Connectionist Modeling of Language: Examples and Implications

David C. Plaut

Departments of Psychology and Computer Science, Carnegie Mellon University
Center for the Neural Basis of Cognition

May 1998

To appear in Banich, M. and Mack, M. (Eds.), *Mind, brain, and language: Multidisciplinary perspectives*.

Researchers interested in human cognitive processes have long used computer simulations to try to identify the principles of cognition. The strategy has been to build computational models that embody putative principles and then to examine how well such models capture human performance in cognitive tasks. Until the 1980's, this effort was undertaken within the context of the "computer metaphor" of mind. Researchers built computational models based on the conceptualization that the human mind operated as though it were a conventional digital computer. However, with the advent of so-called connectionist, neural network, or parallel distributed processing models (Anderson, Silverstein, Ritz, & Jones, 1977; Hinton & Anderson, 1981; McClelland & Rumelhart, 1981; McClelland, Rumelhart, & the PDP Research Group, 1986; Rumelhart, McClelland, & the PDP Research Group, 1986b), researchers began exploring the implications of principles that are more broadly consistent with the style of computation employed by the brain.

In connectionist models, cognitive processes take the form of cooperative and competitive interactions among large numbers of simple, neuron-like processing units. Unit interactions are governed by weighted connections that encode the long-term knowledge of the system and are learned gradually through experience. The activity of some of the units encodes the input to the system; the resulting activity of other units encodes the system's response to that input. The patterns of activity of the remaining, so-called *hidden* units constitute learned, internal representations that mediate between inputs and outputs. While each unit exhibits non-linear spatial and temporal summation, units and connections are not generally considered to be in one-to-one correspondence with actual neurons and synapses. Rather, connectionist systems attempt to capture the essential computational properties of the vast ensembles of real neuronal elements found in the brain, through simulations of smaller networks of units. In this way, the approach is distinct from computational neuroscience (Sejnowski, Koch, & Churchland, 1989), which aims to model the detailed neurophysiology of relatively small groups of neurons. Although the connectionist approach uses physiological data to guide the search for underlying principles, it tends to focus more on overall system function or behavior, attempting to determine what principles of brain-style computation give rise to the cognitive phenomena observed in human behavior. The approach enables developmental, cognitive and neurobiological issues to be addressed within a single, integrated formalism, providing new ways of thinking about how cognitive processes are implemented in the brain and how disorders of brain function lead to disorders of cognition.

The simplest structure for a connectionist network is a *feedforward* architecture, in which information flows unidirectionally from input units to output units, typically via one or more layers of hidden units. Although such networks can provide important insights into many cognitive domains, they are severely limited in their ability to learn and process information over time, and thus are relatively ill-suited for domains, such as language, that involve complex temporal structure (Elman, 1990). A more appropriate type of network for such domains is a *recurrent* architecture, with no a priori constraints on interactions among units. In one type of recurrent network, termed an *attractor* network, units interact in such a way that, in response to a fixed input, the network as a whole gradually settles to a stable pattern of activity representing the network's interpretation of the input (including any associated response). Recurrent networks can also learn to process sequences of inputs and/or to produce sequences of outputs. For example, in a *simple recurrent* network (Cleeremans, Servan-Schreiber, & McClelland, 1989; Elman, 1990, 1991), the internal representation generated by a given element of a sequence is made available as input to provide context for processing subsequent elements. Critically, the internal representations themselves adapt so as to provide and make effective use of this context information, enabling the system to learn to represent and retain relevant information across multiple time scales.

In fact, an issue of central relevance in the study of cognition in general, and language in particular, is the nature of the underlying representation of information. Some connectionist models use *localist* representations, in which individual units stand for familiar entities such as letters, words, concepts, and propositions (e.g., Dell, 1986; McClelland & Rumelhart, 1981). Others use *distributed* representations in which such entities are represented by alternative patterns of activity over large numbers of units rather than by the activity of a single unit (e.g., Hinton & Shallice, 1991; Seidenberg & McClelland, 1989). Although distributed representations are more difficult to think about, they offer a rich and powerful basis for understanding learning, generalization, and the flexibility and productivity of cognition (van Gelder, 1990).

The key to the effectiveness of distributed representations is the use of patterns whose similarity relations correspond to the similarities in the roles the patterns play in cognition, given that similar patterns tend to have similar consequences in connectionist models (see Hinton, McClelland, & Rumelhart, 1986, for discussion).[1] In very simple tasks, the similarities among the representations provided by the environment may be sufficient to guide behavior. However, in most cognitive domains, such as language, the functional relationships that must govern effective performance are often quite different from the similarities among surface forms. For example, the words CAT and CAP look and sound very similar but have entirely unrelated meanings. Consequently, the inputs to the system must be re-represented, perhaps via successive transformations across multiple intermediate layers of units, as new patterns of activity whose relative similarities abstract away from misleading surface similarity and, instead, capture the underlying structure of the domain. Traditional approaches to understanding cognition make very strong and specific assumptions about the structure of these internal representations and the processes that manipulate them. For example, it is often assumed that underlying linguistic knowledge takes the form of explicit rules which operate over discrete, symbolic representations (Chomsky, 1957; Chomsky & Halle, 1968; Fodor & Pylyshyn, 1988; Pinker, 1991) and, moreover, that this knowledge is, in large part, innately specified (Chomsky, 1965; Crain, 1991; Pinker, 1994).

By contrast, the connectionist approach places much greater emphasis on the ability of a system to *learn* effective internal representations. Learning in a connectionist network takes the form of modifying the values of weights on connections between units in response to feedback on the behavior of the network. A variety of specific learning procedures are employed in connectionist research; most that have been applied to cognitive domains, such as back-propagation (Rumelhart, Hinton, & Williams, 1986a) take the form of error correction: Change each weight in a way that reduces the discrepancy between the correct response for a given input and the one actually generated by the system. In this process, internal representations over hidden units are learned by calculating how to change each unit's activation to reduce error and then modifying its incoming weights accordingly. Although it is unlikely that the brain implements back-propagation in any direct sense (Crick, 1989), there are more biologically plausible procedures that are computationally equivalent (see, e.g., O'Reilly, 1996).

The emphasis on learning within the connectionist approach has fundamental implications for the nature of the explanations offered for cognitive behavior. Instead of attempting to stipulate the specific form and content of the knowledge required for performance in a domain, the approach instead stipulates the *tasks* the system must perform, including the nature of the relevant information in the environment, but then leaves it up to learning to develop the necessary internal representations and processes (McClelland, St. John, & Taraban, 1989). In some contexts, the resulting solution may bear a close relationship to more traditional mechanisms, but it is more often the case that learning develops representations and processes which are radically different from those proposed by traditional theories, and which generate novel hypotheses and testable predictions concerning human cognitive behavior.

