A DISTRIBUTED FEATURE MAP MODEL OF THE LEXICON *

Risto Miikkulainen
Artificial Intelligence Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
risto@cs.ucla.edu

Abstract

DISLEX models the human lexical system at the level of physical structures, i.e. maps and pathways. It consists of a semantic memory and a number of modality-specific symbol memories, implemented as feature maps. Distributed representations for the word symbols and their meanings are stored on the maps, and linked with associative connections. The memory organization and the associations are formed in an unsupervised process, based on co-occurrence of the physical symbol and its meaning. DISLEX models processing of ambiguous words, i.e. homonyms and synonyms, and dyslexic errors in input and in production. Lesioning the system produces lexical deficits similar to human aphasia. DISLEX-1 is an AI implementation of the model, which can be used as the lexicon module in distributed natural language processing systems.

1 Introduction

The lexicon in symbolic NLP systems is a list of word symbols and phrasal patterns, with pointers to conceptual memory. The memory contains syntactic and semantic knowledge about the lexicon entry in the form of declarations, or procedures which specify how the word should be interpreted in different environments [29; 1; 6]. This knowledge has been explicitly programmed into the system with specific examples in mind. The symbolic lexicons are intended to model the *processes* of lexical access, not the physical structures that implement the processes. Consequently, these models lack the capacity to account for lexical errors in human performance, as well as lexical deficits in acquired aphasia.

A number of connectionist models of lexical disambiguation have been proposed [5; 25; 7; 12; 8]. These models aim at explaining lexical processing with low-level mechanisms, and can better account for the timing of the process, as well as for certain types of performance errors and deficits. However, they are still primarily process models, detached from the physical structures. They are designed as controlled demonstrations, not as building blocks in larger NLP systems.

*This research was supported in part by an ITA Foundation grant and by fellowships from the Academy of Finland, the Emil Aaltonen Foundation, the Foundation for the Advancement of Technology and the Alfred Kordelin Foundation (Finland).

The main goal of the DISLEX project (DIStributed feature map LEXicon) is to develop a computational model of the human lexical system, which is plausible at the level of *physical structures* such as maps and pathways. The model is based on current cognitive neuroscience theories and accounts for several documented lexical deficits in acquired aphasia and dyslexia. A secondary goal is to build a practical implementation of the model for a distributed story understanding system [19].

In terms of the symbolic lexicon models, DISLEX contains both the symbol memory and the conceptual memory, and implements a mapping between them. However, DISLEX is based on distributed representations of the word symbols and the word semantics. The lexical system is seen more like a filter, which transforms an input word symbol into its semantic representation, and vice versa. The memory organization and the mapping are formed in an unsupervised self-organizing process, based on examples of co-occurrence of the word and its meaning. As a model of the lexical system, DISLEX is in good agreement with Caramazza's theory [3]. The architecture offers a simple explanation to several types of lexical errors and deficits.

2 Overview of DISLEX

DISLEX has separate symbol memories for each input and output modality (figure 1). These memories store distributed representations for the physical word symbols, which are used in communication with the external world. For example, an orthographic word representation for `DOG` consists of the visual form of the letters `D`, `O`, `G`, while the phonological representation stands for the string of phonemes `do:g`. The separation of modality-specific channels is intuitively compelling, since the modalities give rise to different representations, and are processed through different structures [3]. The symbol spaces are not identical across modalities, there are homophones and homographs. Considerable experimental evidence also supports dissociation of the lexical components [3] (section 8).

The semantic memory of DISLEX consists of distributed representations of meanings, called semantic words. The semantic word **dog** (or e.g. **dog32**) refers to a specific animal and contains information such as domestic, mammal, brown color etc. There is a pathway from the semantic memory to the higher level language processing systems, which use semantic representations. The semantic memory

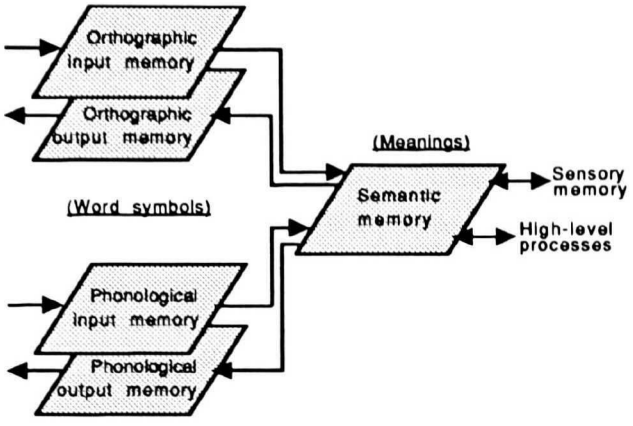

Figure 1: **The DISLEX architecture.** The physical symbol memories are modality and direction specific. The arrows indicate pathways of distributed representations.

