Knowledge and Learning in Natural Language

CHARLES D. YANG

OXFORD
UNIVERSITY PRESS

step in my intellectual development; words of gratitude are not needed for they cannot suffice. Many thanks also to Tom Roeper, for his energy and vast knowledge of child language, and to Patrick Winston, for his good scientific sense and important support during my final days at MIT.

I would like to single out four individuals who have been instrumental to this work. John Frampton and Sam Gutmann made sure my thinking was clear and coherent. Morris Halle taught me everything I know about phonology and devoted an embarrassing amount of time to this work, particularly the chapter on irregular verbs. And finally, Julie Legate read everything I'd ever written and, in many ways, contributed more than anyone else. In addition, I thank her for allowing me to present our joint work in these pages.

Many colleagues and friends examined various parts of the present product at various stages. They are: Steve Anderson, Janet Dean Fodor, Louis Goldstein, Norbert Hornstein, David Lightfoot, Alec Marantz, Fritz Newmeyer, Steven Pinker, Morgan Sonderegger, Jean-Roger Vergnaud, Jill de Villers, Bill Wang, Amy Weinberg, and Ken Wexler. I thank them for encouragement and suggestions. Audiences at many conferences, workshops, and institutions, as well as the students in my seminars at Yale, also helped sharpen some of the ideas developed here. And finally, I am indebted to two anonymous reviewers at Oxford University Press for their good advice on the final manuscript, and John Davey and Jacqueline Smith, my editors, for their patience and support.

This work was funded partially by a graduate fellowship from the National Science Foundation. The Department of Linguistics at Yale provided a warm and supportive environment for the book's completion.

Now, on to the vineyard.

New Haven, Connecticut — Charles D. Yang

1

The Study of Language and Language Acquisition

> We may regard language as a natural phenomenon—an aspect of his biological nature, to be studied in the same manner as, for instance, his anatomy.
>
> Eric H. Lenneberg, *Biological Foundations of Language* (1967), p. vii

1.1 The naturalistic approach to language

Fundamental to modern linguistics is the view that human language is a natural object: our species-specific ability to acquire a language, our tacit knowledge of the enormous complexity of language, and our capacity to use language in free, appropriate, and infinite ways are attributed to a property of the natural world, our brain. This position needs no defense, if one considers the study of language is an empirical inquiry.

It follows, then, as in the study of biological sciences, linguistics aims to identify the abstract properties of the biological object under study—human language—and the mechanisms that govern its organization. This has the goal set in the earliest statements on modern linguistics, Chomsky's *The Logical Structure of Linguistic Theory* (1955). Consider the famous duo:

(1) a. Colorless green ideas sleep furiously.
    b. *Furiously sleep ideas green colorless.

Neither sentence has even a remote chance of being encountered in natural discourse, yet every speaker of English can perceive their differences: while they are both meaningless, (1a) is grammatically well formed, whereas (1b) is not. To understand what precisely this difference is is to give 'a rational account of this behavior, i.e., a theory of the speaker's linguistic intuition . . . the goal of linguistic theory' (Chomsky 1955/1975: 95)—in other words, a psychology, and ultimately, biology of human language.

Once this position—lately dubbed the *biolinguistic* approach (Jenkins 1999, Chomsky 2000)—is accepted, it follows that language, just like all other biological objects, ought to be studied following the standard methodology in natural sciences (Chomsky 1975, 1980, 1986, 1995a). The postulation of innate linguistic knowledge, the Universal Grammar (UG), is a case in point.

One of the major motivations for innateness of linguistic knowledge comes from the Argument from the Poverty of Stimulus (APS) (Chomsky, 1980: 35). A well-known example concerns the *structure dependency* in language syntax and children's knowledge of it in the absence of learning experience (Chomsky 1975, Crain & Nakayama 1987). Forming an interrogative question in English involves inversion of the auxiliary verb and the subject:

(2) a. Is Alex *e* singing a song?
  b. Has Robin *e* finished reading?

It is important to realize that exposure to such sentences underdetermines the correct operation for question formation. There are many possible hypotheses compatible with the language acquisition data in (2):

(3) a. front the first auxiliary verb in the sentence
  b. front the auxiliary verb that most closely follows a noun
  c. front the last auxiliary verb
  d. front the auxiliary verb whose position in the sentence is a prime number
  e. . . .

