

A WORD FREQUENCY DISTRIBUTION
STUDY OF LANGUAGE PRESENTED TO
YOUNG ESL STUDENTS

By

KIM REBANE

B.A. The University of British Columbia, 1980

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Faculty of Education

Department of Language Education

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December 1983

© Kim Rebane, 1983

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of LANGUAGE EDUCATION

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date Jan. 4, 1984

ABSTRACT

The purpose of this study has been to assess how well ESL children are being prepared to communicate in the English language. This was done by comparing the language presented to young ESL learners with the target language (English).

Word frequency was the basis of comparison in this study. The frequency distribution of words in the target language was compared with that of the ESL text series YES!. Published word frequency lists were used to determine how well the sample represented the target language. Comparisons were made on the basis of frequency distribution, high and low frequency words, and similarity to basal reading series designed for young native speakers.

It was found that young ESL learners are being exposed to language that is representative of what is needed to communicate. Results also showed that this language is unlike that of basal reading series which focus on many more repetitions of individual words. Given the different experiences with which young ESL and native learners bring to the task of learning how to read such a difference in the series is necessary.

The results of this study are discussed in terms of the frequency distribution of words, research in the learning of first and second language, and pedagogical implications of the findings.

TABLE OF CONTENTS

	<u>Page No.</u>
I CHAPTER 1: INTRODUCTION	1
Background Information	
The Purpose	
Assumptions and Delimitations of Study	
Justification of Study	
II CHAPTER 2: REVIEW OF LITERATURE	18
III CHAPTER 3: METHODOLOGY	36
Research Design	
Statement of Hypotheses	36
The Sample	39
The Procedure	41
Description of Variables	42
Summary	
IV CHAPTER 4: PRESENTATION OF DATA	49
Descriptive Analysis	
Inferential Analysis	
Summary	
V CHAPTER 5: INTERPRETATIONS AND CONCLUSIONS	85
VI BIBLIOGRAPHY	101
VII APPENDIX	107

LIST OF TABLES

	<u>Page No.</u>
Table 1: Summary of all Types and Tokens Found in the <u>YES!</u> Series	49
Table 2: Percentage of Tokens and Types Accounted for by Divisions of the First 1000 Words	51
Table 3: The Type/Token Relationship for 3 Published Words Lists and the <u>YES!</u> Series List	56
Table 4: A Rank Order List of the 321 Most Frequent Words in the <u>YES!</u> Series	59
Table 5: Pearson Correlation Coefficients for 3 Published Words Lists and the 321 Most Frequent Words in the <u>YES!</u> Series	64
Table 6: Likelihood of Words in the Published Lists having the Same Ranking as Words in the <u>YES!</u> List	65
Table 7: Summary of Word Distributions for the 6 Books in the Series and for the Series	67
Table 8: Pearson Correlation Coefficients for each of the Books in the Series	70
Table 9: Number of High Frequency Words in Each Book	71
Table 10: Recurrence of Least Frequent Words and their Contribution to the Total Word Count	72
Table 11: Distribution of the Least Frequent Words in Each Book of the <u>YES!</u> Series	74
Table 12: A Comparison of Tokens, Types, and Type-Token Ratios for the <u>YES!</u> Series, the <u>Ginn 720 Series</u> , and the <u>MacMillan Series</u>	77
Table 13: Commonality of Words Found on the <u>YES!</u> List with Words Found in the <u>Ginn 720</u> and <u>MacMillan</u> Lists	78
Table 14: Percentage of Word Types Occurring Only Once in Each Series	79
Table 15: Rank List of the 106 Words Unique to the <u>YES!</u> A & B Books	80
Table 16: The Number of New Words Introduced in Each Book of the <u>YES!</u> Series and the "Typical" Number of Words Introduced by a Basal Reading Series	83

LIST OF FIGURES

	<u>Page No.</u>
Figure 1: Percentage of Tokens Accounted for by 35.7% of Types	53
Figure 2: Percentage of Tokens Accounted for by the 500 Most Frequent Words in the <u>YES!</u> Series and 3 Published Words Lists	55
Figure 3: A Graph of the Number of Different Words in Each of the 6 Books in the <u>YES!</u> Series	68

CHAPTER I

An adequate sight word vocabulary is essential for fluent reading (Dolch, 1960). A sight word is one that is immediately recognized by the reader without the necessity of phonic, structural, or contextual analysis. A sight word vocabulary is made up of all the words a reader can immediately identify without taking the time to analyze. Hildreth (1958) said that these instantly recognized words are part of our "word banks". She suggested that one measure of reading maturity was the size of this word bank. A large word bank allows the reader to read faster and more accurately without having to stop and figure out words by such word identification strategies as phonetic or structural analysis.

