

ON THE NATURE AND NURTURE OF LANGUAGE

Elizabeth Bates

University of California, San Diego

Support for the work described here has been provided by NIH/NIDCD-R01-DC00216 (“Cross-linguistic studies in aphasia”), NIH-NIDCD P50 DC1289-9351 (“Origins of communication disorders”), NIH/NINDS P50 NS22343 (“Center for the Study of the Neural Bases of Language and Learning”), NIH 1-R01-AG13474 (“Aging and Bilingualism”), and by a grant from the John D. and Catherine T. MacArthur Foundation Research Network on Early Childhood Transitions.

Please address all correspondence to Elizabeth Bates, Center for Research in Language 0526, University of California at San Diego, La Jolla, CA 92093-0526, or bates@crl.ucsd.edu.

ON THE NATURE AND NURTURE OF LANGUAGE

Elizabeth Bates

Language is the crowning achievement of the human species, and it is something that all normal humans can do. The average man is neither a Shakespeare nor a Caravaggio, but he is capable of fluent speech, even if he cannot paint at all. In fact, the average speaker produces approximately 150 words per minute, each word chosen from somewhere between 20,000 and 40,000 alternatives, at error rates below 0.1%. The average child is already well on her way toward that remarkable level of performance by 5 years of age, with a vocabulary of more than 6000 words and productive control over almost every aspect of sound and grammar in her language.

Given the magnitude of this achievement, and the speed with which we attain it, some theorists have proposed that the capacity for language must be built directly into the human brain, maturing like an arm or a kidney. Others have proposed instead that we have language because we have powerful brains that can learn many things, and because we are extraordinarily social animals who value communication above everything else. Is language innate? Is it learned? Or, alternatively, does language emerge anew in every generation, because it is the best solution to the problems that we care about, problems that only humans can solve? These are the debates that have raged for centuries in the various sciences that study language. They are also variants of a broader debate about the nature of the mind and the process by which minds are constructed in human children.

The first position is called “nativism”, defined as the belief that knowledge originates in human nature. This idea goes back to Plato and Kant, but in modern times it is most clearly associated with the linguist Noam Chomsky (see photograph). Chomsky’s views on this matter are very strong indeed, starting with his first book in 1957, and repeated with great consistency for the next 40 years. Chomsky has explicated the tie between his views on the innateness of language and Plato’s original position on the nature of mind, as follows:

“How can we interpret [Plato’s] proposal in modern terms? A modern variant would be that certain aspects of our knowledge and understanding are innate, part of our biological endowment, genetically determined, on a par with the elements of our common nature that cause us to grow arms and legs rather than wings. This version of the classical doctrine is, I think, essentially correct.” (Chomsky, 1988, p. 4)

He has spent his career developing an influential theory of grammar that is supposed to describe the universal properties underlying the grammars of every language in the world. Because this Universal Grammar

is so abstract, Chomsky believes that it could not be learned at all, stating that

“Linguistic theory, the theory of UG [Universal Grammar]... is an innate property of the human mind.... [and].... the growth of language [is] analogous to the development of a bodily organ”.

Of course Chomsky acknowledges that French children learn French words, Chinese children learn Chinese words, and so on. But he believes that the abstract underlying principles that govern language are not learned at all, arguing that “A general learning theory ... seems to me dubious, unargued, and without any empirical support”.

Because this theory has been so influential in modern linguistics and psycholinguistics, it is important to understand exactly what Chomsky means by “innate.” Everyone would agree that there is something unique about the human brain that makes language possible. But in the absence of evidence to the contrary, that “something” could be nothing other than the fact that our brains are very large, a giant all-purpose computer with trillions of processing elements. Chomsky’s version of the theory of innateness is much stronger than the “big brain” view, and involves two logically and empirically separate claims: that language is innate, and that our brains contain a dedicated, special-purpose learning device that has evolved for language alone. The latter claim is the one that is really controversial, a doctrine that goes under various names including “domain specificity”, “autonomy” and “modularity”.

The second position is called “empiricism”, defined as the belief that knowledge originates in the environment, and comes in through the senses. This approach (also called “behaviorism” and “associationism”) is also an ancient one, going back (at least) to Aristotle, but in modern times it is closely associated with the psychologist B.F. Skinner (see photograph). According to Skinner, there are no limits to what a human being can become, given time, opportunity and the application of very general laws of learning. Humans are capable of language because we have the time, the opportunity and (perhaps) the computing power that is required to learn 50,000 words and the associations that link those words together. Much of the research that has taken place in linguistics, psycholinguistics and neurolinguistics since the 1950’s has been dedicated to proving Skinner wrong, by showing that children and adults go beyond their input, creating novel sentences and (in the case of normal children and brain-damaged adults) peculiar errors that they have never heard before. Chomsky himself has been severe in his criticisms of the behaviorist approach to language, denouncing those who believe that

language can be learned as “grotesquely wrong” (Gelman, 1986).

In their zealous attack on the behaviorist approach, nativists sometimes confuse Skinner’s form of empiricism with a very different approach, alternatively called “interactionism”, “constructivism,” and “emergentism.” This is a much more difficult idea than either nativism or empiricism, and its historical roots are less clear. In the 20th century, the interactionist or constructivist approach has been most closely associated with the psychologist Jean Piaget (see photograph). More recently, it has appeared in a new approach to learning and development in brains and brain-like computers alternatively called “connectionism,” “parallel distributed processing” and “neural networks” (Elman et al., 1996; Rumelhart & McClelland, 1986), and in a related theory of development inspired by the nonlinear dynamical systems of modern physics (Thelen & Smith, 1994). To understand this difficult but important idea, we need to distinguish between two kinds of interactionism: *simple interactions* (black and white make grey) and *emergent form* (black and white get together and something altogether new and different happens).

In an emergentist theory, outcomes can arise for reasons that are not obvious or predictable from any of the individual inputs to the problem. Soap bubbles are round because a sphere is the only possible solution to achieving maximum volume with minimum surface (i.e., their spherical form is not explained by the soap, the water, or the little boy who blows the bubble). The honeycomb in a beehive takes an hexagonal form because that is the stable solution to the problem of packing circles together (i.e., the hexagon is not predictable from the wax, the honey it contains, nor from the packing behavior of an individual bee—see Figure 1). Jean Piaget argued that logic and knowledge emerge in just such a fashion, from successive interactions between sensorimotor activity and a structured world. A similar argument has been made to explain the emergence of grammars, which represent the class of possible solutions to the problem of mapping a rich set of meanings onto a limited speech channel, heavily constrained by the limits of memory, perception and motor planning. Logic and grammar are not given in the world, but neither are they given in the genes. Human beings discovered the principles that comprise logic and grammar, because these principles were the best possible solution to specific problems that other species just simply do not care about, and could not solve even if they did. Proponents of the emergentist view acknowledge that something is innate in the human brain that makes language possible, but that “something” may not be a special-purpose, domain-specific device that evolved for language and language alone. Instead, language may be something that we do with a large and complex brain that evolved to serve the many complex goals of human society and culture (Tomasello & Call, 1997). In other words, language is

a new machine built out of old parts, reconstructed from those parts by every human child.

So the debate today in language research is not about Nature vs. Nurture, but about the “nature of Nature,” that is, whether language is something that we do with an inborn language device, or whether it is the product of (innate) abilities that are not specific to language. In the pages that follow, we will explore current knowledge about the psychology, neurology and development of language from this point of view. We will approach this problem at different levels of the system, from speech sounds to the broader communicative structures of complex discourse. Let us start by defining the different levels of the language system, and then go on to describe how each of these levels is processed by normal adults, acquired by children, and represented in the brain.

I. THE COMPONENT PARTS OF LANGUAGE

Speech as Sound: Phonetics and Phonology

The study of speech sounds can be divided into two subfields: *phonetics* and *phonology*.

Phonetics is the study of speech sounds as physical and psychological events. This includes a huge body of research on the acoustic properties of speech, and the relationship between these acoustic features and the way that speech is perceived and experienced by humans. It also includes the detailed study of speech as a motor system, with a combined emphasis on the anatomy and physiology of speech production. Within the field of phonetics, linguists work side by side with acoustical engineers, experimental psychologists, computer scientists and biomedical researchers.

Phonology is a very different discipline, focused on the abstract representations that underlie speech in both perception and production, within and across human languages. For example, a phonologist may concentrate on the rules that govern the voiced/voiceless contrast in English grammar, e.g., the contrast between the unvoiced “-s” in “cats” and the voiced “-s” in “dogs”. This contrast in plural formation bears an uncanny resemblance to the voiced/unvoiced contrast in English past tense formation, e.g., the contrast between an unvoiced “-ed” in “walked” and a voiced “-ed” in “wagged”. Phonologists seek a maximally general set of rules or principles that can explain similarities of this sort, and generalize to new cases of word formation in a particular language. Hence phonology lies at the interface between phonetics and the other regularities that constitute a human language, one step removed from sound as a physical event.

Some have argued that phonology should not exist as a separate discipline, and that the generalizations discovered by phonologists will ultimately be explained entirely in physical and psychophysical terms. This tends to be the approach taken by emergentists. Others maintain that phonology is a completely independent level of analysis, whose laws cannot be reduced to any combination of physical events. Not surprisingly, this

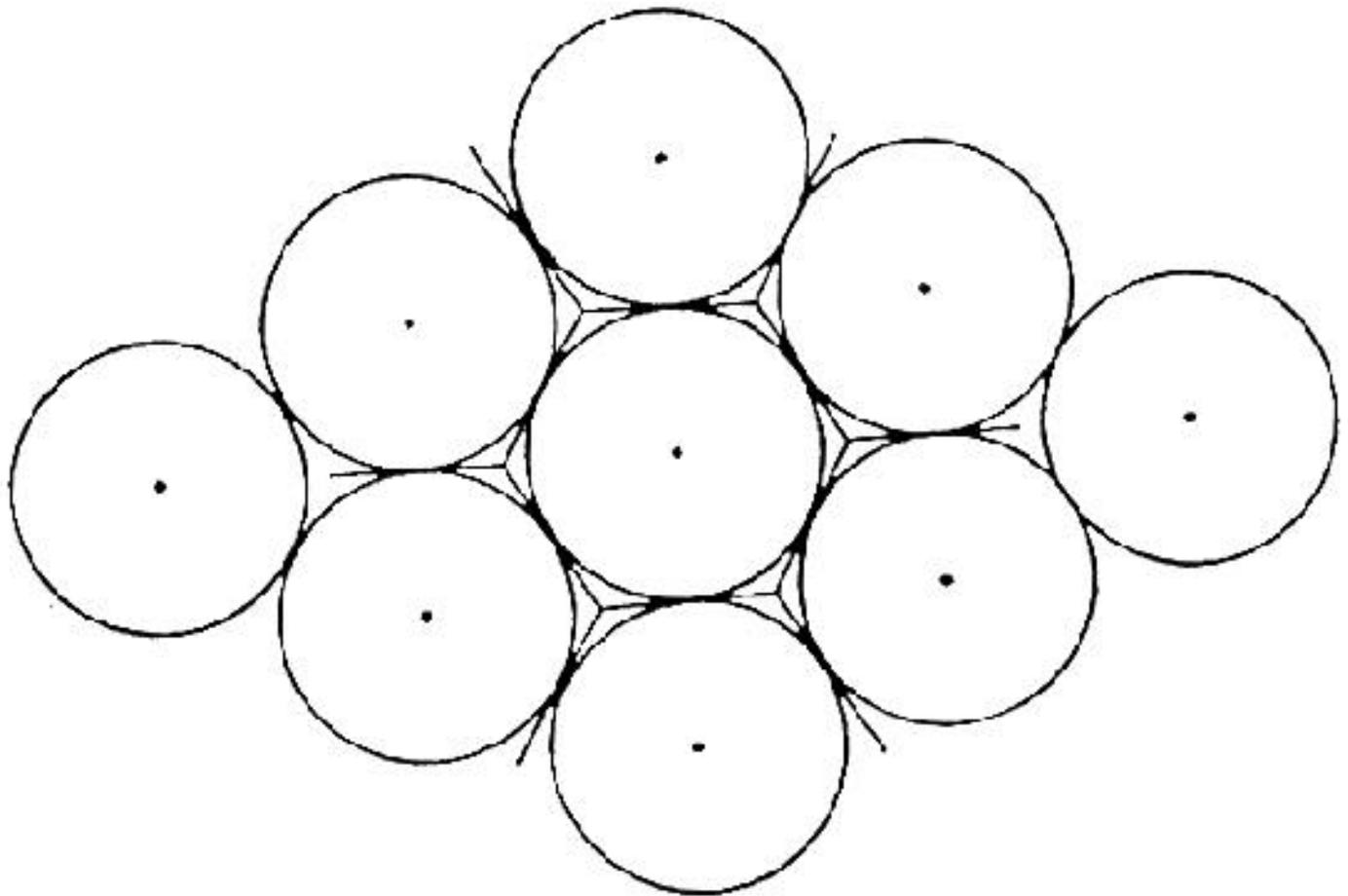


Figure 1. Diagram of hexagonal cells. After Bonanni.

tends to be the approach taken by nativists, especially those who believe that language has its very own dedicated neural machinery. Regardless of one's position on this debate, it is clear that phonetics and phonology are not the same thing. If we analyze speech sounds from a phonetic point of view, based on all the different sounds that a human speech apparatus can make, we come up with approximately 600 possible sound contrasts that languages could use (even more, if we use a really fine-grained system for categorizing sounds). And yet most human languages use no more than 40 contrasts to build words.

To illustrate this point, consider the following contrast between English and French. In English, the aspirated (or "breathy") sound signalled by the letter "h" is used phonologically, e.g., to signal the difference between "at" and "hat". French speakers are perfectly capable of making these sounds, but the contrast created by the presence or absence of aspiration ("h-") is not used to mark a systematic difference between words; instead, it is just a meaningless variation that occurs now and then in fluent speech, largely ignored by listeners. Similarly, the English language has a binary contrast between the sounds signalled by "d" and "t", used to make systematic contrasts like "tune" and "dune." The Thai language has both these contrasts, and in addition it has a third boundary somewhere in between the English "t" and "d". English speakers are able to produce that third boundary; in fact, it is the normal way to pronounce the middle consonant in a word like "butter". The difference is that Thai uses that third contrast phonologically (to make new words), but English only uses it phonetically, as a convenient way to pronounce target phonemes while hurrying from one word to another (also called "allophonic variation"). In our review of studies that focus on the processing, development and neural bases of speech sounds, it will be useful to distinguish between the phonetic approach, and phonological or phonemic approach.

Speech as Meaning: Semantics and the Lexicon

The study of linguistic meaning takes place within a subfield of linguistics called *semantics*. Semantics is also a subdiscipline within philosophy, where the relationship between meaning and formal logic is emphasized. Traditionally semantics can be divided into two areas: *lexical semantics*, focussed on the meanings associated with individual lexical items (i.e., words), and *propositional* or *relational semantics*, focussed on those relational meanings that we typically express with a whole sentence.

Lexical semantics has been studied by linguists from many different schools, ranging from the heavily descriptive work of lexicographers (i.e., "dictionary writers") to theoretical research on lexical meaning and lexical form in widely different schools of formal linguistics and generative grammar (McCawley, 1993). Some of these theorists emphasize the intimate relationship between semantics and grammar, using a

combination of lexical and propositional semantics to explain the various meanings that are codified in the grammar. This is the position taken by many theorists who taken an emergentist approach to language, including specific schools with names like "cognitive grammar," "generative semantics" and/or "linguistic functionalism". Other theorists argue instead for the structural independence of semantics and grammar, a position associated with many of those who espouse a nativist approach to language.

