COGNITIVE SCIENCE Vol 23 (4) 1999, pp. 589–613 ISSN 0364-0213

Grammar-based Connectionist Approaches to Language

PAUL SMOLENSKY

Johns Hopkins University

This article describes an approach to connectionist language research that relies on the development of grammar formalisms rather than computer models. From formulations of the fundamental theoretical commitments of connectionism and of generative grammar, it is argued that these two paradigms are mutually compatible. Integrating the basic assumptions of the paradigms results in formal theories of grammar that centrally incorporate a certain degree of connectionist computation. Two such grammar formalisms—Harmonic Grammar (Legendre, Miyata, & Smolensky, 1990a,b) and Optimality Theory (Prince & Smolensky, 1991, 1993)—are briefly introduced to illustrate grammar-based approaches to connectionist language research. The strengths and weaknesses of grammar-based research and more traditional model-based research are argued to be complementary, suggesting a significant role for both strategies in the spectrum of connectionist language research.

I. INTRODUCTION

This article is addressed to basic methodological issues arising in connectionist research on language. I will attempt to briefly sketch a lengthy argument begun in Smolensky, Legendre, and Miyata (1992) and presented in detail in Smolensky and Legendre (in progress). In many places, I will try to articulate personal viewpoints and, to a very limited degree, justify them. The focus is on two main claims. The first is that there are *two* general styles of research that both deserve a central place in connectionist approaches to language. The first, *model-based research*, is well established. The second, *grammar-based research*, is less so. Each approach, I will argue, has important strengths that are lacking in the other. The second main claim is that the time has come to stop regarding generative grammar and connectionist approaches to language as incompatible research paradigms. Each has significant potential for contributing to the other.

Direct all correspondence to: Paul Smolensky, Department of Cognitive Science, Johns Hopkins University, Krieger Hall 239A, 3400 N. Charles St., Baltimore, MD 21218-2685; E-mail: smolensky@jhu.edu

I will suggest a view of the core theoretical commitments of the two paradigms, connectionism and generative linguistics, and argue that these commitments combine to support a coherent and fruitful research program in connectionist-grounded generative grammar. It is my belief, although I will not attempt to justify it in detail here, that the core commitments I identify are indeed consensus beliefs of the connectionist and generative linguistics research communities.

Going beyond the core commitments, individual researchers have further commitments which are often not mutually compatible, and these competing scientific hypotheses must of course be adjudicated by theoretical and empirical arguments. But at this level, competition between incompatible hypotheses is readily found among generative grammarians themselves, or among connectionists themselves, as well as between generative linguists and connectionists. Thus it seems to me more accurate to regard the current *scientific* debates about language as individual conflicts between individual hypotheses, rather than a war between two unified paradigms, Connectionism and Generative Grammar.

II. COMMITMENTS OF CONNECTIONISM

The Parallel Distributed Processing (PDP) school of connectionism is founded, it seems to me, on the following general principles (Rumelhart, McClelland, & the PDP Research Group, 1986):

(1) Fundamental commitments of connectionism: The PDP Principles
  a. Mental representations are distributed patterns of numerical activity.
  b. Mental processes are massively parallel transformations of activity patterns by patterns of numerical connections.
  c. Knowledge acquisition results from the interaction of
    i. innate learning rules
    ii. innate architectural features
    iii. modification of connection strengths with experience

Testing the implications of these fundamental principles is a challenge in part because of their great generality: depending on how they are instantiated, they can be used to support a number of contradictory claims concerning fundamental issues such as modularity and nativism. This diversity of potential implications of the PDP Principles was already rather clearly in evidence in the earliest PDP models, as the representative citations in the following paragraphs show.

Consider the first PDP Principle: *Mental representations are distributed patterns of numerical activity.* This can easily be seen as entailing that mental representations are crucially graded (non-discrete). Indeed, this is the default case illustrated by many connectionist models, including the majority of the early ones discussed in Rumelhart, McClelland, and the PDP Research Group (1986) and McClelland, Rumelhart, and the PDP Research Group (1986).

However, this same PDP Principle is consistent with the claim that mental representations are discrete, as seen in a number of classes of connectionist model. Most obviously, representations are discrete in networks with discrete-valued units (e.g., the original Boltzmann Machine—Hinton & Sejnowski, 1983—and Harmony Theory—Smolensky, 1983, architectures). But discreteness also plays a crucial role in the important class of models with continuous units that converge to discrete representations (e.g., Anderson, Silverstein, Ritz, & Jones, 1977; Rumelhart, Smolensky, McClelland, & Hinton, 1986), including models with 'winner-take-all' sub-networks (e.g., Grossberg, 1976; Feldman & Ballard, 1982; Rumelhart & Zipser, 1985; Mozer, 1991). And of course discreteness of representations is also a central property of a number of connectionist techniques for embedding symbolic structures as patterns of activity (e.g., Touretzky, 1986; Touretzky & Hinton, 1988; Dolan, 1989; Pollack, 1990; Smolensky, 1990).

The conclusion must be that the PDP Principle concerning representations, (1a), is consistent with both crucial discreteness and crucial non-discreteness of mental representations.

In similar vein, consider the second PDP Principle, (1b): *Mental processes are massively parallel transformations of activity patterns by patterns of numerical connections.* This can readily be viewed as entailing that mental processing is highly interactive, non-modular, non-sequential. Indeed, this is the case for "classic" PDP models (e.g., Sejnowski & Rosenberg, 1987) in which one layer of input units, containing information of many types, projects directly to one output layer of units (possibly through a hidden layer), with connectivity unrestricted.

But clearly the PDP Principle is also consistent with the claim that mental processing is modular or sequential. Modularity of a certain type is central to connectionist models in which different types of information are represented over different groups of units, and in which restricted connectivity between groups of units allows only certain types of information to interact directly (e.g., Mozer, 1991; Plaut & Shallice, 1994). Modularity is the heart of networks that learn to specialize different sub-networks for different sub-tasks (e.g., Jacobs, Jordan, Nowlan, & Hinton, 1991). And sequentiality is crucial for many applications of Simple Recurrent Networks (Elman, 1990) and related recurrent architectures, including the early model of Jordan (1986).

Again, we must conclude that the PDP Principle governing processing, (1b), is consistent with modular or non-modular processing, and processing with or without essential sequentiality.

Finally, consider the most controversial of the PDP Principles, (1c): *Knowledge acquisition results from the interaction of innate learning rules, innate architectural features, and modification of connection strengths with experience.* On first glance, this principle would seem to imply that knowledge acquisition consists entirely in the statistical associations gathered through experience with some task. And this does characterize those early connectionist *tabula rasa* models with simple Hebbian-like learning rules (e.g., Kohonen, 1977; Stone, 1986) and input/output representations that have no built-in domain structure (e.g., local representations as in Hinton, 1986).

However much of the most noteworthy progress in connectionist learning is more appropriately characterized by another claim: connectionist knowledge acquisition consists in fitting to data the parameters of task-specific knowledge models. To varying degrees, this describes somewhat more recent networks (e.g., Rumelhart, Durbin, Golden, & Chauvin, 1996; Smolensky, 1996b) in which more sophisticated error functions that embody Occam's Razor force simplicity of knowledge to compete with closeness-of-fit to data. It also describes networks with specialized activation functions, connectivity patterns, and learning rules that provide task-appropriate biases to the learning process (e.g., McMillan, Mozer, & Smolensky, 1992), and models in which input/output representations reflect, even implicitly, task-specific regularities (e.g., see Lachter & Bever, 1988, and Pinker & Prince, 1988, on Rumelhart & McClelland, 1986).

Thus a commitment to the PDP Principles of (1) does not *per se* constitute a commitment regarding the degree to which discreteness, modularity, or innate learning bias applies to human cognition. This conclusion is no surprise to the practicing connectionist, of course, but it seems to deserve considerably more recognition than it tends to get in polemical debates between pro- and anti-connectionists. The reason for bringing it up here, however, is to point out that *the indeterminism of the basic connectionist commitments toward most central issues of cognitive theory forces a major choice of research strategy*. The most popular choice is this.

