DOCUMENT RESUME

ED 280 005 CS 008 733

AUTHOR Frederiksen, John R.
TITLE The Development of Computer-Based Instructional Systems for Training Essential Components of Reading. Final Report.
INSTITUTION BBN Labs, Inc., Cambridge, MA.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
REPORT NO BBN-6465
PUB DATE Mar 87
CONTRACT 400-81-0030; ONR-N00014-83-C-0639
NOTE 119p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Bilingual Students; *Computer Assisted Instruction; *Reading Ability; *Reading Comprehension; Reading Difficulties; *Reading Instruction; Reading Processes; Reading Research; Reading Skills; Secondary Education; Skill Development

ABSTRACT

A series of studies conducted to examine deficits in reading comprehension skills of low ability, young adult readers and to evaluate computer-based systems that have been designed to improve skills deficiencies in comprehension are described in this report. The studies described in the first section of the report: (1) examined readers' use of semantic entailments (such as the action "murdered" entailing an agent case, "the killer," and a patient case, "the victim") in drawing inferences from text; (2) investigated readers' use of relational terms such as causal and adversative connectives in gaining an understanding of higher order semantic relations among clauses and sentences within a text; and (3) focused on the problem of analyzing referential relations within a text. This section concludes with a discussion of the findings of the studies, which suggest that the subjects, high school students, need to develop efficient techniques for mapping referents. Studies described in the second section address the transfer of skill developed in using context for accessing concepts to the performance of high level comprehension tasks, and the use of component-based training for improving reading skills of low ability readers whose first language is not the language of instruction. The findings reported in this section suggest that bilingual students can benefit from computer-based training focusing on the development of automatic skills for both decoding and encoding orthographic information. (FL)

BBN Laboratories Incorporated

A Subsidiary of Bolt Beranek and Newman Inc.

Report No. 6465

Final Report on the Development of Computer-based Instructional Systems for Training Essential Components of Reading

John R. Frederiksen

March 1987

Prepared for:
Office of Naval Research

Approved for public release; distribution unlimited

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE:

3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S): Report No. 6465

5. MONITORING ORGANIZATION REPORT NUMBER(S):

6a. NAME OF PERFORMING ORGANIZATION: BBN Laboratories

6b. OFFICE SYMBOL *(If applicable)*:

7a. NAME OF MONITORING ORGANIZATION:

6c. ADDRESS *(City, State, and ZIP Code)*:
10 Moulton Street
Cambridge, MA 02238

7b. ADDRESS *(City, State, and ZIP Code)*:

8a. NAME OF FUNDING / SPONSORING ORGANIZATION: Office of Naval Research

8b. OFFICE SYMBOL *(If applicable)*:

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-83-C-0639

8c. ADDRESS *(City, State, and ZIP Code)*:
Personnel and Training Research Program
800 North Quincy Street
Arlington, VA 22217

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
			NR-154-519

11. TITLE *(Include Security Classification)*:
Final Report on the Development of Computer-base Instructional Systems for Training Essential Components of Reading

12. PERSONAL AUTHOR(S): John R. Frederiksen

13a. TYPE OF REPORT: Final Report

13b. TIME COVERED: FROM 7/83 TO 2/87

14. DATE OF REPORT *(Year, Month, Day)*: March 1987

15. PAGE COUNT: 115

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*:
Training technologies, cognitive skill acquisition, reading Skills, reading comprehension, text comprehension, computer-based instruction

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This final report describes (1) studies we have conducted of deficits in reading comprehension skills of low ability, young adult readers, and (2) evaluation studies of computer-based training systems which have been designed to improve essential components of reading shown in prior work to be sources of skill deficiency in such readers.

Three experimental studies were carried out investigating reader deficiencies in the inferential processing of texts. The first study (Rosebery, 1986) examined readers' use of semantic entailments (such as the action "murdered" entailing an agent case ("the killer") and a patient case ("the victim")) in drawing inferences from text. The second study (Warren, 1986) investigated readers' use of relational terms such as causal and adversative connectives (examples are, respectively, "as a result" and "although") in gaining an understanding of high order semantic relations among (continued)

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: Dr. Michael Shafto

22b. TELEPHONE *(Include Area Code)*: (202) 696-4596

22c. OFFICE SYMBOL: ONR1142PT

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

clauses/sentences within a text. The third study focused on the problem of analyzing referential relations within a text. Texts were constructed containing one or more antecedent noun phrases and one or more anaphoric words (a pronoun or lexical substitute). Over the set of texts, we varied the number of antecedent noun phrases, the topical status of the pronoun's referent, referential continuity, the ambiguity of the semantic context of the pronoun, and the syntactic agreement among antecedents and the pronoun. The subject's task was to read each text and supply the correct referent for a pronoun when requested. The results showed that the problem of tracing referential relations represents a substantial source of comprehension difficulty. Subjects generally had difficulty in supplying referents for pronouns when the needed referent is not a topic, and when alternative antecedents (or possible referents) are available. They had the greatest difficulty when the semantic context of the referring term (the pronoun) was written to be ambiguous. While subjects had difficulty in mapping pronouns to their referents, they had no difficulty in mapping lexical substitutes (that is, synonyms, near synonyms, or superordinates) to their referents. These results suggest that subjects need to develop efficient techniques for mapping referents, and, more particularly, to learn to base this search on the activation of concepts in semantic memory based upon the context in which the referring pronoun occurs.

The evaluation studies address two issues: (1) the transfer of skill developed in using context in accessing concepts to the performance of high level comprehension tasks, particularly, the tracing of anaphoric relations; and (2) the use of component-based training for improving reading skills of low ability readers whose first language is not the language of instruction (here, English).

In the first training study, we examined the way in which developing subjects' skill in using contextual information to access word meanings would aid them in performing a high level comprehension task. In this transfer task, subjects were required to provide referents for pronouns that occurred near the end of a two and or three sentence text. We predicted on the basis of a theory of reference mapping that improvements in a subjects' general ability to activate concepts on the basis of their semantic context should lead to an enhancement in their ability to map referents for pronouns, when the topical and semantic context of the pronoun unambiguously allows the selection of a particular referent. Conversely, such training should not lead to improved performance when the context of the pronoun is made ambiguous or when there is a conflict between semantic and topical constraints. In the training study, subjects who were poor to average readers were given general training in the use of semantic context to gain access to concepts, using a new training game called "Defender". The results of the experiment confirmed these predictions and thus supported the theory of skill transfer used to generate them.

The second training study employed bilingual hispanic subjects, who were trained (1) in perceptual skills for encoding English orthographic units, and (2) in skills for decoding English words, using game-like computer training environments focusing on each of these skill components. The subjects were given pre- and post-tests of the perceptual and decoding skills, using both English and Spanish test materials. The skills of the bilingual subjects improved greatly as a result of training. These improvements, in both perceptual encoding and word decoding skills, were as large as those of monolingual English subjects who were given similar training in an earlier study. This finding supports the use of component-based training of reading skills in subjects whose first language is not the language of training. For the bilingual subjects, the gains in performance on the perceptual task were as large for a Spanish as they were for an English version of the task. This suggests that the skills developed in the perceptual training game are of a general nature rather than language specific. In contrast, measured skill gains following training in decoding were specific to the language of training (English), indicating that these skills are linguistic, rather than more general information-processing capabilities.

Final Report on the Development of Computer-based Instructional Systems for Training Essential Components of Reading

John R. Frederiksen

BBN Laboratories

FINAL REPORT

Contract No. N00014-83-C-0639

Task Authority NR-154-519

Personnel and Training Research Programs

Psychological Sciences Division, Office of Naval Research

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

This final report describes (1) studies we have conducted of deficits in reading comprehension skills of low ability, young adult readers, and (2) evaluation studies of computer-based training systems which have been designed to improve essential components of reading shown in prior work to be sources of skill deficiency in such readers.

Three experimental studies were carried out investigating reader deficiencies in the inferential processing of texts. The first study (Rosebery, 1986) examined readers' use of semantic entailments (such as the action "murdered" entailing an agent case ("the killer") and a patient case ("the victim")) in drawing inferences from text. The second study (Warren, 1986) investigated readers' use of relational terms such as causal and adversative connectives (examples are, respectively, "as a result" and "although") in gaining an understanding of high order semantic relations among clauses/sentences within a text. The third study focused on the problem of analyzing referential relations within a text. Texts were constructed containing one or more antecedent noun phrases and one or more anaphoric words (a pronoun or lexical substitute). Over the set of texts, we varied the number of antecedent noun phrases, the topical status of the pronoun's referent, referential continuity, the ambiguity of the semantic context of the pronoun, and the syntactic agreement among antecedents and the pronoun. The subject's task was to read each text and supply the correct referent for a pronoun when requested. The results showed that the problem of tracing referential relations represents a substantial source of comprehension difficulty. Subjects generally had difficulty in supplying referents for pronouns when the needed referent is not a topic, and when alternative antecedents (or possible referents) are available. They had the greatest difficulty when the semantic context of the referring term (the pronoun) was written to be ambiguous. While subjects had difficulty in mapping pronouns to their referents, they had no difficulty in mapping lexical substitutes (that is, synonyms, near synonyms, or superordinates) to their referents. These results suggest that subjects need to develop efficient techniques for mapping referents, and, more particularly, to learn to base this search on the activation of concepts in semantic memory based upon the context in which the referring pronoun occurs.

The evaluation studies address two issues. (1) the transfer of skill developed in using context in accessing concepts to the performance of high level comprehension tasks, particularly, the tracing of anaphoric relations, and (2) the use of component-based training for improving reading skills of low ability readers whose first language is not the language of instruction (here, English).

In the first training study, we examined the way in which developing subjects' skill in using contextual information to access word meanings would aid them in performing a high level comprehension task. In this transfer task, subjects were required to provide referents for pronouns that occurred near the end of a two and or three sentence text. We predicted on the basis of a theory of reference mapping that improvements in a subjects' general ability to activate concepts on the basis of their semantic context should lead to an enhancement in their ability to map referents for pronouns, when the topical and semantic context of the pronoun unambiguously allows the selection of a particular referent. Conversely, such training should not lead to improved performance when the context of the pronoun is made ambiguous or when there is a conflict between semantic and topical constraints. In the training study, subjects who were poor to average readers were given general training in the use of semantic context to gain access to concepts, using a new training game called "Defender". The results of the experiment confirmed these predictions and thus supported the theory of skill transfer used to generate them.

The second training study employed bilingual hispanic subjects, who were trained (1) in perceptual skills for encoding English orthographic units, and (2) in skills for decoding English words, using game-like computer training environments focusing on each of these skill components. The subjects were given pre- and post-tests of the perceptual and decoding skills, using both English and Spanish test materials. The skills of the bilingual subjects improved greatly as a result of training. These improvements, in both perceptual encoding and word decoding skills, were as large as those of monolingual English subjects who were given similar training in an earlier study. This finding supports the use of component-based training of reading skills in subjects whose first language is not the language of training. For the bilingual subjects, the gains in performance on the perceptual task were as large for a Spanish as they were for an English version of the task. This suggests that the skills developed in the perceptual training game are of a general nature rather than language specific. In contrast, measured skill gains following training in decoding were specific to the language of training (English), indicating that these skills are linguistic, rather than more general information-processing capabilities.

1. STUDIES OF SKILL DEFICIENCIES IN COMPREHENSION

Three experimental studies were carried out investigating reader deficiencies in the inferential processing of texts. As in our earlier work (Frederiksen, 1981a), we sought to identify in these studies sources of skill deficiency among young adult readers that, with training, could have a broad enabling effect on the performance of other critical reading skills.

The first study (Rosebery, 1986) examined readers' use of semantic entailments (such as the action "murdered" entailing an agent case ("the killer") and a patient case ("the victim")) in drawing inferences from text. The second study (Warren, 1986) investigated readers' use of relational terms such as causal and adversative connectives (examples are, respectively, "as a result" and "although") in gaining an understanding of high order semantic relations among clauses/sentences within a text. As each of these studies has been reported elsewhere, they will be summarized in the present report.

The third study examined reader difficulties in mapping continuities of reference within a text. An emphasis in this research was on identifying efficient, potentially automatic methods for analyzing referential relations within a text, methods which enable expert readers to identify referents effortlessly and without conscious attention. Our hypothesis was that these methods for reference mapping involve the use of information derived from the context of the anaphoric term along with (when possible) semantic information derived from the anaphoric term itself in the activation of likely antecedent concepts in semantic memory. According to the theory, when by such means a concept is accessed and is found to be already in a state of activation due to its prior occurrence within the text, the concept can be recognized as the required referent without an attention-demanding search of the reader's model of the prior text. In the study, which we shall report here, we develop evidence of the use of such methods for tracing referential relations in a group of high school age subjects representing a wide range of reading skill levels.

1.1 Sensitivity to Semantic Entailments of Verbs and Adjectives

In order to comprehend, readers must be able to analyze, integrate and make inferences about the information present in text. Recent research suggests that the success of these processes is determined by the interaction of two sources of information: (1) factors in the text that affect its readability and (2) the skill with which the reader performs the various processes that comprise "reading". The purpose of this research was to investigate the way in which two text-based factors, word relationships and surface syntactic structure, interact with reader skill to affect readers' ability to analyze the semantic relationships present in text and to make inferences based on those semantic analyses. The influence of word relationships was assessed by manipulating the degree of semantic entailment between two words in a passage. Entailing words are those that are thought to semantically obligate the presence of an associated case word (e.g., the action "murdered" obligate the presence of an agent case word "the killer" and a patient case word "the victim"; the action "died" does not obligate these case words). The influence of syntactic structure was assessed by manipulating the syntactic class (verb/adjective) in which an entailing word appeared in a passage.

The results demonstrate that the ability of all readers to infer action-case relationships was significantly improved by the presence of entailing words. However, reader skill interacted with entailment to produce significant differences in passage comprehensions. First, skilled readers were more efficient at analyzing the semantic relationships present in text than were less skilled readers. Second, skilled readers were better able than less skilled readers to use other semantic information to enable inferences when entailing words were absent. Finally, less skilled readers appeared to depend more on explicit text-based factors like entailing/verb structures to enable semantic analysis than skilled readers. Theoretical implications of the findings are discussed in the report (Rosebery, 1986), focusing in particular on the interactions that occur between word relationships and reader skill during text comprehension. Implications for instruction in vocabulary and comprehension are discussed.

1.2 Understanding High Order Semantic Relations

An essential element in text comprehension is the reader's ability to integrate newly encountered propositions with those previously encoded, into a coherent model of text meaning. Two bases for propositional integration may be: (a) the network of abstract semantic relations, represented linguistically through conjunctions, verbs and other connectives, and (b) the semantic content of related propositions, represented in, for example, argument repetition, collocation and semantic entailment. The purpose of this research was to develop an understanding of the nature of expertise in using information from these sources in integrating semantically related propositions, and to identify sources of comprehension difficulty for the less skilled reader.

The influence of two types of relations (causal and adversative) and two types of connectives (conjunctions and verbs) on readers' comprehension and on-line integration of related propositions was examined. Contextual influences were examined by comparing readers' comprehension of strongly constrained relations with those that were weakly constrained.

Thirty-two high school students of varying levels of reading ability read 96 passages (64 semantically consistent passages, 32 semantically anomalous). Subjects read each passage, clause by clause (presented via microcomputer), and, at the time they read the final clause, indicated by a key press whether the passage was semantically consistent or anomalous. Accuracy on this task and reading times for each clause were recorded.

The results showed that comprehension and on-line integration of semantic relations were strongly influenced by contextual constraints and by readers' skill in using such constraints. The type of relation also exerted an influence. Adversative relations were more difficult than causal relations. Comprehension of both types of relations also appeared to depend upon reader skill. Readers differed as well in skill in using connectives. Skilled readers were unaffected by connective type, whereas less skilled readers' comprehension appeared to depend upon the marking of a relation by a conjunction. Finally, while skilled readers were equally accurate in comprehending consistent and anomalous passages, less skilled readers were highly inaccurate in judging anomalous passages, even though they took longer to read anomalous passages than consistent passages. Implications for theories of text comprehension are discussed in the report (Warren, 1986), as are implications for instruction.

1.3 Mapping Referential Relations that Occur within a Text

The comprehension of a text as a cohesive series of interrelated propositions requires, at a minimum, that continuity of reference be successfully established by the reader. Thus, in reading successive sentences, pointers must be constructed from anaphoric expressions to their antecedents in the reader's text model. Anaphoric words include pronouns, general words (such as person, thing, affair, event), subordinate terms (such as task, plant), and lexical items that are alternative expressions for an earlier occurring concept (such as model, referring to what might previously have been termed a theory). Examples of these types of anaphora, used for reiterating an earlier idea, are provided by Halliday and Hasan (1976, p. 279) in the following:

S_1: I turned to the ascent of the peak.

	The ascent	(repeated item)	
	The climb	(synonym)	
S_2:	The task	(superordinate)	is perfectly easy.
	The thing	(general word)	
	It	(pronoun)	

In an earlier experiment (Frederiksen, 1981b), we studied one form of reference, pronomial reference, and demonstrated that good and poor readers differ in their ability to locate efficiently the proper referent for a pronoun. Low and middle ability readers showed significantly longer reading times for sentences containing a pronoun compared with reading times for identical sentences in which the pronoun was replaced by its referent (lexical repetition). High ability readers, on the other hand, showed only small (and nonsignificant) differences in reading times for these two conditions. The earlier experiment went on to establish that: (1) when they encounter a pronoun, readers consider multiple antecedents that satisfy syntactic/grammatical constraints (e.g., gender, number) associated with the pronouns; (2) they select a referent noun phrase from among those antecedents on the basis of disambiguating semantic constraints of the sentence frame containing the pronoun; and (3) the topical status of an antecedent has a strong influence on its priority in making such semantic evaluations.

It seems clear that, when reading texts containing pronouns or other forms of anaphora, little conscious attention is devoted to searching one's prior text model for antecedents that represent potential referents, and that this is particularly true when the context of the pronoun clearly points to a particular antecedent (that is, it is unambiguous). In our earlier study, subjects' difficulty in mapping referents for pronouns (as measured by reading times) was at its greatest when the context of the pronoun did not point clearly to a single antecedent. Under that condition, subjects also deliberated for a significantly longer time when they were asked to supply the referent for the pronoun. Thus, there was in the experiment a strong association between the presence/absence of clear contextual clues concerning the identity of the referred-to concept and the degree of effort required by subjects in mapping the referential relation to that concept. Use of contextual information constraining the identity of a pronoun's referent appears to greatly reduce the processing demands of the reference-tracing problem, even for short texts of 2 or 3 sentences.

The purpose of the present research is to investigate further the conditions under which automatic processes may allow an efficient, effortless retrieval of referents for pronouns, and to construct a theory of such reference tracing. The experiment focused on three hypothesized processes that may enable a reader to trace more efficiently referential relationships from anaphora back to their antecedents.

1. Grammatical filters. The first concerns the possibility of grammatical filters, which allow a reader automatically to exclude from consideration those antecedents that do not agree with an anaphoric term in gender and/or number. If the process of reference assignment involves a search of the reader's text model, then this search will examine specific instantiations of concepts within the text, and these instantiated concepts should include gender and number information. The hypothesized search process therefore assumes that these concept instantiations within the text model can be rapidly inspected for general properties such as gender and number, and selectively retrieved on the basis of gender/number specifications. Evidence supporting the use of grammatical filters thus would suggest that reference mapping

involves a search of the text model and selective reinstatement of antecedents within that model, rather than a search of semantic memory, which contains generic concepts (not concept instantiations).

2. Semantic memory search. The second hypothesized process makes use of the relationship or mapping between instantiations of concepts in the text model and concepts in semantic memory. It assumes that concepts within the text model are not fully specified there, but rather that such concept specification relies on pointers to an underlying semantic memory. Concept instances within the text model are sets of such pointers to semantic memory. This implies that reference tracing could be mediated in either of two ways: by a search of the text model, or by a search of semantic memory (Collins & Loftus, 1975). The search of semantic memory is presumed to be semantically driven. Thus, frame-based semantic constraints (semantic information concerning the identity of the antecedent that can be derived from the sentence frame surrounding the anaphoric term, such as, for example, from the case frame associated with the verb; cf. Bruce, 1987) and semantic information derived from the anaphoric word itself (which is readily available when the referring term -- the anaphoric word -- is lexically cohesive with the antecedent to which it refers; i.e., it is a synonym, near synonym, or superordinate word) both provide information capable of supporting such a search process. Search of semantic memory is here regarded as an automatic process demanding few attentional resources (cf., Schneider & Fisk, 1982). The process of search is terminated when a concept is encountered with a pointer to the text model (and which is therefore in a prior state of activation).

3. Topical status. The third hypothesized process concerns the use of information about the current topic of the text. Pronouns typically refer to a current discourse topic, which is also typically the subject of the sentence preceding the pronoun (Gruber, Beardsley, and Caramazza, 1978; Lesgold, Roth, and Curtis, 1979). The topic of a discourse may shift at various points in a discourse (Grimes, 1975). Information concerning the current topic may be carried in its height within the text model (Kintsch & van Digk, 1978; Vipond, 1980) or in patterns of activation within semantic memory, or it may be maintained in a separate list (Kieras, 1981). In any case, the search of semantic memory for a referent could proceed in parallel from currently salient or topicalized antecedents at the same time that a semantically-based search

is undertaken. When either or both processes generate concepts with a state of activation exceeding a threshold, the search is terminated and the resulting antecedent taken as the referent.

The emphasis in this study is on the identification of efficient, and perhaps even automatic, methods for analyzing referential relations. Each of these hypothesized processes may be viewed as a potential contributor to efficient and automatic reference assignment, and provides a possible mechanism contributing to the automatic mapping of pronouns to referents. To summarize, the following questions were addressed in the experiment:

1. In searching their model of prior text, can readers "filter out" from consideration (reinstatement) antecedents that do not agree with a pronoun in gender and/or number?

2. Can semantic memory be used to facilitate the analysis of referential relations when the referring term is semantically related (e.g., is a near synonym, or a superordinate term) to the referred-to antecedent, or when semantic constraints are available associated with the context of a pronoun?

3. Can the topical structure of a text be used to gain rapid access to referents when they are current topics of a discourse?

1.3.1 Method and Subjects

The experimental task used in this study is a reference analysis task similar to that used in earlier work (Frederiksen, 1981b). In this task, subjects read two or three sentence paragraphs, presented on the screen of an IBM PC. When they have finished reading the first sentence, they press the "+" key to advance to the second sentence, and so forth. Their reading times are measured for each sentence and transformed into reading times per syllable for purposes of later analysis. When the "+" key is pressed following the last sentence of a paragraph, an underscore appears beneath a pronoun in that sentence. When this probe occurs, the subject's task is to vocally report the referent for that pronoun as rapidly and accurately as possible. The subject's response latency (from the presentation of the probe to the onset of vocalization) is recorded by the computer, and his or her response is typed into the computer at that time by the experimenter. Since the texts presented vary in length from two to three sentences and their length is unknown to the subject, for all sentences but the first, the subject has uncertainty as to whether or not he or she will be probed for a pronoun's referent. Subjects thus must read very carefully and pay close attention to the mapping of pronominal referents as they read each sentence in order to be able to respond quickly and accurately when the probes occur. In our earlier work (Frederiksen, 1981b), we have found that under these loading conditions, subjects' reading times per syllable when reading sentences containing pronouns are a sensitive index of their time to analyze referents for pronouns. In addition, subjects' errors and latencies in reporting referents provide another source of information concerning difficulty in mapping referential relations under a particular set of textual conditions.

The experimental texts used in the experiment were constructed from a corpus of 65 sentence sets that were explicitly written for this purpose. Each set is composed of 12 sentences of the types shown in Table 1 (a total of 780 sentences were thus prepared). The first sentences of the experimental texts are drawn from the first four sentence types, which vary in (1) the number of antecedent noun phrases (one or two) present, which represent potential referents for a subsequently occurring pronoun; (2) the number of those antecedent noun phrases that are syntactically compatible (i.e., agree in gender and number) with the subsequent pronoun; and (3) the topicality of the noun phrase that is the referent for a forthcoming pronoun (i.e., does the referent occur in the subject or predicate position). The second sentence of each experimental text is drawn from sentence types five through eleven. Each of these contains an anaphoric word which is either (1) a repetition of an antecedent noun phrase occurring in the first sentence of a text, (2) a pronoun which is either the subject or in the predicate referring to one or more noun phrases in the first sentence, or (3) a word that is lexically related to an antecedent in sentence one (a synonym, near synonym, or superordinate). The third sentence of the experimental texts (taken from sentence type twelve), when present, always contains a pronoun referring to an antecedent noun phrase which occurred in the first sentence. Two sample sentence sets illustrating each of the twelve sentence types are given in Appendix A.

For each of the 65 sentence sets, thirteen text forms were constructed following the models shown in Table 2. Examples of the thirteen texts resulting from the application of these models are given in Appendix B for the two sentence sets of

Table 1

Sentence Sets Used in the
Reference Analysis Experiment

Sentence One Alternatives	Sentence Two Alternatives	Sentence Three
1. A_1________.	5. Pr(A_1)_____.	12. Pr(A_1)_____.
2. A_1___A_2_____.	6. A_1________.	
3. A_2___A_1_____.	7. Pr(A_1 or A_2) **	
4. A_1____A_3____.*	8. Syn(A_1)____.**	
	9. A_1________.**	
	10. ______Pr(A_1).	
	11. A_2________.	