Propositional semantics has been dominated primarily by philosophers of language, who are interested in the relationship between the logic that underlies natural language and the range of possible logical systems that have been uncovered in the last two centuries of research on formal reasoning. A *proposition* is defined as a statement that can be judged true or false. The internal structure of a proposition consists of a *predicate* and one or more *arguments* of that predicate. An argument is an entity or "thing" that we would like to make some point about. A one-place predicate is a state, activity or identity that we attribute to a single entity (e.g., we attribute beauty to Mary in the sentence "Mary is beautiful", or we attribute "engineerness" to a particular individual in the sentence "John is an engineer."); an n-place predicate is a relationship that we attribute to two or more entities or things. For example, the verb "to kiss" is a two-place predicate, which establishes an asymmetric relationship of "kissing" to two entities in the sentence "John kisses Mary.", The verb "to give" is a three-place predicate that relates three entities in a proposition expressed by the sentence "John gives Mary a book.." Philosophers tend to worry about how to determine the truth or falsity of propositions, and how we convey (or hide) truth in natural language and/or in artificial languages. Linguists worry about how to characterize or taxonomize the propositional forms that are used in natural language. Psychologists tend instead to worry about the shape and nature of the mental representations that encode propositional knowledge, with developmental psychologists emphasizing the process by which children attain the ability to express this propositional knowledge. Across fields, those who take a nativist approach to the nature of human language tend to emphasize the independence of propositional or combinatorial meaning from the rules for combining words in the grammar; by contrast, the various emergentist schools tend to emphasize both the structural similarity and the causal relationship between propositional meanings and grammatical structure, suggesting that one grows out of the other.

How Sounds and Meanings Come Together: Grammar

The subfield of linguistics that studies how individual words and other sounds are combined to express meaning is called *grammar*. The study of grammar is traditionally divided into two parts: *morphology* and *syntax*.

Morphology refers to the principles governing the construction of complex words and phrases, for lexical and/or grammatical purposes. This field is further divided into two subtypes: *derivational morphology* and *inflectional morphology*.

Derivational morphology deals with the construction of complex content words from simpler components, e.g., derivation of the word "government" from the verb "to govern" and the derivational morpheme "-ment". Some have argued that derivational morphology actually belongs within lexical semantics, and should not be treated within the grammar at all. However, such an alignment between derivational morphology and semantics describes a language like English better than it does richly inflected languages like Greenlandic Eskimo, where a whole sentence may consist of one word with many different derivational and inflectional morphemes.

Inflectional morphology refers to modulations of word structure that have grammatical consequences, modulations that are achieved by *inflection* (e.g., adding an "-ed" to a verb to form the past tense, as in "walked") or by *suppletion* (e.g., substituting the irregular past tense "went" for the present tense "go"). Some linguists would also include within inflectional morphology the study of how free-standing function words (like "have", "by", or "the", for example) are added to individual verbs or nouns to build up complex verb or noun phrases, e.g., the process that expands a verb like "run" into "has been running" or the process that expands a noun like "dog" into a noun phrase like "the dog" or prepositional phrase like "by the dog".

Syntax is defined as the set of principles that govern how words and other morphemes are ordered to form a possible sentence in a given language. For example, the syntax of English contains principles that explain why "John kissed Mary" is a possible sentence while "John has Mary kissed" sounds quite strange. Note that both these sentences would be acceptable in German, so to some extent these rules and constraints are arbitrary. Syntax may also contain principles that describe the relationship between different forms of the same sentence (e.g., the active sentence "John hit Bill" and the passive form "Bill was hit by John"), and ways to nest one sentence inside another (e.g., "The boy that was hit by John hit Bill").

Languages vary a great deal in the degree to which they rely on syntax or morphology to express basic propositional meanings. A particularly good example is the cross-linguistic variation we find in means of expressing a propositional relation called *transitivity* (loosely defined as "who did what to whom"). English uses word order as a regular and reliable cue to sentence meaning (e.g., in the sentence "John kissed a girl", we immediately know that "John" is the actor and "girl" is the receiver of that action). At the same time, English makes relatively little use of inflectional morphology to indicate transitivity or (for that matter) any other important aspect of sentence meaning. For example, there are no markers on "John" or "girl" to tell us who

kissed whom, nor are there any clues to transitivity marked on the verb "kissed". The opposite is true in Hungarian, which has an extremely rich morphological system but a high degree of word order variability. Sentences like "John kissed a girl" can be expressed in almost every possible order in Hungarian, without loss of meaning.

Some linguists have argued that this kind of word order variation is only possible in a language with rich morphological marking. For example, the Hungarian language provides case suffixes on each noun that unambiguously indicate who did what to whom, together with special markers on the verb that agree with the object in definiteness. Hence the Hungarian translation of our English example would be equivalent to "John-actor indefinite-girl-receiver-of-action kissed-indefinite). However, the Chinese language poses a problem for this view: Chinese has no inflectional markings of any kind (e.g., no case markers, no form of agreement), and yet it permits extensive word order variation for stylistic purposes. As a result, Chinese listeners have to rely entirely on probabilistic cues to figure out "who did what to whom", including some combination of word order (i.e., some orders are more likely than others, even though many are possible) and the semantic content of the sentence (e.g., boys are more likely to eat apples than vice-versa). In short, it now seems clear that human languages have solved this mapping problem in a variety of ways.

Chomsky and his followers have defined Universal Grammar as the set of possible forms that the grammar of a natural language can take. There are two ways of looking at such universals: as the *intersect of all human grammars* (i.e., the set of structures that every language has to have) or as the *union of all human grammars* (i.e., the set of possible structures from which each language must choose). Chomsky has always maintained that Universal Grammar is innate, in a form that is idiosyncratic to language. That is, grammar does not "look like" or behave like any other existing cognitive system. However, he has changed his mind across the years on the way in which this innate knowledge is realized in specific languages like Chinese or French. In the early days of generative grammar, the search for universals revolved around the idea of a universal intersect. As the huge variations that exist between languages became more and more obvious, and the intersect got smaller and smaller, Chomsky began to shift his focus from the intersect to the union of possible grammars. In essence, he now assumes that children are born with a set of innate options that define how linguistic objects like nouns and verbs can be put together. The child doesn't really learn grammar (in the sense in which the child might learn chess). Instead, the linguistic environment serves as a "trigger" that selects some options and causes others to wither away. This process is called "parameter setting". Parameter setting may resemble learning, in that it helps to explain why languages look as different as they do and how children move toward their language-specific

targets. However, Chomsky and his followers are convinced that parameter setting (choice from a large stock of innate options) is not the same thing as learning (acquiring a new structure that was never there before learning took place), and that learning in the latter sense plays a limited and perhaps rather trivial role in the development of grammar.

Many theorists disagree with this approach to grammar, along the lines that we have already laid out. Empiricists would argue that parameter setting really is nothing other than garden-variety learning (i.e., children really are taking new things in from the environment, and not just selecting among innate options). Emergentists take yet another approach, somewhere in between parameter setting and learning. Specifically, an emergentist would argue that some combinations of grammatical features are more convenient to process than others. These facts about processing set limits on the class of possible grammars: Some combinations work; some don't. To offer an analogy, why is it that a sparrow can fly but an emu cannot? Does the emu lack "innate flying knowledge," or does it simply lack a relationship between weight and wingspan that is crucial to the flying process? The same logic can be applied to grammar. For example, no language has a grammatical rule in which we turn a statement into a question by running the statement backwards, e.g.,

John hit the ball" --> Ball the hit John?

Chomsky would argue that such a rule does not exist because it is not contained within Universal Grammar. It could exist, but it doesn't. Emergentists would argue that such a rule does not exist because it would be very hard to produce or understand sentences in real time by a forward-backward principle. It might work for sentences that are three or four words long, but our memories would quickly fail beyond that point e.g.,

The boy that kicked the girl hit the ball that Peter bought -->

Bought Peter that ball the hit girl the kicked that boy the?

In other words, the backward rule for question formation doesn't exist because it couldn't exist, not with the kind of memory that we have to work with. Both approaches assume that grammars are the way they are because of the way that the human brain is built. The difference lies not in Nature vs. Nurture, but in the "nature of Nature," i.e., whether this ability is built out of language-specific materials or put together from more general cognitive ingredients.

Language in a Social Context: Pragmatics and Discourse

The various subdisciplines that we have reviewed so far reflect one or more aspects of linguistic form, from sound to words to grammar. **Pragmatics** is defined as the study of language in context, a field within linguistics and philosophy that concentrates instead on language as a form of communication, a tool that we use to accomplish certain social ends (Bates,

1976). Pragmatics is not a well-defined discipline; indeed, some have called it the wastebasket of linguistic theory. It includes the study of **speech acts** (a taxonomy of the socially recognized acts of communication that we carry out when we declare, command, question, baptize, curse, promise, marry, etc.), **presuppositions** (the background information that is necessary for a given speech act to work, e.g., the subtext that underlies a pernicious question like "Have you stopped beating your wife?"), and **conversational postulates** (principles governing conversation as a social activity, e.g., the set of signals that regulate turn-taking, and tacit knowledge of whether we have said too much or too little to make a particular point).

Pragmatics also contains the study of **discourse**. This includes the comparative study of discourse types (e.g., how to construct a paragraph, a story, or a joke), and the study of **text cohesion**, i.e., the way we use individual linguistic devices like conjunctions ("and", "so"), pronouns ("he", "she", "that one there"), definite articles ("the" versus "a") and even whole phrases or clauses (e.g., "The man that I told you about...") to tie sentences together, differentiate between old and new information, and maintain the identity of individual elements from one part of a story to another (i.e., **coreference relations**).

It should be obvious that pragmatics is a heterogeneous domain without firm boundaries. Among other things, mastery of linguistic pragmatics entails a great deal of sociocultural information: information about feelings and internal states, knowledge of how the discourse looks from the listener's point of view, and the relationships of power and intimacy between speakers that go into calculations of how polite and/or how explicit we need to be in trying to make a conversational point. Imagine a Martian that lands on earth with a complete knowledge of physics and mathematics, armed with computers that could break any possible code. Despite these powerful tools, it would be impossible for the Martian to figure out why we use language the way we do, unless that Martian also has extensive knowledge of human society and human emotions. For the same reason, this is one area of language where social-emotional disabilities could have a devastating effect on development (e.g., autistic children are especially bad on pragmatic tasks). Nevertheless, some linguists have tried to organize aspects of pragmatics into one or more independent "modules," each with its own innate properties (Sperber & Wilson, 1986). As we shall see later, there has also been a recent effort within neurolinguistics to identify a specific neural locus for the pragmatic aspect of linguistic knowledge.

Now that we have a road map to the component parts of language, let us take a brief tour of each level, reviewing current knowledge of how information at that level is processed by adults, acquired by children, and mediated in the human brain.

II. SPEECH SOUNDS

How Speech is Processed by Normal Adults

The study of speech processing from a psychological perspective began in earnest after World War II, when instruments became available that permitted the detailed analysis of speech as a physical event. The most important of these for research purposes was the sound spectrograph. Unlike the more familiar oscilloscope, which displays sound frequencies over time, the spectrograph displays changes over time in the energy contained within different frequency bands (think of the vertical axis as a car radio, while the horizontal axis displays activity on every station over time). Figure 2 provides an example of a sound spectrogram for the sentence “Is language innate?”—one of the central questions in this field.

This kind of display proved useful not only because it permitted the visual analysis of speech sounds, but also because it became possible to “paint” artificial speech sounds and play them back to determine their effects on perception by a live human being. Initially scientists hoped that this device would form the basis of speech-reading systems for the deaf. All we would have to do (or so it seemed) would be to figure out the “alphabet”, i.e., the visual pattern that corresponds to each of the major phonemes in the language. By a similar argument, it should be possible to create computer systems that understand speech, so that we could simply walk up to a banking machine and tell it our password, the amount of money we want, and so forth. Unfortunately, it wasn’t that simple. As it turns out, there is no clean, isomorphic relation between the speech sounds that native speakers hear and the visual display produced by those sounds. Specifically, the relationship between speech signals and speech perception lacks two critical properties: linearity and invariance.

Linearity refers to the way that speech unfolds in time. If the speech signal had linearity, then there would be an isomorphic relation from left to right between speech-as-signal and speech-as-experience in the speech spectrogram. For example, consider the syllable “da” displayed in the artificial spectrogram in Figure 3. If the speech signal were linear, then the first part of this sound (the “d” component) should correspond to the first part of the spectrogram, and the second part (the “a” component) should correspond to the second part of the same spectrogram. However, if we play these two components separately to a native speaker, they don’t sound anything like two halves of “da”. The vowel sound does indeed sound like a vowel “a”, but the “d” component presented alone (with no vowel context) doesn’t sound like speech at all; it sounds more like the chirp of a small bird or a speaking wheel on a rolling chair. It would appear that our experience of speech involves a certain amount of reordering and integration of the physical signal as it comes in, to create the unified perceptual experience that is so familiar to us all.

Invariance refers to the relationship between the signal and its perception across different contexts. Even though the signal lacks linearity, scientists once hoped that the same portion of the spectrogram that elicits the “d” experience in the context of “di” would also correspond to the “d” experience in the context of “du”. Alas, that has proven not to be the case. As Figure 3 shows, the component responsible for “d” looks entirely different depending on the vowel that follows. Worse still, the “d” component of the syllable “du” looks like the “g” component of the syllable “ga”. In fact, the shape of the visual pattern that corresponds to a constant sound can even vary with the pitch of the speaker’s voice, so that the “da” produced by a small child results in a very different-looking pattern from the “da” produced by a mature adult male.

These problems can be observed in clean, artificial speech stimuli. In fluent, connected speech the problems are even worse (see word perception, below). It seems that native speakers use many different parts of the context to break the speech code. No simple “bottom-up” system of rules is sufficient to accomplish this task. That is why we still don’t have speech readers for the deaf or computers that perceive fluent speech from many different listeners, even though such machines have existed in science fiction for decades.

The problem of speech perception got “curiouser and curiouser” as Lewis Carroll would say, leading a number of speech scientists in the 1960’s to propose that humans accomplish speech perception via a special-purpose device unique to the human brain. For reasons that we will come to shortly, they were also persuaded that this “speech perception device” is innate, up and running in human babies as soon as they are born. It was also suggested that humans process these speech sounds not as acoustic events, but by testing the speech input against possible “motor templates” (i.e., versions of the same speech sound that the listener can produce for himself, a kind of “analysis by synthesis”). This idea, called the Motor Theory of Speech Perception, was offered to explain why the processing of speech is nonlinear and invariant from an acoustic point of view, and why only humans (or so it was believed) are able to perceive speech at all.

For a variety of reasons (some discussed below) this hypothesis has fallen on hard times. Today we find a large number of speech scientists returning to the idea that speech is an acoustic event after all, albeit a very complicated one that is hard to understand by looking at speech spectrograms like the ones in Figures 2-3. For one thing, researchers using a particular type of computational device called a “neural network” have shown that the basic units of speech can be learned after all, even by a rather stupid machine with access to nothing other than raw acoustic speech input (i.e., no “motor templates” to fit against the signal). So the ability to perceive these units does not have to be innate; it can be learned. This brings us to the next point: how speech develops.

Narrow-band sound spectrogram of the sentence:
"Is language innate?"