The correct operation for question formation is, of course, structure-dependent: it involves parsing the sentence into structurally organized phrases, and fronting the auxiliary that follows *the first noun phrase*, which can be arbitrarily long:

(4) a. Is [$_{NP}$ the woman who is sing] *e* happy?
  b. Has [$_{NP}$ the man that is reading a book] *e* had supper?

Hypothesis (3a), which arguably involves simpler mental computation than the correct generalization, yields erroneous predictions:

(5) a. *Is [the woman who *e* singing] is happy?
  b. *Has [the man that *e* finished reading] has finished supper?

But children don't go astray like the creative inductive learner in (3). They stick to the correct operation from very early on, as Crain & Nakayama (1987) showed using elicitation tasks. The children were instructed, 'Ask Jabba if the boy who is watching Mickey Mouse is happy', and no error of the form in (5) was found.

Though sentences like those in (4) may serve to disconfirm hypothesis (3a), they are very rarely if ever encountered by children in normal discourse,[1] not to mention the fact that each of the other incorrect hypotheses in (3) will need to be ruled out by disconfirming evidence. Here lies the logic of the APS:[2] if we know X, and X is underdetermined by learning experience, then X must be innate. The conclusion is then Chomsky's (1975: 33): 'the child's mind . . . contains the instruction: Construct a structure-dependent rule, ignoring all structure-independent rules. The principle of structure-dependence is not learned, but forms part of the conditions for language learning.'

The naturalistic approach can also be seen in the evolution of linguistic theories through successive refinement and revision of ideas as their conceptual and empirical flaws are revealed. For example, the 1960s language-particular and construction-specific transformational rules, while descriptively powerful, are inadequate when viewed in a biological context. The complexity and

[1] In section 4.2, we will rely on corpus statistics from Legate (1999) and Legate & Yang (in press) to make this remark precise, and to address some recent challenges to the APS by Sampson (1989) and Pullum (1996).

[2] See Crain (1991) for several similar cases, and numerous others in the child language literature.

unrestrictiveness of rules made the acquisition of language wildly difficult: the learner had a vast (and perhaps an infinite) space of hypotheses to entertain. The search for a plausible theory of language acquisition, coupled with comparative linguistic studies, led to the Principles and Parameters (P&P) framework (Chomsky 1981), which suggests that all languages obey a universal (and putatively innate) set of tightly constrained principles, whereas variations across constructions and particular languages—the choices that a child learner has to make during language acquisition—are attributed to a small number of parametric choices.

The present book is a study of language development in children. From a biological perspective, the development of language, like the development of other organic systems, is an interaction between internal and external factors; specifically, between the child's internal knowledge of linguistic structures and the external linguistic experience he receives. Drawing insights from the study of biological evolution, we will put forth a model that make this interaction precise, by embedding a theory of knowledge, the Universal Grammar (UG), into a theory of learning from data. In particular, we propose that language acquisition be modeled as a population of 'grammars', competing to match the external linguistic experiences, much in the manner of natural selection. The justification of this approach will take the naturalistic approach just as in the justification of innate linguistic knowledge: we will provide evidence—conceptual, mathematical, and empirical, and from a number of independent areas of linguistic research, including the acquisition of syntax, the acquisition of phonology, and historical language change—to show that without the postulated model, an adequate explanation of these empirical cases is not possible.

But before we dive into details, some methodological remarks on the study of language acquisition.

1.2 The structure of language acquisition

At the most abstract level, language acquisition can be modeled thus:

(6) $\mathcal{L}: (S_o, E) \rightarrow S_T$

A learning function or algorithm $\mathcal{L}$ maps the initial state of the learner, S_o, to the terminal state S_T, on the basis of experience E in the environment. Language acquisition research attempts to give an explicit account of this process.

1.2.1 *Formal sufficiency*

The acquisition model must be *causal* and *concrete*. Explanation of language acquisition is not complete with a mere description of child language, no matter how accurate or insightful, without an explicit account of the mechanism responsible for how language develops over time, the learning function $\mathcal{L}$. It is often claimed in the literature that children just 'pick up' their language, or that children's linguistic competence is identical to adults. Such statements, if devoid of a serious effort at some learning-theoretic account of *how* this is achieved, reveal irresponsibility rather than ignorance.

