fact (since all we have to play with, to account for S_{FINAL} , is $S_{INITIAL}$ and the primary data). Let us say we attribute to S_{FINAL} the knowledge that relative clauses (such as “that hit *Mary*” in 6a) are islands—nothing can “leave” them. Nothing in a relative clause can move to form a question in English because to do so would be to take something out of the relative clause (for example, the object of *hit* in 6a) and move it to the front of the sentence (in the form of *who* in 6b). Talk of movement here is merely a way of representing the fact that English word order is, in the normal case, subject-verb-object, while if we pose a question about the object it appears at the front of the sentence. Since $S_{INITIAL}$ is the state of the human being prior to language acquisition, it is the same for all humans, regardless of what language they are acquiring. Thus, when we attribute something to $S_{INITIAL}$, as in this case, we are tacitly making predictions about language universals: If a language has relative clauses and if it has any rules that *move* things, then it cannot move them out of relative clauses. Thus, we can check our hypothesis about $S_{INITIAL}$ by looking at other languages, and in turn refine the hypothesis.⁴ If our hypothesis is confirmed, the correct grammar for English is one that has the Content Question Rule as part of it (which is all that could have been learned from the primary data of English). The restriction displayed in 6a and 6b is part of the innate knowledge the child brings to the task of language acquisition ($S_{INITIAL}$) and thus also a language universal (not part of the grammar of English alone).

The approach to language acquisition that Chomsky gave rise to, then, was one which, for purposes of scientific theory building, ignored the real-world circumstances of language acquisition and assumed that language was acquired simultaneously on the basis of the primary data. This was done in order to formally demonstrate that the logical problem of language acquisition is not solvable solely by induction from the primary data, that the solution involves nontrivial assumptions about $S_{INITIAL}$, and thus about biology. This is, in itself, an important result. Indeed, all science involves such idealizations. (For example, the law of gravitation says that all bodies fall at an equal rate. This, however, is false unless we ignore friction caused by air particles in our atmosphere. If we ignore friction we get an elegant and general law that holds not only in our world but throughout the universe.) It is possible, of course, that the idealization we have made is wrong in the sense that the problem would have a quite different solution if we added back the real-world variables. This

is not true for the law of gravitation. Many have claimed, however, that Chomsky's idealization is wrong, although no one has come close to showing this. Moreover, Chomsky has always assumed that S_{INTERNAL} must be knowledge-specific to *language*, not knowledge—however rich—about learning, problem solving, or processing in general. Again, though many have claimed that this is wrong, no one has come close to showing convincingly that it is (see Piatelli-Palmarini, 1980).⁵

One can, however, go beyond the logical problem of language acquisition and ask, Can the grammar arrived at by a linguistic theory in search of explanatory adequacy (for example, English) be successfully incorporated into *realistic* models of language acquisition, comprehension, and production? ("Realistic" here means a model that does not ignore or put to one side the actual circumstances of language acquisition, language production, and language perception.) This is something like asking what happens to the law of gravity when we consider its operation in a world with friction. Chomsky (1965) referred to this problem as the question of feasibility.

Psycholinguistics initially had little success in incorporating generative-transformational grammars (based on the so-called *standard theory* of Chomsky, 1965) in realistic "on-line" (real-time) models of language production and processing (Fodor, Bever, & Garrett, 1974). For instance, linguistics once proposed that a sentence like "Wouldn't John be hurt by that remark?" was derived from an underlying structure for a sentence like "That remark would not hurt John" by a series of transformations that progressively deformed a structure for the latter sentence into one for the former: Passive ("John would not be hurt by that remark"), Negative Contraction ("John wouldn't be hurt by that remark"), and Question Formation ("Wouldn't John be hurt by that remark?"). It was then hypothesized by psycholinguists that the more transformations that applied in the derivation of a particular sentence, the longer it would take to comprehend it (on the theory that one had to "unwrap" the transformations to recover the deep structure, the level of the grammar that determined semantic representation). This turned out to be false, as shown by experimental research. In addition, it turned out to be difficult to assign any role at all to transformations in the process of actually producing or comprehending a sentence, though they did capture nicely the static knowledge native speakers had about the ways various types of sentences were related to each other (for example, "John would be hurt by that remark" is the passive of "That remark would hurt John").