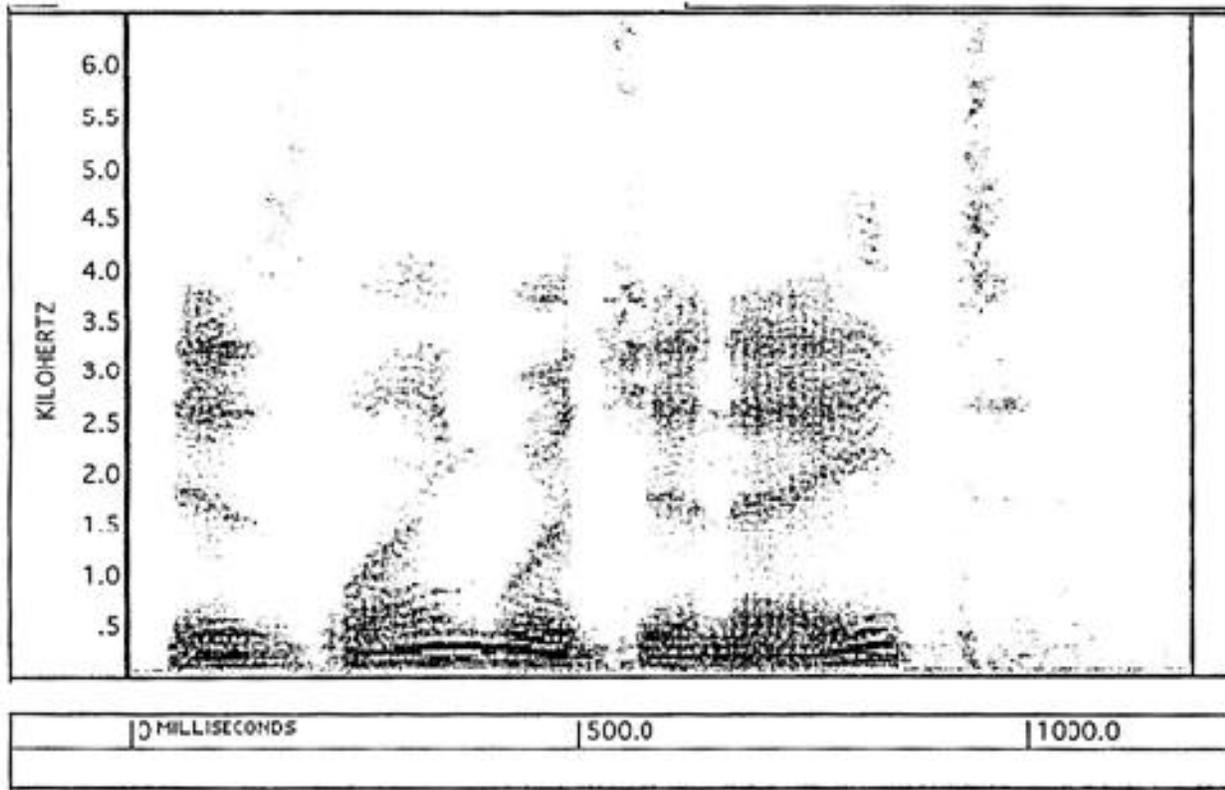
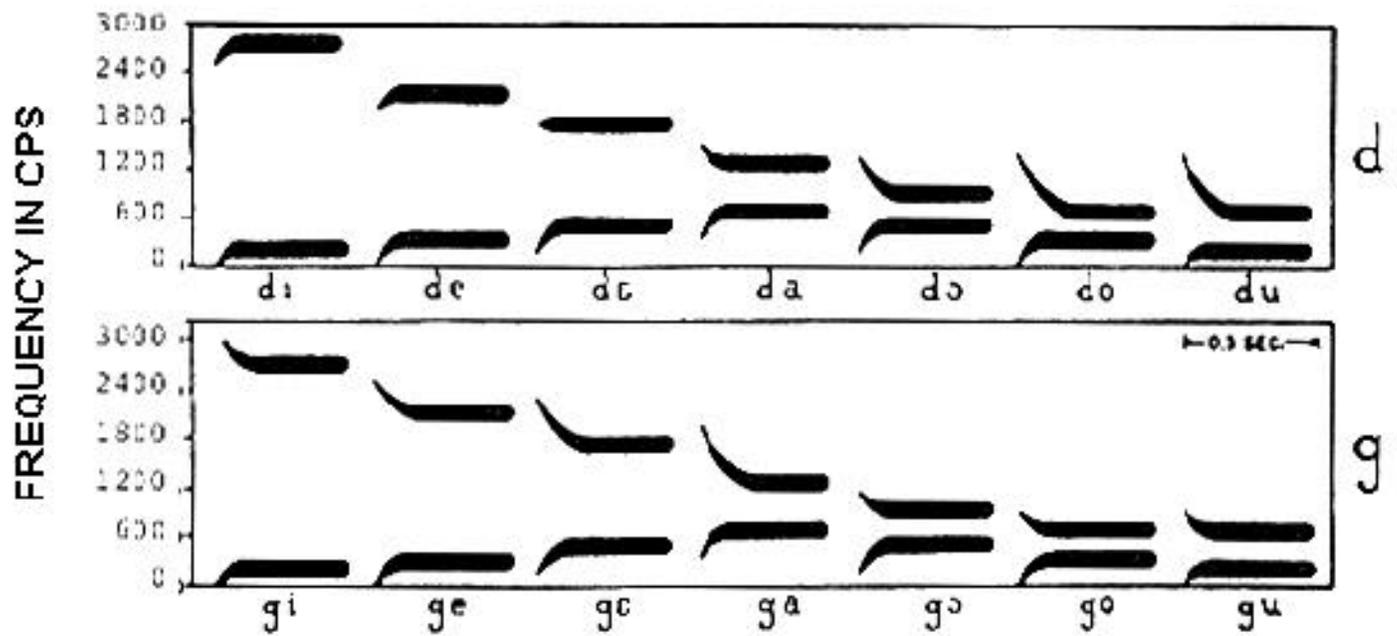


Figure 2



Schematized sound spectrograms for two consonants in different vowel environments.

(Adapted from Aslin, Pisoni, & Jusczyk, 1983)

FIGURE 3

How Speech Sounds Develop in Children

Speech Perception. A series of clever techniques has been developed to determine the set of phonetic/phonemic contrasts that are perceived by preverbal infants. These include High-Amplitude Sucking (capitalizing on the fact that infants tend to suck vigorously when they are attending to an interesting or novel stimulus), habituation and dishabituation (relying on the tendency for small infants to “orient” or re-attend when they perceive an interesting change in auditory or visual input), and operant generalization (e.g., training an infant to turn her head to the sounds from one speech category but not another, a technique that permits the investigator to map out the boundaries between categories from the infant’s point of view). (For reviews of research using these techniques, see Aslin, Jusczyk, & Pisoni, in press). Although the techniques have remained constant over a period of 20 years or more, our understanding of the infant’s initial repertoire and the way that it changes over time has undergone substantial revision.

In 1971, Peter Eimas and colleagues published an important paper showing that human infants are able to perceive contrasts between speech sounds like “pa” and “ba” (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Even more importantly, they were able to show that infants hear these sounds categorically. To illustrate this point, the reader is kindly requested to place one hand on her throat and the other (index finger raised) just in front of her mouth. Alternate back and forth between the sounds “pa” and “ba”, and you will notice that the mouth opens before the vocal chords rattle in making a “pa”, while the time between vocal chord vibrations and lip opening is much shorter when making a “ba.” This variable, called Voice Onset Time or VOT, is a continuous dimension in physics but a discontinuous one in human perception. That is, normal listeners hear a sharp and discontinuous boundary somewhere around +20 (20 msec between mouth opening and voice onset); prior to that boundary, the “ba” tokens sound very much alike, and after that point the “pa” tokens are difficult to distinguish, but at that boundary a dramatic shift can be heard. To find out whether human infants have such a boundary, Eimas et al. used the high-amplitude sucking procedure, which means that they set out to habituate (literally “bore”) infants with a series of stimuli from one category (e.g., “ba”), and then presented with new versions in which the VOT is shifted gradually toward the adult boundary. Sucking returned with a vengeance (“Hey, this is new!”), a sharp change right at the same border at which human adults hear a consonantal contrast.

Does this mean that the ability to hear categorical distinctions in speech is innate? Yes, it probably does. Does it also mean that this ability is based on an innate processor that has evolved exclusively for speech? Eimas et al. thought so, but history has shown that they were wrong. Kuhl & Miller (1975) made a discovery that was devastating for the nativist approach: Chinchillas can also hear the boundary between

consonants, and they hear it categorically, with a boundary at around the same place where humans hear it. This finding has now been replicated in various species, with various methods, looking at many different aspects of the speech signal. In other words, categorical speech perception is not domain specific, and did not evolve in the service of speech. The ear did not evolve to meet the human mouth; rather, human speech has evolved to take advantage of distinctions that were already present in the mammalian auditory system. This is a particularly clear illustration of the difference between innateness and domain specificity outlined in the introduction.

Since the discovery that categorical speech perception and phenomena are not peculiar to speech, the focus of interest in research on infant speech perception has shifted away from interest in the initial state to interest in the process by which children tune their perception to fit the peculiarities of their own native language. We now know that newborns can hear virtually every phonetic contrast used by human languages. Indeed, they can hear things that are no longer perceivable to an adult. For example, Japanese listeners find it very difficult to hear the contrast between “ra” and “la”, but Japanese infants have no trouble at all with that distinction. When do we lose (or suppress) the ability to hear sounds that are not in our language? Current evidence suggests that the suppression of non-native speech sounds begins somewhere between 8-10 months of age—the point at which most infants start to display systematic evidence of word comprehension. There is, it seems, no such thing as a free lunch: In order to “tune in” to language-specific sounds in order to extract their meaning, the child must “tune out” to those phonetic variations that are not used in the language.

This does not mean, however, that children are “language neutral” before 10 months of age. Studies have now shown that infants are learning something about the sounds of their native language in utero! French children turn preferentially to listen to French in the first hours and days of life, a preference that is *not* shown by newborns who were bathed in a different language during the last trimester. Whatever the French may believe about the universal appeal of their language, it is not preferred universally at birth. Something about the sound patterns of one’s native language is penetrating the womb, and the ability to learn something about those patterns is present in the last trimester of pregnancy. However, that “something” is probably rather vague and imprecise. Kuhl and her colleagues have shown that a much more precise form of language-specific learning takes place between birth and six months of age, in the form of a preference for prototypical vowel sounds in that child’s language. Furthermore, the number of “preferred vowels” is tuned along language-specific lines by six months of age. Language-specific information about consonants appears to come in somewhat later, and probably coincides with

the suppression of non-native contrasts and the ability to comprehend words.

To summarize, human infants start out with a universal ability to hear all the speech sounds used by any natural language. This ability is innate, but it is apparently not specific to speech (nor specific to humans). Learning about the speech signal begins as soon as the auditory system is functional (somewhere in the last trimester of pregnancy), and it proceeds systematically across the first year of life until children are finally able to weed out irrelevant sounds and tune into the specific phonological boundaries of their native language. At that point, mere speech turns into real language, i.e., the ability to turn sound into meaning.

Speech production. This view of the development of speech perception is complemented by findings on the development of speech production (for details see Menn & Stoel-Gammon, 1995).

In the first two months, the sounds produced by human infants are reflexive in nature, “vegetative sounds” that are tied to specific internal states (e.g., crying). Between 2-6 months, infants begin to produce vowel sounds (i.e., cooing and sound play). So-called canonical or reduplicative babbling starts between 6-8 months in most children: babbling in short segments or in longer strings that are now punctuated by consonants (e.g., “dadada”). In the 6-12-month window, babbling “drifts” toward the particular sound patterns of the child’s native language (so that adult listeners can distinguish between the babbling of Chinese, French and Arabic infants). However, we still do not know what features of the infants’ babble lead to this discrimination (i.e., whether it is based on consonants, syllable structure and/or the intonational characteristics of infant speech sounds). In fact, some investigators insist that production of consonants may be relatively immune to language-specific effects until the second year of life.

Around 10 months of age, some children begin to produce “word-like sounds”, used in relatively consistent ways in particular contexts (e.g., “nana” as a sound made in requests; “bam!” pronounced in games of knocking over toys). From this point on (if not before), infant phonological development is strongly influenced by other aspects of language learning (i.e., grammar and the lexicon). There is considerable variability between infants in the particular speech sounds that they prefer. However, there is clear continuity from prespeech babble to first words in an individual infant’s “favorite sounds”. This finding contradicts a famous prediction by the linguist Roman Jakobson, who believed that prespeech babble and meaningful speech are discontinuous. Phonological development has a strong influence on the first words that children try to produce (i.e., they will avoid the use of words that they cannot pronounce, and collect new words as soon as they develop an appropriate “phonological template” for those words). Conversely, lexical development has a strong influence on the sounds that a child produces; specifically, the child’s

“favorite phonemes” tend to derive from the sounds that are present in his first and favorite words. In fact, children appear to treat these lexical/phonological prototypes like a kind of basecamp, exploring the world of sound in various directions without losing sight of home.

Phonological development interacts with lexical and grammatical development for at least three years. For example, children who have difficulty with a particular sound (e.g., the sibilant “-s”) appear to postpone productive use of grammatical inflections that contain that sound (e.g., the plural). A rather different lexical/phonological interaction is illustrated by many cases in which the “same” speech sound is produced correctly in one word context but incorrectly in another (e.g., the child may say “guck” for “duck”, but have no trouble pronouncing the “d” in “doll”). This is due, we think, to articulatory facts: It is hard to move the speech apparatus back and forth between the dental position (in “d”) and the glottal position (in the “k” part of “duck”), so the child compromises by using one position only (“guck”). That is, the child may be capable of producing all the relevant sounds in isolation, but finds it hard to produce them in the combinations required for certain word targets.

After 3 years of age, when lexical and grammatical development have “settled down”, phonology also becomes more stable and systematic: Either the child produces no obvious errors at all, or s/he may persist in the same phonological error (e.g., a difficulty pronouncing “r” and “l”) regardless of lexical context, for many more years. The remainder of lexical development from 3 years to adulthood can generally be summarized as an increase in fluency, including a phenomenon called “coarticulation”, in which those sounds that will be produced later on in an utterance are anticipated by moving the mouth into position on an earlier speech sound (hence the “b” in “bee” is qualitatively different from the “b” in “boat”).

To summarize, the development of speech as a sound system begins at or perhaps before birth (in speech perception), and continues into the adult years (e.g., with steady increases in fluency and coarticulation throughout the first two decades of life). However, there is one point in phonetic and phonological development that can be viewed as a kind of watershed: 8-10 months, marked by changes in perception (e.g., the inhibition of non-native speech sounds) and changes in production (e.g., the onset of canonical babbling and “phonological drift”). The timing of these milestones in speech may be related to some important events in human brain development that occur around the same time, including the onset of synaptogenesis (a “burst” in synaptic growth that begins around 8 months and peaks somewhere between 2-3 years of age), together with evidence for changes in metabolic activity within the frontal lobes, and an increase in frontal control over other cortical and subcortical functions (Elman et al., 1996). This interesting correlation between brain and behavioral development is not restricted to changes in

speech; indeed, the 8-10-month period is marked by dramatic changes in many different cognitive and social domains, including developments in tool use, categorization and memory for objects, imitation, and intentional communication via gestures (see Volterra, this volume). In other words, the most dramatic moments in speech development appear to be linked to change outside the boundaries of language, further evidence that our capacity for language depends on nonlinguistic factors. This brings us to the next point: a brief review of current evidence on the neural substrates of speech.

Brain Bases of Speech Perception and Production

In this section, and in all the sections on the brain bases of language that will follow, the discussion will be divided to reflect data from the two principal methodologies of neurolinguistics and cognitive neuroscience: evidence from patients with unilateral lesions (a very old method) and evidence from the application of functional brain-imaging techniques to language processing in normal people (a brand-new method).

