1 Functionalism and the Competition Model

Elizabeth Bates and Brian MacWhinney

The Competition Model is a framework for the crosslinguistic study of language use. It is designed to capture facts about the comprehension, production, and acquisition of language by real human beings, across a variety of qualitatively and quantitatively distinct language types. Our own work on the Competition Model has been illuminated by the insights of a particular class of linguistic theories called “functional grammar.” But the goals of linguistic and psycholinguistic research are often different, revolving around the now-classic distinction between competence and performance. Competence refers to the abstract knowledge of language possessed by an ideal speaker-listener, removed from the constraints and inconveniences of real-time language use. Performance refers to the actual process of language use by real people in real situations. As many linguists and psycholinguists have noted, there is no necessary and direct relation between competence and performance models (Fodor, Bever, & Garrett, 1974; cf. Bresnan, 1978). In principle, a given performance model may be compatible with a variety of competence models, and a given competence model may be compatible with many different characterizations of performance. The success of a performance model must be evaluated in its own terms. Can it predict when a speaker will use one form of expression and not another? Can it account for the use that listeners make of linguistic devices in comprehension? Can it account for the order of acquisition of grammatical devices in children? Can it relate language use to general aspects of human cognition and still account for specifically linguistic aspects of processing? Finally, and perhaps most importantly, can this model account for processing and acquisition in any given language no matter how much the structure of that language differs from that of English?

The particular performance model that we will offer here has grown out of a tradition in linguistics and psycholinguistics called functionalism. Linguistic functionalism is different in many ways from the functionalism of behaviorist psychology, and it is almost entirely in opposition intellectually to the functionalism of philosophy. Perhaps its closest relatives in other disciplines are in the “constructivism” of mathematics and psychology. Linguistic functionalism can be defined as the belief that “the forms of natural languages are created, governed, constrained, acquired and used in the service of communicative functions” (Bates & MacWhinney, 1982; MacWhinney, Bates, & Kliegl, 1984).
This approach is the natural alternative to theories of language that postulate a severe separation between structure and function, and/or theories that attempt to describe and explain structural facts sui generis, without reference to the constraints on form that are imposed by the goals of communication and the capabilities and limitations of human information processing.

A variety of competence models have been proposed within the functionalist tradition, including Eastern European functionalism (Dezsö, 1972; Driver & Fried, 1987; Firth, 1951; Mathesius, 1939), British functionalism (Halliday, 1966, 1967, 1968), generative semantics (Chafe, 1971; Fillmore, 1968), discourse analysis (Chafe, 1981; Givón, 1984; Li, 1975, 1976, 1977; Hopper & Thompson, 1980, 1984), Cognitive Grammar (Langacker, 1987; Lakoff, 1987), Construction Grammar (Fillmore, 1987), Role and Reference Grammar (Foley & Van Valin, 1984), and functionalist explanations couched within the formalisms of “standard theory” (Kuno, 1986). For the sake of simplicity, we will refer to these otherwise rather disparate linguistic approaches with the single term “functional grammar.”

Although functional grammars are not designed to account for real-time processing, they are most compatible with highly interactive models of performance, that is, models in which different sources of information are integrated on equal footing, as rapidly as possible. They are less compatible with models of performance in which different linguistic data types (phonological, morphological, syntactic, semantic) are suberved by distinct and encapsulated processors that communicate only after they have completed their domain-specific analyses. Such “modular” approaches to performance are more compatible with “modular” theories of competence, that is, linguistic theories that emphasize the autonomy of various components and subcomponents of the grammar (cf. Berwick & Weinberg, 1984; Bresnan, 1982; Pinker, 1984).

The Competition Model is more compatible with functional/cognitive grammars. However, we have found over the years that functionalism means different things to different people. Much of this confusion arises from the different ways in which the term “functionalism” is used by philosophers, psychologists, and linguists. To dispel some of this confusion, we think that it would be useful to spell out in detail the particular functionalist assumptions that have motivated our work— with special emphasis on the role of processing constraints, and on the quantitative principles that separate our approach from many other functionalist theories. Then we will provide a more precise account of the model itself, together with some illustrations of the data that support our current claims.

**What Is Functionalism?**

The term “functionalism” appears repeatedly in Western intellectual history, but its uses from one field to another are often so divergent that it is difficult to find a single unifying theme. In philosophy, we find the term used to refer to a position which attempts to separate the human thought from the neuronal hardware. This brand of functionalism views concepts in terms of mathematical functions which, once defined, may be related to one another in terms of their pure conceptual form. This form of analysis can be used to defend the relevance of Artificial Intelligence to Cognitive Science. According to this viewpoint, although the hardware of a digital computer is quite different from that of the brain, the functions it computes can be combined into systems that may well operate in ways that are quite analogous (Pylyshyn, 1984). This type of philosophical functionalism goes directly against many of the assumptions of the Competition Model. When we characterize our position as “functionalist,” we are certainly not thinking about this type of functionalism.

The term “functionalism” was also used to describe schools of empiricist psychology that were prominent in the first part of this century. Positivists like Dewey and Tolman and behaviorists like Skinner and Watson were all characterized as functionalists. Here the term “function” is being used to refer not to mathematical functions, but to the activities of the organism. These functionalists looked at ways in which the organism “functions” within either a real environment or a laboratory environment. The Competition Model also looks closely at relations between the environment and the functioning of the organism, but we are not using the word “function” in this way. Rather, we are using the word “function” to refer to “purpose” or “goal.” When we view language as a functional system, we are not simply saying that it is a system of activities. Rather we are saying that it is a goal-directed system of activities. There is no sharp line between these two senses of “function.” Accordingly, our approach attempts to provide a unification of both the positivist emphasis on cue validity (Brunswik, 1956) and the cognitivist emphasis on goals and plans (Newell & Simon, 1972).

Within linguistics and psycholinguistics, both of these last two types of “functionalism” are clearly opposed to the doctrine of Chomskian generative grammar. Chomsky (1957) argues forcefully against attempts to relate sentence structure to aspects of the environment, and Chomsky (1975) argues for a kind of autonomy of syntax that would cut it off from the pressures of communicative functions. In the Chomskian vision, language is pure and autonomous, unconstrained and unshaped by purpose or function. Although the Competition Model does not posit the simple sorts of relations between communicative function and language proposed by Skinner (1957), neither does it accept the total divorce of language from communicative function embodied in Chomskian linguistics. Instead, it considers the relation between language form and language function as the major empirical phenomenon to be characterized by psycholinguistic theory.

Four major themes unify the particular functionalist approach to language
that we will follow here: (1) cognition as the basis for language universals, (2) grammars as solutions to the mapping problem, (3) biology as providing the roots of function, and (4) quantitative analyses as ways of understanding qualitative variation. Let us first examine these four themes, and then move on to consider different levels and types of functional claims in linguistics and psycholinguistics.

Cognition, Language Universals, and Linguistic Relativism

We assume that the human biological apparatus is a constant across cultures. All normal adults have essentially the same apparatus for perception, articulation, learning, and memory, and we share a common set of social concerns. Human cognition and emotion provide the basic meanings and communicative intentions that any natural language must encode, together with a universal set of processing constraints that sharply delimit the way that meanings and intentions can be mapped onto a real-time stream of gestures and/or sounds. In other words, language universals derive from universal properties of the human mind.

The claim that there are cognitive universals underlying linguistic form does not mean that all human beings think alike. First, there are individual differences in the development of this basic apparatus; these differences exist within cultures, rather than across cultures. Second, cultures vary widely and the beliefs and practices of individuals vary accordingly. Culture-specific information is invariably reflected in the language spoken within a given community, and the rapid changes in culture arising from migration, cultural contact, or social change often result in rapid changes in vocabulary, idiomatic expressions, and sometimes even in changes in the more conservative morphosyntactic elements of the language. Nevertheless, there are certain basic categories of perception and thought that all natural languages must deal with at every point in their history: principles of motion, space and time, and principles of human action and intention. All natural languages have had to evolve some means of encoding distinctions among objects, qualities and events, modes of organizing events in time and space, human attitudes about those objects and events, and human attitudes toward one another. They have also necessarily evolved ways of encoding functions inherent in the communication process itself, that is, the identification of referents, the establishment of a given referent as a discourse topic, the process of making points or comments about particular topics, mechanisms for shifting and/or subordinating topics, and devices that help to create cohesion across the discourse as a whole. Every language is under constant pressure to develop and maintain ways of expressing these cognitive and communicative universals. It is not our goal in this paper to provide a complete account of those cognitive universals that are expressed in language. However, some useful accounts of parts of the universal system of cognitive representations are provided by Langacker (1987), Miller and Johnson-Laird (1976), Jackendoff (1983), Herskovits (1986), and Talmy (1977). Overviews of the universal perceptual-motor apparatus are given by Ladefoged (1980), Lenneberg (1967), and Lieberman (1975).

Note that we have stressed the roles of both cognitive content and cognitive processes in determining the possible forms that a language can take. This focus on process is crucial to our argument. Grammar is not a simple reflection of meaning. Content alone is not sufficient to explain why languages look the way they do, and the exigencies of real-time processing can ultimately result in forms that look relatively opaque (i.e., forms that have no obvious meaning or communicative goal). This, we think, is one of the major sources of misunderstanding in our field. A strong proponent of the autonomous grammar approach once asked us, "If you are right in your belief in the functional determination of grammar, why don't all natural languages look alike?" Other critics have, in all sincerity, accused us of not believing in grammar at all! Of course we believe in grammar, and in grammatical diversity. Indeed, we think that attention to grammatical diversity is long overdue in psycholinguistics. We are not trying to replace grammar with cognition; but we are trying to explain grammar in cognitive terms. We are convinced that Universal Grammar can ultimately be explained without recourse to a special "language organ" that takes up where cognition leaves off. Human cognition is the wellspring of language universals, setting limits on the form that any natural language can and must take. These are the same phenomena that some linguists have called "Universal Grammar," or UG. Certainly, a functionalist account will differ in detail from theories that are divorced from function, but the basic subject matter is the same: the universal properties of human language. Hopefully, this view will emerge more clearly as we elaborate on the interplay between content and form.

Nativists are not the only opponents of a functionalist approach to grammar. At the opposite empiricist extreme, we find followers of the Sapir-Whorf hypothesis (Bloom, 1981; Sapir, 1921; Whorf, 1967) who are willing to accept a strong form of linguistic relativism that questions the experiential basis of even the most fundamental segmentations of reality by the human mind. Whorfians argue that, insofar as languages can vary radically in their structure, human thought can also vary markedly. For these scholars, there are no "universal" principles of cognition or perception; rather, the categories of thought derive from the categories of language itself. Unlike the Chomskians, Whorfians recognize the importance of the relation between language form and language function. However, they see form as the determiner of function. For this reason, Whorfianism is often called "linguistic determinism." The Competition Model only rejects the most extreme forms of linguistic determinism. These extreme forms deny the possibility of universals of cognitive and communicative
functioning. In our opinion, evidence for such universals is so overwhelming that the strongest forms of linguistic determinism cannot be accepted. Moreover, experimental tests of strong linguistic determinism have never been successful (Osgood & Sebeok, 1965).

Weaker forms of linguistic determinism, such as those discussed in Osgood and Sebeok (1965) or Bowerman (1985), are compatible with the Competition Model. As we will see later, the Competition Model assumes a mix of form-driven and function-driven learning. Linguistic forms are themselves a part of the world within which the organism functions and to which it must adapt. Forms may orient the organism to encode relations between cues that it already perceives. Forms may even induce the organism to piece together new concepts from universal primitives. However, it is unlikely that linguistic forms actually induce the learner to form new cognitive primitives. Cognitive primitives are too important to the organism to be left to the whims of social convention.

**Grammars as Solutions to the Mapping Problem**

Grammars can be viewed as a class of solutions to the problem of mapping nonlinear meanings onto a highly constrained linear medium whose only devices are word order, lexical marking, and suprasegmentals. The universal and culture-specific contents of cognition interact with universal constraints on human information processing, creating a complex multivectorial problem space (Bates & MacWhinney, 1982; Karmiloff-Smith, 1984). In fact, there may be no perfect and stable pathway through this problem space. As Slobin (1982) has pointed out, many processing constraints are in direct competition; hence stability in one area may create instability in another. The situation can be compared to the eternal competition for funds between federal agencies and programs. The charge to spend money for “guns” is in constant competition with the charge to spend money for “butter.” Every fiscal budget is a compromise between these two charges and every new attempt to hammer out a budget requires a new resolution of these competing forces. Within language, the charge to be “quick and easy” is invariably opposed to the charge to be “clear.” From the listener’s point of view, a given linguistic marker will signal its meaning most efficiently if it is consistent, salient, and unique. For the listener, homonymy is a barrier to understanding and any deletion of phonological content can damage comprehension. The more detail the speaker provides, the easier the listener’s job. From the speaker’s point of view, homonymy facilitates retrieval and phonological weakenings and deletions cut down on articulatory effort. Composing utterances to meet the listener’s needs requires both cognitive and articulatory effort. Hence, the clear and perceivable markers that evolve for comprehension are often subject to erosion in the service of rapid and efficient speech output.

**Functionalism and the Competition Model**

There is a constant interplay between the forces of language erosion and language creation. Often a new marker will begin as a preposition or postposition. Over time, it will start to function as a prefix or a suffix and begin to undergo phonological erosion or leveling. At the same time, it may undergo functional reinterpretation (MacWhinney, in press; this volume). For example, the Hungarian accusative suffix developed out of a postposed element marking definiteness. As this element became identified with the marking of the role of the noun, a new way of marking definiteness emerged, this time as a preposed article. If a given marker were to express one and only one meaning, reinterpretations of this sort would be more puzzling. Why give up a marker of definiteness only to create another? But a central claim in the Competition Model is that forms often express a variety of correlated functions. This “peaceful coexistence” of functions helps both the listener and the hearer by allowing a small set of structures and markers to serve a myriad of related purposes. This then serves to build a certain adaptive instability into language, since any particular coalition of meanings may eventually collapse and require reinterpretation. Grammars are thus viewed as a set of partial solutions to the mapping problem, each representing one pathway through the constraints imposed by cognitive content and cognitive processing. No solution is perfect, and each one is constantly subject to change; but every grammar used by a community of human adults and acquired by their children has to meet certain implicit but implacable limits of tolerance.

This last point is eloquently illustrated by the inventions and adjustments that take place during creolization, the process by which a pidgin code formerly used only by traders suddenly has to serve as a full-fledged native language for a community of users. As Bickerton (1984), Sankoff (1980), and other Creole scholars have shown, pidgin codes subjected to the pressures of full language use evolve within one generation the grammatical forms that are needed to express complex ideas within a cohesive oral narrative. A similar process has been observed in the acquisition of American Sign Language by deaf native speakers. Deaf children often receive their sign language input in a grammatically incomplete form—either from hearing teachers who acquired the language in adulthood, or from deaf parents who also acquired ASL late in life. Newport and Meier (1985) have shown how deaf children systematize this impoverished input, regularizing certain principles and extending others to cover new cases. Goldin-Meadow (1982) has studied the most extreme case of this type where deaf children with hearing parents create a small pidgin gestural system in order to express their most basic communicative intentions. Two or three generations of this process are apparently sufficient to turn an unsystematic gestural system into a full-fledged human language.

For Creole scholars like Bickerton (1984), the rapid and predictable course of creolization can only be explained with recourse to an innate and domain-
specific "language bioprogram," much like the universal grammar envisioned by Chomsky (1965) and Lightfoot (1982). We would argue, instead, that creolization (and related diachronic phenomena) reflect formal constraints that inhere in the mapping problem itself. Consider the forces that operate to create a soap bubble: the sphere that forms is the inevitable result of an attempt to simultaneously attain maximum volume with minimum surface area. Similarly, universal aspects of human grammar may emerge inevitably whenever universal categories of thought have to be mapped efficiently onto a limited channel.

Functionalism and Biology

All this is not to say that the Competition Model denies the importance of biology. Indeed, there has been a great deal of misunderstanding on this particular point. We think that much of this misunderstanding comes from a failure to distinguish between innateness and domain-specificity. The innateness issue has to do with the extent to which human language is determined by the unique biological heritage of our species. But this biological heritage may include many capacities that are not unique to language itself: our large and facile brain, our particular social organization, our protracted infancy, and a variety of unknown factors that may contribute in indirect but very important ways to the problem of mapping universal meanings onto a limited channel, and to the particular solutions that we have found to that problem. Hence, the human capacity for language could be both innate and species-specific, and yet involve no mechanisms that evolved specifically and uniquely for language itself. Language could be viewed as a new machine constructed entirely out of old parts (Bates, 1979).

Piaget made this critique of Chomsky's linguistic nativism in 1958 (Piaget, 1970) and the force of his criticism has not been diminished by the years. At issue is the difference between innateness and inevitability. Outcomes that are inevitable on structural grounds do not have to be innate. Indeed, an outcome that is insured by problem-solving constraints may be much more robust across a range of situations than an outcome based entirely on direct genetic guidance. In an edited volume documenting an historic encounter between Piaget and Chomsky (Piaget, Chomsky, & Piatelli-Palmarini, 1980), we find Piaget criticized as an "antibiological behaviorist." Our own position has been misunderstood in the same way. In this light, we would like to state as explicitly as possible that we view grammar as a biological system. However, the universal properties of grammar are only indirectly innate, being based on interactions among innate categories and processes that are not specific to language. In other words, we believe in the innateness of language, but we are skeptical about the degree of domain-specificity that is required to account for the structure and acquisition of natural languages.

Functionalism and the Competition Model

Qualitative and Quantitative Variation

Like other biological systems, natural languages display a great deal of intraspecies variation. Although the constraints imposed by the mapping problem are heavy, and the class of possible solutions is finite, the number and range of language types that are possible seems to be quite extensive. Languages differ qualitatively, in the presence or absence of certain linguistic devices (e.g., word order constraints, case marking), but they also differ quantitatively, in the extent to which the "same" linguistic device is used at all and in the range of functional roles that the "same" linguistic device has come to serve.

We will give a number of examples of quantitative differences between languages throughout this volume. One particularly important example has to do with the relative strength of word order versus subject-verb agreement as cues to sentence meaning. In English, word order is a strict and highly valid cue to sentence interpretation and that order is usually Subject-Verb-Object or SVO. In Italian, on the other hand, word order can be varied extensively for pragmatic purposes. This is illustrated in the following imaginary restaurant dialogue taken from Bates, MacWhinney, and Smith (1983):

1. SVO: Io mangerei un primo. (I would eat a first course.)
2. OSV: La pastasciutta Franco la prende sempre qui. (Pasta Franco it orders always here.)
3. VSO: Allora, mangio anche io la pastasciutta. (Well then, am eating also I pasta.)
4. VOS: Ha consigliato la lasagna qui Franco, no? (Has recommended the lasagna here Franco, no?)
5. OVS: No, la lasagna l'ha consigliata Elizabeth. (No, the lasagna it has recommended Elizabeth.)
6. SOV: Allora, io gli spaghetti prendo. (In that case, I the spaghetti am having.)