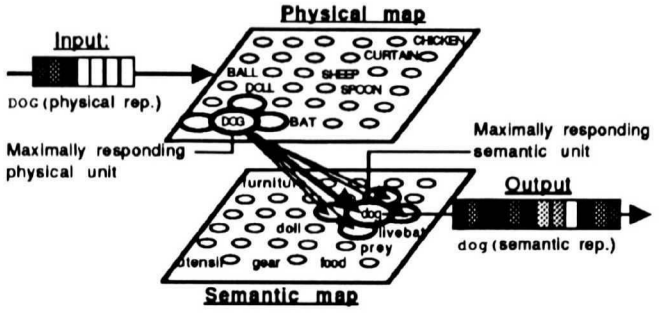

Figure 2: **Physical and semantic feature maps.** The physical input word DOG is transformed into the semantic representation of **dog**. The representations are vectors of real values between 0 and 1, shown by gray-scale coding. The size of the unit indicates the strength of its response. Only a few strongest associative connections are shown.

is also connected to the sensory memory, which contains visual images of objects and other sensory information. This pathway allows nonlinguistic access to the semantic memory, and provides the means for symbol grounding. The semantic word representation contains sensory information about the word referent, and the abstract word meaning originating from the high-level processes (ID and content, see [22]).

The physical and semantic memories are implemented as feature maps (figure 2). There is one map for each input and output modality and one for the semantic memory. The maps lay out each high-dimensional representation space on a 2-D area so that the similarities between words become visible. Physical words with similar form, e.g. **BALL, DOLL** are represented by nearby units in a physical map. In the semantic map, semantic words with similar content, e.g. **livebat, prey** are mapped near each other.

The physical maps are densely connected to the semantic map with associative connections. A localized activity pattern representing a symbol in the physical input map will cause a localized activity pattern to form in the semantic map, representing the meaning of the symbol (figure 2). Similarly, an active meaning activates a symbol in the physical output map. The lexicon thus transforms a physical input representation into a semantic output representation, and vice versa, and serves as an input/output filter for language processing. The physical and semantic maps are organized and the associative connections between them are formed simultaneously in an unsupervised learning process.

3 The DISLEX-1 simulation

DISLEX-1 is an AI implementation of DISLEX, designed as the lexicon module for a distributed neural network story understanding system [19]. DISLEX-1 contains a single physical modality, and the same representation space is used for both input and output. Figure 2 displays the basic architecture of DISLEX-1. Associative connections exist in both directions (the connections from semantic to physical map are omitted from the figure), and the transformation depicted in the figure can be reversed. This is a practical design for an AI module, and illustrates the basic principles and properties of the model.

DISLEX-1 was trained with data from a sentence processing experiment [17; 21] (figure 3). In the remainder of the paper, the mechanisms and properties of DISLEX are discussed, using the DISLEX-1 simulation as an example.

4 Representations

4.1 Physical representations

A central assumption in DISLEX is that the representations in each physical modality reflect the similarities within that modality. For example, the orthographic representations for DOG and DOC are very similar, but less so in the phonological domain.

The DISLEX-1 architecture concentrates on the orthographic modality. A simple encoding scheme was used to build the distributed representations for the written words. Each character was given a value between 0 and 1 according to its darkness, i.e. how many pixels are black in its bitmap representation. The darkness values of the word's characters were then concatenated into one representation vector (figure 3). This simple representation adequately reflects the visual similarities of the orthographic word symbols.

4.2 Semantic representations

The semantic representation is a distributed representation of the meaning of the word. Semantic representations are used internally for processing in cognitive models, and they should facilitate inferencing, expectations, generalizations etc. [15; 22]. A possible solution is to compose the representation from an ID part, representing the sensory referent of the word, and a content part, which encodes the processing properties of the word in relation to other words [22]

