Behavior Research Methods, Instruments, & Computers
1998, 30 (2), 264-271

Abstract noun classification: Using a neural network to match word context and word meaning

KATJA WIEMER-HASTINGS
University of Memphis, Memphis, Tennessee

Psychologists have used artificial neural networks for a few decades to simulate perception, language acquisition, and other cognitive processes. This paper discusses the use of artificial neural networks in research on semantics—in particular, in the investigation of abstract noun meanings. It is widely acknowledged that a word's meaning varies with its contexts of use, but it is a complex task to identify which context elements are relevant to a word's meaning. The present study illustrates how connectionist networks can be used to examine this problem. A simple feedforward network learned to distinguish among six abstract nouns, on the basis of characteristics of their contexts, in a corpus of randomly selected naturalistic sentences.

Since Rumelhart, McClelland, and the PDP Research Group published their work in 1986, neural networks have been used in a variety of psychological research areas to test theories and to simulate psychological processes. These areas include language and speech processing, memory processes, visual perception, and other cognitive processes. Examples of neural networks in psycholinguistic research are (1) the acquisition of past tense morphology (Hoeffner, 1996; Rumelhart & McClelland, 1986), (2) syntactic parsing models (see, e.g., Mikkulainen, 1996), (3) word disambiguation (Eizirik, Barbosa, & Mendes, 1993; Gallant, 1991; Waltz & Pollack, 1985), (4) comprehension of connected discourse (Kintsch, 1988, 1998), and (5) the sequencing of speech act categories (Graesser, Swamer, & Hu, 1997). However, researchers have not yet used a network to model the relationship between lexical entities and the features of their context. The research in this paper was designed to fill this gap.

The traditional approach to concept learning and representation is based on the assumption that concepts can be decomposed into features (Katz & Fodor, 1963; Miller & Johnson-Laird, 1976). However, researchers have recently argued for a more flexible representation of concepts, in which a concept contains features that are dependent on a given context. Barsalou (1982; Barsalou & Medin, 1986), for example, has demonstrated that each particular context selectively activates context-dependent features of concepts that are used for concept construction in working memory. Anderson (1990) contends that a word has a different meaning (or sense) in every new context of use, which implies absolute context dependency of word meaning.

Miller (1991) and Miller and Charles (1991) have claimed that words are similar to the extent to which they occur in similar contexts. He has argued that context information is stored with concepts in the mental lexicon. Miller specifies four types of such context information—namely, collocation, semantic, syntactic, and pragmatic context.

Evidence for context dependence is also apparent in tasks employing the cloze method (Bloom & Fischler, 1980; Hamberger, Friedman, & Rosen, 1996; Schwanenflugel, 1986), which was introduced by Taylor (1953). In this method, subjects receive a sentence in which a word has been replaced by a blank, and they are asked to fill in a word that best fits the context. Some contexts provide enough information for the subject to fill in the correct missing word into the blank. However, this result does not unarguably demonstrate that the meaning of the missing word can be derived from its context. When considering Miller's (1991; Miller & Charles, 1991) types of context information, the ability to fill in a correct word may be explained by collocation information—that is, associations of words that frequently cooccur in sentences—rather than by the semantic context. In order to investigate the relation of context and word meaning, it consequently is necessary to examine a broader set of context features.

In the case of abstract nouns (e.g., *goal*, *idea*), the necessity of a context-sensitive representation is even more evident. Abstract nouns have no directly perceivable reference (see, e.g., Schwanenflugel, 1991). It is, therefore, difficult, perhaps even impossible, to decompose these concepts into features. The question, then, is how one can know what abstract nouns mean.

The author gratefully acknowledges Arthur Graesser, Jim Hoeffner, and three anonymous reviewers for very helpful comments on earlier versions of this paper and Arthur Graesser for valuable discussions related to the study. Thanks to Phil Goodman for his help with the software and for his kind permission to use material from the manual and program in this article. Correspondence should be addressed to K. Wiemer-Hastings, Department of Psychology, The University of Memphis, Campus Box 526400, Memphis, TN 38152-6400 (e-mail: kwiemer@cc.memphis.edu).