* A_1 and A_3 differ in gender/number, while A_1 or A_2 agree in gender and number.

** The sentence frame containing the anaphoric word is ambiguous, allowing reference to either A_1 or A_2 within sentence 1.

Note: Sentence groups (1-4), (5, 6, and 10), and (7-9) are each based upon a common sentence and represent transformations of that common base sentence.

Table 2

Texts Assembled from the Sentence Sets*

Text Form	Sentence One	Sentence Two	Sentence Three
1	(2) A_1___A_2___.	(5) $Pr(A_1)$_____.	--
2	(3) A_2___A_1___.	(6) A_1__________.	(12) $Pr(A_1)$__.
3	(3) A_2___A_1___.	(5) $Pr(A_1)$______.	--
4	(1) A_1________.	(5) $Pr(A_1)$______.	--
5	(2) A_1___A_2___.	(7) $Pr(A_1$ or $A_2)$__.**	--
6	(3) A_2___A_1___.	(10) _______$Pr(A_1)$.	--
7	(2) A_1___A_2___.	(11) A_2________.	(12) $Pr(A_1)$__.
8	(4) A_1___A_3___.	(5) $Pr(A_1)$______.	--
9	(3) A_2___A_1___.	(9) A_1__________.**	(12) $Pr(A_1)$__.
10	(3) A_2___A_1___.	(8) $Syn(A_1)$______.**	(12) $Pr(A_1)$__.
11	(2) A_1___A_2___.	(5) $Pr(A_1)$______.	(12) $Pr(A_1)$__.
12	(3) A_2___A_1___.	(7) $Pr(A_1$ or $A_2)$__.**	(12) $Pr(A_1)$__.
13	(2) A_1___A_2___.	(10) _______$Pr(A_1)$.	(12) $Pr(A_1)$__.

* The number in parentheses are the sentence numbers.

**The sentence frame is ambiguous, allowing reference to either A_1 or A_2.

Appendix A. For example, text form twelve begins with a sentence containing two antecedent noun phrases that agree syntactically with one another, it includes a second sentence containing a pronoun that, in its sentential context, could refer to either of the two antecedents, and it concludes with a sentence containing a pronoun that refers to the non-topicalized antecedent of sentence one. In this case, by looking at subjects' reading times and errors in supplying referents for the pronoun in sentence three, one can make inferences about the assignment of a referent for the ambiguous pronoun in sentence two. (If the referent assigned to the pronoun in sentence two is A_2, then subjects will make more errors and require greater reading times in assigning the correct referent for the pronoun in sentence three, which is A_1.) By comparing performance measures for contrasting text forms, a number of specific hypotheses concerning sources of difficulty in understanding text reference can be studied. A total of eleven such comparisons were planned, and these are shown in Table 3.

Each subject received 65 two and three sentence texts, derived by assigning a particular text form to each sentence set in the corpus. Thus, each subject was given five exemplars of each text form over the course of the experiment. The assignment of text forms to particular sentence sets was partially counterbalanced within the experiment in the following way: Five versions of the experimental materials were created, each of which differed in the assignment of sentence sets to particular text forms. The 65 sentence sets were divided into five successive blocks of 13 sentences, and within each block the 13 text forms were assigned in one of five random orders. Subjects were then assigned randomly to each of these five versions. This means that "subject variance" in the analyses of variance to be carried out will actually reflect both individual differences among subjects and variability due to textual materials. The statistical tests will therefore be conservative in that they reflect the generalizability of an effect with respect to both the subject and text populations.

Subjects. A total of 22 high school students served as subjects in this experiment. They ranged from the 3rd to the 99th percentile on the Gates-MacGinitie Reading Test. For purposes of analysis, the subjects were divided into three groups on the basis of their test scores: the "high" reading ability group included subjects between the 99th and the 86th percentiles, the "middle" group included subjects between the 85th and the 58th percentiles, and the "low" group included subjects

Table 3
Planned Comparisons Among Text Forms
(Percentage of Correct Referents and Average Reading Time per Syllable are Given in Parentheses)

Contrast	Text Form	Sentence One	Sentence Two	Sentence Three
1. Pronominal Ref. vs. Repeated Antecedent	2	A_2—A_1—.	A_1——. (256 msec)	Pr(A_1)—.
	3	A_2—A_1—.	Pr(A_1)—. (293 msec)	--
2. Number of Antecedents	1	A_1—A_2—.	Pr(A_1)—. (82%, 261 msec)	--
	4	A_1——.	Pr(A_1)—. (93%, 259 msec)	--
3. Grammatical Filtering of Antecedents	1	A_1—A_2—.	Pr(A_1)—. (82%, 261 msec)	--
	8	A_1—A_3—.*	Pr(A_1)—. (86%, 302 msec)	--
4. Topicality of Referent	1	A_1—A_2—.	Pr(A_1)—. (82%, 261 msec)	--
	3	A_2—A_1—.	Pr(A_1)—. (73%, 293 msec)	--
5. Proximity of Topic	2	A_2—A_1—.	A_1——.	Pr(A_1)—. (90%, 235 msec)
	7	A_1—A_2—.	A_2——.	Pr(A_1)—. (70%, 264 msec)

Table 3
(continued)

Contrast	Text Form	Sentence One	Sentence Two	Sentence Three
6. Continuity of Reference	7	A_1—A_2—.	A_2——.	$Pr(A_1)$—. (70%, 234 msec)
	11	A_1—A_2—.	$Pr(A_1)$—.	$Pr(A_1)$—. (81%, 241 msec)
	13	A_1—A_2—.	—$Pr(A_1)$.	$Pr(A_1)$—. (83%, 261 msec)
7. Topicalizing within an Intervening Sentence	2	A_2—A_1—.	A_1——.	$Pr(A_1)$—. (90%, 235 msec)
	3	A_2—A_1—.	$Pr(A_1)$—. (73%, 293 msec)	
8. Ambiguity of Reference	1	A_1—A_2—.	$Pr(A_1)$—. (82%, 261 msec)	--
	5	A_1—A_2—.	$Pr(A_1$ or $A_2)$—.** (94%, 327 msec)	--
9. Default Referent Assignment	7	A_1—A_2—.	A_2——.	$Pr(A_1)$—. (70%, 264 msec)
	12	A_2—A_1—.	$Pr(A_1$ or $A_2)$—.**	$Pr(A_1)$—. (57%, 249 msec)

Table 3
(continued)

Contrast	Text Form	Sentence One	Sentence Two	Sentence Three
10. Semantic Memory Activation	9	A_2—A_1—.	A_1——.**	$Pr(A_1)$—. (83%, 283 msec)
	10	A_2—A_1—.	$Syn(A_1)$—.**	$Pr(A_1)$—. (76%, 278 msec)
	12	A_2—A_1—.	$Pr(A_1$ or $A_2)$—.**	$Pr(A_1)$—. (57%, 249 msec)
11. Frame-Based Semantic Constraints	3	A_2—A_1—.	$Pr(A_1)$—. (73%, 293 msec)	--
	6	A_2—A_1—.	—$Pr(A_1)$. (66%, 320 msec)	--

* Antecedents A_1 and A_2 agree in gender and number with the critical pronoun, while A_3 does not.

** The sentence frame is ambiguous, allowing reference to either A_1 or A_2.

between the 52nd and the 3rd percentiles.[1]

1.3.2 Results

Data analyses focused on a set of contrasts among textual conditions designed to allow us to study different possible sources of difficulty readers may experience in analyzing referential relations within a text. These include:

1. An assessment of the additional time required and difficulty associated with analyzing a pronominal reference.
2. The effect of increasing the number of potential antecedents agreeing in gender and number with a pronoun on subjects' difficulty in determining which is the correct referent.
3. Readers' use of gender and number to screen out alternative antecedent noun phrases from consideration as they analyze referential relations.
4. The effect of varying the topicality of the referred to noun phrase.
5. The effect of altering the topical status of an antecedent across sentences within a text, through repetition of the antecedent noun phrase or through pronominal reference.
6. The effect of increasing the contextual ambiguity of an anaphoric term; does the subject have greater difficulty when the context of a pronoun does not semantically constrain the referent for the pronoun, but instead allows several antecedents to be semantically compatible with the pronoun? What is the default referent assigned to a pronoun when the semantic context is nonspecific?
7. Readers use of semantic information contained in a lexical substitute to gain rapid access to the referred-to antecedent noun phrase when the context of the lexical substitute is by itself nonspecific (is compatible with the selection of either of two antecedent terms as referents).
8. The generality of readers' use of semantic context in tracing referential relations. Do readers use context only when it precedes the pronoun within a sentence, or are they equally likely to use contextual information when the context follows the pronoun within the sentence?

---

[1]With the exception of one subject scoring at the 3rd percentile, the subjects in this group ranged from the 52nd to the 35th percentile.

The eleven contrasts among text forms we used to focus on these issues are presented in Table 3. For each contrast, an analysis of variance was carried out which included reading ability group (high, middle, or low) as a subject grouping factor, and the alternative text forms as a repeated measures factor. Dependent variables included (1) the percent of correct pronoun referents supplied, (2) reading time per syllable for (in most cases) the sentence containing the pronoun whose referent was probed, and (3) response latency in reporting the pronoun. The means of the first two variables are given in Table 3 for each text condition employed in the contrast. Means for the third dependent variable will be reported in the text, where appropriate. In addition, differences among reader groups will be presented when they are significant. The results for each contrast in Table 3 will be discussed in turn.

1. Pronominal reference vs. repeated antecedents. Subjects generally required more time to read sentences which contain a problem of reference assignment (293 msec/syllable) than they did sentences in which the referring term (pronoun) was replaced by its antecedent (256 msec/syllable) ($F_{1,19}$=4.77, p=.04). There was also a marginal main effect of reading ability in this analysis ($F_{2,19}$=2.15, p=.14). Subjects in the high ability group had lower reading times (186 msec/syllable) than subjects in the middle (320 msec/syllable) and low ability (313 msec/syllable) groups. We should note also that a relatively high proportion of errors resulted when the second sentence of the text required subjects to understand a pronoun-referent relation (27%). Thus, even with simple two-sentence texts, it appears that the problem of reference tracing constitutes a significant source of comprehension failure among high school age readers. This confirms the results of our earlier study (Frederiksen, 1981b), where in addition we found a significant interaction between reading ability and time to process referential relations.

2. Multiple antecedents. As in the earlier study, we varied the number of antecedents present in the initial sentence of a text that agree with the subsequent pronoun in gender/number. The referent for the pronoun was the topic of the first sentence. Unlike the earlier study, we found no significant difference in reading time for the two text types. However, there was a marginally significant difference in percent of correct referents supplied ($F_{1,19}$=3.16, p=.09). Increasing the number of competing antecedents caused a decrease in percent correct from 93% when there was only one antecedent to 82% when there were two antecedents. Finally, there was a

marginal main effect of reading ability group on reading time ($F_{2,19}$=2.24, p=.13). The mean reading time for the high ability group (162 msec/syllable) was smaller than that for the middle (300 msec/syllable) and low ability (311 msec/syllable) groups.

3. Grammatical filtering of antecedents. There were no significant differences in accuracy in identifying referents when the gender/number of the second, non-topical antecedent of sentence one was made either to agree or not to agree with the pronoun. The subjects' accuracy when the second antecedent did not agree with the pronoun (86%) was closer to that obtained when the two antecedents agreed with the pronoun (82%) than to that obtained when there was only one antecedent present (93%). There was, however, a significant effect of antecedent agreement on subjects' reading times ($F_{1,19}$=10.67, p=.004). Eliminating the gender/number agreement of the second, non-topical antecedent resulted in an increase in reading time beyond that required when the second antecedent agreed with the first. The evidence thus contradicts the hypothesis that subjects can "filter" antecedents that are not in agreement with the pronoun, and in this way increase their efficiency of processing.

4. Topicality of the referent. In this contrast, we compared two-sentence texts in which we varied the position of the referent within sentence one, placing it in the topical (or subject) position or in a nontopical (predicate) position. While there was a marginally significant effect of these changes in text form on subjects' accuracy ($F_{1,19}$=2.56, p=.13), there was a highly significant effect on their reading time ($F_{1,19}$=10.58, p=.004). When the referent was moved from the topical to the nontopical position, there was a decrease in accuracy of 9% and an increase in reading time of 32 msec/syllable. There was also an increase in subjects' latency in reporting referents for pronouns of 388 msec, although this increase was not statistically significant ($F_{1,19}$=1.88, p=.19). Thus, topical status appears to render an antecedent concept more readily available for selection as a referent when such a text reference occurs.

5. Proximity of topic. In this contrast, we compared a text form (2) in which the referent is not the original topic but is topicalized in a second sentence with a text form (7) in which the reverse is the case: the referent is the topic of the initial sentence but is not the topic of the second, intervening sentence. In each case, the critical pronoun occurred in the third sentence. The results show strongly that it is

the current topical status of an antecedent that determines its availability in tracing referential relations. There were significantly fewer errors ($F_{1,19}$=11.63, p=.003), significantly faster reading times ($F_{1,19}$=4.82, p=.04), and significantly shorter report latencies ($F_{1,19}$=4.85, p=.04) (1327 msec for text form 2, and 1597 msec for text form 7) when the referent was the immediate topic than when it was the topic of the more distant sentence. Thus, the topical status of the referent appears to vary at different points in the text, and it is the current topic that has priority in reference assignment. This could be due to increased activation accorded the topic within semantic memory, or to its having a priority position within the text model (Kintsch & van Dijk, 1978; Vipond, 1980; Kieras, 1981).

In this analysis, there were significant reader ability differences in accuracies when subjects were probed for pronoun referents ($F_{2,19}$=6.87, p=.006). The high ability readers were generally more accurate (91%) than either the middle group (74%) or the low ability group of readers (75%).

6. Continuity of reference. In this contrast, we compared three sentence texts (forms 11 and 13) in which continuity of reference was maintained, with texts in which it was not (7). In the first two of these text forms, the critical pronoun occurring in the third sentence is also used within the second or intervening sentence, and in both sentences it refers to the topicalized antecedent in sentence one. In text form 11, the pronoun is in the subject position within sentence two, while in text form 13 it is in the predicate position. We found that subjects were significantly more accurate when there was a prior use of the pronoun to establish referential continuity ($F_{2,38}$=3.45, p=.04). The mean percent correct was 82% when there was continuity of reference, while it was only 70% when referential continuity was lacking. However, under these conditions subjects' accuracy did not depend on the position of the pronoun within the intervening sentence. (It was 81% when the pronoun was in the subject position, and 83% when the pronoun was in the predicate position.) Aside from these differences in accuracy, there were no significant differences in performance among these text forms, in reading times or in report latencies. (We should point out that in all three of these text forms, the referent was the topic of the first sentence.) These results differ from those of our earlier experiment (Frederiksen, 1981b), where we found that the facilitating effect of the prior reference was found only when the pronoun occurred as the subject of the intervening sentence and thus maintained the topical status of the referred-to concept.

7. Topicalizing within an intervening sentence. In this contrast we examined the effect of topicalizing the referent of a pronoun which was not in. ne topic of the paragraph by repeating it as the subject within an intervening sentence. There were significant effects of this manipulation on both subjects' accuracy ($F_{1,19}$=9.65, p=.006) and reading time ($F_{1,19}$=17.04, p=.0006). Topicalizing the pronoun's referent within an intervening sentence resulted in a 17% higher accuracy than that obtained when the referent was not topicalized. Reading times were also reduced from 293 msec/syllable for the non-topicalized case to 235 msec/syllable for the topicalized case.

The three reading ability groups differed in their accuracies in supplying referents, which were, respectively, 91%, 81%, and 72% for the hig' dle and low groups. However, the main effect of reading ability was only marginally s. ificant ($F_{2,19}$=2.30, p=.13).

8. Ambiguity of reference. This contrast was included in order to confirm our earlier finding (Frederiksen, 1981b) that it is the information contained in the context frame of a pronoun that enables subjects to identify the intended referent with minimal effort and attention. We compared our "standard" two sentence case in which a pronoun in the second sentence refers unambiguously to an antecedent (here a topicalized antecedent) with a case in which the context of the pronoun in the second sentence does not unambiguously point to a single referent. These ambiguous sentences were written so that either of two antecedents in sentence one would fit within the semantic context of the second sentence. This can be regarded as an extreme example of a situation in which the contextual constraints, which generally will enable a direct access to the referent, are weak or nonexistent. In the analysis of variance, there were significant differences between the two text forms, both in subjects' reading times ($F_{1,19}$=8.34, p=.009), and in their accuracies in reporting pronoun referents ($F_{1,19}$=4.45, p=.05). The mean reading time for ambiguous contexts was 327 msec/syllable, which is the highest value we obtained across the entire set of text forms employed in the experiment. We should point out that the lower rate of errors for ambiguous contexts (94% correct, compared with 82% for the standard context) reflects the fact that in the former case either of the two antecedent noun phrases of sentence one is correct, and was therefore scored as correct.

There was also a marginally significant main effect of reading ability on reading times in this analysis ($F_{2,19}$=2.58, p=.10). Subjects in the high ability group had faster reading times (178 msec) that either of the other two subject groups (360 msec and 338 msec, respectively, for the middle and low groups). Thus, we have additional evidence that high ability readers are more efficient in processing referential relations than the other groups.

9. Default referent assignment in an ambiguous context. This contrast sought to address the question of how a referent is assigned to a pronoun in the absence of a contextual basis for making a selection. The text forms used in this comparison are given in Table 3. In text form 12, the second sentence is ambiguous. Our hypothesis was that, since contextual constraints are lacking, the default assignment within this sentence is the currently topicalized antecedent, namely A_2. Evidence for this hypothesis can be found in the effects of such a reference assignment on a subsequent reference to the initially non-topicalized antecedent. If the referent initially assigned to the pronoun in sentence two is A_2, then subjects should have difficulty in assigning an alternative referent for the pronoun in the subsequent sentence, which is A_1. Thus, their performance should resemble that for text form 7, in which the antecedent A_2 has been explicitly presented in the second sentence. The analysis of variance revealed a significant main effect of text form on subjects' accuracy in reporting the correct referent ($F_{1,19}$=4.68, p=.04). Subjects' accuracy when the ambiguous pronoun was present in sentence two (57%) was actually lower than that when the alternative antecedent was presented within the intervening sentence (70%).[2] These results support the claim that when the context is ambiguous and provides no countervening evidence, the current topic is assigned as the referent for the pronoun. On the other hand, when constraining semantic information is available within the pronoun's context, the bias towards selecting the current topic is mitigated and the alternative, semantically consistent antecedent is more likely to be selected (for instance, the correct antecedent was reported 73% of the time for text form 3). The mechanism for mapping referents is therefore capable of combining semantic information derived from context with information concerning the topical status of concepts within the text.

---

[2]The reading time for this condition was lower than that for the alternative condition, suggesting that subjects were at times simply reporting the prior referent of the pronoun without considering the semantic context.

In addition to the main effect of text form on accuracy, there was a marginally significant main effect of reading ability group ($F_{2,19}$=2.83, p=.08), and an interaction between text form and reader ability ($F_{2,19}$=3.31, p=.06). Subjects in the high ability group (86%) were markedly superior to the other groups (57% and 67%, respectively for the middle and low ability groups) in supplying the correct referents when the pronoun had not occurred earlier in the text (text form 7). However, when the pronoun had occurred earlier in the text (text form 12) and was (we have inferred) assigned to a different referent than that to be reported, all groups performance was poor, with the performance of the low ability group (45%) markedly lower than that for the high (63%) and middle (66%) groups. The lowest ability group tended to report the earlier referent of the pronoun as the referent for the pronoun in the final sentence, even though it was semantically inappropriate in that context.

10. Semantic memory activation. In addition to information derived from the context in which a pronoun or other anaphoric word appears, when the anaphoric term is a lexical substitute (such as a synonym or superordinate), there is semantic information available within the anaphoric term itself and this provides a basis for tracing referential relations. Consider how skilled writers try to vary their language in referring to a concept in order to avoid sounding too repetitious or to shorten their references. (For example, as in our earlier example, they will use words like "the task" to stand for earlier occurring concepts such as "the ascent of the peak".) Such writing practices assume that using such lexical substitutes will not lead to errors in reference tracing or to greater difficulty in comprehension. These assumptions thus presuppose an efficient, automatic mechanism for tracing referential relationships, through the overlapping of word meanings, as well as through the use of contextual and topical information.

The present contrast among text forms was designed to allows us to study the use of semantic relationships among words as a separate contributer to the efficient mapping of referential relations. We introduced a text form (10) in which a lexical substitute (generally a synonym, near synonym, or superordinate) replaced the pronoun within a context that was otherwise ambiguous. The logic of our predictions was similar to that of the preceding contrast. Within text form 10, the mapping of the synonym in sentence two to antecedent A_1 should produce a low error rate in the mapping of the pronoun in sentence three to its referent, A_1. Thus, performance should be similar to that for text form 9, in which the antecedent A_1 was actually inserted in place of the synonym within the intervening sentence. And performance for both text forms 9 and 10 should be more accurate than that for text form 12, in which a pronoun replaces the synonym within the ambiguous context of the intervening sentence (since we have seen that the pronoun in this case is assigned to the focalized antecedent A_2). The results supported these hypotheses. There was a significant main effect of text form on accuracy ($F_{2,38}$=13.52, p<.0001). The accuracy for the case where the lexical substitute occurred within the intervening sentence was 76% while it was 83% for repeated antecedent and only 57% for the ambiguous pronoun. There was also a marginally significant effect of text form on reading efficiency ($F_{2,38}$=2.82, p=.07). The reading times were, respectively, 278 msec/syllable, 283 msec/syllable, and 249 msec/syllable. Note that the reading times for the first two conditions were nearly identical. The low reading time of 249 msec/syllable for the case where the pronoun had been previously used to refer to an alternative referent (where the accuracy was only 57%) suggests a speed-accuracy trade-off -- subjects at times simply reported the prior referent of the pronoun without considering the semantic context. The high degree of similarity in performance for the text forms involving the synonym and the repeated antecedent supports the hypothesis that activation of concepts in semantic memory can contribute to efficient tracing of referential relations.

11. Frame-based semantic constraints. We have reviewed evidence that information derived from the context frame of a pronoun can be used to efficiently map referential relations. The final contrast among text forms was included in order to determine the conditions under which contextual information is used in this way. In one condition, the pronoun preceded the context, while in the other, it followed to a large measure the context. If in the former case subjects rely on a search of their prior text model to immediately reinstate candidate referents and then select from among these as contextual information becomes available, reading times should be elevated, since the absence of contextual information precludes an efficient process for referent selection through the use of semantic memory. If, on the other hand, subjects postpone the assignment of a referent until contextual information becomes available in order to use more efficient search processes, then reading efficiency should be independent of the position of the pronoun. In the analysis of reading times, there was a significant effect of text form ($F_{1,19}$=6.60, p=.02), and a significant

interaction of reading ability and text form ($F_{2,19}=7.17$, $p=.005$). For subjects in the middle ability group, reading times were longer when the pronoun occurred at the end of the final sentence (424 msec/syllable), rather than at the beginning (335 msec/syllable). The other two groups of readers had similar reading times for either position of the pronoun (they were 208 and 201 msec/syllable, respectively, for the high ability group, and 328 and 337 msec/syllable for the low ability group). In addition, there was a significant effect of text form on subjects' latencies for reporting the correct referent ($F_{1,19}=9.68$, $p=.006$). Subjects generally took longer to report a referent when the pronoun occurred at the end of the sentence (3760 msec) than when it occurred at the beginning (1837 msec). This suggests that when the pronoun occurs at the end of the sentence, subjects may be continuing the process of reference mapping after they have pressed the button at the end of the sentence. Neither of these results offers support for the hypothesis that contextual information is used in assigning referents only when it precedes the pronoun. Instead, readers appear to process the words of a sentence in chunks rather than serially, and these chunks are large enough to encompass the contextual constraints on the identity of a pronoun's referent.

1.3.3 Summary and Discussion

The experimental subjects were found to require greater time to process a sentence which contains a problem of reference analysis than one that does not. In this study, while there were marginally greater errors when multiple antecedents were present in the preceding text, there was no compelling evidence for a reduction in the effect of multiple antecedents when the gender/number agreement of the alternative antecedent and pronoun was eliminated. Rather, creating a gender/number mismatch appeared to increase difficulty. Thus, there was no evidence supporting our first hypothesized process that in searching a text model for a referent, subjects can automatically "filter out" from consideration antecedents that do not agree syntactically with the pronoun.

A series of contrasts that we carried out among text conditions showed that subjects had greater difficulty in mapping references when the pronoun referred to a non-topic than when it referred to a topic. When a sentence intervened between the referent and pronoun, subjects were more accurate in assigning referents when there was a continuity of reference from sentence to sentence. When the continuity of reference was established through argument repetition, there was greater difficulty when the referent was not a topic of the intervening sentence than when it maintained the topical status of the referent. However, when continuity of reference was established through pronominal reference, the topicality of the pronoun was not important. Prior consistent use of the pronoun thus facilitated subsequent use of the pronoun, and overroad the topicality effects. However, apart from the special case of prior consistent use of a pronoun, we have seen that the topical structure of a text has an important impact on the process of reference tracing and can influence, through its effect on topicality of antecedents, how efficiently this process is carried out. Note that this impact was felt even when the context of the pronoun contained semantic information that constrained highly the identity of the referent concept. Thus, information concerning topical status and information concerning semantic appropriateness appear to jointly influence subjects' efficiency and accuracy in mapping references. In this experiment, we manipulated the topicality of antecedents by varying the syntactic subject of the initial and subsequent sentences within a paragraph. However, more generally topicality depends upon the text macrostructure as well as upon local sentence structure (Kintsch & van Dijk, 1978; Vipond, 1980; Kieras, 1981). Thus, text macrostructure would be expected to influence ease of reference mapping as well.