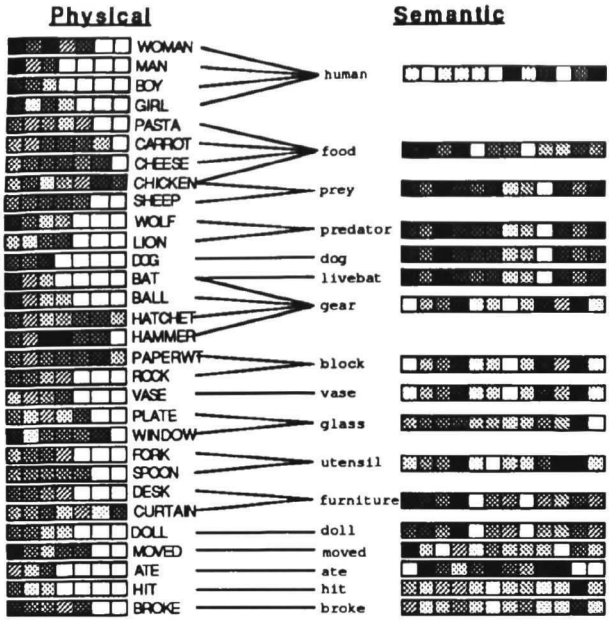

Figure 3: **The training data for DISLEX-1.** The physical representations code the orthographic word symbols, while the semantic representations stand for distinct meanings. Gray-scale boxes indicate component values within 0 and 1. The connections depict the mapping between the symbols and their meanings.

Category	Semantic words
animal	**prey predator 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 animal object**
verb	**hit ate broke moved**

Table 1: **Semantic categories.** Each slot in the sentence templates specifies a category, and can be filled with any semantic word in that category. In other words, the categorization determines how the words are used in the sentences.

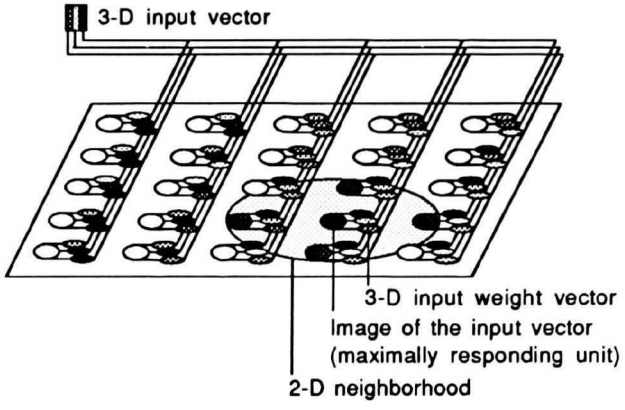

Figure 4: **A self-organizing feature map network.** A mapping is formed from a 3-dimensional input space onto a 2-dimensional network. The values of the input components, weights and the unit output are shown by gray-scale coding.

With the FGREP-mechanism [21] it is possible to extract the processing content of the word from examples of its use, and code it into a distributed representation. An FGREP-module is a three-layer backpropagation network which automatically developes distributed representations for its input items as it is learning a processing task.

For simplicity, and without restricting the generality of the model, the sensory part was omitted from the training data for DISLEX-1. The semantic representations for DISLEX-1 were formed with FGREP in the sentence case-role assignment task. The input to the FGREP network consisted of the syntactic constituents of the sentence and the network was trained to assign the correct semantic case roles to them. The sentences were generated from templates, by filling each slot in the template with a word from a specified category (table 1). The actual sentences and the specifics of the task are not important for this discussion (see [21]). However, the meanings embedded in the semantic representations originate from the categorization in table 1.

The representations that result from the FGREP process reflect the use of the semantic words (figure 3). Words belonging to the same category have a number of uses in common, and their representations become similar. The total usage is different for each word, and consequently, they stand for unique meanings.

5 Word maps

5.1 Topological feature maps

A 2-D topological feature map [13] 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 (figure 4). The map takes an N-dimensional vector as its input, and produces a localized pattern of activity as its output. In other words, an 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.

The ordering of the weight vectors retains the topology of the input space. This means roughly that nearby vectors in the input space are mapped onto nearby units in the map. This is a very useful property, since the complex similarity relationships of the high-dimensional input space become visible on the map.

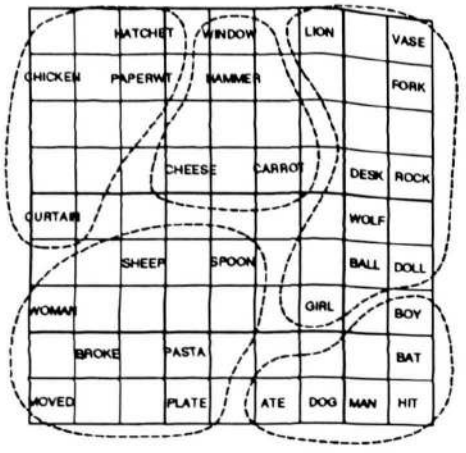

Figure 5: **The physical map.** Each unit in the 9×9 network is represented by a box in the figure. The labels indicate the image unit for each physical word representation. The map is divided into major subareas according to word length.