Table 1
List of Context Features

Module label	Features
World knowledge	work, family/relationship, production, research, politics/law, art, technology, entertainment, communication, education
Ontological status	process, event/state, object, mental entity
Adjective modifier	qualification, comparison, relevance, time duration, positive, negative
Target-related verb	construction, destruction, biology, communication, motion, translocation, cognition, emotion, possession, temporal relation, evaluation, causative, intention, agentive
Case role	agentive, instrumentative, dative, factitive, locative, objective
Event dynamics	change, cause, effect
Time reference	future, presence, past
Syntactic surface features	main clause/subclause, position within clause (beginning/end), modifying article, plural form, modifier quantifier, referential frame of target, person has _____

Note—The features are organized into eight modules, which represent different types of context information.

The theoretical claim of the present study is that abstract concepts depend strongly on context. This view has been previously formulated in the context availability model—for example, by Schwanenflugel and Shoben (1983). According to this model, abstract noun representations are only weakly connected to associated context information and are, therefore, processed more easily when presented in a rich context that activates the associated information in memory. Because the reference of abstract nouns is not visible, it is conceivable that the language learner has to analyze the contexts in which an abstract noun is used in order to identify its reference. For example, in order to understand what *idea* means, it may be necessary to know that having an idea has certain consequences, such as trying a new approach to solving a problem. A language learner must construct the meanings of abstract nouns from the contexts in which they occur. The role of context in the acquisition of word meanings motivates the hypothesis that the meaning of abstract nouns may be *determined*, not just influenced, by the context of use.

It should be noted also that lexicographers derive a word's definition from example sentences in which the defined word is used. This nicely demonstrates that the context of words contains information about their meaning. It is consistent with Wittgenstein's (1953) view that the meaning of words is identical to their use.

There is some psychological evidence that context is critical for abstract nouns. For example, it has been shown that the memory advantage of visualizable material disappears if both abstract and concrete words are embedded in context (Schwanenflugel & Shoben, 1983). In a discussion of word learning, Stahl (1991) discussed some studies that looked at how children derive the meanings of unknown words from context (see, e.g., Elshout-Mohr & van Daalen-Kapteijns, 1987). After exposure to different contexts, the learner ideally starts to decontextualize the word meaning (McKeown, 1985).

The empirical investigation of context-dependent concept meanings presents a challenge. Context analysis is a complex task, and it is unknown which elements of context are relevant. Two recent projects on semantic representation have emphasized the role of word context—namely, HAL (Hyperspace Analogue to Language; see, e.g., Lund & Burgess, 1996) and LSA (Latent Semantic Analysis; see Landauer & Dumais, 1997). In these systems, context is not represented by features per se, as in the network discussed in the present article, but by the cooccurrence of words in contexts. The differences between the present network and these alternative approaches are discussed later, in the Discussion section.

The present study describes a neural network that investigates the role of context in abstract concept representation. In order for context to determine the meaning of a word, the contexts of the word must share specific features. In particular, the contexts of a given word must be sufficiently distinct from the contexts of different, unrelated words, whereas the contexts of similar words (synonyms) should have features in common. In this study, a vector of context features represents context information. The features are listed in Table 1.

The network was designed to assess whether context information is sufficient to determine the meanings of abstract nouns. I do not wish to conclude anything about how humans learn concepts from the network's performance. It has been argued that words are learned from context (see, e.g., Sternberg & Powell, 1983), but this network is not intended to be a model of such human learning. However, it supports such views, by showing that context information is *in principle* sufficient for distinguishing different abstract concepts.

The network was trained to identify the target nouns on the basis of context features. The performance of the network assesses (1) whether the information in the contexts is sufficiently distinctive to classify abstract nouns and (2) whether the included features cover the relevant information.

THE NETWORK

The network examined whether six abstract nouns can be distinguished on the basis of 53 characteristics of

their contexts. The network received 53 context features extracted from particular sentences as input and was trained to identify the abstract noun that occurred in the sentence on the basis of the context features.

The trained network also provided an estimate of the relevance of each of the 53 context features for the network's performance. Theoretically, this can be done in at least three different ways. One way is a lesion simulation study, in which the weights of the connections radiating from an input unit are set to zero, so that the activation from this input unit is blocked. The drop in the performance of the network is an indicator of the relevance of that feature. A similar common approach is a sensitivity analysis. Input values are deleted from the input vector one at a time, and the performance of the resulting network is compared with the performance of the network trained with all of the data. In the present study, 54 different networks would have to be trained to perform this kind of analysis. In one network, all of the 53 features would be included; in the remaining 53 networks, 1 feature at a time would be deleted from the input.