Connectionist models have been applied to the full range of perceptual, cognitive, and motor domains (see McClelland et al., 1986; Quinlan, 1991; McLeod, Plunkett, & Rolls, 1998). It is, however, in their application to language that they have evoked the most interest and controversy (see, e.g., Pinker & Mehler, 1988). This is perhaps not surprising in light of the special role that

[1] This property arises because the input to each unit is typically a weighted sum of the activations of units from which it receives connections. A similar pattern of activity over the sending units, summed across the same weights, will generally produce a similar input to the receiving unit and, hence, a similar activation. This bias toward giving similar responses to similar inputs can be overcome by having large weights on particular connections, but this takes time to develop and will happen only if it is required to perform the task.

language plays in human cognition and culture. It also stems in part from the considerable divergence of goals and methods between linguistic versus psychological approaches to the study of language. This rift goes deeper than a simple dichotomy of emphasizing competence versus performance (Chomsky, 1957); it cuts to the heart of the question of what it means to know and use a language (Seidenberg, 1997). From a connectionist perspective, performance is not an imperfect reflection of some abstract competence, but rather the behavioral manifestation of the internal representations and processes of actual language users: language is as language does. The goal is not to abstract away from performance but to articulate computational principles that account for it.

A major attraction of the connectionist approach to language, apart from its natural relation to neural computation, is that the very same processing mechanisms apply across the full range of linguistic structure, including phonology, morphology, and syntax. The remainder of this paper discusses three specific connectionist models, each applied to one of these levels. The first model (Plaut & Kello, in press) is directed at central issues in phonological development, the second (Joanisse & Seidenberg, 1998) accounts for neuropsychological data in inflectional morphology, and the third (St. John & McClelland, 1990) addresses the integration of syntax and semantics in sentence comprehension. Beyond the use of common computational machinery, these models are all similar in that they they learn internal representations that mediate between input and output surface forms and their underlying meanings. None of them provides a fully adequate account of the relevant phenomena in its domain. Nonetheless, they collectively illustrate both the breadth and depth of the approach. The concluding section highlights some of the limitations of current models and identifies important directions for future research.

Phonological Development: Plaut and Kello (in press)

Phonology is concerned with the sound structure of a language, and with how contrasts in meaning are conveyed by contrasts in the surface forms of words. The development of phonological representations plays a key role in the acquisition of both speech comprehension and production. In comprehension, time-varying acoustic input must be mapped onto a stable representation of the meaning of the utterance. This process poses a considerable challenge to the infant due to the considerable variability in the speech signal across talkers and contexts, and because, at the morphemic level, the relationship of spoken words to their meanings is largely arbitrary. In production, a meaning representation must generate appropriate time-varying articulatory output. Here, the infant must

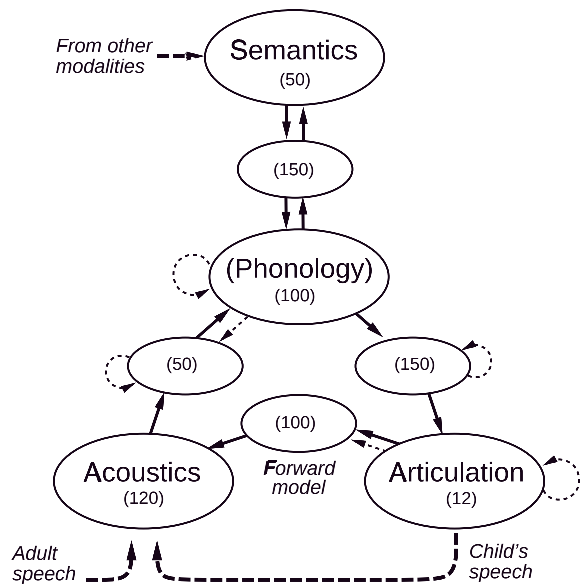

Figure 1. The architecture of the simple recurrent network used by Plaut and Kello (in press). Solid arrows indicate standard projections of full connectivity between groups; dashed arrows indicate projections from "context" units whose states are copied from the previous time step. For the sake of clarity, context projections are depicted as coming from the source of the copied states rather than from separate context units. The number of units in each group is given in parentheses. (Adapted from Plaut & Kello, in press)

learn to produce comprehensible speech without any direct feedback or instruction from caretakers as to what articulatory movements are required to produce particular sound patterns. Moreover, although abilities in comprehension tend to precede those in production (see Jusczyk, 1997; Vihman, 1996, for reviews) these two processes must nonetheless converge on a mutually consistent solution to ensure that the infant comes to speak the same language(s) that he/she hears.

Plaut and Kello (in press) proposed a connectionist framework for phonological development in which phonology is a learned, internal representation that mediates among acoustic, articulatory, and semantic representations in the service of both comprehension and production. In support of the framework, Plaut and Kello developed an implementation in the form of a simple recurrent network, depicted in Figure 1, that learned to understand and produce isolated spoken words in the absence of explicit articulatory feedback

The framework instantiates two key assumptions. The first is that both comprehension and production are subserved by the same underlying phonological representations. These representations develop initially under the pressure of mapping acoustics to semantics in the course of learning to understand adult speech, but become increasingly refined by articulatory factors as skill in production develops. By sharing common underlying phonological representations, structure learned in the service of comprehension is available to guide production (see Vihman, 1996), and refinements driven by the demands of articulation automatically apply within comprehension (see Liberman, 1996).

The second key assumption is that feedback needed to guide the development of speech production is derived from the comprehension system—that is, from the acoustic, phonological, and semantic consequences of the system's own articulations (Locke, 1983; Menn & Stoel-Gammon, 1995; Studdert-Kennedy, 1993). This can be accomplished by first learning an internal *forward* model of the physical processes relating specific articulations to the acoustics they produce (Jordan & Rumelhart, 1992; Perkell et al., 1995). Such a model can be learned by executing a variety of articulations, predicting how they will sound, and then adapting the model based on the discrepancy or error between this prediction and the actual resulting acoustics. In the infant, the forward model is assumed to develop primarily as a result of canonical and variegated babbling in the second half of the first year (see Vihman, 1996, for review, and Houde & Jordan, 1998, for empirical support for the existence of such a forward model).

The importance of learning an articulatory-acoustic forward model is that it can be used to convert acoustic and phonological feedback (i.e, whether an utterance sounded right) into articulatory feedback that can improve speech production. The approach assumes that learning to produce speech takes place largely in the context of attempts to imitate adult speech. In imitation, the system first derives acoustic and phonological representations for an adult utterance during comprehension. It then uses the resulting phonological representation as input to generate a sequence of articulatory gestures. These gestures, when executed, result in acoustics which are then mapped back onto phonology via the comprehension system. The discrepancies between the resulting representations and the original acoustic and phonological representations generated by the adult constitute the error signals that ultimately drive articulatory learning. In order for this to work, however, these "distal" errors in acoustics and phonology must be converted to "proximal" errors in articulation. This is done by propagating phonological and acoustic error across the forward model to derive error signals over articulatory states. These error signals are then used to adapt the production system (i.e., the mapping from stable phonological representations onto articulatory sequences) to better approximate the acoustics and phonology generated by the adult.

The implementation developed by Plaut and Kello necessarily incorporated a number of simplifications to keep computational demands within reasonable limits. Two of these are most critical. First, the implementation used discrete rather than continuous time. The time-varying acoustic input and articulatory output were described in terms of sequences of events marking points of significant change. There were approximately as many events in an utterance as phonemes (plosives, affricates and diphthongs corresponded to two events) although, due to coarticulatory influences, information about a given segment was spread out over a number of adjacent events.