This short and realistic conversation contains all possible orders of Subject, Verb, and Object. Some of these require particular intonation patterns to sound exactly right, and some are definitely better with particular grammatical markers like the object clitic. But all these orders can be found in a large enough sample of free speech, and all of them occur at some point in the input received by Italian children (Bates, 1976).

At one level, the sentences given above simply serve to illustrate well-known qualitative differences between languages: Italian has word order options that do not exist in English at all. However, this qualitative variation also has quantitative implications. We have demonstrated in several different experiments that Italian listeners "trust" word order - even good old-fashioned Subject-Verb-Object order - less than their English counterparts. Given a sentence like "The pencil hits the cow," English listeners from ages 2 to 80 have a strong tendency to pick the pencil as the agent/subject. Given the Italian equivalent La matita colpisce la vacca, Italians are much more likely to choose the cow as the agent/subject. Hence a qualitative difference in the availability of word order
types has a quantitative effect even on that subset of grammatical structures that both languages share (e.g., SVO order).

This point is made clearer still by observing the reaction of subjects to a semigrammatical sentence like “The cow are hitting the horses.” English listeners do not like this sentence very much, but in a forced choice situation they overwhelmingly choose the cow as the subject of the verb. In other words, they “distrust” morphological cues to meaning, preferring to rely on word order instead. Speakers of richly inflected languages with variable word order like Italian or Hungarian do not behave this way at all; they “trust” their morphological cues, and rapidly choose the plural noun as the subject of a plural verb regardless of constituent order.

This division between inflected and uninflected languages is useful but it is still too discrete to account for the variation that we have observed to date. Instead, there are degrees of “word order dependence” or “morphological dependence.” For example, German seems to stand somewhere in between the Italian and English extremes described above: There is more reliance on grammatical morphology than in English, but more reliance on word order than in other richly inflected languages (MacWhinney, Bates, & Kliegl, 1984). Results from Chinese are particularly interesting in this regard (Kilborn & Ito, this volume). Mandarin Chinese has very little inflectional morphology, and no morphemes at all to indicate agent/object relations. Nevertheless, when lexical semantics and word order are placed in competition, native speakers of Chinese base their interpretations on lexical rather than syntactic cues. As we shall see in several chapters to follow, these differences of degree are principled, reliable and predictable. They follow from statistical as well as structural facts in the languages studied to date. For example, the weak word order effects in Chinese sentence comprehension make sense in light of (1) the frequency with which the subject is omitted in Chinese (reducing the utility of preverbal position as a cue to meaning) and (2) the availability of certain marked and noncanonical word orders in both formal and informal speech.

Most of the examples presented in this book come from studies of sentence comprehension. But we have also uncovered some interesting quantitative differences in the domain of sentence production. For example, Bates and Devescovi (this volume) have described some robust differences between Italian and English in the use of relative clauses. The structural options available in the two languages are the same, at least for the set of structures studied by these investigators. In both languages, it is perfectly grammatical to describe a picture of a monkey eating a banana by saying either “A monkey is eating a banana” or “There is a monkey that is eating a banana.” English speakers typically use the first option. By contrast, Italian speakers describing exactly the same pictures, under the same conditions, are three to five times more likely to produce a relative clause. This crosslinguistic difference in relative clause use is already well-established in children by the age of three, and it tends to persist even in elderly patients who have suffered left-hemisphere damage. How can we capture a quantitative difference between two structures that are equally grammatical from a traditional grammatical perspective? To be sure, there are some differences between the two languages in the range of functions that control these particular forms. In particular, Italians appear to use the relative clause as a kind of topic marker. But in addition to (and perhaps because of) these differences in function, there are also clear processing differences between English and Italian in the “accessibility” of the relative clause. Function and frequency codetermine the selection of grammatical forms in sentence production (together with a number of other processing factors like recency and point of commitment – see Bates & Devescovi, this volume, for details).

Our emphasis on quantitative variation is probably the greatest point of divergence between the Competition Model and other functionalist theories. With some exceptions, functional grammars tend to state the relationship between form and meaning in terms that are both direct and discrete: “Use structure X if and only if semantic factors Y and pragmatic conditions Z are met.” From this point of view, human language use should be perfectly predictable from meaning. Imperfect, probabilistic mappings between form and function would then reflect a failure on the part of the linguist (or psycholinguist) to provide an adequate account of the rich array of contextual factors governing the formal device in question. It is important, of course, to dig as deeply as possible into the contextual motivation underlying the use of forms. Work by Gee and Savisir (1985) indicates how much we can sharpen our understanding of the use of linguistic devices by a sensitive study of the context of their uses. However, such studies focus on qualitative aspects of language functioning, seldom recognizing the fundamentally probabilistic nature of language processing.

The emphasis in the Competition Model on probabilistic rules does not mean that we ignore the powerful laws that separate one language from another. After all, the values “0” and “1” do exist even in a probabilistic system, and an adult native speaker may come to know with some certainty that a particular structure is impossible in his or her language. The difference between our characterization of adult knowledge (i.e. “competence to perform”) and the characterizations offered in most competence models lies in our ability to capture the many values that fall between 0 and 1. We describe linguistic representations in terms of a complex set of weighted form–function mappings, a dynamic knowledge base that is constantly subject to change.

This quantitative approach has a number of advantages in the description of a language at a particular point in time (i.e., synchronics); but perhaps more importantly, quantitative description makes it easier to explain linguistic change...
(i.e., diachronics). In the evolution of a single language from one synchronic type to another, in the acquisition of language by children, and in the loss of language by individuals or by whole communities, we need an appropriate means for characterizing what it means to be "in between" structures. A theory phrased entirely in terms of discrete, qualitative rules cannot serve this purpose—whether or not those rules make reference to cognitive content.

There are at least two traditions within functionalist linguistics that do deal directly with quantitative facts about language use. The first is laid out most explicitly by Givón (1979, 1984), although it can also be found in work by Keenan and Comrie (1977), Kuno (1986), Hopper and Thompson (1980), and many others. Givón's approach to quantification is different from the one adopted here, in that it rests on ordinal rather than interval scales. He describes "clines" or "dimensions" of topicality, foregroundedness, and other communicative functions along which different linguistic forms can be ordered. On the foregroundedness cline, explicit noun phrases are higher in foregroundedness than pronouns, pronouns are higher than zero anaphors, and so on. Many of the ordinal claims that result from this kind of scaling have been verified in detailed text analyses in a variety of structurally distinct language types. However, this approach makes no predictions about the distance between different steps on the foregroundedness scale (or any other functional cline). In the terms of measurement theory, clines make predictions only about ordinal relations on scales. We believe it is now possible to go further by specifying precise degrees of correlation between forms and functions. That is, we can now work with interval, rather than ordinal, scales.

An attempt at interval scaling can be found within the variable rule school of sociolinguistics led by Labov. As elaborated in Sankoff (1978), variable rule theory represents an effort to state directly in the grammar the probability that a given rule will apply within a particular social context. Researchers working with variable rule theory have described in detail the statistical variation that can be observed in a large number of linguistic forms, ranging from /u/ and /vou/ in French Canadian to /s/-deletion in Puerto Rican Spanish. These alternations are not random; they cannot be explained by attaching a coin-flipper to a discrete rule. Rather, the observed alternations are quite lawful and are conditioned by a variety of factors in the phonological, syntactic, and discourse context of the alternation. Labov has explained these statistical regularities with a device called a "variable rule," that is, a rule with an explicitly stated weight or probability on its application.

Labov's use of variable rules to estimate parameter strengths is in the best tradition of mathematical psychology. Sankoff's log-linear instantiations of those models are close mathematical cousins of the maximum likelihood models used by psychologists such as Anderson (1982) and Massaro (1987) to study information integration, an approach which in turn serves as the basis of our

Functionalism and the Competition Model

The most recent work in statistical modeling (McDonald, this volume). The cognitive assumptions underlying variable rule theory are quite familiar to psychologists, since so much of psychological theory is stated in terms of activation and strength. Like Labov, psychologists generally believe that it is possible to build a precise theory of imprecise phenomena. Statistical regularities and lawful gradations must be described and explained. They are part of the native speaker's knowledge of his/her language, and they are an important source of information (not just an unfortunate form of noise) for the language learner. Physicists have lived with the fundamental indeterminism of physical reality for most of this century. Perhaps it is high time that linguists and psycholinguists follow suit. Along these lines, the Competition Model can be viewed as the psycholinguistic analog to quantum theory in physics.

Levels of Evidence for Functionalism

Different kinds of functionalist claims require different kinds of evidence. This is a point that we have made in several places (Bates & MacWhinney, 1982; Bates, MacWhinney, & Smith, 1983; Bates & MacWhinney, 1987), but it is sufficiently important that we think it deserves reiterating here. We distinguish four different levels of functionalist claims, ordered from weakest to strongest (in the sense that claims at the higher levels presuppose that claims at the lower levels are true). Level 1 focuses on the role of cognitive and communicative functions in the evolution of language proper, and in the history of individual languages. Claims at this diachronic level need not (as we shall see) have implications for current language use by adults, language acquisition by children, or the proper characterization of grammatical knowledge. Level 2 is a synchronic variant of Level 1, focusing on the causal relationship between form and function in real-time language use by adult speakers of the language. Level 3 presupposes, but also goes beyond Level 2, focusing on the causal role of cognitive and communicative functions in language acquisition by children. Finally, Level 4 is reserved for the claim that facts from Levels 1 through 3 play a direct role in the characterization of adult linguistic competence. The Competition Model includes claims at Levels 1 through 3; functional grammars reside at Level 4. Let us examine these four levels one at a time, in order to clarify exactly what kinds of claims are involved in the Competition Model. This will hopefully help the reader to evaluate the relationship between theory and evidence in the chapters that follow.

Level 1—Language History

Claims at Level 1 constitute a kind of linguistic historical evolutionism, which holds that functional constraints have played a role in determining the forms that grammars take today. Where did the tiger get his stripes? Why do giraffes...
have long necks? Why do grammars have relative clause markers? A great deal of work in functionalist linguistics focuses on the possible functional origins of grammatical structures during language evolution. For example, Givón, Sankoff, and others discuss the development of relativization structures from deictic elements. Many writers discuss how agreement marking on the verb arises from the fusion of pronominal elements to the verbal stem. Givón (1984) and Bybee (1985) discuss the development of aspectual markings and tense markings from general verbs such as “have,” “go,” and “be” that become coopted into particular grammatical functions. Grammaticalization is the process by which relatively transparent lexical combinations become conventionalized into grammatical morphemes; in the course of this process, through use with a larger and larger array of lexical items, the original meaning of the proto-morpheme may become “bleached.”

Historical functionalism is the weakest kind of functional claim, because it carries no necessary consequences for the psychological relationship between form and function in a living native speaker today. Evolution is littered with examples of vestigial forms, structures, and behaviors that were selected at one point to serve a function that is no longer relevant. For example, humans still have wisdom teeth, an appendix, and the Babinski reflex, although none of these structures serve the purposes for which they originally developed. Because we are no longer toothless by the age of 20, wisdom teeth no longer serve the function of providing us with replacements for lost teeth; the appendix no longer serves as a digestive organ, although it may now play a role in the immune system; and the Babinski reflex is no longer needed to prepare us for falling out of trees, since we no longer live in trees. It is both logically and biologically possible that pieces of grammar also have achieved a certain functional autonomy (i.e., vestigiality) in present-day language use. However, we suspect that vestigial forms are relatively rare in language, because there are too many competing pressures on real-time language processing to permit waste. As soon as forms start to seem vestigial, the frugal aspects of the human mind begin to subject them to reinterpretation (MacWhinney, in press). To justify this claim, we need to spend some time carefully analyzing the different kinds of causal relationships that can exist between cognitive functions and grammatical forms.

Icon, Index, and Symbol

Peirce (1932) has offered a taxonomy of sign—meaning relations that is useful in characterizing cause-and-effect relationships at Level 1. According to Peirce, a sign or signal can come to stand for its referent in one of three ways: iconic, indexical, and symbolic determinism. If any of these three relations holds, then there is a functional relationship between the sign and its referent. Some researchers (e.g., Petito, 1987) seem to equate functionalism

with iconicity, believing that evidence against iconicity constitutes evidence against functionalism. This is not true, since a sign is functional whenever any of these three relations hold.

Iconic Determinism. Iconicity refers to a literal physical resemblance between a sign and its meaning. A picture of a house looks like a house; the line drawing of a man or a woman on a restroom door tell us which sex may use the facilities; a picture of a cow at a cattle crossing warns drivers to slow down because a large animal may be crossing the road. Although the idea of iconicity is an appealing one, in reality most iconic signs bear only a partial, often minimal, resemblance to their referents—a fact that becomes clear to the traveler in Europe struggling to decipher international highway signs and symbols next to hotel listings in the Michelin guide. Hieroglyphic writing systems have an iconic base, but the physical resemblance between sign and referent is usually evident only after the learner has been told the meaning (“Oh yeah, I guess that does look sort of like a house...”). There is good reason to believe that iconic relations are rare and evanescent in the evolution of language. In the lexicon of spoken languages, we find identifiable iconic relations in only a handful of onomatopoeic expressions like “sneeze” and “bang.” At the sentence level, it is even harder to find transparent examples of an iconic relationship between meaning and form. One possible candidate is the relationship between degrees of pragmatic emphasis and degrees of phonological contrastive stress (Bolinger, 1986). However, a close analysis shows that contrastive stress also interacts with canonical word order and a host of other grammatical conventions that bear no one-to-one, isomorphic relation to meaning.

The minor role of iconicity in language evolution is particularly clear in studies of sign languages (Klima & Bellugi, 1979; Newport & Meier, 1985). For example, Hoffmeister (1978) has shown that, despite the transparent iconicity of the signs for locative relations like “in,” “on,” and “under,” deaf children make the same kinds of mistakes displayed by hearing children (e.g., they tend to place objects in boxes and on tables, regardless of the language input, evidently relying more on their knowledge of canonical events than the relationship expressed in the language). And Meier (1987) has shown that verb agreement in ASL is determined by morphological rather than iconic factors. The minimal contribution of iconicity to language acquisition and language use has occasionally been cited as evidence against a functional approach to grammar (e.g., Petito, 1987). However, we think this conclusion is unwarranted, because there are still two viable routes for functional determination of form: symbolic and indexical determinism.

Symbolic Determinism. There are many areas where a weaker form of functional determination holds: symbolic determinism. In Peirce’s terminology,
a symbol is a sign that comes to stand for its referent simply because a community of users has agreed to use the sign in this way. There is no "natural", discoverable link between a symbolic sign and its referent, and if the community of users that established such conventions should disappear without a trace, the probability that another community would assign the same sign to the same meaning is astronomically small. In other words, symbolic signs are completely arbitrary. Symbolic determinism appears to be the predominant form of sign-meaning relationship within the lexicons of spoken languages, and it operates within the grammar as well (Langacker, 1987).

But notice that what is arbitrary in a symbolic relation is only the specific choice of a sign to go with a referent. The basic decision to choose a sign to express a particular referent is not arbitrary at all. Rather, it is governed by the pressure of fundamental communicative functions. For example, the Italian language needed some kind of symbol to fill out the conditional paradigm, expressing the counterfactual status of a third person plural verb. Could a language get by without such verb markers? Yes. Chinese has no explicit verb morphemes to convey counterfactual reasoning - a fact that has misled some Western researchers into believing that the Chinese cannot think in counterfactual terms (Bloom, 1981). However, as Au (1983, 1984) has shown, Chinese does make use of lexical markers to indicate the same meanings that are conveyed by modal verbs in English, and by conditional suffixes in Italian. When those markers are removed (as in Bloom’s experiments), Chinese subjects do indeed fail to demonstrate counterfactual reasoning. But when Au provided her Chinese subjects with the appropriate lexical cues, they performed just like English subjects on counterfactual conditionals; conversely, when Au removed the counterfactual markers in her English materials, English subjects performed just like the Chinese in Bloom’s experiments. We accept Au’s conclusion that the concept of underlying counterfactuality is a cognitive/communicative universal. Every natural language will evolve some means of expressing this idea, but the particular means selected to carry out this work may vary markedly from one language to another.

Symbolic determinism is, then, a form of functional causation. Grammatical symbols can carry out communicative work without bearing a literal physical resemblance to their associated meanings or functions. Sometimes the functional base of a grammatical symbol may be difficult to discern (more on this below), but the lack of an iconic relationship between form and function does not mean that no relationship exists at all. This is particularly true when we recall that grammatical devices can evolve in the service of cognitive processing as well as cognitive content. For example, gender markers in German bear only the most indirect and infelicitous relationship to sex. There is no obvious semantic basis for classifying "little girl" as neuter, "sun" as feminine, and "moon" as masculine. And yet, gender markers may be crucial in helping the listener to keep track of referents across a complex passage of discourse (see Kilborn, 1987, for a demonstration of the role of German morphological markers in lexical recognition). Tags of some kind are needed to do the necessary communicative work; it is not obvious why gender was borrowed for that purpose in the first place, but if gender were not available then some other distinction would have to be drafted into service.

Indecisive Determinism. The third kind of sign-referent relation in Peirce’s taxonomy is the most subtle and the most difficult to explain. And yet we think that this form of determinism may play a particularly important role in the evolution of grammar. An index is a sign that comes to stand for its referent not because of physical resemblance, nor because of an arbitrary decision by a community of users, but because the sign participates in some natural and discoverable way with the event or object that it represents. Thunder is an index for a coming storm; smoke is an index for fire; the sound of footsteps in the brush is an index of coming danger for small forest animals. Like icons, but unlike symbols, indices bear a natural relationship to their referents. However, the "natural" relationship between an index and its referent must be discovered. The relation between an icon and its referent is so transparent that no real process of discovery is required. The relation between a symbol and its referent is so arbitrary that there is no preset relation ready to be discovered. Only in the case of the index can we say that there is a preset relation that must be discovered through learning and exploration. We must learn that lightning precedes thunder. Once we have learned and elaborated the connections between these phenomena, the relation then seems natural. However, if we are not aware of the natural link that binds sign and referent together, the relation may seem quite opaque. For a new learner, it may be impossible to distinguish between indices and true symbols (i.e., the arbitrary sign-referent bonds discussed above). Hence, an index may be learned by rote (see Bates, 1979, for a further discussion of this point).