The model must also be *correct*. Given reasonable assumptions about the linguistic data, the duration of learning, the learner's cognitive and computational capacities, and so on, the model must be able to attain the *terminal* state of linguistic knowledge S_T comparable to that of a normal human learner. The correctness of the model must be confirmed by mathematical proof, computer simulation, or other forms of rigorous demonstration. This requirement has traditionally been referred to as the *learnability condition*, which unfortunately carries some misleading connotations. For example, the influential Gold (1967) paradigm of identification in the limit requires that the learner converge onto the 'target' grammar in the linguistic environment. However, this position has little empirical content.[3]

First, language acquisition is the process in which the learner forms an *internalized* knowledge (in his mind), an I-language

[3] I am indebted to Noam Chomsky for many discussions on the issue of learnability.

(Chomsky 1986). Language does not exist in the world (in any scientific sense), but resides in the heads of individual users. Hence there is no external target of learning, and hence no 'learnability' in the traditional sense. Second, section 1.2.2 below documents evidence that child language and adult language appear to be sufficiently different that language acquisition cannot be viewed as recapitulation or approximation of the linguistic expressions produced by adults, or of any external target. And third, in order for language to change, the terminal state attained by children must be different from that of their ancestors. This requires that the learnability condition (in the conventional sense) must fail under certain conditions—in particular (as we shall see in Chapter 5) empirical cases where learners do not converge onto any unique 'language' in the informal and E-language sense of 'English' or 'German', but rather a combination of multiple (I-language) grammars. Language change is a result of changes in this kind of grammar combinations.

1.2.2 *Developmental compatibility*

A model of language acquisition is, after all, a model of reality: it must be compatible with what is known about children's language.

Essential to this requirement is the *quantitativeness* of the model. No matter how much innate linguistic knowledge (S_o) children are endowed with, language still must be acquired from experience (E). And, as we document extensively in this book, not all languages, and not all aspects of a single language, are learned uniformly. As long as this is the case, there remains a possibility that there is something in the input, E, that causes such variations. An adequate model of language acquisition must thus consist of an explicit description of the learning mechanisms, $\mathcal{L}$, that quantify the relation between E, what the learner receives, and S_T, what is acquired. Only then can the respective contribution from S_o and E—nature vs. nurture, in a cliché—to language acquisition be understood with any precision.[4]

This urges us to be serious about quantitative comparisons between the input and the attained product of learning: in our case, quantitative measures of child language and those of adult language. Here, many intriguing and revealing disparities surface. A few examples illustrate this observation and the challenge it poses to an acquisition model.

It is now known that some aspects of the grammar are acquired successfully at a remarkably early age. The placement of finite verbs in French matrix clauses is such an example.

(7) Jean voit *souvent/pas* Marie.
    Jean sees *often/not* Marie.
    'John *often* sees/does *not* see Marie.'

French, in contrast to English, places finite verbs in a position preceding sentential adverbs and negations. Although sentences like (7), indicative of this property of French, are quite rare in adult-to-child speech (7%; estimate based on CHILDES—see MacWhinney & Snow 1985), French children, from as early as can be tested (1;8: Pierce 1989), almost never deviate from the correct form. This discovery has been duplicated in a number of languages with similar properties; see Wexler (1994) and much related work for a survey.

In contrast, some very robustly attested patterns in adult language emerge much later in children. The best-known example is perhaps the phenomenon of subject drop. Children learning English, and other languages that require the presence of a grammatical subject often produce sentences as in (8):

(8) a. (I) help Daddy.
    b. (He) dropped the candy.

Subject drop appears in up to 30% of all sentences around 2;0, and it is not until around 3;0 that they start using subjects at adult

[4] This requirement echoes the quantitative approach that has become dominant in theoretical language acquisition over the past two decades—it is no coincidence that the maturation of theoretical linguistics and the construction of large scale child language databases (MacWhinney & Snow 1985) took place around the same time.

level (Valian 1991), in striking contrast to adult language, where subject is used in almost all sentences.

Perhaps more interestingly, children often produce utterances that are virtually absent in adult speech. One such example that has attracted considerable attention is what is known as the Optional Infinitive (OI) stage (e.g. Weverink 1989, Rizzi 1994, Wexler 1994): children acquiring some languages that morphologically express tense nevertheless produce a significant number of sentences where matrix verbs are non-finite. (9) is an example from child Dutch (Weverink 1989):

(9) pappa schoenen wassen
daddy shoes to-wash
'Daddy washes shoes.'

