Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society

Edited by Ashwin Ram and Kurt Eiselt

August 13 to 16, 1994
Georgia Institute of Technology

(1) Hastings & Lytinen
(2) Lampinen & Faries

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1994 Hillsdale, New Jersey Hove, UK

Objects, actions, nouns, and verbs

Peter M. Hastings
Artificial Intelligence Lab
1101 Beal Avenue
The University of Michigan
Ann Arbor, MI 48109
(313)763-9074
peter@umich.edu

Steven L. Lytinen
DePaul University
Department of Computer Science
243 South Wabash Avenue
Chicago, IL 60604-2302
(312)362-6106
lytinen@cs.depaul.edu

Abstract

This paper describes a lexical acquisition mechanism that was implemented in order to increase the robustness of a Natural Language Processing system. Although the mechanism was not intended to be a cognitive model of children's language acquisition, it demonstrates many similarities with psycholinguistic findings. In particular, the structure of the domain knowledge representation forces the system to take a bipolar approach to learning nouns and verbs. Psycholinguistic studies demonstrate differing treatment of nouns and verbs by children and suggest a structural basis for this difference. The knowledge-level similarities between our system and human linguistic knowledge make it possible to infer that children must adopt a similar strategy to effectively learn word meanings.

Introduction

We have developed an incremental approach to learning word meanings from their usage in example sentences. This approach has been developed and implemented as part of the LINK Natural Language Processing system (Lytinen & Roberts 1989). The learning system, called Camille (Contextual Acquisition Mechanism for Incremental Lexeme Learning), selects candidate word meanings from a collection of frames, stored in a concept hierarchy which is part of LINK's knowledge base.

The original motivation for building Camille was to improve the robustness of LINK. With the addition of the learning mechanism, LINK can process sentences which contain words that are not already defined in the system's knowledge base. Although this original motivation did not include the desire to develop a cognitive model of human language acquisition, it has allowed some insight into the process of human lexical acquisition.[1] In particular, the structure of LINK's knowledge base, along with the nature of the task of acquiring new word meanings, forces Camille to take a different approach for learning nouns than for learning verbs. This dichotomy closely parallels psychological findings about human acquisition of nouns and verbs (Gentner 1978; Goldin-Meadow, Seligman, & Gelman 1976; Behrend 1990). Thus, Camille suggests an explanation for the differences described in the psycholinguistic literature.

Unlike other acquisition systems (Zernik 1987; Granger 1977; Salveter 1979; Selfridge 1986), and in accord with psycholinguistic theories like Syntactic Bootstrapping (Gleitman 1990; Naigles 1990), Camille takes full advantage of LINK's grammatical knowledge in making its inferences. The acquisition procedure takes place during parsing and interacts with the syntactic decisions made by the parser.

This paper describes the nature of the knowledge that Camille uses to make its inferences, the basic process,[2] and the correspondences to developmental psycholinguistics. The last section contains an analysis of the types of inferences that are affected by the structure of the knowledge representation.

The nature of the knowledge

LINK's knowledge representation consists of an inheritance hierarchy of domain-independent and domain-specific concepts, or frames. Figure 1 shows some of LINK's domain-specific object concepts which were used during our participation in ARPA's Third and Fourth Message Understanding Competitions, or MUC-3 and 4 (Sundheim 1992) (the shading will be explained later). The MUC task involved the extraction of information from newspaper articles which described terrorist activities in Latin America; thus, the frames encode knowledge about the domain of terrorism.

The contents of the knowledge base are determined by the requirements of the domain and by the need to cover the input language. The structure is determined by the relative specificity of the individual frames. The most general frames are located near the root of the hierarchy and specific ones are located at the leaves.

Figure 2 shows some of the actions from the terrorism domain. Action concepts provide the relational structure that binds together the representation of the meaning of sentences. These concepts also constrain the types of arguments that can be attached as their slot-fillers (the constraints are also shown in figure 2). The Detonate node, for example, has the constraint that its semantic object must be a descendant of Bomb (or the node Bomb itself). Detonate inherits an additional constraint from its parent Nasty-Action. This constraint specifies that the actor must be a Human.

The nodes in LINK's concept hierarchy serve as its basic units of meaning. For Camille, then, learning the meaning of an unknown word reduces to finding the appropriate node in the hierarchy which best represents the word's meaning — a graph search problem. To drive the search, the semantic constraints, which are normally used to limit attachment of slot-fillers to the head verb, interact with the "evidence" provided

[1]General correspondences of the initial implementation of this system with psycholinguistic findings were described in (Hastings, Lytinen, & Lindsay 1991).

[2]The actual process that Camille uses is described more fully in (Hastings 1994).

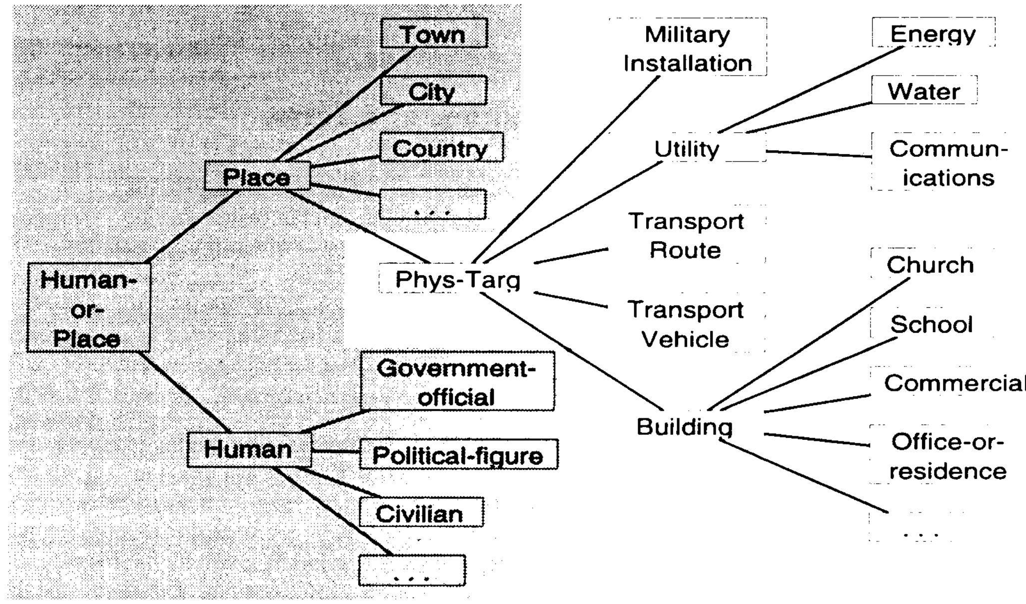

Figure 1: The pruned object tree

by example sentences. But the interaction works in different ways for different classes of words. Nouns (as the heads of noun phrases) normally serve as the slot-fillers of sentences and thus, as the items which are constrained. For example, in the sentence "Terrorists destroyed a flarge," the word "destroy" refers to the concept Destroy which has the constraint [Object = Phys-Targ]. When "flarge" is attached as the Object of the verb, the constraint places an *upper bound* on its interpretation as depicted in figure 1. The shaded-out nodes cannot be a valid interpretation of the meaning of "flarge".