Linguists responded to this problem in several different ways. One response was the development of nontransformational generative grammars that were constructed, in part, with an eye to their possible incorporation into realistic models of language production and comprehension. The most successful of these responses to date is the theory of Lexical Functional Grammar (LFG) developed by Bresnan (1982) and her colleagues. On the other hand, Chomsky and his colleagues never accepted the psycholinguists' interpretation of transformations. They stressed that generative grammar was meant to capture the knowledge native speakers had about their lan-

guage, not how that knowledge actually was put to use. They insisted that their principles would have to be supplemented with various additional psychological principles to handle the way speech is produced and comprehended. Chomsky and his colleagues have, however, made the theory of generative-transformational grammar progressively more abstract. They have argued for a model of science in which the relationship between principles in a theory and the data they attempt to explain is quite abstract, often requiring a long deductive chain of reasoning to relate them. This has led to a theory with a much more constrained use of transformations (Chomsky, 1981, 1984), but one that is also much less accessible than Chomsky's earlier work. The theory, called Government and Binding (GB), is, in fact, probably no less a candidate than Bresnan's for incorporation into a realistic model of language processing though little such work has been done to date.

The concern with incorporating grammatical theory in a realistic model of language acquisition is fairly new. It received its impetus from Hamburger, Wexler, and Culicover's pioneering work on formal learnability theory (Hamburger & Wexler, 1975; Wexler & Culicover, 1980), and, for many linguists, was given a big push by Baker (1979). To see some of the issues that arise, let us consider the Dative Alternation again. When we consider language acquisition in a realistic setting, we discover that there are a number of problems with the primary data, apart from the fact that they are finite. The biggest problem is the lack of "negative evidence." Children can assume that, more often than not, sentences they hear in the primary data are grammatical, but they receive no useful indication of what count as ungrammatical sentences. Children are not, in general, corrected when they make errors, and it is certain that they can acquire language in the absence of any overt correction (see Baker, 1979, for a discussion of this point and citations of the relevant literature). On the other hand, children hear errors in the primary data (speech disfluencies, repetitions, self-corrections, false starts, and so forth), but they are not told that these are errors. More generally, many ungrammatical sentences are never exemplified in the data, and children are never explicitly told that these sentences are ungrammatical. Consider in this light the transformational view of the Dative Alternation (see 3a-c). A child could learn that there was a Dative Alternation for verbs like *tell*, *send*, and *give*, simply by hearing both sorts of sentences in the data. But how would a child learn that the rule does not apply in the case of similar verbs such as *say*, *donate*, and *report*, and therefore that these verbs have to be marked in the lexicon as [- Dative]? The child is never told that these verbs are ungrammatical in the NP NP construction ("Mary reported John the story") and is not corrected if she uses such a construction. At this point in the story there is some debate about the empirical facts. Some (for example, Baker, 1979) claim that children never overgeneralize the dative, that is, say things like "He said me the magic words." Then the question becomes, Why don't they? since they do overgeneralize in other cases (saying *foots* for *feet* or making up novel causatives like "Mommy sweater me"). Others claim that children do overgeneralize the dative (p. 312). Then the problem becomes,

How do they ever recover from the overgeneralization? Why wouldn't they just add this verb to the dative rule and change the language? Such things do happen. For instance, *bead* used to mean *prayer*, but children, seeing monks "saying their beads" while handling a rosary, took *beads* (prayers) to mean *beads* (rosary). Further, adults will accept innovations like "I'll xerox you a copy of the paper."

Obviously the problem of language acquisition is even more acute than we had assumed. In the case of the ungrammatical content question in 6b, there was absolutely nothing in the data to indicate that it was ungrammatical; we simply had to assume a piece of innate knowledge at work. Yet in the case of the Dative Alternation rule, there is evidence for it in the primary data, but the rule still does not look learnable since there is no way for the child to learn its limits. Further, it is hard to see what innate constraint could be in operation here. We are forced to talk about an actual procedure which a child could use to acquire the language in this case, not about an abstract constraint on the operation of a rule.

