

Connectionist Natural Language Processing: The State of the Art

MORTEN H. CHRISTIANSEN

Southern Illinois University

NICK CHATER

University of Warwick

This Special Issue on *Connectionist Models of Human Language Processing* provides an opportunity for an appraisal both of specific connectionist models and of the status and utility of connectionist models of language in general. This introduction provides the background for the papers in the Special Issue. The development of connectionist models of language is traced, from their intellectual origins, to the state of current research. Key themes that arise throughout different areas of connectionist psycholinguistics are highlighted, and recent developments in speech processing, morphology, sentence processing, language production, and reading are described. We argue that connectionist psycholinguistics has already had a significant impact on the psychology of language, and that connectionist models are likely to have an important influence on future research.

I. INTRODUCTION

Connectionist modeling of language processing has been highly controversial. Some have argued that language processing, from phonology to semantics, can be understood in connectionist terms; others have argued that *no* aspects of language can be fully captured by connectionist methods. And the controversy is particularly heated because for many, connectionism is not just an additional method for studying language processing, but an *alternative* to the traditional symbolic accounts. Indeed, the degree to which connectionism supplants, rather than complements, existing approaches to language is itself a matter of debate (see the discussion papers in Part II of this issue). Moreover, the debate over connectionist approaches to language is important as a test of the viability of connectionist models of cognition more generally (Pinker & Prince, 1988).

Direct all correspondence to: Morten H. Christiansen, Department of Psychology, Southern Illinois University, Carbondale, IL 62901-6502; E-Mail: morten@siu.edu or nick.chater@warwick.ac.uk.

This Special Issue aims to provide the basis for an appraisal of the current state of play. In Part I, leading connectionists detail the most recent advances within key areas of language research. Gaskell and Marslen-Wilson present statistical simulations exploring the properties of a distributed connectionist model of speech processing. Plunkett and Juola introduce a new model of English past tense and plural morphology. Tabor and Tanenhaus describe their latest progress in using recurrent networks to model sentence processing within a dynamic framework. Dell, Chang and Griffin report on three models of language production, including a novel network model of syntactic priming. Finally, Plaut presents a new connectionist model of sequential processing in word reading. Part II provides an evaluation of the status and prospects of connectionist psycholinguistics from a range of viewpoints. Seidenberg and MacDonald argue for a radical connectionist approach to language acquisition and processing; Smolensky argues for an integration of connectionist and symbolic approaches; and Steedman assesses connectionist sentence processing from the point of view of the symbolic cognitive science tradition.

In this introduction, we aim to set the scene for the Special Issue, providing a brief historical and theoretical background as well as an update on current research in the specific topic areas outlined below. In *Background*, we sketch the historical and intellectual roots of connectionism and outline some of the key debates concerning connectionist psycholinguistics. We then consider the five central topics considered in Part I below: *Speech Processing*, *Morphology*, *Sentence Processing*, *Language Production*, and *Reading*. These topics illustrate the range of connectionist research on language discussed in more depth in the papers in Part I. They also provide an opportunity to assess the strengths and weaknesses of connectionist methods across this range, setting the stage for the general debate concerning the validity of connectionist methods in Part II. Finally, we sum up and consider the prospects for future connectionist research.

II. BACKGROUND

From the perspective of modern cognitive science, we tend to see theories of human information processing as borrowing from theories of machine information processing. Symbolic processing on general purpose digital computers has been the most successful method of designing practical computers. It is therefore not surprising that cognitive science, including the study of language processing, has aimed to model the mind as a symbol processor.

Historically, however, theories of human thought inspired attempts to build computers, rather than the reverse. Mainstream computer science arises from the view that cognition is symbol processing. This tradition can be traced to Boole's (1854) suggestion that logic and probability theory describe "Laws of Thought", and that reasoning in accordance with these laws can be conducted by following symbolic rules. It runs through Turing's (1936) argument that all thought can be modeled by symbolic operations on a tape (i.e., by a Turing machine), through von Neumann's design for the modern digital computer, to modern computer science, artificial intelligence, generative grammar and symbolic cognitive science.

Connectionism¹ (also known as “parallel distributed processing”, “neural networks” or “neuro-computing”) has a different origin, in attempts to design computers inspired by the brain. McCulloch and Pitts (1943) provided an early and influential idealization of neural function. In the 1950s and 1960s, Ashby (1952), Minsky (1954), Rosenblatt (1962) and others designed computational schemes based on related idealizations. These schemes were of interest because these systems *learned* from experience, rather than being designed. Such “self-organizing” or learning machines therefore seemed plausible as models of learned cognitive abilities, including many aspects of language processing (although e.g., Chomsky, 1965, challenged the extent to which language *is* learned). Throughout this period connectionist and symbolic computation stood as alternative paradigms for modeling intelligence, and it was unclear which would prove to be the most successful. But gradually the symbolic paradigm gained ground, providing powerful models in core domains, such as language (Chomsky, 1965), and problem solving (Newell & Simon, 1972). Connectionism was largely abandoned, particularly in view of the limited power of then current connectionist methods (Minsky & Papert, 1969). But, more recently, some of these limitations have been overcome (e.g., Rumelhart, Hinton & Williams, 1986), re-opening the possibility that connectionism constitutes an alternative to the symbolic model of thought.

What does the “neural inspiration” behind connectionism mean in practice? At a coarse level, the brain consists of a very large number of simple processors, neurons, which are densely interconnected into a complex network. These neurons do not appear to tackle information processing problems alone—rather, large numbers of neurons operate simultaneously and co-operatively to process information. Furthermore, neurons appear to communicate numerical values (encoded by firing rate) rather than symbolic messages, and therefore neurons can be viewed as mapping numerical inputs (from other neurons) onto a numerical output (which is transmitted to other neurons). Connectionist nets mimic these properties: They consist of large numbers of simple processors, known as *units* (or nodes), which are densely interconnected into a complex network, and which operate simultaneously and co-operatively; they transmit numerical values; and the output of a unit is usually assumed to be a function of its inputs. But connectionist nets are not realistic models of the brain (see, e.g., Sejnowski, 1986), either at the level of individual processing unit, which drastically oversimplifies and knowingly falsifies many features of real neurons, or in terms of network structure, which typically bears no relation to brain architecture. One research direction is to seek increasing biological realism (e.g., Koch & Segev, 1989). But in the study of aspects of cognition such as language where few biological constraints are available, research has concentrated instead on modeling human behavior. Thus, data is taken from cognitive psychology, linguistics and cognitive neuropsychology, rather than neuroscience. Here, connectionist nets must compete head-on with symbolic models language processing.