A third procedure, which was used in this study, is called automatic relevance determination (ARD). ARD has been developed by MacKay and Neal (Neal, 1996) and involves a hyperparameter that assesses the penalty assigned to the network's connections. On the basis of the penalties, it estimates the relevance of the variables or features. In NevProp (Goodman, 1996)—the program used in this study—ARD can be used either during training (during which it increases the impact of relevant features and suppresses that of irrelevant ones) or after training. The present study used the latter method. ARD after training can be used "to compute, for the final model, summary estimates of variable relevance and number of well-determined parameters (effective degrees of freedom used by the model)." (Goodman, 1996, p. 152; see section 7.7 on ARD settings for more information).

The advantage of ARD is that it is neither tedious nor time consuming. A disadvantage is that ARD values often vary among networks performing the same task, even if the networks show a similar performance. Therefore, it was necessary to combine ARD with another procedure, in order to test the reliability of its estimates. In this study, a sensitivity analysis also was performed, in which input units with low ARD values were deleted from the training data. The performance of the newly trained network should be comparable to the network with the complete input, if the ARD estimates are reliable.

Training and Test Data

Six abstract nouns were randomly selected from a corpus (Carroll, Davies, & Richman, 1971) as output targets: *concept*, *consultation*, *goal*, *idea*, *impression*, and *wisdom*. Isolated sentences were selected from the NexisLexis database that contained one of the six nouns. For each word, 100 sentences were selected from the database. Semantically depleted sentences, such as "This was the goal," were not included, because they do not contain any information about the word's meaning. The following sentence is an example sentence that was included in the corpus: "A similar concept, tunneling, is being developed as a means to let telecommuters access the company (via) intranet from their own houses."

Given that there were six nouns and 100 sentences for each noun, 600 training cases were constructed from the sentences. Fifty-three semantic and syntactic features were manually extracted for each sentence. The 53 features are listed in Table 1.

Cottrell (1989) and many others have argued that single units (like words) are only of minor relevance for meaning. Instead, one should examine larger constituents, such as clauses, sentences, and discourse. The manner in which a word is interconnected with other parts of a phrase is important semantic information. The contextual information was included in this analysis by coding features of the verb and adjective that are directly related to the abstract nouns, the case roles of the nouns, and the information expressed by the whole sentence (including domains of world knowledge).

A surplus of features was not a problem in this study, because the network is capable of determining, during training, which features are necessary and which are redundant with respect to the correct solution. Those features in the set that were not used in the analysis process by any sentences were deleted from the set.

The context features belong to two general subgroups, semantic and syntactic context features. These are two of the context information types specified by Miller (1991; Miller & Charles, 1991). Most of the modules are semantic. Each sentence was assigned features that corresponded to general world knowledge domains. The domains included were based on their occurrence in the corpus at hand. The ontological status of the target nouns was represented by four types of status, some of which correspond to the types suggested by Vendler (1967).

The adjective and the verb that was directly related to the abstract noun were classified. Adjectives were coded with respect to a few features only. The classification of adjectives is extremely difficult. Atchison (1994) points out that the meaning of adjectives can vary, depending on the related noun (see also Lahav, 1989, for examples). Categories of adjective meaning that depend on the noun are useless in a test of the impact of the adjective's meaning on the noun's meaning, because they require a priori knowledge of the meaning of the noun. The adjectives were classified with respect to their connotation (positive/negative), with respect to whether they specified duration (e.g., *long*) or relevance (e.g., *important*), and with respect to qualification and comparison, which are roughly equivalent to the categories *pertainyms* and *ascriptive* adjectives, as specified by Gross and Miller (1990). Verbs were classified according to verb classification categories from Miller and Johnson-Laird (1976) and according to the analysis of predicates in Graesser

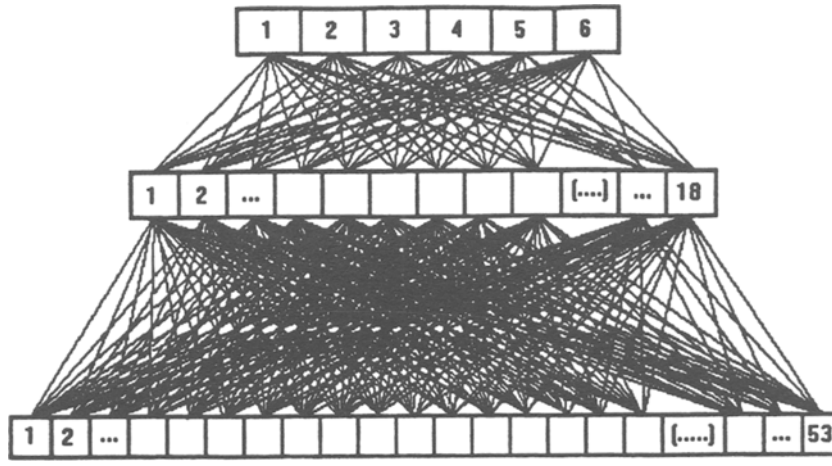

Figure 1. The architecture of the neural network. The lowest level represents the 53 input features; the middle level is a hidden layer with 18 nodes; and the top level contains the six output units corresponding to the six nouns. The units of adjacent levels are fully connected.