(2) Model-based strategy for connectionist research on language
Because the basic connectionist principles (1) are too general to have definitive consequences for key theoretical issues, less vague connectionist proposals are needed. These can be achieved as follows:

a. Choose a *particular* set of cognitive data on which these issues bear (a set of specific input/output pairs).
b. Propose a *specific connectionist model* in which, for the particular data in question, choices are made of a specific input/output representation, specific activation functions and learning algorithms, specific numbers of internal layers of specific sizes and connectivity, etc.
c. Evaluate the proposed model based on the closeness-of-fit to the data achieved by a computer simulation of the model, and on the internal structure in the model that allows it to achieve its performance.

As I will discuss in Section V, there are many advantages to this research strategy, and it has produced many important results (for an overview, see Christiansen & Chater, this issue). The main point of this article is to argue that there is, however, *another* strategy available within PDP connectionism, and while this new approach has significant limitations, it also has certain advantages lacking in the model-based strategy. This alternative strategy could be formulated at various levels of generality, but given the topic of this Special Issue I will use a formulation directed at the study of language. (For another cognitive domain, say, reasoning, a description of the analogous strategy results from replacing "grammar formalism" with "formal theory of human reasoning," etc.)

(3) Grammar-based strategy for connectionist language research
Because the basic connectionist principles (1) are too general to have definitive consequences for key theoretical issues, less vague connectionist proposals are needed. These can be achieved as follows:
   a. Choose a mathematically precise formulation of the PDP Principles (1).
   b. Derive from these principles a precise but general grammar formalism (or grammatical theory).
   c. To evaluate the proposed formalism, choose a particular class of target empirical generalizations concerning human language.
   d. Apply the grammar formalism to language data instantiating the target generalizations, defining a formal 'account' of the target phenomena.
   e. Compare the degree of explanation of the target generalizations that can be achieved with the new account with that achieved with previous grammar formalisms. ('Explanation' here means deduction of generalizations and particular data from the principles defining the proposed account.)

I would not argue that all, or even most, connectionist research on language should pursue this grammar-based strategy; my claim is only that a central place in connectionist language research should be reserved for a certain amount of such work, because the model- and grammar-based strategies are nicely complementary. What the grammar-based strategy seeks to provide is a way of pursuing the explanatory goals of linguistic theory while incorporating computational insights from connectionist theory concerning mental representation, mental processing, and learning.

In the remainder of this article I will: (§III) briefly illustrate the grammar-based strategy with Harmonic Grammar and Optimality Theory; (§IV) identify some of the central goals and commitments of one approach to linguistic theory, generative grammar, and illustrate how Optimality Theory addresses these goals while introducing certain general connectionist insights; (§V) discuss the complementary strengths and weaknesses of the model- and grammar-based strategies for connectionist research on language.

III. CONNECTIONIST GRAMMAR ILLUSTRATED

If we believe that a mind is an abstract, higher-level description of a brain—as I presume most cognitive scientists do—and if we believe that connectionist networks provide a useful stand-in for a solid theory of neural computation yet to be constructed—as I presume most connectionists do—then it follows that abstract, higher-level descriptions of connectionist computation should provide the basis for theories of mind (Smolensky, 1988). The nature of the abstract, higher-level descriptions of connectionist computation will depend on which types of connectionist networks we adopt: some more discrete than others, some more modular than others, some containing more innate knowledge than others, etc.

Among many such possibilities, here is one.

(4) The Symbolic Approximation
In the domain of language, the patterns of activation constituting mental representations admit abstract, higher-level descriptions that are closely approximated by the kinds of discrete, abstract structures posited by symbolic linguistic theory.

That this is indeed *a possibility* is suggested by research over the past decade showing how distributed patterns of activity can possess the same abstract properties as symbolic structures like syntactic trees. (One such proposal, tensor product representations, Smolensky, 1990, is mentioned below; for related proposals, see Dolan, 1989; Pollack, 1990; Plate, 1991; see also Tesar & Smolensky, 1994, for relations to temporal-synchrony schemes such as Hummel & Biederman, 1992; Shastri & Ajjaganadde, 1993.) That the Symbolic Approximation is the *correct* possibility is, of course, a working hypothesis; indeed, it seems to me that it must be the hypothesis underlying all symbolic language research that takes linguistic representations to be psychologically real. Evidence for the correctness of this hypothesis comes from the successes of symbolic linguistic theory in explaining the overall structure of human language, successes which I take to be most impressive.

The relevant hypothesis, to be somewhat more precise, is that *some*—not *all*—aspects of the mental representation of linguistic information are well approximated by the abstract symbolic structures posited by linguistic theory. It is agreed by everyone that such representations do not capture all that cognitive theory wants to capture. The claim is that what they *do* capture has great explanatory power.[1]

The approaches to connectionist grammar I will now briefly describe both assume the Symbolic Approximation. This assumption is not required by the grammar-based strategy for connectionist research outlined in (3); indeed, it would be extremely interesting to develop a connectionist grammar based on an alternative formal higher-level description of linguistic representations (for example, less discrete, more image-like, representations, perhaps along the lines of certain proposals of ‘cognitive linguistics,’ e.g. Lakoff, 1987; Langacker, 1987; Talmy, 1988).

Harmonic Grammar

As a first illustration of the connectionist grammar strategy, I indicate explicitly for each step of the strategy (3) how that step is instantiated in *Harmonic Grammar*. The presentation here is brief and informal; more formal versions may be found in Legendre, Miyata, and Smolensky (1990a). We will step through the five parts of the Grammar-based Strategy outlined in (3) and Smolensky, Legendre & Miyata (1992).

a. Choose a mathematically precise formulation of the PDP Principles (1). The first principle, *Mental representations are distributed patterns of numerical activity* (1a), is made more precise via the Symbolic Approximation: linguistic representations are assumed to be patterns of activity in a connectionist network that are well-approximated by tensor product realizations of the types of symbolic structures proposed in symbolic theories. Tensor product representations are a general class of schemes by which structured information is encoded in distributed representations. The distributed pattern realizing a symbolic structure is the sum or superposition of distributed patterns realizing its constituent parts, each of which is a distributed pattern realizing a symbolic filler bound by the tensor (generalized outer) product operation to a distributed pattern realizing its structural role.

The second principle, *Mental processes are massively parallel transformations of activity patterns by patterns of numerical connections* (1b), is more precisely rendered as: linguistic processing is performed by a connectionist network with a ‘harmonic’ architecture, which entails that network outputs maximize Harmony, i.e., optimally satisfy the simultaneous soft constraints encoded in the connections. (The connectionist principle of Harmony maximization, or “energy” minimization, and an appreciation of its significance and generality, emerged from a body of research, including Hopfield, 1982, 1984; Cohen & Grossberg, 1983; Hinton & Sejnowski 1983, 1986; Smolensky, 1983, 1986; and Golden, 1988. For relevant recent review articles, see Hirsch, 1996, and Smolensky, 1996a.)

