1a. Lastname = Mauner
1b. Firstname = Gail
1c. Department = Psychology
1d. UB Card Barcode = 29072006506469
1e. Campus Address = 368B Park Hall
1f. Campus Phone = 645-3650 x 368
1g. Home Phone = 838-6948
1h. E-mail = mauner@acsu.buffalo.edu
1i. Fax = 645-3801
2a. Title = Brain and Language
2b. Call Number = no call #, shelved by title
2c. Volume = 59
2d. Issue = 2
2e. Date = 1997
2f. Pages = 334-366
2g. Author = Miikkulainen, R
2h. Aticle Title = Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon
Deliver = Email
LibraryOwn = HSL
status = Faculty

Sending Host: hamachi.psy.buffalo.edu
Sender's Client: Mozilla/4.75C-CCK-MCD {C-UDP; EBM-APPLE} (Macintosh; U; PPC)
Original URL:

DOC EXP
E-MAIL

BRAIN AND LANGUAGE 59, 334–366 (1997)
ARTICLE NO. BL971820

Dyslexic and Category-Specific Aphasic Impairments in a Self-Organizing Feature Map Model of the Lexicon

Risto Miikkulainen

Department of Computer Sciences, The University of Texas at Austin

DISLEX is an artificial neural network model of the mental lexicon. It was built to test computationally whether the lexicon could consist of separate feature maps for the different lexical modalities and the lexical semantics, connected with ordered pathways. In the model, the orthographic, phonological, and semantic feature maps and the associations between them are formed in an unsupervised process, based on cooccurrence of the lexical symbol and its meaning. After the model is organized, various damage to the lexical system can be simulated, resulting in dyslexic and category-specific aphasic impairments similar to those observed in human patients. © 1997 Academic Press

INTRODUCTION

The human lexical system is believed to be highly modular, consisting of a central semantic component and separate symbol memories for the different input and output modalities (Caramazza, 1988; McCarthy & Warrington, 1990). Such an architecture is intuitively compelling since the modalities give rise to different representations, and they are processed through different neural structures. Considerable experimental evidence also supports the dissociation of lexical components. Modularity therefore forms a good guideline for building a computational model of the lexical system.

How are the individual components implemented in the brain? Not much is known about the structures underlying higher functions such as the lexicon. However, the perceptual mechanisms are very well understood, and they appear to be organized around topological maps. For example, nearby regions in the mammalian primary visual cortex respond to nearby regions in the retina (Hubel & Wiesel, 1959, 1965). Similar topological maps are known to exist in other sensory systems and motor systems as well (Knudsen et al., 1987), and it is quite possible that higher-level information is also represented in a similar manner. However, higher areas of the brain represent abstract information, and it is difficult to establish to what features a particular neuron is sensitive, let alone determine whether the sensitivity of neurons in a particular area forms a topological organization. It has been possible to locate cells that are responsive to particular faces and facial expressions, as well as neurons that respond selectively to different words, and these cells appear to form localized groups (Hasselmo et al., 1989; Heit et al., 1989; Rolls, 1984).

Indirect evidence for localization in the lexical system comes from patients with brain lesions. A number of patients have impairments of specific syntactic or semantic categories, such as concrete words, inanimate objects, and names of fruits and vegetables (Caramazza, 1988; Hart et al., 1985; Warrington & Shallice, 1984). Such impairments could result from localized damage to a topological map that lays out the semantic properties of words.

These observations form the motivation for the DISLEX model of the human lexical system. The main hypothesis to be tested computationally is that the lexical system consists of multiple topological feature maps, each either representing the symbols within one modality or laying out the word semantics. In the experiments reported in this paper, the DISLEX model was first organized based on examples of desired input–output behavior, and then subjected to simulated psycholinguistic experiments under various neural damage. The dyslexic and aphasic behavior observed in the model as a result was consistent with those of human patients. Because such behavior emerges automatically from the DISLEX architecture (and is not programmed in per se), it constitutes computational support for the hypothesis. The model also predicts that form-specific impairments would be possible in the human lexical system.

The orthographic and semantic components of DISLEX were used as the lexicon for the DISCERN subsymbolic story processing system (Miikkulainen, 1993). This paper describes the first full implementation of DISLEX, including the phonological modality as well. Below, an overview of DISLEX is first given. The orthographic, phonological, and semantic representations used in the model are reviewed, followed by an analysis of the topological maps in DISLEX and the mechanisms for associating lexical symbols with their meanings. The behavior of DISLEX is illustrated focusing on priming and disambiguation, dyslexic impairments, and category-specific aphasic impairments. A discussion of the limitations of the model and future research directions concludes the paper.

OVERVIEW OF THE DISLEX MODEL

DISLEX consists of two main parts: memories for the lexical symbols in the different input and output modalities, and the memory for the lexical

This research was supported in part by National Science Foundation Grant IRI-9309273 and by Texas Higher Education Coordinating Board Grant ARP-444. Thanks to Jon Hilbert for obtaining the phonological word representations for DISLEX.

Address reprint requests to Risto Miikkulainen, Department of Computer Sciences, The University of Texas at Austin, Austin TX 78712, risto@cs.utexas.edu.

0093-934X/97 $25.00

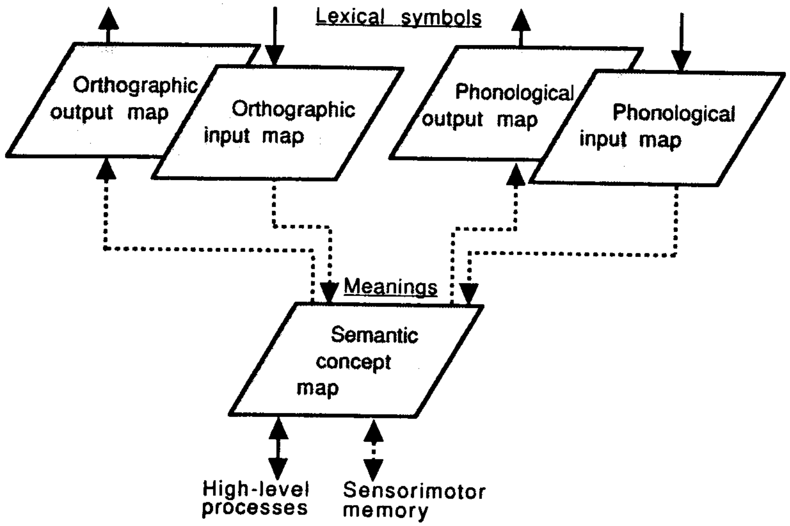


FIG. 1. The DISLEX model of the human lexical system. The lexical symbol memories are modality and direction specific. Dashed lines indicate associative pathways, solid lines propagation of distributed representations.

semantics (Fig. 1). The symbol memories store distributed representations (vectors of gray-scale values between 0 and 1) for the orthographic and phonological word symbols that are used in communication with the external world. For example, the orthographic representation for DOG consists of the visual form of the letters D, O, and G, while the phonological representation stands for the string of phonemes /d/, /Q/, and /g/. The semantic memory consists of distributed representations of distinct concepts. For example, the concept dog refers to a specific animal and contains information such as domestic, mammal, brown, and so on. There is a pathway from the semantic memory to higher-level systems such as language processing and episodic memory, which use the semantic representations. The semantic memory is also connected to sensorimotor memory, which contains visual images of objects and other perceptual and motor information. This pathway allows nonlinguistic access to the semantic memory, and provides a possible means for symbol grounding.

The symbol memories and the semantic memory are implemented as feature maps. There is one map for each input and output modality and one for the semantic memory. Each unit in a feature map represents a word (i.e., a symbol or a concept) in two ways: (1) each unit has an internal parameter vector, also called the input weight vector, which stores a distributed representation for a word, and (2) each unit is a local representation for that word on the map. The maps lay out each high-dimensional distributed representation space on a 2D network so that the similarities between words become

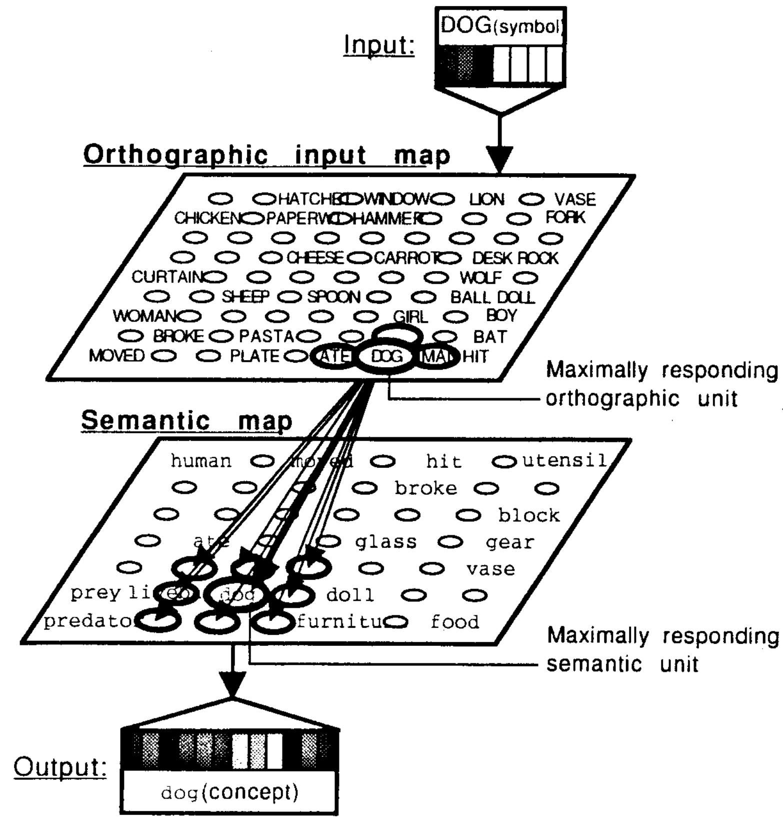


FIG. 2. Lexicon propagation. The orthographic input symbol DOG is translated into the semantic concept dog in this example. The representations are vectors of gray-scale values between 0 and 1, stored in the weights of the feature map units. The size of the unit on the map indicates how strongly it responds. Only a few strongest associative connections of the orthographic input unit DOG (and only that unit) are shown.

apparent (Fig. 2). Lexical symbols with similar form, such as BALL and DOLL, are represented by nearby units in the symbol map. In the semantic map, semantic concepts with similar content, such as livebat and prey, are mapped near each other. When a word representation is input, the units in the map respond according to how similar their input weight vector is to the word representation. The maximally responding unit is taken to represent the input word on the map, and its label indicates the classification of the input. For example, an orthographic input pattern may be recognized this way as an instance of the symbol DOG (Fig. 2).