and Clark (1985). A few categories were added, as required by the corpus. The target nouns were assigned a case according to Fillmore's case grammar (1968).

Besides these modules, contexts contain other obvious kinds of information that are not theoretically grounded but are explicit in the sentences. For the sake of completeness of the analysis, such information was included in the feature set. It includes modules for dynamics (effects, consequences) that are related to the abstract noun, for time reference in the sentence, and for syntactic surface features.

An explicit effort was made only to include features of the *context*—that is, features that could be analyzed even if there was a blank instead of the target word. In particular, the feature group labeled *ontological status* may be misleading in this respect. The ontological status of a noun is a feature of the noun. However, the ontological status of a noun, in many cases, can be derived from the context of the noun. Vendler (1967), for example, has pointed out that particular kinds of nominalizations—that is, nouns that have been derived from a verb or adjective—only occur in particular contexts, or *container sentences*. Examples for such kinds of nominalizations are *processes* or *achievements*. Contexts contain information that is related to these ontological statuses (e.g., *process*). The context puts constraints on what the ontological status of the noun in question can be. In the present analysis, the feature set for a sentence included ontological status information only if the status could be derived from the context. The context can contain explicit information about this—for example, in the sentence "The ____ process will include looking at revamping regulation to make it easier to . . . ." It can also contain implicit information, as in "For about the price of a one-hour ____ with an in-the-flesh landscaper, you can consult . . . . ." In both cases, it can be seen from the context that ____ is a process: It transpires over time. The missing abstract noun in these examples is *consultation*. Whenever the context did not constrain the ontological status, all of the four features for ontological status were set to zero.

Each sentence was represented by a vector of 53 binary values that was fed into the input layer of the network. Each feature was assigned a 1 if it applied, and a 0 if it did not apply. Six output units represented the target nouns. They were either *on* (highest activation) or *off* (lower activation). Two sets of test data were constructed to test the network's generalization. For each target noun, 10–11 additional sentences (altogether, 62) were analyzed in the same way as were the training data. In addition, eight sentences were analyzed that contained synonyms of the target nouns or nouns that were closely related to them in order to test the network's performance on these. The related nouns were listed as synonyms for the target words by WordNet (Miller, 1990). The idea behind this synonym test is that, if word meanings are captured by context, similar words should have similar contexts (Miller, 1991; Miller & Charles, 1991). Accordingly, the trained network should be able to classify sentences that contain synonyms of the target nouns, such as *belief*, as those target nouns that are synonymous, such as *impression*.

Description of the Network

The NevProp software (Goodman, 1996) that was used in this project is available for free, along with an excellent manual, at the following URL address: ftp://unssun.scs.unr.edu/pub/goodman/nevpropdir It runs on UNIX, DOS, and Mac platforms. The program can be used to train feedforward networks with a variable number of hidden layers. NevProp can be used for ARD.

Figure 1 presents the architecture of the network. It has a 53-18-6 fully connected feedforward architecture, with 53 input units, 18 hidden units, and 6 output units.

Table 2
Set-up of the Network in NevProp

Line	Program
	# NP FILE SETTINGS
1	ResultsFile a1.res
2	SaveWeightsFile a1.wts
3	SaveTrainPrdFile a1.ptr
4	SaveTestPrdFile a1.pts
	# DATA FILE SETTINGS
5	ReadTrainFile ani.trn
6	ReadTestFile test.tst
7	NHeaders 1 IDColumn YES
8	StandardizeInputs 1 SaveStandWts NO ImputeMissing median
9	ShuffleData YES
	# CONNECT CALLS
10	Connect 1 53 54 71
11	Connect 54 71 72 77
	# CONFIGURATION SETTINGS
12	Ninputs 53 Nhidden 18 Noutputs 6
13	kNN 0 1ofN YES
14	HiddenUnitType 2 OutputUnitType 2
15	WeightRange 3
	# TRAINING SETTINGS
16	TrainCriterion 0 BiasPenalty NO WeightDecay -0.001
17	OptimizeMethod 1 SigmoidPrimeOffset 0
18	Stochastic NO LearnRate 0.02 SplitLearnRate NO Momentum 0.002
	# BEST-BY-HOLDOUT SETTINGS
19	NHoldout 150 PercentHoldout 0
20	AutoTrain YES MinEpochs 500 BeyondBestEpoch 2
21	NSplits 5 SepBootXVal NO
	# AUTOMATIC RELEVANCE DETERMINATION SETTINGS
22	UseARD YES WhenARD Auto ARDTolerance 0.05 ARDFreq 25
23	GroupSelection Input BiasRelevance NO ARDFactor 1