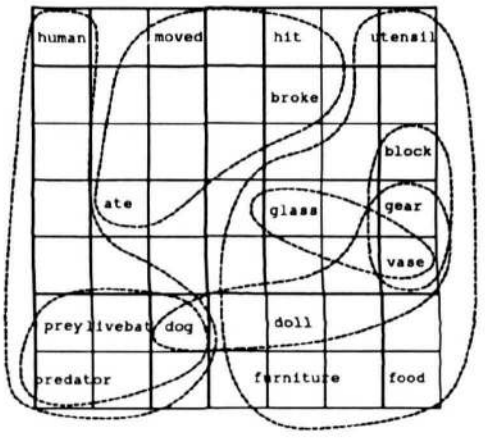

Figure 6: **The semantic map.** The labels on this 7×7 map indicate the maximally responding unit for each semantic word representation. The map is organized according to the semantic categories (table 1).

5.2 Self-organization

The organization of the map, i.e. the assignment of the weight vectors, is formed in an unsupervised learning process [13]. Input items are randomly drawn from the input distribution and presented to the network one at a time (figure 4). The weight vector of the image unit and each unit in its neighborhood are changed towards the input vector, so that these units will produce an even stronger response to the same input in the future. The parallelism of neighboring vectors is increased at each presentation, a process which results in a global order.

The process starts with very large neighborhoods, i.e. weight vectors are changed in large areas. This results in a gross ordering of the map. The size of the neighborhood decreases with time, allowing the map to make finer and finer distinctions between items.

There are several alternatives for implementing the similarity metric, neighborhood selection, and weight change. A biologically plausible process would be based on scalar products of the weight and input vectors, lateral inhibition and redistribution of synaptic resources [14; 20]. These mechanisms can be abstracted and replaced with computationally more efficient ones without obscuring the process itself. The similarity in DISLEX-1 is measured by Euclidian distance, the neighborhood consists of the 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) is

$$\eta_{ij} = \begin{cases} 1.0 - \frac{\|x - \mu_{ij}\|}{\|x - \mu_{max}\|} & \text{if } (i,j) \in N_c(t) \\ 0.0 & \text{otherwise} \end{cases} \quad (1)$$

where μ_{ij} is the unit's weight vector, x is the input vector, $N_c(t)$ is the neighborhood around the maximally responding unit (shrinking with time), and μ_{max} is the weight vector least similar to x in the neighborhood. This forms a nice 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:

$$\Delta\mu_{ij,k} = \begin{cases} \alpha(t)[x_k - \mu_{ij,k}] & \text{if } (i,j) \in N_c(t) \\ 0.0 & \text{otherwise} \end{cases} \quad (2)$$

5.3 Physical and semantic maps

The physical and semantic maps are organized independently, albeit simultaneously, so that associative connections between them can be developed at the same time (see next section). The ordered maps in DISLEX-1 (figures 5 and 6) were obtained in 150 epochs, i.e. by presenting each physical/semantic representation pair (figure 3) to the appropriate map 150 times in random order.

In the self-organizing process, the physical and semantic representations become stored in the weights of the units. For each e.g. physical word, there is an image unit in the physical map, and this unit's weight vector equals the physical representation of that word. The weight vectors of the intermediate units represent combinations of representations. For example, an unlabeled semantic unit between **dog** and **predator** would have features of both domestic and carnivorous animals.

Both maps exhibit hierarchical knowledge organization. Large areas are allocated to different categories of words, and each area is divided into subareas with finer distinctions. The physical map is mainly organized according to the word length. There are separate, adjacent areas for words with 3, 4, 5, 6 and 7 characters. Within these areas, similar words are mapped near each other. For example, **BAT** is mapped between **BOY** and **HIT**, **DOLL** is mapped next to **BALL** etc.

The semantic map has three main areas: verbs, animate objects and inanimate objects. Finer distinctions reveal the semantic categories of table 1. For example, there are subareas for hitters, possessions and fragile-objects, with **vase**, which belongs to all these categories, in the center. Note that the categorization was not directly accessible to the system at any point. It was only manifest in the sentences that were input to the FGREP-mechanism. The

categories were extracted by FGREP, coded into the representations, and finally made visible in the semantic feature map. The final map reflects both the syntactic and semantic properties of the words.