When the semantic context of a pronoun was rendered ambiguous, subjects' effort in resolving the reference problem (as measured by reading time) increased dramatically. Subjects thus appear to depend heavily on semantic information derived from context in identifying, without a high degree of effort, an appropriate referent for the pronoun. This process, we have hypothesized, is driven by semantic information concerning the identity of the referent that is available at the time the pronoun is encountered (i.e., in the context of the pronoun). We found that this relatively automatic process for reference mapping may also make use of semantic information contained in the anaphoric word itself, when such information is available (as in the case of a lexical substitute). In our experiment, when we placed a lexical substitute (a synonym or superordinate of the referent) in a sentential context that was otherwise ambiguous (that is, it would allow multiple antecedents to be selected), our subjects mapped the reference as readily as when the referent was actually repeated within the ambiguous sentence. Thus, it appears that subjects can use semantic information derived from either source in automatically tracing a reference to an earlier occurring concept within the text.

Reader group differences. A difference between the present results and those of the previous study is that we found less evidence for interactions between text form and reader group. This may be due to the smaller sample size of the present study which reduced power, particularly in making between group comparisons, and to smaller differences in the range of abilities represented by the three ability groups in the present study.[3] With respect to subjects' accuracy in supplying referents for pronouns, there was evidence obtained in contrasts 5 and 7 indicating that high ability subjects tend to perform better than the other subjects under conditions where the needed referent is not the current topic within the text. When we compared the difference in mean percent correct for the high ability group with the average for the other two groups taken together, we found in contrast 5 that the difference between these means was 11% when the referent was the current topic, while it was 24% when the referent was not the current topic. Corresponding figures obtained in contrast 7 are 10% in the case of a current topic, and 20% when the referent is not the current topic. Again, in contrast 9, when the referent of the pronoun was not the immediate topic, the mean percent correct was 86% for the high ability group, compared with an average of 62% for the middle and low ability groups (the difference between these means is 24%). In this last contrast, the interaction of reader group and text form was close to achieving significance (p=.06). Thus, our first conclusion is that, with respect to accuracy, the high ability group is distinguished from the other two groups in its ability to use information concerning topical status of concepts in selecting referents for pronouns.[4]

For the second condition within contrast 9 (text form 12), it was the lowest

---

[3]With the exception of one subject (whose reading test score was at the 3rd percentile), the reading ability level of the lowest group covered a range from the 35th to the 52nd percentile with a median at the 44th percentile.

[4]In the prior study, we found evidence that lower ability readers showed larger increases in reading time than did high ability readers when the referent was shifted from a non-topical to a topical position.

ability group that was distinguished from the other two groups in accuracy in supplying referents. In this condition, subjects encountered an ambiguous pronoun within an intervening sentence and tended to assign to it the current topic. When they subsequently encountered the same pronoun (this time used to refer to a different referent), the poor readers showed an inability to use the semantic context of the pronoun to rule out the prior referent of the pronoun as inappropriate. Their error rate in supplying the correct referent was 55%, compared to error rates of 37% and 34% for the other two ability groups. Our second conclusion is, therefore, that the lowest ability readers have a particular deficiency in their ability to use semantic information derived from context in tracing referential relations.

In addition to these reader group differences in accuracy, in three of the contrasts (1,2, and 8) there were significant reader group differences in efficiency of processing, as measured by subjects' reading times. While the conditions in these contrasts were quite variable (they included cases where the referent was a topic or not a topic, where there were multiple antecedents or a single antecedent, and where the context of the pronoun was ambiguous or non-ambiguous), the mean reading times of the three groups were quite consistent across text conditions. In each case, the high ability group was clearly distinguished in processing time from the other two groups. (The mean reading times across these conditions ranged from 162-186 msec/syllable for the high ability group, from 300-360 for the middle group, and from 311-338 for the low ability group.) Note that all of these contrasts involved two sentence texts, and the error rates were relatively low. Thus, under the less difficult of the experimental conditions, where the three ability groups did not differ in accuracy, the high ability readers showed higher reading rates than subjects in the other groups.

In summary, our conclusions are that:

1. The lowest ability subjects have particular difficulty in mapping references when information from the context of a pronoun must be used in re-assigning the referent for a pronoun.

2. The high ability group exceeds the middle and low ability groups in making use of topical information to gain access to the correct referent.

1.3.4 An Activation Model for Reference Mapping

There are two process supporting reference mapping:

1. Search of a text model containing the reader's representation of the text and selection of a referent based upon the match of antecedents with the context frame of the pronoun.

2. Activation of concepts in semantic memory on the basis of the context frame of the anaphoric word and semantic information derivable from the anaphoric word itself.

The high reading times and rates of errors obtained when the context frame of a pronoun does not semantically constrain the referent suggest that the former of these processes is effortful and attention demanding, and that the latter is relatively automatic. The ability to use a general process for searching semantic memory for the purpose of reference assignment during reading depends upon the existence of a linkage between the text model the reader is creating, and concepts in semantic memory. Presumably, these linkages are bidirectional. (1) Tracing a link from the text model to the knowledge base of semantic memory allows for activation of concepts within semantic memory, for the purposes of concept elaboration and inference. (2) Tracing linkages from concepts within semantic memory to their instantiations within the text model enables reference tracing. Finally, the existence of activation in semantic memory due to concept use within a text model facilitates the integration of topical and semantic information in reference assignment. Activation of concepts is presumed to based upon either semantic information or information concerning topical structure.

Referents for a pronoun can be generated by initiating parallel searches of semantic memory, one on the basis of semantic information (either derived from context or from the anaphoric word itself) and the other on the basis of topical information (from currently activated concepts within semantic memory -- which are linked with concepts in the text model). The searches proceed from their starting information through a process of spreading activation (Collins & Loftus, 1975). The strength of activation is, in the case of the semantic search, determined by the degree of contextual support and, in the case of the topical search, by the level of importance in the topical structure at the time the search is initiated. When activation of a concept resulting from these search processes reaches a threshold level, the concept emerges as the referent assignment for the pronoun. The set of concepts activated as a result of these two search processes constitutes the set of possible references for the anaphoric term in question. When a single concept is so activated, it is adopted automatically as the correct referent for the anaphoric term. When two or more concepts are activated, a more effortful, deliberative process is hypothesized to be invoked to resolve the conflict.

When the set of activated concepts contains a single concept, reference assignment is fast and effortless. For example, when the anaphoric term is a synonym or a repetition of the referent, reference tracing will be based on the direct (that is, with little activation spread) and immediate activation of the referred-to concept based on semantic information contained in the anaphoric term itself. When the anaphoric term is a pronoun, reference assignment will be automatic when the context constrains highly the referent for the pronoun, and when the required referent is currently receivir¸ topical emphasis within the text. However, if there is a conflict between activation due to topical and contextual information, multiple antecedent concepts may be activated and a more effortful, deliberative process may be needed to choose among them. If, on the other hand, the semantic information available is non-selective (for example, if the context of the pronoun contains only general information), we would expect reference assignment to be based more heavily upon topicality constraints, since concepts will be activated principally due to topicality relations and only marginally to semantic constraints. However, if the contextual information available is ambiguous (that is, it supports two or more particular antecedents), several concepts will be activated on semantic grounds and, again, a deliberative process must be invoked to settle (if possible) on a "best" referent.[5] The hypothesized automatic process for mapping referential relations contrasts with the attention-demanding search process which is posited for scanning the text model (cf. Giron, Kellogg, Posner, & Yee, 1985). It is our contention that searches of a text model can be avoided for the most part since pronominal references within natural text are typically to antecedents i nearby clauses (Giron, et al., 1985) and typically to subjects of those clauses.

---

[5]Subjects in our experiments at times actually reported both possible antecedents as referents when the semantic context was ambiguous.

On the basis of this model for reference assignment and on the reading ability differences we have observed in this and in previous research, we feel that a worthwhile approach to developing such automatic skills for reference mapping will be to train subjects in the use of the semantic information within a sentence context to guide their search of semantic memory in gaining access to concepts that are consistent with the sentential context. If the model we have suggested for automatic tracing of referential relations is correct, then we will expect to find transfer of training to the performance of a reference tracing task, for those textual conditions in which an automatic process has been hypothesized to be operative. Such a training study was carried out within the current project, and will be presented in the following section.

2. EFFECTS OF TRAINING IN THE USE OF CONTEXT ON REFERENTIAL ANALYSIS

In the previous section we have sketched a mechanism that can perform an automatic assignment of referents for anaphoric terms under a wide variety of textual conditions. According to the theory, activation of concepts in semantic memory is initiated on the basis of the semantic context of the referring term (the anaphoric word), the current topic, and semantic information contained in the anaphoric word itself (for example, when it is a lexical substitute). The referent for an anaphoric word is chosen by selecting the concept(s) receiving the greatest activation on the basis of these sources. When two or more concepts are thus selected, a more effortful, deliberative process is invoked to resolve the conflict. This deliberative process becomes involved, for example, when there is contextual support for two or more antecedent noun phrased as referents (as was the case when the sentence containing the pronoun was written to be ambiguous), or when there is a conflict between concepts activated on the basis of topicality and context (as was the case when the referred-to antecedent was not the topic of the previous sentence).

To test this theory, we carried out a training study in which we sought to develop subjects' ability to use the information in a context frame in gaining access to concepts which occur within that context. The training task we employed focused on semantic entailments within a single sentence (such as those associated with a case frame; cf. Fillmore, 1968, Bruce, 1987) and their use in constraining the "semantic space" associated with a missing word within the sentence. After training, subjects were then tested for transfer of this ability to use contextual constraints to the performance of a reference mapping task in which subjects were required to supply referents for pronouns appearing near the end of a two or three sentence text. Since the training task does not involve practice in reference mapping, it is the <u>indirect</u> effects of the skill acquired during training that can exert an influence on transfer task performance (cf., Frederiksen Warren, & Rosebery 1985a,b). In particular, training in the use of context to prime categories in semantic memory should improve performance on reference mapping problems for which an activation-based selection process is sufficient, and should not improve performance otherwise.

These predictions can be made more precise after we examine the training and transfer tasks more closely. The training task we created incorporates sentences that (at first tightly and later less tightly) constrain the identity of a word within them. During training, this word is presented to subjects in a visually degraded form so that it cannot be recognized without a close consideration of the semantic context in which it occurs. The subjects' task during training is to make a judgement of the semantic acceptability of the word in its context, as quickly as was possible. The word at first appears in a highly degraded form (only a small percentage of pixels of each letter are turned on) and then slowly increases in visual clarity (more and more pixels are turned on). Therefore, to respond early subjects have to integrate information derived from context with visual information they have received. This training task thus develops subjects' ability to integrate contextual information with visual information in the activation of relevant concepts in semantic memory, so as to efficiently gain access to word meanings.

Since such a skill is hypothesized to mediate the automatic mapping of referents for pronouns we predicted that, as a result of training, subjects who were initially deficient in such a skill would show improvement in efficiency of reference tracing for those textual conditions in which automatic reference tracing is in theory operative, but not for texts where such an automatic mechanism is inoperative. To test this prediction, we constructed two parallel versions of a reference tracing task, to be administered before and after training in the use of context. This transfer task included four text conditions used in the previous experiment. These included two (text forms 1 and 2) in which contextual and topical constraints were consistent, and two in which these constraints led to a conflict among antecedent concepts. In one case (text form 3), the context constrained one antecedent concept and the topical constraint supported another antecedent. In the other case (text form 5), the semantic constraints were ambiguous, pointing to two antecedents. In both cases, since there is a conflict among antecedents, a more effortful, deliberative process is hypothesized to be operative due to the fact that the activation patterns do not allow the selection of a single antecedent. We therefore predicted that there would be significant reductions in reading times and increases in accuracy in reporting referents of pronouns for text forms 1 and 2, and no improvement in either measure of performance for text forms 3 and 5. This differential transfer of performance would not be predicted if the activation model for reference mapping were incorrect, and the only basis for transfer of training were general improvements in reading speed/comprehension.

2.1 Method and Subjects

2.1.1 Defender: A Training Environment for Context Utilization

The Defender training environment is based upon a contextual priming task in which the subject must read and evaluate words presented in a sentence context. The target words are presented in a visually degraded condition, in order to encourage subjects to make use of semantic information present in the context, which constrains the word's identity. For each context sentence, two or more target words are presented successively, each on a separate trial of the training game. In this way, we sought to acquaint subjects with the "family" of words or concepts that are appropriate within a given context.

For each context sentence, the early target words were of lower frequency[6] and the later targets were of higher frequency. Initially in training, the target sentences presented were strongly constraining, allowing only words representing a single concept. Later in training, the context sentences were less strongly constraining, allowing words representing a family of concepts. For instance, an example of a strongly constraining sentence is:

> "The new minister managed to spark a controversy in his very first _____."

For this context, a low probability target word was "service", while the high probability target word was "sermon". (Examples of foils or inappropriate words were, in this case, "veranda" and "mold".) An example of a lowly constraining context sentence is:

> "To no one's surprise, she completely captivated the audience with that final _____."

Low probability target words included "encore", and "remark", while the high probability target was "song". In all, subjects received 129 high constraining sentences and 106 low constraining sentences. There were an average of 4.6 stimulus words for each context sentence (48% low probability words, 19% high probability words, and 33% foils), making a total of 1,085 stimulus words.

---

[6] Frequencies of words in context were determined empirically; see Frederiksen, Warren, & Rosebery, 1985b.

A representation of the computer screen during the Defender game is shown in Figure 2-1. At the top of the screen is the context sentence in which a word is missing. After the subject has read the sentence (at whatever pace he or she desires), the subject presses the space bar to initiate the presentation of the target word, which emerges from the "chute" below the sentence. The word appears in a series of exposures, each below the previous one. Initially, only a small proportion of the pixels for each character of the word are available (for example, 24 out of 64 possible). On each successive exposure, an additional 2 pixels of each character are turned on. The subjects' task is to judge the suitability of the word in its context, and to respond as early as possible (that is, at as low a pixel density as possible), consistent with making a proper evaluation of its meaning.

A game theme has been created for the training task which centers on the sending and receiving of code messages in order to protect friendly spaceships from enemies on a space station. The player is a spy on the station (represented by the ground at the bottom of the screen). In order to tell if an incoming ship is a "friend" or an "alien", English messages with a word missing are sent to incoming ships, which must in turn respond with a correct filler word. Only friendly ships will be able to respond with an semantically appropriate word (since only they have a clear knowledge of life on earth). As the spaceship comes closer, the words become clearer and clearer. When a semantically acceptable word is returned by a friend ship, the subject must correctly judge it as early as possible. When it is correctly judged, a warning signal is sent from the antenna on the left of the screen. The ship will then escape after delivering a fuel tank to the player. If the player is incorrect, the friend ship will crash. If the ship is an alien ship, the word that is returned will not be a semantically appropriate word. If the alien ship is correctly recognized, a special "all clear" beam is sent from the antenna on the right and the alien ship will move off without landing. If however it succeeds in landing, a tank of fuel is stolen from the player. The player's score depends upon how early he or she recognizes the "friend" (i.e. contextually appropriate) words, and on whether or not the "alien" (i.e., inappropriate) words are correctly rejected (but not on how early they are judged, since context cannot prime the recognition of such words).

The starting pixel density on each trial and the time interval between exposures are adjusted dynamically based upon the subjects' earlier performance. (The rules

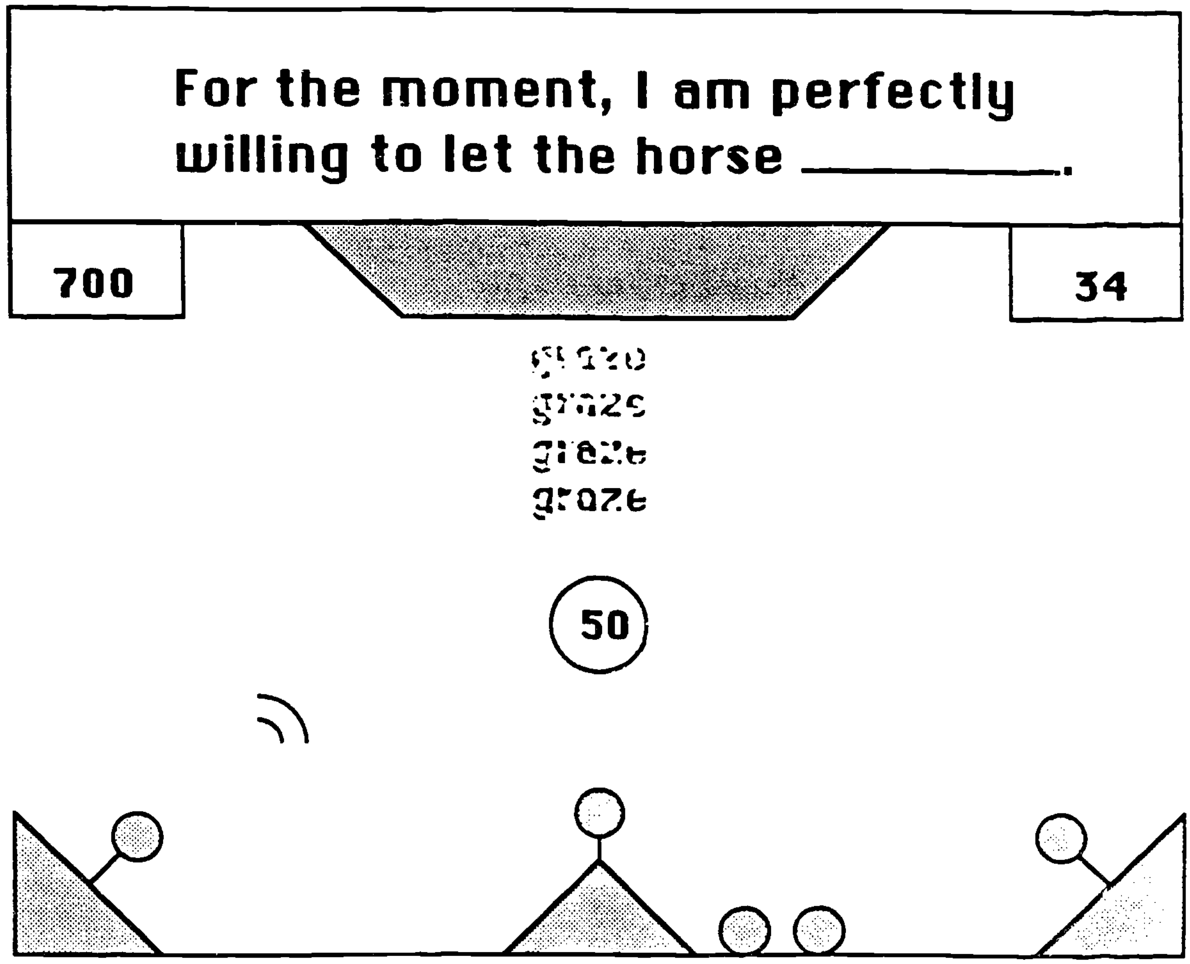

Figure 2-1: A representation of the computer screen during the Defender game.

used in making these adjustments are described in Appendix C, which contains a more full description of the Defender game.) Performance measures obtained during training include (1) the mean pixel densities for judging high and low probability words and foils, and (2) the mean percentage of correct responses. These performance measures were calculated for each block (day) of practice. Each subject received a total of eleven practice blocks, each lasting approximately one hour.

Subjects. The subjects were eight high school students who scored below the fiftieth percentile on the Gates–MacGinitie Reading Test. Their scores ranged from the 3rd to the 48th percentile, with the median at the 26th percentile. The subjects included 5 males and 3 females.

2.1.2 The Reference Mapping (Transfer) Task

The transfer task was derived from the materials used in the experiment on mapping of referential relations described in the previous section. Subjects read two or three sentence texts presented on the screen of an IBM personal computer. Following the final sentence of a text, they were probed for the referent of a pronoun occurring in that sentence. Their task was to orally supply the correct referent, which was an antecedent noun phrase occurring in an earlier sentence. Performance measures included (1) subjects' mean reading time per syllable for the final sentence containing the pronoun whose referent was probed, (2) subjects' mean percentage of correctly supplied referents for pronouns, and (3) subjects' mean report latencies, measured from the occurrence of the probe (an underscore appearing under the pronoun) to the onset of vocalization.

The text forms included in the transfer task are shown in Table 4. In text forms 1 and 2, the referent for the pronoun was unambiguously constrained by the context frame of the final sentence, and was also the current topic of the text. In text form 2, the correct antecedent A_1, while not the topic of the initial sentence, is foregrounded within the second, intervening sentence. In text form 3, the correct antecedent is not the topic of the previous sentence. In text form 4, the context is ambiguous, providing support for both of the antecedents occurring in the previous sentence.

Table 4

Text Types Used in the Reference Mapping Task and Hypothesized Sources of Activation for Each Antecedent

Text Form	Format of Sentences	Activation of Antecedents	Ratio of Activations
1	S_1:A_1 __ A_2 __.	A_1 & A_2 by context.	3:1 or 3.0
		A_1 as topic.	
	S_2: Pr(A_1)___.	A_1 by context.*	
2	S_1: A_2 __ A_1 __.	A_1 & A_2 by context.	4:1 or 4.0
		A_2 as topic.	
	S_2: A_1 ______.	A_1 by context.	
		A_1 as topic.	
	S_3: Pr(A_1)___.	A_1 by context.*	
3	S_1: A_2 __ A_1 __.	A_1& A_2 by context.	2:2 or 1.0
		A_2 as topic.	
	S_2: Pr(A_1)__.	A_1 by context.*	
5	S_1: A_1 __ A_2 __.	A_1& A_2 by context.	3:2 or 1.5
		A_1 as topic.	
	S_2: Pr(A_1 or A_2)__.	A_1& A_2 by context*	

* Hypothesized primary locus of instructional effects.

Applying a very simple-minded model for counting the sources of activation for each of the antecedents (A_1 and A_2), we can develop predictions concerning the relative difficulties of the four text forms assuming that the hypothesized activation model for reference tracing is operative. We count each contextual activation of an antecedent as one unit of activation, and add an additional unit for the antecedent that is the topic at the time of the final sentence.[7] The ratios of activations for antecedents A_1 and A_2 are given in Table 4 for each text form. Text forms 1 and 2 both have a large ratio of activations, indicating that the activation of the correct antecedent A_1 is much stronger than that for the incorrect antecedent A_2. The activation ratio for Text form [illegible] the higher of these two, since the first antecedent received additional activation from its occurrence within the second sentence, where it also was accorded topic status. Text forms 3 and 5 both have low ratios of activation, due to a conflict between contextual and topical constraints (text form 3), or to an ambiguity in the contextual constraints within the sentence containing the pronoun (text form 5). Our hypothesis is that mean reading times will be ordered in the reverse of the order of activation ratios: the higher the ratio of activations, the more readily subjects will select the correct antecedent on the basis of levels of activation. The smaller the ratio, the more difficult it will be for subjects to distinguish the correct antecedent on the basis of activations alone. In this case, subjects will need to invoke a more deliberative process for evaluating which antecedent is the correct referent for the pronoun. With respect to accuracy, subjects should be more accurate in the cases where the ratio of activations allows a selection of a single antecedent than in the case where there is a conflict between two antecedents. However, since in text form 5 either antecedent is correct (and was so scored), accuracy should be high in that case as well.

These predictions concerning the relative difficulty of the four text forms have received some support in the previous experiment. The mean reading times for text forms 2, 1, 3, and 5 were 235 msec, 261 msec, 293 msec, and 327 msec/syllable, respectively. The mean percentages of correct responses were 90% for text form 2, 82% for text form 1, 73% for text form 3, and 94% for text form 5. Thus, for a group

---

[7]The actual values assigned to these sources of constraint are not critical in making these predictions.

of subjects having a relatively high median Gates-MacGinitie Reading Test score (corresponding to the 75th percentile), the predicted rank order of reading times was confirmed.

Our expectation is that for the group of subjects in our training experiment, who were chosen to have lower reading scores than those of the previous experiment (with a median at the 26th percentile), differences in performance prior to training will not be associated in this way with the pattern of contextual activation. These subjects we hypothesize are less apt to base their performance on levels of activation of concepts within semantic memory, and are more likely to depend upon a search of their prior text model. To the extent that their search is optimized by topicality, their performance for text forms 1 and 2 should be superior to that for text forms 3 and 5. Further, their mean reading times should in general be longer than those of our earlier group of subjects cited above, and their levels of accuracy should be lower.

The effects of training on performance for these four text forms should also follow the above predictions. Text forms in which patterns of activation due to contextual information will permit the selection of a referent should show a benefit of training, while those in which there is a conflict among antecedents that are activated should show no such benefit of training. Finally, following training we expect that subjects' performance on the four text forms should conform to the predicted pattern, based on patterns of activation in semantic memory.

2.2 Results

We will first present results obtained during training in the use of context. These results bear on the success of subjects in acquiring such a skill. Then we will present results for the the transfer task. These findings will bear on subjects' use of contextual activations as a basis for mapping of referential relations.