Note—Important lines of the program are explained in the text by referring to the line numbers. Parts of the program have been deleted because they are not directly relevant to the present article. The lines marked with the sign # describe the function of blocks of the program and are not parts of the program. The program structure is part of the NevProp software and is adapted with permission from *NevProp software, Version 3*, by P. Goodman, 1996, Reno: University of Nevada. Copyright 1996 by Philip H. Goodman.

Table 2 demonstrates how this network is set up in NevProp. Lines 10–11 in Table 2 display the connections. Both the hidden and the output units had logistic transfer functions (line 14). The network had to select one output unit as the *correct* one for each case (1ofN, line 13). Abstract nouns are often interrelated in such a way that, occasionally, two different abstract nouns could occur in a similar sentence. However, a previous network that was permitted to come up with several guesses at a time double-classified only about 2% of the cases.

The network was trained with the *Autotrain* procedure (lines 19–21), which will be described in more detail in a later section. It reduces the risk of overfitting the network, which means that a network's performance on the training data becomes excellent but generalizes poorly to new data. The network was trained with the backpropagation algorithm. Further parameter specifications can be read from Table 2.

Training

The network was trained in two phases. In the first phase, only 75% of the training set was used for training. The remaining cases were used to estimate the network's generalization performance during training. The network was trained five times in this way. After each of the five training sessions, the program stored the average square error (ASE) value that occurred when the error for the holdout set was minimal—that is, when the network generalized best. After phase one, the program computed the mean of the five stored error values. This mean ASE served as the target error in training phase two, in which the network was trained with the full data set. Training stopped automatically when the target error was reached. The training procedure is comparable to what presumably happens when a child learns the meaning of an abstract noun: It associates the contexts with the terms.

RESULTS

After training was completed, the network achieved an average ASE of .085 for the 600 cases it had been trained on. This value is computed as the squared difference between the true output values and the network's predicted values, averaged over all the cases.

The generalization of the network was tested with 62 new cases on which the network had not been trained. This test is comparable to a cloze experiment, in which humans have to fill in a missing abstract noun into a sen-

tence frame. The ASE for the test sentence set was .121. The network classified 63% of the test cases correctly. This number is not perfect (100%). However, one has to consider that, for six output units, performance at chance level would be a classification rate of 17%. Thus, 63% correct classification means that the network performs clearly above chance. Also, human performance on the same test would not necessarily be 100%.

The network was also tested on the contexts of *synonyms* of the target nouns. The following terms were selected as synonyms for the target nouns: *belief* for impression, *thought* for concept and idea, *counseling* for consultation, *intention* for goal, *plan* for goal and concept, and *knowledge* and *experience* for wisdom. The network classified six of the eight cases as the synonymous target nouns. The two terms that could not be classified as their synonyms were experience (classified as idea) and knowledge (classified as concept).

In summary, the network could correctly classify the majority of the cases. This means that the features taken from the sentence contexts are, in most cases, sufficient to distinguish among the six target nouns. The results suggest that words can be distinguished and identified on the basis of their contexts.

AUTOMATIC RELEVANCE DETERMINATION

The second question of this study addressed the relevance of the 53 features. ARD was used to estimate the relevance of each of the input features for the performance of the network. ARD computes the input relevance by taking the sum of the squared weights of the *i*th input, divided by the total sum of squared weights. The resulting value indicates the relevance of each feature *relative to* the impact of the other features. The values are given in percent and ranged from 0.14% to 6.17% in this analysis.

It is important to note that the program computed ARD values only for a locally connected version of the network, in which the input features of separate domains (i.e., verb, adjective, versus syntax, etc.) were connected to separate parts of hidden layer. This hidden layer was fully connected to another hidden layer that, in turn, was fully connected to the output unit. The network had a 53-20-10-6 architecture.