In the self-organizing process, the distribution of the weight vectors becomes an approximation of the input vector distribution [13]. This means that the most frequent areas of the input space are represented to greater detail, i.e. more units are allocated to represent these inputs. For example, the representations for the different animals are very similar (figure 3), yet they accommodate a large area in the map.

The two dimensions of the map do not necessarily stand for any recognizable features of the input space. The dimensions develop automatically to facilitate the best discrimination between the input items. As a result, the ordered areas on the map are likely to have complicated and intertwined, rather than linear shapes.

Feature maps have several useful properties for representing lexical information. (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) The map is continuous, and can represent items between established categories. In other words, words can have soft boundaries. (3) 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. Finally, (4) 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.

6 Word associations

6.1 The physical ⇌ semantic mappings

The physical words do not correspond one-to-one to semantic words. Some words have multiple meanings (homonyms), and sometimes the same meaning can be expressed with several different symbols (synonyms). The mapping between the physical and semantic representations is many-to-many.

The training data for DISLEX-1 contained several such ambiguities (figure 3). The physical word **CHICKEN** could mean a living chicken or food. Similarly, **BAT** could be a baseball bat or a living bat. There were also several groups of synonymous words in the data. **MAN, WOMAN, BOY, GIRL** all have the same meaning **human**, **predator** could be **WOLF** or **LION** etc. In the DISLEX model, the many-to-many mapping between the physical words and their meanings is implemented with associative connections between the physical and semantic maps.

6.2 Associative connections

The physical word maps are fully connected to the semantic map with one-directional associative connections (figure 2). There is a connection from each unit in the physical input map to each unit in the semantic map, and from each unit in the semantic map to each unit in the physical output map. The connection weight indicates the strength of the association. The weights are stored as associative output weight vectors per each unit.

The physical and semantic feature maps and the associative connections between them are organized at the same time. The physical pattern for the word is presented to the physical map, and ordinary feature map adaptation takes place. At the same time, the semantic pattern for the same word is input to the semantic map, and the feature map weight vectors in this map are adapted. At this point, both maps display concentrated patterns of activity. DISLEX learns to associate the physical word with its meaning through Hebbian learning. The weights between active units are increased proportional to their activity:

$$\Delta a_{ij,kl} = \alpha(t)\eta_{ij}\eta_{kl} \tag{3}$$

where $a_{ij,kl}$ is the weight between the physical unit (i,j) and the semantic unit (k,l), and η_{ij} and η_{kl} indicate the activities of these units. The associative weight vectors are then normalized, which in effect decreases the weights on all nonactive output connections of the same unit. This corresponds to redistribution of synaptic resources, where the synaptic efficacy is proportional to the square root of the resource [20]. Initially the activity patterns are large, and associative weights are changed in large areas. As the two maps become ordered, the associations become more focused.

For example, DISLEX-1 was trained by simultaneously presenting pairs of physical words and their semantic counterparts from figure 3. The final associative connections form a continuous many-to-many mapping between the two maps. Unambiguous words have focused connections (figures 7a and 8b). If a physical word has several meanings, or one meaning can be expressed with several synonyms, there are several groups of strong connections (figures 7b and 8a). Units located between image units tend to combine the connectivity patterns of nearby words (figure 8a).

7 DISLEX in action

7.1 Transforming representations

A physical word is transformed to its semantic counterpart (and vice versa) through the associative connections. For example in figure 2, the physical representation of **DOG** is input to the physical map, which forms a concentrated activity pattern around the unit labeled **DOG**. The activity propagates through the associative connections (figure 7a) to the semantic map, where a localized activity pattern forms around the unit labeled **dog**. The semantic representation for dog is now output through the weight vector of this unit. In a similar fashion, a semantic representa-

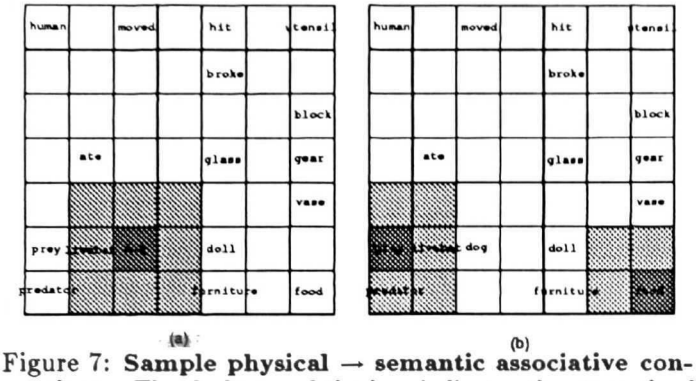

Figure 7: **Sample physical → semantic associative connections.** The darkness of the box indicates the strength of the connection from the physical unit DOG (a) or CHICKEN (b) to the semantic unit. The strongest connections concentrate around the semantic image units. CHICKEN has two possible interpretations, food and prey.