Many so-called iconic signs are actually indices, or at least, mixtures of iconic and indexical relations. Consider, for example, the use of footprints as a clue to the presence of a large animal. The footprint bears an iconic relationship to the animal’s foot; however, the foot itself stands in a part-whole relationship to the animal we are looking for. This part-whole or participatory relationship is in fact an index. Hence the chain of meaning that lies behind the use of a footprint as a signal is a mixture of iconic and indexical relations.

Some more interesting and illuminating examples of mixed sign relations come from American Sign Language. For example, the ASL sign for "nurse" is a cross signed on the upper arm. This sign does not look like a nurse at all. But it does look like the red cross on a nurse’s sleeve, an entity which in turn forms a part of (i.e., participates in) the complex of features that make up our concept
of nurses as a class. There is also a sense in which this mixed iconic/indexical relation is quite arbitrary (i.e., symbolic). Why choose the red cross on the sleeve, and not some other part of the nurse-concept, as the part that will be chosen to stand for the whole? In fact, some signed languages form their signs by choosing an altogether different bit of a complex meaning to serve as the basis for an iconic/indexical sign (e.g., a nurse’s cap). There is, then, a certain degree of arbitrariness in the particular choice of parts to stand for a whole – a kind of sign relationship called “metonymy.”

The possible role of indexical deterministic in signed languages is clear. But what possible role could this kind of causation play in a spoken language? Silverstein (1976a) has argued that much of language is indexical in nature, and that indices are the basis and indeed the definition of that aspect of language called “pragmatics.” The paradigm case of indexicality in language can be found in pronouns and other forms of deictic reference. The pronoun “it” can take on meaning only via its relation to the context in which it is uttered: in the nonverbal context (so-called exophoric reference), in past verb discourse (anaphoric reference) or, less often, in the discourse that is about to be uttered (cataphoric reference). Pronouns provide the most obvious illustration of linguistic indexicality, but many more subtle forms could be mentioned. In general, a relationship between a sign and a meaning is indexical whenever the understanding of that sign is at least partially dependent on aspects of the context in which the sign is used.

We think that the concept of indexical determinism can be usefully extended to several other aspects of language use and language change. These include (1) the metaphorical extension of signs, (2) the reinterpretation of the meaning of signs, and (3) pressures on the shape of signs that arise from processing. In each of these cases, language processing itself produces a “natural” link between form and meaning – one based on participation or indexicality rather than iconicity or physical resemblance.

**Metaphoric Extension.** Metaphoric extension is an extremely common process in language, in which a form is extended to convey a meaning that is only indirectly, partially, or peripherally related to its semantic core (Lakoff & Johnson, 1980; Lakoff, 1987; MacWhinney, in press). We are usually aware of poetic metaphors like “John is a young lion,” somewhat less aware of metaphors that have attained an idiomatic status like “John spilled the beans,” and perhaps even less aware of metaphors in which a concrete sensorimotor process is extended to cover abstract cases like “John was unable to convey his thoughts to her.” The formation of new compound words such as “couch potato,” new derived words such as “teach-in,” and new inflected words such as “hacked” (in the computational sense) all require the active utilization of metaphoric extension. These issues are discussed further by MacWhinney (in press). For

**Present purposes, we would simply like to note that metaphor is a kind of indexical relation.** One meaning or piece of meaning that we would like to express overlaps with or participates in another. The “borrowing” of forms is ensured by (or enabled by) this overlap in meaning.

**Reinterpretation.** Whenever forms begin to lose their functional motivation, reinterpretation works to breathe new life into them. Reinterpretation takes advantage of the fact that natural language is based on richly confounded combinations of overlapping functions and forms (the same situation that is responsible for metaphoric extension, above). The notion of sentential subject serves as a particularly good example of a form–function coalition to illustrate this point. At the surface level, we find several morphosyntactic devices that tend to be assigned to the same element of meaning. In English, each sentence has an element which (1) is expressed as a noun phrase (a pronoun or an explicit noun), (2) agrees with the verb in person and number, (3) tends to appear in preverbal position, and (4) carries nominative case if it surfaces as a pronoun. These morphological and positional facts cooccur so reliably that we tend to think of them as obligatory rules, operating as a block (although this occurrence can break down, legally or illegally – more on this point later).

In addition, we find a set of “optional” formal devices that are also correlated (although less strongly) with the subject coalition. For example, the subject is more likely to be a definite noun phrase, and more likely than other arguments of the verb to be modified by a relative clause.

Agent-mapping devices like nominative case and topic-mapping devices like definiteness are correlated at the surface level because agent meanings and topic meanings are correlated at the semantic level. But where does the semantic correlation come from? Agency and topichood are both important and common meanings in human discourse. Hence, they each deserve high priority access to the “real estate” provided by the grammar. This might lead (and occasionally does lead) to a situation of competition in which semantic and pragmatic roles struggle for control over important grammatical options. Fortunately, however, the rules of agent and topic are usually assigned to the same referent. That is, the agent of an action is more likely than any other semantic role to serve as the topic of a passage of discourse, and the topic of a given discourse is particularly likely to be the protagonist in some human drama (the agent of an action, or a closely related role such as the experiencer of an emotion or state). This kind of semantic/pragmatic overlap no doubt derives from social universals: we like to talk about ourselves and our activities. In fact, in oral discourse the topic is also particularly likely to be the speaker or the listener. This social universal results in a statistical bias toward first or second person subjects – a fact that has also been conventionalized into some grammars, such as so-called “split ergative” languages – see Silverstein, 1976b). For present purposes, the point is that the
statistical overlap between agent and topic at the level of meaning results in a tendency for many natural languages to assign agent and topic marking devices together as a block. Across generations, this cooccurrence of form may lead to a "reinterpretation" of the relation between forms and meanings, particularly when some of the dominant form-meaning relations have become "bleached" through overextension. This is a kind of linguistic guilt-by-association. For example, Bates et al. (1982) have shown that Italian adults tend to interpret definite nouns as the agent of a transitive action. Definiteness is in the subject coalition primarily because of its association with topichood (in particular, the portion of topichood that is contributed by givenness); however, the long association between topic and agent has led to a situation in which definiteness can also be used as a cue to agency. This kind of plurifunctionality is, we will argue, the rule rather than the exception in natural language. One-to-one mappings between form and function are rare. Under the real-time pressures of adult language use, redundancy is welcome. As a result, double function mitigates against the existence of vestigial grammatical devices. For example, suppose the functional link between definiteness and topichood in Italian were to erode over time (for reasons that need not concern us here); the definite/indefinite contrast might be taken over in the service of transitivity, that is, to function as "who did what to whom." MacWhinney (in press) offers some examples in terms of verbal and case marking in Hungarian.

Darwin made extensive use of the principle of double function to salvage the principle of natural selection (Gould, 1983). Near-perfect adaptations like wings or eyes are easy to explain within a creationist framework, but they present problems for the Darwinian point of view. What possible function could have been served by two-thirds of an eye or seventy-five percent of a wing? And yet, if those forms evolved from other forms, something must have kept them going on their way toward their ultimate functions of seeing and flying, respectively. Darwin's answer was that the intermediate forms are maintained by different functions altogether, permitting gradual selection of the new activity over generations. For example, it appears to be the case that wings evolved from membranes that initially served the function of helping to maintain body heat. Indeed, the flapping motion that is part and parcel of flying in the current organism may also have originated from this original heat-maintenance function. Nature is apparently content to pass on half-baked jobs from one generation to another, lending the necessary time for better mechanisms to evolve, as long as the intermediate forms are contributing to survival and reproductive success. Vestigial forms will also be tolerated if they don't get in the way, but they are more likely to be maintained if they can be "recycled" (cf. interpreted) in the service of some other function. If we think of language in these biological terms, then it becomes clearer how some of the more exquisite structures of grammar could have evolved in the first place.

Functionalism and the Competition Model

Processing Determinism. Psycholinguists have often noted that the shape of the communicative act in which a given device participates feeds back on that device to determine or influence its shape. The device will not "look like" its meaning (iconic determinism), but it nevertheless has to adjust to real-time processing pressures if it is going to convey that meaning. The speech event in which the device participates places constraints on where the device should occur in the speech stream (position), and how much time it should take (length).

The marking of relative clauses can be used to illustrate several different forms of processing determinism. The basic function of the relative clause is to identify the head of the relative (Keenan & Comrie, 1977). As MacWhinney and Pléh (in press) note, this function is best served when the language has clear ways of marking the identity of the head of the relative clause and the role of the head in the relative clause. If one were to ignore processing considerations, there are many ways of marking these relations that one could imagine. Consider three marking techniques that are almost never used: (1) the presence of a relative clause could be marked at the beginning of the clause, (2) the identity of the head could be marked by agreement particles on the main verb, and (3) the role of the head in the relative clause could be marked by prefixes on the head. In Downing's (1978) survey of relative clause marking types in 52 languages, we find that none of these three techniques are used. Rather, languages tend to use markings that are far easier to process. The identification of the head of the relative clause is usually done by placing the relative directly before or after the head. However, this particular solution sometimes means that the relative clause may interrupt a main clause. Kuno (1973) and Hakuta (1982) argue that this kind of interruption can be costly for two reasons. First, because relative clauses are longer than most modifiers, the main clause has to be held open for a rather long time. Second, because relative clauses resemble main clauses in many respects, there is a potential for confusion (e.g., which verb goes with which noun). One way to relieve the burden on processing is to transpose the relative clause to the end of the sentence. Many languages do this by attaching special markers to the head to indicate that an extraposition has occurred. This type of marking, along with the various other markers of relative clause structure, can be understood as an Indexical marking whose shape is governed by processing determinism.

In general, there are a small number of high-probability solutions to the relative clause-marking problem. The exact solutions chosen by a particular language depend on the presence of other structures in that language.

Another example of processing determinism is found in the tendency for languages to place their modifiers in one preferred position (i.e., before or after the noun), or in the tendency to line noun and verb inflections up together on one side of the root word (usually as suffixes). As discussed by Venningmann (1974), it may simply be easier in both production and comprehension to locate all forms of a given type in the same general area. This pressure is not an
Functionalism and the Competition Model

that operate across the system as a whole. Such systemic pressures come not from the cognitive content of grammatical forms, but from the many different and often competing demands of perception, production, learning, and memory.

Level 2 – Language Use

At Level 2 we find claims about in vivo causal relationships between form and function, constraints that operate in real-time language use. This is a synchronic version of the diachronic claims at Level 1. We set these out as two separate levels because in fact we cannot assume that old cause-and-effect relationships are still operating in today’s grammar. As we noted earlier, it is at least logically possible that certain forms once served a communicative function but now have no clear communicative role. Such vestigial forms would motivate a distinction between Level 1 functionalism and Level 2 functionalism. But there is also a second and more compelling reason to keep Levels 1 and 2 separate: Many of the functional constraints that we have described so far are large-scale effects that are visible only across many generations, collective effects of communication among many different individuals. We cannot assume that each of these cause-and-effect relationships is detectable, or, for that matter, operative at all in a single speaker. Historical language change may be the linguistic equivalent of population genetics, with facts and principles that depend upon but also transcend knowledge of the individual organism.

Two kinds of methods have been used to establish the existence of “live” cause-and-effect relationships in individual language users. Linguists tend to make use of text analysis, demonstrating that particular target forms are always or typically used only in a certain set of functional/communicative contexts (for example, Zubin, 1977, 1979). Psycholinguists are, instead, more likely to manipulate putative cause-and-effect relationships directly. In studies of sentence production, functional conditions are manipulated as independent variables (e.g., agency, givenness, topicality); the linguistic devices that speakers produce under those conditions serve as the dependent variable(s) (e.g., a particular word-order configuration, case markers, contrastive stress, or pronominalization). This logic is reversed in studies of sentence comprehension: Surface forms are presented to subjects in different competing or converging combinations (i.e., as independent variables); the meanings or interpretations that the subjects derive serve as dependent variables. This kind of design characterizes most of the comprehension and production studies described in this volume. Because the second half of this chapter and many of the other chapters in this book are devoted to an elaboration of Level 2 functionalism, we will not spend our time here repeating discussions that the reader will encounter later.

The various caveats about cause and effect that we raised at Level 1 apply just as seriously at Level 2. First, there are many types of functional determinism...
Functionality and the Competition Model

function correlations in the course of language learning, insofar as they are able
to understand and formulate the relevant dimensions of meaning. In addition,
however, they are also capable of detecting certain distributional facts in the
sound stream even in the absence of an interpretation for those facts. They can
treat sound as an object, just like any perceptual object, and pick up recurring
patterns whether they understand them or not (MacWhinney, 1978, 1982, in
press; Bates, Bretherton, & Snyder, 1988). However, this process is greatly
enhanced when meaning is available (1) to motivate the child's attention to the
relevant facts and (2) to add critical patterned information of its own.

An adequate account of language development will necessarily involve a
mixture of function-driven and form-driven learning. In underscoring the role
that meaning and cognition play in language acquisition, many of us in the
functionalist camp have neglected the other half of the story, that is, the fact
that language learning is also a perceptual-motor problem. The emphasis on
function-driven learning and the relative neglect of form-driven learning in
some of our writings (Bates & MacWhinney, 1979, 1982) has resulted in
some misunderstandings that we would like to begin to rectify here. These
misunderstandings relate to (1) the relation between cues and functions, (2)
the role of cues in distributional learning, and (3) the role of reinforcement in
learning.

Cues and Functions. It has in fact been argued that the acquisition of “pure-
formal” linguistic patterns constitutes clear disconfirmation of the Competition
Model and any other functionalist theory of language learning (Chien & Lust,
1985; Hyams, 1986; Levy, 1983b, Meisel, 1986). We cite these articles only
because they make the analysis of the relevant issues so clear. Many other
papers based on similar analyses could also be cited. Levy (1983b) presents
an argument of this general type based on her analysis of the acquisition of
gender in Hebrew and other languages. However, it turns out that what Levy
calls “formal” patterns are really simple surface phonological cues and not deep
abstract concepts in a formalist universal grammar. As Zubin and Köpcke
(1981, 1986), Köpcke and Zubin (1983, 1984), and Taraban, McDonald, and
MacWhinney (in press) have shown, the surface phonological shapes of words
provide children with a great deal of information about category assignment.
MacWhinney (1978) emphasized the importance of such information in language
learning. However, it is misleading to call surface phonological cues “formal
patterns” since they do not force the child to abstract formal categories based
on the principles of universal grammar. Garden-variety principles of learning
and pattern detection may be sufficient.

Meisel (1986) presents a more general challenge to the functional view of
language acquisition. Observing aspects of grammatical development in a
French-German bilingual child, he demonstrated that there are large differences
at work in adult language use. Second, form and function stand in a many-to-
many relationship. Therefore, we are likely to obtain a distorted view of reality,
if we restrict ourselves to simple one-to-one manipulations in our experimental
designs. Third, we are usually dealing with qualitative rather than quantitative
effects when we manipulate form-function relations. We must be prepared
to accept “maybe” as an answer; in fact, we ought to have theories that are
powerful and precise enough to predict the proper degree of “maybe” in a given
experimental design. This may require us to move toward increasingly complex
designs, employing the powerful multivariate tools that have been developed
in other social sciences (e.g., economics) to model a multilayered reality.
Computer simulation may also become an important tool in the construction
of psycholinguistic theory, permitting researchers to manipulate more variables
than the “eyeball” can handle.

With these caveats in mind, Level 2 research can tell us a great deal about
the “ecological niche” of a given linguistic form. When and why do we use the
passive? What are the functions conveyed by aspect marking? However, this
kind of result does not require us to conclude that form–function mappings play
any role at all in acquisition, nor are we forced to conclude that form–function
correlations belong in a structural description of the language. Such claims
belong at the next two levels.

Level 3 – Language Acquisition

Level 3 functionalism claims that language acquisition is guided by form–
correlations. The child can only exploit form-function correlations in
acquisition if such correlations actually exist. Hence, claims at Level 3
presuppose the plausibility of claims at Level 2. However, a different kind of
evidence is needed to prove that children actually take advantage of these
correlational facts. After all, many of the complex discourse functions that
motivate adult grammar may be entirely opaque to the child. And it is
demonstrably possible for children to learn aspects of language that they do
not understand at all. For example, children acquire linguistic categories like
gender at a relatively young age, despite the fact that these categories have no
obvious semantic/pragmatic base. In the case of linguistic patterns like gender
marking, language acquisition becomes a problem of pattern detection that may
or may not require the application of innate linguistic knowledge. We suspect
that more general principles of pattern detection and distributional learning are
sufficient for the task – but that is an empirical question, perhaps the most
important question in language acquisition research.

In the Competition Model, language acquisition is characterized as cue-driven
distributional analysis. Primary among the various cues the child detects are
those involving form–function correlations. Children take advantage of form–
between the child’s two languages in the age at which forms with the “same” meaning are acquired (see also Slobin, 1973). He assumes, reasonably enough, that the bilingual child brings one and only one set of cognitive prerequisites to bear on learning her two languages. Hence, any difference in onset time between the two target grammars must be due to something other than cognitive content. Up to this point, we agree completely with Meisel’s reasoning. However, he then goes on to conclude that this “something” must be the innate and grammarspecific language acquisition device that Chomsky has foreseen in his writings on Universal Grammar. We see no reason to opt for this particular explanation of the developmental pattern. There may, for example, be differences between the child’s two target languages in the perceptual, motor and/or memory demands associated with “equivalent” linguistic forms. The problem with Meisel’s line of reasoning is that he is comparing a radically simplistic functionalist account with a not-yet-fully-specified autonomous syntactic account. Our work on the Competition Model suggests that the truth lies somewhere in between. Later on we will give examples of “cue cost” effects on learning in languages that are otherwise quite similar, lending support to our interpretation.

Hyams (1986) states the autonomous position in even stronger terms. In reviewing studies of grammatical development in Italian, Hyams notes that Italian children acquire many aspects of gender, number and person agreement at a very young age (see also Bates, 1976; Caselli & Pizzuto, 1988). Furthermore, the Italian child provides evidence on all the appropriate nouns, pronouns, adjectives and verbs – not only those that correspond to a semantic core such as “concrete object” or “physical action.” Given the generality of agreement phenomena in early Italian (i.e., the fact that they transcend concrete semantic categories), Hyams concludes that the child must control abstract notions like “noun phrase” and “verb phrase.” From this conclusion, he goes on to a stronger point: the child could not possibly learn such abstract categories, so they must be given in the innate and language-specific knowledge that children bring to bear on the acquisition problem. But of course this need not follow. Suppose (as we do) that children are simultaneously learning vertical correlations between form and function (i.e., semantic control), and horizontal correlations among surface forms (e.g., the markings that characterize agreement). A number of experimental and computational studies have shown that this kind of simultaneous learning is possible and perhaps desirable (e.g., Vallan & Coulson, 1988; St. John & McClelland, in press). Suppose also that this combined horizontal/vertical learning is carried out by a learning device that accepts partial information. In fact, we know that multiple fallible cues can lead to surprisingly robust learning (Taraban et al., in press). In short, Hyams is mistaken in her belief that imperfect correlations between form and meaning cannot in principle account for the acquisition of grammatical categories.