The second, lexical approach to the Dative Alternation we described above can be the basis of such an account, if in fact children do not overgeneralize the dative rule. When a child hears a sentence such as 1a, she will add to her lexicon the information that *tell* can occur in the environment [___ NP to NP]; when she hears a sentence such as 1b, she will add the additional information that it can occur in an environment like [___ NP NP]. Since for *say* children will only hear sentences like 2a, they will only add [___ NP to PP] to that entry, and never [___ NP NP]. This is essentially to say that children are conservative learners, at least in this case. Of course, it still does not account for the overgeneralizations of the Dative Alternation that some children do appear to make, nor does it explain how they recover from them. This in fact makes the problem nearly intractable. It does demonstrate, however, that considerations about learnability in real time have implications for the shape of grammars.

Pinker's Contribution to Language Acquisition

Pinker's goal is the construction of a theory of acquisition that is realistic—in the sense that it considers language acquisition under the sorts of conditions in which it actually takes place—but still retains the ground Chomsky has won through his formal approach to the logical problem of language acquisition. In addition, his project moves beyond formal (mathematical) proofs that a particular grammar is learnable and attempts to state in explicit step-by-step terms the procedures children use to acquire specific aspects of the language. One of the most illuminating things about such an enterprise is the way it exposes substantive assumptions that are normally tacit in the traditional child language literature.

Let me sketch some of the crucial and highly controversial assumptions Pinker makes prior to his first significant proposal about an actual learning mechanism:

1. The learning theory is stated in terms of several submechanisms that are each

built to acquire a particular class of rules in the grammar. This means Pinker has to pick a particular theory of grammar to work with since different theories postulate different rule types (LFG has no transformations, for example, whereas GB does). Pinker picks LFG, though he never offers a convincing argument that GB would not work as well. The point is that one has to pick a theory. This differentiates learnability theories from almost all extant work in child language, where no formal theory is assumed.

2. He also accepts Chomsky's central assumption that the child has substantive innate language-specific knowledge. Thus, Pinker assumes that the child knows, prior to language acquisition, such task-specific knowledge as the overall structure of the grammar, the formal nature of the sorts of rules it contains, and the primitives from which these rules may be composed (such as *noun* and *verb*). This assumption is obviously controversial, but task-general learning theories for language are nowhere to be found, and theories that assume no innate knowledge (whether language-specific or not) are ruled out by the logical problems of language acquisition above.⁶

3. Under the rubric of what he calls the *continuity assumption*, Pinker assumes that the cognitive and grammatical mechanisms of the child are qualitatively the same as those of the adult (p.7). That is, at any stage of development, the child's grammar is stated in the same formal terms as the adult's. This assumption rules out qualitative maturational changes in the course of linguistic development. It rules out, for instance, the possibility that at the beginning stages of acquisition the child has linguistic rules that are stated in purely semantic terms (agent, patient, instrument, and so on) and that later, through some maturational growth, the child's grammar is stated in syntactic terms (noun, verb, noun phrase, verb phrase, and so forth). Since the adult's rules are storable only in syntactic terms (accepting this as shown by Chomsky), then the child's rules at all stages must also be stated in these terms.

This assumption—that the child's cognitive and linguistic mechanisms are qualitatively the same as the adult's—immediately rules out, for example, Piaget's approach to language development, where the child goes through qualitatively different stages (for example, a stage in which a certain cognitive schema is stated in motoric terms and later in formal operational terms). As Pinker points out, the burden of proof falls on those who would deny the continuity assumption. So far, no one has shown how we would know that underlying mechanisms were significantly different at different stages, and, worse, how the child gets from one set of mechanisms as a whole to another. In Pinker's theory the child simply changes pieces of rules or adds rules to a system already in place. Surely this is the more parsimonious assumption, though one has to admit that the biological world is full of examples of qualitative change in physical development. Without the continuity assumption, a theory of acquisition is probably not attainable given the current state of our knowledge.