The symbol maps are densely connected to the semantic map with one-way associative connections (Fig. 2). Each symbol unit is initially connected to all semantic units, although only a small subset of those connections re-main effective after learning. A localized activity pattern representing a symbol (e.g., DOG) in an input map will cause a localized activity pattern to form in the semantic map, representing the meaning of the symbol. The units in the semantic map have input weight vectors just like the symbol units, and these vectors represent distinct meanings. After the maximally active semantic unit has been found (the one labeled dog), the corresponding semantic representation is obtained from the input weights of this unit (i.e., by propagating its activity through the input weights to the output of the semantic map).

In addition to activation through the associative connections, units on the semantic map can be activated in a normal feature map manner through the input connections. If a particular meaning is to be output, its distributed representation is given as input to the semantic map. A localized activity pattern results, and activation propagates through the associative connections to the output maps. The maximally activated symbol map unit then stands for the symbol corresponding to the input meaning. Its distributed representation is obtained from the input weights of the maximally activated symbol map unit. The lexicon thus transforms a symbol representation into a semantic representation, and vice versa, and serves as an input/output filter for language processing.

The symbol and concept maps are organized and the associative connections between them are formed simultaneously in an unsupervised learning process by presenting the system with cooccurring lexical symbols and their meanings. Before discussing the details of the self-organizing process, let us look at how the symbols and concepts are represented.

REPRESENTING SYMBOLS AND MEANINGS

As customary in artificial neural network models, both the symbols and meanings are represented distributively as feature vectors, or vectors of gray-scale values between 0 and 1. It is the similarities among these vectors that determines the organization of the lexicon, and therefore they must be designed so that they capture the essential similarities in the domain.

Symbol Representations

It is reasonable to assume that the neural representations in each lexical symbol modality reflect the structure of the physical symbols for which they stand. Therefore in DISLEX, the orthographic representations reflect the visual similarities among the written words, and spoken words that sound similar are represented by similar phonological feature vectors. The same orthographic and phonological representations are used for both input and output. While certainly the motor representations for the symbols are different from their perceptual counterparts, it is assumed that they encode essentially the

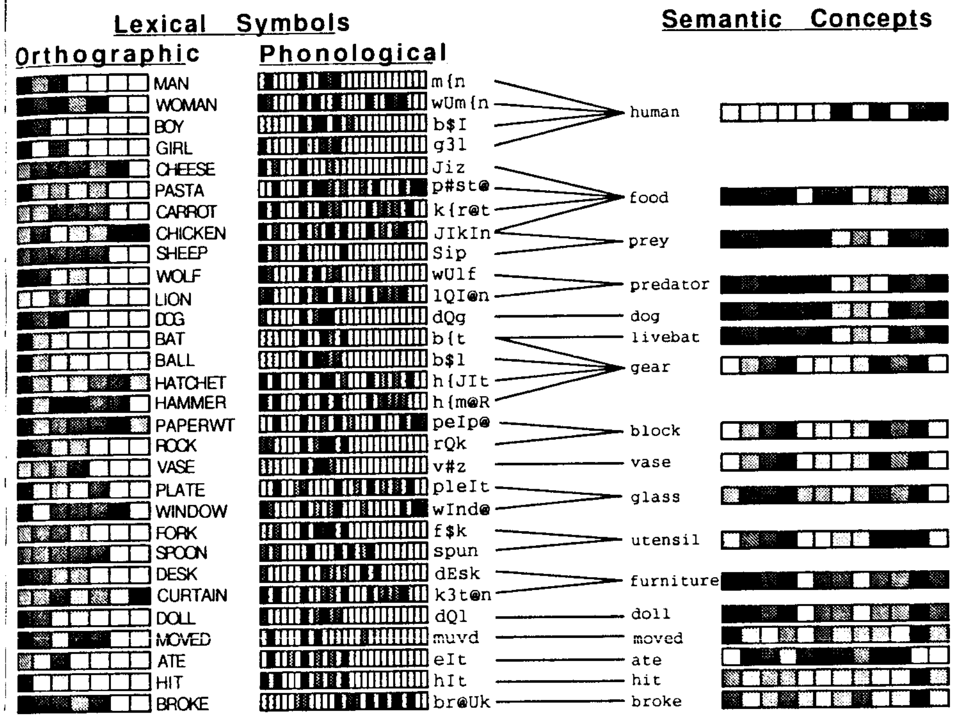


Fig. 3. The training data for the lexicon. Orthographic representations are blurred bitmaps of the orthographic words and phonological representations consist of concatenations of phoneme representations. Concept representations were developed by FGREP in the case-role assignment task and stand for distinct meanings. Gray-scale boxes indicate component values between 0 and 1. The connections depict the mapping between the symbols and their meanings. Many concepts map to several synonymous lexical symbols, and the homonymous symbols CHICKEN and BAT map to two distinct concepts each. The orthographic and phonological symbols correspond one-to-one to each other in this data.

same information and can be approximated by representations of the physical properties of the symbols.

In the orthographic domain, each letter of the alphabet was given a value between 0 and 1 according to its darkness, measured by the number of black pixels in its bitmap representation (Appendix A). The word representation vectors were then formed by concatenating the darkness values of the individual letters (Fig. 3). This encoding scheme is very simple and leaves out many orthographic details: in effect, the representations stand for extremely blurred pictures of the words. This scheme was chosen over more complicated ones for two reasons: (1) In the orthographic domain, there is no obvious more accurate alternative that would capture the similarities any better. (2) This scheme is quite adequate for the DISLEX task and data. Each written word symbol has a unique representation, and similar symbols have similar representations (Figure 3).

A slightly more detailed encoding was employed in the phonological domain, not because more details were necessary, but because such an encoding is standard in this domain. Each phoneme was represented as a feature vector according to the International Phonetic Alphabet, with numeric values coding the features of place and manner of articulation, sound, chromaticity, and sonority (Appendix B). The phoneme representation vectors were then concatenated to form the word representations (Fig. 3). Again, each phonological word symbol has a unique representation in the resulting vectors, and similar words have similar representations.

Concept Representations

The semantic concept representations stand for distinct meanings in the language. Although it is possible to encode meanings by hand as feature vectors (see, e.g., McClelland & Kawamoto, 1986), as was done for symbols, it is difficult to decide what the appropriate semantic features should be.

With the FGREP-mechanism (Miikkulainen, 1993; Miikkulainen & Dyer, 1991), it is possible to derive a distributed encoding automatically, based on examples of how the words are used in the language. An FGREP-network is a three-layer backpropagation (Rumelhart et al., 1986) network, where part of a task is to modify the input representations so that they best support the task. Representations for items that are used in similar ways in the training examples become similar, and in this sense, FGREP representations can be claimed to stand for the meanings of the input items.

The semantic representations for DISLEX were formed with an FGREP network in the sentence case-role assignment task of McClelland and Kawamoto (1986). A number of sentence examples were generated based on a set of templates and semantic categories (listed in Appendix C). Only concepts that represented unique meanings among the lexical symbols were used. For example, MAN, WOMAN, BOY, and GIRL were used exactly the same way in the data, and therefore they were considered instances of the same concept: `human`. On the other hand, CHICKEN had two distinct meanings: `food` and `prey`. Such ambiguities between symbols and meanings were set up intentionally to make the lexicon mapping more interesting. The input to the network consisted of the syntactic assignments of the sentence (e.g., subject = `human`, verb = `ate`, object = `food`), and the network was trained to assign the correct semantic case roles for them (agent = `human`, verb = `ate`, patient = `food`).

Starting with initially random representations and weights, the FGREP network was trained with 0.1 learning rate for 2000 epochs and with 0.05 for an additional 250 epochs, at which point the average output error E_{avg} was 0.015. As a side effect of learning the case-role assignment task, the

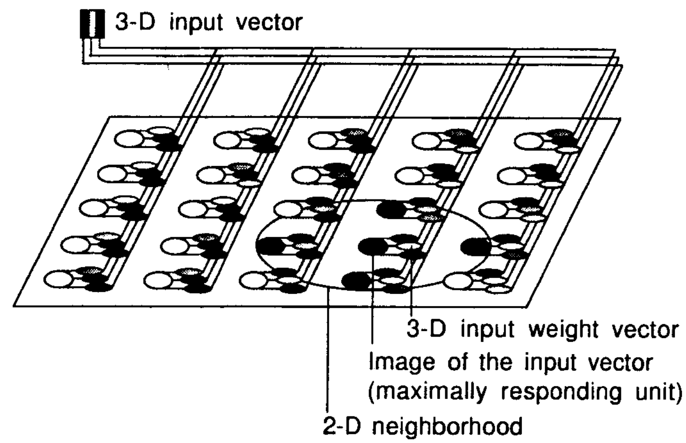


FIG. 4. A self-organizing feature map network. This network implements a mapping from a 3-dimensional input space onto a two-dimensional location in the network. The values of the input components, weights, and the unit output are indicated by gray-scale coding.

network developed representations for the input concepts (Fig. 3). Words that belong to the same semantic category (such as animals, hitters, etc.) have a number of uses in common, and their representations have become similar. The total usage is different for each concept, and consequently their representations are different. They stand for unique meanings.

The ambiguities between linguistic symbols and their meanings are shown explicitly in the many-to-many mapping between unique symbols and unique meanings of Fig. 3. These data were used to organize the lexicon.

SYMBOL AND CONCEPT MEMORIES

The lexicon components are implemented as feature maps. The basic idea of self-organizing feature maps is first briefly reviewed below, followed by a description of the symbol and concept maps in DISLEX and their properties.

Self-Organizing Feature Maps

A 2-D topological feature map (Kohonen, 1989, 1990) implements a topology-preserving mapping from a high-dimensional input space onto a 2-D output space. The map consists of an array of processing units, each with N weight parameters (Fig. 4). The map takes an N-dimensional vector as its input and produces a localized pattern of activity as its output. In other words, the input vector is mapped onto a location on the map.

Each processing unit receives the same input vector and produces one output value. The response is proportional to the similarity of the input vector and the unit's weight vector. The unit with the largest output value constitutes the image of the input vector on the map. The weight vectors are ordered in such a way that the output activity smoothly decreases with the distance from the image unit, forming a localized response (an activity "bubble").

The weight vectors approximate specific items of the input space in such a way that topological relations are retained. This means roughly that nearby vectors in the input space are mapped onto nearby units on the map. This is a very useful property, because the complex similarity relationships of the high-dimensional input space (such as a word representations) become visible on the map.

The organization of the map (i.e., the assignment of the weight vectors) is formed in an unsupervised learning process (Kohonen, 1982b, 1989). The input items are randomly drawn from the input distribution and presented to the network one at a time (Fig. 4). The map responds to each vector by developing a localized activity pattern. The weight vector of the maximally responding unit and each unit in its neighborhood are changed toward the input vector, so that these units will produce an even stronger response to the same input in the future. This way, the map adapts in two ways at each presentation: (1) the weight vectors become better approximations of the input vectors, and (2) neighboring weight vectors become more similar. Together these two adaptation processes eventually force the weight vectors to become an ordered map of the input space. The process begins with very large neighborhoods, that is, the weight vectors change in large areas. This results in a gross ordering of the map. The size of the neighborhood and the learning rate decrease with time, allowing the map to make finer and finer distinctions between items.