Chien and Lust (1985) devised an interesting experimental test of the functionalist claims of the Competition Model. Their test was based on the imitation of sentences in Chinese that show equi-NP deletion. They gave children Chinese sentences like “Xiaohua, father like father watch TV,” which means something like “As for Xiaohua, (her) father likes (her) father to watch TV.” Then they asked children to imitate these sentences and found that, beginning around age 3, they tended to delete the second mention of the subject noun “father” rather than the topic noun “Xiaohua.” They take these results as evidence that young children have an abstract category of “subject” that is not identified “with closely related and highly accessible semantic/pragmatic concepts” (p. 1773). While accepting the results of this study as reasonable, we believe that they provide little evidence against the functional analysis of the subject category. Both Bates and MacWhinney (1982) and MacWhinney (1977) note that there is no reason to think that an agent cannot coexist in the same sentence. It is true that the first argument of the verb “like” is not as agential as the first argument of a verb like “hit.” However, it still contains many elements of the cluster of meanings which we refer to as “agent,” including “perspective” and “source of volition.”

There is an important leap that Chien and Lust, Levy, Meisel, and Hyams all make in their arguments: children acquire regularities in their input that have no transparent semantic base; therefore, learning must be based on pure linguistic categories that are free of cognitive content. Since it is not clear where else linguistic categories can come from, they must be innate. Hence, the existence of distributional learning serves as evidence in favor of innateness! We think this logic is flawed and the conclusions are incorrect. But we can also understand why researchers might have been led to such reasoning, reacting to earlier more unidimensional formulations about the “critical” role of meaning in the acquisition of grammar (Bloom, 1974; Bowerman, 1973; Schlesinger, 1974).

Form-Driven Learning. In order to correct these misunderstandings, we would like to first clarify our own position on the role of form-driven learning in language acquisition. In our earlier papers we made a distinction between two types of distributional learning. The first type is what MacWhinney (1975, 1982) called “item-based frames.” A good example of an item-based frame is the English indefinite article. As Katz, Baker, and Macnamara (1974) discovered, very young children can use the absence of the article before a word to infer that word is probably a proper noun, rather than a common noun. If the children are told “show me Zav,” they tend to hand the experimenter a figure of a human. If they are told “show me a zav,” they tend to hand the experimenter a nonhuman object. MacWhinney (1982, 1985, in press) argued that children were using this syntactic frame to “abduce” the semantics of the nonce word “zav.” Working with the acquisition of gender and case marking in German,
MacWhinney (1978) noted that children use a similar mechanism to learn the genders of new nouns. Although this type of learning is clearly form-based, it relies on the existence of a clear syntactic/semantic relation between the operator word and the word whose semantics are being learned.

In our early papers on the functional bases of syntax (Bates & MacWhinney, 1979, 1982; MacWhinney, 1982) we noted the importance of this simple type of form-driven learning, while questioning the importance of a more extensive sort of distributional learning scheme proposed by Maratsos and his colleagues (Maratsos & Chalkley, 1980; Maratsos, 1982). Maratsos argued that children could extract abstract linguistic classes entirely on the basis of distributional evidence — without reference to correlations between form and meaning, and without reference to innate clues about the structure of linguistic categories. To illustrate, take the syntactic category “verb.” Verbs tend to occur within reliable paradigms: The same lexical item that ends in “-ing” in one context will also end in “-ed” in another context. These elements will also have certain positional privileges relative to members of other paradigms (e.g., nouns, prepositions). Maratsos argued that children are capable of detecting these regularities in their data, without mapping those regularities onto any a priori category. At first this pattern of correlations may be extended only to particular lexical items, perhaps only to those words that carried the information in the first place. However, with enough experience of this kind the child can learn to make systematic predictions: Given a nonsense word that ends in “-ing” in a particular context, the child can predict that the same nonsense word will end in “-ed” in another context. The child does not even have to know the meaning of the word to reach this conclusion. The prediction is based, instead, purely on distributional facts. At some point late in this process we can say that the child possesses the abstract category “verb.” Note that this is quite similar to the “item and arrangement” approach to grammatical categories, and the “substitution and contrast” approach to grammatical induction proposed by American structuralists before the onslaught of generative grammar (e.g., Harris, 1951; see Gazdar et al., 1985, for a modern version of phrase structure grammar).

To make a long story short, as both Maratsos and we developed our ideas further, the gap between our respective formulations continually narrowed. Maratsos increasingly stressed the role of meaning as providing cues used by the distributional learning mechanism. We became increasingly comfortable with a general distributional learning mechanism within which the various cues to which we had drawn attention could play a role. Recent developments in connectionist theory (Rumelhart & McClelland, 1986) have provided us with an algorithmic way of understanding how various cue types can interact within a general framework of distributional/cue-driven learning. Within this new framework, our earlier differences with Maratsos become nothing more than debates about the relative strengths of cues at different points in development.

Functionalism and the Competition Model

We see now that the important issue is not whether learning is driven by form or by function. The answer to that question is that it is driven in both ways. The basic scheme for learning is a meaning-driven distributional analysis. The important issue now is whether learning can take place without a priori semantic or syntactic categories. Here Maratsos agrees with us in questioning the quick recourse to nativist explanations.

Language acquisition is a perceptual-motor problem. The child is trying to extract patterns, islands of regularity that can be isolated within a noisy sound stream. This is the perceptual side of the problem, and it is subject to all the vicissitudes of perceptual learning and pattern recognition in any domain (much more on this below, and in other chapters throughout this volume). Once patterns have been isolated, the child will also try to reproduce them. This is the motor side of the problem, and it can only be understood within a much broader theory of motor learning. All of our claims about form-function mapping presuppose this perceptual-motor framework. Forms exist, and they must be perceived before any mapping can occur at all (Golinkoff et al., 1987). We do not see why this admission should lead one to question the functionalist enterprise. Indeed, within a highly interactive theory we should expect to find intimate causal relationships among phonetic, articulatory, grammatical and semantic facts.

What these various controversies have served to obscure is the fact that a matrix of correlated features will be acquired more easily if it includes meanings and uses that are transparent to the child. This is the essence of form-function mapping, the meaning-driven aspect of distributional learning. Functional correlations facilitate the acquisition of form for two reasons. First, regular mappings between form and function constitute a large part of the information that is available to the child. Second, communicative functions can drive the learner in a rather special way, by directing attention to regularities in the linguistic environment. The child is scanning the input for ways to convey interests and needs, trying to extract information that will help in predicting the behavior and attitudes of other people.

Reinforcement. Meaning contributes to the acquisition of grammar by providing both information and motivation. This emphasis on the motivation of grammar makes some of our colleagues nervous; they fear that it smacks of American behaviorism from James to Skinner. In fact, we reject the Skinnerian view that reward and punishment determine learning. We are more comfortable with the treatment of reinforcement offered by Tolman (1922). Tolman was viewed as an outlaw and a mystic by many of his contemporaries, because he argued that rats and other beings (e.g., people) operate on the basis of their expectations and their theory of the world. In Tolman’s cognitive learning theory, reinforcement facilitates learning by motivating the organism to attend
to regularities in the environment. The learning itself is driven by these attended regularities, and not by arousal, drive reduction, or any other direct and mechanistic definition of reinforcement (cf. Hull, 1943).

But even within this framework, we must insist that reinforcement means something very special in our species. As Piaget has argued, children are inherently motivated to learn, and will do so in the absence of any reinforcement other than the sheer joy of finding out how things work. In a socially motivated species like ours, this includes a drive to find out how other people work. Human children want to behave like others in their community and engage in imitation, with or without an extrinsic reward. They also want to know how other people think, and begin to build a "theory of mind" as early as the second year of life (Bretherton, McNew, & Beeghly-Smith, 1981). From this point of view, language is precious stuff indeed, providing the key to successful participation in a society. This is surely enough motivation even for such a daunting task as the acquisition of German gender, or acquisition of the mysteries of WH-movement.

Level 4 – Competence

It is possible to accept functionalist claims at Levels 1 through 3 without accepting the further claim that these form–function relations must be stated within the grammar, that is, within an abstract description of the rules and regularities that comprise the ideal speaker-listener's knowledge of his or her language. From a nativist perspective (e.g., Pinker, 1984, 1987), it is quite possible to argue that children use semantic bootstrapping to discover certain boundaries in their language; once those boundaries are discovered (e.g., once they know what nouns and verbs look like in this language), they can then instantiate a whole set of innate categories and principles. From this point on, the old semantic prototype becomes irrelevant (sloughed off like the tadpole's tail or the chrysalis from which the butterfly emerges).

We tend instead to favor an alternative view of semantic bootstrapping. Semantic/pragmatic prototypes remain at the core of all or most linguistic categories, showing up in a wide array of adult linguistic behaviors (Bates & MacWhinney, 1979 and 1982; see below for details). Does this mean that we espouse Level 4 functionalism, i.e., the claim that linguistic competence must be described in functional terms? We cannot answer that question at this point. No functionalist grammar has yet been formulated with enough internal detail to constitute a model of competence in any human language. Indeed, there is no grammatical theory of any sort that has yet met this challenge, and debates about what a competence model should look like are all currently based on nothing but fragmentary analyses. In lieu of a convincing competence model, we have contented ourselves with a characterization of competence to perform,

describing the kinds of representations that could underlie the Level 2 and Level 3 findings in our crosslinguistic work to date.

As we shall see in more detail below, the native speaker's latent knowledge can be described in terms of a network of weighted connections: correlations between forms and functions, correlations among forms themselves, and correlations or points of overlap between particular communicative functions or meanings. Borrowing a term from Marr (1982), who described aspects of visual perception in terms of a "2.5D sketch" (a two-dimensional description annotated with three-dimensional information), we could characterize the representational component of the Competition Model as Functionalist Level 3.5. It is a compromise that stands somewhere midway between a description of linguistic behavior and a characterization of competence.

This is our position on the matter right now. But let us put the same question a different way. Could the representational component of a performance grammar ever replace the Level 4 competence models offered by linguists? In considering this question, we are drawn to a useful set of contrasts offered by Pinker and Prince (1988) in their critique of the connectionist approach to language acquisition described in Rumelhart and McClelland (1987). Connectionist models provide a characterization of the "microstructure" of cognition, a subsymbolic level of analysis in which knowledge is represented in terms of many simple "on-off" units that are massively interconnected. This level of organization is intentionally "brainlike," in that it has certain properties in common with the organization of neural nets. To some extent, these systems can mimic the operation of traditional cognitive models, which are instead inspired by the organization of digital computers with discrete symbols and rules or procedures for manipulating those symbols. And yet it is not at all obvious where these "rules" and "symbols" are located in a connectionist model. At best they are emergent properties of lower-level units; at worst, they may play no role at all. Pinker and Prince describe three different relationships that could hold between a connectionist model and traditional rule-based accounts of cognition.

1. Implementational connectionism: This term refers to an approach in which there is an isomorphic relationship between the facts expressed in a connectionist format and at least some subset of the facts expressed in a traditional rule-based symbolic mode. Connectionist representations add nothing new (except to show how symbols and rules might be implemented in a neural net), and new facts could still be discovered at the symbolic level using traditional representations and traditional investigative techniques. This is plainly the most palatable version of connectionism for Pinker and Prince, since it poses no challenge to the modes of inquiry that they value.

2. Eliminative connectionism: This term refers to the belief that connectionist accounts completely eliminate the need for higher-order descriptions. Notions like "symbol" and "rule" are viewed as epiphenomena of facts at the subsymbolic level. They may provide a convenient shorthand for the description of finer-grained phenomena, but they have no explanatory value. This is the version of connectionism that Pinker and Prince find most objectionable (see also Podor & Pylyshyn, 1988).
3. Symbolic-revisionist connectionism: This phrase refers to an approach that recognizes the value of descriptions at both the symbolic and the sub-symbolic level. However, it is argued that the two levels are not equivalent. In particular, discoveries at the sub-symbolic level can constrain the class of psychologically and biologically plausible symbolic theories.

The limitations of a purely qualitative approach are particularly clear in some recent attempts to apply Chomsky’s theory of principles and parameters to language acquisition data (Chomsky, 1982 and 1986; Hyams, 1986; Roepke and Williams, 1987). Parameter theory represents the first serious attempt within generative grammar to provide a typology of natural languages. This typology is based on a small set of structural dimensions or “parameters” along which natural languages may vary; each parameter represents two or more possible forms that a supposedly innate principle can take. For example, the null subject parameter captures the fact that natural languages either do or do not permit subject ellipsis in a free-standing declarative sentence. Other parameters that have been proposed include order relations among the constituents of a phrase (e.g., “head first” or “head final”), and the presence or absence of a canonical order for the major constituents of a sentence (e.g., “configurational” versus “nonconfigurational”). Although the list of possible parameters is still open (i.e., some await discovery; others may prove not to deserve parametic status), it is generally believed that the list is finite and rather small; furthermore, most investigators hope that the parameters will prove to be orthogonal (i.e., setting along one parameter is independent of setting along another). Parameter theory is, then, an excellent example of a theory based entirely on discrete, qualitative principles. There is still some controversy within generative grammar circles concerning the “right” application of parameter theory to language learning (e.g., whether or not all children begin with a universal default setting, and/or whether parameters “mature”), but most interested theorists agree that a parameter setting process of some kind would greatly simplify the child’s learning problem. In order to figure out what kind of language s/he has to learn, the child simply has to listen for the right kind of positive evidence (e.g., the referentially empty pronoun in a sentence like “It is raining,” a phenomenon that supposedly occurs only in a language with obligatory subjects). When these “triggers” are encountered, the child can immediately draw a whole series of conclusions about the nature of her/his target language.

Unfortunately, there is a very bad fit between this theory and the messy empirical facts of language acquisition. In its original form, parameter theory requires all children to begin with the same default setting (i.e., the so-called subset principle). Within this framework, it is difficult to understand how Turkish, Italian and English children could look so different at the stage of first word combinations (or, for that matter, in the one-word stage – Slobin, 1985). Worse still, any serious version of parameter theory predicts abrupt transitions, developments in which a whole host of correlated phenomena appear simultaneously, with no turning back. For example, Hyams (1986) claims that English children begin with the default assumption that they are learning a pro-drop language; this explains why subject omission is possible in Stage I English. Children supposedly “switch” to the obligatory subject setting only after they have encountered the appropriate triggering evidence (e.g. referentially empty pronouns; modal auxiliaries that can be moved out of the verb phrase). At a global level, the evidence supports Hyams’s argument: some time before their fourth birthday, normal English children produce empty pronouns, acquire the auxiliary system, and stop omitting obligatory subjects. However, parameter setting can be distinguished empirically from garden-variety learning only if these events occur in a “special” time-locked fashion: triggers precede parameter setting, and once the parameter is set, the relevant phenomena are set in place simultaneously and very fast. These requirements are not met in the existing data for English. For example, Maratsos (1988) has shown that English children cease to omit the subject many months before they have acquired the auxiliary system. Conversely, Loeb and Leonard (1988) have shown that some English children fail to produce obligatory subjects even though they demonstrate productive control over “empty” subject pronouns (e.g., the “it” in “it is cold”). In short, the facts of English language learning support a model in which these diverse aspects of grammar are acquired gradually, in a potentially variable order. They “hang together” no more or less than any of the other morphosyntactic changes that take place between the ages of 20 and 40 months. (See Hyams, 1987, for a revised proposal, in which the triggering evidence for obligatory subjects in English comes from the nonuniformity of verb conjugation – a proposal which, in our view, suffers from even more serious empirical problems).

Finally, if we assume that English and Italian children have the “same” setting at Stage I, then we should expect them to omit the subject at a similar rate. However, our own data for English suggest that very young children produce the subject more than fifty percent of the time in their first word combinations; by contrast, very young Italian children may produce no lexical subjects at all for the first weeks or months of multiword speech (Bates, 1976). In other words, English and Italian children differ on the pro-drop parameter before the relevant parameter is set. These results are not compatible with Hyams’s theory, but they are compatible with a quantitative theory like ours, i.e., a theory in which children are sensitive to the statistical properties of their native language from the earliest stages of language learning. The Competition Model does have nonlinear mechanisms that allow for abrupt transitions when they occur, but it can also account for the much more general phenomenon of gradual, statistically driven change.

In the ongoing argument about the relationship between symbolic and sub-
symbolic structure, some of our colleagues have suggested that the relationship is analogous to the contrast between Newtonian and quantum mechanics. Deterministic Newtonian mechanics suffice for most macro-level problems (e.g., building a bridge or launching a missile); indeterministic quantum mechanics come into play only for micro-level problems (e.g., research on subatomic particles). This analogy is quite appealing to those who prefer an implementational form of connectionism, because it makes the subsymbolic approach irrelevant for most problems in cognitive science. But we think another analogy from physics may be more useful: the contrast between wave and particle theories of light. It is now generally agreed that this is a false opposition. Light displays both wave and particle properties, and both must be considered simultaneously for a complete account of the relevant physical and mathematical facts. In the same vein, a complete account of language may require simultaneous consideration of its symbolic and subsymbolic properties. This may be the most useful way of characterizing symbolic-revisionist connectionism.

Our approach to the relationship between rules and probabilistic representations corresponds best to the symbolic-revisionist approach. We recognize the value of linguistic descriptions that are cast in a traditional rule-based format. But we also believe that the underlying facts of human performance constrain the set of competence models that are likely to lead to real discoveries. This is the major reason why we cast our lot not with the "ruling party" of MIT generative grammar, but with the functionalist "loyal opposition."

The Competition Model

Before describing the basic tenets of the Competition Model, we should explain why we call this a model and not a theory. We are trying to account for a range of crosslinguistic data within a unified framework, but the framework itself cannot be disconfirmed in any single critical experiment. A model must instead be evaluated in terms of (1) its overall coherence, (2) its heuristic value in inspiring further research, and (3) its performance compared with competing accounts. Components of a model are tested one at a time, and if predictions fail, modifications are made and new concepts are introduced. Ideally, these new concepts should have an independent motivation (i.e., they should not be added merely to save the model), and they should lead to new discoveries in their own right. But the framework as a whole will be rejected only if (1) it loses coherence, weighted down by circularities and ad hoc assumptions, (2) it loses its heuristic value, and/or (3) a better account of the same phenomena comes along.