The ARD values uncover some interesting trends. For example, high values were obtained for features that represent the general knowledge domain and the verb classification categories. Interestingly, all of the syntactic features had low ARD values.

In order to test the reliability of the ARD estimates, especially for the fully connected network, six features with ARD values below 1% were deleted from the training data set, and another fully connected network was trained. The network's parameters were identical to the first, except for the number of input and hidden units. The hidden layer was reduced to 12 units, with a resulting network architecture of 47-12-6. The network achieved an ASE of .167 on the training cases. On the test cases, the network had an ASE of .167. Only 51% of the test cases were classified correctly, which is 12% less than the network trained with all features but still well above chance level. Also, seven of the eight synonym cases were classified correctly by this network.

By using ARD and sensitivity analysis in combination, the relevance of the individual features can be identified, and the network can be fine-tuned. That is, one can approximate the minimal set of context features, on the basis of which the network can do the classification task.

DISCUSSION

The results of the network model support the claim that context contains information that is relevant to the meaning of abstract nouns. The 53 (47) features covered an important part of this context information. These results are an encouraging beginning. This network was trained with only 100 sentences per noun. Presumably, the performance of the network can be improved with a larger training corpus and with an extension and improvement of the feature set, motivated by additional theories in linguistics and cognition.

The findings for synonyms support the view that similarity of words' meanings is reflected in (or based on) their verbal contexts. However, the set of cases tested so far is too small to defend this claim more strongly. This finding needs to be replicated with a larger set of test cases. However, the present data are a promising first demonstration of meaning representation in context.

Regarding ARD, all of the features contributed to the network's performance to some extent. However, the results suggest that the information that is needed in order to distinguish among the abstract nouns is mostly covered by some of the semantic features (i.e., verbs and world knowledge domains). The dominant role of semantic features perhaps is not surprising, since the task is to distinguish different meanings. Also, semantic features mostly were selected for the task. The finding is consistent with the views of Landauer and Laham (1997) and of other researchers who have questioned the relevance of syntax in information processing, on the basis of the performance of systems such as LSA and HAL.

Within the semantic features, world knowledge domains and verb categories contained the features that received the highest relevance values. The ARD values should perhaps not be the basis for conclusions at this point, because the estimates tend to vary in different training sessions for similar networks. However, the results from an independently conducted discriminant analysis support the finding that these modules are most relevant.

DISCRIMINANT ANALYSIS: AN ALTERNATIVE?

The network assessed whether the abstract nouns could be distinguished on the basis of context features. As an alternative approach, a discriminant analysis could be run on the training data. Discriminant analyses assess

how well cases can be classified with respect to a grouping variable, on the basis of a set of predictor variables. In this case, a discriminant analysis would look at how well the 53 variables could predict which abstract noun a case represents. The abstract nouns would be different values in the grouping variable. A discriminant analysis of the 600 training cases was performed, and 88.2% of the cases were classified correctly. Note that this is below the classification rate of the network on these training cases after training, which was 98.5%. Evidently, the network does something beyond what discriminant analysis offers. This may be due to the computational advantages of networks. For example, in a network, different transfer functions can be used to modify the impact of the input values on the output.

It is difficult to assess, on the basis of the results from the discriminant analysis, how well the variables would help to classify new sentences. One way would be to use the loadings that were obtained from the analysis for the prediction of the new cases. However, such a procedure is not available in current statistical programs. The network offers an advantage in this respect. It would be interesting to see whether the prediction of the new cases by the loadings from the discriminant analysis would be better or worse than the generalization performance of the network.

The discriminant analysis provides an alternative approach to assessing the relevance of the features. Separate analyses were performed in which the abstract noun was predicted by all of the features, as opposed to only by features representing world knowledge, verb classification, or by a combination of the latter two. Recall that these two modules of features got the highest ARD estimates. The discriminant analysis correctly classified 44% of cases on the basis of verb features, 45.5% on the basis of world knowledge, and 63% on the basis of both. The latter value suggests that the two modules do most of the classification work, which is consistent with the high ARD values in these modules.

GENERAL DISCUSSION

This study used a neural network to test whether the context of abstract nouns contains sufficient information to distinguish among a set of different abstract nouns. The results suggest that it does. In particular, semantic context information is of importance, whereas the syntactic surface information that was included in the feature set did not appear to be crucial for the task.

The results in the synonym test show that the network, in fact, classifies the nouns according to their *meaning*. Because of the small number of test sentences, the extent of this generalization cannot be evaluated at this point.