2.2.1 Defender Training

A repeated measures analysis of variance was carried out for two dependent variables: (1) the mean pixel density at the time of the subject's response, and (2) the percentage of correct responses obtained during training. There were two factors in the analysis: Training Block (blocks 1-6 vs. blocks 7-11), and Stimulus Probability (high probability target words, low probability target words, and foils or words that are unrelated to the context). The mean pixel densities obtained by subjects early and later in training are shown in Figure 2-2 for each type of stimulus word. There were significant effects of Training ($F_{1,7}$=6.74, p=.04) and Stimulus Type ($F_{2,14}$=14.94, p=.0003) on the mean pixel density required for word recognition. The overall average pixel density decreased from 37.2 in the first six blocks of training to 28.4 in the last five blocks of training. This reduction in amount of visual information required occurred despite the fact that early in training the context sentences were highly constraining, while later in training they were predominately lowly constraining. At each stage in training, the average pixel densities required were lower for words that are context related (they averaged 31.8 and 32.3, respectively, for the high and low probability target words), than for unrelated words (34.3).

The subjects' accuracy in judging the semantic appropriateness of target words and foils is shown in Figure 2-3. While there were significant differences attributable to a stimulus word's probability of occurring in the context frame of the sentence ($F_{2,14}$=13.78. p=.0005), there were no changes in accuracy as a result of training ($F_{1,7}$=.004). While the mean percentages of correct responses did not differ significantly for high (89.8%) and low probability (86.9%) target words, both of these values exceeded that for foils (62.5%); the corresponding t-tests were t_7=6.89, p<.0005 and t_7=6.23, p<.0005, respectively. The significantly lower accuracy shown by subjects in judging the appropriateness of foils suggests that subjects had a tendency to trade speed of responding against accuracy in that case. When a stimulus word was not constrained by the context, subjects tended to respond before they had recognized the word, guessing as to its semantic appropriateness.

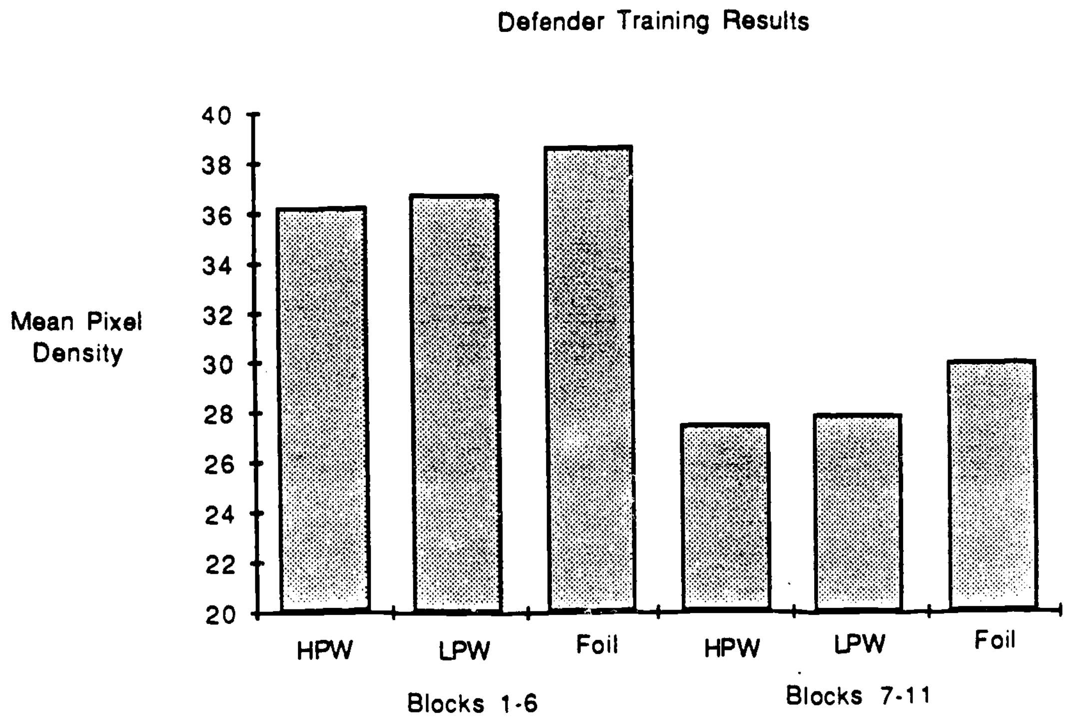

Figure 2-2: Mean pixel densities for high probability words, low probability words, and foils presented earlier and later in Defender training.

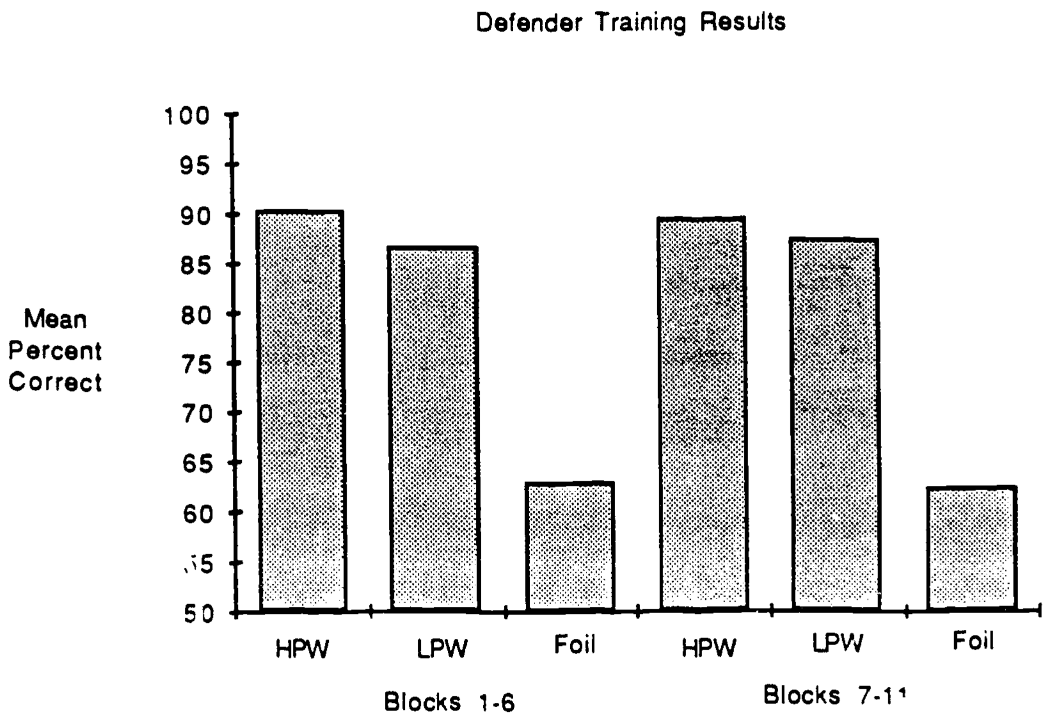

Figure 2-3: Mean percentages of correct responses for high frequency words, low frequency words, and foils presented earlier and later in Defender training.

2.2.1.1 Individual Differences

While the above statistics characterize the group as a whole, there was evidence of individual differences in the strategies subjects adopted during Defender training. Data for individual subjects are given in Table 5. We classified subjects into strategy groups on the basis of these performance data, and an inspection of their daily performance (both their mean pixel density and accuracy). The subjects appear to fall into three strategy groups, based upon the goals we infer they have adopted.

1. Speed and accuracy goals. Four subjects (YW, LW, AP, and JM) adopted a goal of improving both speed and accuracy. These subjects decreased the average pixel density they required (or, in the case of subject AP, maintained an already low average pixel density) while maintaining or increasing their accuracy. A representative performance operating characteristic for this group of subjects is shown in Figure 2-4. The points plotted in this figure represent the mean pixel density and accuracy for each block (day) of training. This subject can be seen to have improved both his speed and accuracy as a result of training.

2. Accuracy bias. Two subjects (PM and MM) displayed a bias towards improving their accuracy rather than trying to reduce the amount of visual information they received. (The first of these subjects might arguably have been included in the speed and accuracy group, since the amount of visual information he required was extremely low, even during the early part of training.) The performance operating characteristic for the second of these subjects is shown in Figure 2-5. It can be seen that, while there was little improvement in the mean pixel density he required, his accuracy increased substantially over the course of training.

3. Speed-accuracy trade-off. Two subjects (SC and SP) adopted a strategy of trading accuracy for speed of responding. Both of these subjects showed dramatic drops in average pixel density. One of these subjects (SP) showed a general drop in accuracy for both targets and foils. The other subject (SC) showed a decrease in accuracy for foils alone. Since foils constituted only 33% of the stimulus items, he evidently discovered that he could keep his overall error rate low by responding "Yes" or "Acceptable" to items when he did not have enough visual information to identify them. Since his mean pixel densities in the later half of training were 10 and below, it is quite clear that he tended to respond without attempting to read the stimulus

Table 5
Mean Percentage Correct Responses for Each Subject Given Defender Training (Mean Pixel Densities are in parentheses)

		Blocks 1 - 6			Blocks 7 - 11			
Subject	Reading Test Percentile	High Prob. Word	Low Prob. Word	Foil	High Prob. Word	Low Prob. Word	Foil	Strategy
PM	48	93 (24)	87 (24)	54 (28)	89 (25)	94 (25)	71 (29)	AB
SC	39	89 (24)	89 (24)	52 (25)	83 (10)	96 (9)	25 (9)	SAT
MM	33	93 (35)	89 (35)	73 (39)	92 (35)	90 (36)	87 (41)	AB
YW	32	83 (44)	76 (45)	78 (46)	92 (37)	82 (39)	81 (41)	S&A
LW	20	93 (51)	84 (52)	89 (53)	95 (38)	93 (40)	92 (42)	S&A
AP	18	91 (19)	93 (20)	26 (22)	94 (20)	73 (20)	49 (23)	S&A
JM	3	88 (49)	84 (49)	66 (50)	89 (33)	89 (34)	52 (34)	S&A
SP	3	93 (45)	89 (45)	67 (48)	82 (21)	82 (21)	41 (21)	SAT

* S&A = Speed and Accuracy Goals (4 Ss);
AB = Accuracy Bias (2 Ss);
SAT = Speed-Accuracy Trade-off (2 Ss).

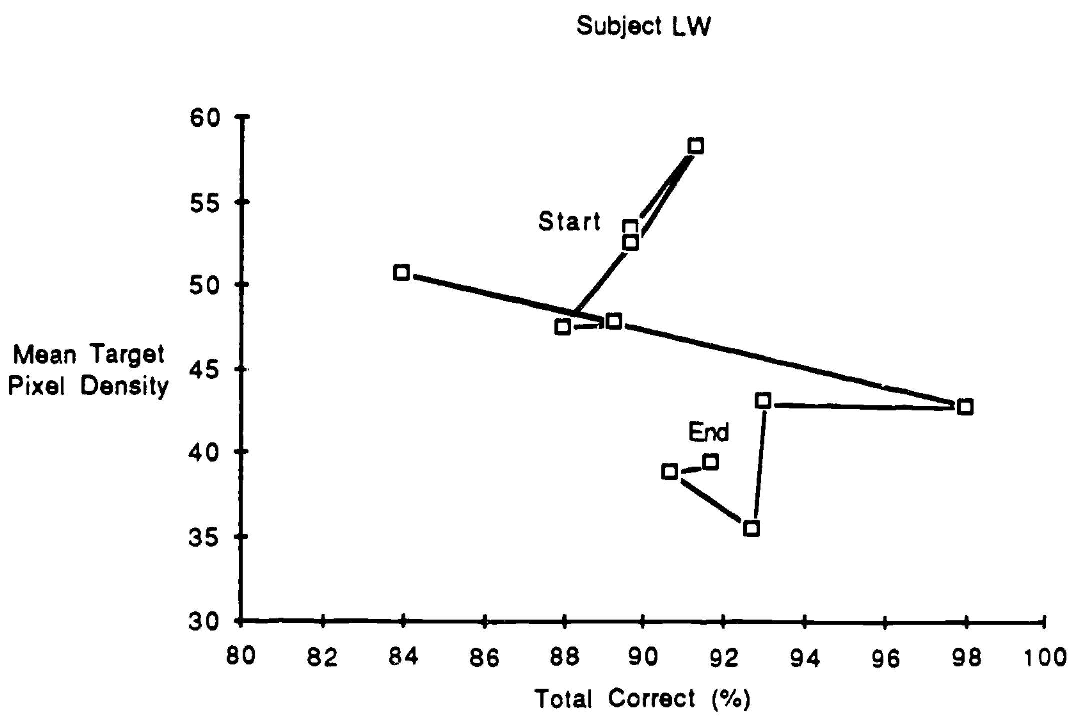

Figure 2-4: A representative performance operating characteristic for a subject who maintained speed and accuracy goals.

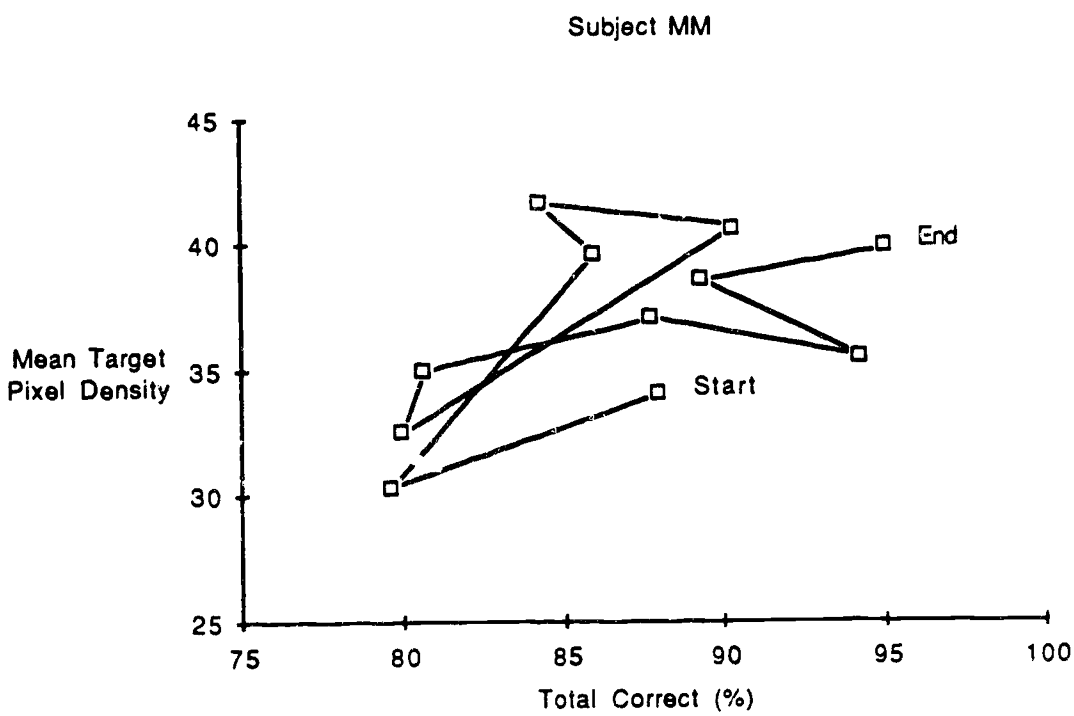

Figure 2-5: A representative performance operating characteristic for a subject who adopted an accuracy bias.

words. The performance operating characteristic for this subject is given in Figure 2-6.

In summary, while there were individual differences in strategy, the majority of subjects showed improvement in their performance, by decreasing the amount of visual information they required and/or by improving in their accuracy of responding. All but two subjects (YW and LW) had higher rates of errors in responding to foils than to target words. Thus, even though the Defender game did not penalize subjects for responding more slowly when the incoming ship (word) corresponded to a foil, subjects nonetheless continued to respond early, perhaps due to the pressures of the game. They did not realize that a strategy of waiting until the word was recognizable would be preferable, since it was only errors in judging foils, not speed of responding, that had an influence on their score.

2.2.2 The Transfer Task: Reference Mapping

The reference mapping task was administered to each subject before and after training. The task included the four text forms outlined in Table 5. These included two forms in which contextual and topical information supported the activation/selection of a single antecedent as referent, a form in which there was a conflict between activation based upon topic and that based upon context, and a form in which the contextual activation was ambiguous, supporting two antecedents. Repeated measures analyses of variance were carried out for each of three dependent variables: (1) the mean reading time per syllable for the final sentence of each text which contained a pronoun, (2) the percent of correct responses in supplying referents for that pronoun, and (3) the mean vocalization onset latency in supplying referents for pronouns.

Reading times. The mean reading times for sentences containing the pronoun are shown in Figure 2-7 for each text form, before and after Defender training. In evaluating our hypotheses, a series of planned comparisons were made of (1) differences among text forms prior to Defender training, (2) the effects of training for each text form, and (3) differences among text forms following training. In the first comparison, we found no significant differences among the four text forms prior to training. The means for the text two text forms in which the referent was currently a

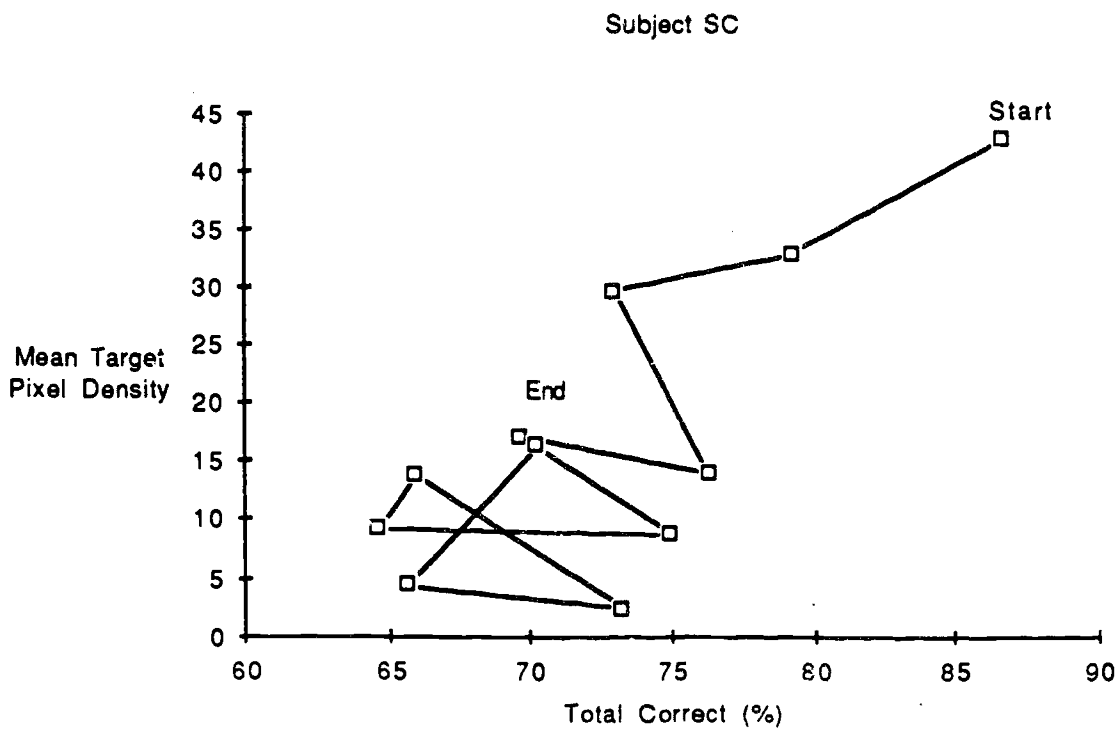

Figure 2-6: A representative performance operating characteristic for a subject who adopted a strategy of trading speed for accuracy.

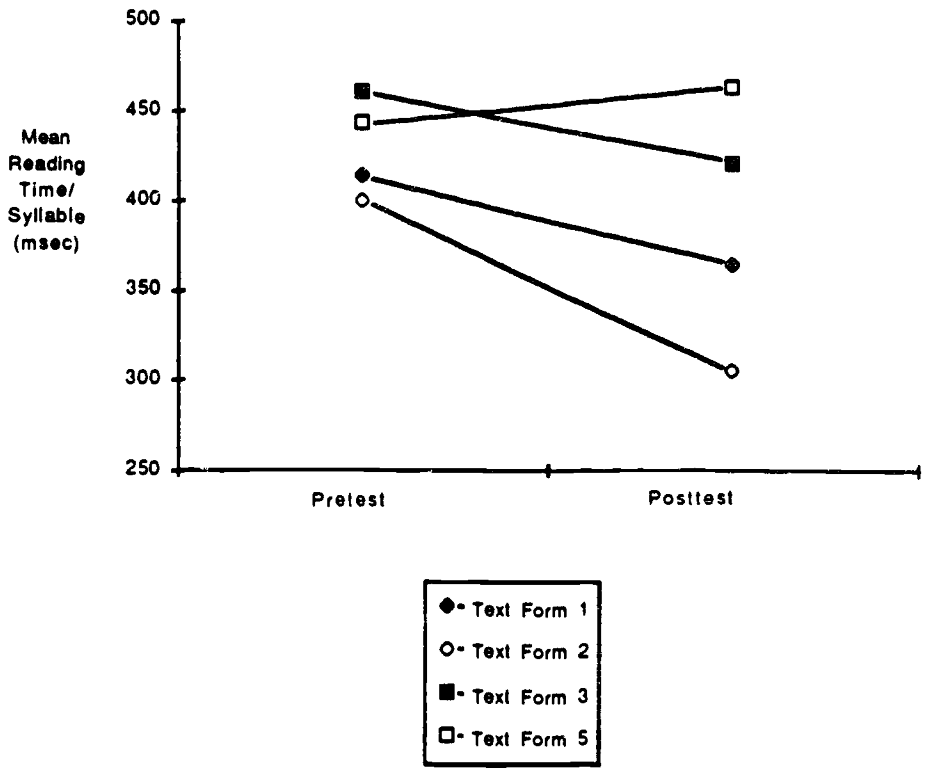

Figure 2-7: Mean reading times obtained in the reference mapping task before and after Defender training.

topic (413 msec/syllable for text form 1 and 400 msec/syllable for text form 2) were smaller than those for the text forms in which the referent was not a topic (460 msec/syllable) or was ambiguous (442 msec/syllable), but not significantly so. These mean reading times were all substantially longer than those obtained in the earlier experiment by a group of generally higher ability readers, which ranged from 235 msec/syllable to 327 msec/syllable. Thus, prior to training the subjects in the training experiment showed no evidence of an efficient, automatic process for tracing referential relations.

In the second set of comparisons, we tested the hypothesis that there would be significant improvements in reading times as a result of Defender training for text forms (1 and 2) in which the activation of the correct antecedent, on the basis of contextual and topical information, was substantially greater than that of the alternative antecedent. In the analysis, there were significant or nearly significant decreases in reading times for both text forms 1 and 2 as a result of training (t_7= 2.14, p=.03 for text form 2, and t_7= 1.55, p=.08 for text form 1). The mean decreases in latency for these text forms were 49 msec and 96 msec, respectively. In contrast, there were no significant effects of training on reading times for the other two text forms, for which the net levels of activation due to context and topic do not discriminate clearly between the two antecedents. The effects of Defender training were thus confined to those text conditions for which the mechanism of semantic activation could be applied, as was hypothesized.

Finally, in the last of the planned comparisons we tested for differences in the posttest reading times among the four text forms. There were significant differences among text forms in this analysis, ($F_{3,21}$=3.95, p=.02). In pairwise comparisons, we found that there were significant differences among mean reading times for all text forms except forms 3 and 5. Moreover, the ordering of these posttest means followed the predicted pattern. Text form 2, which provided the clearest distinction in activations among antecedents, had the fastest reading time (304 msec), followed by that for text form 1 (364 msec), and finally those for text forms 3 (421 msec) and 5 (463 msec), for which the differential activation of the two antecedents was hypothesized to be minimal. Note finally that text form 5 contains an ambiguous context, and that an increase in subjects' reliance on context-based activations might be expected in this case to cause even greater difficulty in mapping referents. An increase in mean reading time for this text form did occur, although it was not statistically significant.

Accuracy. The second set of analyses we carried out were on the subjects' mean percentages of correct referents supplied by subjects. These results are presented in Figure 2-8. There were again significant or nearly significant improvements in performance as a result of training for text forms 1 (t_7=2.36, p=.02) and 2 (t_7=1.67, p=.07), but not for the other text conditions. In addition, in both the pretest and posttest, there were significant differences among text forms ($F_{3,21}$=12.20, p<.001 for the pretest, and $F_{3,21}$=18.80, p<.001 for the posttest). Performance for text form 5, in which either of the antecedents was scored as correct, was uniformly high (above 90%), while that for text form 3, in which there was a conflict between topical and contextual information, was close to 50% correct. Training in the use of context thus does not appear to reduce errors in comprehension due to a failure in reference assignment for cases where the referred-to antecedent is low in the topical structure. It is interesting to note that writers generally appear to avoid pronominal reference and instead use forms of lexical reference when the antecedent they are referring to is distant within the text (cf. Biron, Kellogg, Posner, & Yee, 1985, Table XII). Such a practice also appears to be warrented when the referent is low in the topical structure.

