Learning to represent and understand locative prepositional phrases

Cynthia Cosic
Paul Munro

University of Pittsburgh
Department of Information Science

1. Introduction

One of the principal problems in natural language processing (NLP) is the role of context in understanding the sense of a particular word in a particular situation. Therefore the connectionist approach, which has fared quite well in accounting for context sensitivity, has become an attractive approach to NLP (eg. Cottrell & Small, 1983). This report presents a network that learns the spatial relationships that can be communicated through the use of a locative preposition. The goals of this network were to both "decode" and encode" prepositional phrases. (Herskovits, 1986) The decoding task involves understanding the spatial relationship between two nouns (N1 and N2) described by a locative preposition (LP); consider, for example, the different uses of the LP *in* accessed by the phrases *water in glass* and *crack in glass*; in the first case, N2 is acting as a container for N1 whereas in the second N1 is within the substance of N2. The encoding task is the converse problem: finding the appropriate preposition to express a given spatial relationship between two particular objects.

In addition to discovering whether the network could learn to encode and decode prepositional phrases, another area of interest was studying the context-free meanings of the prepositions themselves. The ability of connectionist networks to complete partial patterns (e.g. McClelland, Rummelhart, and Hinton, 1986) makes it possible to study the response of the network to a preposition alone, in the absence of the nouns. In this case, the question of whether these meanings were the same as the ideal meanings defined by Herskovits (1986) was investigated. Herskovits's (1986) theory of the meanings of locative prepositional phrases is built around the notion of an "ideal meaning", which is like a prototype. (Herskovits, 1986, p. 39) It is constructed out of the "perceptually salient" characteristics of a set of objects and their relationship (Herskovits, 1986, p. 54). Her ideal meanings of the prepositions *in*, *on*, and *at* are:

> The ideal meaning of *in*:
> The inclusion of a geometric construct in a one-, two-, or three-dimensional geometric construct. (Herskovits, 1986, p. 48)
>
> The ideal meaning of *on*:
> For a geometric construct X to be contiguous with a line or surface Y; if Y is the surface of an object Oy and X is the space occupied by another object Ox, for Oy to support Ox. (Herskovits, 1986, p. 49)
>
> The ideal meaning of *at*:
> For one point to coincide with another. (Herskovits, 1986, p. 50)

Finally, the way the network grouped the nouns used as arguments to the preposition was studied. The network was designed to build its own internal representation for each noun. Once again, Herskovits's (1986) theory of locative prepositional phrases provides clues as to now the nouns might be divided by the system. While Herskovits feels that the ideal meaning is at the core of each use of a preposition, she also states that in actual use the meaning is often changed or shifted in some manner. (Herskovits, 1986, p. 39) A particular use of a prepositional phrase, then, often has a meaning similar to (but not exactly the same as) the ideal meaning of the preposition. The structure she uses to hold the relationship between an ideal meaning of a preposition and the meaning embodied in a particular instance of using a preposition is called a use type. (Herskovits, 1986, p. 87) We did not represent all the use types in our simulation; in particular, our current architecture cannot support those which require more than two arguments to the preposition, and we have chosen to exclude abstract relationships such as membership (e.g. student in school) in our initial studies. The use types that were represented in the stimulus set suggested possible ways that the network might categorize the nouns. These divisions are shown in Table 1.

Institution	
Gap	
Part/Whole	
Geometric Entity	Point Line Plane
Spatial Entity	Geographic Location Landmark Generic Location
Physical Object	Offers support Is a Media Has an Outline One of a group Container Usable Artifact Person

Table 1. Possible divisions of nouns, corresponding to distinctions required by Herskovits.

2. Network architecture

The network currently under study is feed-forward, with an input layer partitioned into four banks, three intermediate banks, and an output layer partitioned into two banks, organized as in Figure 1.

The network is a variation of the "encoder architecture" introduced by Ackley, Hinton, and Sejnowski (1986), which is trained to replicate patterns of input at the output layer. A pattern of the form N1-LP-N2 or N1-SP-N2 is presented to the network by activating from three to five units: one unit in bank A corresponding to N1, a noun; one unit in bank D corresponding to N2, a second noun; and either one unit in bank C, corresponding to the LP, the locative prepostion, or from one to