Second, the implementation used artificial rather than real speech. Acoustic events were encoded in terms of ten variables: the first three formants (1–3) and their rates of change (4–6), amount of frication (7) and bursting (8), loudness (9), and degree of jaw openness (10). The last variable is, strictly speaking, visual/proprioceptive rather than acoustic but has been shown to be a important source of information in infant speech acquisition (Locke, 1995; Meltzoff & Moore, 1977). Articulatory events were similarly encoded in terms of six variables: degree of oral (1) and nasal constriction (2), place of oral constriction (3), tongue height (4) and backness (5), and amount of voicing (6). Finally, the physical processes relating articulation to acoustics were approximated by a set of complex, coupled equations that map any combination of values for the articulatory variables onto the corresponding values for the acoustic variables. Considerable effort was spent to make these representations and equations as realistic as possible while staying within the constraints of computational efficiency.

In the simulation, the value of each articulatory variable was represented by the difference in activity between contrasting units (corresponding to the ends of the continuum). Each acoustic value was represented by the mean of a Gaussian pattern of activity over a bank of twelve units; the total activity of the Gaussian encoded the strength of the information. For illustration purposes, Figure 2 shows the articulatory and acoustic representations for the closure and release of the /p/ in the word SPIN.

The training vocabulary for the network was the 400 highest frequency monosyllabic nouns and verbs in the Brown corpus (Kučera & Francis, 1967) with at most four phonemes (mean = 3.42). Words were selected for presentation during training in proportion to a logarithmic function of their frequency of occurrence.

The network underwent three kinds of training episodes: babbling, comprehension, and imitation. Intensional naming is also mentioned because the network was tested on this task even though it was not trained on it explicitly.

Babbling. Babbling served to train the articulatory-acoustic forward model (see Figure 1). Pseudo-random

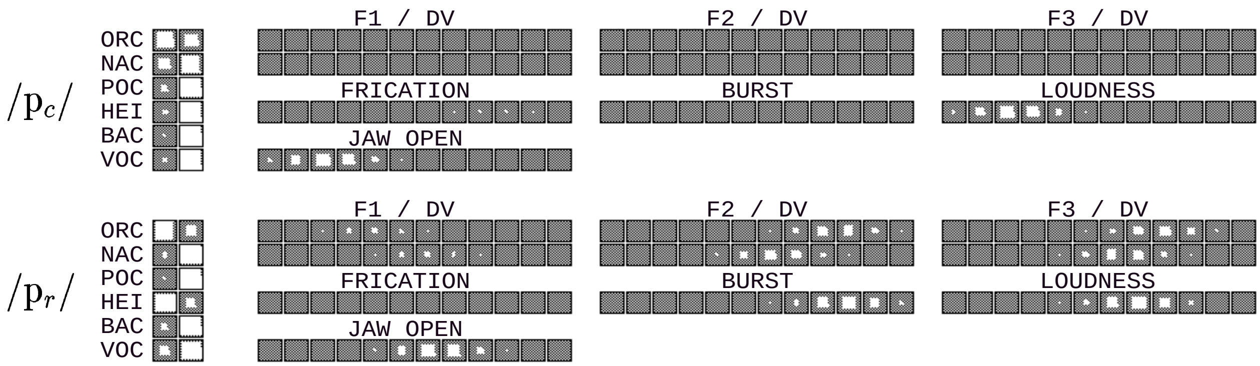

Figure 2. The articulatory (left) and acoustic (right) events corresponding to the closure and release of /p/ in an utterance of the word SPIN. (Adapted from Plaut & Kello, in press)

articulatory sequences, designed to mimic a bias toward mandibular (jaw) oscillation in infants (MacNeilage & Davis, 1990), were generated and passed through the articulation-to-acoustics equations to produce a sequence of acoustic patterns. The articulations also served as input to the forward model, which generated a sequence of predicted acoustic patterns. The discrepancy or error between the actual and predicted acoustics at each step was then back-propagated through the forward model and used to adjust its connection weights to improve its ability to predict the acoustic outcomes of the given articulations. In this way, the forward model gradually learned to mimic the physical mapping from articulatory sequences to acoustic sequences (as instantiated by the articulation-to-acoustics equations).

Comprehension. Comprehension involved deriving the semantic representation of a word from the acoustic sequence produced by an adult utterance of the word. Adult utterances were generated by applying the articulation-to-acoustics equations to the sequences of canonical articulatory events for words, subject to intrinsic variability and coarticulation. Each resulting sequence was then mapped from acoustics via phonology to semantics, and the error between the generated semantics at each step and the correct semantics for the word was back-propagated to change the weights between acoustics and semantics.[2] Gradually, the network learned to activate the correct semantic pattern for the acoustic sequences corresponding to each word; in doing so, the final pattern of activity over phonology constituted the network's internal phonological representation of the word.

Imitation. Imitation involved using a phonological representation derived from an adult utterance as input to drive articulation, and comparing the resulting acoustic and phonological representations with those of the adult utterance. Specifically, after hearing an adult utterance, the network used its derived phonological representation as input to generate a sequence of articulatory representations. This sequence was then mapped both by the forward model to generate predicted acoustics, and by the articulation-to-acoustics equations to generate actual acoustics. The latter were in turn mapped via the comprehension system to phonology (and semantics). The error between the acoustic and phonological representations generated by the network and those generated by the adult was then back-propagated from phonology to acoustics and then back across the forward model to derive error feedback for articulation. (Note that the forward model plays the critical role here of converting acoustic error into articulatory error.) This feedback was then back-propagated to phonology and used to modify the weights in the production system to improve its ability to reproduce the acoustics and phonology of the adult utterance.

Note that, in learning to imitate, the network is provided only with the acoustics of adult utterances. It must learn to adapt its own articulations based solely on how similar to the adult utterances they *sound*.

Intentional naming. Intentional naming involved generating an articulatory sequence given the semantic representation of a word as input. Although the network was not trained specifically to perform this task, it can be tested on it in a way that is similar to imitation. The only difference is that the initial phonological representation is generated from semantics top-down rather than from an adult utterance bottom-up.

The network was tested for its ability to comprehend, imitate, and intentionally name words after every 500,000 word presentations, up to a total of 3.5M (million) presentations.[3] Figure 3 shows the levels of correct perfor-

[2] The semantic representations were generated artificially to cluster into categories and assigned to words randomly (see Plaut & Kello, in press, for details). Although the relationship between the surface forms of words and their meanings was arbitrary, no attempt was made to approximate the actual meanings of the words themselves.

[3] Although this may seem like an excessive amount of training, children speak up to 14,000 words per day (Wagner, 1985), or over 5 million words per year.

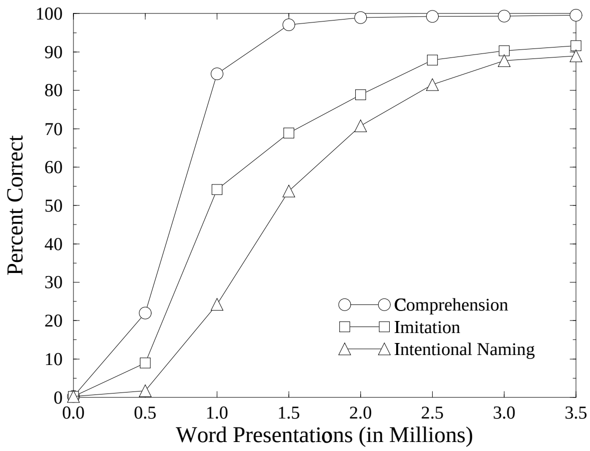

Figure 3. Correct performance of the Plaut and Kello (in press) network on comprehension, imitation, and intentional naming, over the course of training. (Adapted from Plaut & Kello, in press)

mance on these tasks over the course of training. Performance was measured in terms of whether the semantics generated by the network was more similar to the correct semantics of the word than to those of any other word. Comprehension performance improved relatively rapidly, reaching 84.3% correct by 1M word presentations and 99.6% by 3.5M presentations. This level of performance is impressive given the lack of systematicity in the mapping between acoustics and semantics and the considerable intrinsic variability of adult utterances. Relative to comprehension, production developed more slowly: the network was only 54.2% correct at imitation by 1M presentations, although it did achieve 91.7% correct by 3.5M presentations. Intentional naming was slightly poorer than imitation throughout training, eventually reaching 89.0% correct. This is not surprising as the task involves mapping through the entire network and was not trained explicitly.