For unknown verbs, however, the situation is quite different. Because they usually map to the actions in the domain, the verbs *apply* the constraints. Thus, the constraints place a *lower bound* on the interpretation of unknown verbs. The shaded areas of figure 2 show the concepts that are ruled out for an example sentence like "Terrorists froobled the headquarters." Possible meanings of "frooble" which are consistent with this sentence include any action which allows "headquarters" as its semantic object; thus, actions with constraints that are specific enough to exclude "headquarters" are not candidates for the meaning of "frooble."

Since input to Camille comes from the results of LINK's parses of example sentences from the domain, Camille does not receive negative examples as part of the learning process. Because of this, the learning problem is underconstrained for both nouns and verbs, in opposite ways. Example sentences provide an upper bound in the concept hierarchy on candidate hypotheses for nouns, but no lower bound, while the reverse is true for verbs. Because of this dichotomy, Camille must have different strategies for learning verbs and learning nouns. They can be stated most succinctly as follows:

- For nouns, choose the most general hypothesis which is consistent with the example sentence(s)
- For verbs, choose the most specific hypothesis which is consistent with the example(s)

In essence, because evidence only provides one kind of constraint (a lower or upper bound) for both nouns and verbs, Camille must form hypotheses for the meanings of unknown words which can be falsified by future examples. This means choosing the hypotheses which are closest to the bound provided by the evidence. More abstractly, then Camille's learning strategy could be stated as, "Always choose the concept(s) closest to the boundary." If this strategy were not pursued, future evidence would not allow Camille to further refine its hypotheses, thus resulting in overly general verb definitions and overly specific noun definitions.

It is important to note that the difference in information provided by example sentences for unknown nouns and verbs is is not just an artifact of LINK's knowledge representation structure. It is due to a fundamental principle of language. Because actions serve as the relational elements of sentence structure, they are the only logical place for the constraints to reside; thus contextual information about the meaning of a verb is intrinsically different from the information provided by context for unknown nouns. This difference is prescribed by the nature of the knowledge and it is consonant with psycholinguistic theories which will be described below. Without this two-part strategy, Camille could never learn those concepts closest to the edge because additional examples could never disconfirm the initial hypotheses. The next section briefly describes Camille's inference procedure.

Camille's inference procedure

As mentioned above, Camille takes a different approach to learning nouns than to learning verbs, but both start the same way. When Camille does not know the definition of a word, it enters a generic definition into the parse for various parts of speech. The combination of morphological and syntactic constraints is almost always sufficient to determine the correct lexical category for an unknown word. The rest of this section

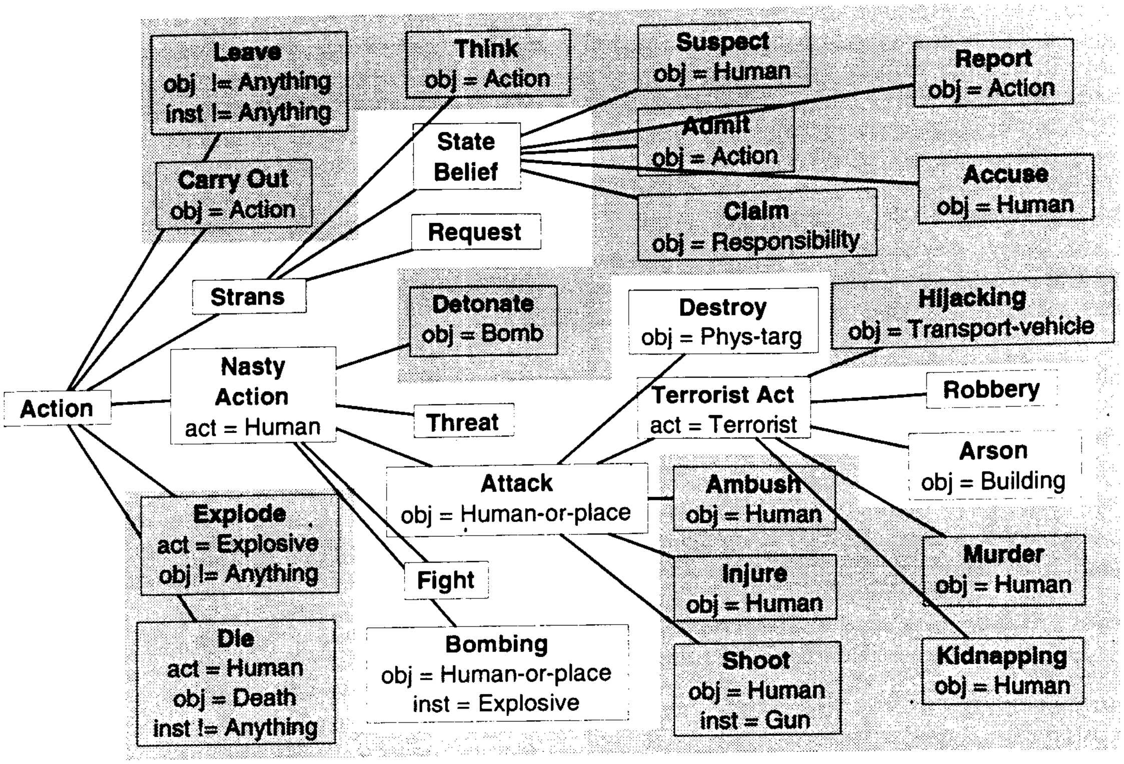

Figure 2: The pruned action tree

contrasts the learning of noun meanings and the learning of verb meanings.