There are several alternatives for implementing similarity metric, neighborhood selection, and weight change in feature maps. A biologically plausible process would be based on weighted sum of the input, lateral inhibition and redistribution of synaptic resources (Kohonen, 1982b; Sirosh & Miikkulainen, 1994). These mechanisms can be abstracted and replaced with computationally more efficient ones without obscuring the process itself. The similarity in DISLEX is measured by Euclidian distance, the neighborhood consists of a square area around the maximally responding unit, and the weight changes are proportional to the Euclidian difference. More specifically, the output η_{ij} of unit (i,j) in a lexicon map is

$$\eta_{ij} = \begin{cases} 1 - \dfrac{\|\mathbf{x} - \mathbf{m}_{ij}\| - d_{\min}}{d_{\max} - d_{\min}} & \text{if } (i,j) \in N_c, \\ 0 & \text{otherwise,} \end{cases} \qquad (1)$$

where $\mathbf{x}$ is the symbol or concept representation vector, $\mathbf{m}_{ij}$ is the weight vector of unit (i,j), N_c is the neighborhood around the image unit c (defined as the set of units within a certain vertical and horizontal distance from c), and d_{min} is the smallest and d_{max} the largest distance of $\mathbf{x}$ to a unit in the neighborhood. This formula generates a regular concentrated activity pattern around the maximally responding unit.

With $\alpha(t)$ as the gain, the weight components are changed according to the input vector–weight vector difference:

$$\mu_{ij,k}(t+1) = \begin{cases} \mu_{ij,k}(t) + \alpha(t)[\xi_k(t) - \mu_{ij,k}(t)] & \text{if } (i,j) \in N_c(t), \\ \mu_{ij,k}(t) & \text{otherwise,} \end{cases} \qquad (2)$$

where the neighborhood $N_c(t)$ shrinks with time.

Symbol and Concept Maps

The symbol and concept maps in DISLEX were organized independently and simultaneously, so that associative connections between them could be developed at the same time (as discussed in the next section). Because the same symbol representations were used for both input and output in each modality, the input and output maps developed the same order. This common organization is referred to as the orthographic map and the phonological map below.

During self-organization, each lexical symbol ↔ semantic concept representation pair (figure 3) was presented to the appropriate maps 150 times in random order. The same learning rate $\alpha(t)$ was used for all maps and associative connections (Eqs. 2 and 3). The learning rate was linearly decreased from 0.1 to 0.05 during the first 50 epochs, then to 0.0 during the remaining 100 epochs. At the same time, the neighborhood radii on all maps were decreased from 4 to 1 and then from 1 to 0.

In the self-organizing process, the symbol and concept representations become stored in the weights of the feature map units. Each orthographic and phonological symbol has an image unit on the appropriate symbol map, and this unit's weight vector equals the representation for that word. Semantic concepts are represented in the same manner. The weight vectors of intermediate units represent combinations of representations. For example, an unlabeled semantic unit between `dog` and `predator` has features of both domestic and carnivorous animals.

All final maps exhibit hierarchical knowledge organization (Fig. 5). Large areas are allocated to different categories of words, and each area is divided into subareas with finer distinctions. The symbol maps become mainly organized according to word length. There are separate, adjacent areas for orthographic symbols with 3, 4, 5, 6, and 7 characters, and words with 3, 4, and 5 phonemes. Within these areas, similar words are mapped near each other. For example, BAT is mapped between BOY and HIT and DOLL is located next to BALL in the orthographic map. Similarly in the phonological map, /dQg/ and /dQ1/ and /b$I/ and /b$1/ are mapped near each other.

The semantic map has three main areas: verbs, animate objects, and inanimate objects. Finer distinctions reveal the semantic categories used in generating the sentence examples for FGREP (Table 5). For example, there are subareas for hitters, possessions, and fragile-objects, with vase, which belongs to all these categories, at the center. Note that the categorization was not directly accessible to the FGREP network or the feature map at any point. It was only implicitly represented by the sentences that were input to the FGREP network. The categories were extracted by FGREP, coded into the representations, and finally visualized on the semantic feature map. The final map reflects both the syntactic and the semantic properties of the words.

In the self-organizing process, the distribution of the weight vectors becomes an approximation of the input vector distribution (Kohonen, 1982a, 1989; Ritter, 1991; Ritter & Schulten, 1986). More weight vectors are allocated to dense areas of the input space, and as a result these areas are magnified (represented to greater detail) on the map. This can be clearly seen in the word maps. For example, the semantic representations for the different animals are very similar, spanning only a very small part of the representation space (Fig. 3), yet a relatively large area is allocated for animals on the semantic map.

The two dimensions of the map do not necessarily stand for any recognizable features of the input space. They develop automatically to facilitate best discrimination between input items. The map tries to approximate high-dimensional similarities with space-filling (Peano) surfaces and tries to fill the whole area of the map with data. As a result, the ordered areas on the map are likely to have complicated and intertwined, rather than compact and regular, shapes. This is the case in all the maps in DISLEX.

Feature Maps as Lexicon Components

To conclude, self-organizing feature maps have several properties that make them a good model for the lexical system:

1. The classification performed by a feature map is based on a large number of parameters (the weight components), making it very robust. Incomplete or somewhat erroneous word representations can be correctly recognized.

2. Once an inexact word symbol or concept is recognized, it is possible to recover its exact representation from the weights of the image unit. In other words, categorical perception can be modeled.

3. The map tends to be continuous, containing many intermediate units that represent items between established categories. In other words, words can have soft boundaries.

4. Several items can be active on the map at the same time, which means that different alternatives (synonyms, or ambiguous meanings) can be represented distinctly and in parallel. With connections between different maps, many-to-many mappings are possible.

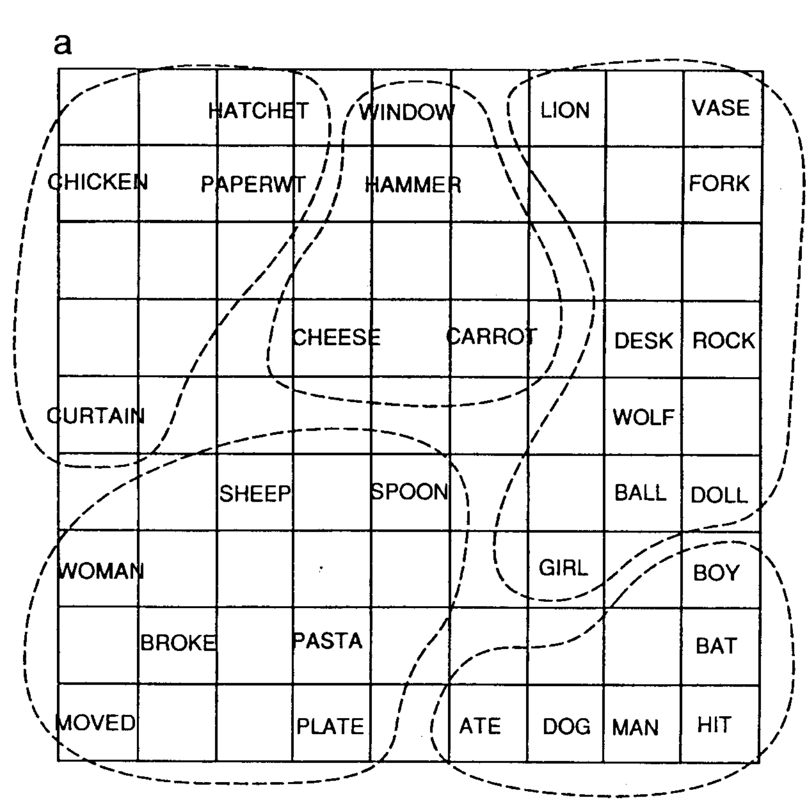
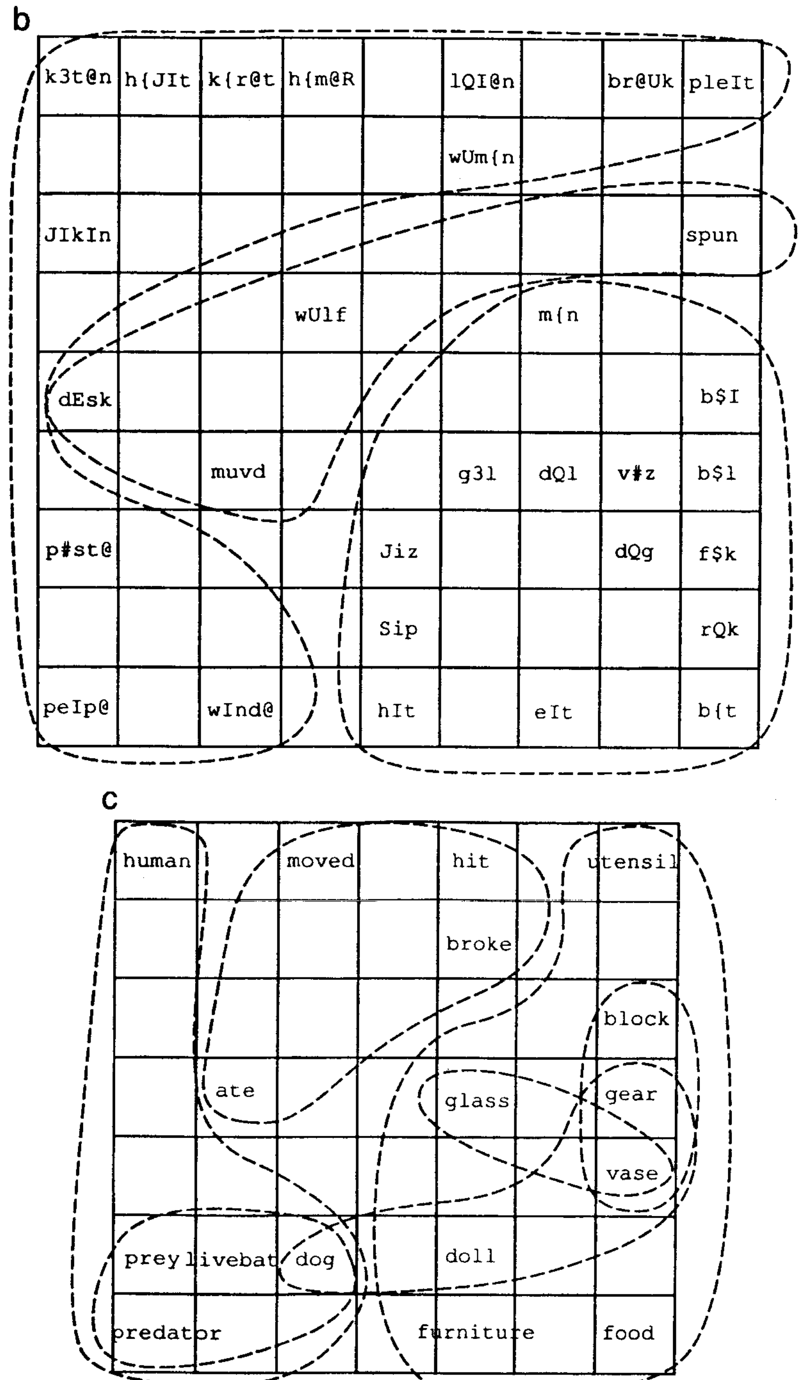


FIG. 5. The orthographic, phonological, and semantic maps. The input and output maps in each modality have the same order, shown here only once in (a) and (b). (a) Orthographic map. Each unit in the 9×9 network is represented by a box, and the labels indicate the image unit for each symbol representation. The map is divided into major subareas according to word length. (b) Phonological map. The labels indicate the images for each phonological word representation. Again, the word length is the major ordering factor. (c) Semantic map. The labels on this 7×7 map indicate the maximally responding unit for each concept representation. The map is organized according to the semantic categories (Table 5).