The final principle is, *Knowledge acquisition results from the interaction of innate learning rules, innate architectural features, and modification of connection strengths with experience* (1c). More precisely, we assume that the connections in the language-processing network have been adjusted so that the higher the Harmony of a linguistic structure, the more well-formed it is in the language; specifically, the Harmony-to-well-formedness function is assumed to be a monotonically-increasing logistic function.

b. Derive from these principles a precise but general grammar formalism (or grammatical theory). Given the assumptions in a., it can be shown that the language-processing network has the following property. At a higher level of description, the output of the network corresponds to a symbolic structure that optimally satisfies a set of soft constraints of the form: “if a structure contains a constituent of type *i*, then it must [or must not] contain a constituent of type *j* (strength: $H_{i,j}$).” A particular set of such constraints defines a *Harmonic Grammar.*

c. To evaluate the proposed formalism, choose a particular class of target generalizations concerning human language. In a wide range of languages, intransitive verbs divide into two classes according to whether their argument Noun Phrase displays “object-like” or “subject-like” properties (e.g., in *the river froze,* the argument of ‘froze,’ ‘the river,’ displays object-like syntactic properties relative to the more subject-like argument of ‘shone’ in *the river shone*; Levin & Rappaport–Hovav, 1995). Which type of behavior is displayed by the arguments of a verb correlate in complex ways with various syntactic and semantic properties of the verb and argument, and precisely characterizing these correlations as grammatical principles often proves problematic.

d. Apply the grammar formalism to linguistic data instantiating the target generalizations, defining a formal “account” of the target phenomena. A variety of French intransitive verbs and sentence structures illustrate the correlations referred to in c. A set of soft constraints defining a Harmonic Grammar is developed to account for the overall acceptability pattern of the French sentences. A typical soft constraint is “if a structure contains a verb describing an event with an inherent endpoint, then it must not contain a subject-like argument.” These rules, with their strengths, capture the correlations of interest; the Harmonic Grammar formalism defines formally how the conflicting demands of these rules are combined to make precise predictions of sentence well-formedness.

e. Compare the degree of explanation of the target generalizations that can be achieved with the new account with that achieved with previous grammar formalisms. (‘Explanation’ here means deduction of generalizations and particular data from the principles defining the proposed account.) The Harmonic Grammar account allows a complexity of interaction between syntactic and semantic constraints that better fits the French data, while elucidating the nature and grammatical role of the general correlations. The proposed soft constraints provide the means for precise prediction of the well-formedness of particular sentences, as well as deductive links to the general correlations to be explained.

Optimality Theory

A grammar formalism conceptually related to Harmonic Grammar, *Optimality Theory* (Prince & Smolensky, 1991, 1993, 1997), is considerably more restrictive, and, as we see in the next section, therefore better suited to the explanatory goals of linguistic theory. The relevant principles of the theory are summarized in (5) below, using the traditional generative concept of the ‘input’ and ‘output’ of a grammar; for current purposes, we can roughly take the input to be the ‘intended interpretation’ a speaker wishes to convey, and the output to be the actual linguistic structure which expresses that interpretation in the language (for a more accurate characterization, see Legendre, Smolensky & Wilson, 1998).

(5) Optimality Theory

a. Given an input, the grammar produces as output the linguistic structure that maximizes Harmony.

b. The Harmony of a potential output is the degree to which it simultaneously satisfies a set of violable constraints on linguistic well-formedness (including constraints requiring that the output faithfully express the input).

c. The constraints have different strengths, determining which take priority when constraints conflict.

d. The grammar of a language is a ranking of constraints from strongest to weakest; a higher-ranked constraint has absolute priority over all lower-ranked constraints.

e. The set of possible outputs, and the set of constraints, is the same in all languages; grammars of languages differ only in the way constraints are ranked.

The additional restrictiveness of Optimality Theory (OT) over Harmonic Grammar comes from restricting the interactions among constraints to those that can be achieved by ranking (as opposed to arbitrary numerical strengths), and from the principle that the grammatical constraints, and the possible outputs, are the same in all languages. These additional restrictions reflect important *empirical* generalizations about language, some of which are long-standing observations of linguists, others of which were discovered in the process of developing OT; more on this in the next section.

In (6), connectionism is related to the fundamental principles that define OT and differentiate it from other generative theories of grammar. (See also Prince & Smolensky, 1993; chap. 10.)

(6) Fundamental defining principles of OT, and their relation to connectionism

Principles deriving from connectionism

a. *Optimality*. The correct output representation is the one that maximizes Harmony.

b. *Containment*. Competition for optimality is between outputs that include the given input. (Clamping the input units restricts the optimization in a network to those patterns including the input.)

c. *Parallelism.* Harmony measures the degree of simultaneous satisfaction of constraints. (Connectionist optimization is parallel: the constraints encoded in the connections all apply simultaneously to a potential output.)

d. *Interactionism*. The complexity of patterns of grammaticality comes not from individual constraints, which are relatively simple and general, but from the mutual interaction of multiple constraints. (Each connection in a network is a simple, general constraint on the co-activity of the units it connects; complex behavior emerges only from the interaction of many constraints.)

e. *Conflict*. Constraints conflict: it is typically impossible to simultaneously satisfy them all. (Positive and negative connections typically put conflicting pressures on a unit’s activity.)

f. *Domination*. Constraint conflict is resolved via a notion of differential strength: stronger constraints prevail over weaker ones in cases of conflict.

g. *Minimal violability*. Correct outputs typically violate some constraints (because of e.), but do so only to the minimal degree needed to satisfy stronger constraints.

h. *Learning requires determination of constraint strengths*. Acquiring the grammar of a particular language requires determining the relative strengths of constraints in the target language.

Principles not deriving from connectionism

i. *Strictness of domination*. Each constraint is stronger than all weaker constraints combined. (Corresponds to a strong restriction on the numerical constraint strengths, and makes it possible to determine optimality without numerical computation.)

j. *Universality*. The constraints are the same in all human grammars. (Corresponds to a strong restriction on the content of the constraints, presumably to be explained eventually by the interaction of certain innate biases and experience.)

The OT conception of grammar embodied in the first eight principles (6a–h) directly reflect basic connectionist computational principles. The last two principles (6i–j), however, are unexpected from a connectionist perspective. These two principles reflect empirical discoveries about the similarities and differences among human grammars, to be discussed in the next section. These surprising principles would appear to have strong implications for the connectionist foundations of OT, but these potentially important implications remain to be explored in future research.

In the final section, I will briefly consider the relative strengths and weaknesses of the model-based and grammar-based strategies for connectionist language research. But the strengths of the grammar-based connectionist strategy are closely tied to the goals of grammar-based research more generally, so I digress to discuss these in the next section. In the process I will argue that generative grammar and connectionism are not incompatible research paradigms, the second main claim of this article.

IV. OPTIMALITY THEORY AND GENERATIVE GRAMMAR

Within generative linguistics, a grammar is taken to be a mathematical function: given an input—for us, roughly, an intended interpretation—the grammar determines an output—the linguistic structure that expresses that interpretation. The grammar in the mind of an English speaker is the knowledge which entails that "which theory did Noam trash?" is the English formulation of a particular question in a particular discourse context, a question which would be rendered in other languages (with appropriate substitution of the corresponding words): "Noam which theory trashed?," "Noam trashed which theory?," "which theory trashed Noam?," etc. A speaker uses her grammar every time she utters or hears a sentence; popular misconceptions notwithstanding, grammars are for producing and comprehending words and sentences—they are not for producing meta-linguistic grammaticality judgments (although they can be indirectly pressed into service for that purpose).

Among the most basic generalizations molding the enterprise of generative grammar are those listed in Table 1; shown also are the roles generative grammar assigns to grammatical theory in response to these generalizations (see Archangeli, 1997, for a recent pedagogical exposition). The final column of Table 1 indicates schematically how Optimality Theory meets the demands of a generative theory of grammar. By using optimal satisfaction of simultaneous conflicting violable constraints as the computational mechanism, OT offers novel proposals for solving the basic problems of grammatical theory, and new types of explanations of the central generalizations of linguistics.