4. Finally, Pinker assumes that a child, even when she does not know the structure of a sentence, can infer the meaning of adults' utterances from their physical and discourse contexts and from the meanings of individual words in the sentence. He

actually makes some quite specific assumptions about what semantic knowledge a child has prior to the acquisition of syntax: a child can extract the meanings of predicates and their associated arguments (nominals); the relations of these arguments to the predicate; the potentially grammatically relevant semantic features of the sentence participants (their number, person, gender, and so forth) and of the proposition as a whole (tense, aspect, modality, and so forth); and the discourse features of the speech act (whether it is, for example, declarative, interrogative, exclamative, negative, or emphatic). This is tantamount to assuming a *language of thought* that essentially has its own structure (*grammar*). It renders the learnability question one of how a child “translates” from this language of thought to the language she is learning. In context, a child might understand that certain “pieces of meaning” are being talked about—for instance, that an action of throwing is being talked about, and that throwing involves an animate agent, an action, and a patient. She knows that such meaningful features as the animateness, personhood, number, and sex of the participants may be linguistically relevant (the sort of thing that languages can encode). She knows that languages have nouns and verbs, and, from context, she knows that a statement is being made, not a question asked.

But, even knowing all this beforehand, and even assuming that a child can segment the speech stream, how does she figure out what pieces in a string of words (such as “The boy threw the rock”) encode the various pieces of meaning she has recognized to be relevant from context? It does no good to know that an action of throwing is being talked about if you do not know which word encodes this concept. It does no good to know that languages have nouns if you do not know which are the nouns in the incoming speech stream. This, then, is the first problem. To solve it, Pinker (following Grimshaw, 1981; Macnamara, 1982) proposes his first acquisition mechanism. In adult grammar there is no one-to-one correspondence between semantic categories and syntactic ones. Not all nouns name persons, places, or things (*honesty, a throw, dizziness*); not all verbs name actions or changes of state (*know, see*); not all subjects are agents and not all objects are patients (for example, in “John received a blow on the head,” “John” is the patient). However, let’s assume that the child makes the assumption that syntactic categories *do* match semantic ones one-for-one *until proven otherwise*. Thus the child has a scheme that says, for instance, that the prototypical noun names a person or thing, the prototypical verb names an action or change of state, the prototypical subject names an agent or causer, the prototypical object names a patient. Then the child simply has to find some instances in the input of sentences where these correspondences hold (and can ignore the rest for the time being). Faced with a string of words such as “The boy threw the rock,” the child can assign syntactic categories and grammatical relations through these prototypicality schemes. Although in the earliest stages of acquisition children ignore function words such as articles (the indefinite article *a* and the definite article *the*), for the sake of simplicity we will assume that the child has a prototypicality scheme that associates definiteness in discourse with the category ART (a device which actually comes into play later in acquisition).

Given from context and
child's semantic
knowledge:

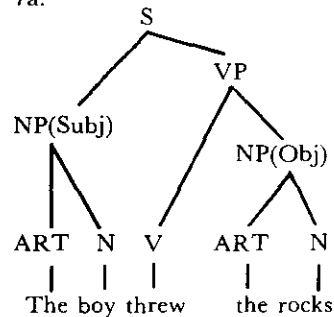
Definite	Agent	Action	Definite	Patient
the	boy	threw	the	rock

Assigned grammatical
relations and syntactic
categories, on the basis
of prototypicality:

ART	N(Subj)	VERB	ART	N(Obj)
-----	---------	------	-----	--------

Pinker calls this procedure *semantic bootstrapping* (see also Grimshaw, 1981; Macnamara, 1982). It is essentially a way into the system. As a next step, we assume that the theory of phrase structure is part of the child's innate linguistic knowledge. This tells the child that languages have phrases, that each type of phrase (noun phrase, verb phrase, adjective phrase) is headed by (organized around) a matching grammatical category (noun, verb, adjective), and that phrases can contain certain other sorts of units beside their heads (noun phrases can contain ARTs, for example). The child, of course, doesn't know what order words will have in phrases because this differs across languages. Using this knowledge about phrase structure, and the representation the child has induced from semantic bootstrapping, the child can now induce some of the phrase structure rules of the language, assigning a tree to the input and storing phrase structure rules that will make predictions about other possible sentences. Additionally, she can add certain words to her lexicon (see 7a and 7b below). In LFG, phrase structure rules are annotated with symbols for grammatical relations.

7a.



7b.