Response latencies. Finally, an analysis was carried out of subjects' latencies in reporting referents for pronouns. Differential increases in reaction times for alternative text forms would suggest that subjects may under some conditions be postponing their assignments of referents until after they have indicated that they have completed the sentence. Such a result could complicate the interpretation of reading times we have presented above. The mean response times are presented in Figure 2-9. While there were marginally significant decreases in response times for the four text conditions as a result of training (t_7=1.65, p=.07 for text form 1, t_7=1.37, p=.11 for text form 2, t_7=1.41, p=.10 for text form 3, and t_7=1.61, p=.08 for text form 5), there were no significant differences among the four text forms in either the pretest or the posttest. The reductions in reading time that occurred following training suggest that subjects are increasingly trying to make reference assignments for pronouns while they are reading the critical sentences.

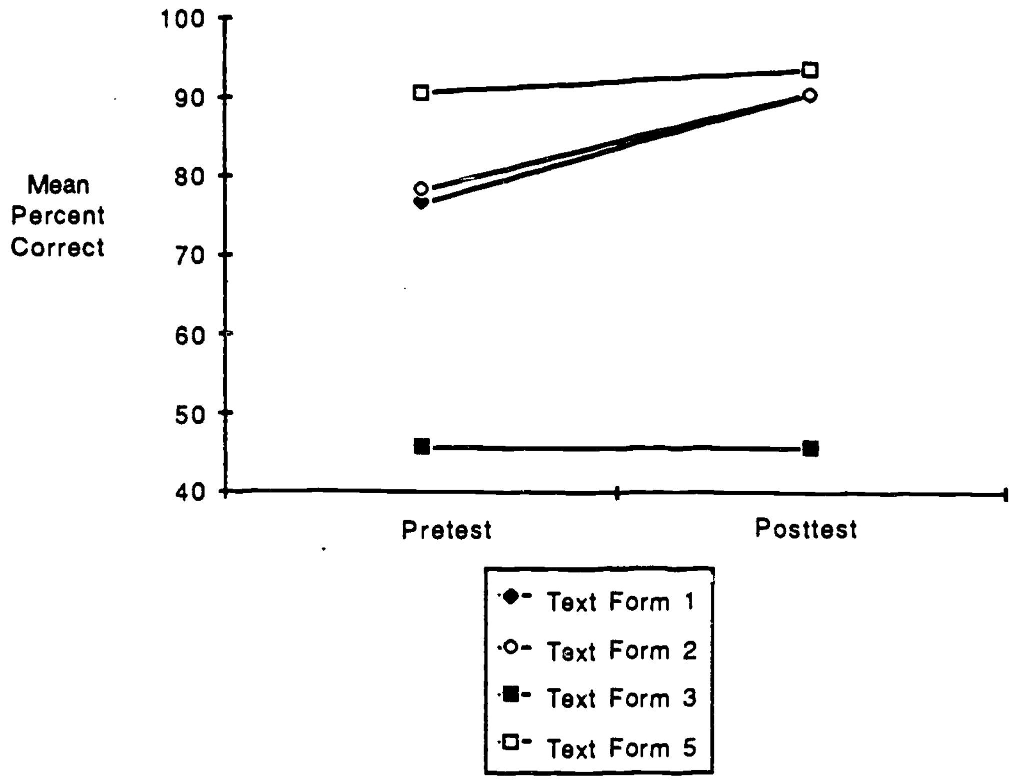

Figure 2-8: Mean percentages of correct referent assignments obtained in the reference mapping task before and after Defender training.

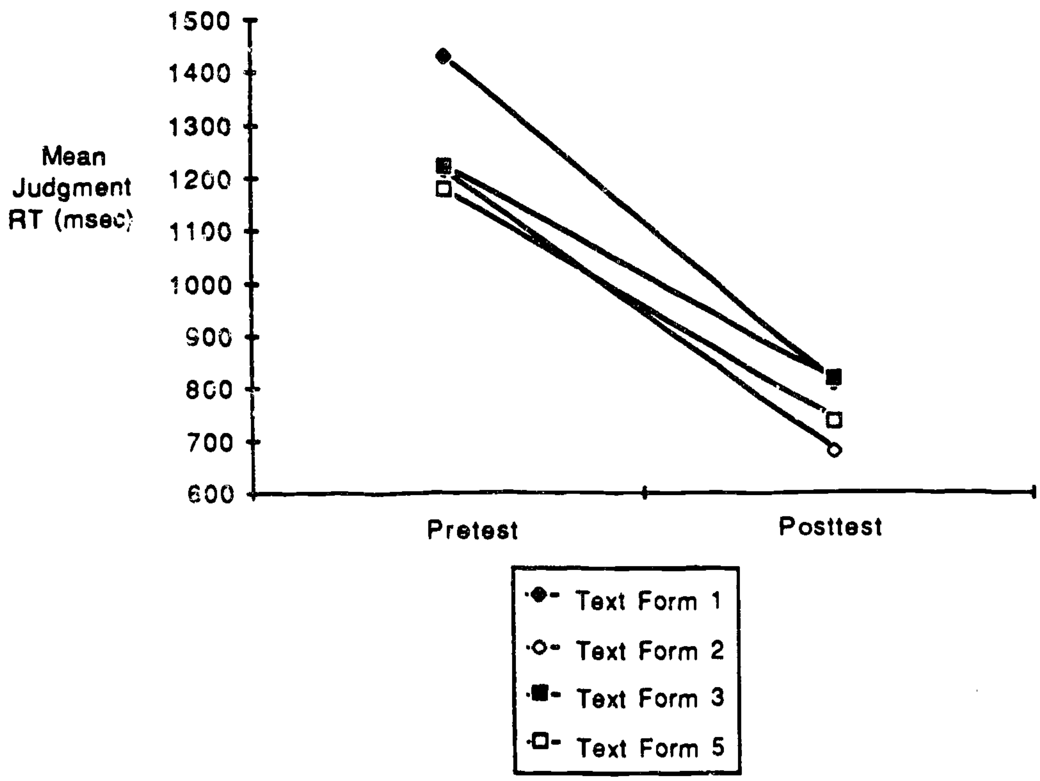

Figure 2-9: Mean judgement times obtained in the reference mapping task before and after Defender training.

2.3 Discussion

The results of the training study provide evidence for transfer from the training of a component skill to the performance of a reading task involving an important aspect of comprehension -- the successful mapping of referential relations. As in our earlier work (Frederiksen, Warren, & Rosebery, 1985a, b) which focused on the transfer of perceptual and decoding skills, the patterns of transfer in the present study followed lines predicted on the basis of an interactive reading theory. Such a theory emphasizes the lines of enablement among processing components of reading, with particular attention paid to their contribution to the efficiency of reading (Frederiksen & Warren, 1986). As in our previous work, the goal was to use training studies to verify specific theoretical predictions. In the present instance, a model of the use of parallel processes of activation within semantic memory for tracing referential relations was developed and used to predict (1) what critical enabling skills could be expected to contribute to such a mechanism for reference tracing, and (2) under what conditions such enablements should appear.

The evidence supports the view that parallel processes of activation within semantic memory can mediate an important comprehension skill (reference tracing) under realistic textual conditions, and that they are therefore useful skills to acquire. The evidence also suggests that there are textual conditions in which automatic processes for reference mapping are not applicable, and that these conditions should be avoided in writing clear prose. For example, from the standpoint of text design, when pronouns are employed they should be presented in a nonambiguous context. When a writer seeks to refer to an antecedent that is not currently a topic, he or she generally should use lexical substitutes or argument repetition in order to avoid the potential comprehension problem created when there is a conflict between topic and context in the assignment of a referent.

While transfer of skill acquired in the use of context was found to take place, from the standpoint of training we would like to maximize the impact of such a skill on the performance of other reading activities, such as reference tracing and understanding high order relations. A learning strategy in which components are first trained individually and then integrated within the context of a "whole task" environment has been found to be effective in other domains (Frederiksen & White, in press). Thus, it would be worthwhile to follow training on a skill component such as that addressed in Defender with further training in a practice environment in which the application of that skill to other reading situations could be accomplished. The reference-tracing task represents one such potential training task for integrating component performance within a task environment involving high level comprehension. The Defender game should thus be regarded as one in a series of componential training tasks focusing on critical enabling skills in reading and on their integration within a general reading context. Such an approach has been followed by Schwartz (1986), whoe trained primary school children using three computer games, including letter matching, word/pseudoword matching, a word/pseudoword pronunciation task, and, importantly, an integration task in which the skill components could be applied in a sentence reading context in which both speed and accuracy are required (here, in identifying if the last word is or is not anamolous). They found significant improvement in comprehension scores (as measured by a CLOZE text) for students who initially scored below the median on the reading comprehension test. Moreover, these gains were significantly greater than those for similar students who were given training using DISTAR (Science Research Associates, 1983). Our conclusion is that, to be most effective, a componential approach to developing reading skills must address both the development of individual components and their integration.

The development of skills for the contextual activation of concepts and the application of such skills in the automatic assignment of pronominal referents could have a considerable impact on subjects' ability to comprehend text, since a correct and effortless assignment of referents is important in understanding cohesive texts and may also be important in facilitating the understanding of high order relations within a text, as has been argued elsewhere (Frederiksen & Warren, 1986). Analyses we have carried out of clauses within a text that are linked through high order causal, temporal, or conditional relations revealed that such clauses almost invariably share referents (usually multiple referents). As Kintsch and van Dijk (1978) have pointed out, in order to analyze such high order inter-clause relations, the relevant prior clauses must be reinstated into working memory. Reinstatement constitutes a necessary, if not sufficient condition for understanding the high order relations among clauses. Thus, one mechanism for reinstatement is that of reference tracing. While Kintsch and van Dijk consider primarily texts containing repeated antecedents, the mechanism we have described for automatic mapping of referents provides a more general process which can apply to anaphoric terms other than repeated lexical items. These automatic processes for tracing reference relations can therefore lead to the reinstatement of clauses that are likely to be linked through higher order semantic relations to the clause currently being processed by the reader. For these theoretical reasons, we feel that additional training focusing on the analysis of referential relations within a text represents a prime candidate for further instructional research.

3. EVALUATION OF COMPONENT-BASED TRAINING FOR BILINGUALS

The bilingual study we have undertaken seeks to answer two research questions: (1) Which reading skill components are language-dependent systems?, and (2) How effective are training systems which focus on particular components for developing reading skills of bilingual trainees? The general method used to address each of these questions is based upon training studies in which individual skill components are developed through the use of computer training environments that focus on particular reading skills. The language dependence/independence issue is addressed by measuring performance on reading components before and after training using linguistic materials drawn from two languages. If training effects transfer to both languages, we will have evidence that the skill developed is a language independent skill. If training effects are apparent only for the trained language, the evidence supports the hypothesis of language dependence. The degree of improvement in performance on component-specific tasks will provide evidence concerning the effectiveness of training in a bilingual population.

Recent studies of skill deficiencies in bilingual subjects' performance in reading within their primary or secondary language have found that such subjects show differences in the availability of automatic process to aid in word recognition. Differences have been found, for example, in subjects' ability to take advantage of semantic context (Favreau & Segalowitz, 1984; Duran, 1985). Favreau, Komoda, & Segalowitz (1980) have implicated subjects' sensitivity to orthographic as well as syntactic redundancies of the language. Oller & Tullius (1973) have found differences in fixation durations for textual materials from the primary and secondary language for which comprehension is equivalent, which suggests that there are differences in the time to process orthographic information. Duran (1985) has advocated for bilingual subjects training focusing on sharpening discrimination and speed of discrimination of word features in English which might be confused with word features stemming from the non-English language. In the present study, we have attempted to take a step in this direction by evaluating within a hispanic bilingual population two training systems that focus on the development of automatic skills in encoding orthographic information and the decoding of such information in word recognition.

The work we have carried out in support of the bilingual study centers on three groups of activities: (1) Development of computer training environments that focus on individual reading components; (2) Development of criterion measures of reading skill components in two languages; and (3) Experimental studies of the effects of training particular components on the performance of criterion tasks presented in two languages.

1. Development of computer training environments. Attention was devoted to improving two prototype training systems developed in our earlier ONR contract (Frederiksen, Warren, & Rosebery, 1985a,b). In particular, we have modified the training system dealing with perceptual encoding skills, SPEED. The modified system has a built-in instructional monitor and provides better graphics and feedback concerning performance. We have also modified a second system, RACER, which focuses on the decoding skill component. The enhanced system makes extensive use of computer-generated speech in order to provide feedback concerning word pronunciations. It also incorporates enhanced graphics and includes an instructional monitor. Finally, training materials have been developed aimed at developing knowledge of particular phonic principles. The addition of such materials was found to be important in training very low ability readers in our earlier study, and was judged to be essential if the training technique was to be successful with subjects whose native language was not English.

2. Development of criterion measures of reading skill components. Criterion measures developed in the previous ONR work have been modified for administration using the IBM personal computer. In addition, Spanish versions of each task have been created in support of the research plan for evaluating the language dependence/independence of skills that are trained. These criterion tasks include (1) detecting multiletter units within words, and (2) decoding words of varying difficulty.

3. Experimental studies of skill transfer. A training study has been carried out using bilingual subjects. Transfer of skills developed during instruction is evaluated by administering criterion tasks before and after training. Such transfer studies have been completed for the SPEED and RACER training systems.

Based upon our earlier experimental evaluations of SPEED and RACER (Frederiksen, Warren, & Rosebery, 1985a, b), hypotheses can be stated concerning the language specificity of the skills developed in each game. The SPEED game was found in this research to support development of a parallel encoding of multiletter units in which attention is distributed across the array of letters of a word. As a result of training, multiletter units embedded within words were detected as readily as those occurring at the beginnings of words. Moreover, the effects of training were found to transfer fully to test units that were never actually presented to subjects during training, which strongly implies that a general encoding skill had been developed. Finally, there was evidence for the transfer of training to a reading task (an oral reading task), suggesting that the skill developed is a general one enabling of other skills. For these reasons, our hypotheses for the present study are that (1) bilingual subjects will show substantial skill development when trained using units and test words from their non-native language, and (2) their training will transfer, not only to English units not included in training, but also to Spanish units occurring within Spanish words.

The evidence developed in our earlier study of RACER (Frederiksen, Warren, & Rosebery, 1985b) supported the interpretation that the skill developed was an automatic decoding of orthographic forms into their phonological correspondents. The strongest evidence for this interpretation was that subjects' decoding latencies for words and pseudowords that were matched in orthographic form were equivalent after training. It was also found that the difference in decoding latencies for one- and two-syllable items were essentially eliminated as a result of training. Our hypothesis for the current study is therefore that, since the skill development involves learning language-specific associations/rules relating orthographic units to pronunciations, the effects of RACER training should be limited to the language of training. By including materials focussing on particular phonic principles, we can expect hispanic subjects to be successful in acquiring such skills within the language of training, English.

3.1 Method and Subjects

3.1.1 Subjects

Hispanic bilingual subjects with English as a second language were identified for testing on the basis of recommendation from Bilingual specialists at several local high schools. We sought subjects whose reading level in Spanish was average and whose reading level in English (as measured by the Gates-MacGinite Reading Test) was at or below the 30th percentile. To assess Spanish reading ability, the Prueba de Lectura, Nivel 3, Nivel 4 (Guidance Testing Associates, 1962) was administered. This test did not serve as a criterion for selection but as a baseline measure of subjects' skill in reading Spanish. A total of 11 subjects were selected, all of whom participated in the SPEED evaluation study, and 9 of whom participated in the RACER evaluation. Scores of these subjects ranged from the 1st through the 30th percentile on the Gates-MacGinitie, with a median score at the 7th percentile. Their scores on the Prueba de Lectura ranged from the 27th to the 67th percentile, with a median at the 57th percentile. Nine of the subjects were female and 2 were male.

3.1.2 The SPEED Training Task

3.1.2.1 Game Format and Design Specifications

The perceptual units training system is based upon the SPEED game used in previous research (Frederiksen, Warren, & Rosebery, 1985a). A diagram of the computer screen for the new implementation of the SPEED game is given in Figure 3-1. In this game, subjects are presented target multiletter units (e.g., IST) and then confronted with a series of rapidly occurring words within a window, some of which contain the target unit, some of which contain a similar-appearing unit (called a similar foil; e.g., INT), and some of which contain no such similar-appearing unit (called a dissimilar foil). The subject's task is to indicate with a button press whether or not each test word contains the target unit. The dynamics of the game are determined by the subject's performance. Initially, stimulus words are presented at a moderate rate of 60 words per minute (wpm). Each time a player responds correctly to a word, the rate at which the words are presented is increased by 2 wpm, and each

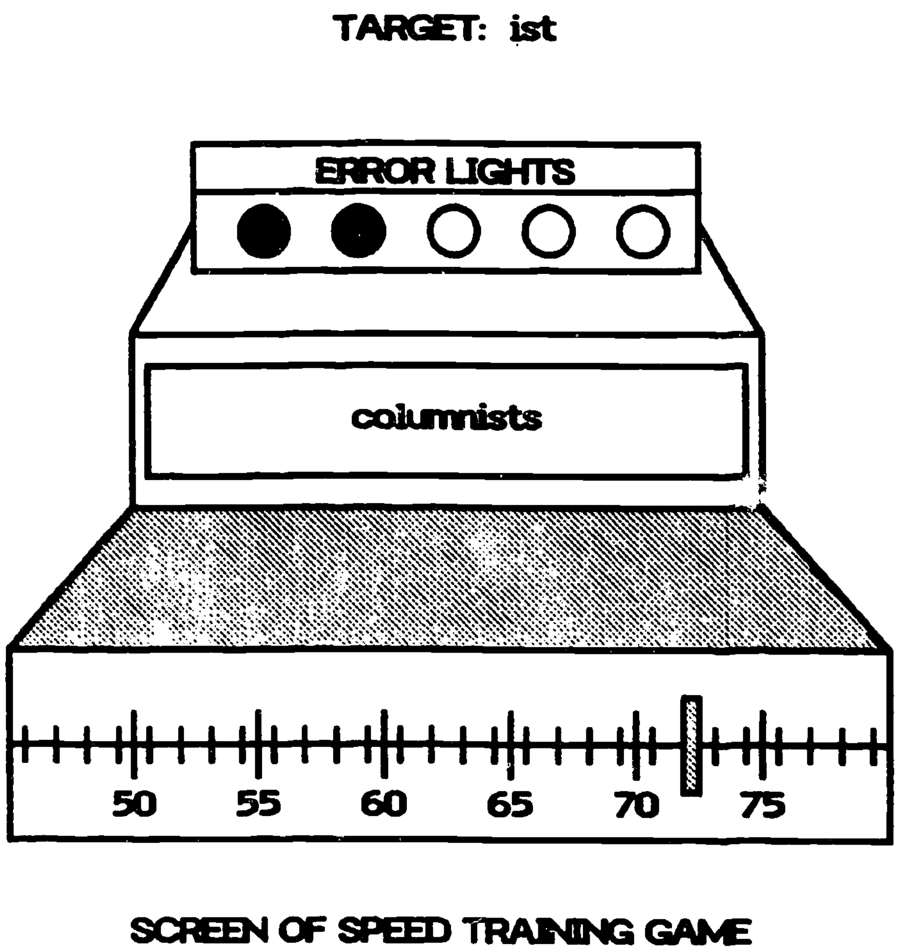

Figure 3-1: A representation of the computer screen for the new implementation of the SPEED game.

time the response is incorrect, the rate of presentation is decreased by 2 wpm. At the bottom of the screen there is located a speedometer which displays the current rate of presentation of test words in the window. Initially, the speed is set at the low end of the speedometer. The goal of the game is to correctly identify the presence/absence of target units often enough to accelerate the rate of presentation until it has reached a goal speed (the speed at the top of the speedometer), which is 50 wpm above the starting speed. Above the stimulus window are five error lights. Each time an error is committed, a light comes on. A maximum of five lights can be on, and any error committed while there are five lights on results in a "crash" -- the end game run. However, whenever the subject makes a correct response, a light on the speedometer is turned off, negating one of his or her previous errors. The error lights thus act as a warning to the subject that it is a good idea to slow down a little until errors are under control. The subject is in these ways placed in a speed-accuracy bind, in which he or she must increase the rate of correct responding in order to accelerate the speedometer, while at the same time not allowinging too many consecutive errors. Further details of the game format are described in the earlier report (Frederiksen, Warren, & Rosebery, 1985a). The main change in game format incorporated in the new implementation concerns the events that occur when the subject reaches the goal speed on the speedometer. In the new version of the game, a run on a unit no longer necessarily ends when a trainee reaches the designated goal speed. The subject now has the option of extending the run beyond that goal speed, in order to see how far it is possible to go before six consecutive errors are made. Each time this option is taken, the goal speed on the speedometer is increased by 10 units. Note finally that when five errors occur after a goal speed has been reached, it is not considered to be a crash.

3.1.2.2 Sequencing of Materials

A set of rules has been incorporated into a monitor program incorporated within the SPEED system, in order to automate instructional decisions, thereby making SPEED instructor-independent.

In the SPEED system a hierarchical series of instructional decisions must be made: 1) what units should be presented at any give time; 2) how many training runs should be held for each unit; and 3) what starting and goal speed should be employed on a given run. Associated with each of these decisions is a performance criterion: one for determining when to raise or lower starting and goal speeds from run to run on a unit, one for deciding when to reintroduce a given unit, and mastery criteria for determining when to drop a unit from the training set and when to terminate training itself.

1. Selection of units. In the monitor, six units are maintained in an Active Instructional Group (AIG) at any give time. Selection of the initial set of units for the AIG is based on stratified random sampling without replacement from a pool of 42 units, the stratification ensuring a proportional representation of easy (prefix) and more difficult (non-prefix) units within the AIG. When a unit is deleted from the AIG, a new unit from the same subgroup (prefix, non-prefix) is randomly selected to replace it.

Training runs for units currently in the AIG are presented in a random "round-robin" fashion: that is, the current six units are randomly ordered and presented as one training block of six runs. A new random ordering is developed for the next training block, etc. When a unit is deleted from the AIG, its replacement will be substituted in the subsequent training block. Trainees may opt to discontinue training after any run and the monitor will begin the next session at the point at which training was discontinued.

2. Setting rates of unit presentation. The starting speed for the initial run of the first unit to be introduced is arbitrarily set at 60. The starting speed for the initial run of the second unit is based on the weighted mean RT for correct responses in the initial run of the first unit. The starting speed for the initial run of the third unit is based on the weighted mean RT for correct responses in the initial runs of units 1 and 2. The number of units contributing to the mean RT is incremented with each successive run until a maximum of 4 such units is reached. From then on, the starting speed of each newly introduced unit is based on the weighted mean RT for correct responses in the initial run for the 4 immediately preceding units.

The starting speed for the second and subsequent runs of an individual unit are calculated from the highest speed attained in the preceding run of that unit in the following way: The goal speed for a subsequent run of a unit is 20 speedometer units above the highest speed attained on the preceding run of that unit. The starting speed is 50 units below the goal speed.

3. Criteria for terminating training. The mastery criterion for terminating training on a unit is set initially at 130. This is be adjusted upwards as trainees become more experienced with SPEED following a rule based upon the mean number of runs needed to achieve the criterion for the 4 previously introduced units. The criterion is be increased by 20 whenever the mean number of runs needed to reach criterion drops to 2 or less.

Training on the SPEED system is terminated when a speed of 150 or better is achieved on the initial run of 5 consecutively introduced units.

3.1.3 The RACER Training Task

The decoding training system is based upon the RACER game used in previous research (Frederiksen, Warren, Rosebery., 1985a). In re-implementing the game on an IBM personal computer, changes were made (1) to the game format and design specifications, so that feedback about pronunciation immediately follows a trainee's pronunciation response; (2) to the set of materials, in order to expose students to a set of basic phonic principles in addition to the sequence of graded materials used in the old RACER system that cover a range of difficulty; and (3) to the scoring rules, to ensure that students sufficiently master the materials at one level of difficulty before going on to the next. Finally, a tool for displaying student performance within and across levels of difficulty was created.