The central column of Table 1 constitutes, I believe, the core of the central commitments defining generative grammar. None of these seems to me inconsistent with con-

TABLE 1
Some Central Generalizations Shaping the Generative Approach to Grammatical Theory

	Generalizations	Role of Theory	Optimality Theory
a.	Grammars of widely-scattered languages of the world share a tremendous amount of commonality.	Grammatical theories must identify these common principles.	All human grammars share a specified set of well-formedness constraints.
b.	With respect to some particular aspect of linguistic structure, the grammars of languages differ, but in a remarkably limited number of ways.	Grammatical theories must identify exactly the possible modes of variation across languages.	Language-particular grammars are different rankings of the same constraints.
c.	The principles common across languages are connected to the observed data in complex ways.	Grammatical theories must provide formal accounts of the complex connection between general linguistic principles and the data of a particular language.	The structures of a language are those that maximize Harmony, i.e., those that optimally satisfy the constraints as ranked by the language's grammar.
d.	Relative to the astronomical number of generalizations children *could* draw based on the data of their language, they converge remarkably quickly toward a correct grammar.	Grammatical theories must provide formal accounts of how a correct grammar can be efficiently learned.	The space of possible OT grammars is sufficiently restricted and well-structured that algorithms for identifying a correct ranking can be formulated, and efficiency results obtained.

nectionist principles. Indeed, in addition to Optimality Theory, other connectionist-based approaches to generative grammar can be identified, for example, the Linear Dynamic Model (Goldsmith & Larson, 1990). (See also Touretzky & Wheeler, 1990, a connectionist implementation of a proposal by Lakoff, 1993.) The Linear Dynamic Model is a framework proposed for syllabification and stress assignment based on a particular connectionist architecture. This framework was used to argue the importance of graded linguistic representations and of numerical weights for encoding phonological grammars; yet this model has received serious consideration within generative linguistics. (Indeed, Alan Prince—known to most connectionists for his critical evaluation with Steven Pinker of early claims about grammar based on connectionist modeling, Pinker & Prince, 1988—has studied this model in great mathematical detail in order to evaluate its linguistic implications: Prince, 1993. The name 'Linear Dynamic Models' derives in fact from Prince's formal analysis of the linear dynamics of these networks. As part of his analysis, Prince proposes an alternative version of the model which is radically *more* continuous.) It should not be surprising that the Linear Dynamic Model was taken seriously by generative linguists, because, despite the major breaks with mainstream phonology concerning the nature of linguistic representations and grammatical knowledge, this work nonetheless attempts to address the central issues identified in Table 1.

The commitments of generative grammar are not inconsistent with graded representations, grammars encoded as numerical weights, or, indeed, probabilistic models, non-modular architecture, or theories in which all language-specific knowledge is acquired by induction. The commitments of generative grammar are to explain the overarching generalizations about human language summarized in Table 1, and if there is in the generative literature a strong preponderance of grammatical formalisms that rely on discrete representations, sequential symbol manipulation, modular non-probabilistic rule systems, and a highly constrained role for learning from the environment, it is because those working assumptions have enabled substantial progress in addressing the issues of Table 1. Other proposals that also do so—be they based in numerical representations and knowledge like the Linear Dynamic Model, or parallel satisfaction of soft constraints like Optimality Theory—will be seriously considered, and evaluated on the quality of the explanations that they provide for the key generalizations in Table 1 and their many particular manifestations in the world's languages. At least that is what I believe the historical record shows.

In this section I have tried to explain why generative grammar and connectionism should not be considered incompatible research paradigms. The practice of generative grammar requires addressing a certain body of generalizations and providing precise theories of grammatical knowledge that offer some explanation of these generalizations. Connectionism provides proposals for the computational architecture underlying grammar: proposals concerning representation, processing, and learning. The grammar-based strategy for connectionist research on language provides a way of integrating connectionism with generative grammar. In the final section, I give reasons why such an integrated research program is a valuable complement to model-based connectionism.

V. COMPLEMENTARITY OF MODEL- AND GRAMMAR-BASED STRATEGIES FOR CONNECTIONIST LANGUAGE RESEARCH

Two criteria are useful for bringing out the complementary strengths and weaknesses of the model- and grammar-based strategies for connectionist language research. The first criterion concerns the feasibility of incorporating connectionist computational proposals into language research; the second, the feasibility of providing explanations of empirical generalizations about language.

Feasibility of Incorporating Connectionist Computational Proposals into Language Research

It will not have escaped anyone's attention that while Optimality Theory replaces generative grammar's previous computational architecture—sequential rule application and hard constraints—with parallel optimization over soft constraints (6), OT fails to incorporate many other important features of connectionism: graded representations, probabilistic processing, and associationist learning, to name a few. That OT lacks these features, in my personal view, is not per se a virtue—although what I find most striking is the impact on grammatical theory that can result from incorporating even a *single* connectionist conception, parallel optimization over soft constraints. It is my personal hope that the grammar-based strategy for connectionist language research will develop further grammatical theories that successfully incorporate more features of the connectionist computational architecture—but one has to start somewhere.

The real point here is that incorporating connectionist computational features into grammar-based research is particularly difficult, because what is needed is a precise characterization of the *higher-level consequences* of these connectionist features—precise enough that these consequences can be used as the fundamental principles of a formal grammatical theory. The connections between Optimality Theory and connectionist theory arise from two connectionist features the higher-level consequences of which can be connected to grammar: tensor product representations—connecting the low-level structure of certain distributed patterns of activity to their higher-level structure as symbolic representations—and Harmony maximization—connecting the low-level structure of certain activation-spreading rules to their higher-level structure as parallel optimization over soft constraints. Waiting to be exploited, I believe, are many more such connections linking lower- and higher-level structure in neural networks (indeed, hopes of furthering this aspect of connectionist theory was my primary motivation in assembling a book: Smolensky, Mozer, & Rumelhart, 1996).

The model-based strategy suffers from no such limitation on the incorporation of features of connectionist computation into language research. Not relying on a mathematical characterization of the higher-level consequences of lower-level network assumptions, the model-based strategy uses computer simulation to explore the consequences for specific linguistic data of specific computational proposals. This has opened up new ways of conceptualizing linguistic representations via distributed representations, suggested new ways of computing linguistically-relevant functions using massively parallel processing, and deepened our perspective on language learning, for example, by providing provocative glimpses into the self-organizing capacities of sophisticated inductive systems. It is my view that this work is still in the exploratory stage, with the main issues remaining open: Can connectionist networks capture real linguistic knowledge without, to a considerable extent, implementing symbolic representations and rules? Can connectionist learning provide an account of linguistic acquisition that is not an implementation of innate-knowledge-triggered-by-experience? While I consider these questions largely unresolved by modeling research to date, I also believe that our understanding of the issues has been significantly sharpened and deepened by modeling work.

It seems likely to me that model-based research will always play a central role in connectionist approaches to cognitive science because the computer-simulation-based study of various features of connectionist computation will likely continue to be decades ahead of strong mathematical results concerning those features. (At the same time, it seems clear that mathematical results, for their part, have frequently precipitated the development of many new connectionist techniques.)

Feasibility of Providing Explanations of Empirical Generalizations about Language

What, then, is the relative advantage of *grammar*-based connectionist research? The principal advantage, I will argue, concerns scientific explanation. Since this is a notoriously controversial topic, I will attempt to ground my argument in a (very!) concrete little example.

Consider the French word for small (masculine), *petit*. Oversimplifying slightly, it is pronounced with a final *t* [pətit] only when it is followed by a vowel-initial word; otherwise, the pronunciation is [pəti]. How are we to understand this quite typical example of context-sensitive phonological alternation?

It might well prove interesting to build a connectionist model that learned the contextually-appropriate pronunciation of this word—and others like it—from examples. Suppose this done. We are now positioned to address some very interesting questions about data-driven acquisition of phonology. What kind of initial structure in the representations and architecture of the network, and what kind of data, allows learning procedures to master this pattern of alternation?

Another kind of question is extremely important, however: putting aside questions about the learning process itself, we pass to questions of *exactly what knowledge* the network has acquired. That is, we now want an explanation of just how the word *petit* is represented (in the model's 'lexicon'), and how the network's connections (its 'grammar') manage to produce the right pronunciation at the right time.

Why is this type of question important? Isn't it enough to show that the trained network generalizes correctly to, say, 96% of unfamiliar words? Do we really need to characterize the 'lexicon' and 'grammar' that the network has learned?

Perhaps not. But in that case, we can't conclude that connectionism has provided an *alternative* way to *explain* acquisition and use of knowledge of language.[2] This is so, I believe, for at least the following three reasons.

First, the network may not provide an 'explanation' in an acceptable sense of the word because the degree of success that it does achieve may be due to aspects of the simulation that are not actually theoretical commitments of the implicit connectionist theory of language being proposed (McCloskey, 1992). For example, aspects of the input/output representations may provide important biases, reflecting principles of symbolic linguistic theory, and these biases may be more crucial than any general connectionist principles in allowing the network to perform reasonably well. Thus, it is important to understand what knowledge the network actually has and what the critical sources of that knowledge are.