5. The differences of the most frequent input items are magnified in the mapping, i.e., the variations of the most common word meanings or surface forms are more finely discriminated.

6. The self-organizing process requires no supervision and makes no assumptions on the form or content of the words. The properties of the representations which provide the best discrimination are determined automatically.

In the following sections, it will be shown how the topological organization of the map leads to plausible dyslexic and aphasic behavior under simulated damage. However, first we need to discuss how the associative connections between maps translate symbols to concepts and vice versa.



ASSOCIATING SYMBOLS AND CONCEPTS

The mapping between symbols and concepts is many-to-many. Some words have multiple meanings (homonyms), and sometimes the same meaning can be expressed with several different symbols (synonyms). For example, in the DISLEX training data the lexical symbol CHICKEN could mean a living chicken or food. Similarly, BAT could be a baseball bat or a living bat. There are also several groups of synonymous words in the data. For example, MAN, WOMAN, BOY, and GIRL all have the same meaning `human`, and WOLF and LION are both `predators`. The many-to-many mapping between symbols and meanings is implemented with associative connections between the symbol and concept maps.

There is a unidirectional associative connection from each unit in the orthographic and phonological input maps to each unit in the semantic map, and from each unit in the semantic map to each unit in the orthographic and phonological output maps. The connection weight indicates the strength of the association between the symbol and the concept.

The symbol and concept maps and the associative connections between them are organized simultaneously by presenting examples of symbol–concept pairs (listed in Fig. 3). Such a training scheme models the training data in the real world. In any particular processing context, only one of the synonyms or homonyms is active, but different mappings are possible at different times. The many-to-many mapping must be learned from these individual examples.

The distributed representation for the symbol is presented to the appropriate symbol map, which develops a localized activity pattern around the image unit (Eq. 1). Ordinary feature map adaptation then takes place within the neighborhood. At the same time, the representation for the corresponding concept is input to the semantic map, which develops a similar localized response, and the feature map weight vectors adapt within the neighborhood. At this point, both maps display localized patterns of activity. The lexicon learns to associate them by their cooccurrence, that is, through Hebbian learning (Hebb, 1949; Hertz et al., 1991; Gustafsson & Wigström, 1988). The weights between active units are increased proportional to their activity:

$$\Delta w_{ij,uv} = \alpha(t)\eta_{S,ij}\eta_{D,uv}, \tag{3}$$

where $w_{ij,uv}$ is the unidirectional weight between the source map unit at location (i,j) (either symbol or concept) and the destination unit at (u,v) (concept or symbol), and $\eta_{S,ij}$ and $\eta_{D,uv}$ indicate the activities of these units. As is common with Hebbian learning, the associative weight vectors are then normalized:

$$w_{ij,uv}(t+1) = \frac{w_{ij,uv}(t) + \Delta w_{ij,uv}}{\{\sum_{u,v}[w_{ij,uv}(t) + \Delta w_{ij,uv}]^2\}^{1/2}}. \tag{4}$$

Normalization is carried out over all associative connections of the source unit, and its effect is to decrease the strengths of the connections to less active units. The process corresponds to redistribution of synaptic resources, where the synaptic efficacy is proportional to the square root of the resource (Sirosh & Miikkulainen, 1994).

Initially, the activity patterns on the symbol and semantic maps are large, and associative weights are changed in large areas. As the maps become ordered, the associations become gradually more focused. The final associative connections form a continuous many-to-many mapping between the maps. Unambiguous symbols and concepts have focused connections, as shown in Fig. 6. If a symbol has several meanings, or one meaning can be expressed with several synonyms, there are several groups of strong connections (Fig. 7). Units located between image units tend to combine the connectivity patterns of nearby words (Fig. 7).

The associative connections are responsible for translating a symbol to its semantic counterpart, and vice versa. The activity in one map propagates through the connections and causes an activity pattern to form in the other map:

$$\eta_{D,uv} = g(y_{uv}) = g\left(\sum_{i,j} w_{ij,uv}\eta_{S,ij}\right), \tag{5}$$

where $w_{ij,uv}$ stands for the weight between the source map unit (i,j) and the destination map unit (u,v), and $\eta_{S,ij}$ and $\eta_{D,uv}$ indicate the activities of these units. The activation function $g(y) = y/y_{max}$, where y_{max} is the largest of the weighted sums y to the destination map. This function scales the activity linearly within 0 and 1, approximating focusing the initial response through lateral inhibition. The output representation is obtained from the input weights of the maximally responding unit by propagating the activity (which is equal to 1) through its weight vector to the output of the lexicon.

For example, in Fig. 2, the orthographic representation of DOG is input to the orthographic input map, which forms a concentrated activity pattern around the unit labeled DOG. The activity propagates through the associative connections of all active units (Fig. 6) to the semantic map, where a localized activity pattern forms around the unit labeled dog. The semantic representation for dog is then obtained by propagating activation through the weight vector of this unit. In a similar fashion, a phonological input can be translated to the corresponding concept, or a concept to its orthographic or phonological counterpart.

The behavior of the system is very robust. Even if the input pattern is noisy or incomplete, it is usually mapped on the correct unit. Even if this does not happen, the associative connections of the intermediate units provide a mapping that is close enough, so that the correct meaning or symbol can be retrieved with top-down priming.

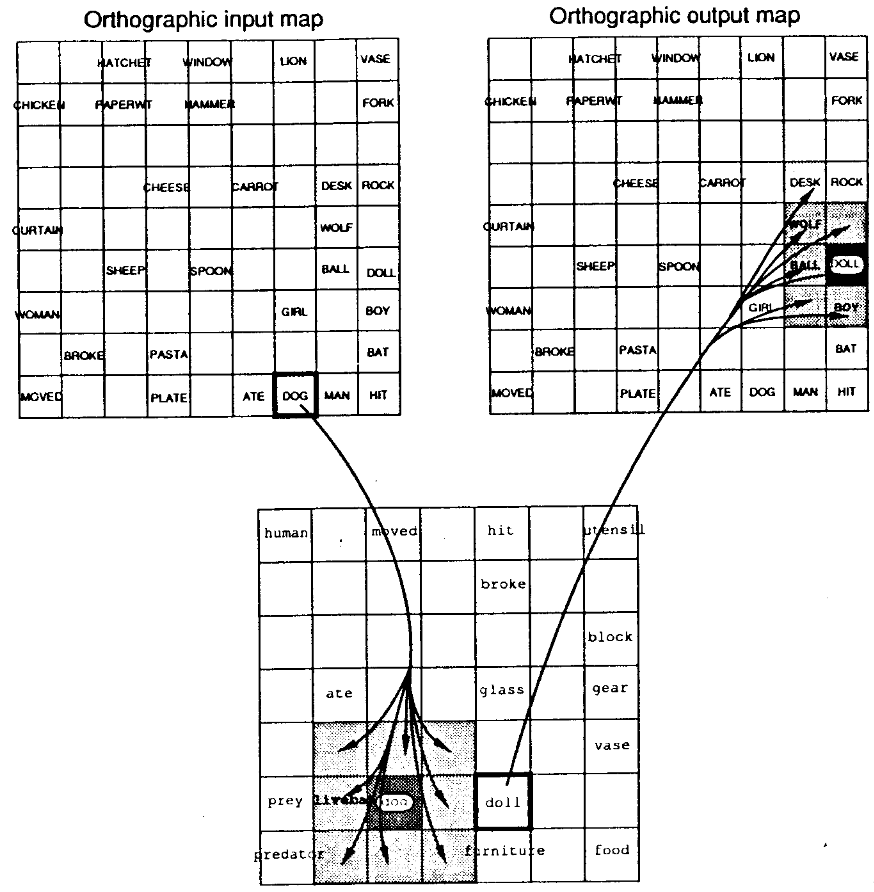


FIG. 6. Sample unambiguous associative mappings. Shown here are the active connections $w_{DOG,uv} > 0$ from the orthographic input unit DOG to the semantic map and the active connections $w_{doll,ij} > 0$ from the semantic unit doll to the orthographic output map. The darkness of the box indicates the strength of the connection to the unit. The strongest connections concentrate around the image units but tend to activate nearby representations as well. For example, in noisy conditions the input might be understood as livebat instead of dog, or the symbol BALL might be output instead of DOLL, resulting in dyslexic behavior.

PRIMING AND DISAMBIGUATION

When an ambiguous symbol is input to the lexicon, all possible meanings are activated at the same time (Fig. 7). Such behavior is consistent with experimental results on lexical access. For example, Swinney (1979) showed that in sentence processing, all meanings of ambiguous words are initially activated upon reading the word, although after reading three more syllables, only the correct meaning for the current context remains active.

