//24

Brain and Language **68**, 110–117 (1999)
Article ID brln.1999.2106, available online at http://www.idealibrary.com on IDEAL®

The Influence of Morphological Regularities on the Dynamics of a Connectionist Network

Jay G. Rueckl and Michal Raveh

University of Connecticut and Haskins Laboratories

The effects of morphological regularities on the behavior of connectionist networks were studied by training identical networks on orthographic-semantic mappings that either contained such regularities or did not. Morphological regularities had a substantial impact on both the amount of training needed to learn a mapping and the number of words that could be included in the training set. A variety of analyses demonstrated how morphological regularities structure the organization and componentiality of a network's internal representations. © 1999 Academic Press

Key Words: connectionism; morphology; word identification.

In the past decade or so, a vigorous debate has been waged concerning the relative merits of connectionist and symbolic theories of cognition. One focus of this debate has concerned the processes underlying the production of morphologically complex words, and in particular English past tense forms (c.f., Pinker & Prince, 1988; Plunkett & Marchman, 1991; Rumelhart & McClelland, 1986). At issue has been the degree to which a network that is trained on the mapping from present- to past-tense forms, and hence comes to embody the statistical regularities of this mapping in its patterns of connectivity, provides an adequate account of the relevant experimental data.

A related body of data that has only begun to enter into this debate concerns the *perception* of morphologically complex words. It is well established that readers are influenced by the morphological structure of the words that they read (Feldman, 1991; Henderson, 1985). From a symbolic perspective, these findings are generally interpreted as revealing that the mental lexicon is organized morphologically, either in that words are represented in

This research was supported by Grant HD-01994 from the National Institute of Child Health and Development to Haskins Laboratories. Portions of this research were presented at the Annual Meeting of the Psychonomics Society, Philadelphia, PA., November 1997.

Address correspondence and reprint requests to Jay G. Rueckl, Department of Psychology, Box U-20, University of Connecticut, Storrs CT 06269. E-mail: rueckl@psych.psy.uconn.edu.

0093-934X/99 $30.00

terms of their morphological components (e.g., Taft & Forster, 1975), or in terms of the patterns of interconnectivity among morphologically related lexical entries (e.g., Feldman & Fowler, 1987). However, in several recent papers (Raveh & Rueckl, 1999; Rueckl et al., 1997) we have argued that effects of morphological structure in reading occur because statistical regularities related to morphology influence the dynamics of the processes that map representations of orthographic and phonological form to representations of meaning.

The support for this argument rests in part on empirical phenomena that are predictable from the general principles characterizing the behavior of connectionist networks, but are relatively problematic for structural accounts (see Rueckl et al., 1997, for the details). However, a more thorough evaluation of this position will require the demonstration that an implementation of these principles can provide a detailed account of a wide variety of relevant phenomena. The simulations reported here represent a first step in the development of such an account. These simulations were designed with two primary goals in mind. First, previous simulations emphasizing the influence of statistical regularities on a network's behavior (e.g., models of past-tense acquisition and models of word naming in reading, e.g. Plaut et al., 1996) have usually involved domains that are highly systematic in the sense that similar inputs are generally mapped onto similar outputs.[1] In contrast, with the exception of morphological relatives, the mapping from form to meaning is fundamentally arbitrary (i.e., similarity in form is uncorrelated with similarity in meaning). Thus, one purpose of the present simulations is to ask whether a network can exploit morphological regularities which occur against the backdrop of an otherwise unstructured mapping, or whether instead the potential influence of such regularities is overpowered by the arbitrary character of the mapping as a whole.

Assuming that the effects of such regularities are observable in a network's behavior, the second goal was to shed light on the processes underlying these effects. By taking advantage of several techniques that have been developed to "peek inside" a network's dynamics, we sought to understand how morphological regularities shape the organization of the internal "hidden" states that mediate the mapping from spelling to meaning.