Learning noun meanings

The process for learning nouns is fairly straightforward. When the noun is attached as an argument of a verb, any semantic constraints from that verb are applied to the noun. As mentioned above, this provides a single concept which is the root of a subtree that contains the valid interpretations for the word's meaning. Because the system wants to infer the most general noun meanings, this upper bound node is taken as the hypothesis for the word's meaning. Further examples can further specify the meaning.

As previously mentioned, given the sentence, "Terrorists destroyed a flarge," Camille can logically deduce that "flarge" must refer to some type of Phys-Targ. In practice, Camille infers that "flarge" means exactly the concept Phys-Targ. With a later sentence like, "Mary hijacked the flarge," Camille can further refine the hypothesis, inferring that "flarge" refers to Transport-Vehicle.

Learning verb meanings

For verbs, the process works the other way around. It is more difficult, however, because the system must search the action hierarchy for a node with appropriate constraints for the current slot fillers. This defines a lower bound on the set of possible meanings as described above. Unfortunately, the tree structure produces many concepts along the lower bound (8 nodes in figure 2). Camille ranks these hypotheses by computing the distance between their constraints and the actual slot filler in the example sentence. Camille keeps only the tightest fits, i.e. the concepts whose constraints most closely match the example. These concepts are the most specific valid hypotheses *and* the most falsifiable ones. Camille's learning method is incremental, so later examples of the word's use can either confirm the original hypothesis, or conflict with that hypothesis, triggering further search.

As an example of Camille's verb acquisition procedure, consider the sentence given above, "Terrorists froobled the headquarters." When LINK attaches "headquarters", which is represented by the Building concept, as the semantic object of the verb "frooble", the search for a meaning for "frooble" begins. As previously mentioned, the system can deductively rule out those concepts that were shaded out in figure 2. In order to *induce* a specific, falsifiable hypothesis, Camille prefers the concept Arson because its constraint that the object is a Building most closely matches the evidence.[3] The attachment of "Terrorists" as the actor of the verb is consistent with this hypothesis, so it is stored as the tentative definition for "frooble".

Given an additional example of the word's use like, "Joe froobled a pedestrian," Camille realizes that it has made an inappropriate guess because Arson's object constraint is violated by "pedestrian". Camille is incremental, so it takes the previous hypothesis as the starting point of the search for a better guess. The system climbs the tree (because any

[3]Because the object constraint for Arson is also the node Building, the distance between the constraint and the filler is 0. The object constraint for the Destroy node, on the other hand, is Phys-Targ which has a distance of 1 from Building.

more specific concept would also necessarily have constraints that were violated) and then tries to find another specific hypothesis that can take both a Civilian and a Building as its object. The node Robbery with its inherited object constraint Human-or-Place satisfies the search conditions and becomes the new hypothesis for the meaning of "frooble".

Empirical testing

The system was empirically evaluated on a set of 50 sentences chosen randomly from the MUC terrorism corpus.[4] The test sentences were quite long and complex, with an average length of 23 words, for example:

> Lopez Albujar, who left his post at the Ministry in May 1989, was riddled with bullets as he was getting out of his car in the Lima residential district of San Isidro.

To test Camille's learning algorithm, the definitions of all the verbs[5] in this set were removed from the lexicon. Then each of the sentences was processed by the system, and the learned definitions of these verbs were compared to the original definitions. Despite the complexity of the sentences and the low number of repetitions of each verb (average 2.7 occurrences), the system was able to infer the appropriate meanings for 8 of the 15 verbs, or 53%.[6]

Psychological connections

When psychologists first started studying the types of words that children acquired first, a striking observation was made. Children learn nouns well before and at a faster rate than they learn verbs. This prompted several studies of the differences in acquisition between the various types of words and theories about what causes those differences. These studies suggest that differences in the knowledge representation for nouns and verbs force children to use different mechanisms to learn them.

Gentner (1978, pp. 988-989) cites several studies that describe differences in acquisition between nouns and verbs. Some of these studies showed that young children's initial vocabulary consists entirely of nouns with verbs slowly making their way in. Others showed that the first verbs took almost twice as long to appear as the first nouns. A study of comprehension and production by Goldin-Meadow, Seligman, and Gelman (1976) showed two stages of early lexical development. In both stages, many more nouns than verbs were comprehended. Only a portion of the comprehended verbs were produced in the second stage, and none were produced in the first. Finally, Gentner described an additional study that demonstrated that the differences in acquisition are not just attributable to differences in the frequency of verbs versus nouns that the child hears. Even when presented with made-up nouns and verbs, and when balancing the presentation of these new words, children first used verbs an average of 8 months after starting to use nouns.[7]

In a slightly different vein, Behrend conducted in-depth studies of different types of verbs to compare children's comprehension and production among these various verbs (1990). The types of verbs that he studied were those that described actions (e.g. "squeeze", "pound"), results ("flatten", "break"), and instruments ("hammer"). He found that when labelling actions ("What is the person doing?"), children are more likely to use an instrument verb than an action verb. This seems strange for two reasons. First, because children use relatively little evidence and quickly infer mappings from words to meanings, they should make the best possible guess about the meaning of the word given what they know. All other things being equal, this should correspond to the type of verb that occurs most frequently in the language. But instrument verbs are far less frequent than action verbs are. Second, instrument verbs carry more information than action verbs and are therefore more specific. Thus the children in the experiments were labeling the events with the *most specific* label possible. This contradicts the results found in acquisition of nouns, which demonstrate that "specific subordinate terms are used much less frequently than basic-level terms as labels for familiar objects." (Behrend 1990, p. 694).

These results point to a fundamental difference between children's acquisition of nouns and verbs, one that can be most readily explained by differences in the underlying mental representation of the different concepts. Gentner calls the basis for this difference the "referential / relational" distinction. Nouns normally refer to objects or "thinglike elements." Objects tend to be highly constrained by the physical world. Hence, similar objects share almost all the same attributes. On the other hand, verbs tend to express relationships between objects or changes in those relationships. Relationships are more abstract and more difficult to describe with a set of attributes. The representation that Gentner espouses to represent these varying concepts is a semantic net, in which meanings are built up compositionally by referring to more basic elements of meaning.

Gentner's representation focuses on the representation of single nodes of meaning. Behrend supports Huttenlocher and Lui's proposal (1979) that these differences in behavior are caused by the overall structure of the representation. They suggest that objects are organized in a structured hierarchy so that nearby elements share many of the same features. The relational elements that are expressed by verbs are represented in a matrix structure with nodes connecting across the various object hierarchies. Graesser, Hopkinson, and Schmid (Graesser, Hopkinson, & Schmid 1987) have recently done

[4]The test set was kept relatively small to simulate a sparse-input learning task. The assumption is that most of the lexical definitions have been entered as part of the knowledge engineering of the system. Words that were overlooked in this process are not likely to be encountered frequently by the system, so the testing set contains a small number of examples of each word.