Second, for all we know, whatever correct performance the network displays, or whatever explanation of the phenomena the account provides, may arise because the network partially implements some symbolic system of linguistic representations and rules. Proponents of symbolic cognitive theories believe that symbolic rules and representations are somehow implemented in neural networks—perhaps they're right, and our network has managed to perform just such an implementation. Or, at least, this may be so to the limited degree required to achieve a 96% success rate on its limited data. Unless we have adequate understanding of the network's knowledge at the higher, more abstract levels where symbolic rules and representations *could possibly* reside, we have no grounds to argue that the network's success counts as evidence *against* symbolic theory.

Third, it is possible (indeed, I believe, highly likely) that what the network has learned has little or no relevance to any general conclusions about linguistics knowledge. Surely it would be a staggering modeling feat to develop a network capable of learning the full phonology of any known human language. In the imaginable future, all models will confront an infinitesimal fraction of this data. The important point to note is that, in contrast, theories of knowledge of language developed in theoretical linguistics *are* informed by, and responsible to, an *overwhelmingly* larger body of cross-linguistic data than any feasible connectionist model.

When a theoretical linguist analyzes a set of data *D* of the scope that might feasibly be presented to a network for training, the linguist's analysis is shaped by a huge mass of *implicit data* in addition to the particular data *D* explicitly under scrutiny. This point will be developed further below, but to briefly illustrate with the case of *petit*, the data informing contemporary symbolic linguistic analysis of the pronunciation of *petit* includes other phonological processes in French that have little superficial relation to the issue of final-consonant pronunciation, the pronunciation of final consonants in Australian languages, the syllabification of word-internal consonants in all known languages, the pronunciation of certain vowels in Slavic languages, and much more. All this is the implicit data that informs the linguist's analysis of *petit*; a proposed analysis is responsible for generalizing, in the relevant respects, to all this implicit data. Thus a large number of logically possible analyses of *D* are rejected by the linguist because such analyses would have no hope of generalizing beyond *D* to other phonological phenomena, and other languages. Successfully accounting for *D* in isolation is usually quite easy; what is hard is accounting for *D* in a way that employs representations and processes that might conceivably generalize beyond *D*.

Now it is imaginable that we have in our heads one network for handling one data set *D*, a separate network for another data set D', yet another for D'', ... Such an approach to linguistic knowledge is hyper-modular, in a sense favored by no proponent of modularity within linguistics; the burden of proof is on the connectionist modeler to show that it could conceivably work and, if so, that it provides the best explanation. Theoretical linguistic methodology, by contrast, imposes the constant methodological constraint that proposed analyses exploit every means of generalization possible, and thus the burden of generality is borne within each analysis, and not by splicing together a collection of separate analyses, each employing knowledge of highly restricted generality. Thus, unless we have some understanding of what knowledge a network is using to cope with a very limited body of data *D*, we have no basis for believing that this knowledge will generalize beyond *D*. Indeed, there would be no reason to believe the network has solved a relevantly difficult problem: accounting for the data *D* in a way that generalizes well beyond *D*. Furthermore, what makes these problems difficult is often a relatively small set of challenging cases; which cases are the challenging ones depends on the analysis. Thus, unless we understand how the network is analyzing *D*, for all we know, the challenging cases amount to only 4% of the data, and these are just the cases our network gets wrong; that is, the network has simply failed to handle the pattern that makes *D* difficult in the first place.

Thus, it seems to me that the model-based strategy provides a researcher two options. The first is to be content with a model that accounts for 96% of its data, and with only very fragmentary understanding of what knowledge the network has that allows it to exhibit this performance. In this case, for the reasons outlined in the previous paragraphs, I believe the model cannot be seen as providing strong evidence one way or the other concerning the question of whether connectionism provides a viable alternative to symbolic linguistic theory.

The second option is to determine what knowledge the network has, and to show (1) that this knowledge derives from genuine principles of a connectionist theory of language, and not from arbitrary implementation details of the simulation; (2) that it is not the case that this knowledge is (partially) successful only because it (partially) implements a symbolic linguistic theory; and (3) that this knowledge encodes regularities of considerable (cross-)linguistic generality, and correctly explains the patterns in the data that make it interestingly challenging to explain. Armed with such a characterization of the knowledge in the network, we could take its success as evidence that connectionist principles lead to knowledge of language that is not an implementation of a symbolic theory, and of sufficient generality to be significant: that is, as evidence that connectionist principles can provide an alternative to symbolic accounts of knowledge of language.

But achieving such a strong understanding of the knowledge residing in a network that is sophisticated enough to perform linguistically complex tasks is a very tall order, well beyond the current state of the art. Unfortunately, while some significant progress has been made over the past decade, our ability to understand the knowledge in networks of any sophistication remains quite rudimentary. So suppose we have trained a network to pronounce *petit* and related French words, achieving 96% generalization to novel words. It seems reasonable to expect—based on the past decade's experience with such experiments—that with a great deal of insight, skill, and persistence (and a bit of luck), a researcher *might* be able to analyze the trained network to the point of being able to explain why, given the connection weights, and the representation of individual words like *petit* in various contexts, the correct pronunciation behavior *mathematically follows*. It does not seem at all likely, however, that a researcher could produce more than highly fragmentary explanations at any greater level of generality about the lexical and grammatical knowledge of the network and its consequent behavior.

Now how does such depth and breadth of explanation of this tiny bit of phonology compare with that provided by basic linguistic theory? To provide some perspective on this question, Table 2 summarizes some aspects of an explanation of *petit*'s behavior that emerge from contemporary generative phonology.

Starting with row *a* of Table 2, "pətit ∼ pəti" means that the two pronunciations of *petit* are alternants; we have already seen a generalization concerning the contexts in which each form appears. In beginning the generative explanation, the first thing to note is that this behavior is not a phenomenon peculiar to French; in as far-flung a language as

TABLE 2
A *petit* Explanation

Level of Generality	French	Lardil
a. word in target language	pətit ~ pəti	
b. word in another language		ŋaluk ~ ŋalu
c. segment in language	$t \sim \varnothing\,]_{\mathrm{Wd}}$ (certain words)	$k \sim \varnothing\,]_{\mathrm{Stem}}$ (always)
d. segment class in language	$C \sim \varnothing\,]_{\mathrm{Wd}}$ (certain words)	$C \sim \varnothing\,]_{\mathrm{Stem}}$ (certain Cs)
e. universal	Patterns of 'defective segments' cross-linguistically; C ~ ∅ in syllable coda cross-linguistically, especially 'worse' Cs	
f. markedness	Cs avoided in syllable coda in: adult inventories, adult phonology, child language, language change . . .	

Lardil, an Australian aboriginal language, we discover that the noun stem for 'story' is pronounced [*η*aluk] when followed by a vowel-initial suffix, but [*η*alu] otherwise: the final *k* behaves like the final *t* of *petit* (row *b*). (Indeed, even English has a tiny case of such behavior, in the indefinite article *an* ~ *a*.) The next observation is that in French this behavior is not limited to *petit*; the final *t* in many—but not all—French words behaves the same way. In row *c*, this is written: '$t \sim \varnothing\,]_{\mathrm{Wd}}$', that is, *t* alternates with silence before the right word-boundary, denoted '$]_{\mathrm{Wd}}$'. Again, the situation in Lardil is parallel: before a stem-boundary, *k* alternates with silence; this time, for all words.

The next observation (row *d*) is that in French this behavior is not limited to *t*: there is a class of final consonants that all behave the same way. Again, in Lardil, *k* is merely one member of a class of consonants that can appear only before a vowel. In Lardil, this class can be loosely characterized as: all consonants specifying a place of articulation different from that of a *simple coronal*—a single occlusion created by the tongue tip in the general area of the alveolar ridge (see Prince & Smolensky, 1993, chap. 7, and the literature cited therein.)