Although one hopes that these two lines of evidence will ultimately converge, yielding a unified view of brain organization for language, we should not be surprised to find that they often yield different results. Studies investigating the effects of focal lesions on language behavior can tell us what regions of the brain are *necessary* for normal language use. Studies that employ brain-imaging techniques in normals can tell us what regions of the brain *participate* in normal language use. These are not necessarily the same thing.

Even more important for our purposes here, lesion studies and neural imaging techniques cannot tell us where language or any other higher cognitive function is located, i.e., where the relevant knowledge "lives," independent of any specific task. The plug on the wall behind a television set is necessary for the television's normal function (just try unplugging your television, and you will see how important it is). It is also fair to say that the plug participates actively in the process by which pictures are displayed. That doesn't mean that the picture is located in the plug! Indeed, it doesn't even mean that the picture passes through the plug on its way to the screen. Localization studies are controversial—and they deserve to be! Figure 4 displays one version of the phrenological map of Gall and Spurzheim, proposed in the 18th century, and still the best-known and most ridiculed version of the idea that higher faculties are located within discrete areas of the brain. Although this particular map of the brain is not taken seriously anymore, modern variants of the phrenological doctrine are still around, e.g., proposals that free will lives in frontal cortex, faces live in the temporal lobe, and language lives in two places on the left side of the brain, one in the front (called Broca's area) and another in the back (called Wernicke's area). As we shall see throughout this brief review, there is no

phrenological view of language that can account for current evidence from lesion studies, or from neural imaging of the working brain. Rather, language seems to be an event that is staged by many different areas of the brain, a complex process that is not located in a single place. Having said that, it should also be noted that some places are more important than others, even if they should not be viewed as the "language box." The dancer's many skills are not located in her feet, but her feet are certainly more important than a number of other body parts. In the same vein, some areas of the brain have proven to be particularly important for normal language use, even though we should not conclude that language is located there. With those warnings in mind, let us take a brief look at the literature on the brain bases of speech perception and production.

Lesion studies of speech. We have known for a very long time that injuries to the head can impair the ability to perceive and produce speech. Indeed, the first case that appeared in the Edmund Smith Surgical Papyrus, attributed to the Egyptian Imhotep. However, little progress was made beyond that simple observation until the 19th century. In 1861, Paul Broca observed a patient called "Tan" who appeared to understand the speech that other people directed to him; however, Tan was completely incapable of meaningful speech production, restricted entirely to the single syllable for which he was named. This patient died and came to autopsy a few days after Broca tested him. An image of that brain (preserved for posterity) appears in Figure 5. Casual observation of this figure reveals a massive cavity in the third convolution of the left frontal lobe, a region that is now known as Broca's area. Broca and his colleagues proposed that the capacity for speech output resides in this region of the brain.

Across the next few decades, European investigators set out in search of other sites for the language faculty. The most prominent of these was Carl Wernicke, who described a different lesion that seemed to be responsible for severe deficits in comprehension, in patients who are nevertheless capable of fluent speech. This region (now known as Wernicke's area) lay in the left hemisphere as well, along the superior temporal gyrus close to the junction of the temporal, parietal and occipital lobes. It was proposed that this region is the site of speech perception, connected to Broca's area in the front of the brain by a series of fibres called the arcuate fasciculus. Patients who have damage to the fibre bundle only should prove to be incapable of repeating words that they hear, even though they are able to produce spontaneous speech and understand most of the speech that they hear. This third syndrome (called "conduction aphasia") was proposed on the basis of Wernicke's theory, and in the next few years a number of investigators claimed to have found evidence for its existence. Building on this model of brain organization for language, additional areas were proposed to underlie reading (to explain "alexia") and writing (responsible for "agraphia"), and arguments raged about the relative



Photograph of the lesioned brain of Broca's first patient, "Tan".
(From Signoret, Castaigne, Lhermitte, Abelanet, & Lavoret, 1984)

FIGURE 6

separability or dissociability of the emerging aphasia syndromes (e.g., is there such a thing as alexia without agraphia?).

This neophrenological view has had its critics at every point in the modern history of aphasiology, including Freud's famous book "On aphasia", which ridicules the Wernicke-Lichtheim model (Freud, 1891/1953), and Head's witty and influential critique of localizationists, whom he referred to as "The Diagram Makers" (Head, 1926). The localizationist view fell on hard times in the period between 1930-1960, the Behaviorist Era in psychology when emphasis was given to the role of learning and the plasticity of the brain. But it was revived with a vengeance in the 1960's, due in part to Norman Geschwind's influential writings and to the strong nativist approach to language and the mind proposed by Chomsky and his followers.

Localizationist views continue to wax and wane, but they seem to be approaching a new low point today, due (ironically) to the greater precision offered by magnetic resonance imaging and other techniques for determining the precise location of the lesions associated with aphasia syndromes. Simply put, the classical story of lesion-syndrome mapping is falling apart. For example, Dronkers (1996) has shown that lesions to Broca's area are neither necessary nor sufficient for the speech output impairments that define Broca's aphasia. In fact, the only region of the brain that seems to be inextricably tied to speech output deficits is an area called the insula, hidden in the folds between the frontal and temporal lobe. This area is crucial, but its contribution may lie at a relatively low level, mediating kinaesthetic feedback from the face and mouth.

A similar story may hold for speech perception. There is no question that comprehension can be disrupted by lesions to the temporal lobe in a mature adult. However, the nature and locus of this disruption are not at all clear. Of all the symptoms that affect speech perception, the most severe is a syndrome called "pure word deafness." Individuals with this affliction are completely unable to recognize spoken words, even though they do respond to sound and can (in many cases) correctly classify meaningful environmental sounds (e.g., matching the sound of a dog barking to the picture of a dog). This is not a deficit to lexical semantics (see below), because some individuals with this affliction can understand the same words in a written form. Because such individuals are not deaf, it is tempting to speculate that pure word deafness represents the loss of a localized brain structure that exists only for speech. However, there are two reasons to be cautious before we accept such a conclusion. First, the lesions responsible for word deafness are bilateral (i.e., wide ranges of auditory cortex must be damaged on both sides). Hence they do not follow the usual left/right asymmetry observed in language-related syndromes. Second, it is difficult on logical grounds to distinguish between a speech/nonspeech contrast (a domain-specific distinction) and a complex/simple

contrast (a domain-general distinction). As we noted earlier, speech is an exceedingly complex auditory event. There are no other meaningful environmental sounds that achieve anything close to this level of complexity (dogs barking, bells ringing, etc.). Hence it is quite possible that the lesions responsible for word deafness have their effect by creating a global, nonspecific degradation in auditory processing, one that is severe enough to preclude speech but not severe enough to block recognition of other, simpler auditory events.

This brings us to a different but related point. There is a developmental syndrome called "congenital dysphasia" or "specific language impairment", a deficit in which children are markedly delayed in language development in the absence of any other syndrome that could account for this delay (e.g., no evidence of mental retardation, deafness, frank neurological impairments like cerebral palsy, or severe socio-emotional deficits like those that occur in autism). Some theorists believe that this is a domain-specific syndrome, one that provides evidence for the independence of language from other cognitive abilities (see Grammar, below). However, other theorists have proposed that this form of language delay is the by-product of subtle deficits in auditory processing that are not specific to language, but impair language more than any other aspect of behavior. This claim is still quite controversial, but evidence is mounting in its favor (Bishop, 1997; Leonard, 1997). If this argument is correct, then we need to rethink the neat division between components that we have followed here so far, in favor of a theory in which auditory deficits lead to deficits in the perception of speech, which lead in turn to deficits in language learning.

Functional Brain-Imaging Studies of Speech. With the arrival of new tools like positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), we are able at last to observe the normal brain at work. If the phrenological approach to brain organization for language were correct, then it should be just a matter of time before we locate the areas dedicated to each and every component of language. However, the results that have been obtained to date are very discouraging for the phrenological view.

Starting with speech perception, Poeppel (1996) has reviewed six pioneering studies of phonological processing using PET. Because phonology is much closer to the physics of speech than abstract domains like semantics and grammar, we might expect the first breakthroughs in brain-language mapping to occur in this domain. However, Poeppel notes that there is virtually no overlap across these six studies in the regions that appear to be most active during the processing of speech sounds! To be sure, these studies (and many others) generally find greater activation in the left hemisphere than the right, although many studies of language activation do find evidence for some right-hemisphere involvement. In addition, the frontal and

temporal lobes are generally more active than other regions, especially around the Sylvian fissure (“perisylvian cortex”), which includes Broca’s and Wernicke’s areas. Although lesion studies and brain-imaging studies of normals both implicate perisylvian cortex, many other areas show up as well, with marked variations from one study to another. Most importantly for our purposes here, there is no evidence from these studies for a single “phonological processing center” that is activated by all phonological processing tasks.

Related evidence comes from studies of speech production (including covert speech, without literal movements of the mouth). Here too, left-hemisphere activation invariably exceeds activation on the right, and perisylvian areas are typical regions of high activation (Toga, Frackowiak, & Mazziotta, 1996). Interestingly, the left insula is the one region that emerges most often in fMRI and PET studies of speech production, a complement to Dronkers’ findings for aphasic patients with speech production deficits. The insula is an area of cortex buried deep in the folds between temporal and frontal cortex. Although its role is still not fully understood, the insula appears to be particularly important in the mediation of kinaesthetic feedback from the various articulators (i.e., moving parts of the body), and the area implicated in speech output deficits is one that is believed to play a role in the mediation of feedback from the face and mouth. Aside from this one relatively low-level candidate, no single region has emerged to be crowned as “the speech production center”.

One particularly interesting study in this regard focused on the various subregions that comprise Broca’s area and adjacent cortex (Erhard, Kato, Strick, & Ugurbil, 1996). Functional magnetic resonance imaging (fMRI) was used to compare activation within and across the Broca complex, in subjects who were asked to produce covert speech movements, simple and complex nonspeech movements of the mouth, and finger movements at varying levels of complexity. Although many of the subcomponents of Broca’s complex were active for speech, all of these components participate to a similar extent in at least one nonspeech task. In other words, there is no area in the frontal region that is active only for speech.

These findings for speech illustrate an emerging theme in functional brain imagery research, revolving around task specificity, rather than domain specificity. That is, patterns of localization or activation seem to vary depending on such factors as the amount and kind of memory required for a task, its relative level of difficulty and familiarity to the subject, its demands on attention, the presence or absence of a need to suppress a competing response, whether covert motor activity is required, and so forth. These domain-general but task-specific factors show up in study after study, with both linguistic and nonlinguistic materials (e.g., an area in frontal cortex called the anterior cingulate shows up in study after study when a task is very new, and very hard). Does this mean that domain-specific functions

“move” from one area to another? Perhaps, but it is more likely that “movement” and “location” are both the wrong metaphors. We may need to revise our thinking about brain organization for speech and other functions along entirely different lines. My hand takes very different configurations depending on the task that I set out to accomplish: to pick up a pin, pick up a heavy book, or push a heavy box against the wall. A “muscle activation” study of my hand within each task would yield a markedly different distribution of activity for each of these tasks. And yet it does not add very much to the discussion to refer to the first configuration as a “pin processor”, the second as a “book processor”, and so on. In much the same way, we may use the distributed resources of our brains in very different ways depending on the task that we are trying to accomplish. Some low-level components probably are hard-wired and task-independent (e.g., the areas of cortex that are fed directly by the auditory nerve, or that portion of the insula that handles kinaesthetic feedback from the mouth and face). Once we move above this level, however, we should perhaps expect to find highly variable and distributed patterns of activity in conjunction with linguistic tasks. We will return to this theme later, when we consider the brain bases of other language levels.

III. WORDS AND GRAMMAR

How Words and Sentences are Processed by Normal Adults

The major issue of concern in the study of word and sentence processing is similar to the issue that divides linguists. On one side, we find investigators who view lexical and grammatical processing as independent mental activities, handled by separate mental/neural mechanisms (e.g., Fodor, 1983). To be sure, these two modules have to be integrated at some point in processing, but their interaction can only take place after each module has completed its work. On the other side, we find investigators who view word recognition and grammatical analysis as two sides of a single complex process: Word recognition is “penetrated” by sentence-level information (e.g., Elman & McClelland, 1986), and sentence processing is profoundly influenced by the nature of the words contained within each sentence (MacDonald, Pearlmuter, & Seidenberg, 1994). This split in psycholinguistics between modularists and interactionists mirrors the split in theoretical linguistics between proponents of syntactic autonomy (e.g., Chomsky, 1957) and theorists who emphasize the semantic and conceptual nature of grammar (e.g., Langacker, 1987).

In the 1960’s-1970’s, when the autonomy view prevailed, efforts were made to develop real-time processing models of language comprehension and production (i.e., performance) that implemented the same modular structure proposed in various formulations of Chomsky’s generative grammar (i.e., competence). (For a review, see Fodor, Bever, & Garrett, 1974). The comprehension variants had a kind of “assembly line”

structure, with linguistic inputs passed in a serial fashion from one module to another (phonetic --> phonological --> grammatical --> semantic). Production models looked very similar, with arrows running in the opposite direction (semantic --> grammatical --> phonological --> phonetic). According to this "assembly line" approach, each of these processes is unidirectional. Hence it should not be possible for higher-level information in the sentence to influence the process by which we recognize individual words during comprehension, and it should not be possible for information about the sounds in the sentence to influence the process by which we choose individual words during production.

The assumption of unidirectionality underwent a serious challenge during the late 1970's to the early 1980's, especially in the study of comprehension. A veritable cottage industry of studies appeared showing "top-down" context effects on the early stages of word recognition, raising serious doubts about this fixed serial architecture. For example, Samuel (1981) presented subjects with auditory sentences that led up to auditory word targets like "meal" or "wheel". In that study, the initial phoneme that disambiguates between two possible words was replaced with a brief burst of noise (like a quick cough), so that the words "meal" and "wheel" were both replaced by "(NOISE)-eel". Under these conditions, subjects readily perceived the "-eel" sound as "meal" in a dinner context and "wheel" in a transportation context, often without noticing the cough at all.

In response to all these demonstrations of context effects, proponents of the modular view countered with studies demonstrating temporal constraints on the use of top-down information during the word recognition process (Onifer & Swinney, 1981), suggesting that the process really is modular and unidirectional, but only for a very brief moment in time. An influential example comes from experiments in which semantically ambiguous words like "bug" are presented within an auditory sentence context favoring only one of its two meanings, e.g.,

Insect context (bug = insect):

"Because they had found a number of roaches and spiders in the room, experts were called in to check the room for bugs...."

Espionage context (bug = hidden microphone):

"Because they were concerned about electronic surveillance, the experts were called in to check the room for bugs...."

Shortly after the ambiguous word is presented, subjects see a either real word or a nonsense word presented visually on the computer screen, and are asked to decide as quickly as possible if the target is a real word or not (i.e., a lexical decision task). The real-word targets included words that are related to the "primed" or contextually appropriate meaning of "bug" (e.g., SPY

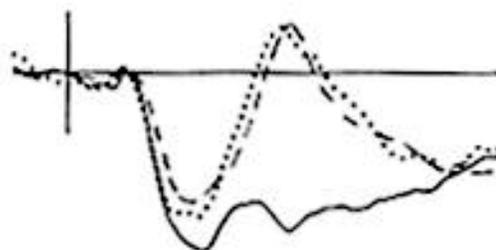
in an espionage context), words that are related to the contextually inappropriate meaning of "bug" (e.g., ANT in an espionage context), and control words that are not related to either meaning (e.g., MOP in an espionage context). Evidence of semantic activation or "priming" is obtained if subjects react faster to a word related to "bug" than they react to the unrelated control. If the lexicon is modular, and uninfluenced by higher-level context, then there should be a short period of time in which SPY and ANT are both faster than MOP. On the other hand, if the lexicon is penetrated by context in the early stages, then SPY should be faster than ANT, and ANT should be no faster than the unrelated word MOP.