In the pursuit of these goals, we conducted a series of simulations comparing the behavior of pairs of networks that differed only in the characteristics of the input–output mappings they were required to learn. For some networks, the mapping from orthographic input patterns to semantic output patterns was structured by morphological regularities; for the remaining networks, morphological regularities were eliminated, and hence the relation

[1] This generalization holds even when "exception words" (e.g., "made" in the case of past tense, "pint" in the case of word naming) are considered, because in virtually all cases exception words are exceptional in only one or two of their components.

between form and meaning was completely arbitrary. Thus, networks of the latter sort provide a control condition against which the effects of morphological regularities on both the behavior of a network and its internal dynamics can be evaluated.

METHOD

Identical three-layer (input, hidden, output) networks were trained on morphologically and arbitrarily structured form-meaning mappings. (As a simplification, the simulations focus only on the mapping from orthography to semantics, and thus the contribution of phonological processes was ignored.) Each network was trained using the back-propagation learning algorithm (Rumelhart et al., 1986), and simulations were conducted with different sets of starting weights, varying numbers of hidden units, stimulus sets of different sizes, and, for stimulus sets of a given size, different sets of input and output patterns. Due to space limitations, only the results from simulations with 40 hidden units and 200-word training sets are reported in detail, but these results are representative of the qualitative pattern that was consistently obtained across a wide range of parameters.

In the morphologically structured condition, the training vocabulary included a number of three-letter stems. For each stem, three morphologically related forms were created by concatenating the stem with three different one-letter suffixes. (The same three suffixes were conjoined with each stem.) The inputs were represented by position-specific letter nodes (McClelland & Rumelhart, 1981) such that the input representation of a stem included three active nodes (one for each letter), and the input pattern for an inflected form included four active nodes.

The construction of the output patterns began by pairing each stem with a semantic pattern. These patterns were formed by assigning a randomly selected subset of 22 output nodes an activation of 1, with the remaining nodes assigned an activation of 0. The meaning of an inflected form was represented by the activation of the semantic pattern associated with the stem, together with the activation of an additional output node that was uniquely and consistently paired with a given suffix.

The vocabularies for the networks trained on arbitrary mappings were formed by randomly repairing the input and output patterns used in the structured mappings. This method assured that the morphologically and arbitrarily structured mappings differed only in terms of the *inter*-level structure of the mapping between representations of form and meaning, and not in terms of the *intra*-level correlations among the patterns.

RESULTS

The structure of the orthographic-semantic mapping had a striking effect on both the learning rate and the capacity of the network (i.e., the size of the training set that could be learned to criterion). Networks trained on arbitrary mappings required approximately an order of magnitude more training for performance to reach criterion than did networks trained on morphologically structured mappings (see Table 1a). Similarly, for a given number of hidden units, the capacity of a network was roughly an order of magnitude smaller when the mapping from form to meaning was arbitrarily structured than when the mapping was structured by morphological regularities. For example, with 40 hidden units and arbitrary mappings, the 200-word training sets push the limits of a network's capacity. In contrast, networks of the same

TABLE 1
The Results of Simulations Comparing Morphologically Structured and Arbitrary Mappings from Form to Meaning

Measure	Mapping: Morphologically structured	Mapping: Arbitrary
(a) Learning rate		
Sweeps to criterion	79	2212
(b) Between-level correlations		
Input–output	.25	.00
Input–hidden	.83	.65
Hidden–output	.37	.10
(c) Hidden-level correlations		
Stem-(pseudo)relative	.92	.50
(d) Contribution correlations		
Stem	.99	.51
Suffix	.95	.44

size could easily learn morphologically structured mappings including 1000 words (the largest training sets we employed).

These findings demonstrate (perhaps not surprisingly) that the structure provided by morphological regularities has a strong influence on a network's global behavior. However, by themselves these results say little about how a network capitalizes on such morphological regularities. Because this question is central to the network account, we conducted a number of additional analyses intended to shed light on the manner in which statistical regularities shape a network's dynamics. The primary goal of these analyses was to determine how morphological regularities influence the organization of the hidden-layer representations that mediate the mapping from form to meaning.