We noted that the relative merits of connectionist and symbolic models of language are hotly debated. But should they be in competition at all? Advocates of symbolic models of language processing assume that symbolic processes are somehow implemented in the brain: They too are connectionists, at the level of *implementation*. They assume that

language processing can be described both at the psychological level, in terms of symbol processing, and at an implementational level, in neuroscientific terms (to which connectionism approximates). If this is right, then connectionist modeling should start with symbol processing models of language processing, and implement these in connectionist nets. Advocates of this view (Fodor & Pylyshyn, 1988; Pinker & Prince, 1988) typically assume that it implies that symbolic modeling is entirely autonomous from connectionism; symbolic theories set the goalposts for connectionism, but not the reverse. Chater and Oaksford (1990) have argued that, even according to this view, there will be a two-way influence between symbolic and connectionist theories, since many symbolic accounts can be ruled out precisely because they could not be neurally implemented to run in real time. But most connectionists in the field of language processing have a more radical agenda: To challenge, rather than reimplement, the symbolic approach.

Before discussing research in the key domains discussed in this Special Issue, we set out some recurring themes in discussion of the value of the connectionist approach to language:

Learning

Connectionist nets typically learn from experience, rather than being fully prespecified by a designer. By contrast, symbolic models of language processing are typically fully prespecified and do not learn.

Generalization

Few aspects of language are simple enough to be learned by rote. The ability to generalize to novel cases is thus a critical test for many connectionist models.

Representation

Because connectionist nets learn, their internal codes are devised by the network to be appropriate for the task. Developing methods for understanding these codes is an important research problem. Whereas internal codes may be learned, the inputs and outputs to a network generally use a code specified by the designer. The choice of code can be crucial in determining network performance. How these codes relate to standard symbolic representations of language is contentious.

Rules versus Exceptions

Many aspects of language exhibit “quasi-regularities”—regularities which usually hold, but which admit exceptions. In a symbolic framework, quasi-regularities may be captured by symbolic rules, associated with explicit lists of exceptions. Symbolic processing models often incorporate this distinction by having separate mechanisms for regular and exceptional cases. In contrast, connectionist nets may provide a single mechanism which

can learn general rules, and their exceptions. The viability of such “single route” models has been a major point of controversy, although it is not intrinsic to connectionism. One or both separate mechanisms for rules and exceptions could themselves be modeled in connectionist terms (Pinker, 1991; Coltheart, Curtis, Atkins & Haller, 1993). A further question is whether networks really learn rules at all, or merely approximate rule-like behavior. Opinions differ on whether the latter is an important positive proposal, which may lead to a revision of the role of rules in linguistics (Rumelhart & McClelland, 1986; Smolensky, 1988; but cf. Smolensky, this issue), or whether it is fatal to connectionist models of language (Pinker & Prince, 1988).

With these general issues in mind, we consider the five core domains which are the focus of discussion in Part I of this Special Issue.

III. SPEECH PROCESSING

Connectionist modeling of speech processing was initiated by the influential TRACE model (McClelland & Elman, 1986). This model has an *interactive activation* architecture: It consists of a sequence of “layers” of units. Units in the first layer are specific to phonetic features, units in the second layer to phonemes, and units in the third layer to words. Within and between layers, there are inhibitory connections between units which stand for incompatible states of affairs. For example, there are inhibitory connections between word units, so that “candidate” words compete. Similarly, excitatory connections exist between units that stand for mutually reinforcing states of affairs. In addition to the standard interactive activation architecture, which we shall encounter repeatedly below, TRACE includes a feature to deal with the temporal dimension of speech: There are many copies of the entire network, standing for different points in time in the utterance, with appropriate connections between the units in each copy. Unlike later models, TRACE is completely prespecified (i.e., it does not learn).

The *interactive* character of TRACE embodies a controversial theoretical claim. Many researchers assume that speech processing involves the successive computation of increasingly abstract levels of representation, and assume no feedback from more abstract to less abstract levels. This kind of account is sometimes known as “bottom-up” and can also be realized in connectionist networks, as we shall see below. TRACE, by contrast, allows information to flow both bottom-up and top-down. Whether speech processing is bottom-up or interactive is highly controversial, and the same debate rages in the reading literature and throughout perception (e.g., Fodor, 1983; Marr, 1982).

TRACE captures a wide range of empirical data, such as the apparent influence of lexical context on phoneme identification, and the categorical aspects of phoneme perception. In addition, TRACE makes empirical predictions which appear to be incompatible with any bottom-up model. In natural speech, the pronunciation of a phoneme is altered by the surrounding phonemes: This is known as coarticulation. The speech processing system takes account of this in phoneme recognition—this is called “compensation for coarticulation” (CFC). CFC appears to provide a way of detecting whether lexical information feeds back, top-down, to the phoneme level. Elman and McClelland

(1988) considered CFC *across* word boundaries, for example, a word-final /s/ influencing a word-initial /t/ as in *Christmas tapes*. If the lexical level feeds back to phoneme level, the compensation of the /t/ should still occur when the /s/ relies on lexically driven phoneme restoration for its identity (i.e., in an experimental condition in which the identity of /s/ in *Christmas* is obscured, the /s/ should be restored and thus CFC should proceed as normal). TRACE does indeed make this prediction; and it is not obvious that a bottom-up account of speech perception could make the same prediction. Elman and McClelland (1988) conducted the crucial experiment and confirmed TRACE's prediction.

It has recently been argued, however, that bottom-up connectionist models can, despite appearances, capture these results. Norris (1993) trained a *simple recurrent network* (SRN) (introduced by Elman, 1990—see Steedman, this issue, for a description of this architecture) on input and output consisting of words (from a 12 word lexicon) presented one phoneme at a time. Input phonemes were represented by vectors of phonetic features, and these features could have intermediate values, corresponding to ambiguous phonemes. The output layer consisted of one unit for each phoneme. When the net received input with an ambiguous first word-final phoneme and ambiguous initial segments of the second word, a parallel to CFC was observed: The percentages of /t/ and /k/ responses to the first phoneme of the second word depended on the identity of the first word, as in Elman and McClelland's experiment. But the explanation for this pattern of results cannot be top-down influence from units representing words, because the net has no words units and is, in any case, purely bottom-up.