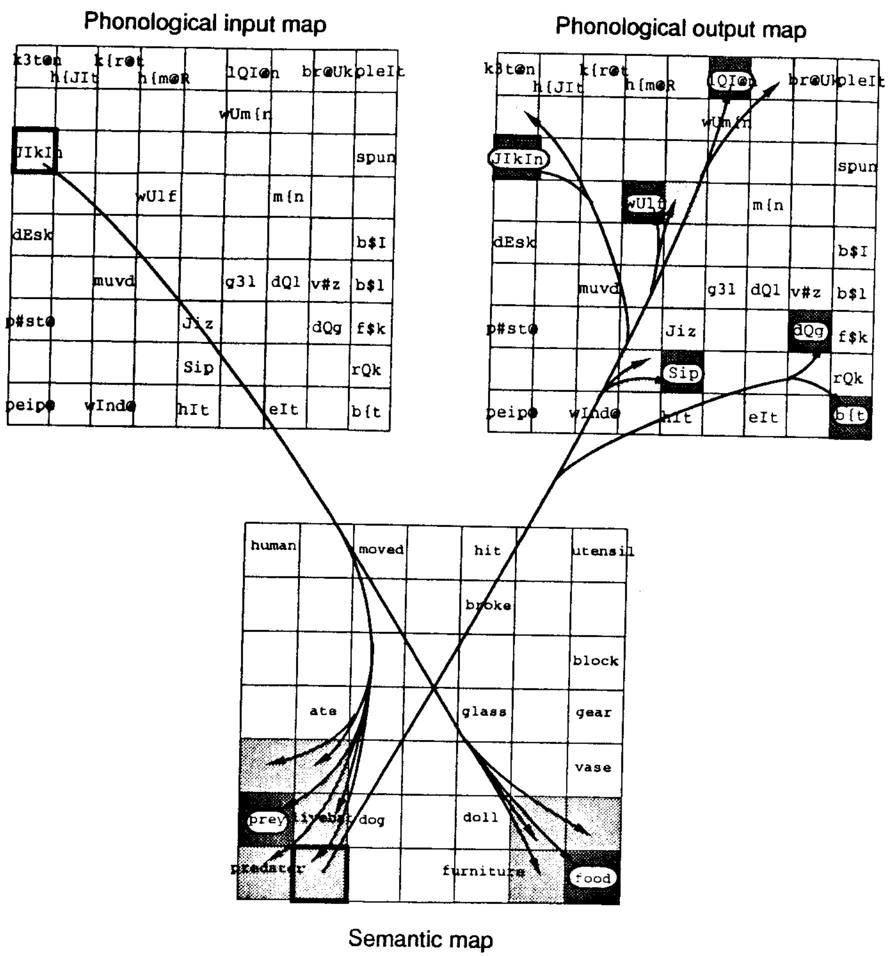


FIG. 7. Sample ambiguous associative mappings. The semantic map shows the active connections from the phonological input unit /JIkIn/ (chicken), which has two possible interpretations, food and prey. A priming process is required to select between them. At right, the active connections from the intermediate unit next to prey, livebat, dog, and predator to the phonological output map are shown. Possible output symbols include all animal names /JIkIn/, /1QI@n/, /Sip/, /wUlf/, /dQg/, and /b{t/.

To select the correct representation in DISLEX, a top-down priming mechanism combined with competition among the map units can be employed. In addition to the associative activation, the semantic map receives priming activation through its input connections (Fig. 8). Each unit (i,j) combines the two activations in its response:

$$\eta_{ij} = \sigma([1 - p]y_{A,ij} + py_{I,ij}), \tag{6}$$

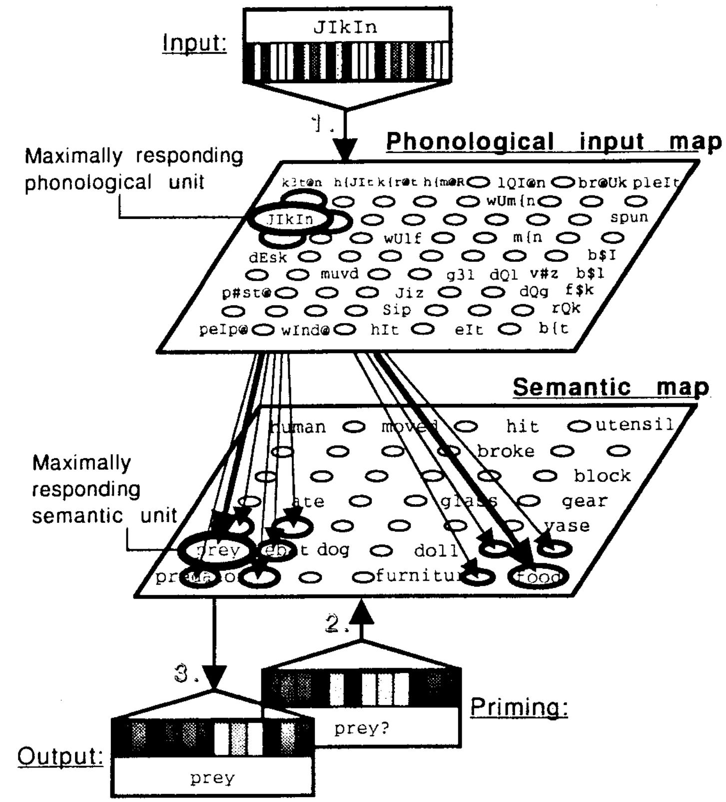


FIG. 8. Priming and disambiguation. 1. An ambiguous symbol /JIkIn/ (chicken) is input to the phonological input map. The activity propagates through the associative connections of /JIkIn/ to the semantic map, turning on prey and food, the two possible meanings associated with /JIkIn/. 2. At the same time, priming activation is input to the semantic map through its input connections. 3. The unit representing prey receives the largest total activation, turns off the other units, and sends the pattern for prey to the output of the map.

where $y_{A,ij}$ indicates activation due to associative input and $y_{I,ij}$ due to the priming input, the parameter $p: 0 \le p \le 1$ determines the strength of the priming, and σ is the standard sigmoid activation function. Due to the priming activation, the unit representing the correct meaning now responds more strongly than the other units. The representation stored in its weights is propagated to the output of the map.

Such priming input could originate from the high-level parsing processes. For example, the expectations generated by the FGREP parsing network (Section 3.2) could serve as a possible source. After inputting The predator ate the, the FGREP network generates a strong expectation for prey. When the phonological symbol /JIkIn/ is input, it is mapped on the unit labeled /JIkIn/, whose associative connections activate `prey` and `food` equally in the semantic map (Fig. 8). The expectation pattern, which is close to the representation for `prey`, is input to the semantic map and the resulting activity is combined with the activity propagated through the associative connections. As a result, the `prey` unit becomes most highly activated and is selected as the output of the lexicon.

The weights on the associative connections learn to represent statistical likelihoods of the associations. A very frequently active connection becomes stronger than a rare connection. For example, if most of the occurrences of /JIkIn/ in the training data had been paired up with `food`, the /JIkIn/ unit would tend to activate the `food` unit more than the `prey` unit. By default, the `food` meaning would be selected, and stronger priming for `prey` would be required to override it.

The current implementation of DISLEX simply selects and outputs the representation stored at the maximally responding unit. The selection could also be implemented with lateral inhibition. The units on the map would be connected laterally with inhibitory weights, and the initial activation would propagate through these connections, implementing cooperation and competition between units. In this process, the activation would gradually settle into a localized response (Sirosh & Miikkulainen, 1994). The settling times should correspond to the reaction times observed in humans (such as those described by, e.g., Simpson & Burgess, 1985). High-frequency words should have shorter reaction times, and these times could be changed with priming. With several equally likely interpretations, settling would generally take longer. The ambiguity effect, where a word with multiple meanings is recognized faster than an unambiguous word (Balota et al., 1991; Jastrzembski, 1981; Jastrzembski & Stanners, 1975), could result from proximity of initial activation as proposed by Joordens and Besner (1994). Such a dynamic implementation of the lexicon feature maps is an important direction of future research.

DYSLEXIC ERRORS AND SEMANTIC SLIPS

In dyslexia, words are often confused with semantically or visually/aurally similar ones. The lexicon architecture is well suited for modeling such behavior. If the system performance is degraded, for example, by adding noise to the connections, two basic types of input and production errors occur.

In production, noise in the input connections to the semantic map may cause a semantic representation to be classified incorrectly. As a result, a word with a similar but incorrect meaning would be produced, corresponding to a semantic error in deep dyslexia. For example, the representation for `dog` may be accidentally mapped on the intermediate unit among `dog`, `livebat`, `predator`, and `prey`. Instead of /dQg/, one of the animals /b{t/, /Sip/, /JIkIn/, /wUlf/, or /lQI@n/ could be produced (Fig. 7). Or, due to noise in the associative output connections, the activity in the semantic map may propagate incorrectly to the symbol map. In this case, a word with a similar orthographic or phonological form but a different meaning would be output, modeling surface dyslexic behavior. For example, BALL is likely to be generated instead of DOLL in noisy propagation to the orthographic map (Fig. 9).

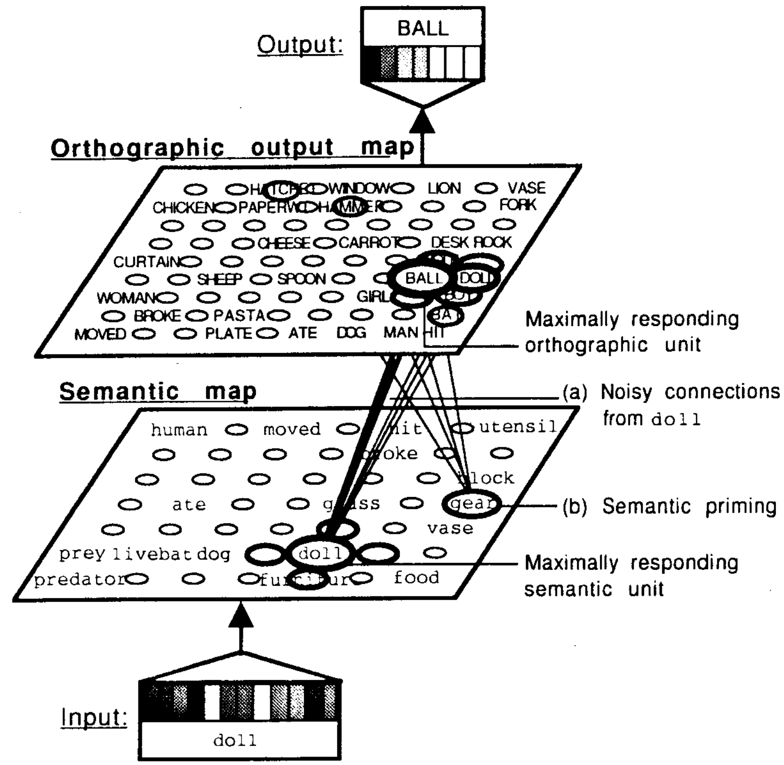


Fig. 9. An example of dyslexic behavior. If the associative propagation is noisy, BALL may be output instead of DOLL (a). Priming or residual activation on `gear` has a similar effect (b).

Visual, phonological, and semantic errors may also occur during input. If an orthographic representation is mapped incorrectly on a nearby unit on the orthographic map, a visual error results, corresponding to seeing the word incorrectly. For example, DOLL may be input as BALL. The activity in the symbol map may also propagate incorrectly to a nearby unit in the semantic map, in which case, for example, /JIkIn/ could be understood semantically as `livebat` (Fig. 7).