Non-finite root sentences like (9) are ungrammatical in adult Dutch and thus appear very infrequently in acquisition data. Yet OI sentences are robustly used by children for an extended period of time, before they gradually disappear by 2;6 or later.

These quantitative disparities between child and adult language represent a considerable difficulty for empiricist learning models such as neural networks. The problem is, as pointed out by Fodor & Pylyshyn (1988), that learning models without prior knowledge (e.g. UG) can do no more than recapitulate the statistical distribution of the input data. It is therefore unclear how a statistical learning model can duplicate the developmental patterns in child language. That is, during the course of learning,[5]

(10) a. The model must not produce certain patterns that are in principle compatible with the input but never attested (the argument from the poverty of stimulus).
b. The model must not produce certain patterns abundant in the input (the subject drop phenomenon).
c. The model must produce certain patterns that are never attested in the input (the Optional Infinitive phenomenon).

[5] Note that there is no obvious extralinguistic reason why the early acquisitions are intrinsically 'simpler' to learn than the late acquisitions. For instance, both the obligatory use of subject in English and the placement of finite verbs before/after negation and adverbs involve a binary choice.

Even with the assumption of innate UG, which can be viewed as a kind of prior knowledge from a learning-theoretic perspective, it is not clear how such quantitative disparities can be explained. As will be discussed in Chapter 2, previous formal models of acquisition in the UG tradition in general have not begun to address these questions. The model developed in this study intends to fill this gap.

Finally, quantitative modeling is important to the development of linguistics at large. At the foundation of every 'hard' science is a formal model with which quantitative data can be explained and quantitative predictions can be made and checked. Biology did not come of age until the twin pillars of biological sciences, Mendelian genetics and Darwinian evolution, were successfully integrated into the mathematical theory of population genetics—part of the Modern Synthesis (Mayr & Provine 1980)—where evolutionary change can be explicitly and quantitatively expressed by its internal genetic basis and external environmental conditions.[6] If language development is a biological process, it would certainly be desirable for the interplay between internal linguistic knowledge and external linguistic experience to be quantitatively modeled with formalization.

1.2.3 *Explanatory continuity*

Because child language apparently differs from adult language, it is thus essential for an acquisition model to make some choices on explaining such differences. The condition of *explanatory continuity* proposed here imposes some restrictions, or, to be more precise, heuristics, on making these choices.

Explanatory Continuity is an instantiation of the well-known Continuity Hypothesis (Macnamara 1982, Pinker 1984), with roots dating back to Jakobson (1941), Halle (1962), and Chomsky (1965). The Continuity Hypothesis says that, without evidence to

[6] See Lewontin (1996) and Maynard Smith (1989) for two particularly insightful introductions to population genetic theories.

the contrary, children's cognitive system is assumed to be identical to that of adults. Since child and adult languages differ, there are two possibilities:

(11) a. Children and adults differ in linguistic performance.
b. Children and adults differ in grammatical competence.

An influential view holds that child competence (e.g. grammar) is identical to adult competence (Pinker 1984). This necessarily leads to a performance-based explanation for child acquisition. There is no question that (11a) is, at some level, true: children are more prone to performance errors than adults, as their memory, processing, and articulation capacities are still underdeveloped. To be sure, adult linguistic performance is affected by these factors as well. However, if and when both approaches are descriptively adequate, there are reasons to prefer competence-based explanations.

Parsimony is the obvious, and primary, reason. By definition, performance involves the interaction between the competence system and other cognitive/perceptual systems. In addition, competence is one of the few components in linguistic performance of which our theoretical understanding has some depth. This is partially because grammatical competence is to a large degree isolated from other cognitive systems—the so-called autonomy of syntax—and is thus more directly accessible to investigation. The tests used for competence studies, often in the form of native speakers' grammatical intuition, can be carefully controlled and evaluated. Finally, and empirically, child language differs from adult language in very specific ways, which do not seem to follow from any general kind of deficit in children's performance.[7] For example, it has been shown that there is much data in child subject drop that does not follow from performance limitation explanations; see e.g. Hyams & Wexler (1993), Roeper & Rohrbacher (1994), Bromberg & Wexler (1995). In Chapter 3, we will show that a theory of English past tense learning based on

[7] Obviously, this claim can only be established on a case-by-case basis.

memory lapses (Pinker 1999) fails to explain much of the developmental data reported in Marcus et al. (1992). Phonological rules and structures in irregular verbs must be taken into account to obtain a fuller explanation. And in Chapter 4, we will see additional developmental data from several studies of children's syntax, including the subject drop phenomenon, to show the empirical problems with the performance-based approach.