The next step (row *e*) is to observe that the behavior of French final consonants like *t* in *petit* is part of a much more general pattern of 'defective segments' seen in a number of the world's languages; for example, Slavic languages have *vowels* that come and go. The general pattern is that defective segments provide phonological material which can be present or absent, depending on whether the resulting phonological structure would be 'better' with or without that material. In the case of final consonants, the relevant notion of 'better' is this: a consonant is better placed at the beginning of a syllable (the *onset*), rather than the end of a syllable (the *coda*). French defective final consonants appear when they would start a syllable (including the initial vowel of the following word); they disappear when they would fall into a syllable coda. In Lardil, the syllable coda (unlike the onset) can never contain the 'worst' consonants: those specifying a place of articulation different from that of a simple coronal.

The general notion that certain linguistic structures are 'better' than others was developed in the 1930's by Jakobson, Trubetzkoy, and others of the Prague Linguistics Circle under the name *markedness*, the 'worse' structures being called *marked* (as by a scarlet letter; Trubetzkoy, 1939/1969; Jakobson, 1941/1968). Thus, syllable codas are marked relative to syllable onsets; simple coronal consonants are *un*marked relative to other consonants. The Praguian view is that markedness pervades all aspects of language (row *f*). Marked structures are avoided altogether in certain languages, i.e., entirely absent from some languages' inventories of possible structures (e.g., languages in which *no* syllables have codas; Blevins, 1995). In other languages, these same marked structures may appear, but only under strongly restricted conditions (e.g., in Lardil, codas can appear only if they contain unmarked consonants). In such languages, marked structures are avoided 'when possible' through phonological alternations (e.g., the final *t* of *petit*). Furthermore, marked structures are avoided by changes across time to the language itself (in Old French, the final *t* of *petit* was pronounced even in coda position). Moreover, Jakobson believed marked structures to be later-acquired by children; first to be lost in aphasia; and, presumably, harder to process on-line.

Exploiting the notion of markedness, the explanation suggested in Table 2 weaves the tiny thread of *t*'s behavior in *petit* into a web of cross-linguistic empirical generalizations. *t*'s behavior is explained as the avoidance of syllable codas; this in turn is then woven together with the tendency of children to avoid syllable codas, all seen as instances of how avoidance of markedness of all sorts—from syllable codas and non-coronal consonants in phonology through plural number in morphology to passive voice in syntax—pervades all of language: intact and disordered, adult and child.

The notion of linguistic markedness parallels in important respects the notion of (negative) Harmony (or "energy") in connectionist networks. Optimality Theory's basic hypothesis, that the output of the grammar is the structure that maximizes Harmony relative to the given input, can be viewed as a formalization of the notion that languages avoid more marked—less harmonic—structures. Optimality Theory's formal calculus of markedness—a version of parallel soft constraint satisfaction, Harmony maximization—has for the first time enabled markedness to provide the very core of a generative theory of grammar.

At all the levels of explanation identified in Table 2, OT analyses have provided a formal means to deduce from the basic principles of the theory both detailed language-particular patterns and overarching empirical generalizations. In OT, a marked linguistic structure is formalized as one that violates one or more universal well-formedness constraints. To express the universal markedness of syllable codas, a universal constraint on the well-formedness of syllables is posited: NoCoda, 'Syllables do not have codas' (Prince & Smolensky, 1993, chap. 6). Similarly, universal constraints express the universal unmarkedness of simple coronal consonants relative to other consonants (Prince & Smolensky, 1993, chap. 9; Smolensky, 1993). The possible phonetic bases of such constraints in articulation and perception is one facet of OT research, but what concerns us now is not what gives rise to these constraints, but rather, what the constraints give rise to: precise accounting of the forms of particular words in particular contexts in particular languages (rows *a–d*), and the universal generalizations of which language-particular facts are instances (rows *e–f*).

Along with other hypothesized universal constraints, NoCoda enables correct prediction of the different context-dependent phonological forms of *petit* and other French words with final defective consonants (Tranel, 1994). Such analysis formally captures the explanation that such consonants are present or absent depending on which option gives the 'better'—more harmonic, less marked—form, entailing avoidance of Harmony-lowering violations of NoCoda. NoCoda plays a formally parallel role in the prediction of the context-dependent forms of Lardil words like *ŋaluk* (Prince & Smolensky, 1993, chap. 7), formally capturing the commonality shared by final consonants in French and Lardil (rows *a–d* of Table 2).

Moving to a still more universal level (row *e* of Table 2), an OT analysis of the property that distinguishes defective French consonants from ordinary consonants makes it possible to formally deduce a universal typology which spells out the cross-linguistic possibilities for the behavior of defective material (Zoll, 1996). This typology situates French defective consonants in a universal picture that connects them to the defective vowels of Slavic and even to defective material in African languages, consisting only in a pitch tone. In another vein, the Lardil prohibition against placing the least harmonic consonants in the least harmonic syllable position (coda) is seen as one instance of a formal pattern of Harmony maximization—'banning the worst of the worst'—which is evident in many linguistic contexts, including, for example, requirements in African languages that vowels in the same word have the same tongue-root configuration (Prince & Smolensky, 1993; Smolensky, 1993, 1997).

At a yet more general level, the overarching generalizations concerning markedness summarized in row *f* of Table 2 can be formally deduced within OT. Once any markedness-defining constraint C (e.g., NoCoda) is recognized as a universal constraint, the general computational principles of OT take over and many logical consequences follow.

According to the general OT theory of cross-linguistic variation, languages will differ in how highly-ranked C is in their grammars. There will be some languages in which C is very highly ranked, with the effect that marked structures violating C will never be most harmonic; they will never appear in the language. Thus, there will be languages that ban C-violating structures from their inventories of possible structures, but no languages that ban C-satisfying structures. (For the case C = NoCoda, this derives the typological fact that among the world's languages there are some that prohibit syllable codas but none that require them—and just the opposite for syllable onsets; Jakobson, 1962; Prince & Smolensky, 1993, chap. 6).

In other languages, C will be less dominant; C will be outranked by other conflicting constraints that are relevant in some contexts but not in others. In such a language, C-violations will be possible in the language in some contexts, but avoided in others by phonological alternations: marked, C-violating structures will not be optimal, except in those limited contexts where conflicting constraints out-ranking C are relevant.

Furthermore, according to a general OT learning theory (Tesar & Smolensky, 1996, 1998, to appear; Tesar, 1998; Smolensky, 1996c, to appear), markedness-defining constraints C are initially high-ranked, so children's early grammars allow only unmarked structures to be produced; only after such a constraint has been demoted in rank during learning can children produce marked, C-violating structures. (When C = NoCoda, this derives the tendency of children's initial syllables to lack codas.)

OT analysis of language change as re-ranking of constraints over time (e.g., Zubritskaya, 1997) predicts that when a constraint C assumes higher rank, the language loses C-marked structures. And at the frontier of current OT research, in studies of on-line sentence processing, an incremental-optimization theory of parsing connects processing difficulty to relative Harmony or markedness of syntactic structures (Stevenson & Smolensky, 1997), and current studies of phonological errors by aphasic patients seek to model damage as relative promotion of markedness constraints, so aphasic productions lack certain marked structures produced by the intact grammar (Goldrick, Rapp, & Smolensky, 1999).

It will be a long time, surely, before our ability to analyze specific connectionist models trained on specific linguistic data reaches the depths and breadths necessary to deduce from basic principles the range of generalizations summarized in Table 2. But a product of grammar-based connectionist research, Optimality Theory, is already able to do so. And that is the primary reason why—despite the significant limitations on the aspects of connectionist theory that can today be successfully exploited—it seems to me that the Grammar-based Strategy is a valuable complement to the Model-based Strategy in the arsenal of connectionist approaches to language.

VI. SUMMARY

Debates about the relation of connectionism to language often seem to take it for granted that it is possible to identify two broad theories of language, Connectionism and Generative Grammar, and that these theories are locked in deep scientific conflict. I believe this to be quite false. It is not difficult to identify a particular connectionist proposal about language which conflicts with a particular generative proposal about the same aspect of language—the classic debate about the acquisition of the English past tense (Rumelhart & McClelland, 1986; Pinker & Prince 1988; Lachter & Bever, 1988) may be such a case. But, equally, it is easy to identify two connectionist proposals that conflict as accounts of a common phenomenon, and even easier to identify pairs of conflicting generative proposals. There are plenty of disagreements about language to go around; the question is, do the theories divide into two camps, Connectionism and Generative Grammar, with fundamentally incompatible commitments?