Combinations of the four basic error types are also possible, resulting in visual-then-semantic or semantic-then-visual errors (such as sympathy ↔

orchestra) and their phonological counterparts. For example, although SPOON is not visually or semantically similar to BAT, such a confusion could take place if SPOON was first visually mistaken for SHEEP, and the activity then propagated incorrectly to `livebat`. DISLEX can also explain why dyslexic errors are often both visually/phonologically and semantically related to the correct word. For example, if `dog` is presented to the semantic map under noisy conditions, a localized response around the intermediate unit among `dog`, `livebat`, `predator`, and `prey` might develop. Through the associative connections of these units, all the symbols representing animals would be activated in the phonological map. Each symbol receives activation from at least two semantic units: their own semantic unit (e.g., /dQg/ from `dog`, /lQI@n/ from `predator`), and the winning intermediate unit (as shown in figure 7). However, /dQg/ and /b{t/ are so close in the phonological map that they both receive activation from both `dog` and `livebat`. Therefore, if /dQg/ does not receive the highest activation, then /b{t/ most likely will. In other words, /b{t/ would be output because it is both semantically and phonologically similar to /dQg/.

Similar visual, phonological, and semantic errors, as well as combined visual-and-semantic and visual-then-semantic errors and their phonological counterparts, have been well documented in patients with various forms of dyslexia (Caramazza,1988; Coltheart et al., 1988a). They also occur in noisy, stressful, and overload situations in normal human performance. Such behavior can be explained by the above mechanisms, lending support to the multiple feature map lexicon architecture.

With priming, it would also be possible to model another interesting type of performance error: the semantic (Freudian) slip (see e.g., Aitchison 1987; Freud 1926/1958). Such errors occur when very strong semantic priming interferes with the output function. For example, if `doll` was input to the semantic map where `gear` is also active due to the simultaneous or residual priming, the activity would propagate through the associative connections of both (Fig. 9). As a result, the symbol BALL would receive the strongest activation, and would be output instead of DOLL. The two output symbols are similar, but the meaning of BALL reveals the hidden semantic priming.

CATEGORY-SPECIFIC APHASIC IMPAIRMENTS

The DISLEX architecture is consistent with recent cognitive neuropsychology theories of the human lexical system, such as those of Caramazza (1988), Warrington (1975), and Warrington and McCarthy (1987). Many observed lexical deficits in acquired aphasia have straightforward explanations in the model.

Aphasia is a language-processing disorder that typically results from a well-localized damage to the central nervous system, such as cerebral infarction, brain tumor, or contusion (Damasio, 1981). A common feature of aphasic impairments is category specificity. A patient may have selective difficulty or selective preservation of words that belong to a specific syntactic or semantic category. In certain patients the lexical access to function words is selectively impaired, in other cases the patient has trouble with verbs (Caramazza, 1988; Coltheart et al., 1988a). More specific impairments often occur in semantic hierarchies. Some patients have trouble with concrete words, inanimate objects, or indoor objects (Warrington & McCarthy, 1983; Warrington & Shallice, 1984; Yamadori & Albert, 1973), or even with classes as specific as names of fruits and vegetables (Hart et al., 1985). In some cases, categories such as letters, body parts, and colors are selectively preserved (Goodglass et al., 1986).

Deficits of this kind can be explained by the topological organization of the semantic memory. The semantic map in DISLEX is hierarchically organized and reflects both the syntactic and the semantic properties of the words. Localized lesions to the map that damage units or their connections would produce selective impairments like the above.

In some cases the impairments cover all modalities, sometimes they are limited only to verbal input or output, or even only to the orthographic or phonological domain. This suggests that the semantic memory, visual input, and verbal input and output modalities are represented in separate structures. For example, some patients were unable to access specific meanings from verbally as well as visually (with pictures) presented cues (Warrington, 1975; Warrington & Shallice, 1984). This implies that the semantic memory itself had been damaged. Another patient could not give definitions for aurally presented names of living things such as "dolphin," although he was able to describe other objects. But when shown a picture of a dolphin, he could name it and give an accurate verbal description of it (McCarthy & Warrington, 1988). This suggests that the visual pathway to the semantic memory, the semantic memory itself, and the verbal output were preserved, but the verbal access to the semantic memory had been damaged. In another case, the patient was unable to name fruits and vegetables, although he was able to match their names with pictures, and classify them correctly when their names were presented aurally (Hart et al., 1985). In other words, his semantic memory and verbal input were preserved, but the verbal output function was selectively impaired.

The impairment of semantic categories restricted to a single input or output modality can be modeled in DISLEX by damaged pathways between symbol and concept maps. A necessary assumption is that the pathways in DISLEX are not single axons, but consist of interneurons that also exhibit map-like organization. Close to the semantic map the organization is semantic, close to a symbol map it parallels the symbol map. If the pathway is severed close to the semantic map, a semantic impairment within this modality results.

The dissociation of the orthographic and phonological modalities is also well documented in aphasic data. Some patients have deficits only in one of the input or output channels, or different deficits in different channels (Basso et al., 1978; Caramazza, 1988). For example, a patient may have spelling difficulties exclusively in the orthographic output domain (Goodman & Caramazza, 1986; Miceli et al., 1985). The types of errors in orthographic and phonological dyslexia (Section 7) further suggest that the channels are organized according to the physical forms of the words. The DISLEX model predicts that it would also be possible to lose access to specific types of symbols, such as long or short words, as a result of localized damage to a lexical map.

In the aphasic impairments, high-frequency words are often better preserved than rare words (Caramazza, 1988; Newcombe et al., 1965). This is also predicted by the feature map organization. The most common words occupy larger areas in the map, making them more robust against damage.

DISCUSSION

An important characteristic of the DISLEX model is that its performance directly depends on the physical organization of the hardware. Noise can be added to it and it can be locally lesioned, and it displays deficits similar to those of human dyslexics and aphasics. This suggests that the model captures some of the physical structures underlying the lexical system in the brain. Its verification could therefore serve as a starting point for various neuropsychological experiments. The central assumption, and the most important to verify, is that the symbols and meanings are laid out on maps where different units are selectively sensitive to different words in the data. Indeed, recently it was found that neurons in the hippocampus respond selectively to visually presented words (Heit et al., 1989). It would be important to find out whether these selectivities form a map-like organization. Next, if such maps could be found for the different modalities, it would perhaps be possible to verify that they are connected with ordered pathways.

DISLEX still finesses much of the fine neural structure, and the mapping to the neuron level is nontrivial. The units and connections in the model do not necessarily correspond one-to-one to neurons and synapses, but rather to connected groups of neurons. For example, the weight vectors in the semantic map are used both for input and output, which is not a plausible model of synaptic efficacies. However, these two-way connections could be implemented with tightly interconnected (or phase-locking) groups of neurons in the brain. Whether such groups can serve as the basic units of information processing would need to be confirmed experimentally.

A number of other connectionist models of lexical access and lexical disambiguation have been proposed recently (Bookman, 1989; Cottrell & Small, 1983; Gallant, 1991; Gasser, 1988; Gigley, 1988; Kawamoto, 1988; Sharkey, 1989; Small, 1990; Waltz & Pollack, 1985). These models aim at explaining lexical processing with low-level mechanisms, focusing on the

timing of the process as well as on certain types of performance errors and deficits. They are primarily process models, detached from the physical structures, and designed as controlled demonstrations of how disambiguation could be carried out in the lexical system. One model that shares many of the goals of DISLEX is that of Hinton and Shallice (1991), further developed by Plaut (1991) and Plaut and Shallice (1993). In this model, an orthographic word representation is mapped to a semantic feature representation of the word meaning, and on to a phonological representation. An essential part of the model is that the semantic representation layer is recurrent (trained through backpropagation through time, Rumelhart et al. 1986, or as a deterministic Boltzmann machine, Hinton 1989). A noisy orthographic input representation causes initial activity in the semantic representation layer, which then settles into one of the attractor states representing a meaning. The network can be lesioned by deleting units and connections and by adding noise to the connections. As a result, the attractor basins are distorted and words are sometimes mapped to incorrect semantics in a manner that represents the types of errors observed in human deep dyslexia.

Although DISLEX and the attractor model are based on very different principles, they account for much of the same data. Hinton, Plaut, and Shallice have addressed a wider range of dyslexic phenomena in their work, including effects of word abstractness. On the other hand, DISLEX can account for many category-specific impairments in acquired aphasia. Further computational experiments are necessary to compare the merits and disadvantages of the two approaches. Experimental results supporting neural maps vs distributed and dynamic symbol and meaning representations would also help in verifying the assumptions of the two models.

An important computational validation of DISLEX was performed as part of the DISCERN system (Miikkulainen, 1993). In DISCERN, subsymbolic neural network models of parsing, generation, episodic memory, and the lexicon are brought together into a large artificial intelligence system that learns to read, paraphrase, and answer questions about script-based stories. In DISCERN the components, including the DISLEX lexicon, are not just models of isolated cognitive phenomena; they are shown to be sufficient computational constituents for generating complex high-level language processing behavior.

The orthographic and phonological pathways to semantic representations are very clearly separated in DISLEX. There is some evidence, however, that orthographic access to semantics is at least partially affected by phonology (Coltheart et al., 1988b). For example, Van Orden et al. (1988) found that nonwords such as `sute` and words with ambiguous phonological representations such as `hare` were often categorized according to their phonological representation as `a piece of clothing` or `a part of the human body`, although the orthographic representation alone would not activate those meanings. These results could be explained by an automatic process that associates pronunciation to the text (or "reads aloud") in the background. Such a process could be modeled through associative connections from the orthographic to the phonological map. Activation of an orthographic representation would activate its phonological counterpart, which in turn would send activation to the semantic map. It would be interesting to find out whether such associations exist between other modalities as well; they could easily be incorporated into the DISLEX architecture.

DISLEX is currently a model of single-word processing. It does not have special mechanisms for representing and processing phrasal structures or morphology. The model can deal with structured expressions in two ways: (1) Most common morphological forms and idioms, such as `nationalism` or `The Big Apple` can be represented like words, as single entries in the lexicon. Different morphological forms of the same word are mapped nearby on the semantic map and slips between forms are possible. (2) More complex phrases and unusual, constructive forms such as `kick the bucket` or `nonpreemptive` can be represented by their constituents in the lexicon, and the responsibility for parsing/generating them lies within the sentence-processing modules. These mechanisms together give a rough but fairly plausible account of human performance (as described, e.g., by Aitchison 1987). However, it seems likely that people initially process the constituents of a new form or phrase separately, but after extensive practice the expression becomes a single unanalyzable entry in the lexicon (Stemberger, 1985). How this learning process could be modeled in DISLEX is an open question. Also, dyslexic and aphasic data suggest that morphology is an independent system that can be selectively impaired or preserved (Caramazza, 1988; Coltheart et al., 1988a). In some cases, inflectional affixes are processed incorrectly while the stems are preserved (Gleason, 1978), in others, the patient has trouble producing appropriate word stems but demonstrates correct inflection of the resulting nonwords (Caplan et al., 1972). Such dissociations cannot be easily explained by the current architecture.