Thus, the network achieved quite good performance at both comprehension and production. The fact that comprehension precedes production in the model stems directly from the fact that learning within the production system is driven by comparisons over representations within the comprehension system. The excellent performance on imitation demonstrates that feedback from the comprehension system via a learned forward model can provide effective guidance for articulatory development.

Plaut and Kello also carried out an analysis of the errors produced by the network. In general, the network showed a strong bias toward phonological similarity in its errors compared with the chance rate, for both comprehension and imitation. At the phoneme level, there were far more errors on consonants than on vowels and, among consonants, a relatively higher error rate on fricatives, affricates (e.g., /ʧ/) and /ŋ/ (as in RING). These errors involved both additions and deletions; when they were deleted, they were often replaced by a plosive. In fact, plosives accounted for over half of the total number of insertions. By contrast, the liquids /r/ and /l/ were deleted occasionally, but never inserted. These characteristics are in broad agreement with the properties of early child speech errors (e.g. Ingram, 1976).

In summary, Plaut and Kello (in press) developed a connectionist framework in which phonology is a learned internal representation mediating both comprehension and production, and in which comprehension provides production with error feedback via a learned articulatory-acoustic forward model. An implementation of the framework, in the form of a simple recurrent network, learned to comprehend, imitate, and intentionally name a corpus of 400 monosyllabic words. Moreover, the speech errors produced by the network showed similar tendencies as those of young children. Although only a first step, the results suggest that the approach may ultimately form the basis for a comprehensive account of phonological development.

Inflectional Morphology: Joanisse and Seidenberg (1998)

The second example of a connectionist model of language processing is from recent work by Joanisse and Seidenberg (1998) in the domain of English inflectional morphology. The past-tense system of English verbs is a classic example of a *quasi-regular* domain, in which the relationship between inputs and outputs is systematic but admits many exceptions. Thus, there is a single regular "rule" (add –ed; e.g., WALK ⇒"walked") and only about 150 exceptions, grouped into several clusters of similar items that undergo a similar change (e.g., SING ⇒"sang", DRINK ⇒"drank") along with a very small number of very high-frequency, arbitrary forms (e.g., GO ⇒"went"; Bybee & Slobin, 1982).

The traditional view of the language system (e.g., Pinker, 1984, 1991) is that the systematic aspects of language are represented and processed in the form of an explicit set of rules. Given that most domains are only partially systematic, however, a separate mechanism (e.g., an associative network; Pinker, 1991) is required to handle the exceptions. The distinction between a rule-based mechanism and an exception mechanism, each operating according to different computational principles, forms the central tenet of so-called "dual-route" theories of language.

Rumelhart and McClelland (1986) argued for an alternative view of language in which all items coexist within a single system whose representations and processing reflect the relative degree of *consistency* in the mappings for different items (also see Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, & Patterson, 1996). They developed a connectionist model that learned a direct association between the phonology of all types of verb stems and the phonology of their past-tense forms. Pinker and Prince (1988) and Lachter and Bever (1988), however, pointed out numerous deficiencies in the model's performance and in some of its specific assumptions, and argued more generally that the applicability of connectionist mechanisms in language is fundamentally limited (also see Fodor & Pylyshyn, 1988). Subsequent simulation work has addressed many of the specific limitations of the Rumelhart and McClelland model (Cottrell & Plunkett, 1991, 1995; Daugherty & Seidenberg, 1992; Hoeffner, 1992; MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1991, 1993, 1996) and has extended the approach to address language disorders (Hoeffner & McClelland, 1993; Marchman, 1993) and language change (Hare & Elman, 1995).

More recently, proponents of dual-route theories have identified neuropsychological dissociations in processing regular versus irregular inflectional morphology, both in the performance of brain-damaged patients (Marslen-Wilson & Tyler, 1997; Ullman et al., 1997) and in the regional cerebral blood flow of normal subjects (Jaeger et al., 1996, although see Seidenberg & Hoeffner, 1998, for criticism). These dissociations have been interpreted by these authors as supporting the existence of separate mechanisms for rule-governed versus exceptional items. For example, Ullman et al. (1997) found that patients with Alzheimer's disease were relatively impaired in generating the past tense of irregular verbs (60% correct) compared with regular verbs (89% correct) and novel verbs (e.g., CUG; 84% correct). By contrast, patients with Parkinson's disease were relatively impaired on the novel verbs (65% correct) compared with both the regular and irregular verbs (80 and 88% correct, respectively). A similar contrast in performance on novel versus irregular verbs held among aphasic patients with either posterior lesions (novel 85%; irregular 71%) or anterior lesions (novel 5%; irregular 69%). Ullman and colleagues interpreted these findings as implicating two separate mechanisms: a posterior "mental dictionary" needed to retrieve irregular inflections, and a frontal/basal-ganglia grammatical rule system needed to inflect novel verbs.

An alternative account is that the double dissociation of novel versus irregular inflectional morphology is due to damage to different types of information within a single mechanism that processes all types of items (also see Plaut, 1995). In particular, irregular morphology may be particularly sensitive to semantic damage whereas novel inflections may be particularly sensitive to phonological damage. In fact, the same proposal as been made in the domain of word reading (Patterson & Hodges, 1992; Patterson & Marcel, 1992; Plaut et al., 1996) where analogous dissociations occur: surface dyslexic patients (see Patterson, Coltheart, & Marshall, 1985) are impaired in pronouncing exception words (e.g., PINT) but not pseudowords (e.g., RINT), whereas phonological dyslexic patients (see Coltheart, 1996) are impaired in pronouncing pseudowords relative to both regular and exception words. In fact, there is independent evidence for semantic impairments with posterior (temporal) involvement in Alzheimer's disease (Schwartz, 1990) and in surface dyslexic patients (Graham, Hodges, & Patterson, 1994), and for phonological impairments with frontal involvement in Parkinson's patients (e.g., Grossman, Carvell, Stern, Gollump, & Hurtig, 1992) and in phonological dyslexic patients (Patterson & Marcel, 1992).