3.1.3.1 Game format and design specifications.

The "playing board" for the game is shown in Figure 3-2. It consists of a race track displayed across the bottom of the screen on which two runners, the trainee (or player) and the Computerman compete. The goal of the game is for the player's runner to cross the finish line first. The "track" is divided into 20 steps, each corresponding to a word the player must read and pronounce. The player advances his runner by pronouncing each word, which moves the runner one step forward on the track. The Computerman runs at a constant rate based on the player's mean race

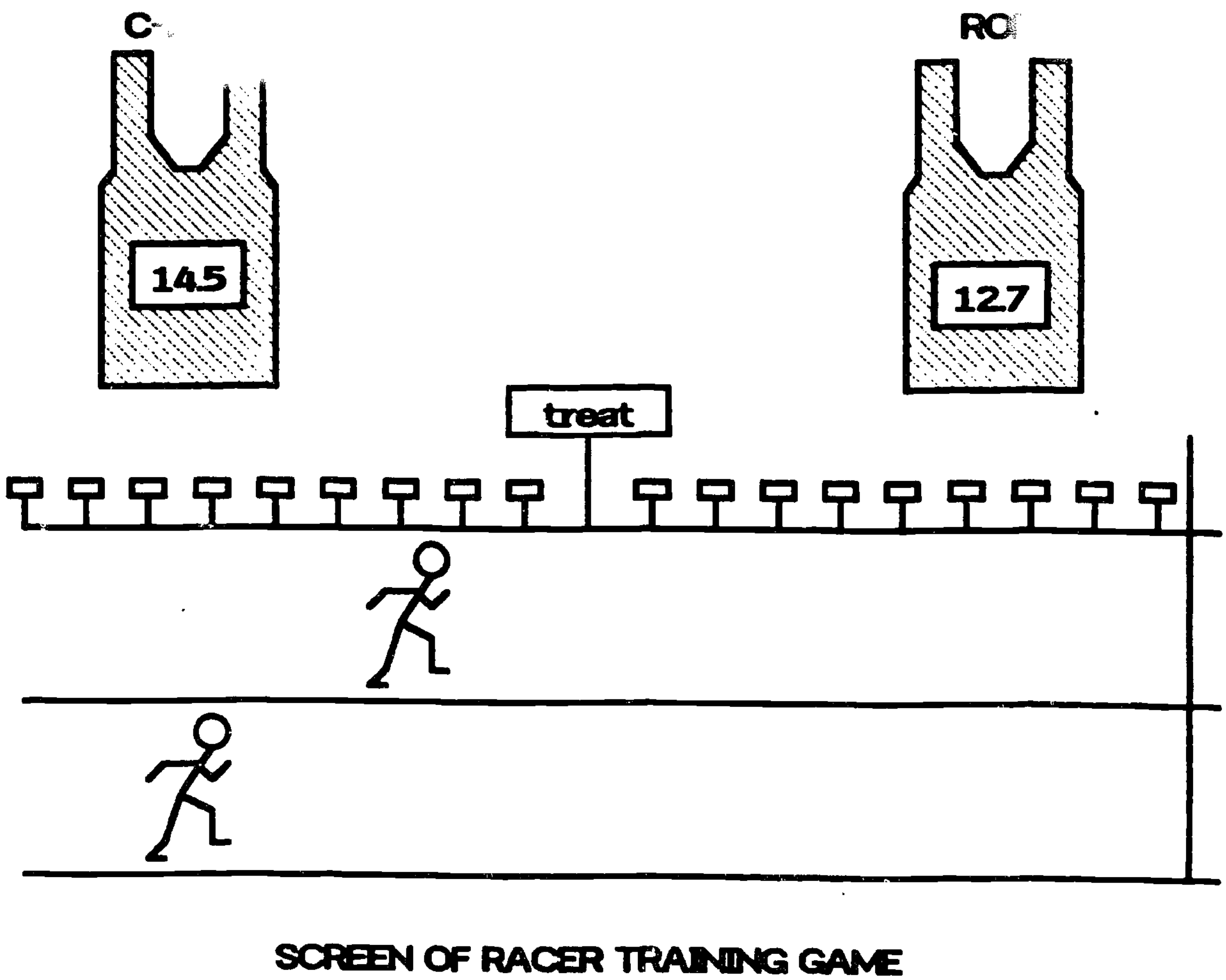

Figure 3-2: A representation of the computer screen for the new implementation of the RACER game.

time over the last three races. If the player pronounces a word faster than the Computerman, his/her runner will gain on the computerman. If the player is unsure of a word and spends excessive time decoding and pronouncing it, the Computerman will gain on the player's runner. Thus, to win the game, the player must respond more quickly than his or her own average rate of responding over the last three games played. The race ends when one of the runners crosses the finish line.

There are twenty pre-set flag locations along the race track. As the player's runner approaches each location, a flag pops up displaying a word that the player must pronounce into a microphone. The word is displayed until the trainee initiates a pronunciation. At this point, the word is masked for 250 msec, and then the display is erased and the sequence of words continues. Thus, only one flag and stimulus word appear on the screen at a time. Since words appear in a linear sequence in close proximity to one another on the track, the player's eye remains close to the "action" of the runners. Also displayed at the top of the screen are two running jerseys, one that shows the student's current mean race time (the time he is racing against) and one that counts off the elapsed time of the race currently underway. The race is accompanied by sound effects that reflect its pace.

Randomly (up to 5 times) throughout the race and immediately after a trainee's pronunciation, the action is momentarily stopped and a word is presented aurally via the Intex Talker speech synthesizer. The player must indicate with a key press whether the word he or she hears is the one just read or a sound-alike distractor. If the player responds correctly, his or her runner advances to the next flag location. If the player makes an error, the computerman advances one step, while the player's runner remains at the current location and the next word appears at that same location. After the player has read the initial 20 race words, any words for which there has been an error will reappear, in order to give the player additional practice on troublesome items. Thus, students receive immediate feedback about pronunciation accuracy and can discover the origins of decoding errors. In this implementation of RACER, as in the last, students are encouraged to be both efficient and accurate in their pronunciations.

Finally, a "Request Pronunciation" option has been incorporated to discourage players from guessing at unfamiliar words. To hear an unknown word, the player can press a key while the word is displayed. At this point, the ra is momentarily suspended, the displayed word is outlined on the screen to indicate that the race has stopped, and the word is pronounced by the Intex Talker. When the outline disappears, the player must pronounce the word he has just heard as quickly as p ble

3.1.3.2 Sequencing of Materials.

Ov a ser of aces, the words used in the RACER game are sequen ed in their difficulty decod accomplished by creating a series of dictionaries for use by RACER that represent va ety of decoding principles, word lengths, and frequencies. The words used in a given RACER run could be sampled from a single dictionary, or they could be sampled, in designated proportions, from two or more dictionaries. Students began RACER training with words drawn from a single dictionary, which contains words of a particular form chosen to illustrate and provide practice with a particular decoding principle. The six dictionaries used in this sequence contain one-syllable words representing each of the following basic phonic principles: (1) simple short vowels, (2) simple long vowels and silent -e markers, (3) vowel digraphs, (4) r-controlled vowels, (5) initial consonant blends, and (6) terminal consonant blends. Following these dictionaries focussing on particular phonic principles, subjects continued their RACER training with words of graded difficulty that were drawn from two or more dictionaries. The set of dictionaries[8] from which words were drawn included, in addition to the six dictionaries containing one-syllable words, seven additional dictionaries made up of words of high, moderate, and low frequency, having lengths of either two, three, or four syllables[9]. For example, subjects were first given a mixture of 20% high frequency, 2-syllable words, 50% 1-syllable words having initial consonant blends, and 30% 1-syllable words with terminal consonant blends. These proportions were next changed to 60% high frequency, 2-syllable words

[8] Each dictionary entry is accompanied by a pronunciation code for use by the Intex-Talker. These codes were obtained from a DEC Talk speech synthesis device, and then translated to a code useable by the Intext Talker. Altogether, there are 13 dictionaries, containing approximately 8000 words in all. Maintaining separate dictionaries allows an instructor the option of specifying the sampling of frequencies of words from various dictionaries to be used in a RACER game at any particular point in training.

[9] For four syllable words, only the high frequency level was included.

and 40% 1-syllable words containing the two sorts of consonant blends. The sequence of materials continued, slowly increasing the proportions of longer and lower frequency words. The total number of levels, including the initial six levels containing "pure" examples of phonic principles, was 27, with the last level made up entirely of 4-syllable words of high frequency.

The student's progress through the levels of materials is determined by his or her performance. A student is considered to have "mastered" a given level of difficulty when he or she wins four of six consecutive races. When this occurs, the instructional monitor allows the student to move on to the next level of difficulty.

At the end of each race, the student is presented with the "win window" screen. This screen displays the student's best time to date, the time of his/her last race, the number of races run at a particular difficulty level, the number of races won at that level, and the "win window" itself. The "win window" represents the outcomes of the last six races the student has completed with a set of lights. Each race is represented by a "light", which is "on" if that race was won. In order to advance to the next level of difficulty, a student must win four of six consecutive races. This means that four of six lights in the window must be lit. Thus, by looking at the window, a student knows exactly what must be done to advance to the next level. When a level of difficulty is completed, the screen flashes and musical sound effects are played for the student. After each race, the trainee has the option of continuing training, viewing his performance record to date, or stopping training for the day.

The program graphically displays a student's performance (1) within a given level of difficulty or (2) across the entire sequence of training. By pressing a key, the student can scroll to the left (to see races run early) or to the right (to see races run recently) and thus examine his or her performance over a series of races.

3.1.4 Evaluation Tasks

In addition to the standardized reading measures administered at the beginning of the experiment, the following criterion measures were developed for use in evaluating skill acquisition within each of the games. Both English and Spanish versions of each criterion task were developed, and for each version, two alternate forms were developed to allow repeated testing before and after training.

3.1.4.1 The Unit Detection Task

The unit detection task, administered before and after SPEED training, was used to assess improvement in perceptual skill components resulting from SPEED training. Subjects were tested using both Spanish and English versions of the task. The unit detection task is similar to that used in an evaluation of an earlier version of the SPEED game (Frederiksen, Warren, & Rosebery, 1985a). On each trial, a multiletter target unit (e.g., COL) is identified in advance for a subject, whose task is to monitor for its presence in a series of 32 stimulus words that are rapidly presented within a window on the screen of an IBM personal computer. The target units and stimulus words are either English or Spanish, depending upon the version of the task. Within the series of stimulus words presented, 20 words contain the target unit while 12 do not (6 contain a unit similar to the target, and the other 6 contain no similar units). When a target unit is present within a stimulus word, it occurs equally often in an initial or medial position within the word. There are in all 24 target units represented in each version of the task. In the English version, half of the target units are ones that have received training in the SPEED game, and half are not. In the Spanish version of the task, a third of the units are similar to English units that receive training, a third are similar to English units but not those that receive training, and a third are not at all similar to English units (e.g., LLA). Finally, in both versions the units very in length (2 and 3 letters).

A trial begins with the identification of a target unit. The subject's task is to decide for each stimulus word that appears whether on not it contains the target unit and press the appropriate response key. Stimulus words are presented in a random order. The trial for a subsequent stimulus word does not start until the subject has completed a response to the previous item. Each word trial begins with the presentation of a blank screen for 1000 msec. The stimulus word is then displayed for 200 msec and is immediately followed by a mask that is also displayed for 200 msec. The screen then remains blank until the trainee responds. The subject's reaction time is measured from the onset of the stimulus. The accuracy of response is also scored.

3.1.4.2 Pronunciation Task

English and Spanish versions of the word pronunciation task were administered before and after RACER training, in order to evaluate improvement in decoding skill resulting from such training. The pronunciation tasks are designed to measure the speed and accuracy with which a subject pronounces each word in a series of test words which differ in their difficulty of decoding.

The English version of the task is similar to that used in our prior evaluation study (Frederiksen, Warren, & Rosebery, 1985b), adapted for use on the IBM PC. The list of words contains 76 items representing 19 orthographic patterns common in English (these include: CVCC, CVCE, CVVC, CCVC, CVCCC, CVCCE, CVVCC, CCVCC, CCVCE, CCVVC, CVVCCC, CCVCCC, CV-CV, CV-CXX, CVC-CV, CV-CVXX, CVC-CXX, CVV-CXX, and CCV-CXX, where C stands for a consonant, V for a vowel, E is the letter "e", X stands for any letter, and a dash represents a syllable break). For each orthographic form, high and low frequency words were equally represented (Carol, Davies, & Richman, 1971). Thus, the words vary in syllabic length and frequency class as well as in the types of vowels and consonant blends they contain. Two versions of the task were used, one as a pretest and the other as a posttest.

The subject's task is to pronounce each word as as soon as he or she can, as it is displayed on the monitor screen. Each response is judged for accuracy by an experimenter at the time it is pronounced, and the accuracy score is entered into the computer. Each word trial begins with an arrow focussing the subject's attention at the appropriate point in the display. After 1000 msec, the stimulus word appears for 250 msec. The trial for the subsequent word only begins after the subject has responded and the experimenter has entered the accuracy information. Reaction times are recorded from the onset of the stimulus word to the onset of the subject's vocalization.[10] To ensure accuracy in comparing latencies for words of different forms, words were matched across forms on their initial phonemes.

The Spanish version of the word pronunciation task is similar to the English task

---

[10]Vocalization latencies were measured by examining information at the cassette port of the computer, to which the amplified speech signal was directly connected, and testing for the presence of a sustained input.

in all respects except for the forms of stimulus words employed. Six orthographic patterns that occur regularly in Spanish were identified. These patterns include the following: CVCV (e.g., beso), CVVCV (e.g., bueno), CCVCV (e.g., bledo), CVCCV (e.g., barco), CVVCCV (e.g., biarca), and CCVCCVC (e.g., blindar). These patterns represent various combinations of simple vowels, vowel diphthongs, single consonants and consonant blends. Based upon these patterns, word lists that reflect variations in number of syllables (two and three) and frequency (high and low) were constructed. Because an empirically-based word frequency measure does not to our knowledge exist for Spanish, word frequency was established on the basis of ratings by native speakers (Carroll, 1971). A total of 166 words were evenly divided into two lists, one of which was used as a pretest and the other as a posttest.

3.1.5 Schedule of Training

Subjects were first administered Spanish and English versions of the unit detection task. They then received training using the SPEED game. Training continued until 20 units had been mastered or until a criterion of 150 words per minute had been reached on the initial runs of five consecutively introduced units. They were then administered alternate forms of the Unit Detection Task in Spanish and English, followed by the Spanish and English Word Pronunciation Tasks. Subjects next were given RACER training, which continued until the 27 levels of difficulty had been covered. Finally, they were administered alternate forms of the Spanish and English word pronunciation tasks. The design thus provided pretests and posttests of the skills targeted in the SPEED and RACER games. For the SPEED game, the unit detection task served as the criterion measure. For the RACER game, the word pronunciation task was the criterion. The evaluation of training was based upon analyses of performance on the criterion tasks.

3.2 Results

3.2.1 SPEED Evaluation

To evaluate the effects of SPEED training on the development of the perceptual and attentional skills it addressed, a series of repeated measures analyses of variance were carried out of performance on the criterion task, the unit detection task. In the first set of analyses, we compared the effects of SPEED training on subjects' overall performance in detecting target units. We compared performance when the target units are present in the stimulus word (which are called targets in the analysis) with cases where the stimulus word contains instead an orthographically similar unit (called a similar foil) or contains only orthographically dissimilar units (called a dissimilar foil). The analysis thus included as factors Training (pretest vs. posttest) and Item Type (targets, similar foils, and dissimilar foils). The analysis was carried out for two dependent variables: unit detection latency and percent of correct detections. There were separate analyses for the English and Spanish versions of the unit detection task. The goal was to ascertain the effects of training on perceptual skill components of reading for subjects whose primary language was not the language of training. A second goal was to determine the dependency of skills acquired on the language of training. If the effect of training is the development of general perceptual and attentional skills as we have argued elsewhere (Frederiksen, Warren, & Rosebery, 1985a), then the effects of training should be independent of the language in which the skill is evaluated.

Results for the analyses of targets, similar, and dissimilar foils are shown in Figures 3-3 through 3-6, for the English and Spanish versions of the unit detection task. In the analysis of mean unit detection latencies (Figures 3-3 and 3-4), there were significant effects of training for both the English ($F_{1,10}$=25.66, p=.0005) and Spanish ($F_{1,10}$=9.78, p=.011) versions of the task. In each case, there was a sizeable reduction in mean detection latencies with training, averaging 298 msec for the English task and 366 msec for the Spanish task. Moreover, following training the mean latencies were similar for both English and Spanish units (they averaged 740 msec for English units and 794 msec for Spanish units). In contrast, before training latencies for detecting English units (1037 msec) were smaller than those for detecting Spanish

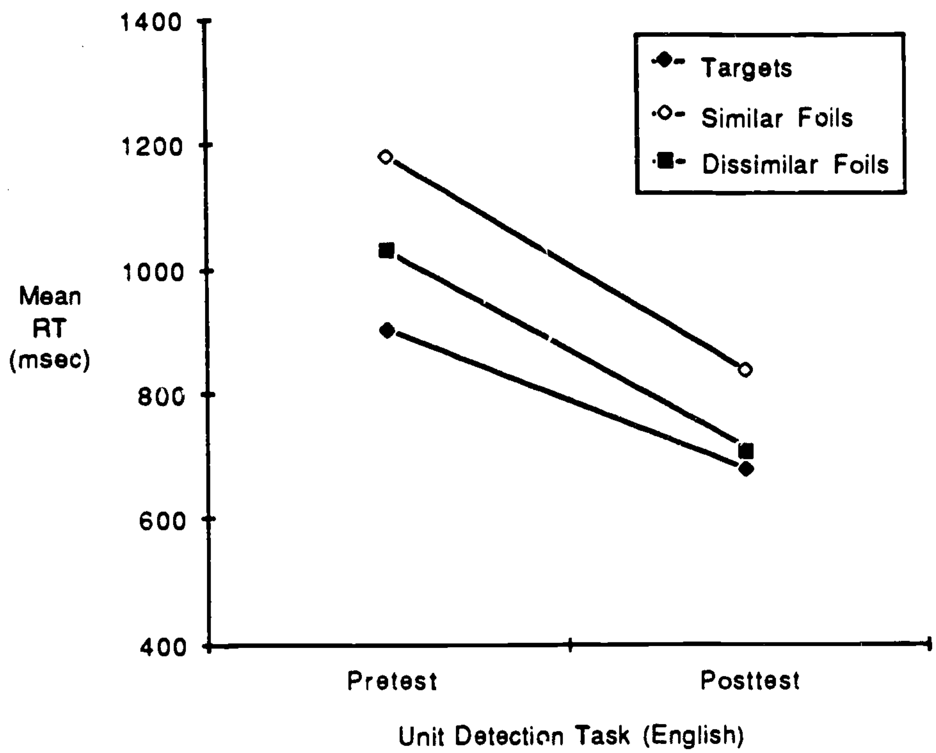

Figure 3-3: Mean latencies obtained in the English version of the unit detection task before and after SPEED training.

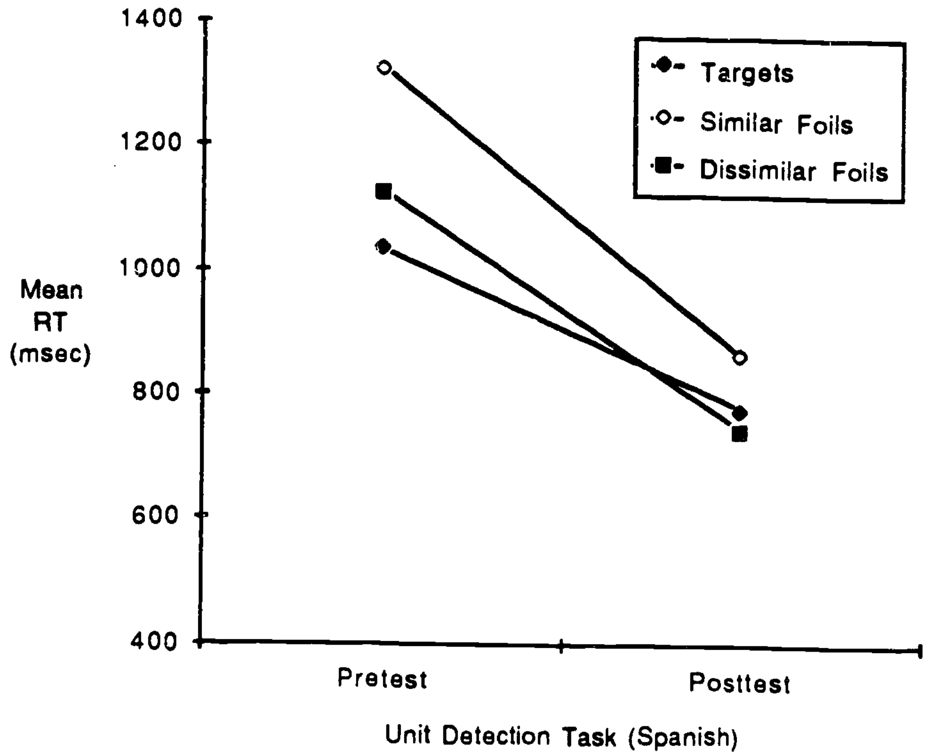

Figure 3-4: Mean latencies obtained in the Spanish version of the unit detection task before and after SPEED training.

units (1160 msec). Furthermore, in each case there were significant differences in the effects of training for targets, similar, and dissimilar foils. The Training by Item Type interaction was significant for both the English ($F_{2,20}=3.55$, $p=.05$) and Spanish ($F_{2,20}=5.98$, $p=.009$) versions of the task. In each case, the effect of training depended upon the overall item difficulties: they were greatest for similar foils, next greatest for dissimilar foils, and least for targets. The main effect of Item Type was also significant for both versions of the task ($F_{2,20}=18.53$, $p<.0001$ for the English version, and $F_{2,20}=12.63$, $p<.0001$ for the Spanish version).

In the analyses of the percentage of correct unit detections (Figures 3-5 and 3-6), there were significant increases in accuracy as a result of training for both the English and Spanish versions of the task. The main effect of Training yielded $F_{1,10}=9.78$, $p=.01$ for the English task and $F_{1,10}=4.60$, $p=.06$ for the Spanish task. For the English task, the mean accuracy increased from 85.7% in the pretest to 90.8% in the posttest, and for the Spanish task the accuracy increased from 87.2% to 89.7%. There were also significant differences due to Item Type ($F_{2,20}=54.95$, $p<.0001$ for the English task, and $F_{2,20}=85.19$, $p<.0001$ for the Spanish task). In each case, the majority of errors were false positives made to similar foils. In neither case, however, was there a significant interaction between Training and Item Type.

A second set of analyses of the unit detection data were carried out to study the effects of training on the detection of target units of varying length (two or three letters) and position (initial or medial position within the stimulus word). In addition, half of the units were actually included in training and half were not. The goal of these analyses was to test for changes in the effects of unit length and unit position that occurred as a result of training. Reductions in unit length and position effects are suggestive of a shift in the subjects' mode of processing from a strategy of serially scanning each stimulus word and testing letter sets against a memory of the target unit to a parallel attentional strategy in which evidence for the target unit's presence emerges from a parallel encoding of information within the stimulus item. It is such a parallel encoding of orthographic information that appears to be characteristic of more able readers (Frederiksen, 1977, 1981a). Finally, if such a general perceptual/attentional skill is developed in SPEED training, the effects of training should be equivalent regardless of whether or not the units tested were actually trained.

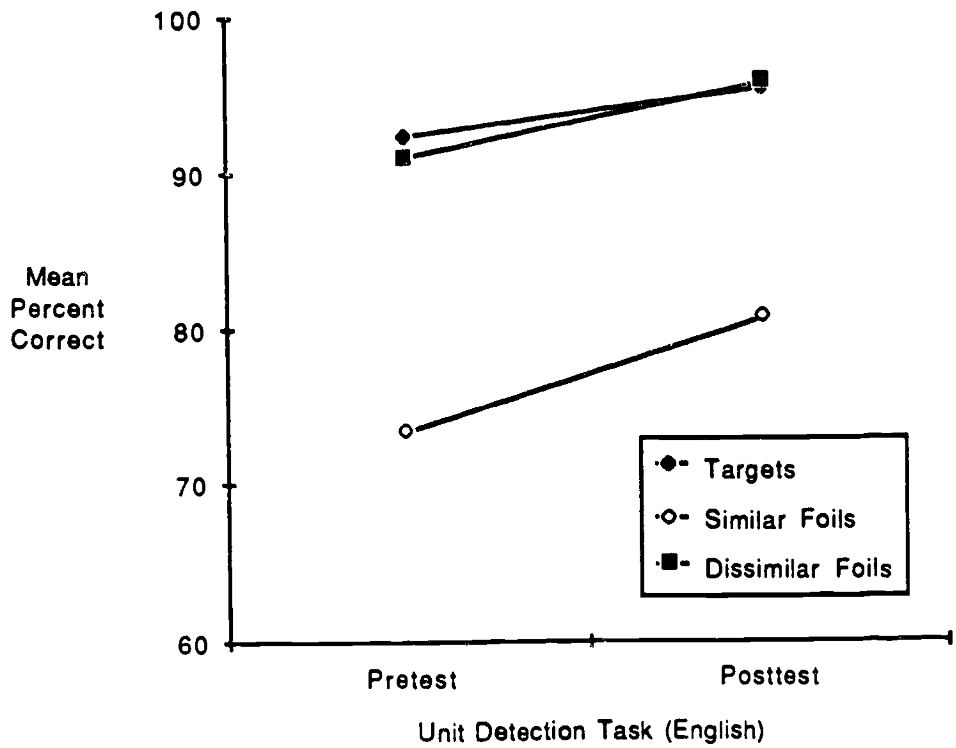

Figure 3-5: Mean percentages of correct responses in the English version of the unit detection task before and after SPEED training.