The first round of results using this technique seemed to support the modular view. If the prime and target are separated by at least 750 msec, priming is observed only for the contextually appropriate meaning (i.e., selective access); however, if the prime and target are very close together in time (250 msec or less), priming is observed for both meanings of the ambiguous word (i.e., exhaustive access). These results were interpreted as support for a two-stage model of word recognition: a "bottom-up" stage that is unaffected by context, and a later "top-down" stage when contextual constraints can apply.

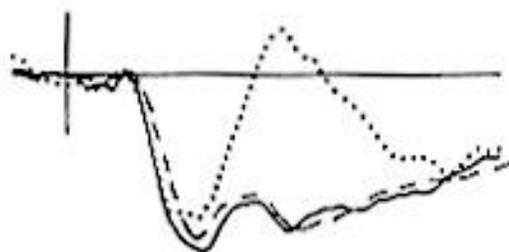
Although the exhaustive-access finding has been replicated in many different laboratories, its interpretation is still controversial. For example, some investigators have shown that exhaustive access fails to appear on the second presentation of an ambiguous word, or in very strong contexts favoring the dominant meaning of the word. An especially serious challenge comes from a study by Van Petten and Kutas (1991), who used similar materials to study the event-related scalp potentials (i.e., ERPs, or "brain waves") associated with contextually appropriate, contextually inappropriate and control words at long and short time intervals (700 vs. 200 msec between prime and target). Their results provide a very different story than the one obtained in simple reaction time studies, suggesting that there are actually three stages involved in the processing of ambiguous words, instead of just two. Figure 6 illustrates the Van Petten and Kutas results when the prime and target are separated by only 200 msec (the window in which lexical processing is supposed to be independent of context), compared with two hypothetical outcomes. Once again, we are using an example in which the ambiguous word BUG appears in an espionage context. If the selective-access view is correct, and context does penetrate the lexicon, then brain waves to the contextually relevant word (e.g., SPY) ought to show a *positive* wave (where positive is plotted downward, according to the conventions of this field); brain waves to the contextually irrelevant word (e.g., ANT) ought to look no different from an unrelated and unexpected control word (e.g., MOP), with both eliciting a *negative* wave called the N400 (plotted upward). If the modular, exhaustive-access account is correct, and context cannot penetrate the lexicon, then any word that is related lexically to the ambiguous

Possible vs. observed outcomes in the brainwaves observed for ambiguous words that do or do not fit the context. (200-msec condition - from Van Petten & Kutas, 1995).

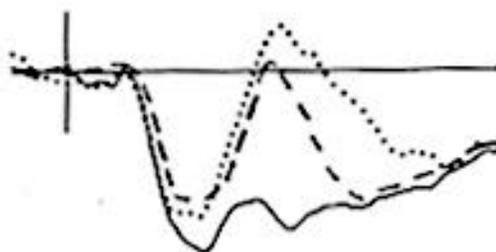
Selective access:
context penetrates
the lexicon



Exhaustive access:
lexicon is modular



Observed outcome:
selective access
precedes exhaustive
access



..... *Unrelated*
----- *Contextually irrelevant*
———— *Contextually relevant*

FIGURE 6

prime (e.g., either SPY or ANT) ought to show a positive (“primed”) wave, compared with an unexpected (“unprimed”) control word. The observed outcome was more compatible with a selective-access view, but with an interesting variation, also plotted in Figure 6. In the very first moments of word recognition, the contextually inappropriate word (e.g., ANT) behaves just like an unrelated control (e.g., MOP), moving in a negative direction (plotted upward). However, later on in the sequence (around 400 msec), the contextually irrelevant word starts to move in a positive direction, as though the subject had just noticed, after the fact, that there was some kind of additional relationship (e.g., BUG ... ANT? ... Oh yeah, ANT!!). None of this occurs with the longer 700-millisecond window, where results are fit perfectly by the context-driven selective-access model. These complex findings suggest that we may need a three-stage model to account for processing of ambiguous words, paraphrased as

SELECTIVE PRIMING -->
 EXHAUSTIVE PRIMING -->
 CONTEXTUAL SELECTION

The fact that context effects actually precede exhaustive priming proves that context effects *can* penetrate the earliest stages of lexical processing, strong evidence against the classic modular view. However, the fact that exhaustive priming does appear for a very short time, within the shortest time window, suggests that the lexicon does have "a mind of its own", i.e., a stubborn tendency to activate irrelevant material at a local level, even though relevance wins out in the long run.

To summarize, evidence in favor of context effects on word recognition has continued to mount in the last few years, with both reaction time and electrophysiological measures. However, the lexicon does behave rather stupidly now and then, activating irrelevant meanings of words as they come in, as if it had no idea what was going on outside. Does this prove that lexical processing takes place in an independent module? Perhaps not. Kawamoto (1988) has conducted simulations of lexical access in artificial neural networks in which there is no modular border between sentence- and word-level processing. He has shown that exhaustive access can and does occur under some circumstances even in a fully interactive model, depending on differences in the rise time and course of activation for different items under different timing conditions. In other words, irrelevant meanings can “shoot off” from time to time even in a fully interconnected system. It has been suggested that this kind of “local stupidity” is useful to the language-processing system, because it provides a kind of back-up activation, just in case the most probable meaning turns out to be wrong at some point further downstream. After all, people do occasionally say very strange things, and we have to be prepared to hear them, even within a “top-down,” contextually guided system.

This evidence for an early interaction between sentence-level and word-level information is only

indirectly related to the relationship between lexical processing and grammar per se. That is, because a sentence contains both meaning and grammar, a sentential effect on word recognition could be caused by the semantic content of the sentence (both propositional and lexical semantics), leaving the Great Border between grammar and the lexicon intact. Are the processes of word recognition and/or word retrieval directly affected by grammatical context alone?

A number of early studies looking at grammatical priming in English have obtained weak effects or no effects at all on measures of lexical access. In a summary of the literature on priming in spoken-word recognition, Tanenhaus and Lucas (1987) conclude that “On the basis of the evidence reviewed ...it seems likely that syntactic context does not influence prelexical processing” (p. 223). However, more recent studies in languages with rich morphological marking have obtained robust evidence for grammatical priming (Bates & Goodman, in press). This includes effects of gender priming on lexical decision and gating in French, on word repetition and gender classification in Italian, and on picture naming in Spanish and German. Studies of lexical decision in Serbo-Croatian provide evidence for both gender and case priming, with real word and nonword primes that carry morphological markings that are either congruent or incongruent with the target word. These and other studies show that grammatical context can have a significant effect on lexical access within the very short temporal windows that are usually associated with early and automatic priming effects that interest proponents of modularity. In other words, grammatical and lexical processes interact very early, in intricate patterns of the sort that we would expect if they are taking place within a single, unified system governed by common laws.

This conclusion marks the beginning rather than the end of interesting research in word and sentence processing, because it opens the way for detailed cross-linguistic studies of the processes by which words and grammar interact during real-time language processing. The modular account can be viewed as an accidental by-product of the fact that language processing research has been dominated by English-speaking researchers for the last 30 years. English is unusual among the world’s languages in its paucity of inflectional morphology, and in the degree to which word order is rigidly preserved. In a language of this kind, it does seem feasible to entertain a model in which words are selected independently of sentence frames, and then put together by the grammar like beads on a string, with just a few minor adjustments in the surface form of the words to assure morphological agreement (e.g., “The dogs walk” vs. “The dog walks”). In richly inflected languages like Russian, Italian, Hebrew or Greenlandic Eskimo, it is difficult to see how such a modular account could possibly work. Grammatical facts that occur early in the sentence place heavy constraints on the words that must be recognized or produced later on, and the words that we recognize or produce at the beginning of an

utterance influence detailed aspects of word selection and grammatical agreement across the rest of the sentence. A model in which words and grammar interact intimately at every stage in processing would be more parsimonious for a language of this kind. As the modularity/interactionism debate begins to ebb in the field of psycholinguistics, rich and detailed comparative studies of language processing are starting to appear across dramatically different language families, marking the beginning of an exciting new era in this field.

How Words and Sentences Develop in Children

The modularity/interactionism debate has also been a dominant theme in the study of lexical and grammatical development, interacting with the overarching debate among empiricists, nativists and constructivists.

From one point of view, the course of early language development seems to provide a *prima facie* case for linguistic modularity, with sounds, words and grammar each coming in on separate developmental schedules (Table 1). Children begin their linguistic careers with babble, starting with vowels (somewhere around 3-4 months, on average) and ending with combinations of vowels and consonants of increasing complexity (usually between 6-8 months). Understanding of words typically begins between 8-10 months, but production of meaningful speech emerges some time around 12 months, on average. After this, most children spend many weeks or months producing single-word utterances. At first their rate of vocabulary growth is very slow, but one typically sees a "burst" or acceleration in the rate of vocabulary growth somewhere between 16-20 months. First word combinations usually appear between 18-20 months, although they tend to be rather spare and telegraphic (at least in English — see Table 2 for examples). Somewhere between 24-30 months, most children show a kind of "second burst", a flowering of morphosyntax that Roger Brown has characterized as "the ivy coming in between the bricks." Between 3-4 years of age, most normal children have mastered the basic morphological and syntactic structures of their language, using them correctly and productively in novel contexts. From this point on, lexical and grammatical development consist primarily in the tuning and amplification of the language system: adding more words, becoming more fluent and efficient in the process by which words and grammatical constructions are accessed in real time, and learning how to use the grammar to create larger discourse units (e.g., writing essays, telling stories, participating in a long and complex conversation).

This picture of language development in English has been documented extensively (for reviews see Aslin, Jusczyk & Pisoni, in press; Fletcher & MacWhinney, 1995). Of course the textbook story is not exactly the same in every language (Slobin, 1985-1997), and perfectly healthy children can vary markedly in rate and style of development through these milestones (Bates,

Bretherton, & Snyder, 1988). At a global level, however, the passage from sounds to words to grammar appears to be a universal of child language development.

A quick look at the relative timing and shape of growth within word comprehension, word production and grammar can be seen in Figures 7, 8 and 9 (from Fenson et al., 1994). The median (50th percentile) in each of these figures confirms that textbook summary of average onset times that we have just recited: Comprehension gets off the ground (on average) between 8-10 months, production generally starts off between 12-13 months (with a sharp acceleration between 16-20 months), and grammar shows its peak growth between 24-30 months. At the same time, however, these figures show that there is massive variation around the group average, even among perfectly normal, healthy middle-class children. Similar results have now been obtained for more than a dozen languages, including American Sign Language (a language that develops with the eyes and hands, instead of the ear and mouth). In every language that has been studied to date, investigators report the same average onset times, the same patterns of growth, and the same range of individual variation illustrated in Figures 7-9.

But what about the relationship between these modalities? Are these separate systems, or different windows on a unified developmental process? A more direct comparison of the onset and growth of words and grammar can be found in Figure 10 (from Bates & Goodman, in press). In this figure, we have expressed development for the average (median) child in terms of the percent of available items that have been mastered at each available time point from 8-30 months. Assuming for a moment that we have a right to compare the proportional growth of apples and oranges, it shows that word comprehension, word production and grammar each follow a similar nonlinear pattern of growth across this age range. However, the respective "zones of acceleration" for each domain are separated by many weeks or months.

Is this a discontinuous passage, as modular/nativist theories would predict? Of course no one has ever proposed that grammar can begin in the absence of words! Any grammatical device is going to have to have a certain amount of lexical material to work on. The real question is: Just how tight are the correlations between lexical and grammatical development in the second and third year of life? Are these components dissociable, and if so, to what extent? How much lexical material is needed to build a grammatical system? Can grammar get off the ground and go its separate way once a minimum number of words is reached (e.g., 50-100 words, the modal vocabulary size when first word combinations appear)? Or will we observe a constant and lawful interchange between lexical and grammatical development, of the sort that one would expect if words and grammar are two sides of the same system?

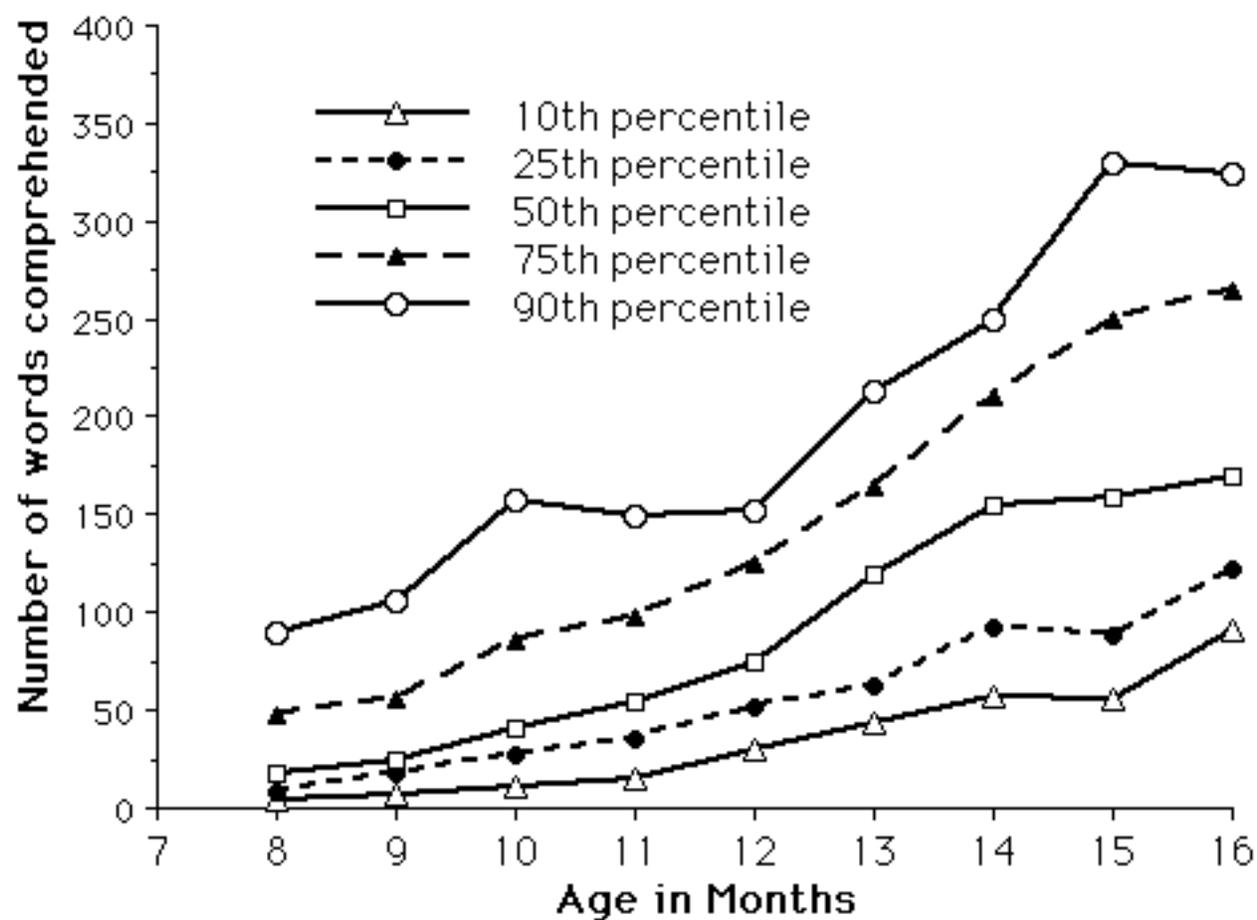
Our reading of the evidence suggests that the latter view is correct. In fact, the function that governs the