Norris' small scale example is suggestive, but the question remains: Would a bottom-up net trained on natural speech show the same effect? Shillcock, Lindsey, Levy and Chater (1992) trained a recurrent network (a close variant of the SRN) on phonologically transcribed conversational English, where inputs and outputs to the network were represented in terms of phonetic features. As in Norris' simulations, there was no lexical level of representation, and processing was strictly bottom-up. Nonetheless, phoneme restoration followed the pattern that Elman and McClelland explained by lexical influence.

How can bottom-up processes mimic lexical effects? Shillcock et al. (1992) argue that restoration occurs on the basis of statistical regularities at the phonemic level, rather than lexical influence. It just happens that the words used in Elman and McClelland's (1988) experiment were more statistically regular at the phonemic level than the non-words with which they were contrasted. This was confirmed by a statistical analysis of the corpus of natural speech on which Shillcock et al.'s net was trained.

Further evidence for the ability of bottom-up models to accommodate apparently lexical effects on speech processing was provided by Gaskell, Hare and Marslen-Wilson (1995). They trained an SRN version of the Shillcock et al. model to map a systematically altered featural representation of speech onto a canonical representation of the same speech, and found that the network showed evidence of lexical abstraction (i.e., tolerating systematic phonetic variation, but not random change). More recently, Gaskell and Marslen-Wilson (1997) have added a new dimension to the debate, presenting an SRN network in which sequentially presented phonetic input for each word were mapped onto corresponding distributed representations of phonological surface form and semantics.

Based on the ability of the network to model the integration of partial cues to phonetic identity and the time course of lexical access, they suggested that distributed models may provide a better explanation of speech perception than their localist counterparts (e.g., TRACE). An important challenge for such distributed models is to accommodate the simultaneous activation of multiple lexical candidates necessitated by the temporal ambiguity of the speech input (e.g., /kæp/ could be the beginning of both *captain* and *captive*). The coactivation of several lexical candidates in a distributed model results in a semantic “blend” vector. Through statistical analyses of vector spaces, Gaskell and Marslen-Wilson (this issue) investigate the properties of such semantic blends, and apply the results to explain some recent empirical speech perception data.

The interactive *vs.* bottom-up debate illustrates how the introduction of connectionist models has led to unexpected theoretical predictions, and promoted further empirical research seeking to provide definitive evidence for either the interactive (e.g., Samuel, 1997) or the bottom-up approach (e.g., Pitt & McQueen, 1998).

IV. MORPHOLOGY

One of the connectionist models that has created the most controversy is Rumelhart and McClelland's (1986) model of the learning of English past tense. The English past tense is a quasi-regular mapping, traditionally assumed to require two symbolic routes. This dual route account appears to be backed up by a U-shaped pattern of acquisition—oversimplifying, children appear initially correct with irregulars, then fail due to over-regularization, and finally, re-establish the irregulars correctly. This has traditionally been explained by assuming that the child initially uses a memorization route, which is then overtaken by a rule-based route, and finally the correct balance between the two is established.

Rumelhart and McClelland (1986) argued that this pattern can, however, be explained using a single processing route. They trained a single layer network to map from roots to past tense forms of words, using the perceptron learning algorithm (Rosenblatt, 1962). They used a “wickelfeature” representation, which encodes triples of consecutive elements in the phoneme string.

When first trained on 10 verbs, and then exposed to 420 verbs, the net approximated a U-shaped learning curve. After the first training stage, the network performed perfectly on the 10 verbs getting irregulars and regulars correct. Early in the second stage, it tended to regularize irregular verbs while getting regulars correct. Finally, towards the end of training the network again approached perfect performance on all verbs.

The model has, however, faced considerable criticism. First, the wickelfeature representation has been attacked (e.g., Pinker & Prince, 1988), and later models have switched to other styles of representation. Second, and more fundamental, the U-shaped learning appears to be an artifact of suddenly increasing the total number of verbs (from 10 to 420), a discontinuity which has no developmental justification (Pinker & Prince, 1988).

Plunkett and Marchman (1991), however, have shown U-shaped learning for a net trained with a *fixed* training set. They used a feed-forward network with a hidden unit

layer, trained on a vocabulary of artificial verb stems and past tense forms, patterned on regularities in the English past tense. With a constant training vocabulary, they obtained classical U-shaped learning, and also observed various selective micro U-shaped developmental patterns found in children's behavior. For example, the net was able to simulate a number of subregularities between the phonological form of a verb stem and its past tense form (e.g., *sleep* → *slept*, *keep* → *kept*) not captured by Rumelhart and McClelland's (1986) model. Subsequently, Plunkett and Marchman (1993) also obtained similar results using an incremental, and perhaps more realistic, training regime. Following initial training on 20 verbs, the vocabulary was gradually increased to 500 verbs. This incremental training regime significantly improved the net's overall performance. This work also suggested an intriguing theoretical claim: That a critical mass of verbs is needed before a change from rote-learning (memorization) to system-building (rule-like generalization behavior) can occur. Plunkett and Juola (this issue) find a similar critical mass effect in their model of English noun and verb morphology. They analyzed the developmental trajectory of a feed-forward network trained to produce the plural form for 2280 nouns, and the past tense form for 946 verbs. The model exhibited patterns of U-shaped development for both nouns and verbs (with noun inflections acquired earlier than verb inflections), and also demonstrated a strong tendency to regularize deverbal nouns and denominal verbs.

However, Prasada and Pinker (1993) have argued that connectionist models implicitly depend on an artifact of the idiosyncratic frequency statistics of English. They focus on the *default* inflection of words (e.g., *-ed* suffixation of English regular verbs). The default inflection of a word is assumed to be independent of its phonological shape and to occur unless the word is marked as irregular. Prasada and Pinker argue that connectionist models generalize according to frequency and surface similarity. Regular English verbs have a high type frequency but a relatively low token frequency, allowing a network to construct a broadly defined default category. By contrast, irregulars have low type frequency and high token frequency, which permits the memorization of the irregular past tenses in terms of a number of narrow phonological subcategories (e.g., one for the *i-a* alternation in *sing* → *sang*, *ring* → *rang*, another for the *o-e* alternation in *grow* → *grew*, *blow* → *blew*, etc.). Prasada and Pinker (1993) show that the default generalization in Rumelhart and McClelland's (1986) model depends on a similar frequency distribution in the model's training set. They furthermore contend that no connectionist model can accommodate default generalization for a class of words which have both low type frequency and low token frequency, such as the default inflection of plural nouns in German (see Clahsen, Rothweiler, Woest & Marcus, 1993; Marcus, Brinkmann, Clahsen, Wiese, Woest & Pinker, 1995). If true, such lack of cross-linguistic validity would pose serious problems for connectionist models of morphology.