The first set of analyses examined the relationship between similarity in form or meaning and similarity at the hidden level. If patterns of activation are treated as vectors, the similarity of two patterns is given by their correlation: identical patterns have a correlation of 1, and orthogonal patterns have a correlation of 0. Thus, for each pair of words separate measures of similarity can be taken at the orthographic, hidden, and semantic levels. The correlation of these correlations provides an index of the degree to which similarity at one level of representation is related to similarity at another level.

As can be seen in Table 1b, the between-level correlations were consistently higher for networks trained on morphologically structured mappings. The difference in the input–output correlations merely reflects the manner in which the stimulus sets were constructed: Morphological regularities create a correlation between similarity in form and meaning, whereas in arbitrary mappings, form and meaning are unrelated. The input-hidden and hidden-output correlations are of more theoretical interest in that they reveal the self-organizing properties of adaptive networks. More systematic mappings (in the present case the consequence of morphological regularities) create a stronger relationship between the organization of the hidden-layer representations and the similarity structure of the input and output patterns (also see Plaut et al., 1996; Rueckl et al., 1989).

Between-level correlations provide a nonspecific index of the effects of morphology on the organization of the hidden-layer representations. That is, although the obtained differences on these measures are clearly the product of morphological regularities, both morphologically related and morphologically unrelated pairs of words contribute to these measures. A more specific index of the effects of morphological regularities can be obtained by measuring the similarity of the hidden representations of a root word and its affixed forms. As Table 1c illustrates, the hidden patterns for morphological relatives (in morphologically structured mappings) were substantially more similar than were the representations of word pairs of comparable orthographic similarity in the absence of any morphological relations (i.e., in arbitrary mappings). Thus, the presence of morphological regularities engenders an organization of the hidden-layer representations in which the patterns for morphological relatives are clustered together in the hidden-layer state space.

A final set of analyses examined the componentiality of the hidden representations. These *contribution correlation* analyses (Plaut et al., 1996) provide a measure of the degree to which the morphological components of a word are represented independently of one another. Determining the contribution correlation of the stem of a morphologically complex word requires two comparisons. First, the hidden pattern for the whole word is compared to the hidden pattern in a baseline condition where the input pattern consists of the suffix alone. The difference between these patterns represents the contribution of the stem to the representation of the whole word. Second, the hidden pattern for an input consisting of the stem alone is compared to the hidden pattern in a baseline condition where no input is presented (i.e., where all the orthographic nodes have an activation of 0). The correlation of these differences provides a measure of the degree to which the contribution of the stem to the hidden representation is context-independent: The higher the correlation, the greater the degree to which the hidden representation of the stem is independent of the representation of the suffix. The context-sensitivity of the representation of a suffix can be determined by an analogous procedure.

The results of these analyses are presented in Table 1d. The contribution correlations were much higher for networks trained on morphologically structured mappings than for networks trained on arbitrary mappings. These results indicate that networks that learn morphologically structured mappings represent the morphological constituents of a word in a relatively context-independent fashion. Put another way, such networks make use of *compo-*

nential representations (Plaut et al., 1996), in which stem and suffixes are represented by distinct hidden-layer subpatterns.

GENERAL DISCUSSION

Morphological structure has a strong influence on the dynamics of networks that learn to associate orthographic and semantic representations. As the simulations demonstrate, morphological regularities increase the rate of learning and allow for larger vocabularies. These behavioral effects arise because the organization of the network's internal representations reflects statistical regularities in the environment. In the present case, morphological regularities allow the similarity structure of the hidden representations to mirror the organization of the orthographic and semantic representations. Consequently, morphologically related words are represented by similar hidden representations, and in the case of multimorphemic words these representations have a componential structure such that the pattern for a word is formed by the superimposition of subpatterns corresponding to each of its morphological constituents.