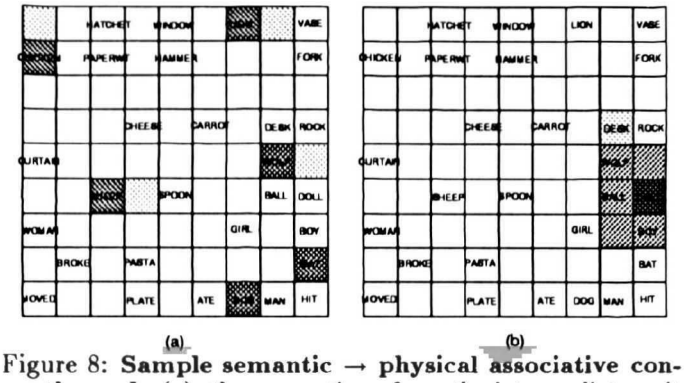

Figure 8: **Sample semantic → physical associative connections.** In (a), the connections from the intermediate unit between dog, livebat, predator and prey are shown. Possible output symbols include all animal names CHICKEN, SHEEP, WOLF, LION, BAT and DOG. In (b), weak connections from doll to nearby units might cause BALL to be output instead of DOLL in noisy conditions.

tion can be transformed to its physical counterpart. The associative connections are different in the two directions, but the same feature map weight vectors are used for both input and output.

The behaviour 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.

7.2 Priming

When an ambiguous physical or semantic representation is input to the lexicon, all possible meanings (or symbols) are activated at the same time (figures 7b and 8a). A top-down priming mechanism is employed to select the correct representation. In addition to the associative activity, the map receives priming activation through its input connections. The activities add up, selecting one of the possible interpretations. If the priming arrives after a short delay, all alternatives are briefly active before one of them is selected. This complies with experimental results [24], which indicate that all meanings of ambiguous words are activated upon reading the word.

The expectations generated by the FGREP mechanism provide a possible source for semantic priming. After reading **The wolf ate the**, the FGREP network generates a strong expectation for **prey** [22]. When the physical symbol CHICKEN is read in, both the **food** and **prey** units are initially equally active in the semantic map (figure 7b). The expectation pattern, which is close to the representation for **prey**, is input to the semantic map and summed up with the activity propagated through the associative connections. As a result, the **prey** unit receives the strongest activity and becomes selected.

The weights on the associative connections represent statistical likelihoods of the associations. A very frequently active connection is much stronger than a rare connection. For example, if most of the occurrences of CHICKEN in training DISLEX-1 would have been paired up with **prey**, the CHICKEN unit would tend to activate the **prey** unit much more than the **food** unit. By default, the **prey** meaning would be selected, and stronger priming for **food** would be required to override it.

DISLEX-1 simply selects and outputs the representation stored at the maximally responding unit. The selection could also be implemented with lateral inhibition, where the map settles into a localized response around the maximally responding unit [20]. The settling times would most likely correspond to the reaction times observed in humans [23]. High-frequency words would have shorter reaction times, and these times could be changed with priming. With several equally likely interpretations, settling would take longer.

7.3 Errors

The DISLEX architecture is well suited into modeling dyslexic performance errors. If the system performance is degraded e.g. by adding noise to the connections, two types of input errors and two types of production errors are observed.

In the input, a physical representation may be mapped incorrectly on a nearby unit in the physical map. This corresponds to reading or hearing the word incorrectly. For example, DOLL may be input as BALL (figure 5). The activity in the physical map may also propagate incorrectly to a nearby unit in the semantic map, in which case e.g. CHICKEN would be understood semantically as livebat (figure 7b).

Analogously in production, a semantic input representation can be classified incorrectly, and a word with a similar but incorrect meaning is produced. For example, if the semantic pattern for **block** is accidentally mapped on **vase** (figure 6), the output reads VASE instead of, say, PAPERWT. Or, the activity in the semantic map may be propagated incorrectly to the physical map, and a word with a similar surface form but different meaning is output. This means generating BALL instead of DOLL (figure 8b).