From these three points of view, the Competition Model has fared rather well so far. We began with a set of broad assumptions about the functional bases of language, the importance of crosslinguistic comparison, and the need to account for quantitative as well as qualitative facts about language structure and language use (cf. Bates, 1976; MacWhinney, 1978; Bates & MacWhinney, 1979, 1982). To account for our first demonstrations of crosslinguistic variation in performance, we appealed to a single quantitative principle: cue validity, that is, the information value of a given linguistic device as a cue to an underlying meaning or intention (MacWhinney et al., 1984). With this single principle, we were able to quantify with some precision the degree of informativeness that the "same" structure carries in two different languages. We knew that this minimalist approach was much too simple; no single principle, however powerful, can account for all the complexities of language processing. The real test of our model has come from the failures of cue validity, failures that have helped us to discover a range of interesting constraints. Many of these new constraints fall within a general category that we have called cue cost, that is, processing limitations that mute or augment the speaker-listener's ability to take advantage of different information types. Others have to do with different aspects of cue validity itself, including the contrast between overall cue validity and conflict validity - a discovery that helps us to understand how relatively rare phenomena in the language can play a role in acquisition and processing, providing an explanation for certain late developments that were difficult to explain in any other way. For present purposes, the important point is that these new constraints all have independent motivation. They have enriched the model, but they do not exist solely to keep the model going.

After many years of research, much of it summarized in this volume, the Competition Model is now competitive. It will undoubtedly be replaced by or merged with its competitors at some point down the road, when enough crosslinguistic data have accumulated to challenge the statements offered here. Indeed, many of the authors in the present volume have pointed out fundamental problems that do not yet have adequate solutions. But, if we are successful in inspiring more crosslinguistic research on language processing, then the model has served its purpose. We will describe the model and its supporting data in three sections: the structure of the system, processing dynamics, and acquisition.

The Structure of the System

Our understanding of the structure of the language processing system can be expressed in terms of five key concepts: two-level structure, direct mapping between these levels, cue validity, cue strength, and coalition between forms and functions.

Two-Level Structure

Only two levels of informational structure are specified a priori in this model: a functional level (where all the meanings and intentions to be expressed in an
utterance are represented) and a formal level (where all the surface forms or expressive devices available in the language are represented). We assume that mappings between these two levels are as direct as possible. Intervening layers will emerge only when they are essential for processing to take place in real time. These layers are the products of the meaning-driven distributional analysis described previously, and neither their content nor their structure is specified in advance of language learning.

The two-layer system of form–function mappings that we have described elsewhere constitutes a kind of perceptron (Rosenblatt, 1959; Minsky & Papert, 1969; Rumelhart & McClelland, 1986). It is known that perceptrons can learn any linearly separable combination of inputs and outputs, and that they cannot learn a host of other important second- or third-order relationships such as (for example) the relationship of “exclusive or.” In our work with Italian, Hungarian, and Dutch, we have encountered a variety of compound cues to sentence interpretation that seem to defy a linearly separable representation. This means that the two-layer approach that we have specified elsewhere has to be modified to permit the emergence of intervening layers.

To illustrate, consider the interaction of word order and contrastive stress in Italian (Bates et al., 1982; MacWhinney et al., 1984). For adult Italians, word order cues are essentially uninterpretable without additional information of some kind. As we have noted, Italians will attend preferentially to lexical and/or morphological cues if these are available; if the sentence is semantically and morphologically ambiguous, then pragmatic and/or prosodic cues come into play. However, word order and stress cues are never evaluated separately; instead, they are interpreted jointly in complex configurations, as follows (keeping in mind that these results are always probabilistic):

<table>
<thead>
<tr>
<th>Default Stress</th>
<th>First Noun Stressed</th>
<th>Second Noun Stressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVN</td>
<td>SVO</td>
<td>OVS</td>
</tr>
<tr>
<td>NNV</td>
<td>random</td>
<td>SOV</td>
</tr>
<tr>
<td>VNN</td>
<td>random</td>
<td>VOS</td>
</tr>
</tbody>
</table>

Notice that stress and word order do not combine in a linear fashion. The “same” prosodic cue can take on different meanings depending on its syntactic environment; conversely, the “same” word order information can take on new meaning, or even reverse its meaning, when stress cues are added. This nonlinear relationship is also supported by the reaction time data: stress speeds up performance on NNV and VNN, but slows down performance on canonical NVN strings. We suggest that the relationship between contrastive stress and word order in Italian constitutes a kind of exclusive-or problem. These cues simply do not “add up” in the orthogonal fashion required for learning in a two-layer perceptron. There is no single formula, no one weighting of the cue “stressed noun” that will yield the above pattern. The interpretation of stress depends entirely upon the word order environment in which it occurs. This is not the only possible configuration; we find equally complex interactions between word order and stress in German, but with different results. Kail (this volume) reports similar nonlinear interactions between word order and pronominal object clitics in French and Italian: the mere presence of a morphologically ambiguous clitic in an NVN suspends the usual SVO interpretation. This kind of cue interaction appears to be widespread in adult sentence comprehension.

There are a variety of ways in which we can complicate our simple two-layer model in order to represent such compound cues. Of these, the most attractive is the “hidden unit” model of Rumelhart, Hinton, and Williams (1986). The “back propagation” algorithm of Rumelhart et al. provides a robust learning rule of the type that Minsky and Papert thought was only possible for the two-layer perceptron. In our work on the acquisition of German case and gender marking, we (Taraban et al., in press) have found that systems utilizing hidden units can provide us with powerful ways of understanding developmental patterns. The acknowledgment that intervening layers can emerge through learning is a new feature of the model (cf. Bates & MacWhinney, 1979, 1982, 1987; MacWhinney, 1978, 1982). In this regard, it is interesting that the above order/stress configurations are acquired late by Italian children (Bates et al., 1984). They begin with a linear interpretation of each variable (e.g., treating stress as a cue to the agent role regardless of word order environment), and reach the above nonlinear combinations sometime after the age of five. Kail and Devescovi (personal communication) report a similar late onset for the nonlinear interactions between word order and object clitics.

**Direct Mapping**

In the Competition Model the mapping between form and function is stated as directly as possible — in two layers or (as described above) with intervening layers that are the minimum required for learning in a particular language. Since the principle of direct mapping has been a source of some confusion, it may be worth our while to underscore what this principle does and does not mean within the Competition Model.

First, direct mapping does not mean that relationships between form and function are necessarily one to one; indeed, we assume that one-to-one mappings are rare in natural languages, which are instead composed primarily of many-to-many relationships (polysemy, homophony, syncretism — see coalitions, below).

Second, direct mapping also does not refer to the direction in which information flows during sentence interpretation, that is, the question of serial versus parallel processing. A great deal of research in psycholinguistics from 1960 to 1980 focused on precisely this issue: which stages occur first in processing, and which stages can be skipped. In a typical modular model of sentence processing of the sort that underlay most psycholinguistic research in the 1960s
(Fodor, Bever, & Garrett, 1974, Massaro, 1975), processing was assumed to go through these stages: peripheral auditory processing, phonemic encoding, lexical look-up, phraseal construction, full sentence parsing, and then semantic interpretation. These processes were not supposed to interact and were supposed to occur in a serial pipelined fashion. In the 1970s, a large number of studies appeared demonstrating context effects on processing at both the phonological and syntactic levels, effects that made this serial pipelined flow of information difficult to defend (Warren & Warren, 1970; Marslen-Wilson, 1975).

Some researchers believed that the subsequent abandonment of serial processing signaled the end of a modular view to language processing, in favor of an approach in which meanings are mapped onto sound strings directly, without passing through separate processing units. However, alternative accounts are now available that salvage autonomy/modularity within a more parallel system (Swinney, 1979; Tunenhaus, Leiman, & Seidenberg, 1979; Tunenhaus, Carlson, & Seidenberg, 1985). McClelland’s Cascade Model (1979) is of this type. In a cascade model, given a complex Input A, the phonetic/phonological module goes to work in extracting recognizable packages of sound, and passes those patterns directly to the other modules in a rapid and constant cascade. After this point, several alternatives are possible (and they are not mutually exclusive). The grammatical module may go to work directly on that part of the phonetic input that it recognizes (e.g., closed-class morphemes), while the lexical module works on all the meaningful units that are passed up from the phonological processor; alternatively, the grammatical processor may have to wait just a short time for the lexical/semantic processor to recognize possible information about word class (e.g., nouns versus verbs) and/or to separate out closed- and open-class information, passing these relevant bits onto the grammar in another rapid and continuous cascade. Both the lexical and the grammatical processor may in turn pass their products on to a general information processor that assigns final interpretations; or they may employ some kind of a shared blackboard on which partial products are written until they have a coherent parse. The main point is that, in a cascade-type model, modular processes can go on in parallel; cascades may impose short time delays as information passes from one module to another, but these delays may be so brief that they defy measurement. For example, it may be that modularity obtains for a period of only one millisecond.

The principle of direct mapping is not proposed as an alternative to serial processing, but rather as an alternative to certain forms of modularity. Modularity postulates the computational independence of data sources and computational differences between the various modules that have evolved to deal with each data source. By contrast, the principle of direct mapping emphasizes (1) the mixed nature of input to the language processor, and (2) the homogeneity of processing across different data types. By mixed data types we mean that the language processor can make use of compound cues that cross traditional boundaries (e.g., segmental phonology, suprasegmental phonology, morphology, the lexicon, and positional frames). The above example of word order/stress configurations in Italian illustrates this claim. By homogeneous processing we mean that different sources of information (morphological, phonological, lexical, and syntactic) are processed in a similar fashion, via a common set of recognition and retrieval mechanisms. This may be regarded as a strong form of lexicalism. That is, the native speaker learns to map phrasal/prosodic configurations onto propositions, using exactly the same learning principles and representational mechanisms that he uses to map single words onto their meanings (see Bates, Breherton & Snyder, 1988, and Bates & Wulfeck, this volume).

**Cue Validity**

The major predictive construct in the Competition Model is cue validity. Following Brunswik (1956) and Gibson (1966), we argue that human beings possess psychological mechanisms that lead them to act in accordance with the validity or information value of cues in their ecology. Validity is an objective property of the cue itself, a property of the perceptual environment relative to some organismic state. Because of this, cue validity is not a circular notion; it can be measured directly in samples of spoken or written language, and used to derive predictions concerning language processing by adults and/or language acquisition by children.

MacWhinney (1978), MacWhinney, Pléh, & Bates (1985), and McDonald (1986, this volume) have analyzed cue validity into three components:

1. **Availability** represents the extent to which a cue is there when you need it. In the scheme of McDonald (1986), availability is best expressed numerically as the ratio of the cases in which the cue is available over the total number of cases in a task domain. For example, the availability of the cue of preverbal position is very high in English, but relatively low in Italian. This reflects the fact that subjects are frequently omitted in Italian, leaving many verbs in sentence-initial position.

2. **Reliability** represents the degree to which a cue leads to the correct interpretation when you count on it. Reliability can be expressed numerically as a ratio of the cases in which a cue leads to the correct conclusion over the number of cases in which it is available. For example, preverbal position is a highly reliable cue in English, where it is almost always assigned to the agent of a transitive action; it is a very unreliable cue in Italian (when it is available at all), since OV and SOV constructions are both possible and likely. (Overall cue validity is defined as the product of availability times reliability.) Given the reliability and availability calculations described above, this necessarily means that the cue validity of preverbal position is very high in English and very low in Italian—a fact that is reflected in the performance of English and Italian listeners in our sentence comprehension experiments.

3. **Conflict validity** is yet another way in which the validity of a cue can be measured. When estimating conflict validity, we look only at those cases where two or more cues conflict. These relatively infrequent competition situations are the denominator for calculations of conflict validity. For any given source of information, the conflict validity estimate is the number of competition situations in which that cue "wins," i.e., leads to a correct interpretation, divided by the number of competition...
situations in which that cue participates. Overall validity and conflict validity are two mathematically distinct and partially dissociable measures of information value.

Because they are theoretically and empirically distinct, these two validity estimates can account for different aspects of language processing and language acquisition.

The distinction between overall validity and conflict validity constitutes one of the most important discoveries in our fifteen years of crosslinguistic research. Overall validity can explain many phenomena, but some puzzling exceptions remain that can only be explained by considering the way that cues behave in conflict situations. As we shall see below, conflict validity has been particularly helpful in explaining certain late and/or U-shaped developments in children, and in explaining how relatively infrequent structures can influence adult performance. The two forms of cue validity also have a principled and very natural realization within a connectionist architecture, as by-products of learning with hidden units (Taraban et al., in press).

Cue Strength

To model the organism's knowledge about the validity of information, we postulate a subjective property of the organism—called cue strength. This is a quintessentially connectionist notion, referring to the probability or weight that the organism attaches to a given piece of information relative to some goal or meaning with which it is associated. In other words, cue strength is the weight on the connection between two units.

In our psycholinguistic instantiation of this idea, each link between a given surface form and an underlying function is given a weight or strength. With this kind of mechanism, no sharp line is drawn between probabilistic tendencies and deterministic rules. An obligatory relationship between form and function is nothing other than a connection whose strength approaches unity. This permits us to capture statistical differences between adult speakers of different languages, and it permits us to describe facts about language change (in language history and/or in language learning) in gradual and probabilistic terms; we are not forced to postulate a series of all-or-none decisions in which parameters are definitively set and rules are added or dropped.

We view development as the process whereby, under ideal conditions, the value of cue strength converges on the value of cue validity. As a result, the order of importance of cues to meaning for adult speakers ought to closely reflect cue validity estimates. This simple prediction has been confirmed repeatedly in our own crosslinguistic studies, and in work by our colleagues around the world. However, in many cases, estimates of the relative validity of each cue type have been based on the knowledge of informal language use in each language, and not on precise tabulations of cue validity (in texts). Therefore, these studies can only test ordinal predictions (e.g., the prediction that word order will be stronger than subject-verb agreement in English, with the opposite pattern occurring in Italian and German—MacWhinney, Bates, & Kliegl, 1984). In more recent studies (especially McDonald, 1984, this volume; Sokolov, this volume), estimates of cue validity have been made directly in sample texts. This permits the experimenter to work with a more precise interval scale, predicting not only the rank order of cues in each language but also the relative distance between ranks. Both approaches have met with considerable success, although interval scaling provides a much stronger and more interesting test of the model.

Table 1.1 summarizes the results of our experiments on sentence comprehension in adult native speakers of different languages. In almost all of these experiments, adult native speakers were presented with a series of simple transitive sentences composed of two concrete nouns and a transitive action verb. On each item, they are asked to decide which of the two nouns is the actor/subject (i.e., “Whodunit?”). Although the linguistic devices that are tested vary from one experiment to another, the materials always include some orthogonal set of competing and/or converging cues to sentence meaning: different levels of word order, animacy/reversibility, grammatical morphology (including case marking, subject-verb agreement, agreement between objects and clitic pronouns, and reflexive markers), prosodic contrasts, and in a few experiments, different forms of topicalization. The impact of cue validity on performance is assessed by evaluating the overall variance accounted for by each independent variable, as well as the extent to which each variable contributes to determining the “winner” in situations of competition and/or cooperation (including “conspiracies” of two or more weak cues against one strong one). McDonald and MacWhinney (this volume) have carried the estimation of cue strength considerably further, using maximum likelihood procedures to construct statistical models of cue interaction, models which provide a single goodness-of-fit statistic for the performance of individuals or groups in our sentence processing experiments.

Before describing the data obtained with this method, a few words are necessary to defend the method itself. In order to test the predictions of the Competition Model, we have to examine situations of both competition and convergence. This requires orthogonalization of factors that are usually confounded in natural language use (cf. Massaro, 1987). In some experiments, in some languages, the resulting stimuli contain a mix of grammatical and semigrannatical stimuli. For example, English experiments may include sentences like “The horse are hitting the cows”; sentences in Hungarian or Turkish may include a large number of ungrammatical sentences with no case contrast (e.g., with both nouns taking nominative case marking). It has been argued that these semigrannatical sentences may evoke abnormal processing strategies that are unrelated to the processes used in normal sentence comprehension.

Given the important role that this particular method has played in our crosslinguistic research, we have taken this criticism seriously. Our answer
Table 1.1. Order of importance of cues to actor assignment across languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Cue Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>SVO &gt; VOS, OSV &gt; Animacy, Agreement &gt; Stress, Topic</td>
</tr>
<tr>
<td>Adults:</td>
<td>SVO &gt; Animacy &gt; Agreement &gt; NNV, VNN, Stress</td>
</tr>
<tr>
<td>5 - 7</td>
<td>SVO &gt; Animacy &gt; Agreement &gt; NNV, VNN, Stress</td>
</tr>
<tr>
<td>Under 5</td>
<td>SVO &gt; Animacy &gt; Stress, SOV, VSO &gt; Agreement</td>
</tr>
<tr>
<td>Italian</td>
<td>SV Agreement &gt; Clitic Agreement &gt; Animacy &gt; SVO &gt; Stress, Topic</td>
</tr>
<tr>
<td>Adults:</td>
<td>SV Agreement &gt; Clitic Agreement &gt; Animacy &gt; SVO &gt; Stress, Topic</td>
</tr>
<tr>
<td>(NNV, VNN interpretable only in combination with stress, clitics)</td>
<td></td>
</tr>
<tr>
<td>Under 7</td>
<td>Animacy &gt; SVO &gt; SV Agreement &gt; Clitic Agreement &gt; SOV, VSO</td>
</tr>
<tr>
<td>(no interactions of NNV,VNN with stress, clitics)</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>SV Agreement &gt; Clitic Agreement &gt; Animacy &gt; SVO &gt; Stress</td>
</tr>
<tr>
<td>Adults:</td>
<td>SV Agreement &gt; Clitic Agreement &gt; Animacy &gt; SVO &gt; Stress</td>
</tr>
<tr>
<td>Under 6</td>
<td>SVO &gt; Animacy &gt; SOV, SOV (agreement not tested)</td>
</tr>
<tr>
<td>Spanish</td>
<td>Accusative preposition &gt; SV Agreement &gt; Clitic Agreement &gt; Word order</td>
</tr>
<tr>
<td>(animacy not tested)</td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>Case &gt; Agreement &gt; Animacy &gt; SOV, VSO, SVO</td>
</tr>
<tr>
<td>Adults:</td>
<td>Case &gt; Agreement &gt; Animacy &gt; SOV, VSO, SVO</td>
</tr>
<tr>
<td>Dutch</td>
<td>Case &gt; SVO &gt; Animacy</td>
</tr>
<tr>
<td>Adults:</td>
<td>Case &gt; SVO &gt; Animacy</td>
</tr>
<tr>
<td>Under 10</td>
<td>Case &gt; Word Order (animacy not tested)</td>
</tr>
</tbody>
</table>

Functionalism and the Competition Model

Table 1.1. (cont.)