The study demonstrates that a neural network is a useful tool for exploring the relevance of contextual features in a word's context to the word's meaning. The required manual analysis of contexts is tedious and not a perfectly reliable procedure. Obviously, this is a disadvantage, in comparison to other approaches.

As mentioned earlier, the general idea underlying this neural network is related to the systems HAL and LSA. They demonstrate that semantic information is *in the data*—in particular, in the context of words—and can be constructed from natural language. There are, of course, a lot of differences between the approaches. Most obvious is that it is difficult to test the present network on a larger corpus of training data and for more than a small set of nouns, because all the input features result from manual analysis. Instead of including a detailed comparison here, I outline only what I take to be a striking *conceptual* difference between the approaches.

In the discussed neural network, the input is a finite set of features that are abstract. Every sentence is evaluated with respect to all of the features. In HAL and LSA, the input is *raw data*—that is, a corpus of unanalyzed language. Both systems compute their semantic representations on the basis of the cooccurrences of words with other words or contexts. The network's results are interesting, because they give us an idea of what it is in the context that is related to a concept. In contrast, the dimensions in HAL or LSA are not grounded in a referential theory. The performance of HAL and LSA is impressive, because the systems are based on natural text units and because they demonstrate that concept knowledge can be acquired by the systems via comparatively simple computations on word-by-word or word-by-context matrices of cooccurrence. An obvious advantage to this procedure is that the systems work without information extraction by humans.

At present, it is uncertain whether a network can be trained to distinguish more than the 6 abstract nouns on the basis of the features used in this study. It is possible that this approach does not work for a set of, say, 60 abstract nouns rather than 6. Studies with a larger number of abstract nouns are being planned.

A network of this type may be used to address a number of problems in the future. First, it can be used to investigate synonym relationships. The closer two similar words are in meaning, the higher the likelihood that they would be classified as the same output node. Also, it would be informative to test whether a network can be trained to distinguish among a set of similar as opposed to a set of different abstract nouns.

Alternative classification systems can be compared with regard to their exhaustiveness and efficiency in the analysis process and to their effectiveness in providing relevant contextual features. For example, two verb classification systems—such as Schank's set of primitives (1972), as opposed to the categories suggested by Miller and Johnson-Laird (1976)—could be included as features in the context analysis. The feature values for a corpus of sentences that result from the two analyses could be used to train two alternative networks and the performances could be compared.

The network can simulate similarity ratings provided by human subjects to estimate the similarity of words. For example, the closest related word would be the first choice of the network (ideally, the word itself), the next

similar in meaning would get the second highest activation, and so forth. If the results matched human similarity ratings, this would suggest that humans use context information similar to that used in the network when judging similarity.

With respect to abstract nouns, the network shows that the contexts of abstract nouns contain information that is sufficient to distinguish among nouns of different meaning and that the relationship between abstract nouns and their contexts of use is strong. This finding supports the hypothesis that the meanings of abstract nouns are determined by their contexts. Such a view is tenable, because the network demonstrates that the context information would, in principle, be sufficient to enable the network both to distinguish between concepts and to identify concepts in the majority of cases.

REFERENCES

ANDERSON, R. C. (1990). Inferences about word meaning. In A. C. Graesser & G. H. Bower (Eds.), *Inferences and text comprehension* (pp. 1-16). San Diego: Academic Press.

ATCHISON, J. (1994). *Words in the mind*. Oxford: Blackwell.

BARSALOU, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, **10**, 82-93.

BARSALOU, L. W., & MEDIN, D. L. (1986). Concepts: Static definitions or context-dependent representations. *Cahiers de Psychologie Cognitive*, **6**, 187-202.

BLOOM, P. A., & FISCHLER, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition*, **8**, 631-642.

CARROLL, J. B., DAVIES, P., & RICHMAN, B. (1971). *Word frequency book*. Boston: Houghton Mifflin.

COTTRELL, W. G. (1989). Toward connectionist semantics. In Y. Wilks (Ed.), *Theoretical issues in natural language processing* (pp. 64-72). Hillsdale, NJ: Erlbaum.

EIZIRIK, L. M. R., BARBOSA, V. C., & MENDES, S. B. T. (1993). A Bayesian-network approach to lexical disambiguation. *Cognitive Science*, **17**, 257-283.

ELSHOUT-MOHR, M., & VAN DAALEN-KAPTEIJNS, M. (1987). Cognitive processes in learning word meanings. In M. G. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 53-72). Hillsdale, NJ: Erlbaum.