However, recent connectionist work has addressed minority defaults. Hare, Elman and Daugherty (1995) trained a multi-layer feed-forward network (with additional "clean-up" units—see Plaut, this issue, for an explanation) to map between phonological representations of stems and past tenses for a set of verbs representative of very early Old English. The training set consisted of five classes of irregular verbs plus one class of regular

verbs—each class containing the same number of words. Thus, words taking the default generalization *-ed* formed a minority (i.e., only 17%). But the net learned the appropriate default behavior even when faced with a low-frequency default class. Indeed, it appears that generalization in neural networks may not be strictly dependent on similarity to known items. Hare et al.'s results show that if the non-default (irregular) classes have a sufficient degree of internal structure, default generalization may be promoted by the lack of similarity to known items.

Moreover, Hahn and Nakisa (in press) provide problems for the dual route approach. They compared connectionist and other implementations of rule and memorization routes, against a single memorization route, and found that performance was consistently superior when the rule-route was *not* used, on a comprehensive sample of German nouns.

Finally, rule-like and frequency-independent default generalization may not be as pressing a problem for connectionist models as Clahsen et al. (1993) and Marcus et al. (1995) claim. Reanalyzing data concerning German noun inflection (in combination with additional data from Arabic and Hausa), Bybee (1995) showed that default generalization is sensitive to type frequency and does not seem to be entirely rule-like. This pattern may fit better with the kind of default generalization in connectionist nets rather than with the rigid defaults of symbolic models.

The issue of whether humans employ a single, connectionist mechanism for morphological processing is far from settled. Connectionist models fit a wide range of developmental and linguistic data. And even opponents of connectionist models typically concede that a connectionist mechanism may explain the complex patterns found in the “irregular” cases. The controversial question is whether a single connectionist mechanism can simultaneously deal with both regular and the irregular cases, or whether the regular cases can only be generated by a distinct route involving symbolic rules. Future work is likely to involve further connectionist modeling of cross-linguistic phenomena as well as more detailed fits with developmental data.

V. SENTENCE PROCESSING

Sentence processing provides a considerable challenge for connectionist research. In view of the difficulty of the problem, much early work “hand-wired” symbolic structures into the network architecture (e.g., Fianty, 1985; McClelland & Kawamoto, 1986; Miyata, Smolensky & Legendre, 1993; Small, Cottrell & Shastri, 1982). Such connectionist *re-implementations* of symbolic systems can have interesting computational properties and may be illuminating regarding the viability of a particular style of symbolic model for distributed computation (Chater & Oaksford, 1990). But most connectionist research has a larger goal: To provide alternatives to symbolic accounts of syntactic processing.

Two classes of models potentially provide such alternatives, both of which *learn* to process language from experience, rather than implementing a prespecified set of symbolic rules. The first, less ambitious, class (e.g., Hanson & Kegler, 1987; Howells, 1988; Stolcke, 1991) learns to parse “tagged” sentences. Nets are trained on sentences, each associated with a particular grammatical structure, and the task is to assign the appropriate

grammatical structures to novel sentences. Thus, much linguistic structure is not learned by observation, but is built into the training items. These models are related to statistical approaches to language learning such as stochastic context-free grammars (e.g., Charniak, 1993) in which probabilities of grammar rules in a prespecified context-free grammar are learned from a corpus of parsed sentences.

The second, more ambitious, class of models, which includes Tabor and Tanenhaus (this issue), attempts the much harder task of learning syntactic structure from sequences of words. The most influential approach of this kind is due to Elman (1991, 1993), who trained an SRN to predict the next input word, for sentences generated by a small context-free grammar. This grammar involved subject noun/verb agreement, variations in verb argument structure (i.e., intransitive, transitive, optionally transitive), and subject and object relative clauses (allowing multiple embeddings with complex long-distance dependencies). These simulations suggested that an SRN can acquire some of the grammatical regularities underlying a grammar. In addition, Elman's SRN showed some similarities with human behavior on center-embedded structures. Christiansen (1994, 1999) extended this work, using more complex grammars involving prenominal genitives, prepositional modifications of noun phrases, noun phrase conjunctions, and sentential complements, in addition to the grammatical features used by Elman. One of the grammars moreover incorporated cross-dependencies, a weakly context-sensitive structure found in Dutch and Swiss-German. Christiansen found that SRNs could learn these more complex grammars, and moreover, that the SRNs exhibit the same qualitative processing difficulties as humans do on similar constructions. The nets also showed sophisticated generalization abilities, overriding local word co-occurrence statistics while complying with structural constraints at the constituent level (Christiansen & Chater, 1994).

Current models of syntax typically use "toy" fragments of grammar and small vocabularies. Aside from raising the question of the viability of scaling-up, this makes it difficult to provide detailed fits with empirical data. Nonetheless, some attempts have recently been made toward fitting existing data and deriving new empirical predictions from the models. For example, Tabor, Juliano and Tanenhaus (1997) present an SRN-based dynamic parsing model which fits reading time data concerning the interaction of lexical and structural constraints on the resolution of temporary syntactic ambiguities (i.e., garden path effects) in sentence comprehension. MacDonald and Christiansen (in press) provide SRN simulations of reading time data concerning the differential processing of singly center-embedded subject and object relative clauses by good and poor comprehenders. Finally, Christiansen (1999; Christiansen & Chater, 1999) describes an SRN trained on recursive sentence structures, which fits grammaticality ratings data from several behavioral experiments. He also derives novel predictions about the processing of sentences involving multiple prenominal genitives, multiple prepositional phrase modifications of nouns, and doubly center-embedded object relative clauses, which have subsequently been empirically confirmed (Christiansen & MacDonald, 1999).

Overall, connectionist models of syntactic processing are at an early stage of development. Tabor and Tanenhaus (this issue) advance this work by extending their dynamic parsing model to account for some recent empirical findings concerning semantic effects

in sentence processing. They also propose a new approach to the distinction between syntactic and semantic incongruity. However, more research is required to decide whether promising initial results can be scaled up to deal with the complexities of real language, or whether a purely connectionist approach is beset by fundamental limitations, so that connectionism can only succeed by providing reimplementations of symbolic methods (see the papers in Part II of this issue for further discussion).

VI. LANGUAGE PRODUCTION

In connectionist research, as in the psychology of language in general, there is relatively little work on language *production*. However, some important steps have been taken, most notably by Dell and colleagues. Dell's (1986) spreading activation model was an important early production model. The model was presented as a sentence production model, but high level morphological, syntactic, and semantic processing were not implemented. The implemented part of the model was concerned with moving from the choice of word to be spoken, to finding the phonological encoding of that word.