[5]The empirical testing focused on learning verb meanings because they were more difficult to acquire.

[6]The system was also tested in another domain in which the sentences were very simple (average length: 4). Because the parser produced many more correct parses, the system scored much better, 71%.

[7]These results are tempered somewhat by recent suggestions that the findings may be specific to the English language. Gopnik and Choi (1990), in a study of the correlation between linguistic and cognitive development, cite studies that Korean- and Japanese-speaking children show a higher use of verbs during the one-word stage than English-speaking children do. They attribute this difference in behavior to structural differences in the languages. Fernald and Morikawa (1993) found evidence that maternal speech to the children was responsible for the increased use of verbs by Japanese-speaking children.

experimental testing to support this hypothesis. The subjects were asked to sort sets of words by similarity. The findings suggested that people tend to sort nouns hierarchically while verbs were less structured and more "cross-classified." The cross-classification of Camille's concepts is discussed in the next section.

Camille's implications for learning

The first important point to make is that Camille's knowledge representation is not just an artifact of the particular system. The IS-A inheritance hierarchy is used widely in Artificial Intelligence, and for good reason. The structure facilitates generalization, an ability that is key to learning and reasoning. It also provides efficient storage of information. Furthermore, various psychological studies support the existence of hierarchical structures in the brain ((Kaplan, Weaver, & French 1990) and (Keil 1991), for example).

The use of a hierarchy to represent objects also seems intuitively correct. For representing actions and relations, the utility of this structure is not so obvious. As previously pointed out, although some psycholinguistic researchers have postulated a hierarchical scheme for their representation, recently the focus has turned to more "matrix-like" schemes. But the latter approach seems to conflict with the observation about the nature of constraints provided by the input and whether an upper bound or lower bound is created on the set of possible meanings for an unknown word. It is just not clear what "lower bound" would mean in a matrix-type organization.

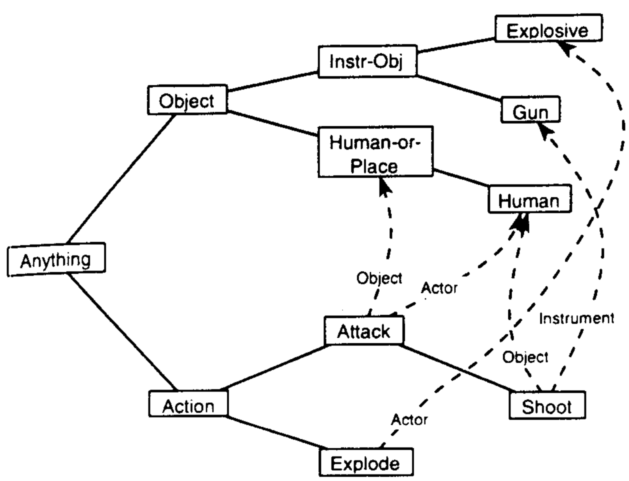

Figure 3: Matrix-like organization of action concepts

On close inspection, it appears that Camille's representation has the best of both worlds. If the slot-filler constraints are displayed graphically (see figure 3 where the solid lines represent paths in the IS-A hierarchy and the dashed lines represent constraints on the actions), it is apparent that the connections do (as Huttenlocher and Lui put it) cut across the various parts of the hierarchy. This leads to the question, "Does it make sense to have the additional structure imposed by enforcing a hierarchical structure on *actions*?" The answer appears to be yes, for the same reasons given for object representation above. The hierarchy has representational strength — it allows for efficient storage of the attributes and constraints of actions.

The question that remains then is, "What does this imply about human concept organization and the learning process?" For one thing, it lends support to the idea that there can be multiple organization structures within the brain. There are clear advantages to having both types of concept representation. In addition, it suggests that learning could proceed in one of two ways. If a child realizes that her idea of what a word means is wrong, she should look for concepts that are closely related in the hierarchy. If the child's hypothesized *constraints* for the word are wrong, she should change those constraints based on the structure of the hierarchy that is selected by the matrix links.

Conclusion

The structure of the concept knowledge that is used for inferring word meaning from context places constraints on how the learning process can proceed. Some form of these constraints should also apply to human lexical acquisition. The behavior of the computer model can therefore suggest explanations for psycholinguistic findings about differing treatment of nouns and verbs. It can also suggest other avenues of exploration to allow us to find out more about how children learn language.

Although Camille sheds some light on the difference in nature of learning nouns and verbs, in several ways the model is overly simple when compared to human lexical acquisition. First, the learning strategy as described here assumes that an unknown word has only one meaning. We have done preliminary work on extending the learning algorithm for ambiguous words. As described in (Hastings & Lytinen in press), the acquisition of ambigous words also has implications for the differing treatment of nouns and verbs. If the two-part strategy described here is not followed, the system cannot determine which words should have multiple definitions.

Second, the Camille model assumes the pre-existence of concepts which correspond to the meanings of new words. This is an overly simplistic assumption. It may be that learning a new word sometimes requires the addition of a new node in the concept hierarchy as well. Another extension of Camille (Hastings 1994) addresses concept acquisition within the context of a lexical acquisition system. In accordance with the results described herein, object concepts were found to be much easier to learn than action concepts.

Another limitation of the Camille system was brought to light by testing on larger data sets. Given more examples of the unknown words, the incremental system would be expected to increase the quality of its inferences. Instead the percentage of correct inferences decreased. This was due primarily to noise created by incorrect partial parses. As previously mentioned, the sentences in the terrorism domain were quite complex, and often the LINK parser could only return a partial parse tree. If the fragments included incorrect role assignments (e.g. "... Michigan froobled ..." instead of "... terrorists from Michigan froobled ... "), Camille misinterpreted this as justification for changing its previous hypothesis. Thus, the parser has a strong effect on the lexical acquisition process which must be further studied.

Even with these additional considerations, however, the essential differences between the tasks of learning nouns and verbs still remain in our model, and are intrinsic to the lexical acquisition task. This is due to the differences in usage of nouns and verbs. Thus, it seems inevitable that a computer system, or a human learning, must adopt different strategies when learning these two types of words.