If we tentatively reject (11a) as, at least, a less favorable research strategy, we must rely on (11b) to explain child language. But exactly how is child competence different from adult competence? Here again are two possibilities:

(12) a. Child competence and adult competence are qualitatively different.
b. Child competence and adult competence are quantitatively different.

(12a) says that child language is subject to different rules and constraints from adult language. For example, it could be that some linguistic principle operates differently in children from adults, or a piece of grammatical knowledge is absent in younger children but becomes available as a matter of biological maturation (Gleitman 1981, Felix 1987, Borer & Wexler 1987).

It is important to realize that there is nothing unprincipled in postulating a discontinuous competence system to explain child language. If children systematically produce linguistic expressions that defy UG (as understood via adult competence analysis), we can only conclude that their language is governed by different laws. However, in the absence of a concrete theory of how linguistic competence matures (12a) runs the risk of 'anything goes'. It must therefore remain a last resort only when (12b)—the approach that relies on adult competence, for which we do have concrete theories—is shown to be false.[8] More specifically, we must not confuse the difference between child language and *adult*

[8] This must be determined for individual problems, although when maturational accounts have been proposed, often non-maturational explanations of the empirical data have not been conclusively ruled out. For example, Borer & Wexler's proposal (1987) that certain A-chains mature have been called into question by many researchers (e.g. Pinker et al. 1987, Demuth 1989, Crain 1991, Allen 1996, Fox & Grodzinsky 1998).

language with the difference between child language and *Universal Grammar*. That is, while (part of) child language may not fall under the grammatical system the child eventually attains, it is possible that it falls under some *other*, equally principled grammatical system allowed by UG. (Indeed, this is the approach taken in the present study.)

This leaves us with (12b), which, in combination with (11b), gives the strongest realization of the Continuity Hypothesis: that child language is subject to the same principles and constraints in adult language, and that every utterance in child language is potentially an utterance in adult language. The difference between child and adult languages is due to differences in the *organization* of a continuous grammatical system. This position further splits into two directions:

(13) a. Child language reflects a *unique* potential adult language.
b. Child language consists of a *collection* of potential adult languages.

(13a), the dominant view ('triggering') in theoretical language acquisition will be rejected in Chapter 2. Our proposal takes the position of (13b): child language in development reflects a statistical combination of possible grammars allowed by UG, only some of which are eventually retained when language acquisition ends. This perspective will be elaborated in the rest of this book, where we examine how it measures up against the criteria of formal sufficiency, developmental compatibility, and explanatory continuity.

1.3 A road map

This book is organized as follows.

Chapter 2 first gives a short but critical review of previous approaches to language acquisition. After an encounter with the populational and variational thinking in biological evolution that inspired this work, we propose to model language acquisition as a population of competing grammars, whose distribution changes in response to the linguistic evidence presented to the learner. We will give a precise formulation of this idea, and study its formal/computational properties with respect to the condition of *formal sufficiency*.

Chapter 3 applies the model to one of the biggest developmental problems in language, the learning of English past tense. It will be shown that irregular verbs are organized into classes, each of which is defined by special phonological rules, and that learning an irregular verb involves the competition between the designated special rule and the default *-ed* rule. Again, quantitative predictions are made and checked against children's performance on irregular verbs. Along the way we will develop a critique of Pinker and his colleagues' *Words and Rules* model (Pinker 1999), which holds that irregular verbs are individually and directly memorized as associated pairs of root and past tense forms.

Chapter 4 continues to subject the model to the *developmental compatibility* test by looking at the acquisition of syntax. First, crosslinguistic evidence will be presented to highlight the model's ability to make quantitative predictions based on adult-to-child corpus statistics. In addition, a number of major empirical cases in child language will be examined, including the acquisition of word order in a number of languages, the subject drop phenomenon, and Verb Second.

Chapter 5 extends the acquisition model to the study of language change. The quantitativeness of the acquisition model allows one to view language change as the change in the distribution of grammars in successive generations of learners. This can again incorporate the statistical properties of historical texts in an evolving, dynamic system. We apply the model of language change to explain the loss of Verb Second in Old French and Old English.

Chapter 6 concludes with a discussion on the implications of the acquisition model in a broad context of linguistic and cognitive science research.