S → NP_{Subj} VP
NP → ART N
VP → V NP_{Obj}

Lexicon:

<i>the</i> :	ART:	Definite = +
<i>boy</i> :	N:	Pred = "dog"
<i>rock</i> :	N:	Pred = "rock"
<i>threw</i> :	V:	Pred = "threw (Subj, Obj)" agent patient

Once a basic scaffolding of semantically induced rules and lexical items is in place, semantically neutral (nonprototypical) items and rules can be learned by observing their distribution within the known structures. For example, hearing a sentence such as "The situation frightened John," the child could learn that *situation* is a noun and that it heads noun phrases because it patterns in the same way as the previous class of prototypical nouns. Pinker refers to this process as *structure-dependent distributional learning*.

The role of semantic bootstrapping can be seen in another way. Syntactic notions, such as *subject*, in reality stand for a correlated set of phenomena. For instance, across languages, subjects tend to display a variety of correlated properties: (a) they can stand for an agent, whatever else they can stand for (that is, all languages will allow an agent to be the subject of a sentence, some will allow other semantic relations to be the subject—as in “John received a blow,” where the patient is the subject—while others allow only agents to be subject; (b) they can stand for a topic, whatever else they can stand for; (c) they can trigger agreement on the verb; (d) they can be unexpressed in complements (“John wanted ____ to win”); (e) they are never expressed as reflexives (*Himself loves John); (f) they can be unexpressed in conjoined clauses (“John likes soup and ____ hates meat”); and (g) they can be relativized, even if nothing else in the language can (“the boy who died”), and so on. In fact, this is how a linguist identifies subjects across languages. Pinker’s bootstrapping hypothesis assumes that the child knows, prior to language acquisition, this clustering of properties and that, in order to start the acquisition process, she looks for a prototypical agent, identifies this as a subject, sets up the phrase structure rule for subjects (fixes word order), and then predicts that all of the properties of subjects will follow (until proven otherwise in a particular language). Notice that the claim is *not* that *subject* is a semantic notion nor that the grammar contains rules that are purely semantic ($S \rightarrow \text{Agent} + \text{Action} + \text{Patient}$) at any point. The child uses semantic knowledge to break into a formal system. Once in, the child knows what other formal properties can cluster with this semantic property.

It may seem counterintuitive to attribute to young children who utter only two- and three-word combinations (as in Brown’s Stage I, 1973, for example), syntactic rules which will generate much more complicated structures. Because of the restricted sorts of sentences young children utter, it has been claimed that children’s early rules ought to be stated in semantic terms ($\text{Sentence} \rightarrow \text{Agent} + \text{Action}$) or that subject and objects are not drawn from a unitary syntactic category (NP) but rather are different entities ($\text{Sentence} \rightarrow \text{Pronoun or Animate Noun} + \text{Verb} + \text{Inanimate Noun}$). Pinker’s bootstrapping hypothesis claims that children use actionhood as a cue that a word is a verb, and agenthood and patienthood as cues to determine which NP is its subject (the agent) and which its object (the patient). Thus, he predicts that children’s earliest (at least first) verbs will indeed denote actions whose subjects and objects are agents and patients respectively. However, this semantic pattern will not continue past the earliest utterances, since it is merely a key into the formal system. Through a close study of the available developmental data, Pinker shows that children do not restrict their grammars in semantic terms. Early grammars are not restricted to animate preverbal elements (subjects) and inanimate postverbal ones (objects)—children produce enough of the opposite possibilities to demonstrate that they have grammars that can generate the full range of possibilities (grammars that must appeal, then, to a unitary formal notion like *noun phrase*). Pinker shows that postverbal elements are expanded using all the same structural combinations (noun, pro, art + noun, adj + noun, poss + noun) regardless of the semantic role they

are playing (patient, object of experience, predicate nominal, locative). Furthermore, Pinker cites work by Brown which shows that children at Brown's Stages I and II produce virtually all the possible combinations of elements that a simple phrase structure grammar would predict (with the same structural possibilities for subject and object, that is, NP), though only in strings of less than five words.⁷

Pinker concludes that the fragmentary nature of early speech is due to performance factors (how the child uses rules) and not the rule system the child has. While many have found it "obvious" that children's grammars must be stated in semantic or pragmatic terms, Pinker demonstrates that this is not obvious. If, in fact, it is true that only purely formal (syntactic) rules can account for the full range of children's early language production, then Pinker's case that children have syntactic rules from the start, and the case for a universal grammar statable in structural terms, is strengthened.