References

Behrend, D. 1990. The development of verb concepts: Children's use of verbs to label familiar and novel events. *Child Development* 61:681–696.

Fernald, A., and Morikawa, H. 1993. Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development* 64:637–656.

Gentner, D. 1978. On relational meaning: The acquisition of verb meaning. *Child Development* 49:988–998.

Gleitman, L. 1990. The structural sources of verb meaning. *Language Acquisition* I(1):3–55.

Goldin-Meadow, S.; Seligman, M.; and Gelman, R. 1976. Language in the two-year-old. *Cognition* 4:189–202.

Gopnik, A., and Choi, S. 1990. Do linguistic differences lead to cognitive differences? a cross-linguistic study of semantic and cognitive development. *First Language* 10:199–215.

Graesser, A.; Hopkinson, P.; and Schmid, C. 1987. Differences in interconcept organization between nouns and verbs. *Journal of Memory and Language* 26:242–253.

Granger, R. 1977. Foul-up: A program that figures out meanings of words from context. In *Proceedings of Fifth International Joint Conference on Artificial Intelligence.*

Hastings, P., and Lytinen, S. in press. The ups and downs of lexical acquisition. In *Proceedings of the* 12^{th} *National Conference on Artificial Intelligence.*

Hastings, P.; Lytinen, S.; and Lindsay, R. 1991. Learning words: Computers and kids. In Hammond, K., and Gentner, D., eds., *Proceedings of the 13th Annual Conference of the Cognitive Science Society*, 251–256. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hastings, P. 1994. *Automatic Acquisition of Word Meaning from Context.* Ph.D. Dissertation, University of Michigan, Ann Arbor, MI.

Huttenlocher, J., and Lui, F. 1979. The semantic organization of some simple nouns and verbs. *Journal of verbal learning and verbal behavior* 18:141–162.

Kaplan, S.; Weaver, M.; and French, R. 1990. *Active symbols and internal models: Towards a cognitive connectionism.* London: Springer Verlag. Springer Series on Artificial Intelligence and Society.

Keil, F. 1991. Theories, concepts, and the acquisition of word meaning. In Byrnes, J. P., and Gelman, S. A., eds., *Perspectives on language and thought: Interrelations in development.* Cambridge: Cambridge University Press.

Lytinen, S., and Roberts, S. 1989. Unifying linguistic knowledge. AI Laboratory, Univ of Michigan, Ann Arbor, MI 48109.

Naigles, L. 1990. Children use syntax to learn verb meanings. *Journal of Child Language* 17:357–374.

Salveter, S. 1979. Inferring conceptual graphs. *Cognitive Science* 3:141–166.

Selfridge, M. 1986. A computer model of child language learning. *Artificial Intelligence* 29:171–216.

Sundheim, B. 1992. Overview of the fourth message understanding evaluation and conference. In *Proceedings of the Fourth Message Understanding Conference.* San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Zernik, U. 1987. Strategies in language acquisitions: Learning phrases from examples in context. Technical Report UCLA-AI-87-1, UCLA.

Levels of Semantic Constraint and Learning Novel Words

James M. Lampinen
Department of Psychology
Northwestern University
Evanston, Illinois,60201
jim@swim-two-birds.psych.nwu.edu

Jeremiah M. Faries
Department of Psychology and the Institute for the Learning Sciences
Northwestern University
Evanston, Illinois,60201
jerry@swim-two-birds.psych.nwu.edu

Abstract

A common method of teaching vocabulary involves presenting students with new words in context and having the students derive the meaning of these words based on contextual cues. Beck, McKeown and McCaslin (1983) have argued that the contexts used to teach new words should be highly constraining. Although highly constraining contexts avoid ambiguity they do not present the learner with the necessity of combining contextual and word specific information and thus practicing skills needed for general comprehension. We suggest that a superior method of teaching is to relax the amount of contextual constraint because to optimize the learning from the presentation of a sentence the student must use both top down and bottom up processes to discover the meaning of the sentence, thus integrating two sources of knowledge about the word. The present research compares knowledge and use of newly learned words between students who learned the new words using three encounters with highly constraining contexts, three encounters with moderately constraining contexts or three progressively less constraining contexts. Students were given definitional and comprehension tests both immediately after study and at a one week delay. The results suggest that repeated encounters with moderately constraining contexts are superior to repeated encounters with highly constraining contexts.

When people learn new words they typically do so by combining information about the way the word is used in the context of the utterance with information from their pre-established knowledge of the word. In fact, most words are learned contextually and incidentally (e.g., Nagy and Herman, 1987). Moreover, comprehension of word meanings almost always occurs within a context and arguably the processes used to understand general contextual meaning contribute greatly to the activation of what we think of as word meaning. Identification of words is easier and faster when preceded by a relevant context (Schuberth and Eimas, 1977) suggesting that contextual information is an important component for processing individual words. Learning words in the absence of these general comprehension processes (e.g., presenting word-definition pairs to study) may produce an inferior record of word meaning insofar as that word can be used and comprehended quickly and naturally (although such learning may facilitate assimilation for words when encountered in context). Researchers, therefore, who are interested in directly teaching word meanings have explored ways to use context as a teaching method (Kolich, 1991).

Debate about the merits of contextual versus definitional methods (see Stahl and Fairbanks, 1986 for an excellent review of this literature), has often amounted to comparing optimal strategies within one broad instructional method with suboptimal strategies within another broad instructional method (see Hall, 1988). Since any successful approach to vocabulary instruction is likely to include both contextual and definitional components (Stahl and Fairbanks, 1986), it may be more constructive to investigate the boundary conditions that determine the effectiveness of both contextual and definitional methods, rather than to pit one against the other.

The goal of our present research is to determine the set of conditions under which contextual methods of instruction are likely to be most effective. One important variable in the construction of study sentences is how tightly the sentence constrains the meaning of a target word. If one were to replace the target word with a blank line (as in a cloze test) and the meaning intended by the blank line were easily inferred, the sentence would be highly constraining. It has been suggested that contexts used in teaching new words should, in fact, be highly constraining (Beck, McKeown and McCaslin, 1983). They suggest that the contexts which students use to learn new words should quite narrowly limit the number of possible definitions of the new word. A supporting argument for the use of highly constraining contexts is that they make word meaning clear and unambiguous and thus easier to derive. On the other hand, Schank (1982) has argued that students who study a word in a highly constraining context do not have to pay any attention at all to the new word since the meaning of the utterance is clear without knowing the word. If they make any association between the derived meaning and the new word it is incidental and not required by the task. In a study of eye fixation patterns, Ehrlich and Rayner (1981) found that target words from highly constrained passages were not fixated on or were fixated on more briefly than the same words in a poorly constrained passage. Further, the situations in which knowing a word's meaning is most crucial are precisely those situations in which the word's meaning is least likely to be deduced from context. This suggests that moderately constraining contexts might be superior to highly constraining contexts because it is only in moderately constraining contexts that the reader needs to integrate contextual information with prior knowledge about the word.