<table>
<thead>
<tr>
<th>Language</th>
<th>Cue Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serbo-Croatian</td>
<td>Case &gt; Agreement &gt; Animacy &gt; SVO, VSO, SOV</td>
</tr>
<tr>
<td>Adults:</td>
<td>Case &gt; Agreement &gt; Animacy &gt; SVO, VSO, SOV</td>
</tr>
<tr>
<td>Under 5</td>
<td>Animacy &gt; Case &gt; SVO, VSO, SOV &gt; Agreement</td>
</tr>
<tr>
<td>Hungarian</td>
<td>Case &gt; SV &gt; Agreement &gt; SVO, SOV &gt; Animacy &gt; V - O agreement</td>
</tr>
<tr>
<td>Adults:</td>
<td>Case &gt; SV &gt; Agreement &gt; SVO, SOV &gt; Animacy &gt; V - O agreement</td>
</tr>
<tr>
<td>Under 2</td>
<td>Animacy &gt; Case &gt; SVO &gt; Stress (agreement not tested)</td>
</tr>
<tr>
<td>Turkish</td>
<td>Case &gt; Animacy &gt; Word Order</td>
</tr>
<tr>
<td>Adults:</td>
<td>Case &gt; Animacy &gt; Word Order</td>
</tr>
<tr>
<td>Under 2</td>
<td>Case &gt; Word Order (animacy not tested)</td>
</tr>
<tr>
<td>Hebrew</td>
<td>Case &gt; Agreement &gt; Order</td>
</tr>
<tr>
<td>Adults:</td>
<td>Case &gt; Agreement &gt; Order</td>
</tr>
<tr>
<td>Under 10</td>
<td>Case &gt; Order &gt; Agreement</td>
</tr>
<tr>
<td>Warlpiri</td>
<td>Case &gt; Animacy &gt; Order</td>
</tr>
<tr>
<td>Adults:</td>
<td>Case &gt; Animacy &gt; Order</td>
</tr>
<tr>
<td>Under 5</td>
<td>Animacy &gt; Case &gt; Order</td>
</tr>
<tr>
<td>Chinese</td>
<td>Case &gt; Animacy &gt; Order</td>
</tr>
<tr>
<td>Adults:</td>
<td>Case &gt; Animacy &gt; SVO</td>
</tr>
<tr>
<td>Japanese</td>
<td>Case &gt; Animacy &gt; SOV</td>
</tr>
<tr>
<td>Adults:</td>
<td>Case &gt; Animacy &gt; SOV</td>
</tr>
</tbody>
</table>

...of noun-phrases marked in the possessive. MacWhinney et al. first examined the relative contributions of case marking, word order, and semantics using stimuli that included ungrammatical case-ambiguous sentences like “The horse-nominal is kicking the cow-nominal.” These results were compared with those of a second experiment with a similar design, using possessive sentences in which case-ambiguous combinations are grammatical (e.g., “Your red-possessive is hitting my blue-possessive”). The main effects and interactions among variables were quite similar in these two experiments, suggesting that the presence of semigrammatical stimuli does not evoke strategies that are qualitatively different from those observed with normal sentences only. Finally, we have also conducted a number of studies varying methodological factors: randomized designs compared with blocked designs, with or without instructions that focus subjects' attention on their processing strategies, with
or without instructions forcing subjects to respond as quickly as possible. The
crosslinguistic results that we have obtained to date appear to be quite robust,
replicating across all these methodological conditions. Furthermore, most of
our crosslinguistic effects are extraordinarily large, accounting for 20–50% of
the experimental variance. Strategies that are this robust must reflect something
solid and real. We are convinced that our subjects respond in this laboratory
situation by assimilating both grammatical and ungrammatical sentences to the
representations and processes used in everyday life.

As shown in Table 1.1, many different hierarchies of cue strength have
now been observed across languages. The functional differences that we have
discovered so far suggest the need for a new functional taxonomy of languages,
a taxonomy that may not always correspond to the formal distinctions of
typological linguistics. Traditional taxonomies are based upon contrasts such as
the presence or absence of case marking, the pro-drop parameter (e.g., whether
or not a language can omit the subject of free-standing declarative sentences),
and the default word order used in a language (including the postulation of
certain “non-configurational” languages that may have no basic word order at
all). Our data present problems for all these classic dichotomies.

First, word order has proven to be a more important cue in German (a case-
marked language with rich inflectional morphology) than it is in Mandarin
Chinese (a language with very little grammatical morphology overall, and
no inflections of any kind to mark transitive relations). This pattern is not
compatible with a sharp dichotomy between case-marked and non-case-marked
languages.

Second, findings about French, Italian and English present problems for a
theory that emphasizes the pro-drop parameter (Rizzi, 1982). These are all
languages without case inflections, and in which SVO is the basic or default
word order. However, Italian permits considerably more word order variation,
and sentences without overt subjects are both permissible and frequent. French
and English both require overt subjects, including dummy subjects like the “it”
in “it is raining.” Furthermore, at least for sentences without clitic pronouns (see
below), deviations from canonical SVO are rare in French. Nevertheless, Kail
and her colleagues have shown that French is functionally more similar to Italian
than to English (Kail, this volume, McDonald and MacWhinney, this volume).
In particular, semantic and morphological cues are much more important than
word order for sentence interpretation by French adults. We will offer some
explanations for this pattern shortly; for now, the point is that these findings on
linguistic performance do not fit a formal classification based on the pro-drop
parameter.

Yet another surprise comes from experimental results in Warlpiri (Bavin
& Shopen, this volume). Warlpiri is a so-called nonconfigurational language
in which word order should in principle have no effect at all on sentence
processing. It is true that Warlpiri adults pay more attention to morphological
and/or semantic cues than to word order; however, they also have a significant
tendency to choose the first noun in NVN constructions, particularly when
no other contrasts are available. In this respect, Warlpiri is not functionally
very different from Italian or French. In fact, English is the one truly exotic
language – the only language we have studied to date in which word order is
the most important determinant of sentence meaning across all tested morphological
and semantic/pragmatic conditions. Insofar as ninety percent or more of the
existing research in psycholinguistics is based on the performance of English
subjects, this is a worrisome finding indeed.

Coalitions and Prototypes

What we have described so far is the relationship between a single cue (e.g.,
preverbal position) and a single function or meaning (e.g., the actor role). This
kind of relationship is depicted in Figure 1.1a. In natural languages, such
mappings of a single form onto a single function are quite rare. Rather,
languages make extensive use of polysemy, thereby producing grammatical
systems in which the same form can map onto several functions, while the
same function can map onto several forms. For example, preverbal position is
associated not only with the actor role, but also with the potentially disessachable
notion of discourse topic. These two functions, in turn, are also associated
with other surface forms – for example, agreement with the verb in person and
number. Taken together, these many-to-many mappings comprise a series of
subsystems which we refer to as coalitions. A very simple coalition of this sort
is illustrated in Figure 1.1b.

The paradigm case of a many-to-many mapping, exploited in most of our
experimental work to date, is the mapping underlying the category of “sentence
subject.” This mapping is based heavily on the activation of concepts such as
“agency” and “topicality.” By labeling a node in a figure such as Figure 1.1b
as expressing “agency” we may seem to be indicating that we view “agency”
as an impenetrable unitary concept. In fact, we realize that concepts such as
“agency” have a fine-grained and distributed internal structure of their own.
In particular, (agency) is an emergent grouping of such underlying features as
humaneness, animacy, intentionality, and motion. None of these features may be
either necessary nor sufficient to activate the agent role (Hopper & Thompson,
1980). The particulars of language processing and development require that we
dig continually deeper into the microstructure of these grammatical categories.
It is not enough to focus simply on the connections between functions and
forms. In order to fully understand the structure of coalitions like the “subject”
coalition, we need to also look at the correlations and connections between
functions and between forms. We will call the connections between forms and

Fig. 1.1. Three ways of viewing form-function mapping.

functions and the connections between forms or between functions. The vertical connections indicate that all forms on a level are potentially connected with double horizontal lines. At the level of horizontal form–form correlations, we can specify the probability that any given pair of forms will cooccur: the probability that a noun mapped onto preverbal position will also agree with the verb in person and number, receive nominative case, occur with definite determiners, and so forth. These correlations among forms are particularly likely to vary from one language to another. For example, in English, preverbal position and subject–verb agreement tend to be assigned as blocks, in Italian the correlation between these two forms is considerably lower. Subjects are more likely to be definite in both English and Italian, but definiteness is much more highly correlated with the other subject devices in Italian. These distributional facts comprise an important part of the input to small children, and as we have noted earlier, it is possible in principle for children to detect such form–form correlations without necessarily detecting their associated meanings.

At the level of horizontal function–function correlations, we specify the probability that any given pair of functions will cooccur: the probability that an entity serving as the topic for a particular comment is also the agent of a transitive action, similar to the speaker with a similar perspective on the situation, and so forth. Languages do not choose these functional correlations randomly. Instead, these correlations reflect natural tendencies in human experience that tend to recur across natural languages. For example, the correlation between agent and topic reflects a high-probability tendency for speakers in all cultures to talk about their own activities and the actions of those who are near and dear to them. Hence, horizontal correlations at the functional level are considerably less likely to vary from one language to another.

Between these respective “horizontal” levels of form and function, there are also a set of “vertical” weightings that represent the probability that an individual function will be mapped with a particular form. With these connections we can, for example, capture the following differences between Italian and English:

1. the probability that the topic of a discourse will be assigned preverbal position (more common in Italian),
2. the probability that the agent will be assigned preverbal position (more common in English), and
3. the probability that the agent/topic will be marked with definite determiners (more common in Italian).

Again, note that many of these language differences are statistical rather than obligatory in nature. Although such performance facts typically fall outside the purview of a grammar based on rules, they are exactly the kind of fact that we are trying to capture in our model.

Finally, as we noted earlier, it may be necessary to postulate more than two layers in order to capture nonorthogonal relationships like the exclusive-OR in order to solve the perceptron problem. We have argued that such intervening layers may develop through experience, in neutral, “hidden units” such as those described by Rumelhart, Hinton and Williams (1986) or Takahara et al. in press; or also MacWhinney et al., this volume). The organization of stress and word order in Italian is a case in point, as is the relationship between pronominal object clitics and word order in Italian and French. Figure 1.2 illustrates a hypothetical fragment of the subject system in Italian, demonstrating how word order and stress cues may be organized together into blocks (i.e., “grammogens”), units which in turn serve as the input to assignment of the actor role.
This can occur, for example, when we need to topicalize "the ball" even though "John" did the hitting. In such cases the grammar has to determine which of the two elements should "win" access to devices like preverbal positioning and verb agreement. We have classified solutions to this problem into two basic types: "compromise" and "divide the spoils."

A typical compromise solution provided by both English and Italian, is selection of the passive (Bates & Devescovi, this volume): the patient/topic, "ball" wins access to the major subject devices, but the agent is placed in a special "by clause" that signals its continued semantic role. In terms of prototype theory, this is the kind of "hedging" and category-mixing that often occurs when categorization decisions have to be made for peripheral members. In a sense, it is a sentence-level analog to word-level expressions like "An ostrich really is a bird," designed to mark explicitly the peripheral status of a category assignment.

Topicalization is an illustration of a typical "divide the spoils" solution. In topicalization in Italian, preverbal position is assigned to the topicalized patient, but verb agreement is still assigned to the agent. In other words, the set of surface devices comprising "subject" is split up and assigned to separate elements. This kind of splitting does occur in English (e.g., in informal constructions like "Now that I'd really love to see"), but such forms are much less frequent and considerably more constrained in English than they are in Italian. The highly correlated subject devices in English tend to be assigned as a block whereas the lower correlations among the same devices in Italian permit the coalition to be split up more easily in nonprototypical situations.

The point is that a series of compromises are made in both sentence comprehension and sentence production. The ideal situation does not always hold. In fact, the fully prototypical instance of a category such as "subject" may actually be fairly rare (like the "ideal member" that is extracted but never taught in studies of artificial category learning such as Posner & Keele, 1968). This is possible because our knowledge of a "prototypic subject" is the emergent property of a great many weightings between individual forms and functions. It is the result of a lifetime of distributional analysis, and not a template derived from any single instance of grammatical learning. To understand how such a system works in real time, we turn to the processing dynamics of the Competition Model.

Processing Dynamics

The model assumes dynamic control of the mapping of form onto function in comprehension, and the mapping of function onto form in production. This mapping is understood to be governed by a system of parallel activation with strength-based conflict resolution much like that found in word-level processing.
models such as Thibadeau, Just, and Carpenter (1982) or McClelland and Rumelhart (1981). The Competition Model extends these word-based models to the sentential level to account for assignment of grammatical roles and other parsing decisions in comprehension.

To illustrate what we mean by this, consider the word-recognition system modeled by McClelland and Rumelhart, to account for experimental findings by Glushko (1979). People know how to pronounce a "new" or nonexistent string like MAVE, even though they have never seen it before. Usually they will pronounce it to rhyme with CAVE, but occasionally they will pronounce it to rhyme with HAVE. Glushko suggested that speakers make their decisions not by applying an abstract set of phonological rules, but by a process of analogy. Specifically, when the letter string MAVE appears, all of the existing words that overlap partially with this nonsense string are activated simultaneously. Each of the real word candidates has a basic activation level reflecting (at least in part) its baseline frequency in the language. The decision about how to pronounce this nonsense input, and the time taken to reach that decision, emerge out of the competition among all of these partially overlapping "demon": HAVE, as well as CAVE, SAVE, RAVE, MANE, MATE, CANE, and so forth. In the case of MAVE, the high-frequency candidate HAVE does occasionally win out, but it is usually overwhelmed by the greater number of word candidates with a long vowel. In general, decisions are a combined product of the number of different types in the competition pool, and the activation-weights associated with each type.

A phrase-structure analog to this process can account for a number of robust phenomena in our crosslinguistic sentence comprehension data. As we have noted, there are massive differences between languages in the way that listeners respond to word order variations. Most of these differences follow directly from calculations of cue validity, as described above. For example, Italians make greater use of both semantic contrasts and subject-verb agreement than their English counterparts; English listeners make greater use of word order than any other cue. In addition, however, we have also uncovered some interesting new information about the specific strategies used in different languages to deal with word order configurations. For example, English listeners not only have an overwhelming SVO strategy to deal with NVN sentences; they also have very strong and reliable VOS and OSV strategies to deal with the two respective noncanonical word order types. For every language that we have looked at, we have discovered some kind of word order bias that do not follow in any straightforward or obvious way from known facts about basic word order types in the language. That is, the word order bias cannot be attributed to any single phrase structure "template."

However, we discovered that these word order and/or order/stress biases can be accounted for by the parallel activation of all the partially overlapping word types in the language. That is, if a phrase structure analog to the MAVE example is going on, we should get exactly the results that we have obtained in all of the languages examined so far. Suppose, for example, that we line up all the possible phrase structures for talking about "John hit the ball" in English. Ignoring morphology, looking only at the "islands" of constituent ordering that we see (e.g., treating a cleft sentence like "The one who hit the ball was John" as a VOS), it is clear that SVO is the statistically predominant ordering in English, followed by OSV and VOS. Carrying out a similar exercise in Italian, we find a much weaker bias toward VSO, with essentially random probabilities for SOV versus OSV and for VOS versus VSO — unless we take stress into account. For French, alternative word orders are possible only in the presence of a complex set of clitic markers. And yet, in experiments by Kail (this volume), French listeners behave almost exactly like Italians, distorting word order and making decisions primarily on the basis of semantic and morphological cues — even though Kail did not provide clitics to "release" word order variation. The behavior of French adults with respect to phrase structure possibilities makes sense only if they are engaged in a process that involves a competition among partially as well as completely overlapping phrase structure candidates. In other words, the cliticized word order variations of French are somehow "echoing" in the listener's mind and influencing a decision about how to treat a novel sentence stimulus with clitic markers.

To what extent is a parallel system of this sort affected by the relative frequency of phrase structure types? A complete answer to this question must await more detailed calculations of the statistical structure of informal language use than we have available at this time. However, it is worth pointing out that absolute frequency may be of limited importance in the kind of nonlinear system of activation and competition that we have postulated here. As we will lay out in more detail in the section on acquisition below, conflict validity may be more important for adult speakers than overall cue validity. For example, our native informants assure us that word order variations are considerably less frequent in French than they are in Italian — even within the kinds of cliticized structures described by Kail (this volume). If performance were driven entirely by frequency, we should expect more differences between Italian and French adults than we have observed so far. The same should be true even if performance were based on the more complex notion of overall cue validity, because overall validity is at least a partial function of the availability (hence, frequency) of a cue. However, if adults are strongly affected by conflict validity (i.e., the trustworthiness of a cue in competition situations), then the absolute frequency of certain word order variations may be far less important than the clarity with which a prediction based on word order is disconfirmed in a reliable subset of sentence types.
Ongoing Updating

The processes of comprehension and production have to unfold in real time. There are constraints on the order in which cues are made available, and the order in which interpretations based on those cues are activated and selected.

In order to control the interaction of the various cues that impinge on sentence comprehension, we believe that the parsing system engages in an ongoing updating of assignments of nouns to case roles. For example, when parsing a sentence such as “The dogs are chasing the cat,” the assignment of “dogs” as the agent is first promoted by its appearance as the initial noun. Then the fact that “are chasing” agrees with “dogs” in number further supports this assignment. Finally, when the singular noun cat appears postverbally, its binding to the object case role further supports the candidacy of “dogs” as the agent. Thus, at each point in sentence processing the mapping from the lexical item “dogs” to the agent role is updated. In this particular case, each updating increased the strength of this assignment. In other cases — particularly in languages that permit a great deal of word order variation — assignments may wax and wane in strength across the course of sentence processing. As a result, the competition pool expands and contracts across the course of sentence understanding, until a solution is reached. Because each language designs cues to permit ongoing updating, and because competing parses are maintained in parallel, the need for backtracking is minimized (Marcus, 1980).

The processes of competition and ongoing updating must take a somewhat different form in sentence production (Bates & Devescovi, this volume). In comprehension, the listener is not “in the driver’s seat.” Unsure of the speaker’s intended meaning, the listener must be prepared to reject an early interpretation in favor of one that proves to be more appropriate later on. Hence, the degradation of “losers” may be gradual, permitting a less probable interpretation to come back to life if it is needed (this is why listeners are not compelled to follow the same garden path over and over again; they can learn from experience when forced to backtrack). In production, the speaker knows (more or less) what meaning is intended, making possible early commitments to a particular set of form–function mappings. Furthermore, if the utterance is going to come out right (i.e., as a relatively well-formed structure in the language), commitments will have to be made rather quickly — the sooner the better. There are only so many ways to express the same thought in a given language, and the pool of alternatives shrinks rapidly from the first word that the speaker selects (Bock, 1982). Although we adopt the view that comprehension and production make use of the same system of representations, we acknowledge that the real-time exigencies of processing may be quite different — a point that is important for language acquisition as well as adult language use and may prove to be particularly important in accounting for language breakdown in brain-damaged adults (Bates & Wulfeck, this volume).