Joanisse and Seidenberg (1998) developed a connectionist simulation of inflectional morphology in support of this account. The architecture of their network is shown in Figure 4. Note that it is broadly similar to the framework proposed by Plaut and Kello (in press): spoken input interacts via a common internal representation both with semantics and with spoken output. Here, though, the surface forms of words are represented in more abstract form. Input and output phonology are represented in terms of sequences of phonemes of the form `CCCVVCCVC` (where some slots may be empty). Thus, the past tense of STOP is `-sta-pt---` and the past tense of WANT is `--wa-nt-Id`. Within each of the nine slots, a phoneme is coded in terms of 18 phonetic features, yielding a total of 162 units. Verb meanings were not encoded explicitly; rather, each verb was assigned a localist representation of a single unit. An additional "past-tense" unit in semantics indicated that the phonological output of the verb should be the past (as opposed to present) tense. The network was trained on 600 randomly selected verbs (weighted by their frequency), including 64 irregular verbs; it thus contained 601 semantic units. In addition, the network contained two groups of 20 "clean-up" units, one interacting with semantics and one with output phonology. These groups are additional hidden units that learn to help semantics and output phonology settle into correct, stable (attractor) states.

The network was trained with back-propagation through time (Rumelhart et al., 1986a) to perform four tasks (on the indicated proportion of trials):

Hearing (40%): Given the input phonology of a verb (present or past), activate the corresponding semantic unit (and the past-tense unit if it was in the past tense).

Repeating (30%): Given the input phonology of a verb, reproduce it over output phonology. To facilitate

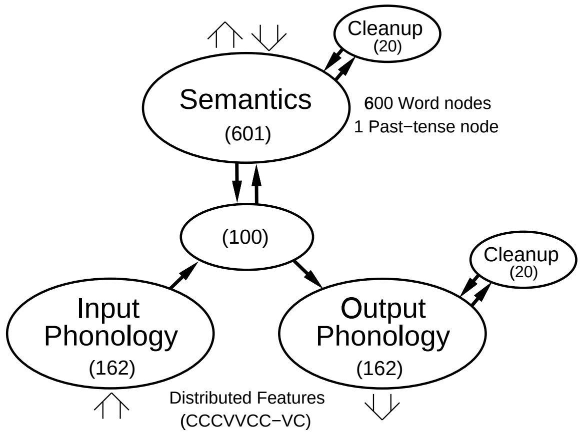

Figure 4. The architecture of the network used by Joanisse and Seidenberg (1998) to model inflectional morphology. The large arrows identify input and output layers. The number of units in each layer is indicated in parentheses. (Adapted from Joanisse & Seidenberg, 1998)

learning English phonology, in addition to the 600 verbs in semantics, the network was trained to repeat an additional 596 verbs in both present and past tense.

Speaking (20%): Given the semantics of a verb (possibly including the past-tense unit), generate the correct output phonology.

Transforming (10%): Given the input phonology of a verb in present tense and activation of the past-tense unit in semantics, generate the past-tense output phonology of the verb.

After 2.7 million training trials, the network was 99.5% correct on hearing, 98.2% correct on repetition, 99.8% correct on speaking, and 99.3% correct on transforming present to past. In addition, the network was 85% correct when tested for its ability to transform the phonology of the novel verbs from the Ullman et al. (1997) study. Thus, although the network did not contain separate mechanisms for regular versus irregular morphology, it nonetheless was capable of highly accurate processing of both types of verbs, as well as reasonably accurate generalization in transforming novel verbs.

Joanisse and Seidenberg then tested their network's performance after either semantic or phonological lesions. Semantic lesions involved adding Gaussian noise with *SD* = 0.22 to the activations of the semantic units and randomly eliminating 22% of their connections from clean-up units. Phonological lesions involved adding Gaussian noise with *SD* = 0.30 to the activation of the output phonology units and randomly eliminating 15% of their connections from clean-up units. Figure 5 shows the performance of the model, averaged over 10 instances of each such lesion, in transforming the present to past tense of regular, irregular, and novel verbs. Also included are Ullman et al.'s (1997) data for the corresponding patient groups. As with Alzheimer's patients, semantic lesions in the model have a much more detrimental effect on irregular verbs than on regular or novel verbs. By contrast, in Parkinson's patients and for the model after phonological lesions, novel verbs are the most impaired (although the dissociation is not as strong as for semantic lesions).

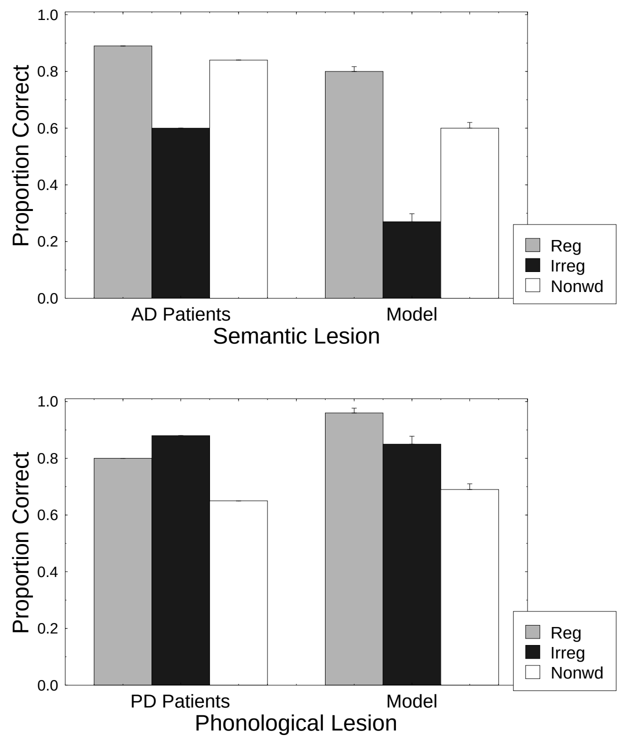

Figure 5. The effects of lesions to semantics (top) or to phonology (bottom) on the performance of the Joanisse and Seidenberg (1998) network in transforming regular, irregular, and novel verbs, and the corresponding data for patients with Alzheimer's disease (AD) or Parkinson's disease (PD) from Ullman et al. (1997). (Adapted from Joanisse & Seidenberg, 1998)

In the model, semantic damage impairs irregulars the most because interactions with semantics are required to override the strong consistency in the regular inflection for these items. In contrast, phonological damage impairs novel verbs the most because, unlike both regular and irregular verbs, such verbs receive no support from interactions with semantics. Thus, due to these learned specializations, lesions to semantics versus phonology in the model replicate the empirical double dissociation of performance in inflecting irregular versus novel verbs that Ullman and colleagues observed among the patients.

Consequently, the dissociations do not provide support for dual-route theories of language processing, and can be accounted for naturally by a distributed connectionist system in which multiple sources of information interact in processing all types of items.

Sentence Comprehension: St. John and McClelland (1990)

Having discussed connectionist models that have been applied to issues in phonology and morphology, we now consider how the approach can provide important insights at the level of the syntax and semantics of sentences. Traditional linguistic theory has focused on grammar as the essential element of linguistic knowledge, abstracting away from semantic and pragmatic influences on performance (Chomsky, 1957, 1965, 1985, 1995). This view has spawned psychological models (e.g., Ferreira & Clifton, 1986; Frazier, 1986; Marcus, 1980) that include an initial syntactic parse which is insensitive to lexical/semantic constraints (apart from word class information). And yet, from a computational point of view, a parser divorced from real-world knowledge runs into a number of difficult problems. Consider the following examples (from McClelland et al., 1989):

1. The spy saw the policeman with a revolver
2. The spy saw the policeman with binoculars
3. The bird saw the birdwatcher with binoculars

As these sentences are structurally identical, the attachment of the prepositional phrase depends solely on the meanings of the words and the relative adequacy of alternative interpretations. In (1) versus (2), only the binoculars are a plausible instrument of seeing, whereas a revolver is more likely to belong to a policeman. In (3), the fact that birdwatchers but not birds often possess and use binoculars reverses the attachment in (2). Indeed, every constituent in a sentence can potentially influence the role assigned to a prepositional phrase (Oden, 1978).