In category-specific impairments, the more general terms are often better preserved (Caramazza et al., 1990; Warrington, 1975). For example, a normal subject would respond faster to "Is a duck a bird?" than to "Is a duck an animal?," but the aphasic patient would find the latter question easier. In some cases, the superordinate categories are accessible when the subordinate categories are not. For example, the patient may be able to classify a canary as a bird, an animal, and a living thing, but could not confirm that it is yellow, small, and a pet (Warrington, 1975). Such data suggest that the semantic memory is hierarchically organized, and specific information is more vulnerable than general information. Unfortunately in the lexicon model, the general terms would be located at the center of more specific terms on the semantic map, and would be equally easy to access and equally likely to be impaired in local damage. Lack of physical hierarchy makes it also difficult to account for certain psychological data on normal processing of category hierarchies. Rosch et al. (1976) and Rosch (1978) demonstrated that names for basic-level categories (such as table or dog) are easier to process than for superordinate and subordinate categories (e.g., furniture, spaniel). For the current model, the level of the category would not make any difference.

However, if the semantic memory was implemented as a hierarchical feature map system (Miikkulainen 1990), the general terms would be represented higher in the hierarchy and could be better preserved. Access to the basic level could be easier than either to the top or bottom of the hierarchy. Such a hierarchical feature map lexicon opens many interesting possibilities and constitutes a most promising direction for future research.

CONCLUSION

The DISLEX model was built to test computationally whether the lexical system could consist of separate topologically organized feature maps for the different modalities and the lexical semantics. The performance characteristics and especially the dyslexic and aphasic behavior exhibited by the model suggest that DISLEX is probably on the right track. The most important direction of future work is to verify some of the assumptions and predictions of the model experimentally and against clinical data. Such interaction between experimental and modeling approaches should lead to better constraints on future models and eventually to a better understanding of the lexical system.

APPENDIX A

Orthographic Representations

The orthographic symbol representations for each word were formed by concatenating the values representing the darkness of each letter into a single vector. The darkness values were obtained by counting the number of black pixels for each letter in the 12pt MacIntosh Geneva font and scaling the number between 0 and 1. The resulting darkness values are listed in Table 1.

Although this encoding scheme is simple, it results in unique representations for all symbols in the training data, and similar symbols have similar representations. With a larger vocabulary, more accurate representations might be needed to make sure they are unique. The actual bitmaps of letters could be used, or bitmaps that have been slightly blurred. Blurring introduces overlap, causing letters that are perceived similar to have more similar representations.

APPENDIX B

Phonological Representations

The phonological word symbols were represented as sequences of phonemes, obtained from the CELEX database at Max Planck Institute for Psy-

TABLE 1
Orthographic Representations

Letter	Value	Letter	Value	Letter	Value	Letter	Value
A	0.481481	H	0.666667	O	0.629630	V	0.370370
B	0.814814	I	0.148148	P	0.592593	W	0.814815
C	0.444444	J	0.296296	Q	0.666667	X	0.444444
D	0.703704	K	0.481481	R	0.703704	Y	0.259259
E	0.703704	L	0.296296	S	0.518519	Z	0.518519
F	0.444444	M	1.000000	T	0.333333		
G	0.740741	N	0.666667	U	0.518519		

Note. The number of black pixels for each letter in the MacIntosh Geneva font was scaled between 0 and 1. Word representations were formed by concatenating the letter values into a single vector.

cholinguistics. Following the International Phonetic Alphabet, each phoneme was classified according to place and manner of articulation, sound, chromaticity, and sonority, and the categorization was translated into a numerical vector. The phoneme representation vectors were then concatenated into the phonological word symbol vectors. The phoneme classifications are listed in Table 2 and the numeric encoding of their feature values in Table 3.

APPENDIX C

Semantic Representations

The semantic representations were obtained with the FGREP method in the task of assigning case roles to the syntactic constituents of the sentence. The sentence templates are listed in Table 4 and the semantic categories in Table 5. The input/output examples were generated from the templates by filling each slot with a concept from a specified category. This data set was obtained from McClelland and Kawamoto (1986) by replacing words that were used in identical ways in their data by a single word. This way, {man, woman, boy, girl} was replaced by human, {cheese, pasta, carrot} by food, {wolf, lion} by predator, {ball, hatchet, hammer} by gear, {paperwt, rock} by block, {plate, window} by glass, {fork, spoon} by utensil, and {desk, curtain} by furniture. In addition, the occurrences of the ambiguous word chicken were replaced by the appropriate unambiguous concepts food and prey, and similarly bat was replaced by livebat and gear. In the resulting data, every concept has a unique and unambiguous usage.

APPENDIX D

DISLEX Code

The code and data for the DISLEX system are available by anonymous ftp from cs.utexas.edu:pub/neural-nets/dilex, or on the World Wide Web, under http://www.cs.utexas.edu/users/nn.

TABLE 2
Phoneme Representations

Label	Example	Place	Manner	Sound	Chromacity	Sonority
I	pIt	none	vowel	voiced	front-center	hi-mid
E	pEt	none	vowel	voiced	front	mid-lo
{	pAt	none	vowel	voiced	front	lo-mid
e	bAy	none	vowel	voiced	front	mid-hi
a	bOut	none	vowel	voiced	front	lo
Q	pOt	none	vowel	voiced	center	lo-mid
V	pUtt	none	vowel	voiced	center-back	mid-lo
U	pUt	none	vowel	voiced	center-back	hi-mid
@	thE	none	vowel	voiced	center	mid
i	hEEd	none	vowel	voiced	front	hi
u	whO'd	none	vowel	voiced	back	hi
3	bURn	none	vowel	voiced	front-center	mid
$	bORn	none	vowel	voiced	back	mid-lo
#	bARn	none	vowel	voiced	center-back	lo
p	Pet	bilabial	stop	unvoiced	none	none
b	Boat	bilabial	stop	voiced	none	none
t	Tot	alveolar	stop	unvoiced	none	none
d	Debt	alveolar	stop	voiced	none	none
k	Ketch	velar	stop	unvoiced	none	none
g	Get	velar	stop	voiced	none	none
N	saNG	velar	nasal	voiced	none	none
m	Met	bilabial	nasal	voiced	none	none
n	Net	alveolar	nasal	voiced	none	none
l	Let	alveolar	lateral	voiced	none	none
r	Row	alveolar	approximant	voiced	none	none
f	For	labio-dental	fricative	frication	none	none
v	Vow	labio-dental	fricative	voiced	none	none
T	THin	dental	fricative	frication	none	none
D	THen	dental	fricative	voiced	none	none
s	Say	alveolar	fricative	frication	none	none
z	laZy	alveolar	fricative	voiced	none	none
S	SHop	palatal-alveolar	fricative	frication	none	none
Z	aZure	palatal-alveolar	fricative	voiced	none	none
j	Yes	palatal	approximant	voiced	none	none
x	loCH	velar	fricative	frication	none	none
h	How	glottal	fricative	aspiration	none	none
w	Why	velar	approximant	voiced	none	none
J	CHeap	palatal-alveolar	stop	frication	none	none
—	juDGe	palatal-alveolar	stop	voiced	none	none

Note. The phoneme label and an example of each phoneme is given, followed by the values of the five features that describe the phoneme. The actual representation vectors were formed by replacing the feature values with their numeric encoding, shown in Table 3.

TABLE 3
Phoneme Features

Place	Manner	Sound	Chromacity	Sonority
0.000 none	0.000 none	0.000 none	0.000 none	0.000 none
0.125 bilabial	0.167 stop	0.250 voiced	0.200 front	0.143 hi
0.250 labio-dental	0.333 fricative	0.500 frication	0.400 front-center	0.286 hi-mid
0.375 dental	0.500 approximant	0.750 unvoiced	0.600 center	0.429 mid-hi
0.500 alveolar	0.667 lateral	1.000 aspiration	0.800 center-back	0.571 mid
0.625 palatal-alveolar	0.833 nasal		1.000 back	0.714 mid-lo
0.750 palatal	1.000 vowel			0.857 lo-mid
0.875 velar				1.000 lo
1.000 glottal				

Note. The values of each feature are represented as real numbers between 0 and 1.

TABLE 4
Sentence Templates

Sentence frame	Correct case roles
1. The human ate.	agent
2. The human ate the food.	agent, patient
3. The human ate the food with the food.	agent, patient, modifier
4. The human ate the food with the utensil.	agent, patient, instrument
5. The animal ate.	agent
6. The predator ate the prey.	agent, patient
7. The human broke the fragileobj.	agent, patient
8. The human broke the fragileobj with the breaker.	agent, patient, instrument
9. The breaker broke the fragileobj.	instrument, patient
10. The animal broke the fragileobj	agent, patient
11. The fragileobj broke.	patient
12. The human hit the thing.	agent, patient
13. The human hit the human with the possession.	agent, patient, modifier
14. The human hit the thing with the hitter.	agent, patient, instrument
15. The hitter hit the thing.	instrument, patient
16. The human moved.	agent, patient
17. The human moved the object.	agent, patient
18. The animal moved.	agent, patient
19. The object moved.	patient

Note. Each frame has one to three concept slots (shown in roman typeface). Each slot has a predetermined case role, shown at right. Each slot can be filled with any of the concepts in the specified category, listed in Table 5. For instance, "The animal broke the fragileobj" generates 4 × 2 different sentences, all with the case-role assignment agent = animal, patient = fragileobj.

TABLE 5
Semantic Categories

Category	Semantic concepts
human	human
food	food
utensil	utensil
animal	prey preditor livebat dog
fragileobj	glass vase
breaker	gear block
hitter	gear block vase
possession	gear vase doll dog
object	gear block vase glass food furniture doll utensil
thing	human prey preditor livebat dog gear block vase glass food furniture doll utensil
verb	hit ate broke moved

Note. Each slot in the sentence templates specifies a category and can be filled with any semantic concept in that category. In other words, the categorization determines how the concepts are used in the sentences.

REFERENCES

Aitchison, J. 1987. *Words in the mind: An introduction to the mental lexicon.* Oxford, UK/New York: Blackwell.

Balota, D. A., Ferraro, F. R., & Connor, L. T. 1991. On the early influence of meaning in word recognition: A review of the literature. In P. Schwanenflugel (Ed.), *The psychology of word meaning.* Pp. 187–222. Hillsdale, NJ: Erlbaum.

Basso, A., Taborelli, A., & Vignolo, L. A. 1978. Dissociated disorders of speaking and writing in aphasia. *Journal of Neurology, Neurosurgery and Psychiatry,* **41,** 526–556.

Bookman, L. A. 1989. A connectionist scheme for modelling context. In D. S. Touretzky, G. E. Hinton, & T. J. Sejnowski (Eds.), *Proceedings of the 1988 connectionist models summer school.* San Mateo, CA: Morgan Kaufmann. Pp. 281–290.

Caplan, D., Keller, L., & Locke, S. 1972. Inflection of neologisms in aphasia. *Brain,* **95,** 169–172.