Dell (1986) used an interactive activation network, like the TRACE model, described above. The net has layers corresponding to morphemes (or lexical nodes), syllables, rimes and consonant clusters, phonemes, and phonetic features. To a first approximation, the nodes are connected bi-directionally between layers, but with no lateral connections within layers. Whereas processing in the TRACE model begins bottom-up from speech input, Dell's model begins top-down, with the activation of a lexical node. Activation then spreads down the network, and then upwards via the feedback connections. At a fixed time (determined by the speaking rate), the nodes with the highest activations are selected for the onset, vowel, and coda slots.

The model accounted for a variety of common speech errors, such as substitutions (e.g., *dog* → *log*), deletions (*dog* → *og*), and additions (*dog* → *drog*). Errors occur when an incorrect node is selected because it becomes more active than the correct node (given the activated lexical node). This may occur due to the feedback connections activating nodes other than those directly corresponding to the initial word node (due to the general spread of activation and differences in resting levels). Alternatively, other words activated as a product of internal noise may interfere with the processing of the network. The model made quantitative predictions concerning the retrieval of phonological forms during production, some of which were later confirmed experimentally (Dell, 1988). More recently, the model has been extended to simulate aphasia (Dell, Schwartz, Martin, Saffran & Gagnon, 1997; Martin, Dell, Saffran & Schwartz, 1994; see also Dell et al., this issue).

Dell's model has had considerable impact on subsequent accounts of speech production, both connectionist (e.g., Harley, 1993) and symbolic (e.g., Levelt, 1989). But the model has limitations, most obviously that it cannot learn. This is psychologically unattractive because lexical information is language-specific, and therefore cannot be innate. Moreover, the inability to learn makes it practically difficult to scale-up the model because each connection must be hand-coded. This problem is addressed by a recent

SRN-based model (Dell, Juliano & Govindjee, 1993) which learned to map lexical items to sequences of phonological segments. The SRN had a small additional modification: The current *output* was “copied back” as additional input to the network (Jordan, 1986) along with the current state of the hidden units. Dell et al. (1993) could account for speech error data without having to build syllabic frames and phonological rules into the network (see Dell et al., this issue, for further discussion; but cf. Dell, Burger & Svec, 1997).

The account of syntactic priming as implicit learning presented in Dell et al. (this issue) can be seen as an extension of this work. This model was trained to generate words given an input message encoding a given proposition using blocks of semantic features (e.g., CHILD and MALE), event roles (e.g., *agent* and *patient*), and action descriptions (e.g., GIVING and WALKING). Dell et al. (this issue) simulated syntactic priming by allowing learning to occur *during* testing. By contrast, most other connectionist models have learning disabled during testing. The ongoing learning created sufficiently robust short-term changes in weight space to ensure priming—even across 10 unrelated sentences. Although the current model focuses on grammatical encoding, it is couched within a broader theoretical framework which provides a first step toward an integrated connectionist account of sentence comprehension and production.

Connectionist models of language production have modeled empirical data on both normal and impaired performance, contributed to fundamental theoretical debates and generated new experimental work. It seems likely that connectionist language production models will have an important role in shaping future research on speech production.

VII. READING

The psychological processes engaged in reading are extremely complex and varied, ranging from early visual processing of the printed word, to syntactic, semantic and pragmatic analysis, to integration with general knowledge. Connectionist models have concentrated on simple aspects of reading: 1) recognizing letters and words from printed text, and 2) word “naming”—i.e., mapping visually presented letter strings onto sequences of sounds. We focus on models of these two processes here.

One of the earliest connectionist models was McClelland and Rumelhart’s (1981) interactive activation model of visual word recognition (see also Rumelhart & McClelland, 1982). This network has three layers of units standing for visual features of letters, letters (in particular positions within the word), and words, and uses the same principles as TRACE, described above, but without the need for a temporal dimension, as the entire word is presented at once.

Word recognition occurs as follows: A visual stimulus is presented, which activates in a probabilistic fashion visual feature units in the first layer. As the features become activated, they send activation via their excitatory and inhibitory connections to the letter units, which, in turn, send activation to the word units. The words compete via their inhibitory connections, and reinforce their component letters via excitatory feedback to the letter level (there is no word-to-letter inhibition). Thus an “interactive” process occurs: Bottom-up information from the visual input is combined with the top-down information

flow from the word units. This process involves a cascade of overlapping and interacting processes: Letter and word recognition do not occur sequentially, but overlap and are mutually constraining.

This model accounted for a variety of phenomena, mainly concerning context effects on letter perception. For example, it captures the facilitation of letter recognition in the context of a word, in comparison to recognition of single letters, or letters embedded in random letter strings. This occurs because partially active words provide a top-down confirmation of the letter identity, and thus they “conspire” to enhance recognition. Similarly, the model explains how degraded letters can be disambiguated by their letter context, how words can be recognized even when *all* their component letters are visually ambiguous, and a range of other effects.

Recent connectionist models of reading have focussed not on word recognition but on word naming, which involves relating written word forms to their pronunciations. The first such model was Sejnowski and Rosenberg’s (1987) NETtalk, which learns to read aloud from text. NETtalk is a two-layer feed-forward net, with input units representing a “window” of consecutive letters of text, and output units representing the network’s suggested pronunciation for the middle letter. The network pronounces a written text by shifting the moving input window across the text, letter by letter, so that the central letter to be pronounced moves onwards a letter at a time. In English orthography, there is not, of course, a one-to-one mapping between letters and phonemes. NETtalk relies on a rather ad hoc strategy to deal with this: In clusters of letters realized by a single phoneme (e.g. “th”, “sh”, “ough”), only one letter is chosen to be mapped onto the speech sound, and the others are not mapped onto any speech sound. NETtalk learns from exposure to text associated with the correct pronunciation using back-propagation (Rumelhart, Hinton & Williams, 1986). Its pronunciation is good enough to be largely comprehensible when fed through a speech synthesizer.

NETtalk was intended as a demonstration of the power of neural networks. The first detailed psychological model of word naming was provided by Seidenberg & McClelland (1989). They also used a feedforward network with a single hidden layer, but they represented the entire written form of the word as input, and the entire phonological form as output. The net was trained on 2897 monosyllabic English words, rather than dealing with unrestricted text like NETtalk. Inputs and outputs used the wickelfeature type of representation, which proved controversial in the context of past tense models, as discussed above.