Let us return to the sorts of problems raised by the Dative Alternation. We saw that many verbs alternate between sentences like "Mary told the magic words to John" and "Mary told John the magic words" but that it is hard to account for how a child learns just which verbs do this and which do not. Here we will discuss, however, a different syntactic structure (causatives), one where the issues are somewhat more tractable, at least in Pinker's framework (pp. 324-347). In English, lexical causatives (cases where the language has a word that includes causation as part of its meaning, as in 8a) can be used in cases of direct or immediate physical causation. Periphrastic causatives (cases where we need to use a verb like *cause* or *make* to express causation) are used for indirect causation (see 8b). Moreover, lexical causation requires that the agent, manner, and goal of causation be conventional or stereotypical for the relevant action. Thus, there are no lexical causatives for verbs such as *laugh* or *vomit*, which lack conventional direct means of causation (see 8c).

- 8a. John broke the window.
- 8b. John caused/made the window break.
- 8c. *John laughed Harry.

Children overgeneralize lexical causatives, producing sentences such as "Sweater me" and "Don't giggle me" that violate the conventionality constraint above or sentences such as "I come it closer so it won't fall," where the adult would use a periphrastic construction ("Make it come closer"). The question, then, is how children come to learn the constraints noted above and to remove the overgeneralized forms they have created, all in the absence of negative evidence (that is, in the absence of being corrected or told that certain cases are ungrammatical).

Pinker offers an interesting solution to this problem, one that is fairly indicative of his overall approach. Somewhat surprisingly, he relates the problem to that of learning "paradigms." For example, in a language that has case markers on its nouns, each noun has a paradigm of forms (in Latin, for instance: *puella* = subject; *puellam* = object; *puellae* = dative; *puellae* = genitive). In these cases, Pinker assumes that while the child may initially use one form (*puella*) for all cases, eventually she realizes

that in the data the noun keeps varying in form. At this point the child has to hypothesize about what is causing this variation. Pinker further assumes that the child initially considers only hypotheses available in universal grammar, and perhaps in some order of likelihood also specified by universal grammar. In a similar vein, Pinker claims that children also learn the various syntactic constructions in which a verb can occur as paradigms. The cells of the paradigm could include active, passive, causative, reflexive, double-object dative, inchoative, middle, and so on, differing somewhat from language to language but drawn from a small universal set.⁸ A child would form such a paradigm for each verb and fill in each cell with an entry corresponding to it which is learned from positive evidence. Below is part of a word-specific paradigm for *open* that may have been formed by a child at a certain stage (*theme* refers to the name of the semantic role of the entity that moves or changes state or location in a sentence; patients are a subclass of themes).

Open

Intransitive	Causative	Passive
Subj V	Subj V Obj	Subj V by NP
Theme	Agent Theme	Theme Agent
The door opened	John opened the door	The door was opened by John
	Subj cause Subj V Obj	
	Causer Agent Theme	
	Bill made John open the door	

Throughout his book Pinker assumes the *unique entry principle*: no construction can be realized by more than one lexical form. The paradigm for *open* violates this principle since the causative cell is realized in two ways. A child can resolve this violation by searching for a new dimension that distinguishes between the two competing forms of the causative. She does this by focusing on context-specific features evident in future usages of one of the forms. For instance, she may notice on a particular occasion that the lexical causative is being used to denote an act of direct physical causation (or conversely, she may notice that the periphrastic form is being used to denote an act of indirect causation) and hypothesize that this is the relevant distinction. This hypothesis will not suffer from disconfirming instances and will yield a new paradigm with a cell for direct causation and one for indirect causation. We still have to account for how the child who has overgeneralized (to "Giggle me," for example) will be able to expunge this entry from her lexicon. Pinker assumes that any entries that have been created by generalizing from the data, and not from direct positive evidence, bear a mark of "preemptability." When the child discovers constraints such as those above, she will remove any items that bear the preemptability mark and that break the constraint. If some item not only had no such mark but also broke the constraint, it would be viewed as a legitimate exception and noted as such. Notice that this proposal has the effect of significantly blurring the line between learning syntax (constructions like the periphrastic causative) and learning morphology (word paradigms like *puella, puellam*).