FILLMORE, C. J. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1-88). New York: Holt, Rinehart & Winston.

GALLANT, S. I. (1991) A practical approach for representing context and for performing word sense disambiguity using neural networks. *Neural Computation*, **3**, 293-309.

GOODMAN, P. (1996). *NevProp software, Version 3*. Reno: University of Nevada, Washoe Medical Center, Department of Internal Medicine.

GRAESSER, A. C., & CLARK, L. T. (1985). *Structures and procedures of implicit knowledge*. Norwood, NJ: Ablex.

GRAESSER, A. C., SWAMER, S. S., & HU, X. (1997). Quantitative discourse psychology. *Discourse Processes*, **23**, 229-263.

GROSS, D., & MILLER, K. J. (1990). Adjectives in WordNet. *International Journal of Lexicography*, **3**, 265-277.

HAMBERGER, M. J., FRIEDMAN, D., & ROSEN, J. (1996). Completion norms collected from younger and older adults for 198 sentence contexts. *Behavior Research Methods, Instruments, & Computers*, **28**, 102-108.

HOEFFNER, J. H. (1996). *Are rules a thing of the past? A single mechanism account of English past tense acquisition and processing*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.

KATZ, J. J., & FODOR, J. A. (1963). The structure of a semantic theory. *Language*, **39**, 170-210.

KINTSCH, W. (1988). The role of knowledge in discourse comprehension: A constructive integration model. *Psychological Review*, **95**, 163-182.

KINTSCH, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Press.

LAHAV, R. (1989). Against compositionality: The case of adjectives. *Philosophical Studies*, **57**, 261-279.

LANDAUER, T. K., & DUMAIS, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **104**, 211-240.

LANDAUER, T. K., & LAHAM, D. (1997, November). *Associative production by LSA: The knowledge in words and passages*. Paper presented at the 36th Annual Meeting of the Psychonomic Society, Philadelphia.

LUND, K., & BURGESS, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, **28**, 203-208.

MCCLELLAND, J. L., RUMELHART, D. E., & THE PDP RESEARCH GROUP (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2*. Psychological and biological models. Cambridge, MA: MIT Press.

MCKEOWN, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly*, **20**, 482-496.

MIKKULAINEN, R. (1996). Subsymbolic case-role analysis of sentences with embedded clauses. *Cognitive Science*, **20**, 47-73.

MILLER, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography*, **3**, 235-312.

MILLER, G. A. (1991). *The science of words*. New York: Scientific American Library.

MILLER, G. A. & CHARLES, W. G. (1991). Contextual constraints of semantic similarity. *Language & Cognitive Processes*, **6**, 1-28.

MILLER, G. A., & JOHNSON-LAIRD, P. N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.

NEAL, R. M. (1996). *Bayesian learning for neural networks*. New York: Springer-Verlag.

RUMELHART, D. E., & MCCLELLAND, J. L. (1986). On learning the past tenses of English Verbs. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 216-271). Cambridge, MA: MIT Press.

RUMELHART, D. E., MCCLELLAND, J. L., & THE PDP RESEARCH GROUP (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations*. Cambridge, MA: MIT Press.

SCHANK, R. C. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, **3**, 552-631.

SCHWANENFLUGEL, P. J. (1986). Completion norms for final words of sentences using a multiple production measure. *Behavior Research Methods, Instruments, & Computers*, **18**, 363-371.

SCHWANENFLUGEL, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meaning* (pp. 223-250). Hillsdale, NJ: Erlbaum.

SCHWANENFLUGEL, P. J., & SHOBEN, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 82-102.

STAHL, S. A. (1991). Beyond the instrumentalist hypothesis: Some relationships between word meanings and comprehension. In P. J. Schwanenflugel (Ed.), *The psychology of word meaning* (pp. 157-186). Hillsdale, NJ: Erlbaum.

STERNBERG, R., & POWELL, J. (1983). Comprehending verbal comprehension. *American Psychologist*, **38**, 878-893.

TAYLOR, W. L. (1953). "Cloze" procedure: A new tool for measuring readability. *Journalism Quarterly*, **30**, 415.

VENDLER, Z. (1967). *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.

WALTZ, D. L., & POLLACK, J. B. (1985). Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science*, **9**, 51-74.

WITTGENSTEIN, L. (1953). *Philosophical investigations*. Oxford: Blackwell.

(Manuscript received October 13, 1997;
revision accepted for publication January 6, 1998.)