The activation of word meaning in everyday use of language typically involves combining top down contextual information with bottom up word specific perceptual information. Vocabulary instruction which uses only highly constraining contexts emphasizes only the use of top down contextual information. Taylor and Taylor (1983) argue that good readers are superior to poor readers because they both use context more efficiently and because they use word specific information more efficiently. Interestingly, they also suggest that good readers do not use context by simply guessing what the target word should be (as might be encouraged by high constraint contexts) but, rather, they "narrow... down alternatives" (as might be encouraged by moderate and low constraint contexts).

A model of word learning should include learning of how the word fits into a variety of contexts as well as the learning of word specific associations. Consider for instance, how word meaning is accessed in an interactive reading model like the one proposed by Adams(1990). Adams proposes three separate processing modules all feeding into a central meaning processor. Phonological, orthographic and context processors all combine to activate word meaning.[1] Accessed meaning then becomes part of the the current text representation available in the context processor. Stahl(1991) has proposed that in learning word meanings from context, the meaning produced by this interaction of contextual and word specific information then becomes linked to the orthographic and phonological representations of the word.[2] This account seems plausible as far as it goes and argues for learning words in a context. However, for the association to be made between the contextual and word specific information it is important that such an association be required as part of the learning task. The strength of the link must depend crucially on how strongly the word is attended to and how important the word's meaning is to the overall meaning of the utterance. As we've seen above, high constraint contexts require readers to attend only minimally to the target word because the context itself is heavily redundant with the meaning of the word. When readers encounter words in moderately constraining contexts, the story is different. In these contexts, the word's meaning can only be derived by combining information from the context processor with information from the orthographic and/or phonological processor. Thus for readers in this situation, the word itself is crucial to the understanding of the passage.

Ideally, to provide practice using the skills they will eventually need, students must encounter words in sentences of varying degrees of contextual constraint and the goal of these encounters should be to encourage students to combine contextual information with word specific information to quickly access word meaning and utterance meaning. At the limit students should even be able to access word meaning when the contextual constraint is weak and the word itself contributes greatly to the intent of the utterance.

Pilot research from our lab suggested that readers learned new words most efficiently when they were provided with study sentences in which the amount of constraint provided was progressively decreased. We reasoned that words to be learned in context should be seen in moderate and low constraint contexts in addition to high constraint contexts and that the presentations of these mixed contexts should progress from high constraint sentences to low constraint sentences. Progressively decreasing contextual constraint, encourages students to integrate new contextual information with information garnered from previous encounters with the word, but it also allows students to encounter the new word in ever more difficult contexts without fear of reaching an impasse, and is thus similar to the scaffolding provided by intelligent tutoring systems.

In the pilot study, subjects were presented with spaced presentations with three example sentences for each of 40 words. These example sentences were either highly constraining, or they progressed from highly constraining to minimally constraining. We found that subjects who studied the new words using the progressively less constraining sentences learned the new words better when measured on a host of definitional and comprehension measures. One difference between subjects in the high constraint and decreasing constraint group was that subjects in the decreasing constraint group encountered at least one moderately constraining sentence. If most of the advantage of the mixed constraint condition comes from requiring subjects to combine contextual information with knowledge gained in previous encounters with the word, it is possible that moderate constraint was the source of the effect of variation. The present experiment was designed to test this possibility. One group of subjects was taught 40 low frequency English words by presenting them with spaced encounters with the words in progressively less constraining contexts. Another group of subjects learned the 40 words

[1] Adams' model is discussed here as an illustrative example. The crucial point is that many models of reading propose that word meaning activation involves both the orthographic and phonological features of the word and expectations generated by the context.

[2] Word meaning is represented in Adams' model in a semantic network. Thus a word's meaning is to be understood in terms of the word's relationship to other concepts. A particular contextual instantiation of a word is likely to pick out only a few of those semantic relations. Moderately constraining contexts require that the reader combine the semantic information garnered in previous encounters with the word with the new semantic information garnered in the current context. Thus moderately constraining contexts encourage the development of a rich integrated conceptual understanding of the new word.

relying exclusively on encounters with the words in three different high constraint contexts. A third group of subjects learned the 40 words relying exclusively on encounters with the words in three different moderate constraint contexts. Beck et al's (1983) notion of a pedagogical context would suggest that students using the highly constraining contexts would learn the new words better. The argument developed above would suggest that students who encounter the words in moderate constraint or in progressively less constraining contexts would do better.

Method

Subjects

Subjects were 85 Northwestern undergraduates who received course credit for their participation.

Materials

Word lists: Two lists of twenty low frequency English words were compiled. These words were rated as unfamiliar by 5 independent subjects. The words were divided into two lists of 20 words so that one list could serve as an interference list for the other list.

Study sentences: For each word five highly constraining sentences were generated, five moderately constraining sentences were generated, and one minimally constraining sentence was generated. An example of the sentences used to teach the word *insalubrious* were:

High: John ate a healthful diet, but Karl's was insalubrious.

Moderate: Researchers believe that there is nothing insalubrious about an occasional glass of wine.

Low: It is not clear from what I've read whether it is insalubrious or not.

The average length in words of sentences in each condition was equated (High: 13.30 (3.68); Moderate: 13.35, (3.36); Low: 13.58, (3.32); $F<1$). Five subjects were presented with these sentence lists and were asked to rank order them based on how likely they would be to use the sentences if they wanted someone to guess the meaning of the target word. Based on these subjects' rankings we selected three different high constraint, three different moderate constraint and one low constraint sentences as study sentences for each word. Subjects in the high constraint condition were presented with three different high constraint sentences. Subjects in the moderate constraint condition were presented with three different moderate constraint sentences. Subjects in the mixed constraint condition were presented with a randomly chosen high constraint sentence, a randomly chosen moderate constraint sentence and a low constraint sentence.