This is an insightful move. A theory of acquisition must account for how the child learns any language, not just English. And across languages we find that what one language does syntactically (for instance, the periphrastic causative in English) another language does morphologically (for instance, many languages have affixes that are added to a verb to signal indirect causation; they use lexical forms like *open* for direct causation; see Comrie, 1981). If the learning mechanism is the same in both cases, we not only account for how the child learns in both cases but begin to account for this sort of variation across languages. The most important implication of Pinker's enterprise is, I believe, the realization that without explicit proposals about learning mechanisms which explain what the child is learning and how exactly she is doing it, attributions to the child are meaningless. For instance, it is possible to collect a large body of data showing that children are sensitive to discourse properties such as topic, focus, new information, and old information. This fact in itself tells us very little. It leaves open many possibilities about what the child's grammar looks like and, thus, amounts to "descriptive adequacy" at best. Pinker demands more—namely, explicit assumptions about the child's grammar and learning mechanisms that explain why the data look the way they do and still render the purely formal, syntactic part of language learnable. He is thus proposing that the study of child language move beyond "descriptive adequacy" and on to "explanatory adequacy." Perhaps it is too early for this demand or perhaps not. Only time will tell. Until then we can only hope that we begin to reach this goal before we are crushed by mountains of data.

Notes

1. Pinker (1979) is an excellent though somewhat technical survey of various formal models of language learning; Pinker's unpublished study (n.d.) is an equally excellent nontechnical discussion. Pinker (1982) is an earlier, fairly technical treatment of the material in his book. While the book represents a number of advances on the article, many readers will find it quite sufficient to give them a feeling for the basic approach. Pinker (1985) is a new and less technical discussion that is probably at this point the best and most accessible place for the interested reader to turn. Slobin has also done extremely important work relevant to learnability theory. It is less technical because not tied to any particular linguistic theory (though less easy to falsify, perhaps, for this same reason). Slobin (1982, in press), among others, is essential reading for anyone interested in language acquisition, as well as learnability.
2. The * in 2b means the sentence is ungrammatical. Notice that this is a structural fact since one can perfectly well figure out what 2b means. Furthermore, there is nothing particularly deviant about this meaning since it is not all that different from the meaning of the grammatical sentence in 1b.
3. I will use the feminine pronominal forms *she* and *her* for the generic pronoun form.
4. Perhaps *relative clause* is not the right term; see, for example, Chomsky (1981).
5. Piatelli-Palmarini (1980) is the record of a fascinating debate between Chomsky and Piaget and their respective followers. Chomsky has made two substantive claims that psychologists have disputed: (1) no one can account for the acquisition of language without making substantive and rich assumptions about innate tacit knowledge somehow encoded in the brain; and (2) that this innate knowledge is specifically germane to language, and not other areas of cognition, and thus we must assume a specially evolved capacity of the mind (somehow represented in the brain) for the acquisition of language. One can accept (1) and still reject (2) by giving an account of language acquisition that draws on innate knowledge relevant to a wide range of domains—for example, perception of the physical world, the construction of space and time, mathematical concepts, systems of social behavior, problem solving, and so forth—and not just language. One can deny both assumptions by constructing a general learning theory that shows how a full range of human knowledge, including language, is acquired by principles that store and process data but does