Conversely, just as word meaning is needed to influence syntactic processes, so sentence-level syntax and semantics must be used to determine word meanings. This can be seen clearly in considering ambiguous words, as in

4. The pitcher threw the ball

in which every content word has multiple meanings in isolation but an unambiguous meaning in context. It also applies to vague or generic words, such as “container,” which can refer to very different types of objects in different contexts (Anderson & Ortony, 1975), as in

5. The container held the apples
6. The container held the cola

Finally, at the extreme end of context dependence are implied constituents which are not even mentioned in the sentence but nonetheless are an important aspect of its meaning. For example, from

7. The boy spread the jelly on the bread

most people infer a knife as instrument (McKoon & Ratcliff, 1981).

These and other considerations have lead a number of researchers to question claims for the autonomy of syntax. Instead, sentence comprehension is envisioned as a constraint satisfaction process in which multiple sources of information from both syntax and semantics are simultaneously brought to bear in constructing the most plausible interpretation of a given utterance (see, e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; McClelland & Kawamoto, 1986; Seidenberg, 1997; Tanenhaus & Trueswell, 1995).

St. John and McClelland (1990; McClelland et al., 1989) developed a connectionist model of sentence comprehension which instantiates this key idea and which, at least in limited form, addresses the challenges raised above. The architecture of the model, in the form of a simple recurrent network, is shown in Figure 6. The task of the network was to take as input a single-clause sentence as a sequence of surface constituents, and to derive an internal representation of the event described by the sentence, termed the *Sentence Gestalt*. Critically, this representation was not predefined but was learned from feedback on its ability to generate appropriate thematic role assignments for the event (given either roles or fillers as “probes”).

Events were organized around actions and had a probabilistic structure. Specifically, each of 14 actions had a specified set of thematic roles, each of which was filled probabilistically by one of the possible constituents. In this process, the selection of fillers for certain roles biased the selection for other roles. For example, for eating events, the busdriver most often ate steak whereas the teacher most often ate soup, although occasionally the reverse occurred. The choice of words in the construction of a sentence describing the event was also probabilistic. The event of a busdriver eating a steak with a knife might be rendered as THE-ADULT ATE THE-FOOD WITH-A-UTENSIL, THE-STEAK WAS-CONSUMED-BY THE-PERSON, SOMEONE ATE SOMETHING, and so on (where the hyphenated phrases are constituents). Overall, given the probabilistic event structures and the lexical and syntactic options for describing events as sentences, there were a total of 120 different events (of which some are much more likely than others) and 22,645 different sentence-event pairs.

During training, sentence-event pairs were generated successively and the constituents of each sentence were

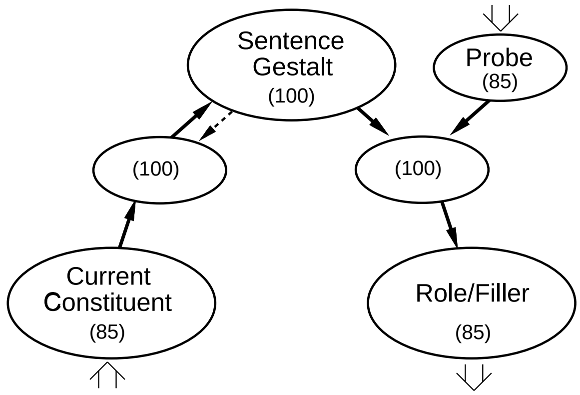

Figure 6. The architecture of the simple recurrent network used by St. John and McClelland (1990) to model sentence comprehension. The number of units in each layer is shown in parentheses. The large arrows identify input and output layers. The dashed arrow indicates a projection from "context" units (omitted for clarity) whose states are copied from the Sentence Gestalt layer for the previous time step. (Adapted from St. John & McClelland, 1990).

presented one at a time over the Current Constituent units (see Figure 6). For each constituent, the network updated its Sentence Gestalt representation and then attempted to use this representation as input to generate the full set of role/filler pairs for the event. Specifically, with the Sentence Gestalt fixed and given either a role or a filler over the "Probe" units, the network had to generate the other element of the pair over the "Role/Filler" units. For example, after the presentation of THE-STEAK in the sentence THE-STEAK WAS-EATEN-BY THE-BUSDRIVER, the network was trained to output, among other things, the agent (busdriver), the patient (steak), the action (eating), and the instrument (fork). It was, of course, impossible for the network to do this with complete accuracy, as these role assignments depend on constituents that have yet to occur or are only implied. Even so, the network could do better than chance—it could attempt to predict missing information based on its experience with the probabilistic dependencies in the event structures. More specifically, it could (and, in fact, did) generate distributions of activity over roles and fillers that approximated their frequency of occurrence over all possible events described by sentences that start with THE-STEAK. Note that these distributions could, in many cases, be strongly biases towards the correct responses. For example, steaks typically fill the patient role in eating events and (in the environment of the network) are most commonly eaten by busdrivers using a fork. In this way, the training procedure encouraged the network to extract as much information as possible as early as possible, in keeping with the principle of *immediate update* (Eberhard, Spivey-Knowlton, & Tanenhaus, 1995; Marslen-Wilson & Tyler, 1980; van Dijk & Kintsch, 1983). Of course, the network also had to learn to revise the Sentence Gestalt appropriately in cases where its predictions were violated, as in THE-STEAK WAS-EATEN-BY THE-TEACHER.

The network was trained on a total of 630,000 sentence-event pairs, in which some pairs occurred frequently and others—particularly those with atypical role assignments—were very rare. Figure 7 shows the performance of the model on sentences of various types as a function of training experience. In general, active voice was learned before passive voice, and syntactic constraints (implied by word order) were learned before semantic constraints (implied by event statistics). By the end of training, when tested on 55 randomly generated sentence-event pairs with unambiguous interpretations, the network was correct in 1699 of 1710 role/filler assignments (99.4% correct).

St. John and McClelland also carried out a number of more specific analyses intended to establish that the network could handle more subtle aspects of sentence comprehension. In general, the network succeeded at using both semantic and syntactic context 1) to disambiguate word meanings (e.g., for THE PITCHER HIT THE BAT WITH THE BAT, assigning flying bat as patient and baseball bat as instrument); 2) to instantiate vague words (e.g., for THE TEACHER KISSED SOMEONE, activating a male of unknown age as patient), and 3) to elaborate implied roles (e.g., for THE TEACHER ATE THE SOUP, activating spoon as the instrument; for THE SCHOOLGIRL ATE), activating a range of foods as possible patients). The network also demonstrated the ability to recover from semantic "garden paths," in which early predictions had to be revised in light of later evidence (see Figure 8).

In summary, St. John and McClelland (1990) present a connectionist model in which semantic and syntactic constraints are integrated to support online sentence comprehension. Although there are significant limitations in the complexity of the language on which the model was trained, it nonetheless instantiates and provides support for a theory of sentence comprehension as probabilistic constraint satisfaction (MacDonald et al., 1994; Seidenberg, 1997). This perspective stands in sharp contrast to traditional linguistic (Chomsky, 1965) and psycholinguistic theories (Ferreira & Clifton, 1986; Frazier, 1986) which espouse a clear separation of grammar from the rest of cognition.