Dependent measures: Subjects were tested at immediate and one week retention intervals.

i) Sentence verification: Subjects read 80 sentences, half of which made sense given the meaning of the target word and half of the sentences did not make sense given the meaning of the target word. The 80 sentences presented at the immediate test were different from the 80 sentences presented at one week delay. The sentences were presented in a different randomized order for each subject. Subjects were instructed to press the "z" key on their keyboard for sentences that made sense and the "/" key for sentences that didn't make sense. Subjects were instructed to respond as quickly as they could while still being accurate.

ii) Definition test: The definition test asked subjects to provide definitions for the 40 words studied. The immediate and delayed tests differed only in that the words were presented in different randomized orders.

Procedure

Subjects were run individually or in small groups. Each student provided us with their score on the verbal portion of the Scholastic Aptitude Test which served as our measure of prior verbal ability. They were seated in front of a computer terminal and were told that they would be learning 40 vocabulary terms and that they would do so by learning the words from context. The student read a sentence and was asked to define the target word based on that sentence. After the student typed in his/her response, the correct definition was presented on the screen for 2 seconds and the subject preceded to the next sentence. Immediately after the students were done with the study phase, they took the tests described above. Subjects returned one week later and retook the alternate versions of the tests.

Results

The results for the immediate accuracy data are shown in figures 1 and 2. As can be seen, they support the claim that medium levels of constraint are superior to high levels of constraint, and this appears to be especially true for high ability students. Any advantage of using progressively decreasing levels of constraint appears to follow from the inclusion of medium constraint sentences in the study sentences. We now describe the results according to the kind of test used.

Immediate testing

Sentence Verification: Accuracy in sentence verification reflects the degree of usable knowledge the subjects had acquired from the study sessions. For YES sentences the only significant predictor was verbal SAT, $F(1,79)=9.775$, $p<.002$. For NO sentences, level of constraint, $F(2,79)=3.425$, $p<.04$, and the interaction between self-rated verbal SAT and level of constraint, $F(2,79)=3.628$, $p<.04$, were both significant predictors of accuracy. Analysis of the response time data rules out the possibility that these results are an artifact of a speed/accuracy trade off because there was no effect of condition on response time, $F<1$. Planned comparisons showed that students who studied the medium constraint sentences, $F(1,54)=6.57$, $p<.02$, and mixed constraint sentences, $F(1,51)=4.11$, $p<.05$, were better at rejecting inappropriate usage than students who studied only high constraint sentences. The advantage of the medium and mixed constraint conditions was greater for higher ability students than it was for lower ability students as demonstrated by the significant interaction between level of constraint and ability for both the high/medium, $F(1,54)=7.04$, $p<.01$, and high/mixed, $F(1,51)=4.11$, $p<.05$, comparisons. There were no significant differences between the medium and mixed constraint conditions. Overall, these results indicate that subjects who studied using either the medium or mixed constraint study sentences were better able to reject inappropriate word usage than were subjects who received only high constraint sentences.

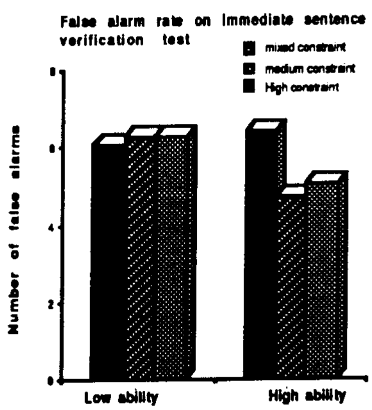

Figure 1: False alarm on immediate sentence verification test

Definitions: This measure indicates how accurately students can provide a definition for words they have learned. Accuracy on the definition test was significantly predicted by level of constraint, $F(2,79)=3.364$, $p<.04$, by verbal SAT, $F(1,79)=6.338$, $p<.02$, and by the interaction between verbal SAT and level of constraint, $F(2,79)=3.855$, $p<.03$. As with the sentence verification test, planned comparisons indicated that subjects in the medium, $F(1,54)=4.505$, $p<.04$, and mixed constraint, $F(1,51)=4.959$, $p<.03$, conditions outperformed those in the high constraint condition and the effect of level of constraint was amplified for high ability level students for both the high/medium, $F(1,54)=5.67$, $p<.02$ and high/mixed, $F(1,51)=5.50$, $p<.03$, comparisons. There were no significant differences between the medium and mixed conditions. So, as with verbal comprehension as measured by sentence verification accuracy, students who learned new words with medium constraint or mixed constraint example sentences were better able to provide definitions than students who learned new words using high constraint sentences.

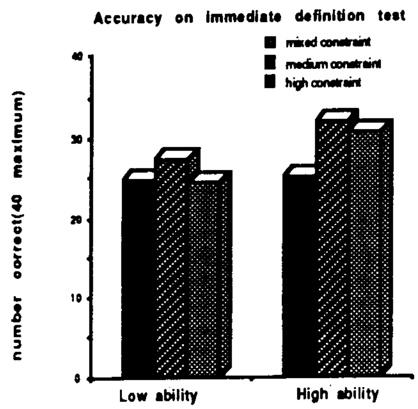

Figure 2: Accuracy on immediate definition test.

Delayed testing

Students returned to be tested again after a one week interval. Again, the results showed an advantage of medium and mixed constraint conditions over the high constraint condition. As before, we present our results by type of test.

Sentence verification: Planned comparisons again revealed a significant advantage of the medium constraint over the high constraint condition on correct NO responses, $F(1,49)=5.51$, $p<.03$ and a significant interaction between constraint level and ability, $F(1,49)=5.36$, $p<.03$. There was also a marginally significant advantage of medium over mixed constraint, $F(1,50)=3.18$, $p<.09$, for correct NO responses on the delayed sentence verification test. The interaction between ability level and level of constraint was also marginally significant, $F(1,50)=3.55$, $p<.07$ and inspection reveals that this is due to poorer performance of low ability students in the mixed constraint condition. There was no difference between subjects in the high constraint and mixed constraint conditions. Thus, while medium levels of constraint improve the ability of students to correctly reject incorrect word usages even at a weeks delay, mixed levels of constraint appear to have a deleterious effect on the ability to reject incorrect usage at a weeks delay.

Definitions: Planned comparisons indicated the same basic pattern of results on the delayed test as on the immediate test. Students who studied using medium constraint sentences outperformed students using high constraint sentences, $F(1,49)=7.22$, $p<.01$, and this advantage was greater for high ability students, $F(1,49)=8.23$, $p<.01$. There was also a marginally significant advantage for the mixed constraint condition over the high constraint condition, $F(1,49)=3.23$, $p<.08$, and a marginally significant interaction between level of constraint and ability, $F(1,49)=3.868$, $p<.06$.