- not assume substantive innate knowledge about any of these domains. The task for Chomsky's critics is fairly clear—and his positions are readily falsifiable. For instance, if anyone could account for how all adult speakers know 6b (*Who(m) did John recognize the truck that hit?) is ungrammatical without appealing to any purely linguistic innate knowledge, they would have falsified assumption (1) in this case, and if they could do so without appealing to any innate knowledge at all they would have falsified (2). No one, as far as I can see, has come remotely close to doing so. Notice also that their account would have to explain why it is the case that whatever blocks this sentence in English blocks the same sort of sentence in other languages, since sentences like this one are indeed also ungrammatical cross-linguistically. One should, however, bear in mind that Chomsky's current account of these matters is much more abstract and covers a far wider range of cases than I have indicated here.
6. "Task-general" learning theories are theories that would appeal only to innate principles relevant to a wide range of domains, not just language. They would deny not that humans have innate knowledge but that they have any innate knowledge germane only to one area of cognition, such as language. For instance, they may claim that humans have innately specified principles germane to problem solving generally and that they use these principles to acquire language as well as solve other sorts of cognitive problems; see n. 3.
 7. Stage I and II are the first two of Brown's five stages of language acquisition. Brown's stages were defined in terms of mean length of utterance (MLU). Stage I children have an MLU of 1.75 words per utterance with an upper bound of 5, while Stage II children have an MLU of 2.5 with an upper bound of 7. Pinker has gone carefully through Brown's copious data to test his claims against what children actually do.
 8. Some examples of the constructions named *Active*: John sold the book to Mary; *Passive*: The book was sold to Mary by John; *Causative*: John made Herman sell the book to Mary; *Reflexive*: John sold himself to Mary; *Double-Object Dative*: John sold Mary the book; *Inchoative*: The book came to be sold to Mary; and *Middle*: Books sell easily. One should keep in mind that other languages have different verb forms for several or all of these constructions, that is, *sell* would change its shape or have a morpheme added to it in each use rather than simply adding new words to the sentence as a whole (like *was* or *made* or *himself*), or changing word order. For example, *yomaserareta* in Japanese is a complex word made up of several parts meaning "was made to read": *yom* (read) *sase* (cause) *rare* (passive) *ta* (past). This verb is a passive causative, realizing in one word what English needs several words for. Thus, we cannot clearly separate learning morphology (the ordering of the different morphemes in complex words) from syntax (the ordering of words in sentences).

References

- Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry*, 10, 533–581.
- Baker, C. L., & McCarthy, J. J. (Eds.). (1981). *The logical problem of language acquisition*. Cambridge: MIT Press.
- Bresnan, J. W. (Ed.). (1982). *The mental representation of grammatical relations*. Cambridge: MIT Press.
- Brown, R. (1973). *A first language: The early stages*. Cambridge: Harvard University Press.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, Netherlands: Foris.
- Chomsky, N. (1984). *Knowledge of language: Its nature, origin, and use*. Unpublished manuscript, Massachusetts Institute of Technology, Cambridge.
- Comrie, B. (1981). *Language universals and linguistic typology*. Chicago: University of Chicago Press.
- Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.
- Givón, T. (1984). *Syntax: A functional-typological introduction*. (Vol. 1). Amsterdam/Philadelphia: John Benjamins.
- Grimshaw, J. (1981). Form, function, and the language acquisition device. In C. L. Baker & J. J. McCarthy (Eds.), *The logical problem of language acquisition* (pp. 165–182). Cambridge: MIT Press.

*Language Issues in Literacy
and Bilingual/Multicultural Education*

- Hamburger, H., & Wexler, K. (1975). A mathematical theory of learning transformational grammar. *Journal of Mathematical Psychology*, 12, 137-177.
- Macnamara, J. (1982). *Names for things: A study of child language*. Cambridge: Bradford Books/MIT Press.
- Piattelli-Palmarini, M. (Ed.). (1980). *Language and learning: The debate between Jean Piaget and Noam Chomsky*. Cambridge: Harvard University Press.
- Pinker, S. (1979). Formal models of language learning. *Cognition*, 1, 217-283.
- Pinker, S. (1982). A theory of the acquisition of lexical interpretive grammars. In J. W. Bresnan (Ed.), *The mental representation of grammatical relations*. (655-726). Cambridge: MIT Press.
- Pinker, S. (1985). Language learnability and children's language: A multidisciplinary approach. In K. Nelson (Ed.), *Children's language* (Vol. 5). Hillsdale, NJ: Erlbaum.
- Pinker, S. (n.d.). *Recent advances in the study of the acquisition of grammar*. Unpublished manuscript, Massachusetts Institute of Technology, Dept. of Psychology and Center for Cognitive Science, Cambridge.
- Ross, J. R. (1967). *Constraints on variables in syntax*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge.
- Slobin, D. I. (1982). Universal and particular in the acquisition of language. In E. Wanner and L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 128-170). Cambridge: Harvard University Press.
- Slobin, D. I. (Ed.). (in press). *The crosslinguistic study of language acquisition*. Hillsdale, NJ: Erlbaum.
- Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition*. Cambridge: MIT Press.