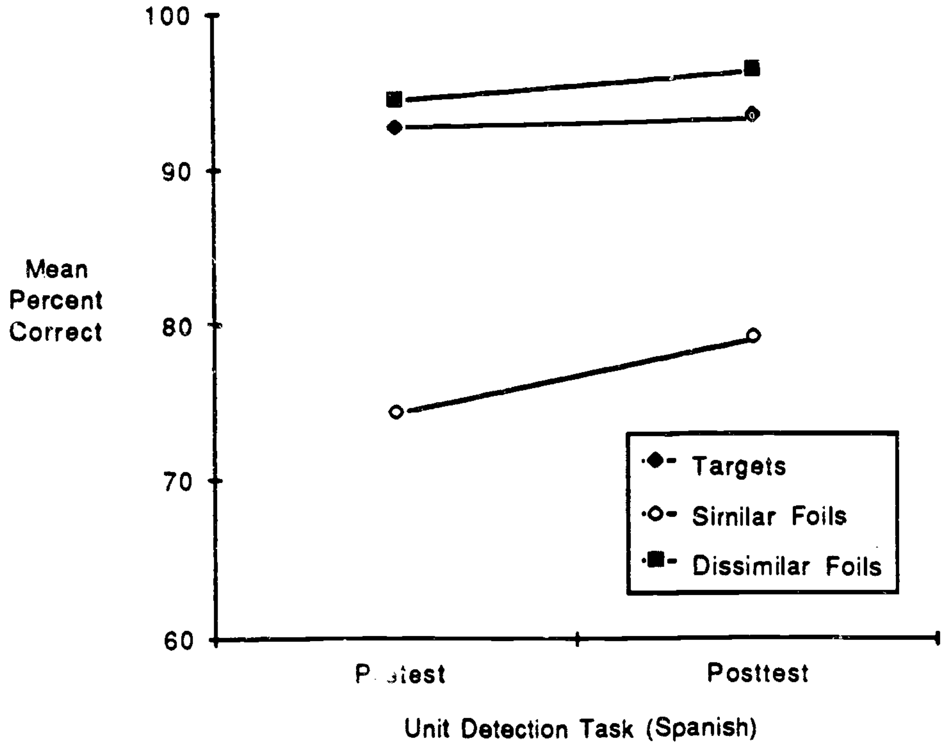

Figure 3-6: Mean percentages of correct responses in the Spanish version of the unit detection task before and after SPEED training.

These analyses were first carried out using subjects' mean unit detection latencies as the dependent variable. These results are shown in Figures 3-7 and 3-8 for the two versions of the unit detection task. For both the English and Spanish versions, there were considerable reductions in detection latencies as a result of training. Furthermore, the effects of target length and position on latency decreased significantly as a result of training. The main effects of training in the English ($F_{1,10}=14.65$, $p=.003$) and Spanish ($F_{1,10}=12.21$, $p=.006$) analyses were highly significant. In the English task, medial units took longer to detect than did initial units, and units of 3 letters took longer to detect than those of two letters, particularly when they occurred in medial positions (the main effects of unit length and position were each significant, with $F_{1,10}=9.38$, $p=.01$ and $F_{1,10}=66.11$, $p<.0001$, respectively, as was the length by position interaction with $F_{1,10}=31.13$, $p=.0002$). These effects of unit position and length were greatly diminished following training. The interactions of Training with Position and Length, and the triple interaction of these three factors were both significant (they were, respectively, $F_{1,10}=11.57$, $p=.007$, $F_{1,10}=5.96$, $p=.03$, and $F_{1,10}=11.37$, $p=.007$). Finally, there was no effect of whether or not the target units were actually present during SPEED training. The improvement in performance was 225 msec for trained units and 209 msec for the untrained units. In the Spanish version of the task, medially presented three-letter units again took longer to detect than two-letter units in that position or than units of any length appearing in the initial position within the stimulus word. While the main effects of unit length and position were not significant, the length by position interaction was significant, with $F_{1,10}=6.56$, $p=.03$. Again, as was the case with the English task, these item differences were reduced as a result of training, although in this case, not significantly so. Finally, there was again no effect of whether or not the target units were similar to English multiletter units that had actually received training. The improvement in performance was 276 msec for trained units and 229 msec for the untrained units.

In addition to the analyses of response latencies, we carried out a similar set of analysis of subjects' target detection accuracies. Results of these analyses are shown in Figures 3-9 and 3-10. For the English version of the task, subjects' accuracies were above 90% in the pretest for all conditions except for medial, three-letter units, where their accuracy was 86%. The main effect of unit length ($F_{1,10}=4.80$, $p=.05$) and position ($F_{1,10}=87.67$, $p<.0001$), and the Length by Position interaction ($F_{1,10}=26.24$, These analyses were first carried out using subjects' mean unit detection latencies as the dependent variable. These results are shown in Figures 3-7 and 3-8 for the two versions of the unit detection task. For both the English and Spanish versions, there were considerable reductions in detection latencies as a result of training. Furthermore, the effects of target length and position on latency decreased significantly as a result of training. The main effects of training in the English ($F_{1,10}$=14.65, p=.003) and Spanish ($F_{1,10}$=12.21, p=.006) analyses were highly significant. In the English task, medial units took longer to detect than did initial units, and units of 3 letters took longer to detect than those of two letters, particularly when they occurred in medial positions (the main effects of unit length and position were each significant, with $F_{1,10}$=9.38, p=.01 and $F_{1,10}$=66.11, p<.0001, respectively, as was the length by position interaction with $F_{1,10}$=31.13, p=.0002). These effects of unit position and length were greatly diminished following training. The interactions of Training with Position and Length, and the triple interaction of these three factors were both significant (they were, respectively, $F_{1,10}$=11.57, p=.007, $F_{1,10}$=5.96, p=.03, and $F_{1,10}$=11.37, p=.007). Finally, there was no effect of whether or not the target units were actually present during SPEED training. The improvement in performance was 225 msec for trained units and 209 msec for the untrained units. In the Spanish version of the task, medially presented three-letter units again took longer to detect than two-letter units in that position or than units of any length appearing in the initial position within the stimulus word. While the main effects of unit length and position were not significant, the length by position interaction was significant, with $F_{1,10}$=6.56, p=.03. Again, as was the case with the English task, these item differences were reduced as a result of training, although in this case, not significantly so. Finally, there was again no effect of whether or not the target units were similar to English multiletter units that had actually received training. The improvement in performance was 276 msec for trained units and 229 msec for the untrained units.

In addition to the analyses of response latencies, we carried out a similar set of analysis of subjects' target detection accuracies. Results of these analyses are shown in Figures 3-9 and 3-10. For the English version of the task, subjects' accuracies were above 90% in the pretest for all conditions except for medial, three-letter units, where their accuracy was 86%. The main effect of unit length ($F_{1,10}$=4.80, p=.05) and position ($F_{1,10}$=87.67, p<.0001), and the Length by Position interaction ($F_{1,10}$=26.24, These analyses were first carried out u latencies as the dependent variable. These resul for the two versions of the unit detection task versions, there were considerable reductions ir training. Furthermore, the effects of target leng significantly as a result of training. The mai ($F_{1,10}$=14.65, p=.003) and Spanish ($F_{1,10}$=12. significant. In the English task, medial units t units, and units of 3 letters took longer to particularly when they occurred in medial posit and position were each significant, with $F_{1,10}$=9 respectively, as was the length by position int These effects of unit position and length were The interactions of Training with Position and these three factors were both significant (they w $F_{1,10}$=5.96, p=.03, and $F_{1,10}$=11.37, p=.007). Fi or not the target units were actually present du in performance was 225 msec for trained units a In the Spanish version of the task, medially pre

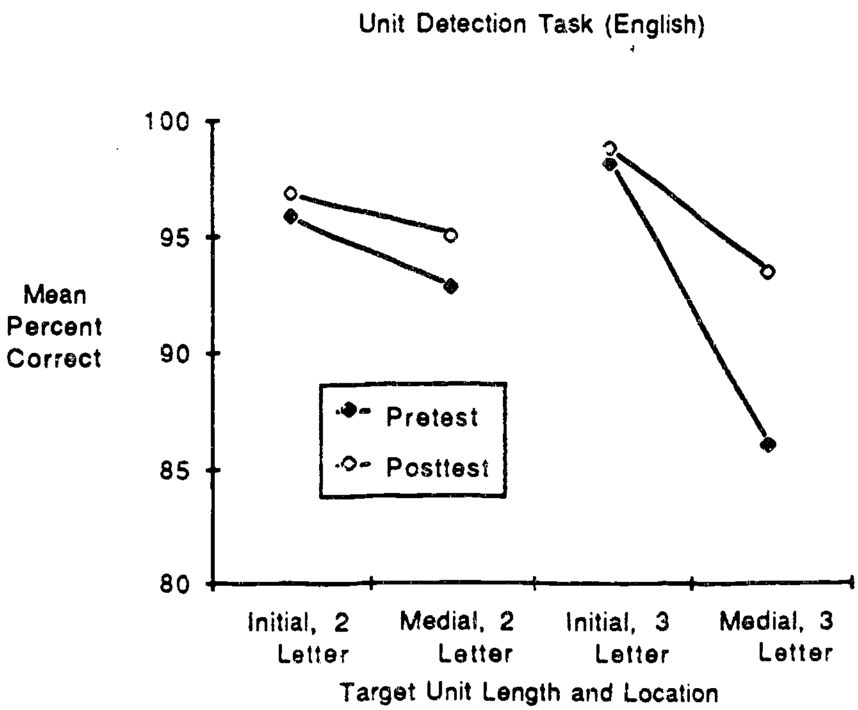

Figure 3-9: Mean percentage of correct responses for English units varying in length and position, before and after SPEED training.

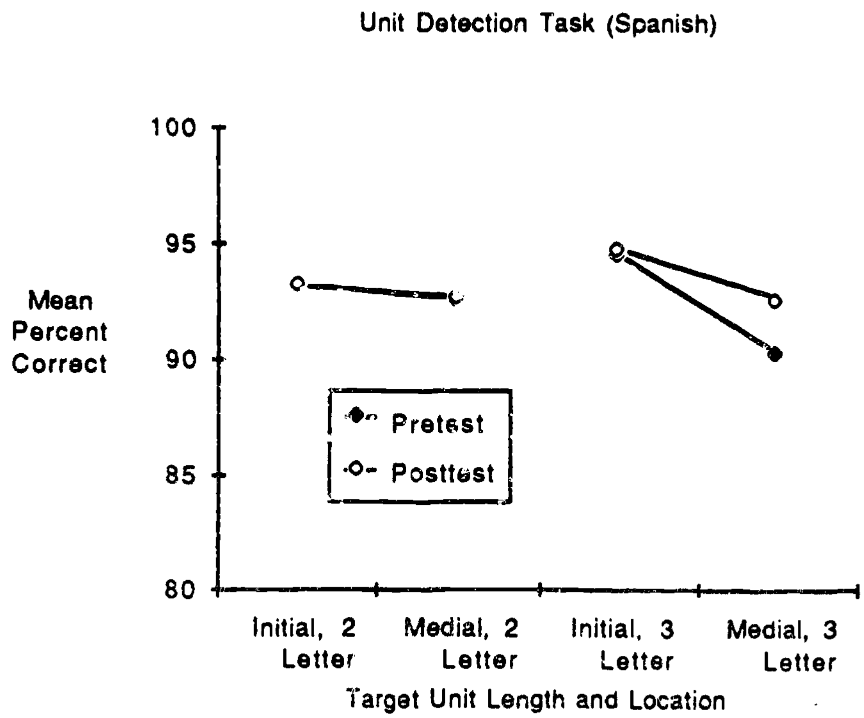

Figure 3-10: Mean percentage of correct responses for Spanish units varying in length and position, before and after SPEED training.

p=.0004) were all significant. As a result of training, the subjects' accuracy increased, particularly for the medial, three-letter unit condition (to 93%). The main effect of training for the English version of the task was significant ($F_{1,10}$=27.131, p=.0004) as were the interactions of Training by Length ($F_{1,10}$=6.43, p=.03) and Training by Position ($F_{1,10}$=31.60, p=.0002). Finally, the effects of training on accuracy were the same for units that were included in SPEED training (3.0%) and those that were not (2.7%).

For the Spanish version of the unit detection task, subjects' accuracy in the pretest was again poorest for the case where 3-letter units were embedded within the stimulus word (90%). For the other conditions, accuracy averaged 93% or greater. The main effect of unit position ($F_{1,10}$=18.96, p=.001) and the interaction of unit length and position ($F_{1,10}$=7.18, p=.02) were both significant. While we again found the effect of training was to reduce error rates, primarily for medially presented 3-letter units, none of the effects of training were significant in this analysis of performance for targets. We should bear in mind that for both the English and Spanish versions of the unit detection task, accuracies in detecting units present within a stimulus word were high both before and after training, and the the main source of errors was the false acceptance of similar foils. Thus, there was little room for improvement in accuracy in detecting target units when they were present. Finally, we should note that again the effects of training were similar for units that were included in SPEED training (.8%) and those that were not (.7%).

Our conclusion is that Hispanic bilingual subjects can profit markedly from training using the SPEED game, even when training is presented in English. We also can conclude that the performance improvements resulting from training represent changes in general perceptual and attentional processes, processes that are not dependent upon particular linguistic systems for their operation. Similar decreases in latency were found for both trained and untrained units and for both English and Spanish versions of the unit detection task, and these were accompanied by increases in accuracy of performance (with reductions in the frequency of false positive responses made to similar foils). These results imply that a general improvement has occurred in the ability of subjects to discriminate and encode multiletter units. Moreover, the reduction in effects of unit length and position within the target word strongly suggest a shift to an attentional strategy in which information from the entire word is processed in parallel. This attentional strategy is independent of whether or not the units and words are the ones actually used in training, and whether or not they are materials from the same language as the language of training. These results confirm the interpretation we have given of skill components developed in SPEED training in our earlier study (Frederiksen, Warren, & Rosebery, 1985a).

3.2.2 RACER Evaluation

The criterion task used in the RACER evaluation was a word pronunciation task, administered in both Spanish and English. The purpose of this task was to evaluate the effects of RACER training, delivered in English, on subjects' word decoding performance in both English and Spanish. Our hypothesis was that word decoding skills addressed in the RACER game involve language-specific rules, and therefore that they should not transfer to the Spanish version of the criterion task. In the pronunciation task, the words varied in frequency and in length (in syllables). High frequency words are more apt to belong to a subjects' sight vocabulary and thus are less likely to require the application of decoding rules. Longer words of a given frequency class are more difficult to decode than shorter words. Therefore, evidence for efficient decoding will consist in a reduction in the performance differences for high and low frequency words and for words that differ in their length. For the English task, word frequency assignments were based upon a standard frequency count (Carroll, Davies, & Richman, 1971), while for the Spanish task they were based upon ratings of frequency. In the Spanish task, words were either 2 or 3 syllables in length, while in the English task, they were either 1 or 2 syllables in length. Two dependent variables were employed: subjects latencies to onset of their pronunciation, and their accuracy of pronunciation as judged by a native speaker of English or Spanish.

Results of the analyses of subjects' vocalization onset latencies are shown in Figures 3-11 and 3-12 for the English and Spanish pronunciation tasks. For the English task, there was a significant reduction in pronunciation latencies as a result of training ($F_{1,8}$=6.53, p=.03). The mean onset latency was 1,286 msec in the pretest and 830 msec in the posttest. The pretest latencies are larger than those we commonly find for monolingual English subjects (which are typically around 700 msec).

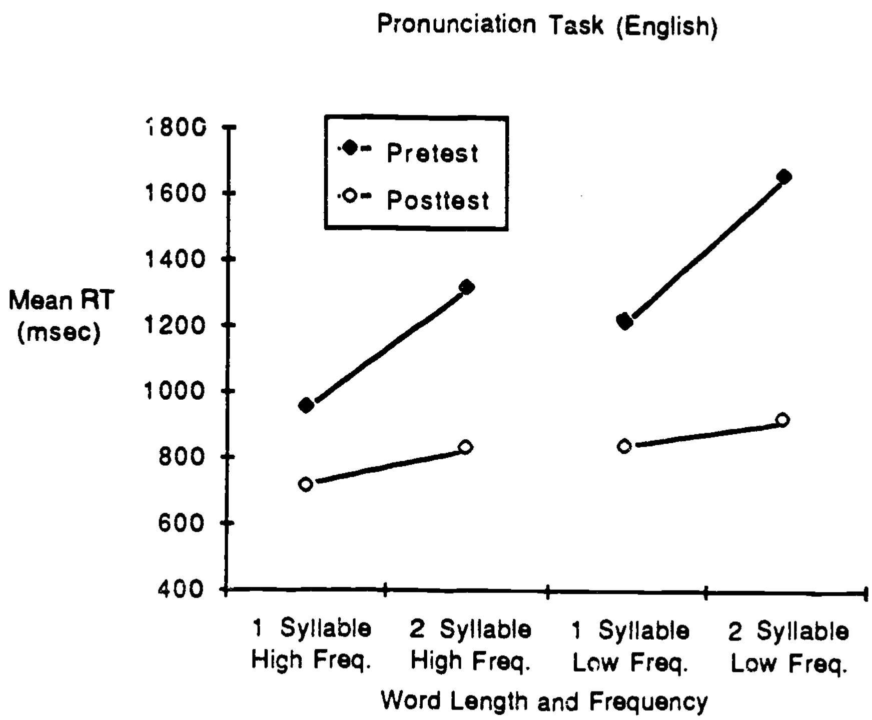

Figure 3-11: Mean vocalization onset latencies for the English pronunciatio task, before and after RACER training.

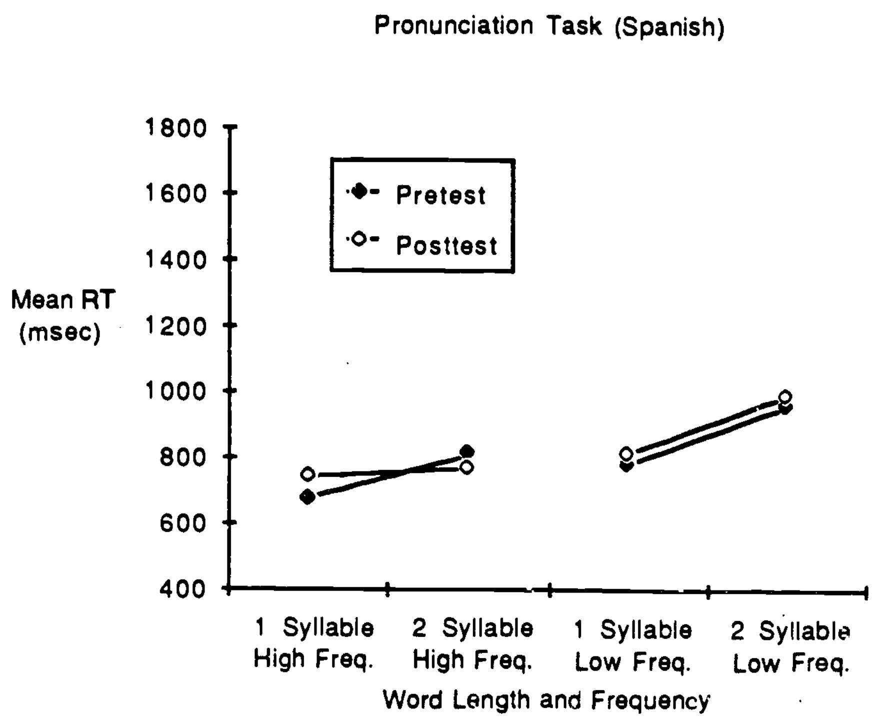

Figure 3–12: Mean vocalization onset latencies for the Spanish pronunciation task, before and after RACER training.

However, the posttest latencies are closer to those of English subjects who have such training (500–600 msec). In addition to this main effect of training, there w significant effect of syllabic length for the pretest (t_8=1.84, p=.05) but not for posttest (t_8=.42). The difference in latencies for 2– and 1–syllable words was msec in the pretest and 93 msec in the posttest. There was also a significant ef of word frequency in the pretest (t_8=2.12, p=.03) but, again, not in the post (t_8=.71). The differences in latencies for low and high frequency words was 305 n in the pretest and 102 msec in the posttest. For the Spanish task, there were demonstrable effects of RACER training on subjects' onset latencies. There however significant syllable ($F_{1,8}$=7.77, p=.02) and frequency effects ($F_{1,8}$=11 p=.01), both before and after training.

It is apparent that bilingual subjects, whose first language is not English, profit from training in decoding using the RACER environment. These gains performance were greater than those shown by English speaking subjects who trained using an earlier version of the RACER game (Frederiksen, Warren, & Roseb 1985b). Following training, subjects depended less on the frequency of the word showed similar onset latencies for 1– and 2–syllable words. These changes indi that the subjects have developed more efficient word decoding skills as a resul training. None of these performance gains extended, however, to Spanish words. range of mean vocalization onset latencies for Spanish words ranged from 683 to msec. Since these values are relatively long compared with those we have found high ability readers of English (which are 600–625 msec; Frederiksen, 1977), it is likely that they represent asymptotic performance for the Spanish task. We there conclude that decoding skill developed using English materials did not transfer to Spanish materials, and that decoding skill is language specific.

In the analyses of subjects' percentages of correct pronunciations (Figures 3 and 3–14), there were no significant effects of training in either the Englisl Spanish versions of the task. In both cases, subjects were less accurate in deco low frequency that high frequency words, and made more errors in decoding the lo than the shorter items. For the English pronunciation task, $F_{1,8}$=4.37, p=.07 for syllable effect and $F_{1,8}$=28.20, p=.0007 for the frequency effect. For the Spa task, $F_{1,8}$=18.24, p=.003 for the syllable effect and $F_{1,8}$=18.27, p=.003 for frequency effect.

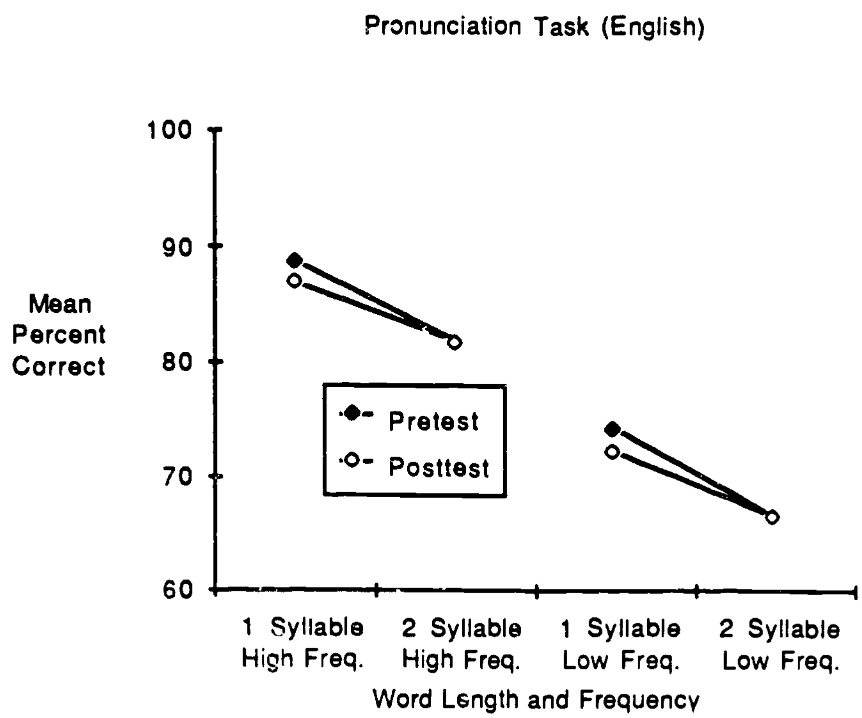

Figure 3—13: Mean percent of words correct for the English pronunciation task, before and after RACER training.

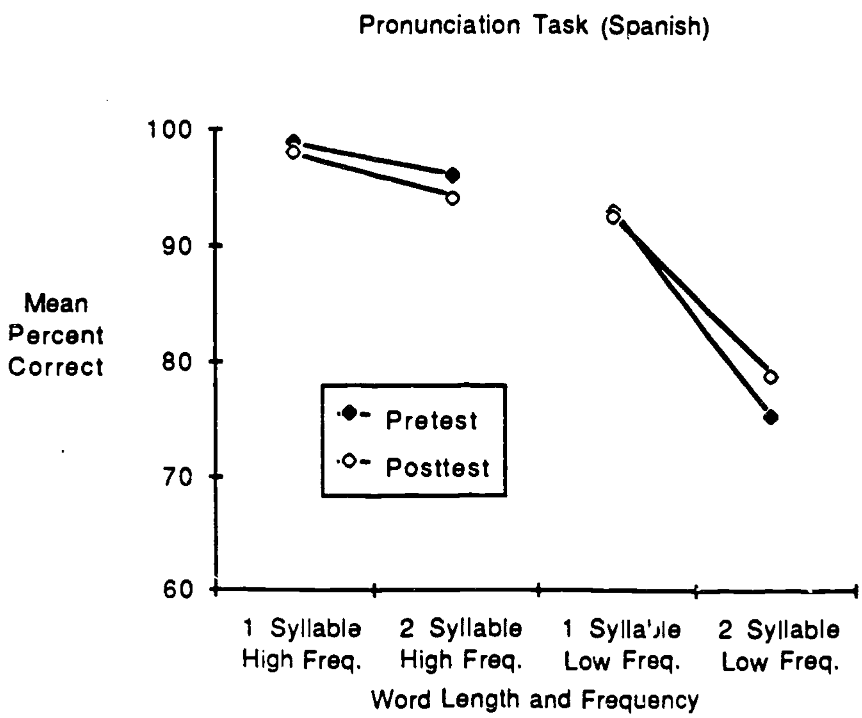

Figure 3-14: Mean percent of words correct for the Spanish pronunciation task, before and after RACER training.

3.3 Discussion

Our conclusion is that Hispanic trainees can benefit from computer-based training focusing on the development of automatic skills for encoding orthographic information and on its decoding. SPEED training succeeded in improving perceptual encoding skills to a level that is comparable to that of English-language subjects given similar training, and the skills developed were generalizable to orthographic units that had not received training and to units in another language (Spanish). RACER training succeeded in developing in Hispanic subjects a degree of automaticity in decoding within English. Indeed, their improvements in onset latencies were comparable to those of monolingual English trainees. However, this improvement in decoding automaticity was not accompanied by a similar improvement in accuracy. While their accuracies for high frequency words were fairly high, they showed particular difficulty with low frequency words. Hispanic bilingual subjects thus appear to need additional instruction to develop a larger English vocabulary beyond that offered implicitly in the sequence of materials used in RACER training. Finally, we conclude that, in contrast to SPEED training, the skills developed in RACER training are language specific, and do not show transfer across languages.