The basic commitments of PDP Connectionism are often taken to entail positions concerning modularity, nativism, and other Big Issues, but it seems to me that brief inspection reveals that while a *particular* PDP proposal may entail such commitments, the broad class of PDP models encompasses proposals that span the spectrum of possible positions on the Issues. Equally, I believe the basic commitments of generative grammar entail no positions on the Big Issues, despite the fact that prominent generative linguists, speaking for themselves, have expressed such commitments.

The view I have tried to sketch here is that PDP connectionism is a commitment to fundamental computational mechanisms, and generative grammar is a commitment to certain types of explanations of certain types of empirical generalizations. These commitments are not in conflict. As I believe recent research shows, it is possible to deploy at least some of the computational mechanisms of PDP connectionism to advance the explanatory goals of generative grammar.

Doing so is not possible using the standard Model-based Strategy for developing concrete proposals within the broad compass of PDP connectionism—at least for the foreseeable future. This is because the types of explanations demanded by generative grammar are not currently feasible within this strategy.

Of course, the Model-based Strategy has produced important advances in cognitive science and will, I believe, continue to do so for a long time to come. While it may not advance the goals of generative grammar, model-based research serves other scientific goals that seem to me at least as important. (For example, I believe it has vastly improved our understanding of data-driven learning.)

But the goals of generative grammar can in fact be advanced by a different, Grammar-based Strategy for developing PDP proposals concerning language. In this strategy, new grammatical theories based upon fundamental PDP computational mechanisms are developed, and replace traditional grammatical theories based upon computational mechanisms such as serial symbol manipulation and hard-constraint satisfaction.

In these early days of exploring the potential of the Grammar-based Strategy for connectionist linguistics, we have succeeded in developing new grammatical formalisms that incorporate only a small part of the full arsenal of connectionist computational principles. But such is the power of the approach that even a small amount of connectionist input, that embodied in Optimality Theory, has already had a major impact on the practice of generative grammar (many OT papers and an extensive OT bibliography can be accessed electronically at the Rutgers Optimality Archive, http://ruccs.rutgers.edu/roa.html).

Looking to the future, two types of prospect can now be discerned. One is the development of new grammatical theories that incorporate additional PDP principles (including theories going beyond the Symbolic Approximation). The other is the advancement of connectionist theories of higher cognition. For Optimality Theory has led to empirical discoveries about universal grammar that are a surprise from the current PDP perspective (6), and these discoveries seem to be telling us that current PDP computational principles are missing something—something quite important for language, at least. It will be most interesting to see what this turns out to be.

Acknowledgments: I am grateful to Géraldine Legendre and Bob Frank for very helpful comments on earlier drafts, to Alan Prince for insightful discussion of the issues, to the editors of the Special Issue for their useful suggestions, and to NSF for financial support through the Learning and Intelligent Systems Initiative.

NOTES

[1] The situation seems highly parallel to what I see as the relation of connectionism to neuroscience. To the criticism that connectionism ignores a tremendous amount of knowledge about the brain, I would respond: “The hypothesis justifying connectionism is that *some*—not *all*—cognitively-relevant aspects of neural

structure are well-approximated by the abstract computational systems posited by connectionist theory. It is agreed by everyone that such systems do not capture all that cognitive theory wants to capture. The claim is that what they *do* capture has great explanatory power."

2. I do not mean to imply that it is incumbent upon all connectionist language research to resolve the question of whether the connectionist account proposed provides an alternative explanation of linguistic generalizations. But if such an alternative explanation is claimed to have been provided by a piece of modeling research, then the concerns expressed in the text are relevant.

REFERENCES

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413−451.

Archangeli, D. (1997). Optimality Theory: An introduction to linguistics in the 1990s. In D. Archangeli & D. T. Langendoen (Eds.), *Optimality theory: An overview* (pp. 1−32). Malden, MA: Blackwell.

Blevins, J. (1995). The syllable in phonological theory. In J. A. Goldsmith, (Ed.), *The handbook of phonological theory*. Cambridge, MA: Blackwell.

Christiansen, M. H. & Chater, N. (1999). Connectionist natural language processing: The state of the art. *Cognitive Science, 23,* 000–000.

Cohen, M. A., & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, 13,* 815−825.

Dolan, C. P. (1989). Tensor manipulation networks: Connectionist and symbolic approaches to comprehension, learning, and planning, Ph.D. thesis. Department of Computer Science, University of California at Los Angeles.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179−211.

Fauconnier, G. (1985). *Mental spaces.* Cambridge, MA: MIT Press.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6,* 205−254.

Golden, R. M. (1988). A unified framework for connectionist systems. *Biological Cybernetics, 59,* 109−120.

Goldrick, M., Rapp, B., & Smolensky, P. (1999). Lexical and postlexical processes in spoken word production. Presented at the Academy of Aphasia Annual Meeting, Venice, Italy. October 25.

Goldsmith, J., & Larson, G. (1990). Local modeling and syllabification. In K. Deaton, M. Noske, & M. Ziolkowski (Eds.), *Papers from the 26th Annual Regional Meeting of the Chicago Linguistic Society* (Part 2). Chicago: Chicago Linguistic Society.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: Part I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23,* 121−134.

Hinton, G. E. (1986). Learning distributed representations of concepts. In *Proceedings of the 8th Annual Conference of the Cognitive Science Society* (pp. 1−12) Hillsdale, NJ: Lawrence Erlbaum.

Hinton, G. E., & Sejnowski, T. J. (1983). Optimal perceptual inference. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 448−453).

Hinton, G. E. & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann Machines. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Foundations* (Vol. 1, pp. 282−317). Cambridge, MA: MIT Press/Bradford Books.

Hirsch, M. W. (1996). Dynamical systems. In P. Smolensky, M. C. Mozer, & D. E. Rumelhart (Eds.), *Mathematical perspectives on neural networks* (pp. 271−323). Mahwah, NJ: Lawrence Erlbaum.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences USA, 79,* 2554−2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences USA, 81,* 3088−3092.

Hummel, H. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480−517.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 15,* 219−250.

Jakobson, R. (1941/1968). *Child language, aphasia and phonological universals.* The Hague: Mouton.

Jakobson, R. (1962). Selected writings I: Phonological studies. The Hague: Mouton.

Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. In *Proceedings of the 8th Annual Conference of the Cognitive Science Society* (pp. 10−17) Hillsdale, NJ: Lawrence Erlbaum.

Kohonen, T. (1977). Associative memory: A system theoretical approach. NY: Springer.

Lachter, J. & Bever, T. G. (1988). The relation between linguistic structure and associative theories of language learning—A constructive critique of some connectionist learning models. *Cognition, 28,* 195−247.

Langacker, R. W. (1987). *Foundations of cognitive grammar, Vol. 1: Theoretical prerequisites.* Stanford, CA.: Stanford University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Lakoff, G. (1993). Cognitive phonology. In J. Goldsmith (Ed.), *The last phonological rule* (pp. 117–145). Chicago: University of Chicago Press.

Legendre, G., Miyata, Y., & Smolensky, P. (1990a). Harmonic Grammar—A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 388−395) Hillsdale, NJ: Lawrence Erlbaum.

Legendre, G., Miyata,, Y. & Smolensky, P. (1990b). Harmonic grammar —a formal multi-level connectionist theory of linguistic well-formedness: An application. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 884−891) Hillsdale, NJ: Lawrence Erlbaum.

Legendre, G., Smolensky, P., & Wilson, C. (1998). When is less more? Faithfulness and minimal links in *wh*-chains. In Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M., & Pesetsky, P. (Eds.) *Is the best good enough? Optimality and competition in syntax* (pp. 249–289). MIT Press and MIT Working Papers in Linguistics. [Available as Technical report JHU-CogSci-96/7, Department of Cognitive Science, Johns Hopkins University, 1996. Rutgers Optimality Archive, ROA-117.].