TABLE 1: MAJOR MILESTONES IN LANGUAGE DEVELOPMENT

0-3 months	INITIAL STATE OF THE SYSTEM <ul style="list-style-type: none">— prefers to listen to sounds in native language— can hear all the phonetic contrasts used in the world's languages— produces only “vegetative” sounds
3-6 months	VOWELS IN PERCEPTION AND PRODUCTION <ul style="list-style-type: none">— cooing, imitation of vowel sounds only— perception of vowels organized along language-specific lines
6-8 months	BABBLING IN CONSONANT-VOWEL SEGMENTS
8-10 months	WORD COMPREHENSION <ul style="list-style-type: none">— starts to lose sensitivity to consonants outside native language
12-13 months	WORD PRODUCTION (NAMING)
16-20 months	WORD COMBINATIONS <ul style="list-style-type: none">— vocabulary acceleration— appearance of relational words (e.g., verbs and adjectives)
24-36 months	GRAMMATICIZATION <ul style="list-style-type: none">— grammatical function words— inflectional morphology— increased complexity of sentence structure
3 years → adulthood	LATE DEVELOPMENTS <ul style="list-style-type: none">— continued vocabulary growth— increased accessibility of rare and/or complex forms— reorganization of sentence-level grammar for discourse purposes

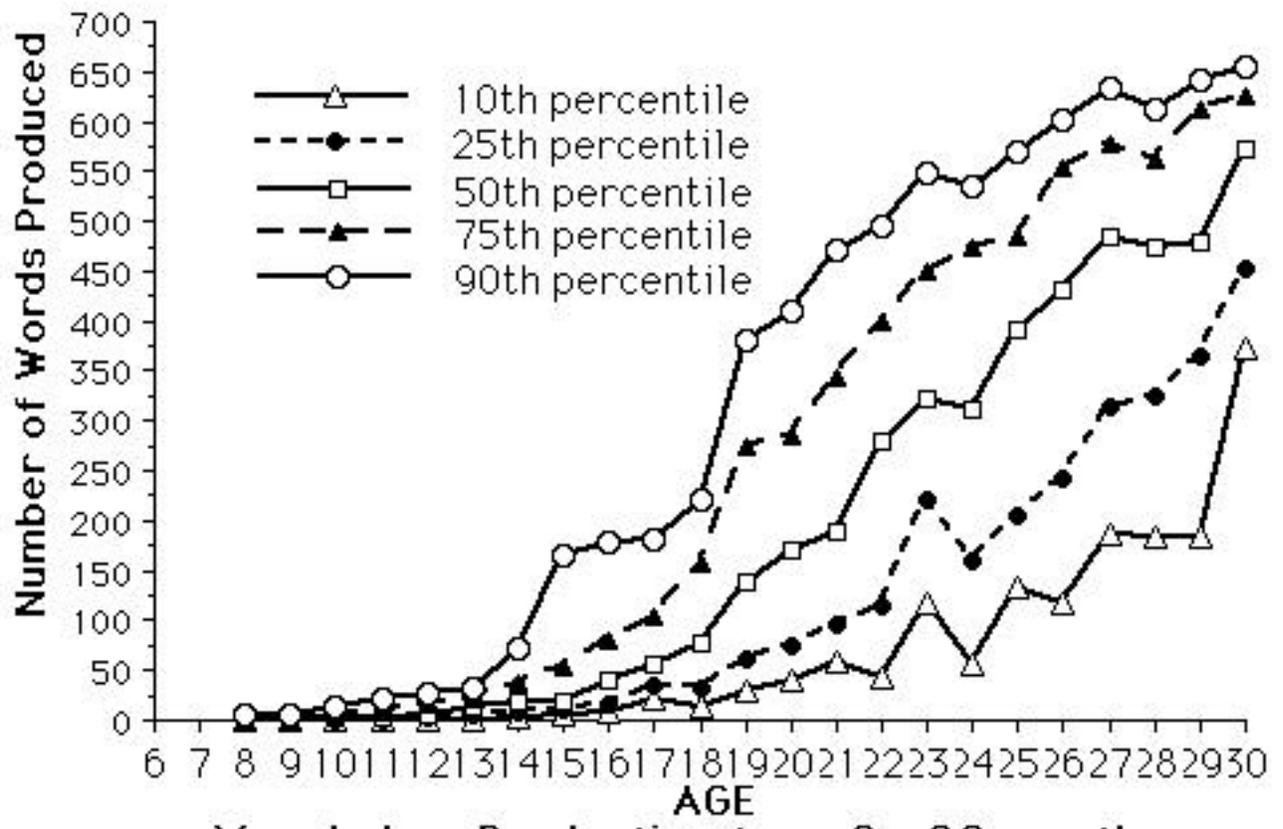
**TABLE 2:
SEMANTIC RELATIONS UNDERLYING CHILDREN'S FIRST WORD COMBINATIONS
ACROSS MANY DIFFERENT LANGUAGES
(adapted from Braine, 1976)**

<u>Semantic Function</u>	<u>English examples</u>	
Attention to X	"See doggie!" "Dat airplane"	
Properties of X	"Mommy pretty"	"Big doggie"
Possession	"Mommy sock"	"My truck"
Plurality or iteration	"Two shoe"	"Round and round"
Recurrence (including requests)	"More truck" "Other cookie"	
Disappearance	"Allgone airplane"	"Daddy bye-bye"
Negation or Refusal	"No bath"	"No bye-bye"
Actor-Action	"Baby cry"	"Mommy do it"
Location	"Baby car"	"Man outside"
Request	"Wanna play it"	"Have dat"



**Number of Words Comprehended
between 8 - 16 months (from >600 cases)**

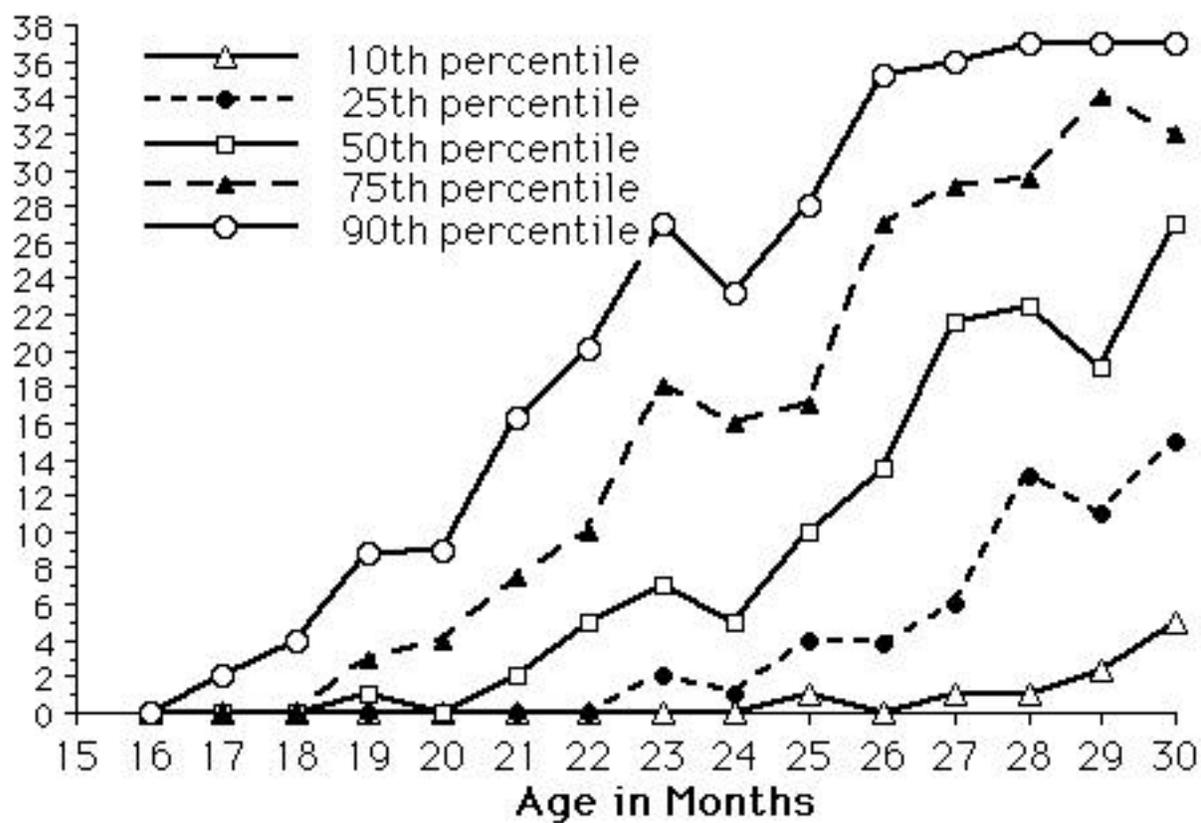
Figure 7



Vocabulary Production from 8 - 30 months
(for > 1800 Cases)

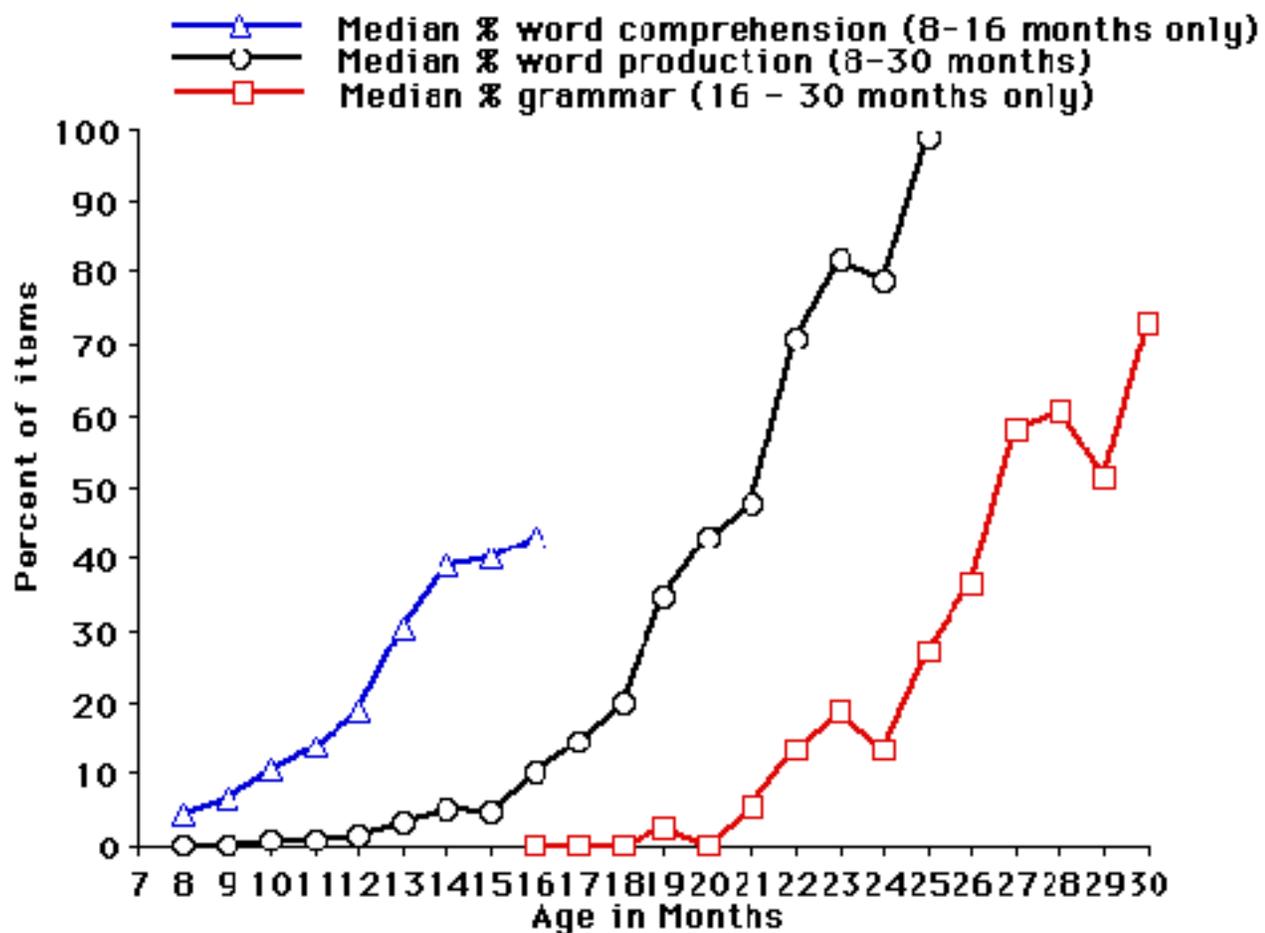
Figure 8

Grammatical Complexity Score (0 - 37)



Grammatical Complexity as a Function of Age (> 1000 cases)

Figure 9



Comparison of Growth in Comprehension, Production and Grammar in a Longitudinal Sample (% items in each category)

Figure 10

relation between lexical and grammatical growth in this age range is so powerful and so consistent that it seems to reflect some kind of developmental law. The successive “bursts” that characterize vocabulary growth and the emergence of grammar can be viewed as different phases of an immense nonlinear wave that starts in the single-word stage and crashes on the shores of grammar a year or so later. An illustration of this powerful relationship is offered in a comparison of Figure 11, which plots the growth of grammar as a function of vocabulary size. It should be clear from this figure that grammatical growth is tightly related to lexical growth, a lawful pattern that is far more regular (and far stronger) than the relationship between grammatical development and chronological age. Of course this kind of correlational finding does not force us to conclude that grammar and vocabulary growth are mediated by the same developmental mechanism. Correlation is not cause. At the very least, however, this powerful correlation suggests that the two have something important in common.

Although there are strong similarities across languages in the lawful growth and interrelation of vocabulary and grammar, there are massive differences between languages in the specific structures that must be acquired. Chinese children are presented with a language that has no grammatical inflections of any kind, compared with Eskimo children who have to learn a language in which an entire sentence may consist of a single word with more than a dozen inflections. Lest we think that life for the Chinese child is easy, that child has to master a system in which the same syllable can take on many different meanings depending on its tone (i.e., its pitch contour). There are four of these tones in Mandarin Chinese, and seven in the Taiwanese dialect, presenting Chinese children with a word-learning challenge quite unlike the ones that face children acquiring Indo-European languages like English or Italian.