The bilingual training study has addressed two components of reading in which students who are not proficient in English show low levels of automaticity: encoding of orthographic units and phonological decoding. Favreau and Segalowitz (1983) and Duran (1985) have provided evidence that such subjects also show deficiencies in their use of semantic information derived from context (in these studies, the contexts were semantically related words). In particular, in Favreau and Segalowitz's study, bilingual subjects whose reading speed was as great in English as in their first language (French) showed a facilitation in their lexical decision latencies when the target word was preceded by a semantically related prime, and no inhibition when the target word was unrelated to the prime. In contrast, subjects whose reading speed in English was not as great as that in French showed only a small such semantic facilitation. This pattern of results, they point out, is strongly suggestive of the operation in the former subject group of an automatic process for using context to gain access to word meanings. Such a process is, of course, the one addressed in the Defender training environment used in our earlier experiments. We feel, therefore, that training of bilingual subjects using the Defender game could address this important skill deficiency and, as we have argued, that such training could be critical in enabling other comprehension skills such as reference assignment.

In proposing such a training study, the question again arises as to the linguistic specificity of the skill addressed. Reviews of research by Duran (1985), McCormack (1977), Dornic (1977), and Lopez (1977) support the hypothesis of a single semantic memory system for words from two languages. Given our theory of the skill developed in the Defender game, this raises the interesting possibility that (1) Defender training delivered in one langage will develop skills which are generalizable to the other language, and (2) such transfer should occur from either the primary to the secondary or the secondary to the primary language. Thus we may speculate that Defender training could initially be delivered in the primary, non-English language and then followed by practice using English materials. In this way, the common underlying skill could be addrssed initially without the additional processing load of word decoding and syntactic parsing in the less familiar language. This is an example of the type of training called for by Duran:

> In the case where bilinguals are not skilled readers in their more familiar language, training of reading skills in the more familiar language may be used as a procedure to prepare for training of reading in the less familiar language (1985, p. 31).

Clearly for such training to be developed, a theory of the cognitive locus of training effects is necessary, as well as an understanding of the similarities in language structures of the two language systems. The present proposal for training in the use of contextual activation within semantic memory has such a conceptual basis and, we feel, merits consideration.

4. ACKNOWLEDGEMENTS

This research was supported primarily by the Office of Naval Research, Personnel and Training Research Program and the Naval Personnel Research and Development Center, under ONR contract N00014-83-C-0639, and also by the National Institute of Education, Contract No. 400-81-0030. In addition to the author, many people contributed to the research. B. Warren and A. Rosebery carried out the initial two studies of skill deficiencies in comprehension, and were instrumental in the development of training and testing materials for the bilingual training study. B. Freeman developed programs for the revised training systems, as well as for the other computer-administered tasks. J. Whinfield was responsible for data collection in the bilingual study. J. Imai helped in the preparation of materials and was responsible for testing of subjects in the study of reference mapping and the Defender evaluation study. Finally, C. Martinez developed the Spanish materials used in the bilingual study and carried out subject testing in Spanish. The work of this project could not have been carried out without the contributions of each of these people.

5. REFERENCES

Bruce, B. (1987). Deep case systems for language understanding. In S. Shapiro (Ed.), Encyclopedia of artificial intelligence. New York: John Wiley &n Sons.

Carroll, J. B., Davies, P., & Richman, B. (1971). Word frequency book. New York: American Heritage.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. Psychological Review, 82, 407–428.

Dornic, S. (1977). Language dominance, spare capacity, and perceived effect in bilinguals. Ergonomics, 1980, 23, 369–377.

Duran, R. (1985). Influences of language skills on bilinguals' problem solving. In S. Chipman, J. Sigel, & R. Glaser (Eds), Thinking and learning skills; Volume 2: Current research and open questions. Hillsdale, N. J.: Lawrence Erlbaum Associates.

Favre M., Komoda, M., & Segalowitz, N. (1980). Second language reading: Implications of the word superiority effect in skilled bilinguals. Canadian Journal of Psychology, 4, 370–381.

Favreau, M., & Segalowitz, N. (1984). Automatic and controlled processes in the first– and second–language reading of fluent bilinguals. Memory & Cognition, 1983, 11(6), 565–574.

Fillmore, C. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart, & Winston.

Frederiksen, J. R. (1977). Assessment of perceptual, decoding, and lexical skills and their relation to reading proficiency. In A. M. Lesgold, J. W. Pellegrino, S. Fokkema, & R. Glaser (Eds.), Cognitive psychology and instruction. New York: Plenum.

Frederiksen, J. R. (1981a). Sources of process interactions in reading. In A. M. Lesgold & C. A. Perfetti (Eds.), Interactive processes in reading. Hillsdale, N. J.: Lawrence Erlbaum Associates.

Frederiksen, J. R. (1981b). Understanding anaphora: Rules used by readers in assigning pronominal referents. Discourse Processes, 4, 323-347.

Frederiksen, J. R., & Warren, B. M. (1986). A cognitive framework for developing expertise in reading. In R. Glaser (Ed.), Advances in instructional psychology. Hillsdale, N. J.: Lawrence Erlbaum Associates.

Frederiksen, J. R., Warren, B. M., & Rosebery, A. S. (1985a). A componential approach to training reading skills: Part 1. Perceptual Units Training. Cognition and Instruction, 2(2), 91-130.

Frederiksen, J. R., Warren, B. M., & Rosebery, A. S. (1985b). A componential approach to training reading skills: Part 2. Decoding and use of context. Cognition and Instruction, 2(3 & 4), 271-338.

Frederiksen, J. R., & White, B. Y. (in press). An approach to training based upon principled decomposition. Acta Psychologica.

Givon, T., Kellogg, W., Posner, M., & Yee, P. (1985). The tracking . referents in discourse: Automated vs. attended processes. Technical Report, University of Oregon, Eugene, Oregon.

Grimes, J. (1975). The Thread of Discourse. The Hague: Molton.

Gruber, E. H.,, Beardsley, W., & Caramazza, A. (1978). Parallel function strategy in pronoun assignment. Cognition, 6, 117-133.

Guidance Testing Associates. (1962). Prueba de Lectura, Form A. St. Mary's University, San Antonio, Texas.

Halliday, M. A. K., & Hasan, R. (1976). Cohesion in English. London: Longmen.

Kieras, D. (1981). The role of major referents and sentence topics in the construction of passage macrostructure. Discourse Processes, 4, 1-15.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Lesgold, A., Roth, S. F., & Curtis, M. E. (1979). Foregrounding effects in discourse comprehension. Journal of Verbal Learning and Verbal Behavior, 18, 291-308.

Lopez, M. (1977). Bilingual memory research: Implications for bilingual education. In J. L. Martinez, Jr. (Ed.), Chicano Psychology. New York: Academic Press.

McCormack, P. D. (1977). Bilingual linguistic memory: The independence-interdependence issue revisited. In P. A. Hornby (Ed.), Bilingualism: Psychological, social, and educational implications. New York: Academic Press.

Oller, J. W., Jr., & Tullius, J. (1973). Reading skills of non-native speakers of English. International Review of Applied Linguistics, 11, 69-80.

Rosebery, A. S. (1986). Semantic and syntactic bases of text comprehension. Technical Report No. 6010, BBN Laboratories, Cambridge, Massachusetts.

Schneider, W., & Fisk, A. D. (1982). Automatic category search and its transfer: Automatic process semantic filtering. Technical report HARL-ONR-8202, University of Illinois, Champaign, Illinois.

Schwartz, S. (1984). Measuring reading competence. New York: Plenum.

Schwartz, S. (1986). A comparison of componential and traditional approaches to training reading skills. Unpublished manuscript, University of Queensland, St. Lucia, Queensland, Australia.

Science Research Associates (1983). DISTAR Reading Mastery, Level 2. Chicago: Author.

Vipond, D. (1980). Micro- and macroprocesses in text comprehension. Journal of Verbal Learning and Verbal Behavior, 19, 276-296.

Warren, B. M. (1986). Aspects of skill in understanding high-order semantic relations. Technical Report No. 6009. BBN Laboratories, Cambridge, Massachusetts.

6. APPENDIX A: SAMPLE SENTENCE SETS USED IN THE PRONOUN REFERENCE EXPERIMENT

Sentence Set 60

1. The human body develops from a single-celled origin.
2. The human body develops from a single, fertilized cell.
3. A single, fertilized cell develops into the human body.
4. The human body develops from the union of male and female reproductive cells.
5. It undergoes progressive changes until the age of 25.
6. The human body undergoes progressive changes until the age of 25.
7. It is a miracle of chemical biological complexity.
8. Man's anatomy is a miracle of chemical and biological complexity.
9. The human body is a miracle of chemical and biological complexity.
10. Until the age of 25 progressive changes alter it.
11. The fertilized cell is produced by the union of a male and female reproductive cells.
12. It follows the basic vertebrate and mammalian pattern of development until adulthood.

Sentence Set 61

1. The nucleus of Halley's comet is chemically intriguing and complex, scientists have recently discovered.
2. The nucleus of Halley's comet is rich with complex and intriguing chemistry, scientists have recently discovered.
3. An intriguing and complex chemistry makes up the nucleus of Halley's comet, scientists have recently discovered.
4. The nucleus of Halley's comet is rich with intriguing and complex chemical structures, scientists have recently discovered.
5. It had never been seen before this March when its picture was taken by five probes.

6. The nucleus had never been seen before this March when its picture was taken by five space probes.

7. It reveals much about the nature of matter in the depths of space.

8. The core reveals much about the nature of matter in the depths of space.

9. The nucleus reveals much about nature of matter in the depths of space.

10. Before this March when five space probes took pictures, no one had ever seen it.

11. The chemistry is of particular interest to scientists who believe that it will provide a detailed account of the molecular composition of the earth in the distant past.

12. It is composed of complex, carbon-based molecules that could account for why it is one of the darkest objects ever seen in the solar systems.

7. APPENDIX B: ESSAY FORMS CREATED FROM THE SAMPLE SENTENCE SETS

Sentence Set 60

1. The human body develops from a single, fertilized cell. It undergoes progressive changes until the age of 25.

2. A single, fertilized cell develops into the human body. The human body undergoes progressive changes until the age of 25. It follows the basic vertebrate and mammalian pattern of development until adulthood.

3. A single, fertilized cell develops into the human body. It undergoes progressive changes until the age of 25.

4. The human body develops from a single-celled origin. It undergoes progressive changes until the age of 25.

5. The human body develops from a single, fertilized cell. It is a miracle of chemical and biological complexity.

6. A single, fertilized cell develops into the human body. Until the age of 25 progressive changes alter it.

7. The human body develops from a single, fertilized cell. The fertilized cell is produced by the union of a male and female reproductive cells. It follows the basic vertebrate and mammalian pattern of development until adulthood.

8. The human body develops from the union of male and female reproductive cells. It undergoes progressive changes until the age of 25.

9. A single, fertilized cell develops into the human body. The human body is a miracle of chemical and biological complexity. It follows the basic vertebrate and mammalian pattern of development until adulthood.

10. A single, fertilized cell develops into the human body. Man's anatomy is a miracle of chemical and biological complexity. It follows the basic vertebrate and mammalian pattern of development until adulthood.

11. The human body develops from a single, fertilized cell. It undergoes progressive changes until the age of 25. It follows the basic vertebrate and mammalian pattern of development until adulthood.

12. A single, fertilized cell develops into the human body. It is a miracle of chemical and biological complexity. It follows the basic vertebrate and mammalian pattern of development until adulthood.

13. The human body develops from a single, fertilized cell. Until the age of 25 progressive changes alter it. It follows the basic vertebrate and mammalian pattern of development until adulthood.

Sentence Set 61

1. The nucleus of Halley's comet is rich with complex and intriguing chemistry, scientists have recently discovered. It had never been seen before this March when its picture was taken by five probes.

2. An intriguing and complex chemistry makes up the nucleus of Halley's comet, scientists have recently discovered. The nucleus had never been seen before this March when its picture was taken by five space probes. It is composed of complex, carbon-based molecules that could count for why it is one of the darkest objects ever seen in the solar system.

3. An intriguing and complex chemistry makes up the nucleus of Halley's comet, scientists have recently discovered. It had never been seen before this March when its picture was taken by five probes.

4. The nucleus of Halley's comet is chemically intriguing and complex, scientists have recently discovered. It had never been seen before this March when its picture was taken by five probes.

5. The nucleus of Halley's comet is rich with complex and intriguing chemistry, scientists have recently discovered. It reveals much about the nature of matter in the depths of space.

6. An intriguing and complex chemistry makes up the nucleus of Halley's comet, scientists have recently discovered. Before this March when five space probes took pictures, no one had ever seen it.

7. The nucleus of Halley's comet is rich with complex and intriguing chemistry, scientists have recently discovered. The chemistry is of particular interest to scientists who believe that it will provide a detailed account of the molecular composition of the earth in the distant past. It is composed of complex, carbon-based molecules that could count for why it is one of the darkest objects ever seen in the solar system.

8. The nucleus of Halley's comet is rich with intriguing and complex chemical structures, scientists have recently discovered. It had never been seen before this March when its picture was taken by five probes.

9. An intriguing and complex chemistry makes up the nucleus of Halley's comet, scientists have recently discovered. The nucleus reveals much about nature of matter in the depths of space. It is composed of complex, carbon-based molecules that could count for why it is one of the darkest objects ever seen in the solar system.

10. An intriguing and complex chemistry makes up the nucleus of Halley's comet, scientists have recently discovered. The core reveals much about the nature of matter in the depths of space. It is composed of complex, carbon-based molecules that could count for why it is one of the darkest objects ever seen in the solar system.

11. The nucleus of Halley's comet is rich with complex and intriguing chemistry, scientists have recently discovered. It had never been seen before this March when its picture was taken by five probes. It is composed of complex, carbon–based molecules that could count for why it is one of the darkest objects ever seen in the solar system.

12. An intriguing and complex chemistry makes up the nucleus of Halley's comet, scientists have recently discovered. It reveals much about the nature of matter in the depths of space. It is composed of complex, carbon–based molecules that could count for why it is one of the darkest objects ever seen in the solar system.

13. The nucleus of Halley's comet is rich with complex and intriguing chemistry, scientists have recently discovered. Before this March when five space probes took pictures, no one had ever seen it. It is composed of complex, carbon–based molecules that could count for why it is one of the darkest objects ever seen in the solar system.

8. APPENDIX C: THE DEFENDER TRAINING TASK

The task, game format and materials in the SKI JUMP training task (Frederiksen et al., 1986) have been changed considerably. The result is a new training system which we have named DEFENDER. What follows is a description of this system.

8.1 Contextual Priming Training Task

We have designed and implemented a task which incorporates two performance measures: pixel density and response time. As in the SKI JUMP training task, the trainee reads a sentence context in which the final word has been omitted. The subject's task is to judge as quickly as possible whether or not a word "fits" that context.

a) Pixel density. The first measure, pixel density, capitalizes on the ability of the IBM PC to mask words dynamically. This allows us to increase or decrease the amount of visual information that is available about a word on a given exposure. Reducing the amount of visual information available about a word forces trainees to use contextual sources of information in combination with visual information to perform the task.

Using this masking technique, we were able to design a task in which a trainee is presented with a series of exposures of a given word, where each subsequent exposure contains an increased amount of visual information about the word. The effect of this display technique is that, with each successive exposure, a word becomes increasingly clear and complete. Thus, if a subject cannot make his judgement on the first exposure, the subject can wait for the second exposure and see a bit more information, or the third exposure and see a bit more, and so on until the word is presented in its entirely. The masking is accomplished by turning on X% of the bits (usually 2 bits of the 64 bits corresponding to the pixels in each character) in a bit map and logically AND-ing the stimulus word display to this. On each successive exposure, an additional 2 pixels on the bit map are turned on on, causing the word to become clearer.

The amount of visual information varies both within a trial, and across trials. Within a trial, a player can see up to 10 spatially separated exposures of the stimulus word (each at a higher level of clarity), followed by additional exposures (superimposed on the tenth display position) which continue until all bits are turned on. The number of pixels (out of 64 per character) in the fourth exposure is referred to as the "mean pixel density" and is used as a reference point in calculating the number of pixels available at the other exposures within a trial. The pixel density of the fourth exposure corresponds to the mean pixel density because the program sets the starting pixel density to a value such that, on the average, the student will identify target words on the fourth exposure.

To illustrate, if a student's current mean pixel density were 34 pixels per bit map, then the number of pixels per letter available at Exposure #4 would be 34. A decrement of 2 pixels per letter is made to calculate the pixel density for each of the three preceding exposures: thus, the number available at Exposure #3 would be 32, and the number available at Exposure #2 would be 30, and #1 would be 28. An increment of two pixels per letter is likewise made to calculate the pixel density for each of the exposures that follow Exposure #4. Because we want to motivate trainees to use less visual information to identify a word than they used on average in previous trials, the scoring system is set up to reward responses to Exposures #1, #2, #3, and #4 more than responses to Exposures #5 and greater.

Using mean pixel density as the point of reference, the number of pixels available across trials is adjusted to reflect changes in the level of a trainee's performance. Depending on the progress a trainee makes with the task, the amount of visual information available across trials is either decreased or increased. The number of pixels available at exposure #4 is based on a trainee's mean performance on the previous three trials. Thus, if a trainee responds on the average after the first, second, or third exposure, the pixel density of Exposure #4 and if he is experiencing difficulty the pixel density at Exposure #4 is increased to equal his current mean performance level. The effect of this rule is that as a trainee becomes more proficient he or she must perform the task with decreasing amounts of visual information.

b) Response time. Response time is the second dependent measure used in the training task. The length of time during which a trainee can respond to an individual exposure is lengthened or shortened depending upon his current level of expertise. Shortening the length of time a trainee has to respond to a given exposure forces the trainee to integrate the information available from contextual and perceptual sources more efficiently. The length of time available for integration also contributes to the level of challenge of the training game.

The period of time available for integrating and responding is referred to as the "response window". The response window is operationally defined as that block of time during which a trainee's key press is considered to be in reaction to exposure n; a key press that occurs after that block of time is considered to be in reaction to exposure n+1. Because the time required to initiate and execute a key press cannot be less than 200 msecs, the response window associated with each exposure extends 200 msec into the display time of the subsequent exposure. Thus, the response window associated with exposure n is comprised of the interval following the nth exposure (but excluding the first 200 msec of that interval) plus 200 msecs of the display interval of exposure n+1.

The response window is updated every fifth correct target trial to reflect a trainee's mean response time to a given exposure. We experimented with various algorithms in order to find one that would track the changes that were occurring in a trainee's level of expertise and yet not be overly sensitive to the performance on any single trial. The following algorithm seemed to meet these specifications: new response window = previous mean response time + (.5 * (new mean response time - previous mean response time)). This has the effect of (a) reducing the response window when on average, a trainee responds faster than his previous mean response time, and (b) increasing the response window when, on average, a trainee responds slower than his or her previous mean response time. This rule encourages trainees to integrate visual and contextual information more efficiently in order to maintain success in the game. Information about this measure is not given to the student.

8.2 Game Format and Graphics

Game format. The game theme centers on sending and receiving code messages in order to protect friendly spaceships from the attractive force of an enemy space station. The player is a spy on the enemy station who has communications equipment which can be used to detect whether an incoming ship is a friend or foe, and in the case of a friend, to warn it off while it is still far enough away to escape the attractive force of the station. The player must decide whether an incoming spaceship is a friend or an enemy. Because the enemy spaceships are disguised, a code message must be used to distinguish friend from enemy.

When a ship approaches, the player hits the space bar to send out an English code message in which the last word is left missing. Friends understand English and can send back words that make sense in the sentence (i.e., targets). Enemies do not understand English well and cannot send back words that make sense (i.e., foils). But enemies are clever; they use an onboard computer to generate English words that they hope will fool the player into letting them get close enough to the station to pose a threat. The player's job is to protect friendly ships by identifying as quickly as possible whether or not the return message (word) makes sense in the context of the coded message, and thus whether an incoming ship is friend or foe.

Due to noise in communication, words coming from an incoming ship are less clear when the ship is distant and become clearer as the ship gets nearer. The player's task is to recognize the incoming word as appropriate (in the case of a friend ship) or inappropriate (in the case of a foe) to the message context as early as is possible, that is, while the ship is far enough away from the station to escape. Thus, the player must recognize words when their pixel density is low. The player's score is the pixel density at which he or she is able to recognize a word and judge its acceptability, and the player must strive to get as low a score as possible. If the player is successful, a warning beam will be sent out and the friend ship will swoop away and avoid crashing into the station. In the process, it will deliver to the player a filled tank of fuel. If the player is unsuccessful, the friend ship will crash, as it cannot escape the attractive force of the station, and the player loses a tank of fuel. It is also important to keep foe ships from landing. If they are sent a special "all clear" beam, they will not land and will disappear into space. However, if they are erroneously sent a warning beam, they will land on the station and steal a tank of fuel. To be successful in the game, the player thus must successfully warn off friends when they are far from the station and send "all clear" signals to the foes to keep them away as well.

To identify a word as a target (and send a warning beam to a friend), the player hits the key marked "YES". To identify a word as a foil (and send an "all clear" beam to a foe), the player hits the key marked "NO". To help the player distinguish when he or she is correct or incorrect, scoring tokens, corresponding to full and empty fuel tanks, are used. Filled in tokens (filled circles) are received when an incoming word from a friendly ship (a target) is successfully judged "correct", while unfilled tokens (open circles) are received when errors are committed to any type of incoming word (target or foil). Filled tokens are most valuable when they are obtained while the ship is still at a distance (and the word is still unclear). The more closed tokens received and the earlier they are received, the greater the effect on the player's score. Correctly responding to return messages from foe ships (e.g., to foils) does not influence the player's score. (This is because the skill being developed, use of context, does not facilitate the early recognition of words that are unrelated to context.) However, errors in rejecting foils do count heavily against the player, so accuracy in judging foils is essential. Error tokens count as though a full ten exposures had been used, and this tends to increase the subject's pixel score.

(b) Graphics. The initial display is of a text window that occupies the top third of the screen and contains a sentence frame. Beneath the center of the text window is a filled-in area called the "word chute", from which the masked words descend. Underneath the bottom-right corner of the text window is a box that will display the subject's lowest mean pixel density to date. When the trainee hits the spacebar, the rest of the display appears: the text window remains in the upper third of the screen, and the game board fills the lower two-thirds. The generator that sends out the "YES" (warning) beams is situated in the left-hand corner of the screen and the generator that sends out the "NO" (all clear) beams is in the right-hand corner. When the trainee hits the spacebar a second time, the first of the exposures drops from the word chute. When the exposure duration for the first exposure has elapsed, the second exposure appears directly underneath the first. This sequence is repeated until the trainee responds or until all ten exposures have been presented. If the subject still hasn't responded, the word will continue to be filled in at the 10th exposure position until the S responds.

When a trainee presses the "YES" key, waves are sent from the generator in the left-hand corner toward the masked word display. The ensuing action and graphics depend on whether the response was correct or incorrect. If "YES" was the correct response, the word is displaying in its entirety, moved to the right and is converted to a filled circle at the right in a location opposite to the exposure of the response. If a "YES" response was incorrect, then the word is displayed in its entirety and crashes down to the tenth exposure location and is converted to an open circle at the far right, indicating an erroneous response. The sequence of events and graphics are similar when the trainee presses the "NO" key, except that the generator in the right-hand corner emits the waves that hit the masked word and a correct "NO" response pushes the word to the left and off the screen with no further consequences. If the "NO" response was incorrect, the word is filled in in its entirety and converted to an open circle. The graphics are accompanied by sound effects that provide additional feedback and reinforcement to the trainee.

After a student has accumulated three circles, either filled or open, a new mean pixel density for the fourth exposure is updated, accompanied by sound effects that reflect whether the change is an increment of a decrement. After every 5 events, the response window is adjusted, unbeknownst to the student.

8.3 Textual Materials

As before, there are two levels of challenge represented in the set of context sentences, those that tightly constrain a single semantic domain and those that permit items from multiple semantic domains. Further, associated with each sentence frame, there are items of high probability and items of low probability. These materials have been edited to (1) reduce the total number of items associated with each sentence frame and (2) widen the range of concepts to be considered within a semantic domain.

(1) Number of items per frame. Because the most important aspect of the training task is the use of contextual information to gain access to word meanings,

the number of response items per sentence has been reduced to an average of 3.9 for high constraining contexts and 5.5 for low constraining contexts. Decreasing the ratio of response items to sentence frames increases the number of different contexts that a trainee will use during training. To achieve this, target words that were highly synonymous with other targets in a given sentence frame were eliminated and the number of high probability items per sentence frame was reduced to one. The number of foils was likewise reduced to maintain the ration of 1 foil to 2 target words (approximately 33%)

The range of concepts. To encourage trainees to consider a wide range of concepts and to rely less on words that they have been previously exposed to in recognizing later-occurring items, highly synonymous targets have been replaced, whenever possible, with contextually-appropriate, less synonymous words.