The net’s performance captured a wide range of experimental data (on the reasonable assumption that the net’s error can be mapped onto response time in experimental paradigms). As with the past tense debate above, a controversial claim concerning this reading model was that it uses a single route to handle quasi-regular mappings. This contrasts with the standard view of reading, which assumes that there are two (non-semantic) routes in reading; a “phonological route”, which applies rules of pronunciation, and a “lexical route” which is simply a list of words and their pronunciations. Regular words can be read using either route; but irregulars must be read by using the lexical route; and non-words must use the phonological route (these will not be known by the lexical

route). Seidenberg and McClelland (1989) claim to have shown that this dual route view is not necessarily correct, because their single route can pronounce both irregular words and non-words. Moreover, they have provided a fully explicit computational model, while previous dual-route theorists had merely sketched the reading system at the level of “boxes and arrows.”

A number of criticisms have been leveled at Seidenberg and McClelland’s account. Besner, Twilley, McCann and Seergobin (1990) have argued that its non-word reading is actually very poor compared with people (though, see Seidenberg & McClelland, 1990). Moreover Coltheart et al. (1993) argued that better performance at non-word reading can be achieved by symbolic learning methods, using the same word-set as Seidenberg and McClelland. Another limitation of the Seidenberg and McClelland model is the use of (log) frequency compression during training. Recently, however, Plaut, McClelland, Seidenberg and Patterson (1996) have shown that a feed-forward network using *actual* word frequencies in the learning process can achieve human level of performance on both word and non-word pronunciation.

As in the past tense debate, the wickelfeature representation has been criticized, leading to alternative representational schemes. For example, Plaut and McClelland (1993) use a localist code which exploits regularities in English orthography and phonology to avoid a completely position-specific representation. Their model learns to read non-words very well, but it does so by building in a lot of knowledge into the representation, rather than having the network learn this knowledge. One could plausibly assume (cf. Plaut et al., 1996) that some of this knowledge is acquired prior to reading acquisition; that is, children normally know how to pronounce words (i.e., talk) before they start learning to read. This idea is explored by Harm, Altmann and Seidenberg (1994) who showed that pretraining a network on phonology can help learning the mapping from orthography to phonology.

One problem with this representational scheme is, however, that it only works for monosyllabic words. Bullinaria (1997), on the other hand, also obtains very high non-word reading performance for words of any length. He gives up the attempt to provide a single route model of reading, and aims to model the phonological route, using a variant of NETtalk, in which orthographic and phonological forms are not pre-aligned by the designer. Instead of having a single output pattern, the network has many output patterns corresponding to all possible alignments between phonology and orthography. All possibilities are considered, and the one that is nearest to the network’s *actual* output is taken as the correct output, and used to adjust the weights. This approach, like NETtalk, uses an input window which moves gradually over the text, producing one phoneme at a time. Hence, a simple phoneme-specific code can be used; the order of the phonemes is implicit in the order in which the network produces them.

A further difficulty for Seidenberg and McClelland’s model is the apparent double dissociation between phonological and lexical reading in acquired dyslexics: Surface dyslexics can read exception words, but not non-words; but phonological dyslexics can pronounce non-words but not irregular words. The standard (although not certain) inference from double dissociation to modularity of function suggests that normal non-word and exception word reading are subserved by distinct systems—leading to a dual-route

model (e.g., Morton & Patterson, 1980). Acquired dyslexia can be simulated by damaging Seidenberg and McClelland's network in various ways (e.g., removing connections or units). Although the results of this damage do have neuropsychological interest (Patterson, Seidenberg & McClelland, 1989), they do not produce this double dissociation. An analogue of surface dyslexia is found (i.e., regulars are preserved), but no analogue of phonological dyslexia is observed. Furthermore, Bullinaria and Chater (1995) have explored a range of rule-exception tasks using feedforward networks trained by back-propagation, and concluded that, while double dissociations between rules and exceptions can occur in single-route models, this appears to occur only in very small scale networks. In large networks, the dissociation in which the rules are damaged but the exceptions are preserved does not occur. It remains possible that some realistic single route model of reading, incorporating factors which have been claimed to be important to connectionist accounts of reading such as word frequency and phonological consistency effects (cf. Plaut et al., 1996) might give rise to the relevant double dissociation. However, Bullinaria and Chater's results indicate that modeling phonological dyslexia is potentially a major challenge for any single route connectionist model of reading.

Single and dual route theorists argue about whether non-word and exception word reading is carried out by a single system, but agree that there is an additional "semantic" route, in which pronunciation is retrieved via a semantic code. This pathway is evidenced by deep dyslexics, who make semantic errors in reading aloud, such as reading the word *peach* aloud as "apricot". Plaut et al. (1996) argue that this route also plays a role in normal reading. In particular, they suggest that a division of labor emerges between the phonological and the semantic pathway during reading acquisition: Roughly, the phonological pathway moves towards a specialization in regular (consistent) orthography-to-phonology mappings at the expense of exception words which are read by the semantic pathway.

The putative effect of the latter pathway was simulated by Plaut et al. (1996) as extra input to the phoneme units in a feedforward network trained to map orthography to phonology. The strength of this external input is frequency-dependent and gradually increases during learning. As a result, the net comes to rely on this extra input. If eliminated (following a simulated lesion to the semantic pathway), the net loses much of its ability to read exception words, but retains good reading of regular words as well as non-words. Thus, Plaut et al. provides a more accurate account of surface dyslexia than Patterson et al. (1989). Conversely, selective damage to the phonological pathway (or to phonology itself) should produce a pattern of deficit resembling phonological dyslexia: Reasonable word reading but impaired non-word reading—but this hypothesis was not tested directly by Plaut et al.²

The Plaut et al. account of surface dyslexia has been challenged by the existence of patients with considerable semantic impairments but who demonstrate a near-normal reading of exception words. Plaut (1997) presents simulation results, suggesting that variations in surface dyslexia may stem from pre-morbid individual differences in the division of labor between the phonological and semantic pathways. In particular, if the

phonological pathway is highly developed prior to lesioning, a pattern of semantic impairment with good exception word reading can be observed in the model.

More recently, connectionist models of reading have been criticized for not capturing certain effects of orthographic length on naming latencies in single word reading. Plaut (this issue) takes up this challenge, presenting an SRN model of sequential processing in reading. Whereas the output of most previous connectionist reading models (except the NETtalk models) generated static phonological representations of *entire* words, this new model pronounces words phoneme-by-phoneme. It also has the ability to refixate on the input when it is unable to pronounce part of a word. The model performs well on words and nonwords, and provides a reasonably good fit with the empirical data on orthographic length effects. These results are encouraging, and suggest that this sequential reading model may provide a first step toward a connectionist account of the temporal aspects of reading.