Summary and Conclusions

Connectionist modeling is attractive as a framework for understanding cognition in general, and language in particular, because it provides an account of the flexibility

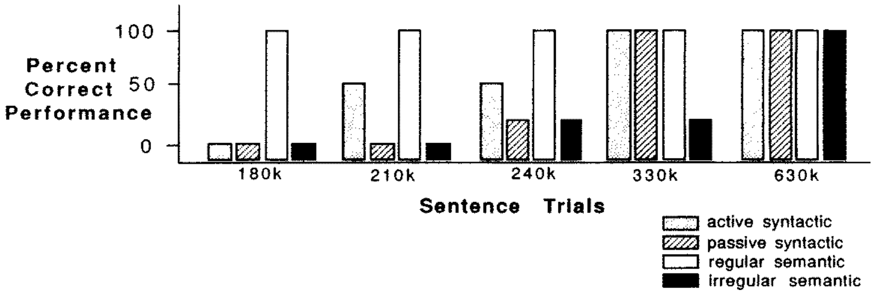

Figure 7. Performance of the St. John and McClelland (1990) network in thematic role assignment for four classes of sentences: *active syntactic* (e.g., THE BUSDRIVER KISSED THE TEACHER), *passive syntactic* (e.g., THE TEACHER WAS KISSED BY THE BUSDRIVER), *regular semantic* (e.g., THE BUSDRIVER ATE THE STEAK), and *irregular semantic* (e.g., THE BUSDRIVER ATE THE SOUP). (Reprinted from St. John & McClelland, 1990).

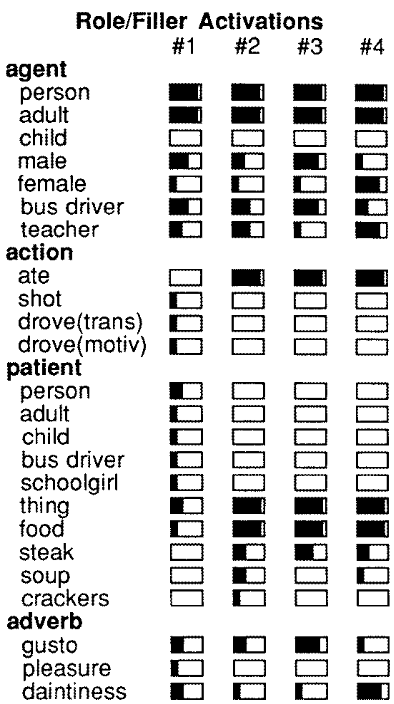

Figure 8. The activations (as black bars) of possible fillers for selected roles generated by the St. John and McClelland (1990) network after processing each of the four constituents in the sentence THE-ADULT$_{\#1}$ ATE$_{\#2}$ THE-STEAK$_{\#3}$ WITH-DAINTINESS$_{\#4}$. After processing THE-STEAK, the network instantiates THE-ADULT as the busdriver, but when WITH-DAINTINESS is encountered, the network must reinterpret THE-ADULT to mean the teacher (given the statistics for eating events). (Reprinted from St. John & McClelland, 1990).

and productivity of human performance through the development of internal representations that capture the underlying structure in a domain, and because it suggests how such representations and processes might actually be learned and carried out by the brain. The current paper discusses three examples of connectionist models, each applied to a different level of language structure. At the phonological level, the Plaut and Kello (in press) model provides an account of how comprehension and production are coordinated in phonological development, and how production can be trained by feedback from the comprehension system via a learned articulatory-acoustic forward model. At the morphological level, the Joanisse and Seidenberg (1998) model demonstrates that neuropsychological dissociations in inflecting regular (and novel) versus irregular English verbs do not implicate separate rule-based and associative mechanisms, but arise naturally from damage to semantic versus phonological processes within a single, distributed system that processes all types of items. At the sentence level, the St. John and McClelland (1990) model illustrates how a system can learn both semantic and syntactic knowledge from its experience with sentences and the events they describe, and bring this knowledge to bear in an online, integrated fashion to construct the most plausible interpretation of a given sentence.

Each of these models has important limitations in its theoretical scope and empirical adequacy. The Plaut and Kello model was applied only to isolated monosyllables which were assigned very abstract distributed semantic representations. Moreover, the articulatory and acoustic representations, and the equations that relate them, provide only a coarse approximation to the richness of the information and constraints in these domains. The Joanisse and Seidenberg model, similarly, employed a limited vocabulary of isolated verbs, highly restrictive phonological representations, and made no attempt to capture similarities among verb meanings. The St. John and McClelland model was trained on sentences restricted to single clauses without embeddings and pre-parsed into syntactic constituents. The use of event structures composed of probabilistic assignment to fixed thematic roles was also highly simplified.

A more general limitation that spans all three models is the approximation of temporal processing in terms of discrete sequences of events. Although networks with continuous-time processing have been applied in language-related domains (e.g., Harm & Seidenberg, in press; Plaut et al., 1996), typically these networks have been trained only to settle to stable attractor states given fixed inputs. An important goal for future work is to establish that such networks can learn to carry out more sophisticated temporal processing, such as interpreting continuously varying acoustic input in speech comprehension, and producing continuous articulatory trajectories in speech production.

Another important goal for future work is to develop models that span linguistic levels. Currently, the most connectionist models of language are restricted to processing single (often monosyllabic) words, whereas models that processes sentences adopt highly simplified (often localist) surface representations for words (also see Elman, 1993; Rohde & Plaut, submitted). In principal, the phonological model of Plaut and Kello could be extended to processes multi-word utterances, and the sentence-level model of St. John and McClelland could be elaborated with more phonologically structured inputs.

It should be clear that none of the three models described in the current paper, nor any other existing connectionist model, accounts for all of the relevant empirical findings in its domain. In considering this, it is important to think of a model as a demonstration of key theoretical principles in the service of supporting an underlying theory, rather than as a proposal for exactly how the human cognitive system operates in every detail. In this respect, the three models are quite successful, although much work remains in refining the principles and in applying them to increasingly realistic tasks.

Connectionist models provide the means of exploring the implications of a set of computational principles that are closely tied to neurophysiology and yet have important implications for cognition. In this way, the approach offers a computational bridge between mind and brain.

Acknowledgments

This research was supported by the National Institute of Mental Health (Grant MH47566) and the National Science Foundation (Grant 9720348). I thank Marlene Behrmann and Marc Joanisse for helpful comments. Correspondence may be directed to David Plaut (plaut@cmu.edu), Mellon Institute 115—CNBC, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh PA 15213–2683.

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.

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles: A problem of polysemy. *Cognitive Psychology*, *7*, 167–180.

Bybee, J. L., & Slobin, D. L. (1982). Rules and schemas in the development and use of the English past tense. *Language*, *58*, 265–289.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1985). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cleeremans, A., Servan-Schreiber, D., & McClelland, J. L. (1989). Finite state automata and simple recurrent networks. *Neural Computation*, *1*, 372–381.

Coltheart, M. (Ed.) (1996). Special issue on Phonological Dyslexia. *Cognitive Neuropsychology*, *13*, 749–934.