Nor should we underestimate the differences that can be observed for children learning different Indo-European language types. Consider the English sentence

Wolves eat sheep

which contains three words and four distinct morphemes (in standard morpheme-counting systems, “wolf” + “-s” constitute two morphemes, but the third person plural verb “eat” and the plural noun “sheep” only count as one morpheme each). The Italian translation of this sentence would be

I lupi mangiano le pecore

The Italian version of this sentence contains five words (articles are obligatory in Italian in this context; they are not obligatory for English). Depending on the measure of morphemic complexity that we choose to use, it also contains somewhere between ten and fourteen morphemes (ten if we count each explicit plural marking on each article, noun and verb, but exclude gender decisions from the count; fourteen if each

gender decision also counts as a separate morphological contrast, on each article and noun). What this means is that, in essence, an Italian child has roughly three times as much grammar to learn as her English age-mates! There are basically two ways that this quantitative difference might influence the learning process:

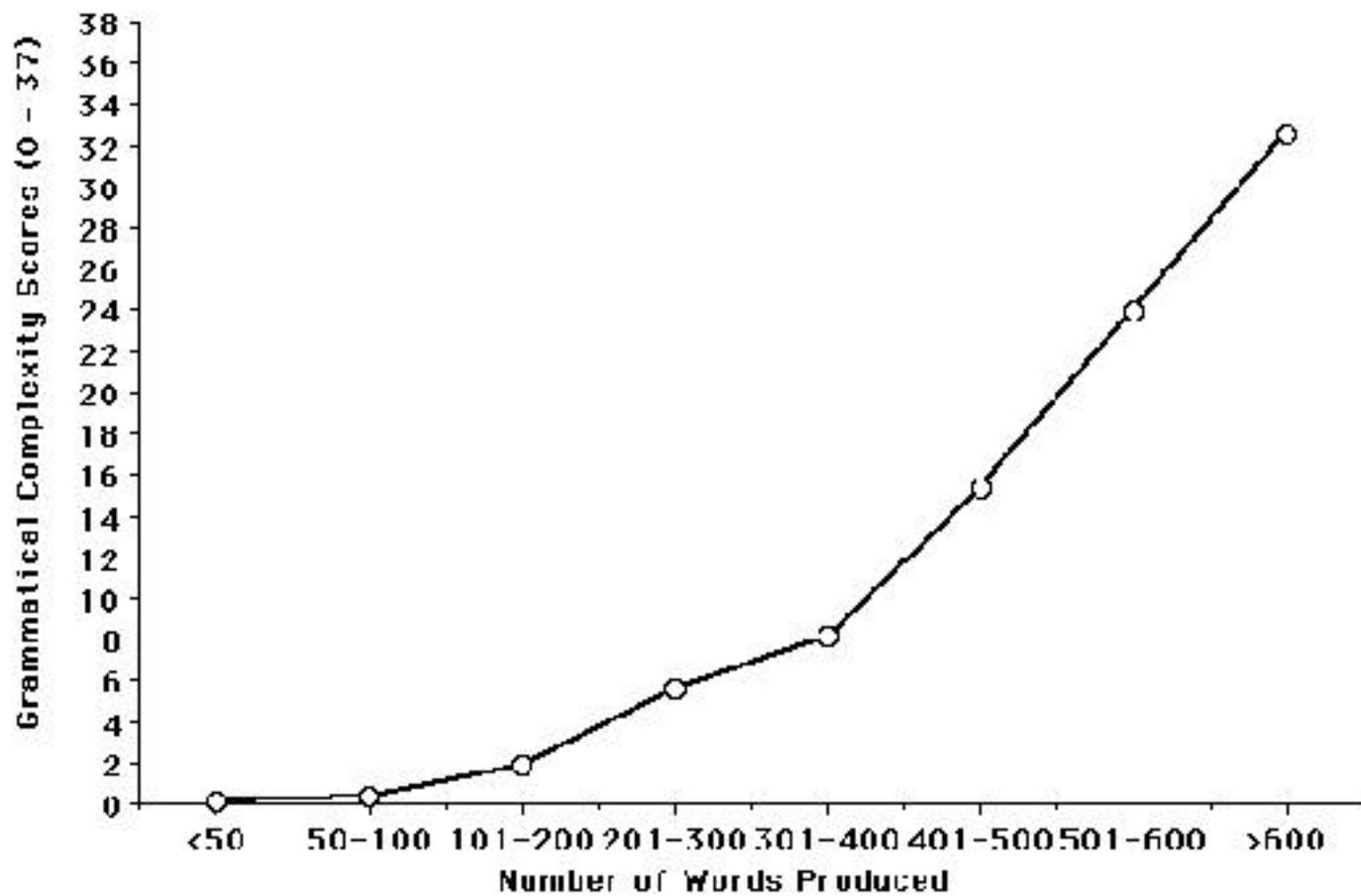
(1) Italian children might acquire morphemes at the same *absolute rate*. If this is true, then it should take approximately three times longer to learn Italian than it does to learn English.

(2) Italian and English children might acquire their respective languages at the same *proportional rate*. If this is true, then Italian and English children should “know” the same proportion of their target grammar at each point in development (e.g., 10% at 20 months, 30% at 24 months, and so forth), and Italian children should display approximately three times more morphology than their English counterparts at every point.

Comparative studies of lexical and grammatical development in English and Italian suggest that the difference between languages is proportional rather than absolute during the phase in which most grammatical contrasts are acquired. As a result of this difference, Italian two-year-olds often sound terribly precocious to native speakers of English, and English two-year-olds may sound mildly retarded to the Italian ear!

My point is that every language presents the child with a different set of problems, and what is “hard” in one language may be “easy” in another. Table 2 presents a summary of the kinds of meanings that children tend to produce in their first word combinations in many different language communities (from Braine, 1976). These are the concerns (e.g., possession, refusal, agent-action, object-location) that preoccupy 2-year-olds in every culture. By contrast, Table 3 shows how very different the first sentences of 2-year-old children can look just a few weeks or months later, as they struggle to master the markedly different structural options available in their language (adapted from Slobin, 1985, 1985-1997). The content of these utterances is universal, but the forms that they must master vary markedly.

How do children handle all these options? Chomsky's answer to the problem of cross-linguistic variation is to propose that all the options in Universal Grammar are innate, so that the child's task is simplified to one of listening carefully for the right “triggers,” setting the appropriate “parameters”. Learning has little or nothing to do with this process. Other investigators have proposed instead that language development really is a form of learning, although it is a much more powerful form of learning than Skinner foresaw in his work on schedules of reinforcement in rats. Recent neural network simulations of the language-learning process provide, at the very least, a kind of “existence proof”, demonstrating that aspects of grammar can be learned by a system of this kind and thus (perhaps) by the human child.



GRAMMATICAL COMPLEXITY AS A FUNCTION OF VOCABULARY SIZE

Figure 11

TABLE 3:
EXAMPLES OF SPEECH BY TWO-YEAR-OLDS IN DIFFERENT LANGUAGES
 (underlining = content words)

English (30 months):

<i>I</i>	<i>wanna</i>	<i>help</i>	<i>wash</i>	<i>car</i>
1st pers. singular	modal indicative	infinitive	infinitive	

Translation: I wanna help wash car.

Italian (24 months):

<i>Lavo</i>	<i>mani</i>	<i>sporche</i>	<i>apri</i>	<i>acqua</i>
<u>Wash</u>	<u>hands</u>	<u>dirty</u>	<u>open</u>	<u>water</u>
1st pers. singular indicative	3rd pers. feminine plural	feminine plural	2nd pers. singular imperative	3rd pers. singular

Translation: I wash hands, dirty, turn on water.

Western Greenlandic (26 months):

<i>anner- punga.....anni- ler-</i>	<i>hurt- about-to</i>	<i>punga</i>
<u>hurt-</u> 1st singular indicative	<u>hurt-</u> about-to	1st singular indicative

Translation: I've hurt myself ... I'm about to hurt myself ...

Mandarin (28 months):

<i>Bu</i>	<i>yao</i>	<i>ba</i>	<i>ta cai-</i>	<i>diao zhege ou</i>
not	<u>want</u>	object- marker	it <u>tear-</u>	down this warning- marker

Translation: Don't tear apart it!

TABLE 3: (continued)
EXAMPLES OF SPEECH BY TWO-YEAR-OLDS
IN DIFFERENT LANGUAGES

Sesotho (32 months):

<i>o-</i>	<i>tla-</i>	<i>hlaj-</i>	<i>uw-</i>	<i>a</i>	<i>ke</i>	<i>tshehlo</i>
class 2	future	<u>stab-</u>	passive mood	by	<u>thorn</u>	class 9
singular			marker			
subject-						
marker						

Translation: You'll get stabbed by a thorn.

Japanese (25 months):

<u>Okashitabe-</u>	<i>ru</i>	<i>tte</i>	<i>say</i>	<u>yut-</u>	<i>ta</i>
<u>Sweets eat</u>	non-	quote-	marker	past	
	past				

Translation: She said that she'll eat sweets.

For example, it has long been known that children learning English tend to produce correct irregular past tense forms (e.g., "went" and "came") for many weeks or months before the appearance of regular marking (e.g., "walked" and "kissed"). Once the regular markings appear, peculiar errors start to appear as well, e.g., terms like "goed" and "comed" that are not available anywhere in the child's input. It has been argued that this kind of U-shaped learning is beyond the capacity of a single learning system, and must be taken as evidence for two separate mechanisms: a rote memorization mechanism (used to acquire words, and to acquire irregular forms) and an independent rule-based mechanism (used to acquire regular morphemes). However, there are now several demonstrations of the same U-shaped developmental patterns in neural networks that use a single mechanism to acquire words and all their inflected forms (Elman et al., 1996). These and other demonstrations of grammatical learning in neural networks suggest that rote learning and creative generalizations can both be accomplished by a single system, if it has the requisite properties. It seems that claims about the "unlearnability" of grammar have to be reconsidered.

Brain Bases of Words and Sentences

Lesion studies. When the basic aphasic syndromes were first outlined by Broca, Wernicke and their colleagues, differences among forms of linguistic breakdown were explained along sensorimotor lines, rooted in rudimentary principles of neuroanatomy. For example, the symptoms associated with damage to a region called Broca's area were referred to collectively as *motor aphasia*: slow and effortful speech, with a reduction in grammatical complexity, despite the apparent preservation of speech comprehension at a clinical level. This definition made sense when we consider the fact that Broca's area lies near the motor strip. Conversely, the symptoms associated with damage to Wernicke's area were defined collectively as a *sensory aphasia*: fluent but empty speech, marked by moderate to severe word-finding problems, in patients with serious problems in speech comprehension. This characterization also made good neuroanatomical sense, because Wernicke's area lies at the interface between auditory cortex and the various association areas that were presumed to mediate or contain word meaning. Isolated problems with repetition were further ascribed to fibers that link Broca's and Wernicke's area; other syndromes involving the selective sparing or impairment of reading or writing were proposed, with speculations about the fibers that connect visual cortex with the classical language areas (for an influential and highly critical historical review, see Head, 1926).

In the period between 1960 and 1980, a revision of this sensorimotor account was proposed (summarized in Kean, 1985). Psychologists and linguists who were strongly influenced by generative grammar sought an account of language breakdown in aphasia that followed the componential analysis of the human language

faculty proposed by Chomsky and his colleagues. This effort was fueled by the discovery that Broca's aphasics do indeed suffer from comprehension deficits: Specifically, these patients display problems in the interpretation of sentences when they are forced to rely entirely on grammatical rather than semantic or pragmatic cues (e.g., they successfully interpret a sentence like "The apple was eaten by the girl", where semantic information is available in the knowledge that girls, but not apples, are capable of eating, but fail on a sentence like "The boy was pushed by the girl", where either noun can perform the action). Because those aspects of grammar that appear to be impaired in Broca's aphasia are precisely the same aspects that are impaired in the patients' expressive speech, the idea was put forth that Broca's aphasia may represent a selective impairment of grammar (in all modalities), in patients who still have spared comprehension and production of lexical and propositional semantics. From this point of view, it also seemed possible to reinterpret the problems associated with Wernicke's aphasia as a selective impairment of semantics (resulting in comprehension breakdown and in word-finding deficits in expressive speech), accompanied by a selective sparing of grammar (evidenced by the patients' fluent but empty speech).

If grammar and lexical semantics can be doubly dissociated by forms of focal brain injury, then it seems fair to conclude that these two components of language are mediated by separate neural systems. It was never entirely obvious how or why the brain ought to be organized in just this way (e.g., why Broca's area, the supposed seat of grammar, ought to be located near the motor strip), but the lack of a compelling link between neurology and neurolinguistics was more than compensated for by the apparent isomorphism between aphasic syndromes and the components predicted by linguistic theory. It looked for a while as if Nature had provided a cunning fit between the components described by linguists and the spatial representation of language in the brain. Indeed, this linguistic approach to aphasia was so successful in its initial stages that it captured the imagination of many neuroscientists, and it worked its way into basic textbook accounts of language breakdown in aphasia.

Although this linguistic partitioning of the brain is very appealing, evidence against it has accumulated in the last 15 years, leaving aphasiologists in search of a third alternative to both the original modality-based account (i.e., motor vs. sensory aphasia) and to the linguistic account (i.e., grammatical vs. lexical deficits). Here is a brief summary of arguments against the neural separation of words and sentences (for more extensive reviews, see Bates & Goodman, in press).

(1) Deficits in word finding (called "anomia") are observed in all forms of aphasia, including Broca's aphasia. This means that there can never be a full-fledged double dissociation between grammar and the lexicon, weakening claims that

the two domains are mediated by separate brain systems.

(2) Deficits in expressive grammar are not unique to agrammatic Broca's aphasia, or to any other clinical group. English-speaking Wernicke's aphasics produce relatively few grammatical errors, compared with English-speaking Broca's aphasics. However, this fact turns out to be an artifact of English! Nonfluent Broca's aphasics tend to err by omission (i.e., leaving out grammatical function words and dropping inflections), while Wernicke's err by substitution (producing the wrong inflection). Because English has so little grammatical morphology, it provides few opportunities for errors of substitution, but it does provide opportunities for function word omission. As a result, Broca's seem to have more severe problems in grammar. However, the grammatical problems of fluent aphasia are easy to detect, and very striking, in richly inflected languages like Italian, German or Hungarian. This is not a new discovery; it was pointed out long ago by Arnold Pick, the first investigator to use the term "agrammatism" (Pick, 1913/1973).

(3) Deficits in receptive grammar are even more pervasive, showing up in Broca's aphasia, Wernicke's aphasia, and in many patient groups who show no signs of grammatical impairment in their speech output. In fact, it is possible to demonstrate profiles of receptive impairment very similar to those observed in aphasia in normal college students who are forced to process sentences under various kinds of stress (e.g., perceptual degradation, time-compressed speech, or cognitive overload). Under such conditions, listeners find it especially difficult to process inflections and grammatical function words, and they also tend to make errors on complex sentence structures like the passive (e.g., "The girl was pushed by the boy") or the object relative (e.g., "It was the girl who the boy pushed"). These aspects of grammar turn out to be the weakest links in the chain of language processing, and for that reason, they are the first to suffer when anything goes wrong.

(4) One might argue that Broca's aphasia is the only "true" form of agrammatism, because these patients show such clear deficits in both expressive and receptive grammar. However, numerous studies have shown that these patients retain knowledge of their grammar, even though they cannot use it efficiently for comprehension or production. For example, Broca's aphasics perform well above chance when they are asked to detect subtle errors of grammar in someone else's speech, and they also show strong language-specific biases in their own comprehension and production. To offer just one example, the article before the noun is marked for case in German, carrying crucial information about who did what to whom. Perhaps for this reason, German Broca's struggle to produce

the article 90% of the time (and they usually get it right), compared with only 30% in English Broca's. This kind of detailed difference can only be explained if we assume that Broca's aphasics still "know" their grammar.

Taken together, these lines of evidence have convinced us that grammar is *not* selectively lost in adult aphasia, leaving vocabulary intact. Instead, grammar and vocabulary tend to break down together, although they can break down in a number of interesting ways.

Neural Imaging Studies. To date, there are very few neural imaging studies of normal adults comparing lexical and grammatical processing, but the few that have been conducted also provide little support for a modular view. Many different parts of the brain are active when language is processed, including areas in the right hemisphere (although these tend to show lower levels of activation in most people). New "language zones" are appearing at a surprising rate, including areas that do not result in aphasia if they are lesioned, e.g., the cerebellum, parts of frontal cortex far from Broca's area, and basal temporal regions on the underside of the cortex. Furthermore, the number and location of the regions implicated in word and sentence processing differ from one individual to another, and change as a function of the task that subjects are asked to perform.