We have seen that connectionist accounts have provided a good fit with data on normal and impaired reading, although points of controversy remain. Moreover, connectionist models have contributed to a re-evaluation of core theoretical issues, such as whether reading is interactive or purely bottom-up, and whether rules and exceptions are dealt with separately or by a single mechanism.

VIII. PROSPECTS FOR CONNECTIONIST NATURAL LANGUAGE PROCESSING

Current connectionist models as exemplified in Part I of this Special Issue, *Progress*, involve drastic simplifications with respect to real natural language. How can connectionist models be 'scaled up' to provide realistic models of human language processing? Part II, *Prospects*, provides three different perspectives on how connectionist models may develop.

Seidenberg and MacDonald (this issue) argue that connectionist models will be able to replace the currently dominant symbolic models of language structure and language processing, throughout the cognitive science of language. They suggest that connectionist models exemplify a *probabilistic*, rather than a rigid, view of language, that requires the foundations of linguistics as well as the cognitive science of language more generally to be radically rethought.

Smolensky (this issue), by contrast, argues that current connectionist models alone cannot handle the full complexity of linguistic structure and language processing. He suggests that progress requires a match between insights from the generative grammar approach in linguistics, and the computational properties of connectionist systems (e.g., constraint satisfaction). He exemplifies this approach with two grammar formalisms inspired by connectionist systems, Harmonic Grammar and Optimality Theory.

Steedman (this issue) argues that claims that connectionist systems can take over the territory of symbolic views of language, such as syntax or semantics, are premature. He suggests that connectionist and symbolic approaches to language and language processing should be viewed as complementary, but as currently dealing with different aspects of

language processing. Nonetheless, Steedman believes that connectionist systems may provide the underlying architecture on which high level symbolic processing occurs.

Whatever the outcome of these important debates, we note that connectionist psycholinguistics has already had an important influence on the cognitive science of language. First, connectionist models have raised the level of theoretical debate in many areas, by challenging theorists of all viewpoints to provide computationally explicit accounts. This has provided the basis for more informed discussions about processing architecture (e.g., single vs. dual route mechanisms, and interactive vs. bottom-up processing). Second, the learning methods used by connectionist models have reinvigorated interest in computational models of language learning (Bates & Elman, 1993). While Chomsky (e.g., 1986) has argued for innate “universal” aspects of language, the vast amount of language-specific information that the child acquires must be learned. Connectionist models may account for how some of this learning occurs. Furthermore, connectionist models provide a test-bed for the learnability of linguistic properties previously assumed to be innate. Finally, the dependence of connectionist models on the statistical properties of their input has contributed to the upsurge of interest in statistical factors in language learning and processing (MacWhinney, Leinbach, Taraban & McDonald, 1989; Redington, Chater & Finch, 1998).

Connectionism has thus already had a considerable influence on the psychology of language. But the final extent of this influence depends on the degree to which practical connectionist models can be developed and extended to deal with complex aspects of language processing in a psychologically realistic way. If realistic connectionist models of language processing can be provided, then the possibility of a radical rethinking not just of the nature of language processing, but of the structure of language itself, may be required. It might be that the ultimate description of language resides in the structure of complex networks, and can only be approximated by rules of grammar. Or perhaps connectionist learning methods do not scale up, and connectionism can only succeed by re-implementing standard symbolic models. The future of connectionist psycholinguistics is therefore likely to have important implications for the theory of language processing and language structure, either in overturning, or reaffirming, traditional psychological and linguistic assumptions. We hope that the papers in this Special Issue will contribute to determining what the future will bring.

NOTES

1. The term “connectionism” as referring to the use of artificial neural networks to model cognition was coined on the pages of this journal by Feldman and Ballard (1982).
2. Harm and Seidenberg (1999) have found the appropriate double dissociation in the context of developmental, rather than acquired, dyslexia.

REFERENCES

- Ashby, W. R. (1952). *Design for a brain*. New York: Wiley.
- Bates, E. A., & Elman, J. L. (1993). Connectionism and the study of change. In M. J. Johnson (Ed.), *Brain development and cognition* (pp. 623–642). Cambridge, MA: Basic Blackwell.

- Besner, D., Twilley, L., McCann, R. S., & Seergobin, K. (1990). On the connection between connectionism and data: Are a few words necessary? *Psychological Review*, *97*, 432–446.
- Boole, G. (1854). *The laws of thought*. London: Macmillan.
- Bullinaria, J. A. (1997). Modelling reading, spelling and past tense learning with artificial neural networks. *Brain and Language*, *59*, 236–266.
- Bullinaria, J. A., & Chater, N. (1995). Connectionist modelling: Implications for neuropsychology. *Language and Cognitive Processes*, *10*, 227–264.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425–455.
- Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: MIT Press.
- Chater, N., & Oaksford, M. (1990). Autonomy, implementation and cognitive architecture: A reply to Fodor and Pylyshyn. *Cognition*, *34*, 93–107.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1986). *Knowledge of language*. New York: Praeger.
- Christiansen, M. H. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure*. Unpublished doctoral dissertation, University of Edinburgh.
- Christiansen, M. H. (1999). *Intrinsic constraints on the processing of recursive sentence structure*. Manuscript in preparation.
- Christiansen, M. H., & Chater, N. (1994). Generalization and connectionist language learning. *Mind and Language*, *9*, 273–287.
- Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*, 157–205.
- Christiansen, M. H., & MacDonald, M. C. (1999). *Processing of recursive sentence structure: Testing predictions from a connectionist model*. Manuscript in preparation.
- Clahsen, H., Rothweiler, M., Woest, A., & Marcus, G. F. (1993). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition*, *45*, 225–255.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, *100*, 589–608.
- Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, *93*, 283–321.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, *27*, 124–142.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*, 123–147.
- Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, *23*, 517–542.
- Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*, 149–195.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*, 801–838.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.
- Elman, J. L. (1991). Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*, 195–225.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71–99.
- Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, *27*, 143–165.
- Fant, M. (1985). *Context-free parsing in connectionist networks* (Tech. Rep. No. TR-174). Rochester, NY: University of Rochester, Department of Computer Science.
- Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, *6*, 205–254.
- Fodor, J. A. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.
- Gaskell, M. G., Hare, M., & Marslen-Wilson, W. D. (1995). A connectionist model of phonological representation in speech perception. *Cognitive Science*, *19*, 407–439.

- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12, 613–656.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science*, 23, 439–462.
- Hahn, U., & Nakisa, R. C. (in press). German inflection: Single or dual route? *Cognitive Psychology*.
- Hanson, S. J., & Kegl, J. (1987). PARSNIP: A connectionist network that learns natural language grammar from exposure to natural language sentences. In *Proceedings of the Eighth Annual Meeting of the Cognitive Science Society* (pp. 106–119). Hillsdale, NJ: Erlbaum.
- Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition*, 56, 61–98.
- Hare, M., Elman, J. L., & Daugherty, K. M. (1995). Default generalization in connectionist networks. *Language and Cognitive Processes*, 10, 601–630.
- Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes*, 8, 291–309.
- Harm, M., Altmann, L., & Seidenberg, M. S. (1994). Using connectionist networks to examine the role of prior constraints in human learning. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 392–396). Hillsdale, NJ: Erlbaum.
- Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition and dyslexia: Insights from connectionist models. *Psychological Review*, 106, 491–528.
- Howells, T. (1988). VITAL, a connectionist parser. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Jordan, M. (1986). *Serial order: A parallel distributed approach* (Tech. Rep. No. 8604). San Diego: University of California, San Diego, Institute for Cognitive Science.
- Koch, C., & Segev, I. (1989) (Eds.). *Methods in neuronal modeling: From synapses to networks*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- MacDonald, M. C., & Christiansen, M. H. (in press). Individual differences without working memory: A reply to Just & Carpenter and Waters & Caplan. *Psychological Review*.
- MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language*, 28, 255–277.
- Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., Woest, A., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, 29, 189–256.
- Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.
- Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasia in deep dysphasia: Testing the consequence of decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language*, 47, 609–660.
- McClelland, J. L., & Elman, J. L. (1986). Interactive processes in speech perception: The TRACE model. In McClelland, J. L., & Rumelhart, D. E. (Eds.), *Parallel distributed processing* (Vol. 2, pp. 58–121). Cambridge, MA: MIT Press.
- McClelland, J. L., & Kawamoto, A. H. (1986). Mechanisms of sentence processing. In J. L. McClelland, & D. E. Rumelhart (Eds.), *Parallel distributed processing* (Vol. 2, pp. 272–325). Cambridge, MA: MIT Press.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375–407.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.
- Minsky, M. (1954). *Neural nets and the brain-model problem*. Unpublished doctoral dissertation, Princeton University, NJ.
- Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Morton, J., & Patterson, K. E. (1980). A new attempt at an interpretation, or, an attempt at a new interpretation. In M. Coltheart, K. E. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 91–118). London: Routledge.
- Miyata, Y., Smolensky, P., & Legendre, G. (1993). Distributed representation and parallel distributed processing of recursive structures. In *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society* (pp. 759–764). Hillsdale, NJ: Erlbaum.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

- Norris, D. G. (1993). Bottom-up connectionist models of 'interaction'. In G. Altmann, & R. Shillcock (Eds.), *Cognitive models of speech processing: The Second Sperlonga Meeting* (pp. 211–234). Hillsdale, NJ: Erlbaum.
- Patterson, K. E., Seidenberg, M. S., & McClelland, J. L. (1989). Connections and disconnections: Acquired dyslexia in a computational model of reading processes. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neuroscience* (pp. 131–181). Oxford: Oxford University Press.
- Pinker, S. (1991). Rules of language. *Science*, 253, 530–535.
- Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73–193.
- Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39, 347–370.
- Plaut, D. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, 12, 767–808.
- Plaut, D. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science*, 23, 543–568.
- Plaut, D., & McClelland, J. L. (1993). Generalization with componential attractors: Word and non-word reading in an attractor network. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 824–829). Hillsdale, NJ: Erlbaum.
- Plaut, D., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science*, 23, 463–490.
- Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, 38, 43–102.
- Plunkett, K., & Marchman, V. (1993). From rote learning to system building. *Cognition*, 48, 21–69.
- Prasada, S., & Pinker, S. (1993). Similarity-based and rule-based generalizations in inflectional morphology. *Language and Cognitive Processes*, 8, 1–56.
- Redington, M., Chater, N., & M., Finch, S. (1998). The potential contribution of distributional information to early syntactic category acquisition. *Cognitive Science*, 22, 425–469.
- Rosenblatt, F. (1962). *Principles of neurodynamics*. New York: Spartan Books.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning internal representations by error propagation*. In McClelland, J. L., & Rumelhart, D. E. (Eds.), *Parallel distributed processing* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effects and some tests and enhancements of the model. *Psychological Review*, 89, 60–94.
- Rumelhart, D. E., & McClelland, J. L. (1986). On learning of past tenses of English verbs. In McClelland, J. L., & Rumelhart, D. E. (Eds.), *Parallel distributed processing*, (Vol. 2, pp. 216–271). Cambridge, MA: MIT Press.
- Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, 32, 97–127.
- Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569–588.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Seidenberg, M. S., & McClelland, J. L. (1990). More words but still no lexicon: Reply to Besner et al. (1990). *Psychological Review*, 97, 447–452.
- Sejnowski, T. J. (1986). Open questions about computation in the cerebral cortex. In McClelland, J. L., & Rumelhart, D. E. (Eds.), *Parallel distributed processing* (Vol. 2, pp. 372–389). Cambridge, MA: MIT Press.
- Sejnowski, T. J., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English text. *Complex Systems*, 1, 145–168.
- Shillcock, R., Lindsey, G., Levy, J., & Chater, N. (1992). A phonologically motivated input representation for the modelling of auditory word perception in continuous speech. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 408–413). Hillsdale, NJ: Erlbaum.

- Small, S. L., Cottrell, G. W., & Shastri, L. (1982). Towards connectionist parsing. In *Proceedings of the National Conference on Artificial Intelligence*. Pittsburgh, PA.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11, 1–74.
- Smolensky, P. (1999). Grammar-based connectionist approaches to language. *Cognitive Science*, 23, 589–613.
- Steedman, M. (1999). Connectionist sentence processing in perspective. *Cognitive Science*, 23, 615–634.
- Stolcke, A. (1991). Syntactic category formation with vector space grammars. In *Proceedings from the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 908–912). Hillsdale, NJ: Erlbaum.
- Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, 12, 211–271.
- Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, 23, 491–515.
- Turing, A. M. (1936). On computable numbers, with an application to the *Entscheidungsproblem*. *Proceedings of the London Mathematical Society, Series 2*, 42, 230–265.