Errors of this kind occur in noisy, stressful or overload situations in normal human performance. They are also documented in patients with deep dyslexia [4; 3]. The observed visual and semantic paralexic errors can be explained by above mechanisms, giving strong support to the physical/semantic feature map architecture.

If priming is used in the model, there is also a possibility for another type of error, the Freudian slip. This occurs when very strong semantic priming interferes with the output function. For example, if `doll` is input to the semantic map, together with simultaneous priming for **gear**, the activity is propagated through the associative connections of both. As a result, the physical `BALL` might receive the strongest activation, and would be output instead of `DOLL`. The output symbols are similar, but the meaning of `BALL` reveals the semantic priming.

8 Modeling aphasia

The DISLEX architecture is in good agreement with the current theories of the human lexical system [3; 27; 26]. Many observed lexical deficits in acquired aphasia have straightforward explanations in the model.

A common feature of the aphasic deficits is category specificity. The patient may have difficulties only with words belonging to a specific syntactic or semantic category. In certain patients the lexical access to e.g. function words is selectively impaired, in other cases the patient has trouble with verbs [3; 4]. More specific impairments seem to occur in semantic hierarchies. Some patients have trouble with e.g. concrete words, or inanimate objects [28], or even as specific classes as names of fruits and vegetables [10].

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 semantic properties of the words. Localized lesions to the map 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 orthographic or phonological domain. This suggests that the semantic memory, visual input, and verbal input/output modalities are represented in separate structures, strongly supporting the distributed DISLEX architecture.

For example, some patients were unable to access the specific meanings from verbally as well as visually (with pictures) presented cues [26; 28]. This implies that the semantic memory itself, i.e. the map, 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 [16]. 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 [10]. In other words, his semantic memory and verbal input were preserved, and the verbal output function was selectively impaired.

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

The dissociation of the orthographic and phonological modalities is also well-documented. Some patients have deficits only in one of the input or output channels, or different deficits in different channels [2]. For example, a patient may have spelling difficulties exclusively in the orthographic output domain [9; 18]. The types of errors in visual and phonological dyslexia (section 7.3) further indicate that the channels are organized according to the physical forms of the words. The DISLEX model predicts that it would be possible to lose access to specific types of physical symbols, as a result of localized damage to a physical map.

In the aphasic impairments, the high-frequency words are often better preserved than rare words. This is also predicted by the feature map organization. The most often occurring words occupy larger areas in the map, making them more robust against damage.

9 Discussion

The DISLEX model can be locally lesioned, and it displays deficits similar to human patients. This suggests that the model successfully represents some of the physical structure underlying the lexical system in the brain. The architecture is based on word maps, where different units are selectively sensitive to different words in the data. Several low-level sensory maps are known to exist in the central nervous system, e.g. retinotopic maps, tonotopic maps, and also tactile and motor maps. Recently it was found that neurons in the hippocampus respond selectively to visually presented words [11]. These response characteristic could be explained by a map-like structure.

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 maps are used both for input and output, which is not a plausible model of the synaptic efficacies. However, these two-way connections could be implemented with tightly interconnected (or phase-locking) groups of neurons in the brain.

The associative connections between two feature maps

learn a many-to-many mapping from one distributed representation space to another, which is hard to do with other neural network mechanisms such as backpropagation. In the maps, several representations can be active at the same time, whereas e.g. in an assembly-based representation all the different alternatives would be combined into a single average representation pattern [22].

DISLEX is primarily a model of single word processing. It does not have special mechanisms for representing and processing phrasal structures and morphology. There are two possible ways of doing this, and it seems that both of them are involved. Common morphological forms and phrases, such as `nationalism` or `The Big Apple` could be represented like words, as single entries in the physical and semantic maps. More complex phrases and unusual, constructive forms, e.g. `kick the bucket` or `non-preemptive` could be represented in the lexicon by their constituents, and parsed/generated by a higher-level language processing module.

10 Conclusion

The DISLEX architecture models the human lexical system at the level of physical structures. The architecture accounts for many observed dyslexic performance errors and lexical deficits in acquired aphasia. DISLEX-1, the AI implementation of the model, can be used as an input/output filter for a natural language processing system, which communicates with the external world with physical symbol representations, but internally processes semantic representations.

References

[1] Yigal Arens. *CLUSTER: An Approach to Contextual Language Understanding.* PhD thesis, Computer Science Division, University of California, Berkeley, 1986.

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

[3] Alfonso Caramazza. Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system. *Annual Reviews in Neuroscience*, 11:395–421, 1988.