Levin, B. and Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax-semantics interface.* Cambridge, MA: MIT Press.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Psychological and biological models*(Vol. 2). Cambridge, MA: MIT Press/Bradford Books.

McCloskey, M. (1992). Networks and theories: The place of connectionism in cognitive science. *Psychological Science, 2,* 387−395.

McMillan, C., Mozer, M., & Smolensky, P. (1992). Rule induction through integrated symbolic and subsymbolic processing. In J. Moody, S. Hanson, & R. Lippman, (Eds.), *Advances in neural information processing systems 4* (pp. 969−76.) [Collected papers of the IEEE Conference on Neural Information Processing Systems—Natural and Synthetic, Denver, Nov. 1991.] San Mateo, CA: Morgan Kaufmann.

Mozer, M. C. (1991). *The perception of multiple objects: A connectionist approach.* Cambridge, MA: MIT Press/Bradford Books.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,* 73−193.

Plate, T. (1991). Holographic Reduced Representations: Convolution algebra for compositional distributed representations. In J. Mylopoulos, & R. Reiter (Eds.), *Proceedings of the 12th International Joint Conference on Artificial Intelligence* (pp. 30−35) San Mateo, CA: Morgan Kaufmann.

Plaut, D., & Shallice, T. (1994). *Connectionist modeling in cognitive neuropsychology: A case study.* Hillsdale, NJ: Lawrence Erlbaum.

Pollack, J. B. (1990). Recursive distributed representations. *Artificial Intelligence, 46,* 77−105.

Prince, A. (1993). *In defense of the number i: Anatomy of a linear dynamical model of linguistic generalizations. Technical Report RuCCS-TR-1,* Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ.

Prince, A., & Smolensky, P. (1991). *Notes on connectionism and Harmony Theory in linguistics. Technical Report CU-CS-533-91:* Department of Computer Science, University of Colorado at Boulder. July. [Notes from the course, ‘Connectionism and Harmony Theory in Linguistics,’ LSA Linguistic Institute, University of California, Santa Cruz; July, 1991.]

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar. Technical Report CU-CS-696-93*: Department of Computer Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ.

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science, 275,* 1604−1610.

Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1996). Backpropagation: The basic theory. In P. Smolensky, M. C. Mozer, & D. E. Rumelhart (Eds.), *Mathematical perspectives on neural networks* (pp. 533−566). Mahwah, NJ: Lawrence Erlbaum.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Psychological and biological models* (Vol. 2, pp. 390−431). Cambridge, MA: MIT Press/Bradford Books.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Foundations* (Vol. 1). Cambridge, MA: MIT Press/ Bradford Books.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in parallel distributed processing. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Psychological and biological models* (Vol. 2, pp. 7–57). Cambridge, MA: MIT Press/Bradford Books.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 9,* 75−112.

Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems 1,* 145–168.

Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rule, variables, and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences, 16,* 417−494.

Smolensky, P. (1983). Schema selection and stochastic inference in modular environments. In *Proceedings of the National Conference on Artificial Intelligence* (pp. 378−382). Los Altos, CA: William Kaufmann.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Foundations* (Vol. 1, pp. 194–281). Cambridge, MA: MIT Press/ Bradford Books.

Smolensky, P. (1988). On the proper treatment of connectionism. *The Behavioral and Brain Sciences, 11,* 1−23.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist networks. *Artificial Intelligence, 46,* 159−216.

Smolensky, P. (1993). Harmony, markedness, and phonological activity. Paper presented at the Rutgers Optimality Workshop−1, Rutgers University. [Rutgers Optimality Archive ROA-87]

Smolensky, P. (1996a). Dynamical perspectives on neural networks. In P. Smolensky, M. C. Mozer, & D. E. Rumelhart (Eds.), *Mathematical perspectives on neural networks* (pp. 245−270). Mahwah, NJ: Lawrence Erlbaum.

Smolensky, P. (1996b). Statistical perspectives on neural networks. In P. Smolensky, M. C. Mozer, & D. E. Rumelhart (Eds.), *Mathematical perspectives on neural networks* (pp. 453−495). Mahwah, NJ: Lawrence Erlbaum.

Smolensky, P. (1996c). On the comprehension/production dilemma in child language. *Linguistic Inquiry, 27,* 720−731.

Smolensky, P. (1997). Constraint interaction in generative grammar II: Local Conjunction (or, Random rules in Universal Grammar). Paper presented at the Hopkins Optimality Theory Conference, University of Maryland Mayfest, Baltimore, Maryland.

Smolensky, P. (to appear). The initial state and 'richness of the base' in Optimality Theory. *Linguistic Inquiry*. [Original version available as Technical Report JHU-CogSci-96/4, Cognitive Science Department, Johns Hopkins University, 1996. Rutgers Optimality Archive ROA-154.].

Smolensky, P., & Legendre, G. (in progress). *Toward. a calculus of the mind/brain: Neural network theory, optimality, and universal grammar.*

Smolensky, P., Legendre, G., & Miyata, Y. (1992). *Principles for an integrated connectionist/symbolic theory of higher cognition. Technical Report CU-CS-600-92*, Department of Computer Science and 92–8, Institute of Cognitive Science. University of Colorado at Boulder.

Smolensky, P., Mozer, M. C., & Rumelhart, D. E. (Eds.). (1996). *Mathematical perspectives on neural networks.* Mahwah, NJ: Lawrence Erlbaum.

Stevenson, S., & P. Smolensky, P. (1997). Extending Optimality Theory to comprehension: Competence and performance. Paper presented at the Architectures and Mechanisms for Language Processing Conference, Edinburgh.

Stone, G. O. (1986). An analysis of the delta rule and the learning of statistical associations. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Foundations* (Vol. 1, pp. 444–459). Cambridge, MA: MIT Press/Bradford Books.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*, 49−100.

Tesar, B. & Smolensky, P. (1994). Synchronous-firing variable binding is spatio-temporal tensor product representation. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*. Atlanta, GA. August.

Tesar, B., & Smolensky, P. (1996). *Learnability in optimality theory (long version). Technical Report JHU-CogSci-96-3,* Department of Cognitive Science, Johns Hopkins University, Baltimore, Md. [Rutgers Optimality Archive ROA-156].

Tesar, B. & Smolensky, P. (1998). Learnability in optimality theory. *Linguistic Inquiry, 29,* 229−268.

Tesar, B. & Smolensky, P. (1998). Learning optimality-theoretic grammars. *Lingua, 106,* 161–196.

Tesar, B. (1998). Error-driven learning in optimality theory via the efficient computation of optimal forms. In: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, & D. Pesetsky (Eds.), *Is the best good enough? Optimality and competition in syntax* (pp. 421−435). Cambridge, MA: MIT Press and MIT Working Papers in Linguistics.

Touretzky, D. S. (1986). BoltzCONS: Reconciling connectionism with the recursive structure of stacks and trees. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp, 522−530). Hillsdale, NJ: Lawrence Erlbaum.

Touretzky, D. S., & Hinton, G. E. (1988). A distributed connectionist production system. *Cognitive Science, 12,* 423−466.

Touretzky, D. S., & Wheeler, D. W. (1990). A computational basis for phonology. In D. S. Touretzky (Ed.), *Advances in neural information processing systems* (Vol. 2, pp. 372−379). San Mateo, CA: Morgan Kaufmann.

Tranel, B. (1994). French liaison and elision revisited: A unified account within Optimality Theory. [Rutgers Optimality Archive ROA-15].

Trubetzkoy. (1939/1969). *Principles of phonology* (translation from the German). Berkeley, CA: University of California Press.

Zoll, C. C. (1996). *Parsing below the segment in a constraint based framework*, Ph.D. thesis. Linguistics Department, University of California, Berkeley, CA [Rutgers Optimality Archive ROA-143].

Zubritskaya, K. (1997). Mechanism of sound change in Optimality Theory. *Language Variation and Change, 9,* 121−148.