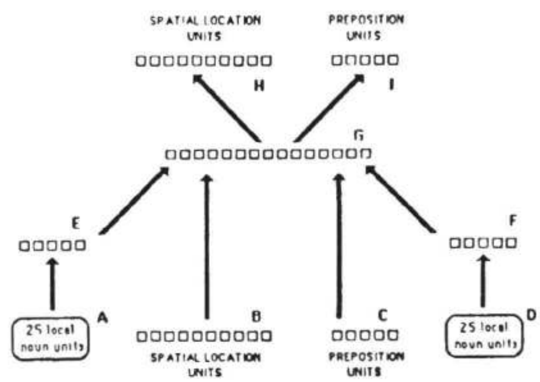

Figure 1. Network architecture for the preposition task. Inputs are presented at the lowest layer, across either banks A, B, and D or A, C, and D. The bold arrows indicate the flow of information from one bank to another. For example, every unit in bank F contributes to the activation of every unit in bank G. After each pattern presentation, the outputs at H and I are compared with the corresponding SP and LP components of the input pattern (banks B and C) to generate an error signal for the learning procedure.

three units in bank B, which correspond to the SP portion of the input, the units that describe the spatial relationship between the two nouns. Table 2 gives a list of all of the units that have been used. Because the network receives incomplete stimuli, and the output level is trained to supply both the LP and SP units, the network can be said to perform a pattern completion task.

Banks E and F receive information exclusively from banks A and D respectively. Therefore, the weight matrix from A to E (let this matrix be denoted [A,E]) must adapt such that representations of N1 are distributed in a manner that facilitates solution of the task. Similarly, F must come to give appropriate representations of N2. Careful study of these representations will be a major component of the task analysis. It should be noted that matrix [A,E] is constrained to be equal to matrix [D,F]; that is, a given noun comes to have the same representation in bank E as it has in bank F, so that important properties of nouns when they occur in the first position may carry over to situations where they occur in the second.

3. Simulation

3125 (25 x 5 x 25) pattern combinations can be formed with the twenty-five nouns and five prepositions; of these, 99 were chosen to constitute a "training corpus". During each training cycle, two input-output pairs were presented to the network. The first half-cycle consisted of presenting N1 on bank A, LP on bank C, and N2 on bank D; an error signal was generated by comparing the resulting response at H and I with the desired response, consisting of SP and LP respectively. This error signal was applied using the back-propagation procedure of Rumelhart, Hinton, and Williams (1986) to adaptively modify the weight values throughout the network. The second half-cycle used the same stimulus pattern except the input consisted of N1 on bank A, SP on bank B, and N2 on bank D. Each unit in the network responds as a semi-linear function of its input values; that is, the inputs were linearly summed to give a net activation value x, from which the response r was calculated using the logistic function:

$$r = \frac{1-e^{-x}}{1+e^{-x}}$$

Noun Units (25) Banks A and D	clouds	lake	campsite	table	book
	sky	river	school	glass	flowers
	plane	road	house	bowl	grass
	boat	city	floor	crack	man
	water	island	room	chip	fish
Semantic Units (10) Bank B	N1 over N2		N1 within border of N2		
	N2 over N1		N1 touching N2		
	N1 at edge of N2		N1 near N2		
	N1 embedded in N2		N1 far from N2		
	N2 contains N1		N2 supports N1		
Preposition Units (5) Bank C	in	at	on	under	above

Table 2. Unit labels for coding input patterns across banks A, B, C, and D

4. Results

The network required from 1,500 to 2,500 passes through the corpus to learn the prepositional phrases with an average error of less than 2%. Various tests were performed on the trained network in order to determine the ideal meaning of each preposition and the network's classification of the various nouns. Specifically, the question of whether the network would follow Herskovits's (1986) theory for the preposition's meanings and the types of nouns that can be arguments to the preposition was studied. In addition, the role of context in determining the meaning of a prepositional phrase, and the responses of the network to novel stimuli were investigated. The following paragraphs will deal with each of these findings separately.

After training, to test whether the network developed ideal meanings consistent with Herskovits (See Introduction), each LP (Bank C) was presented to the network alone, and the SP output (Bank H) was observed. The output representations are shown in Figure 2; filled rectangles indicate positive activity values, and open rectangles indicate negative activities.

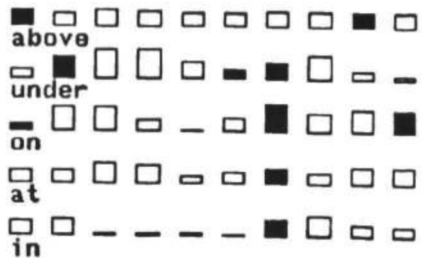

Figure 2. Bank I (spatial) output generated from prepositions alone. After training for from 150,000 to 250,000 pattern presentations (until a very low error is achieved by the network) the system is probed by observing the output of the H (SP) bank to each LP (bank C) input. The positively valued outputs (dark rectangles) correspond to: *in* = embedded in, and touching; *on* = 1 over 2, touching, and supported by; *at* = touching; *under* = 2 over 1, embedded, and touching; and *above* = 1 over 2 and far.