[4] Max Coltheart, Karalyn Patterson, and John C. Marshall, editors. *Deep Dyslexia. International Library of Psychology*, Routledge and Kegan Paul, 1980.

[5] Garrison W. Cottrell and Steven L. Small. A connectionist scheme for modelling word sense disambiguation. *Cognition and Brain Theory*, 6(1):89–120, 1983.

[6] Michael G. Dyer. *In-Depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension.* MIT Press, Cambridge, MA, 1983.

[7] Michael Gasser. *A Connectionist Model of Sentence Generation in a First and Second Language.* PhD thesis, Computer Science Department, UCLA, 1988.

[8] Helen Gigley. Process synchronization, lexical ambiguity resolution and aphasia. In Steven L. Small, Garrison W. Cottrell, and Michael K. Tanenhaus, editors, *Lexical Ambiguity Resolution*, Morgan Kaufmann Publishers, Los Altos, CA, 1988.

[9] R. A. Goodman and Alfonso Caramazza. Aspects of the spelling process: Evidence from a case of acquired dysgraphia. *Language and Cognitive Processes*, 1(4):263–296, 1986.

[10] John Hart, Rita Sloan Berndt, and Alfonso Caramazza. Category-specific naming deficit following cerebral infarction. *Nature*, 316(1):439–440, August 1985.

[11] Gary Heit, Michael E. Smith, and Eric Halgren. Neural encoding of individual words and faces by the human hippocampus and amygdala. *Nature*, (333):773–775, 1989.

[12] Alan H. Kawamoto. Distributed representations of ambiguous words and their resolution in a connectionist network. In Steven L. Small, Garrison W. Cottrell, and Michael K. Tanenhaus, editors, *Lexical Ambiguity Resolution*, Morgan Kaufmann Publishers, 1988.

[13] Teuvo Kohonen. *Self-Organization and Associative Memory*, chapter 5. Springer-Verlag, Berlin; New York, 1984.

[14] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, (43):59–69, 1982.

[15] Geunbae Lee, Margot Flowers, and Michael G. Dyer. Learning distributed representations of conceptual knowledge and their application to script-based story processing. *Connection Science*, 1990. (In press).

[16] Rosaleen A. McCarthy and Elizabeth K. Warrington. Evidence for modality-specific meaning systems in the brain. *Nature*, 334(4):428–430, August 1988.

[17] James L. McClelland and Alan H. Kawamoto. Mechanisms of sentence processing: Assigning roles to constituents. In James L. McClelland and David E. Rumelhart, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume II: Psychological and Biological Models*, MIT Press, 1986.

[18] G. Miceli, M. C. Silveri, and Alfonso Caramazza. Cognitive analysis of a case of pure dysgraphia. *Brain and Language*, 25:187–212, 1985.

[19] Risto Miikkulainen. *A Neural Network Model of Script Processing and Memory.* Technical Report UCLA-AI-90-03, Artificial Intelligence Laboratory, Computer Science Department, University of California, Los Angeles, 1990.

[20] Risto Miikkulainen. *Self-Organizing Process Based on Lateral Inhibition and Weight Redistribution.* Technical Report UCLA-AI-87-16, Artificial Intelligence Laboratory, Computer Science Department, UCLA, 1987.

[21] Risto Miikkulainen and Michael G. Dyer. Encoding input/output representations in connectionist cognitive systems. In David S. Touretzky, Geoffrey E. Hinton, and Terrence J. Sejnowski, editors, *Proceedings of the 1988 Connectionist Models Summer School*, Morgan Kaufmann Publishers, 1989.

[22] Risto Miikkulainen and Michael G. Dyer. Natural language processing with modular neural networks and distributed lexicon. 1989. Submitted to *Cognitive Science.*

[23] Greg B. Simpson and Curt Burgess. Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, 11(1):28–39, 1985.

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

[25] David L. Waltz and Jordan B. Pollack. Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science*, (9):51–74, 1985.

[26] Elizabeth K. Warrington. The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, 27:635–657, 1975.

[27] Elizabeth K. Warrington and Rosaleen A. McCarthy. Categories of knowledge: Further fractionations and an attempted integration. *Brain*, 110:1273–1296, 1987.

[28] Elizabeth K. Warrington and T. Shallice. Category specific semantic impairments. *Brain*, 107:829–854, 1984.

[29] Uri Zernik. *Strategies of Language Acquisition: Learning Phrases from Examples in Context.* PhD thesis, Computer Science Department, University of California, Los Angeles, 1987. Technical Report UCLA-AI-87-1.