Caramazza, A. 1988. Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system. *Annual Review of Neuroscience,* **11,** 395–421.

Caramazza, A., Hillis, A. E., Rapp, B. C., & Romani, C. 1990. The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology,* **7,** 161–189.

Coltheart, M., Patterson, K., & Marshall, J. C. (Eds.) 1988a. *Deep dyslexia.* London/New York: Routledge & Kegan Paul. Second edition.

Coltheart, V., Laxon, V., Rickard, M., & Elton, C. 1988b. Phonological recoding in reading for meaning by adults and children. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14,** 387–397.

Cottrell, G. W., & Small, S. L. 1983. A connectionist scheme for modelling word sense disambiguation. *Cognition and Brain Theory,* **6,** 89–120.

Damasio, A. R. 1981. The nature of aphasia: Signs and symptoms. In Sarno, M. T. (Ed.), *Acquired aphasia.* New York: Academic Press. Pp. 51–65.

Freud, S. 1926/1958. Slips of the tongue. In *The psychopathology of everyday life.* Translated by A. A. Brill. New York: New American Library. Chapter V.

Gallant, S. I. 1991. A practical approach for representing context and for performing word sense disambiguation using neural networks. *Neural Computation,* **3,** 293–309.

Gasser, M. 1988. *A connectionist model of sentence generation in a first and second language.* Ph.D. thesis, Computer Science Department, University of California, Los Angeles. Technical Report UCLA-AI-88-13.

Gigley, H. 1988. Process synchronization, lexical ambiguity resolution and aphasia. In Small, S. L., Cottrell, G. W., & Tanenhaus, M. K. (Eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology & artificial intelligence.* San Mateo, CA: Morgan Kaufmann. Pp. 229–267.

Gleason, J. B. 1978. The acquisition and dissolution of the English inflectional system. In A. Caramazza & E. B. Zurif (Eds.), *Language acquisition and language breakdown: Parallels and divergences.* Baltimore, MD: Johns Hopkins University Press. Pp. 109–120.

Goodglass, H., Wingfield, A., Hyde, M. R., & Theurkauf, J. C. 1986. Category specific dissociations in naming and recognition by aphasic patients. *Cortex,* **22,** 87–102.

Goodman, R. A., & Caramazza, A. 1986. Aspects of the spelling process: Evidence from a case of acquired dysgraphia. *Language and Cognitive Processes,* **1,** 263–296.

Gustafsson, B., & Wigström, H. 1988. Physiological mechanisms underlying long-term potentiation. *Trends in Neurosciences,* **11,** 156–162.

Hart, Jr., J., Berndt, R. S., & Caramazza, A. 1985. Category-specific naming deficit following cerebral infarction. *Nature,* **316,** 439–440.

Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. 1989. The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research,* **32,** 203–218.

Hebb, D. O. 1949. *The organization of behavior: A neuropsychological theory.* New York: Wiley.

Heit, G., Smith, M. E., & Halgren, E. 1989. Neural encoding of individual words and faces by the human hippocampus and amygdala. *Nature,* **333,** 773–775.

Hertz, J., Krogh, A., & Palmer, R. G. 1991. *Introduction to the theory of neural computation.* Reading, MA: Addison-Wesley.

Hinton, G. E. 1989. Deterministic Boltzmann learning performs steepest descent in weight-space. *Neural Computation,* **1,** 143–150.

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

Hubel, D. H., & Wiesel, T. N. 1959. Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology,* **148,** 574–591.

Hubel, D. H., & Wiesel, T. N. 1965. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology,* **28,** 229–289.

Jastrzembski, J. E. 1981. Multiple meanings, number of related meanings, frequency of occurence, and the lexicon. *Cognitive Phychology,* **13,** 278–305.

Jastrzembski, J. E., & Stanners, R. F. 1975. Multiple word meanings and lexical search speed. *Journal of Verbal Learning and Verbal Behavior,* **14,** 534–537.

Joordens, S., & Besner, D. 1994. When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 1051–1062.

Kawamoto, A. H. 1988. Distributed representations of ambiguous words and their resolution in a connectionist network. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution: perspectives from psycholinguistics, neuropsychology & artificial intelligence.* San Mateo, CA: Morgan Kaufmann. Pp. 195–288.

Knudsen, E. I., du Lac, S., & Esterly, S. D. 1987. Computational maps in the brain. In Cowan, W. M., Shooter, E. M., Stevens, C. F., & Thompson, R. F. (Eds.), *Annual review of neuroscience,* Palo Alto: Annual Reviews. Pp. 41–65.

Kohonen, T. 1982a. Analysis of a simple self-organizing process. *Biological Cybernetics,* **44,** 135–140.

Kohonen, T. 1982b. Self-organized formation of topologically correct feature maps. *Biological Cybernetics,* **43,** 59–69.

Kohonen, T. 1989. *Self-organization and associative memory.* Berlin-Heidelberg-New York: Springer. Third edition.

Kohonen, T. 1990. The self-organizing map. *Proceedings of the IEEE,* **78,** 1464–1480.

McCarthy, R. A., & Warrington, E. K. 1988. Evidence for modality-specific meaning systems in the brain. *Nature,* **334,** 428–430.

McCarthy, R. A., & Warrington, E. K. 1990. *Cognitive neuropsychology: A clinical introduction.* New York: Academic Press.

McClelland, J. L., & Kawamoto, A. H. 1986. Mechanisms of sentence processing: Assigning roles to constituents. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, volume 2: Psychological and biological models.* Cambridge, MA: MIT Press. Pp. 272–325.

Miceli, G., Silveri, M. C., & Caramazza, A. 1985. Cognitive analysis of a case of pure dysgraphia. *Brain and Language,* **25,** 187–212.

Miikkulainen, R. 1990. Script recognition with hierarchical feature maps. *Connection Science,* **2,** 83–101.

Miikkulainen, R. 1993. *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory.* Cambridge, MA: MIT Press.

Miikkulainen, R., & Dyer, M. G. 1991. Natural language processing with modular neural networks and distributed lexicon. *Cognitive Science,* **15,** 343–399.

Newcombe, F., Oldfield, R. C., & Wingfield, A. 1965. Object naming by dysphasic patients. *Nature,* **207,** 1217–1218.

Plaut, D. C. 1991. *Connectionist neuropsychology: The breakdown and recovery of behavior in lesioned attractor networks.* Ph.D. thesis, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA. Technical Report CMU-CS-91-185.

Plaut, D. C., & Shallice, T. 1993. Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology,* **10,** 377–500.

Ritter, H. J. 1991. Asymptotic level density for a class of vector quantization processes. *IEEE Transactions on Neural Networks,* **2,** 173–175.

Ritter, H. J., & Schulten, K. J. 1986. On the stationary state of Kohonen's self-organizing sensory mapping. *Biological Cybernetics,* **54,** 99–106.

Rolls, E. T. 1984. Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Human Neurobiology,* **3,** 209–222.

Rosch, E. 1978. Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Erlbaum. Pp. 27–48.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. 1976. Basic objects in natural categories. *Cognitive Psychology,* **8,** 382–439.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986. Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, volume 1: Foundations.* Cambridge, MA: MIT Press. Pp. 318–362.

Sharkey, N. E. 1989. The lexical distance model and word priming. In *Proceedings of the 11th annual conference of the cognitive science society.* Hillsdale, NJ: Erlbaum. Pp. 860–867.

Simpson, G. B., & Burgess, C. 1985. Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance.* **11,** 28–39.

Sirosh, J., & Miikkulainen, R. 1994. Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics,* **71,** 66–78.

Small, S. L. 1990. Learning lexical knowledge in context: Experiments with recurrent feed forward networks. In *Proceedings of the 12th annual conference of the cognitive science society.* Hillsdale, NJ: Erlbaum. Pp. 479–486.

Stemberger, J. P. 1985. *The lexicon in a model of language production.* New York: Garland.

Swinney, D. A. 1979. Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior,* **18,** 645–659.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. 1988. Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14,** 371–386.

Waltz, D. L., & Pollack, J. B. 1985. Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science,* **9,** 51–74.

Warrington, E. K. 1975. The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology,* **27,** 635–657.

Warrington, E. K., & McCarthy, R. A. 1983. Category specific access dysphasia. *Brain,* **106,** 859–878.

Warrington, E. K., & McCarthy, R. A. 1987. Categories of knowledge: Further fractionations and an attempted integration. *Brain,* **110,** 1273–1296.

Warrington, E. K., & Shallice, T. 1984. Category specific semantic impairments. *Brain,* **107,** 829–854.

Yamadori, A., & Albert, M. L. 1973. Word category aphasia. *Cortex,* **9,** 112–125.

BRAIN AND LANGUAGE **59,** 367–389 (1997)
ARTICLE NO. BL971821

The Connectionist Simulation of Aphasic Naming

John F. Wright and Khurshid Ahmad

Artificial Intelligence Group, Department of Mathematical and Computing Sciences, University of Surrey, Guildford, United Kingdom

The simulation of language disorders using interactive activation (IA) networks and connectionist systems is discussed. An existing IA account of aphasic naming is described, in which two network parameters (*decay rate* and *connection strength*) are varied to fit the error production of an aphasic patient. Fairly similar results can be obtained through modification of additional parameters, including the so-called "shared weight increase factor" linking lexical and semantic units. This leads us to consider simulation of aphasic naming using connectionist networks which do not require explicit variation of network parameters. A *modular* connectionist architecture is presented, in which semantic–lexical and phonological knowledge are instantiated using self-organizing Kohonen maps, while connections between them are implemented using Hebbian networks; a linear connectionist network (Madaline) is used to simulate nonword repetition. The Hebbian connections are lesioned in order to reproduce the patient's naming errors.

INTRODUCTION

Connectionist (or neural) networks are computer programs and associated data sets that can be used to simulate a wide range of "real-world" phenomena. In recent years, connectionism—the application of connectionist net-

Address correspondence and reprint requests to Dr. John Wright, Artificial Intelligence Group, Department of Mathematical and Computing Sciences, University of Surrey, Guildford GU2 5XH, UK. E-mail: J.Wright@surrey.ac.uk.

The authors thank Nadine Martin and Eleanor Saffran for their encouragement and their generosity in sharing their data with us. Our interactions with cognitive neuropsychologists such as Gary Dell, Max Coltheart, David Plaut, and Steven Small have been educational and have made us appreciate the depth of, and tensions within, the subject. We thank two anonymous referees for their comments on an earlier draft of this paper. Nadine Martin, one of the named referees, has been very patient with our computational speculation and has been, as usual, very helpful with her suggestions. John Wright is grateful to the Department of Mathematical and Computing Sciences, University of Surrey, for the research studentship that supported him in the period 1992–1995. Much of the work described here was originally presented in August 1994 at the Neurolinguistics and Cognitive Modelling Workshop, University of Turku, Finland.

0093-934X/97 $25.00