The simulations reported here represent the first step toward a network account of experimental findings concerning the effects of morphology on word identification. Building on this initial work, we have more recently demonstrated that the exploitation of morphological regularities is not disrupted by irregular morphology (e.g., *ran–run*) or pseudoaffixation (e.g., *summer;* see Plaut et al., 1996; Rumelhart & McClelland, 1986, for similar results in other domains), and we have begun to explore the effects of factors such as word frequency and semantic transparency on the degree to which hidden representations are organized in a componential fashion. Ultimately, we expect that an understanding of the influence of these and other factors on a network's dynamics will form the basis for a connectionist account of the varied experimental findings (see Feldman, 1991, and Henderson, 1985, for reviews) that reveal the influence of morphological structure on word identification.

We close with a brief discussion of what we see as an important aspect of the connectionist approach that is exemplified by our work. As noted in the Introduction, in virtually all traditional accounts morphological effects are taken to reflect the structural organization of the lexicon. Thus, behavioral organization is explained in terms of internal organization, and the task of the theoretician is to "reverse-engineer" this internal organization on the basis of patterns of behavior. In the connectionist approach, in contrast, although the effects of morphological structure on behavior are thought to be a consequence of the organization of the hidden representations, this internal organization does not correspond to "the structural properties of the lexicon" in any traditional sense. Instead, componential representations of the sort observed in our simulations are "softly assembled," emerging on-line as a consequence of the activation and learning dynamics that control a network's behavior. On this view, internal organization serves a dual role as part of the explanation of behavior and as a phenomenon requiring explanation in its own right, and the task of the theoretician is to identify the forces that give rise to this organization.

The contrast we are drawing is the contrast Köhler (1947) made between "machine theories" and "dynamics," and it is striking that many of the same criticisms that Gestaltists raised about the machine theories of their day parallel the issues that connectionists now raise about structural theories of lexical organization. In our view, network models differ from traditional accounts not only in terms of how they explain behavior, but more fundamentally in terms of their conceptions of psychological explanation. Thus, an important point about the present simulations is that they exemplify the emphasis that connectionist theories place on the dynamic processes that determine, and are determined by, interrelationships among the organization of behavior, the organization of the mind, and the organization of the environment.

REFERENCES

Feldman, L. B. 1991. The contribution of morphology to word recognition. *Psychological Research,* **53,** 33–41.

Feldman, L. B., & Fowler, C. A. 1987. The inflected noun system in Serbo-Croatian: Lexical representation of morphological structure. *Memory and Cognition,* **15**(1), 1–12.

Henderson, L. 1985. Towards a psychology of morphemes. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 15–72). London: Erlbaum.

Kohler, W. 1947. *Gestalt psychology.* New York: Liverright.

McClelland, J. L., & Rumelhart, D. E. 1981. An interactive activation model of context effects in letter perception: Part 1. *Psychological Review,* **88,** 375–407.

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., 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. 1991. U-shaped learning and frequency effects in a multilayered perceptron: Implications for child language acquisition. *Cognition,* **38,** 1–60.

Raveh, M., & Rueckl, J. G. 1999. *Equivalent effects of inflected and derived primes: Long-term morphological priming in fragment completion and lexical decision.* Manuscript submitted for publication.

Rueckl, J. G., Cave, K. R., & Kosslyn, S. M. 1989. Why are "What" and "Where" processed by two cortical visual systems? A computational investigation. *Journal of Cognitive Neuroscience,* **1,** 171–186.

Rueckl, J. G., Mikolinski, M., Raveh, M, Miner, C., & Mars, F. 1997. Morphological priming, fragment completion, and connectionist networks. *Journal of Memory and Language,* **36,** 382–405.

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., Hinton, G., & Williams, R. 1986. Learning internal representations by error propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the microstructure of cognition. Volume I: Foundations.* Cambridge, MA: MIT Press.

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

Stone, G. O., & Van Orden, G. C. 1994. Building a resonance framework for recognition using design and system principles. *Journal of Experimental Psychology: Human Perception and Performance,* **20,** 1248–1268.

Taft, M., & Forster, K. I. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior,* **14,** 638–647.