Although most studies indicate that word and sentence processing elicit comparable patterns of activation (with greater activation over left frontal and temporal cortex), several studies using ERP or fMRI have shown subtle differences in the patterns elicited by semantic violations (e.g., "I take my coffee with milk and dog") and grammatical errors (e.g., "I take my coffee with and milk sugar."). Subtle differences have also emerged between nouns and verbs, content words and function words, and between regular inflections (e.g., "walked") and irregular inflections (e.g., "gave"). Such differences constitute the only evidence to date in favor of some kind of separation in the neural mechanisms responsible for grammar vs. semantics, but they can also be explained in terms of mechanisms that are not specific to language at all. For example, content words like "milk," function words like "and," and long-distance patterns of subject-verb agreement like "The girls.....are talking" differ substantially in their length, phonetic salience, frequency, degree of semantic imagery, and the demands that they make on attention and working memory—dimensions that also affect patterns of activation in nonlinguistic tasks.

To summarize, lesion studies and neural imaging studies of normal speakers both lead to the conclusion that many different parts of the brain participate in word and sentence processing, in patterns that shift in dynamic ways as a function of task complexity and demands on information processing. There is no evidence for a unitary grammar module or a localized neural dictionary. Instead, word and sentence processing appear to be widely distributed and highly variable over tasks and individuals. However, some areas do seem to

be more important than others, especially those areas in the frontal and temporal regions of the left hemisphere that are implicated most often in patients with aphasia.

IV. PRAGMATICS AND DISCOURSE

We defined pragmatics earlier as the study of language in context, as a form of communication used to accomplish certain social ends. Because of its heterogeneity and uncertain borders, pragmatics is difficult to study and resistant to the neat division into processing, development and brain bases that we have used so far to review sounds, words and grammar.

Processing. Within the modular camp, a number of different approaches have emerged. Some investigators have tried to treat pragmatics as a single linguistic module, fed by the output of lower-level phonetic, lexical and grammatical systems, responsible for subtle inferences about the meaning of outputs in a given social context. Others view pragmatics as a collection of separate modules, including special systems for the recognition of emotion (in face and voice), processing of metaphor and irony, discourse coherence, and social reasoning (including a “theory of mind” module that draws conclusions about how other people think and feel). Still others relegate pragmatics to a place outside of the language module altogether, handled by a General Message Processor or Executive Function that also deals with nonlinguistic facts. Each of these approaches has completely different consequences for predictions about processing, development and/or neural mediation. Within the interactive camp, pragmatics is not viewed as a single domain at all. Instead, it can be viewed as the *cause* of linguistic structure, the set of communicative pressures under which all the other linguistic levels have evolved.

Perhaps because the boundaries of pragmatics are so hard to define, much of the existing work on discourse processing is descriptive, concentrating on the way that stories are told and coherence is established and maintained in comprehension and production, without invoking grand theories of the architecture of mind (Gernsbacher, 1994). However, a few studies have addressed the issue of modularity in discourse processing, with special reference to metaphor. Consider a familiar metaphor like “kicked the bucket.” In American English, this is a crude metaphor for death, and it is so familiar that we rarely think of it in literal terms (i.e., no bucket comes to mind). However, it has been suggested that we actually do compute the literal meaning in the first stages of processing, because that meaning is the obligatory product of “bottom-up” lexical and grammatical modules that have no access to the special knowledge base that handles metaphors. By analogy to the contextually irrelevant meanings of ambiguous words like “bug,” these literal interpretations rise up but are quickly eliminated in favor of the more familiar and more appropriate metaphoric interpretation. Although there is some evidence for this kind of “local stupidity,” other studies have suggested instead that the metaphoric interpretation

actually appears earlier than the literal one, evidence against an assembly line view in which each module applies at a separate stage.

Development. There are no clear milestones that define the acquisition of pragmatics. The social uses of language to share information or request help appear in the first year of life, in gesture (e.g., in giving, showing and pointing) and sound (e.g., cries, grunts, sounds of interest or surprise), well before words and sentences emerge to execute the same functions. Social knowledge is at work through the period in which words are acquired. For example, when an adult points at a novel animal and says “giraffe”, children seem to know that this word refers to the whole object, and not to some interesting feature (e.g., its neck) or to the general class to which that object belongs (e.g., animals). Predictably, some investigators believe that these constraints on meaning are innate, while others insist that they emerge through social interaction and learning across the first months of life. Social knowledge is also at work in the acquisition of grammar, defining (for example) the shifting set of possible referents for pronouns like “I” and “you,” and the morphological processes by which verbs are made to agree with the speaker, the listener or someone else in the room. When we decide to say “John is here” vs. “He is here,” we have to make decisions about the listener’s knowledge of the situation (does she know to whom “he” refers?), a decision that requires the ability to distinguish our own perspective from someone else’s point of view. In a language like Italian, the child has to figure out why some people are addressed with the formal pronoun “Lei” while others are addressed with “tu,” a complex social problem that often eludes sophisticated adults trying to acquire Italian as a second language. The point is that the acquisition of pragmatics is a continuous process, representing the interface between social and linguistic knowledge at every stage of development (Bates, 1976).

Brain Bases. Given the diffuse and heterogeneous nature of pragmatics, and the failure of the phrenological approach even at simpler levels of sound and meaning, we should not expect to find evidence for a single “pragmatic processor” anywhere in the human brain. However, there is some evidence to suggest that the right hemisphere plays a special role in some aspects of pragmatics (Joanette & Brownell, 1990), including the perception and expression of emotional content in language, the ability to understand jokes, irony and metaphor, and the ability to produce and comprehend coherent discourse. These are the domains that prove most difficult for adults with right-hemisphere injury, and there is some evidence (however slim) that the right hemisphere is especially active in normal adults on language tasks with emotional content, and/or on the processing of lengthy discourse passages.

This brings us back to a familiar theme: Does the contribution of the right hemisphere constitute evidence for a domain-specific adaptation, or is it the result of

much more general differences between the left and the right hemisphere? For example, the right hemisphere also plays a greater role in the mediation of emotion in nonverbal contexts, and it is implicated in the distribution of attention and the integration of information in nonlinguistic domains. Many of the functions that we group together under the term “pragmatics” have just these attributes: Metaphor and humor involve emotional content, and discourse coherence above the level of the sentence requires sustained attention and information integration. Hence specific patterns of localization for pragmatic functions could be the by-product of more general information-processing differences between the two hemispheres.

One approach to the debate about innateness and domain specificity comes from the study of language development in children with congenital brain injuries, involving sites that lead to specific forms of aphasia when they occur in adults. If it is the case that the human brain contains well-specified, localized processors for separate language functions, then we would expect children with early focal brain injury to display developmental variants of the various aphasic syndromes. This is not the case. In fact, children with early unilateral brain lesions typically go on to acquire language abilities that are well within the normal range (although they do tend to perform lower on a host of tasks than children who are neurologically intact).

An illustration of this point can be seen in Figure 12, which ties together several of the points that we have made throughout this chapter (Kempler, van Lancker, Marchman, & Bates, 1996). Kempler et al. compared children and adults with left- vs. right-hemisphere damage on a measure called the Familiar Phrases Task, in which subjects are asked to point to the picture that matches either a familiar phrase (e.g., “She took a turn for the worse”) or a novel phrase matched for lexical and grammatical complexity (e.g., “She put the book on the table”). Performance on these two kinds of language is expressed in z-scores, which indicate how far individual children and adults deviate from performance by normal age-matched controls. Adult patients show a strong double dissociation on this task, providing evidence for our earlier conclusion about right-hemisphere specialization for metaphors and clichés: Left-hemisphere patients are more impaired across the board (i.e., they are aphasic), but they do especially badly on the novel phrases; right-hemisphere patients are close to normal on novel phrases, but they perform even worse than aphasics on familiar phrases. A strikingly different pattern occurs for brain-injured children. Although these children all sustained their injuries before six months of age, they were 6-12 years old at time of testing. Figure 12 shows that the children are normal for their age on the familiar phrases; they do lag behind their age-mates on novel expressions, but there is no evidence for a left-right difference on any aspect of the task. In fact, findings like these are typical for the focal lesion population. The adult brain is highly differentiated, and lesions can result in

irreversible injuries. The infant brain is far more plastic, and it appears to be capable of significant reorganization when analogous injuries occur. Above all, there is no evidence in this population for an innate, well-localized language faculty.

CONCLUSION

To conclude, we are the only species on the planet capable of a full-blown, fully grammaticized language. This is a significant accomplishment, but it appears to be one that emerges over time, from simpler beginnings. The construction of language is accomplished with a wide range of tools, and it is possible that none of the cognitive, perceptual and social mechanisms that we use in the process have evolved for language alone. Language is a new machine that Nature built out of old parts.

How could this possibly work? If language is not a “mental organ”, based on innate and domain-specific machinery, then how has it come about that we are the only language-learning species? There must be adaptations of some kind that lead us to this extraordinary outcome.

To help us think about the kind of adaptation that may be responsible, consider the giraffe’s neck. Giraffes have the same 24 neckbones that you and I have, but they are elongated to solve the peculiar problems that giraffes are specialized for (i.e., eating leaves high up in the tree). As a result of this particular adaptation, other adaptations were necessary as well, including cardiovascular changes (to pump blood all the way up to the giraffe’s brain), shortening of the hindlegs relative to the forelegs (to ensure that the giraffe does not topple over), and so on. Should we conclude that the giraffe’s neck is a “high-leaf-eating organ”? Not exactly. The giraffe’s neck is still a neck, built out of the same basic blueprint that is used over and over in vertebrates, but with some quantitative adjustments. It still does other kinds of “neck work”, just like the work that necks do in less specialized species, but it has some extra potential for reaching up high in the tree that other necks do not provide. If we insist that the neck is a leaf-reaching organ, then we have to include the rest of the giraffe in that category, including the cardiovascular changes, adjustments in leg length, and so on. I believe that we will ultimately come to see our “language organ” as the result of quantitative adjustments in neural mechanisms that exist in other mammals, permitting us to walk into a problem space that other animals cannot perceive much less solve. However, once it finally appeared on the planet, it is quite likely that language itself began to apply adaptive pressure to the organization of the human brain, just as the leaf-reaching adaptation of the giraffe’s neck applied adaptive pressure to other parts of the giraffe. All of the neural mechanisms that participate in language still do other kinds of work, but they have also grown to meet the language task. Candidates for this category of “language-facilitating mechanisms” might include our social organization, our

extraordinary ability to imitate the things that other people do, our excellence in the segmentation of rapid auditory stimuli, our fascination with joint attention (looking at the same events together, sharing new objects just for the fun of it). These abilities are present in human infants within the first year, and they are clearly involved in the process by which language is acquired. We are smarter than other animals, to be sure, but we also have a special love of symbolic communication that makes language possible.

REFERENCES

- Aslin, R.N., Jusczyk, P.W., & Pisoni, D.B. (in press). Speech and auditory processing during infancy: Constraints on and precursors to language. In W. Damon (Series Ed.) & D. Kuhn & R. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 5. Cognition, perception & language* (5th ed.) New York: Wiley.
- Bates, E. (1976). *Language and context: Studies in the acquisition of pragmatics*. New York: Academic Press.
- Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.
- Bates, E., & Goodman, J. (in press). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. In G. Altmann (Ed.), Special issue on the lexicon, *Language and Cognitive Processes*.
- Bishop, D.V.M. (1997). *Uncommon understanding: Development and disorders of comprehension in children*. Hove, UK: Psychology Press/Erlbaum.
- Braine, M.D.S. (1976). Children's first word combinations. With commentary by Melissa Bowerman. *Monographs of the Society for Research in Child Development*, 41, Serial # 164.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1988). *Language and problems of knowledge*. Cambridge, MA: MIT Press.
- Dronkers, N.F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384, 159-161.
- Eimas, P.D., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 305-6.
- Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press/Bradford Books.
- Elman, J. & McClelland, J. (1986). Interactive processes in speech perception: The TRACE model. In D. Rumelhart & J.L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- Erhard, P., Kato, T., Strick, P.L., & Ugurbil, K. (1996). Functional MRI activation pattern of motor and language tasks in Broca's area (Abstract). *Society for Neuroscience*, 22, 260.2.
- Fenson, L., Dale, P.A., Reznick, J.S., Bates, E., Thal, D., & Pethick, S.J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, Serial No. 242, Vol. 59, No. 5*.
- Fletcher, P., & MacWhinney, B. (Eds.). (1995). *The handbook of child language*. Oxford: Basil Blackwell.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT 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.
- Freud, A. (1953). *On aphasia: A critical study* [E. Stengel, Trans.]. New York: International Universities Press. (Original work published in 1891).
- Gelman, D. (1986, December 15). The mouths of babes: New research into the mysteries of how infants learn to talk. *Newsweek*, 86-88.
- Gernsbacher, M.A. (Ed.) (1994). *Handbook of psycholinguistics*. San Diego: Academic Press.
- Head, H. (1926). *Aphasia and kindred disorders of speech*. Cambridge, UK: Cambridge University Press.
- Joanette, Y., & Brownell, H.H. (Eds.). (1990). *Discourse ability and brain damage: Theoretical and empirical perspectives*. New York : Springer-Verlag.
- Kawamoto, A.H. (1988). Distributed representations of ambiguous words and their resolution in a connectionist network. In S. Small, G. Cottrell, & M. Tanenhaus (Eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence*. San Mateo, CA: Morgan Kaufman.
- Kean, M.-L. (Ed.).(1985). *Agrammatism*. Orlando: Academic Press.
- Kempler, D., van Lancker, D., Marchman, V., & Bates, E. (1996). The effects of childhood vs. adult brain damage on literal and idiomatic language comprehension (Abstract). *Brain and Language*, 55(1), 167-169
- Kempler, D., van Lancker, D., Marchman, V., & Bates, E. (in press). Idiom comprehension in children and adults with unilateral brain damage. *Developmental Neuropsychology*.
- Kuhl, P.K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190(4209), 69-72.
- Langacker, R. (1987). *Foundations of cognitive grammar*. Stanford: Stanford University Press.
- Leonard, L.B. (1997). *Specific language impairment*. Cambridge, MA: MIT Press.

- MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676-703.
- McCawley, J.D. (1993). *Everything that linguists always wanted to know about logic*. Chicago: University of Chicago Press.
- Menn, L., & Stoel-Gammon, C. (1995). Phonological development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language* (pp. 335-359). Oxford: Basil Blackwell.
- Onifer, W., & Swinney, D.A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9(3), 225-236.
- Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain and Language*, 55(3), 352-379.
- Pick, A. (1973). *Aphasia*. (J. Brown, Trans., & Ed.) Springfield, IL: Charles C. Thomas. (Original work published 1913).
- Rumelhart D., & McClelland J.L. (Eds.).(1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- Samuel, A.G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110, 474-494.
- Slobin, D. (Ed.). (1985-1997). *The crosslinguistic study of language acquisition* (Vols. 1-5). Hillsdale, NJ: Erlbaum.
- Sperber, D. & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.
- Tanenhaus, M.K., & Lucas, M.M. (1987). Context effects in lexical processing. In U. Frauenfelder & L.K. Tyler (Eds.), *Spoken word recognition (Cognition special issue)* pp. 213-234. Cambridge, MA: MIT Press.
- Thelen, E., & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
- Toga, A.W., Frackowiak, R.S.J., & Mazziotta, J.C. (Eds.). (1996). *Neuroimage, A Journal of Brain Function, Second International Conference on Functional Mapping of the Human Brain, 3(3) Part 2*. San Diego: Academic Press.
- Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford University Press.
- Van Petten, C., & Kutas, M. (1991). Electrophysiological evidence for the flexibility of lexical processing. In G.B. Simpson, Ed., *Understanding word and sentence*. Amsterdam: Elsevier.