Note that during training, only the preposition *at*, always co-occurs with the same SP representation. While the meanings of *on*, *at*, and *above* do match those that would be expected, the meaning of *in* is not inclusion, as would be predicted. This is probably due to the high number of phrases that have to do with a physical object embedded in some medium, e.g. *the clouds in the sky*. The same is true for the preposition *under*, where the majority of the training examples were phrases such as, *the fish under the water*.

The distributed representations of the nouns were also studied. Because the weight matrices [A,E] and [D,F] were constrained to have the same values, the representation of a noun is the same in banks E and F; these are shown in Figure 3a. We were not able to determine a clear meaning for each unit in the distributed noun representations. However, the similarity structure of this representation was revealed through cluster analysis; the organization imposed upon the hidden units by this network via the learning procedure in order to solve the task (Figure 3b), correlates nicely with the one predicted from Herskovitz's theory (See Table 1).

The output of the network is context sensitive; many of the prepositional phrases used by the network do not have an output representation that corresponds to the preposition's ideal meaning. The output representation of a prepositional phrase is determined not only by the preposition, but also by the particular nouns used in the phrase. For example, the phrase *boat in water* is generally associated with the SP unit *N1 on top of N2* and *N1 supported by N2*, whereas *fish in water* fits much more nicely with *N1 embedded in N2*. To test context sensitivity, all the patterns of one noun and one preposition were presented to the network. At times, these representations are different from not only the preposition's ideal meaning, but also from the final meaning of a full prepositional phrase that incorporates that noun and that preposition. An example of this is shown in Figure 4.