A serious test of these claims about ongoing updating will require a move toward more on-line methods than we have adopted to date, particularly in the domain of receptive processing. In almost all of our sentence comprehension work, decisions are made at the end of the sentence. We have examined end-of-sentence reaction time data in several experiments; in general, the results support the predictions of the model (competition items take longer to resolve, convergence decreases reaction time). But there are some interesting exceptions that suggest a closer look at moment-to-moment decision making.

MacWhinney and Páth (this volume) examined aspects of sentence interpretation in Hungarian, using stimuli in which a competition between cues occurs either early or late in the sentence. They report that reaction times are longer, even though decisions are the same, when the competition occurs relatively late. This result is compatible with the “rich get richer” principle in parallel distributed processing models like the one proposed by McClelland and Rumelhart (1981): The earlier a cue occurs, the sooner it can begin to inhibit its competitors; hence, more activation is required to “beat down” an interpretation that gets off the ground relatively early.

Confirming cues do not necessarily speed the processing of a sentence. Kail (this volume) has shown that the presence of a clitic pronoun can actually slow processing in both French and Spanish even though it eventually aids in the interpretation of the sentence. Devescovi (in progress) reports similar findings for clitics in Italian. An analogous slowing with the presence of a subject–verb agreement cue has been noted by Kail for Spanish, and in our own unpublished data on reaction times in English. However, no slowdown in processing has been noted for use of the agreement cue in French or Italian; in these languages, a subject–verb agreement contrast always results in faster reaction times.

How do we put these contradictory results together? The Competition Model clearly needs to be enriched with on-line information about the way listeners distribute their attention and make predictions across the course of sentence processing. It appears that listeners can often make up their mind more quickly on the basis of incomplete information. Having to pay attention to additional information can slow down processing unless (a) use of that cue is entirely automatic, and/or (b) the cue is expected and eagerly awaited.

These processing effects appear to interact with something that Kail (this volume) and Sridhar (this volume) both call “canonicity.” Canonical, default word orders profit least from additional morphological information, a result that is magnified by (a) the importance of the word order cue in a given language, (b) the relative cue validity of the morphological cue itself, and (c) the probability that these word order and morphological cues should occur together. We have to conclude that cues are not simply added up in a bottom-up fashion. Reaction
times can be slowed by cues that the listener typically does not use (e.g., subject-verb agreement in English), and by unexpected configurations of cues (e.g., clitic pronouns and/or contrastive stress in sentences with default word order – turning a configuration that is usually pragmatically neutral into a marked variant). This slowing occurs even if the sentence interpretation is still a product of combined cue strengths. Effects of this kind require a better understanding of the dynamics of "horizontal" interaction between cues, including top-down inhibitory effects that vary in strength over time.

We are currently turning to a variety of on-line methodologies to investigate these and other aspects of real-time processing in the Competition Model. Kilborn (1987) devised an on-line version of our basic interpretation task, permitting subjects to make decisions as quickly as they can, before the end of the sentence if they prefer. His results for German and English are quite compatible with results obtained off-line: German subjects make their decisions on the basis of morphological information, with assignments occurring as soon as all the relevant information is available; English subjects base their decisions almost entirely on word order, and make those decisions as soon as the word order configuration itself is clear (e.g., immediately after the verb in NVN strings). Kutas and Bates (1988) have investigated event-related brain potentials to semantic and grammatical violations, in English monolinguals and in Spanish-English bilinguals. Their results provide further on-line support for the claim that Spanish and English listeners attend to different aspects of the same sentence input. English listeners show little or no reaction to violations of grammatical morphology (Kutas & Hillyard, 1983). By contrast, Spanish-English bilinguals show large cortical responses to morphological violations, in both of their languages (although the effects appear to be stronger in Spanish).

Other on-line techniques applied by members of our research team include word monitoring (Kilborn, 1987), error detection (Wulfeck, 1987), and a probe-identity task that can be used to test the degree to which a given noun is acceptable as the referent for pronouns and other anaphors (MacDonald & MacWhinney, in preparation). By focusing on the temporal microstructure of language processing, we hope to provide a stronger test of the Competition Model, discovering new constraints on cue validity that are not apparent from off-line methods. This brings us to the issue of processing costs.

Functionalism and the Competition Model

The most obvious limitation on processing is the low-level limitation dictated by the perceptibility of the stimuli. Consider, for example, the example of subject-verb agreement in spoken French, as it is displayed by the contrast between these two sentences:

Elle mange. She eats.
Elles mangent. They+PL eat.PL

For most French verbs, in most conjugations, the clear-cut written difference between the singular and plural form of the verb is entirely inaudible (i.e., mange and mangent are pronounced the same way). Even though the agreement contrast is distributed quite faithfully through written texts of French, it is an imperceptible cue in the oral language. Cue validity means very little if a cue cannot be heard at all.

MacWhinney et al. (1985) have shown that less drastic differences in the perceptibility of cues can also have a significant impact on the way those cues are used in sentence interpretation. In Hungarian, accusative case marking is a highly available and reliable cue. However, sometimes the accusative is hard to detect. When the accusative follows a dental or alveolar consonant, as in motus, it is fairly difficult to identify with certainty. However, when the stem ends with the letter ą as in pipa, the final ą lengthens and changes its height. The distinction between the nominative (pipa) and the accusative (pipa) is quite clear and easily detected. In sentence interpretation experiments with adults and children, this difference in the perceptibility of cues interacted with cue validity, in determining the probability that a Hungarian listener will rely on case information. If case competed with other cues (e.g., word order and semantics), the strong vowel form of the suffix quite clearly "won" the competition; but if the weaker consonant form was involved, listeners would often (though not always) take the "conspiracy" of order and semantics into consideration in making their decision. In such cases, they would often mishear the case marking in a way that matched the default interpretation of the syntactic
structure. MacWhinney et al. suggest that this is a morphological version of the "phoneme restoration effect" of Warren and Warren (1976), for example, the tendency to hear a stimulus like "(cough)-cel" as "wheel" or "meal," depending on the sentence context. Another way of putting it is that a lifetime of not being sure whether a case contrast is there or not has led listeners to "distrust" that cue, even in those instances when it is perceived.

"In a sense, perceivability and availability are similar notions: a cue is not available from the organism's point of view if it cannot be perceived. Remember, however, that we are separating cue validity (an objective property of the linguistic environment) from cue strength (a subjective property of the organism). This move is necessary in order to keep validity from turning into a circular notion. We calculate cue validity from samples of real language use (oral or written). There is nothing in the text itself that tells us which cues are easiest to perceive and to what extent. Indeed, estimates of perceivability may vary markedly from one listener to another (children versus adults, young adults versus older adults with hearing problems - see Bates & Wulfeck, this volume). Some of our ongoing work on "listening through noise" is designed to provide more independent motivation for perceivability and other aspects of cue cost (e.g., Kilborn, 1987).

Assignability

The second "cue cost" factor is what we call "assignability." The principle of ongoing updating outlined above is based on the assumption that the processing system tries to assign cues to meanings as rapidly as possible, integrating each fragment of sound and meaning into one or more larger structures that are compatible with all the information obtained up to that point (Bransford, Barclay, & Franks, 1972; Kintsch, 1974; Van Dijk & Kintsch, 1983). The amount of memory required for integration is relatively low when attachments between units can be made locally (Frazer, Clifton, & Randall, 1983; Ford, Bresnan, & Kaplan, 1982; Frazer, 1985; Small, 1980). Memory load increases when integration must be delayed until more information is received. Cues to sentence meaning can be ranked along a dimension called "assignability," referring to the amount of material that must be held in memory before a meaning assignment can be made.

Case morphology and agreement morphology provide an interesting contrast along this particular dimension of cue cost. In languages like Turkish or Hungarian, (with unambiguously case systems), a case suffix can lead to the assignment of a semantic role as soon as it has been recognized and integrated with its (usually adjacent) noun stem. Ammon and Soblin (1979) and Johnston and Soblin (1979) refer to these as "local cues." Many forms of agreement morphology are, instead, "topological cues." For example, Italians cannot make

Functionality and the Competition Model

The competition in the third person until they have heard the verb and all its associated nouns. If the verb is third person singular, but both the nouns are also third person singular, then the agreement cue is completely ambiguous and provides no information. The same is true for agreement between the object and a clitic pronoun.

If the processing system is under stress, and/or if the processor has limited auditory storage, topological cues or cues that are low in assignability may become so costly to handle that they are abandoned despite their information value. Bates and Wulfeck (this volume) have provided some evidence to suggest that case cues are better preserved than subject-verb agreement in aphasic adults, presumably because of the extra processing load imposed by agreement cues that span several different elements of the sentence. Some implications of assignability for language acquisition by children are discussed in the next section.

Acquisition

Cue-Driven Learning

Within our model, language learning is viewed as a process of acquiring coalitions of form-function mappings, and adjusting the weight of each mapping until it provides an optimal fit to the processing environment. This is quite similar to the process that Gibson (1969) describes as "detection of invariance" and/or "acquired distinctiveness of cues." We have offered some strong predictions about language acquisition across natural languages, claiming that cue validity will determine the order in which grammatical devices are acquired. In other words, children are sensitive from the beginning to the information value of particular perceptual patterns, and will go to work first on those forms that promise a greater "payoff." With some interesting exceptions which we will discuss in more detail below, the data summarized in Table 1.1 provide strong evidence for the role of cue validity in language acquisition.

The same results also disconfirm many putative universals of acquisition. For example, several authors have previously claimed that children will acquire semantic cues to meaning before they understand the word order principles of their language (Strohner & Nelson, 1974). Our data suggest that this proposed universal is much too simple. In our own English-speaking subjects, SVO word order is the first cue to have a significant effect on sentence interpretation (in children as young as 28 months), and canonical order remains the strongest cue to meaning from 2 through 82 years of age (remaining strong even in so-called agrammatic aphasic patients - see Bates & Wulfeck, this volume). By contrast, semantic cues are far stronger than word order at every stage of development in Italian, from 2 through adulthood. Figure 1.3, which is based on data in Bates et al. (1984), plots the percent of experimental variance accounted for by the
respectively main effects of word order and animacy, in each language, at each age.

Another putative universal involves the claim that children will acquire word order cues before they master grammatical morphology (Pinker, 1982). This is certainly true for children acquiring English. However, as Slobin and Bever (1982) have shown, Turkish children have attained adultlike use of case morphology as a cue to sentence meaning by the age of 2; they show no sensitivity at all to canonical SOV word order until 4–5 years of age, and even then the effect remains (as it is for adults) extremely small. Case suffixes are established somewhat later for Serbo-Croatian and Hungarian children, compared with their Turkish counterparts. However, there is no point in development in which canonical word order is more important than case in any of these languages.

A similar analysis can be applied to the use of morphology versus semantics. Figure 1.4 compares the developmental course of these two cue types in Serbo-Croatian and Hungarian (where the amount of experimental variance accounted for by each type of cue is plotted over age). In Serbo-Croatian, where case contrasts are frequently ambiguous, children begin with a bias toward semantic information. Semantically based interpretations drop off and case becomes dominant by 5 years of age. In Hungarian, where case contrasts are very rarely ambiguous, semantic and morphological cues are roughly equivalent in strength at age 3, and case is firmly established as the dominant source of information by age 4. Hence there is a one-year difference between these two languages in the relative strength of case cues. This difference is quite compatible with the differential validity of case in the two languages.

In short, there appears to be no single, universal schedule of development in sentence comprehension. Children may begin by attending to word order, semantics, and/or grammatical morphology, depending on the relative validity of those cues in their native language. We have, however, uncovered a number of surprisingly late developments in sentence interpretation, delays that do not follow in a straightforward fashion from overall cue validity. Many of these late phenomena can be explained by distinguishing between overall cue validity and conflict validity.

**Conflict Validity and Acquisition**

McDonald (1986, this volume) has shown that the course of acquisition cannot be understood without recognizing the importance of the concept of “conflict validity.” The contribution of conflict validity to acquisition represents one of the most important discoveries in our years of joint research. In the earliest stages of learning, children appear to respond primarily on the basis of overall
availability (corresponding roughly to frequency, albeit frequency from the point of view of those meanings that are of interest to the child). After this initial phase, across the first years of language learning, development appears to be controlled primarily by reliability. However, once the bulk of language-specific learning is complete, children begin to fine-tune the system (Karmiloff-Smith & Planck, 1980; Karmiloff-Smith, 1986; McDonald, 1986). They begin to take note of relatively rare situations in which two or more cues compete, and reset form-function mappings to favor those cues that win in such conflict situations. Hence, conflict validity dominates the last phases of language learning, and may in some cases result in U-shaped functions and radical restructuring of the system as a whole (Bowerman, 1982, 1983, 1987).

It may be helpful to think of language learning in terms of the "80/20 rule." This rule holds that, in solving real-life problems, one often gets 80 percent of the gain in 20 percent of the time. This is also true in language. Using availability and reliability as guides, the child can correctly interpret the overwhelming majority of the sentences of his language. However, there often remains a residual set of sentences where full learning depends on attention to particular conflicts between cases which are only rarely encountered. There is evidence that certain later developments in language relate to the problems involved in acquiring a final set of dominance relations. McDonald (1986) has observed a dominance effect in the acquisition of Dutch. In sentences with a pronominal direct object, Dutch allows OVS word order to produce the Dutch equivalent of "Him saw I." For adult subjects, the usually weaker case cue dominates over the strong word order cue in sentences like these. The same relationships present serious problems for Dutch children as old as 8 years. Another example comes from Warlpiri (Bavin & Shopen, this volume), a language in which children must acquire a variety of dominance relations in order to disambiguate case markers. Again, there is evidence here that children find dominance relations hard to learn. Such late developments appear to occur somewhere between 6 years of age and adulthood. This "second phase" of language acquisition is the period in which interpretations based on overall cue validity give way to interpretations based on conflict validity.

This insight can help us to understand a result for French that we initially found very puzzling (Kail, this volume). As we noted earlier, French adults behave very much like Italians in our sentence interpretation experiments: a heavy reliance on grammatical morphology, followed by semantics, with relatively little use of canonical or noncanonical word order strategies. Nevertheless, French children seem to behave much more like young speakers of English: SVO is the first and most important cue to meaning from 2 through 6–7 years of age. Somewhere between 6 years to adulthood, French children must go through a radical reorganization of their comprehension system, switching from word order dominance to a primary reliance on morphology and semantics. This

U-shaped function can be explained if we remember that canonical word order is often violated in French—but only in complex criticized structures within informal adult speech. These segments of the input constitute the conflict cases—in contrast with Italian, where word order violations abound even in very simple sentences spoken to children below the age of two (Bates, 1976). In French, violations of canonical word order violations contribute primarily to conflict validity; in Italian, word order variation influences overall cue validity.

The generality of this progression from overall validity to conflict validity has also been demonstrated in an unpublished experiment by McDonald and MacWhinney on artificial concept learning in adults. In acquiring a set of nonverbal concepts, adult subjects went through the same stages of acquisition that we have just described for children: an initial phase in which responses are controlled primarily by cue availability, followed by a lengthy phase of learning controlled primarily by overall cue validity, with the last phase of learning (i.e., the phase of "fine-tuning") driven primarily by cue dominance in conflict cases (see McDonald, this volume). Hence, the phenomena that we have observed in child language acquisition may reflect a much more general fact about distributional learning.

One way to account for these findings is by simply postulating a shift from overall validity to conflict validity as a general law of learning. However, a deeper and more unified explanation for the same phenomena comes from theories of learning in artificial neural networks (Rumelhart, Hinton, & Williams; 1986). In the Rumelhart et al. formulation, learning takes place through "back propagation" (also called the "generalized delta rule"). Specifically, learning takes place when there is a discrepancy between the desired output (the output presented by the teacher/environment) and the actual output (the output predicted by the system at its current level of learning, in the presence of a given input). At each output node, the degree of discrepancy is noted and propagated back through all input-output connections leading to that node. Each of the intervening weights is adjusted in strength (increased or decreased) in proportion to the degree to which that particular weighted connection was responsible for the final error (i.e., in proportion to "blame"). During the bulk of learning, many adjustments are made on any given trial (i.e., blame is widely shared). In the final stages of learning, far fewer errors are committed. Because the weights are changed only when errors are committed (i.e., when predictions are disconfirmed), relatively rare conflict situations begin to control the late phases of learning. As Taub et al. have shown, the shift from overall cue validity to conflict validity that we have observed in our experiments can be simulated in a connectionist network that employs back propagation. With a constant and unchanging input, and a constant and unchanging learning device, we can obtain U-shaped functions and late developments of the sort that are observed in human language learning (McDonald, this volume; MacWhinney, this volume).
We think that the enriched notion of cue validity can account for many of the late developments observed to date—but not for all of them. Language acquisition is also affected by endogenous changes in information processing capacity. These include changes in cue cost (due to increases in perceptual acuity and/or memory), and a series of conceptual developments that we will refer to with the term functional readiness.

Cue cost in language acquisition

Cue cost is an obvious first place to look for constraints on cue validity in acquisition. Children are less developed organisms than adults, in perception, memory, motor skills. The costs of processing a given linguistic structure might well be expected to weigh more heavily for small children—perhaps enough to blunt the impact of information value on sequences of language learning.

The notion of assignability provides an explanation for one particularly puzzling delay in language learning. In MacWhinney et al. (1984) we have found that subject–verb agreement is an extremely strong cue to sentence interpretation for Italian adults; indeed, this cue appears to be just as important for Italians as case cues are for speakers of German, Hungarian, and other case-inflected languages. Similar patterns are obtained in Italian for agreement between the object and a clitic pronoun. Given these results, we should also expect very early acquisition of agreement by Italian children. And indeed, there is evidence available to suggest that Italian children mark agreement productively and correctly in their own expressive language by 2–3 years of age (although object/clitic agreement may come in somewhat later than subject–verb agreement). Nevertheless, in work still in progress, Devescovi and her colleagues have found that Italian children are unwilling to use agreement contrasts in sentence interpretation. Although there are reliable effects for both agreement contrasts by age 3, the effects are very small; semantic factors play a much greater role (accounting for 2–3 times more variance) until 6–7 years of age. Kail reports a similar late onset for both kinds of agreement in French, suggesting that the late use of agreement may be general phenomena. Finally, Devescovi et al. also describe a striking developmental contrast between case and agreement morphology within a single language. For Serbo-Croatian adults, case morphology is the most important source of sentence meaning; however, if case is ambiguous (which it often is in this language), Serbo-Croatian adults are strongly influenced by gender agreement markers on the verb. For Serbo-Croatian children, case (a local cue) becomes the dominant cue to meaning by age 4; gender agreement (a more distributed cue) is not used reliably for several more years.