Discussion

Beck et. al.(1983) have argued that students should be taught new words using highly constraining contexts. We suggested above that students would do better when they learn words in contexts which require them to integrate information from the current context with information from previous encounters with the word. To test this hypothesis, we taught college students 40 low frequency English words using either three encounters with the words in high constraint contexts, three encounters with words in moderate constraint contexts or using three encounters with the words in progressively less constraining contexts. We have shown that students who study new words in moderate constraint contexts develop a better ability to make use of these new words as measured by both definitional and sentence verification tasks than do students who study words in highly constraining contexts.

The results of this experiment are consistent with the view that teaching new words with moderate constraint contexts is superior to teaching new words in high constraint contexts. Moderate constraint contexts are superior because students are required to combine top down contextual information with bottom up word specific perceptual information as they do during normal reading. Many models of word recognition and meaning activation (e.g., Adams, 1990) hold that word meaning is activated in precisely this way. Our results suggest that learning words in a fashion that uses both sources of information facilitates subsequent use of the acquired meaning. The word knowledge acquired when both sources of information are attended to is more accessible when there is a need to combine contextual information with word specific information, namely, most reading or listening situations.

Our results also suggest that the optimal level of constraint to use in example sentences is contingent upon the ability level of the students. High ability level students appear to show the largest gains from encountering words in moderately constraining contexts. We can rule out prior knowledge of the particular word's meaning because performance prior knowledge would also have elevated scores in the high constraint condition. Van Daalen-Kapteihns and Elshout-Mohr(1981) suggested that lower ability level students developed holistic models of word meaning based on minimal contextual encounters. Since these models are holistic they are difficult to modify in light of new information. Since learning word meaning from less constraining contexts requires the acceptance of a provisional model of word meaning followed by updating that model as new information becomes available, lower ability level students may be less able to take advantage of moderate constraint contexts since their models are less modifiable. Nevertheless, it should be noted that at all ability levels, moderately constraining sentences are at least as efficacious as highly constraining sentences.

The present study also suggests that our finding of advantages for progressively less constraining contexts over high constraint contexts may have been due primarily to the presence of a moderately constraining sentence in the progressively less constraining study list. This is especially clear in the delayed accuracy results where subjects in the moderate constraint condition continue to outperform subjects in the high constraint condition, but where subjects in the mixed constraint condition no longer significantly outperform those in high constraint condition. It is possible that the decreasing constraint sentences we used made jumps which were too large to be effective. The final example sentence, for instance, may have been too sparse for a third encounter. On the other hand, one potential benefit of decreasing constraint may be in situations where the definition is not provided. We are currently comparing these methods under conditions in which definitions are not provided as part of the study procedure and this may show an advantage for progressively decreasing constraint.

Teaching vocabulary is an important part of learning a language for the first time or as a second language. Word meaning knowledge is an important component of general comprehension. Indeed vocabulary knowledge is highly correlated with general comprehension abilities(Anderson and Freebody, 1981). Further, it is reasonable to assume that incidental acquisition of vocabulary is dependent upon current word knowledge, such that learning new words should increase the ability of students to learn other new words in the future. Beck et. al.(1982) have found such incidental gains in students who have participated in their vocabulary building curriculum.

Beck et. al.(1983) advised educators that "not all contexts are created equal." They argued that the contexts used to teach new words should be contexts which highly constrain the meanings of the new words. While this may be good advice for the elementary school teachers they were addressing, our research suggests that less constraining contexts may be superior for older students. Students need to learn to quickly access and use word meaning when reading or listening to normal paced speech. To do that, they need to strengthen the same kinds of general comprehension processes during study that they will use during reading. Our research suggests that there is some advantage to be gained from providing study sentences which are either moderately constraining or which vary in constraint. Using moderately constraining or progressively less constraining contexts results in a better integration of multiple encounters with novel words and consequently, more usable word knowledge.

Acknowledgements

The authors would like to thank Jim Hall, Mitsumi Imai and Laura Namy for their comments on earlier drafts of this manuscript. Thank-you also to Julie Leffelman and Linda Song for their help running the experiment and entering the data.

References

Adams, M.J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge MA: MIT Press.

Anderson, R. and Freebody, P.(1981). Vocabulary knowledge. In J. Guthrie(Ed). *Comprehension and Teaching: Research Reviews.* 77-117, Newark DE: W.W. Norton.

Beck,I.L., McKeown, M.G., and McCaslin, E.S. (1983). Vocabulary development: All contexts are not created equal. Elementary School Journal, **83,** 177-181.

Beck, I.L., Perfetti, C.A. and McKeown, M.G.(1982) Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, **74,** 506-521.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, **22,** 75-87.

Hall ,J.W. (1988) On the utility of the keyword mnemonic for vocabulary learning. *Journal of Experimental Psychology*, **80,** 554-562.

Kolich, E.M.(1991). Effects of computer-assisted vocabulary training on word knowledge. *Journal of Educational Research*, **84,** 177-182.

Nagy W. and Herman, P. (1987) Breadth and depth of vocabulary knowledge. In M.G. McKeown and M.E. Curtis(Eds).*The Nature of Vocabulary Acquisition.* Lawrence Erlbaum, Hillsdale NJ pp. 19-35.

Schank, R.C.(1982). *Reading and Understanding: Teaching from the Perspective of Artificial Intelligence.* Lawrence Erlbaum, Hillsdale, NJ.

Schuberth, R. E., & Eimas, P. D.(1977). Effects of context on the classification of words and nonwords.*Journal of Experimental Psychology:Human Perception and Performance*, **3,** 27-36.

Stahl, S.A.(1991). Beyond the instrumentatlist hypothesis: Some relationships between word meanings and comprehension. In P.J. Schwanenflugel (Ed). *The Psychology of Word Meanings.* Lawrence Erlbaum, Hillsdale, NJ.

Stahl, S.A. and Fairbanks, M.(1986)The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, **56,** 72-110.

Taylor, I. and Taylor, M. M., (1983) *The Psychology of Reading.* Academic Press, NY.

Van Daalen-Kapteijns and Elshout-Mohr(1981). The acquisition of word meanings as a cognitive learning process. *Journal of Verbal Learning and Verbal Behavior*, **20,** 386-399.