a

b

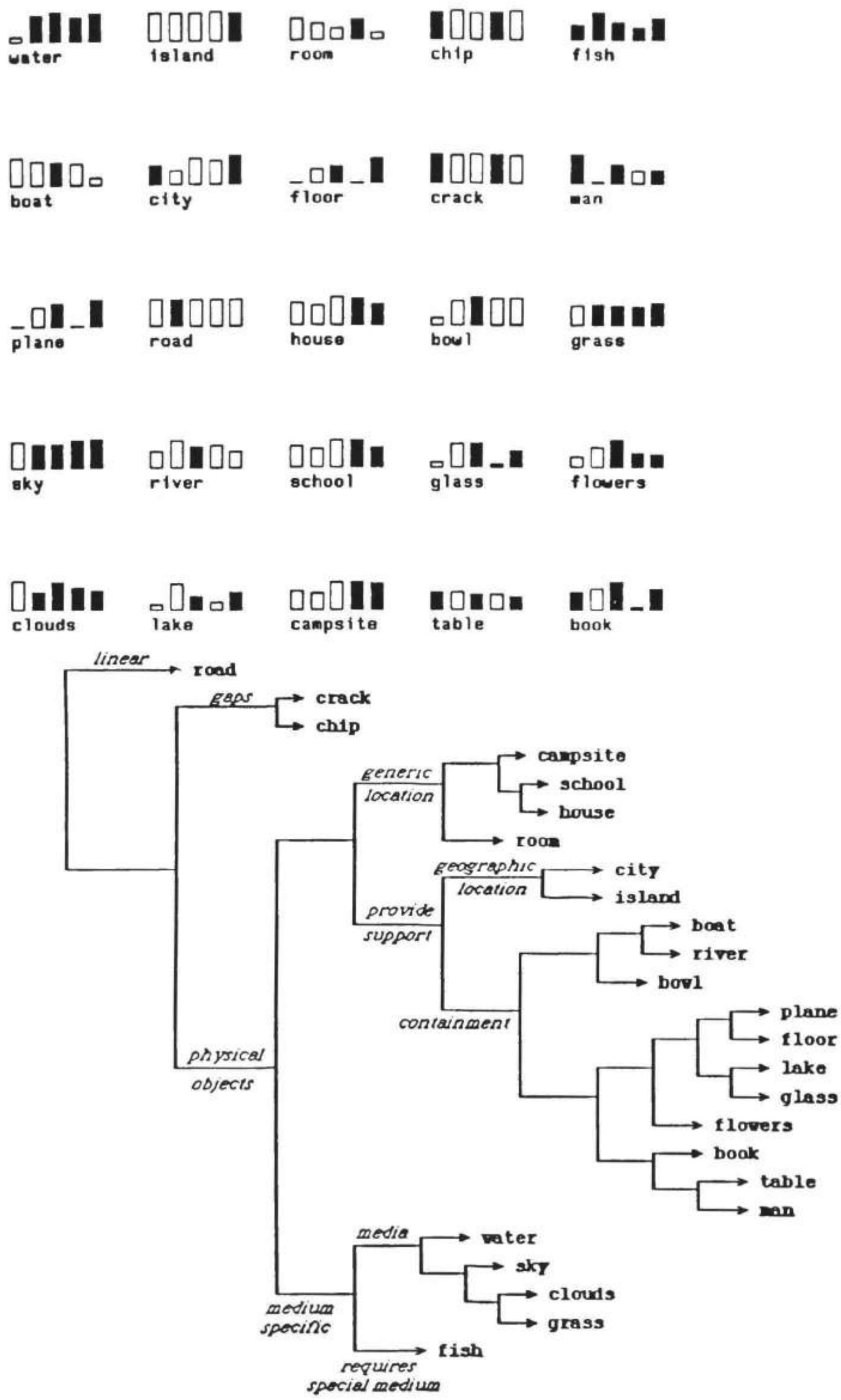

Figure 3. Distributed representations of the nouns. In order to accomplish the task successfully, the learning procedure was forced under this architecture to "discover" suitable distributed representations for the nouns. (a) the noun representations; (b) the tree represents the annotated results of a cluster analysis on these representations.

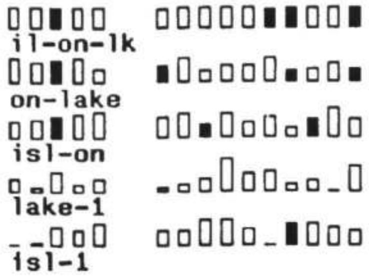

Figure 4. Context Sensitivity. The network's output in response to each stimulus is different. The response to each word alone is different from that given by one noun and a preposition, and from the response to the full phrase.

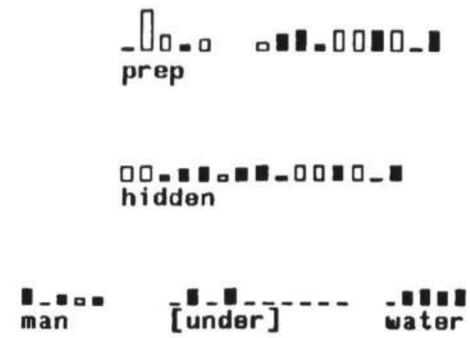

Figure 5. Network processing of the input *man [under] water*. The response of each layer in the network to the input is depicted. Note that the output units for both the preposition *under* and the preposition *in* are weakly activated.

The trained network was also run on a stimulus set of thirty-three novel phrases. Twenty-three were interpreted correctly. One of the more interesting results (Figure 5) was the response to the phrase *man [under] water*, where the brackets denote the appropriate semantic representation for *under* ("N2 over N1" and "N1 embedded in N2"). When the network was given N1, N2, and SP, the preposition chosen by the network was *under*. In addition, however, another preposition received activation (although less activation than *under*): *in*. Some of the results that were different from the correct response were plausible interpretations. For example, the phrase *grass on water* evokes either of two images -- *grass floating on water*, or *grass standing on the edge of the water*. The network, when given N1, N2, and LP gave the semantic units for the latter relationship, which initially surprised us, since we had not considered that interpretation before.

5. Discussion

The network was able to solve both the encoding and decoding task with negligible error. To solve the problem, reasonable meanings for the prepositions alone, and reasonable distributed representations for the nouns were developed. Because of this, the network was able to reasonably handle many of the novel inputs. Perhaps the most interesting aspect of this network lies in the fact that it is used to express the *semantic content* of any of a number of locative prepositions in a variety of contexts. We plan to explore the following conjecture: This semantic representation is language independent, therefore the mapping of semantic content to preposition is learned in both directions (i.e. prepositions can be *decoded* into semantics and semantics can be *encoded* into prepositions), semantic representations can be used to translate locative prepositions from one language to another.

Thus, we plan to explore the potential application of this work to machine translation. Other possible directions for extending this work include [1] increasing the scale of the model system (both in terms of the words used and the number of input patterns), [2] alternative architectures that can be used for variations on the encoding and decoding task, including metaphorical uses of prepositions, and [3] extending the uses of the prepositions beyond concrete spatial relationships to more metaphorical roles.