The general picture appears to be one in which local morphology is acquired early, in accord with its cue validity, whereas more topological agreement cues are not used in sentence interpretation until 5–8 years of age—even if those cues are high in information value. Slobin (1985) has proposed a bias toward local marking as a general operating principle, in both comprehension and production (cf. Ammon & Slobin, 1979). However, given the disparity between production of subject–verb agreement (an early phenomenon) and use of the same forms in comprehension (a late phenomenon), Devescovi and Kail (this volume) both argue for an explanation based on the cue cost of agreement contrasts in receptive processing, in particular the memory costs involved in using cues that are low in assignability. In sentence production, the speaker presumably knows in advance "who did what to whom" and assignment of agreement markers can be made early, without waiting or holding extra material in memory. In sentence comprehension, the listener may not have prior knowledge about "who did what to whom." Under these circumstances, children may prefer to make tentative interpretations as soon as possible, using some combination of word order and semantic information, without holding elements in memory and comparing them to determine whether a usable agreement cue is available. This explanation leads to an interesting hypothesis in Italian: The memory demands of subject–verb agreement apply only for the third person; first and second person cues can be unambiguously assigned, the moment they are encountered, to the speaker and the hearer respectively. This fact leads to the prediction that Italian children will make use of first and second person agreement reliably before three years of age, on a schedule that parallels the appearance of these markers in their own speech production. Pilot results with a small number of Italian two- to three-year-olds and support to this prediction (Devescovi, research in progress).

Perceivability is another source of cue cost that could in principle explain exceptions to cue validity in the development of sentence comprehension. MacWhinney et al. (1985) investigated the use of case suffixes by Hungarian children, comparing the two forms of the accusative described earlier: a -t suffix following a strong vowel (high in perceivability), and a -t suffix following a consonant cluster (low in perceivability). The more perceivable version of the case suffix was established approximately one-year earlier in sentence comprehension. The same finding may explain why there is approximately a one-year overall difference between Hungarian and Turkish children in the onset of case morphology. Case is regularly and unambiguously marked in both these languages; however, the low perceivability of Hungarian case morphology after consonant clusters may delay the completion of case paradigms by Hungarian children compared with their Turkish age mates.

We are currently pursuing some extensions of the perceivability notion, in studies of the acquisition of clitic particles in Italian and French. These preverbal object markers are equally informative in both languages. However, French and Italian differ in the perceptual salience of object clitics in certain verb phrase environments. For example, the final s on the French plural object clitic les is silent in most environments; however, if the clitic is followed by a word
beginning in a vowel (e.g., the participial construction *ont fait*, as in *les ont fait*), then the final s is pronounced — thus rendering the object clitic particularly salient. Compare this with the Italian *li hanno fatto*, in which the vowel-final object clitic is much less salient (blending into the vowel-initial sound of the participle that follows). If perceptual salience augments or degrades the effects of cue validity, then we should find differences between French and Italian in the relative strength of the clitic cue in a participial environment.

There are also some interesting differences between French and Italian in the perceptual similarity among object clitics, full subject pronouns, and articles. For example, the sentence initial sound *le* could signal at least three things in Italian: a third person feminine object clitic (as in *le vede* or “them-feminine sees”), a third person feminine article before a noun (as in *le donne* or “the women”) and/or the first segment of a third person singular subject pronoun (as in *lei viene* or “she is coming”). Some preliminary evidence on pronoun interpretation by Italian children suggests that these potential ambiguities cause them considerable difficulty in the early stages. Confusions between object clitics and articles are also possible in French, but confusions between the object clitic and full subject pronouns are much less likely (i.e., compare *le* with the subject pronouns *lui* and *elle*). We are interested in determining the extent to which subtle phonological factors like these influence the acquisition of clitic forms as a cue to underlying semantic relations. Obviously this complicates any predictions based on cue validity, since a whole catalog of phonological factors would have to be taken into consideration. But if that is the way things are, so be it. Interactions between phonology and syntax may be inconvenient from a theoretical point of view, but they are certainly testable, and may play an important role in the acquisition of grammar.

Functional Readiness in Acquisition

Cue validity is calculated from the objective distribution of forms and their inferred meanings in a sample of text. Even if we assume that children can perceive all of the surface variations in their input language (i.e., horizontal correlations), we cannot assume that they are sensitive from the very beginning to all the meanings signaled by those forms (i.e., vertical correlations). A form–function mapping that is objectively available may be subjectively unavailable if the child has no understanding at all of the function, meaning, or intention signaled by a given linguistic form. The principle of functional readiness refers to the need for certain functions to develop before mappings from form to function can be made.

Consider, for example, the distribution of conditional and subjunctive markers in Italian. Although such forms are relatively rare in adult speech to children (compared with other verb inflections), they are no less frequent than many lexical items that children acquire well before the age of three. And they are associated quite clearly with their respective meanings. Nevertheless, Bates (1976) has shown that the conditional and subjunctive inflections are not mastered by Italian children until somewhere between 4 and 6 years of age. This fact can be explained in part by the cognitive difficulty of the counterfactual notions that underlie both of these verb paradigms. A similar argument can be made for the order of acquisition of locative terms within and across languages (Johnston & Slobin, 1979; Johnston, 1984), and for the relatively late appearance of devices for achieving discourse cohesion (Karmiloff-Smith & Planch, 1980). Children need some kind of guide to tell them when and where a given word, morpheme, or phrase structure should be used. If the children have no idea at all why adults are using a particular sound contrast, they may (a) fail to perceive it in the first place, (b) perceive it, but fail to integrate it well enough into memory for retrieval to take place, or (c) perceive and store the form as an arbitrary associate of a few lexical items (resulting in considerable undergeneralization).

This point may seem to contradict our earlier insistence that distributional learning is possible even in the absence of semantic interpretation. We argued that children could in principle treat distributions of sound like any other perceptual object, detecting and acquiring correlations between morphemes that have no clearcut meaning. Indeed, we must acknowledge such a process to account for the acquisition of German gender paradigms and/or the early appearance of certain forms of agreement marking in Italian. But notice that these latter phenomena are all tied to the lexicon, providing “hacks” on which to hang morphophonological variation. The child is interested in using a given word because she is interested in talking about the referent for that word. Reference provides a functional anchor for the word itself; the variations in sound associated with that word (gender, agreement, etc.) are detected, acquired, and reproduced in the service of reference.

Artificial language learning experiments by Morgan, Meier, and Newport (1987) and Valian and Coulson (1988) are particularly interesting in this regard. These authors presented adults with strings of symbols that obey a set of arbitrary combinatorial rules. These rules had to be learned under one of three conditions: (1) the formal relations in the grammar mapped onto semantic relations in an artificial object world, (2) the formal relations themselves had no semantic base, but each symbol had a constant referent in the same object world, or (3) the grammar had neither semantic nor referential correlates in the object world. Their results suggest that semantic relations may not be necessary for grammatical induction, but constant reference is very important for successful learning.

We suggest that the functional readiness principle is particularly important for sentence-level phenomena that have no clear referential base — that is, for syntactic principles which range across a wide variety of referents, as opposed to lexical/morphological regularities that can be tracked across the input because
they have a recognizable referential "anchor." The child's attention is drawn to regularities of gender and other agreement phenomena because she is interested in the words that incidentally encode those regularities. The same child may not notice or use a phrase structure option in the language, if that phrase structure type is used by adults primarily for discourse purposes that are completely opaque to the child.

We suggest that functional readiness (a property of the child) interacts with conflict validity (a property of the environment). Specifically, changes in cognitive capacity may cause the child to notice conflict cases that she never noticed before. Prior to that point, the subjective input (i.e., the input attended to by the child) may actually be much smaller than the objective input (i.e., the input from the adult's point of view). With this idea in mind, we can provide a unified account for a range of late developments that have appeared in our crosslinguistic research to date:

1. a period around 4–5 years of age in which children overgeneralize some aspect of canonical word order (albeit in different ways, and to different degrees, from one language to another),
2. the shift away from word order dominance toward semantic and morphological strategies in French children after 6 years of age,
3. the appearance of VOS and OSV "second-noun strategies" in English children after age 6, and
4. the late appearance of word order and stress configurations in Italian children after age 6.

The lack of order/stress interactions in Italian is particularly puzzling, since the children are certainly sensitive to the absolute presence of contrastive stress. In fact, children in both Italian and French show a small but reliable tendency in the first years to choose the stressed element as the subject—exactly the opposite of the adult pattern. We have suggested elsewhere that this is essentially a nonlinguistic strategy: If you don't know what else to do, choose the noun that the experimenter said loudest. But it does show that children can hear the stress manipulation; why does it take them so long to use that information?

In the absence of these secondary adult word order patterns, Italian children develop a secondary strategy of their own: Generalize the first noun strategy derived from SVO to other word order types. This strategy starts between 4 and 5 years of age, and does not drop out until after the age of 9, when we finally see reliable use of stress to interpret NNV and VNN sentences. During this period of "word order overgeneralization," Italian children actually make more use of word order and less use of animacy than their adult counterparts (although word order never actually "wins out" over lexical semantics).

Kail (this volume) observed an even more dramatic version of word order overgeneralization in French children (discussed above). A much weaker version of first-noun overgeneralization has been noticed in Chinese (Miao, 1981), Hungarian (MacWhinney et al., 1985), Serbo-Croatian (Smith, personal communication), and Turkish (Slobin & Bever, 1982). In these studies, the overgeneralization of first-noun choice also starts around 4 years of age, although it drops out considerably earlier than it does in Italian or French. No such overgeneralization occurs in our data on English children. Other investigators have reported a tendency for English children to generalize SVO strategies to the passive—a strategy which, interestingly, also peaks between 4 and 5. But the first-noun bias is apparently not extended to NNV and VNN constructions in our research. Bates et al. (1984) suggest that the first-noun tendency is somehow blocked in English, because the English children are sensitive at some level to the factors that create VOS and OSV biases in adults. We will return shortly to an explanation of how this might occur.

Although these patterns vary markedly in size, range, and time of offset, some form of word order overgeneralization seems to occur reliably around age 5 across structurally and functionally distinct language types. The rise and fall of word order overgeneralization can be explained if we assume that the pool of phrase structure types impinging on sentence comprehension undergoes a series of changes during development. In the first phase (peaking between 4–5), evidence for basic or canonical word order accumulates and reaches a saturation point. Around age 5, there is a new influx of phrase structure types, including many complex phrase structures that violate canonical word order. These new entries in the competition pool serve to cut back the strength of canonical word order—to different degrees, and in different ways, depending on the input language. In part, this is the story that we told above, concerning a shift from overall validity to conflict validity. However, we think the reliable timing of this shift across languages can be explained if we also invoke the principle of functional readiness. In particular, we argue that children around age 5–6 begin to hear and to construct longer passages of oral and written discourse. This new contact with narrative speech brings the child into further contact with the sentential forms that adults use to create cohesive texts. Many of the phrase structure configurations that have a late effect on comprehension in our data fall into this category: relatives, clefts and related phrase structure types in English that are responsible for adult second-noun strategies, clefticized phrase structures in French that are implicated in the shift from order dominance to a morphological strategy, and word order/stress configurations in Italian.

Karmiloff-Smith (1979) provides independent evidence for a shift around 6 years of age, from a sentence-based grammar organized to express the semantics of simple events, to an intersentential grammar organized around the demands of discourse. For example, she describes a series of interesting errors that English and French children make around this age in the use of pronouns and determiners. These errors on structures that are produced flawlessly by much younger children can be explained on functional grounds: younger children use the same forms "exophorically" (to describe referents that are physically and
psychologically present), whereas the older children are breaking into the more demanding “anaphoric” uses of the same linguistic forms (to describe multiple referents in a connected story, and/or to remind the listener of referents and events that occurred at another time and place). The same argument may apply to certain complex surface configurations that operate as a block (e.g., clefts), in the service not of sentence-level “event semantics” (e.g., who did what to whom) but rather, intersentential “discourse-level” semantics (e.g., how does this proposition contrast with a presupposition derived from previous discourse?).

If this argument is correct, then we have to predict parallels between comprehension and production in (a) the peak use of canonical phrase structures around age 5, and (b) a subsequent increase in discourse-based phrase structures after age 6. We do not yet know whether this is the case. It is certainly true that preschool children reliably produce only a subset of the possible phrase structure types in their language. However, we need more evidence about such things as the range and frequency of order/stress configurations for child speakers in “free word order” languages like Italian and Hungarian.

We also need to explain why English children fail to overgeneralize SVO to the same extent as children in the other languages studied to date. If they do not have the phrase structures responsible for the OSV and VOS biases of English adults, then why should any “blocking” occur at all? We may have to postulate a “two-tiered” membership in the competition pool of either word or phrase structure candidates. Active members in the pool are members that have been functionally assimilated (as defined above). These candidates exert the greatest force in a parsing decision. However, the child may well retain some memory of sentence types that he or she heard but failed to understand. If enough of these exceptions accumulate, they may serve to block certain generalizations that would be possible in their absence. This “second tier” notion is similar to the “waiting room” idea discussed by Ammon and Slobin (1979), and to the “file of unknown forms” in MacWhinney (1978).

The functional readiness principle can handle a variety of exceptions to cue validity. It is really a very simple notion, tantamount to “What you don’t know can’t hurt you (or help you).” And it may interact in interesting ways with the objective principles of overall validity and conflict validity. However, this principle can become completely circular if it is not used with caution. We need independent evidence for any and all claims that a given grammatical structure is “conceptually difficult” — evidence other than the mere fact that the structure in question appears late in development. Crosslinguistic research is one particularly helpful strategy in this regard, enabling us to separate conceptual difficulty (a dimension that should, in principle, apply equally across natural languages) from variations in structure. This is of course a major motivation behind the crosslinguistic methods pioneered by Slobin and his colleagues (Slobin, 1985). We hope that we have made some contributions here as well.

**Conclusion**

We suggested at the outset that a model must be evaluated along three dimensions: (1) its internal coherence (i.e., can it respond to the data without invoking ad hoc assumptions and/or circularities?), (2) its heuristic value, and (3) how well it compares with competing models in accounting for the same range of data. Let us now consider, briefly, how well the Competition Model stacks up against these three criteria.

With regard to internal coherence, the Competition Model is now considerably more explicit and detailed than it was more than ten years ago (when we first presented our joint work at the 1978 State of the Art conference organized by Gleitman and Wanner — see papers in Gleitman & Wanner, 1982). The principle of cue validity explains an extraordinary range of data, particularly when it is enriched by the distinction between overall validity and conflict validity. A number of interesting constraints on this powerful principle have also been discovered, particularly in the domain of cue cost. All of these constraints have an independent motivation, and they can all be investigated thoroughly and systematically (e.g., degrees of perceivability, assignability, or functional readiness). We have offered a series of explicit and testable claims about the structure of the system, processing dynamics and acquisition: Some of these claims are now precise enough to permit mathematical modeling (McDonald & MacWhinney, this volume), some progress has been made in the development of a realistic parser (MacWhinney, 1988), and there are now simulations of morphological learning that are consistent with the Competition Model (MacWhinney, this volume).

The heuristic value of the model in its present form also seems to be well-established. In addition to the research on normal processing and acquisition described here, the Competition Model has been applied successfully to language breakdown in adult aphasics (Bates & Wulfeck, this volume) and to sentence processing in bilinguals (Kilborn & Ito, this volume). Some further developments in the study of sentence production are described by Bates and Devescovi (this volume). Kail (this volume) underscores the importance of on-line investigations; her point is well taken, and first steps in this direction seem to be quite promising (Kilborn, 1987; Wulfeck, 1987; McDonald & MacWhinney, in press; Kutas & Bates, 1988). At the moment, at least, there seems to be no shortage of new ideas within this research framework — one of the major criteria for evaluating the success and utility of a model.

Finally, the Competition Model fares reasonably well against its competitors — primarily because there are no competitors. To our knowledge, there is no other comprehensive account of linguistic performance across natural languages. This is largely because no one else is trying to do what we are trying to do. There
are three research currents that seem to speak to some of the same issues, but in reality the goals of these three other enterprises are quite different. A number of linguistic theories make interesting claims about crosslinguistic variations in competence. Such theories confine themselves to the study of the "core grammar," often focusing on structures involved in anaphora and case relations. Good examples of the application of linguistic principles to the study of competence in different languages can be found in the work on anaphora collected in Lust (1987) or the work on parameters found in Roepke and Williams (1987). This work attempts to look at variations in linguistic structure with an eye toward identifying a set of parameters which characterize the variation between languages. Having identified these parameters, workers in this tradition then hope that it might be possible to construct an identification procedure that would induce the proper grammars on the basis of sentences of the language and the innate constraints of universal grammar. This work seems to be a useful way of developing formal linguistic theory. However, it tells us little about linguistic performance and places relatively slight emphasis on the experimental verification of hypotheses. This is not a criticism of this line of work. Rather it is simply an observation that it is directed toward entirely different goals.

A much smaller group of investigators has used competence models to make strong claims about linguistic performance. The work of Berwick (1987) and Hyams (1986) is in this tradition. In principle it might be possible to use competence models to predict the quantitative data that are central to language acquisition. However, in practice, researchers trained in the tradition of competence theory have little understanding of the basic principles of experimentation and quantitative analysis that would be necessary to make strong links between competence and performance. Without such skills, these applications are necessarily doomed to failure when they come in contact with the realities of variability in language acquisition and processing data.

A third, extremely vital tradition is represented by the research collected in Slobin (1985). This tradition uses data on spontaneously occurring errors and early uses to abstract a set of "universal operating principles." These operating principles provide the child with biases toward preferring some grammatical forms more than others and hints about the best place to start in the process of acquiring a natural language. Although they seem to have considerable crosslinguistic generality, these operating principles were never intended to serve as a coherent theory of linguistic performance, in children or in adults. Rather, they constitute an ecologically valid set of strategies from which an integrated processing model can be constructed. In this way, work on operating principles can feed directly into work on the Competition Model.

The ultimate duty of a research model is to die for its science, providing hypotheses that are flatly and roundly disconfirmed. We fully expect that to happen to us as well, probably very soon. In the meantime, we hope that the success of our enterprise (and its failures) will encourage investigators to apply the crosslinguistic method in psycholinguistic research. To describe, predict, and explain human linguistic performance, we have to look outside the boundaries of English — an exotic code which is, according to Arnold Pick (1913), "An essentially formless language of high standing."