Cottrell, G. W., & Plunkett, K. (1991). Learning the past tense in a recurrent network: Acquiring the mapping from meaning to sounds. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 328–333). Hillsdale, NJ: Erlbaum.

Cottrell, G. W., & Plunkett, K. (1995). Acquiring the mapping from meanings to sounds. *Connection Science*, *6*, 379–412.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences*, *14*, 597–650.

Crick, F. H. C. (1989). The recent excitement about neural networks. *Nature*, *337*, 129–132.

Daugherty, K., & Seidenberg, M. S. (1992). Rules or connections? The past tense revisited. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 259–264). Hillsdale, NJ: Erlbaum.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

Eberhard, K. M., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, *24*, 409.

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

Elman, J. L. (1991). Distributed representations, 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.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–

71.

Frazier, L. (1986). Theories of sentence processing. In J. Garfield (Ed.), *Modularity in knowledge representation and natural language processing*. Cambridge, MA: MIT Press.

Graham, K. S., Hodges, J. R., & Patterson, K. (1994). The relationship between comprehension and oral reading in progressive fluent aphasia. *Neuropsychologia*, *32*, 299–316.

Grossman, M., Carvell, S., Stern, M., Gollump, S., & Hurtig, H. (1992). Sentence comprehension in Parkinson's Disease: The role of attention and memory. *Brain and Language*, *42*, 347–384.

Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition*, *56*, 61–98.

Harm, M., & Seidenberg, M. S. (in press). Phonological representations, reading, and dyslexia: Insights from a connectionist model. *Psychological Review*.

Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 77–109). Cambridge, MA: MIT Press.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98*, 74–95.

Hoeffner, J. (1992). Are rules a thing of the past? The acquisition of verbal morphology by an attractor network. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 861–866). Hillsdale, NJ: Erlbaum.

Hoeffner, J. H., & McClelland, J. L. (1993). Can a perceptual processing deficit explain the impairment of inflectional morphology in developmental dysphasia? A computational investigation. In E. V. Clark (Ed.), *Proceedings of the 25th Annual Child Language Research Forum* (pp. 38–49). Stanford, CA: Center for the Study of Language and Information.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaption in speech production. *Science*, *279*, 1213–1215.

Ingram, D. (1976). *Phonological disability in children*. London: Edward Arnold.

Jaeger, J. J., Lockwood, A. H., Kemmerer, D. L., Van Valin, Jr., R. D., Murphy, B. W., & Khalak, H. G. (1996). A positron emission tomographic study of regular and irregular verb morphology in English. *Language*, *72*, 451–497.

Joanisse, M., & Seidenberg, M. S. (1998, April). *Dissociations between rule-governed forms and exceptions: A connectionist account.* Paper presented at the 5th Annual Conference of the Cognitive Neuroscience Society, San Francisco CA.

Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, *16*, 307–354.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

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

Liberman, A. M. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.

Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Press.

Locke, J. L. (1995). Development of the capacity for spoken language. In P. Fletcher, & B. MacWhinney (Eds.), *The handbook of child language* (pp. 278–302). Oxford: Blackwell.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

MacNeilage, P. F., & Davis, B. L. (1990). Acquisition of speech production: The achievement of segmental independence. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling*. Dordrecht: Kluwer Academic.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, *40*, 121–153.

Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience*, *5*, 215–234.

Marcus, M. P. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71.

Marslen-Wilson, W. D., & Tyler, L. K. (1997). Dissociating types of mental computation. *Nature*, *387*, 592–594.

McClelland, J. L., & Kawamoto, A. H. (1986). Mechanisms of sentence processing: Assigning roles to constituents of sentences. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (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.

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

McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes*, *4*, 287–335.

McKoon, G., & Ratcliff, R. (1981). The comprehension processes and memory structures involved in instrumental inference. *Journal of Verbal Learning and Verbal Behaviour*, *20*, 671–682.

McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelilng of cognitive processes.* Oxford, UK: Oxford University Press.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*, 75–78.

Menn, L., & Stoel-Gammon, C. (1995). Phonological development. In P. Fletcher, & B. MacWhinney (Eds.), *The handbook of child language* (pp. 335–359). Oxford: Blackwell.

Oden, G. (1978). Semantic constraints and judged preference for interpretations of ambiguous sentences. *Memory and Cognition*, *6*, 26–37.

O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, *8*, 895–938.

Patterson, K., Coltheart, M., & Marshall, J. C. (Eds.). (1985). *Surface dyslexia*. Hillsdale, NJ: Erlbaum.

Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia*, *30*, 1025–1040.

Patterson, K., & Marcel, A. J. (1992). Phonological ALEXIA or PHONOLOGICAL alexia? In J. Alegria, D. Holender, J. Junça de Morais, & M. Radeau (Eds.), *Analytic approaches to human cognition* (pp. 259–274). New York: Elsevier.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1995). Goal-based speech motor control: A theoretical framework and some preliminary data. *Journal of Phonetics*, *23*, 23–35.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1991). Rules of language. *Science*, *253*, 530–535.

Pinker, S. (1994). *The language instinct*. New York: Morrow.

Pinker, S., & Mehler, J. (Eds.). (1988). *Connections and symbols*. Cambridge, MA: MIT Press.

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

Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *17*, 291–321.

Plaut, D. C., & Kello, C. T. (in press). The interplay of speech comprehension and production in phonological development: A forward modeling approach. In B. MacWhinney (Ed.), *The emergence of language*. Mahweh, NJ: Erlbaum.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.

Plunkett, K., & Marchman, V. A. (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. A. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, *48*, 21–69.

Plunkett, K., & Marchman, V. A. (1996). Learning from a connectionist model of the acquisition of the English past tense. *Cognition*, *61*, 299–308.

Quinlan, P. (1991). *Connectionism and psychology: A psychological perspective on new connectionist research*. Chicago: University of Chicago Press.

Rohde, D. L. T., & Plaut, D. C. (submitted). *Language acquisition in the absence of explicit negative evidence: How important is starting small?* Manuscript submitted for publication.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986a). Learning representations by back-propagating errors. *Nature*, *323*, 533–536.

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 (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

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

Schwartz, M. F. (Ed.). (1990). *Modular deficits in Alzheimer-type dementia.* Cambridge, MA: MIT Press.

Seidenberg, M. S. (1997). Language acquistion and use: Learning and applying probabilistic constraints. *Science*, *275*, 1599–1603.

Seidenberg, M. S., & Hoeffner, J. H. (1998). Evaluating behavioral and neuroimaging data on past tense processing. *Language*, *74*, 104–122.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Sejnowski, T. J., Koch, C., & Churchland, P. S. (1989). Computational neuroscience. *Science*, *241*, 1299–1306.

St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*, 217–257.

Studdert-Kennedy, M. (1993). Discovering phonetic function. *Journal of Phonetics*, *21*, 147–155.

Tanenhaus, M. K., & Trueswell, J. (1995). Sentence processing. In P. Eimas, & J. L. Miller (Eds.), *Handbook of perception and cognition: Language*. New York: Academic Press.

Ullman, M. T., Corkin, S., Coppola, M., Hicock, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, *9*, 266–276.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, *14*, 355–384.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child.* Oxford: Blackwell.

Wagner, K. R. (1985). How much do children say in a day? *Journal of Child Language*, *12*, 475–487.