A large sight vocabulary allows the reader to proceed through the reading material in the manner which has been described by Kenneth Goodman. Goodman (1967) views reading as a "psycholinguistic guessing game" whereby the competent reader actively involves himself in the selective information-seeking process of determining meaning. However, this view of the reader describes only the individual who has a great deal of experience with the language in that he is able to recognize and remember the most productive language cues in order to make predictions and associations.

When teaching children how to read we are really asking them to analyze and then synthesize the print on the page. While there has been a great deal of discussion about what that 'unit' is, most experts in the field recognize the importance of the word as the central meaningful unit for learning how to read. One method of teaching reading which is employed at some point in most reading programs is the Whole-Word or "Look and Say" method. This method focuses on teaching children a number of words by sight in an effort to foster early successful reading experiences. The key to this method is that the words chosen for such learning are meaningful. That is, the child already knows what the word means because he uses it in speech. Thus, early reading instruction is not concerned with teaching new words or meanings but, rather, with developing a recognition or sight vocabulary of words whose meanings are already familiar to the child (Causey, 1958). This sight vocabulary is built up through repeated exposure via seeing the word printed, saying and talking about the word, using the word orally, defining the word, and copying the word (Dauzat and Dauzat, 1981).

Wayne Otto, Robert Rude, and D.L. Spiegel (1979) have pointed out four important reasons for children to develop a large sight vocabulary. First, if the child has to concentrate on every individual word, he will fail to comprehend the whole passage because his limited memory span will not permit him to make meaningful connections between the various parts. Second, an adequate sight vocabulary places limitations on the reader's use of the important word identification cue of context. Otto, Rude, and Spiegel (1979) suggest that the reader

needs to be able to recognize 95 percent of the words in the passage in order for the material to be truly meaningful. A third advantage to having a large sight vocabulary is that the sight words can act as catalysts for teaching phonic skills. Since not all words can be taught as sight words, it is necessary to have some method whereby the reader can figure out the word without help. Finally, for words that are phonically irregular (i.e. they can not be sounded out), the sight word strategy is indispensable. The sight word approach is a useful strategy in that initial success at reading a short story with many word repetitions gives the child a feeling of confidence and enthusiasm to continue.

It has already been noted that initial sight vocabulary is developed on the basis of what the child already knows. That is, words that are already within the child's experience are learned first. Many word lists have been developed in an effort to guide teachers and authors of books in deciding which words are most frequent and therefore should become part of the reader's sight vocabulary. Such lists are based on spoken vocabulary (The International Kindergarten Union Study, 1928; Murphy's "Spontaneous Speaking Vocabulary of Children in Primary Grades," 1957), written vocabulary (Rinsland's Basic Vocabulary of Elementary School Children, 1945; Hillerich's "240 Starter Words," 1974), and printed vocabulary - those words found in reading material (Carroll et al. Word Frequency Book, 1971; Johnson's "Basic Vocabulary for Beginning Reading, 1971; Harris and Jacobson's Basic Elementary Reading Vocabularies, 1972). These lists are based on relative frequencies of words used in the medium which is being studied (be it reading material, samples of writing, oral material, or a combina-

tion thereof). The basic premise of this method of rating words is that the child will encounter certain words more often and therefore needs to be familiar with them so that the words do not interfere with learning to read fluently.

Basal reading series have used such published word lists to develop children's sight vocabulary. These series are carefully structured so as to present words that the child is familiar with before those with which he is unfamiliar. Dolch (1960) outlined the relationship between the child who is beginning to read and the typical reading series that were being used at that time. He said that while the child comes to the first grade with a meaning vocabulary of several thousand words and a sight vocabulary of perhaps fifty words, pre-primers assume that the child has no sight vocabulary at all. The words that the child has seen repeatedly on signs, labels, and TV are not exploited as they should be. Furthermore, Dolch also said that the adding of new words throughout the text series was not really done on the basis of any sound pedagogical strategy. The only criterion for the vocabulary structure seemed to be that the fewer new words there were, the easier the book was to read.

Dolch (1960) went on to describe the vocabulary load of such series. Naturally, the number of new words per book increases from level to level within a series. Typically, the pre-primer has 50 new words; the primer, 100 new words; the first year books, 150 new words; the second year books, 400 new words; and the third year books, 600 new words (Dolch, 1960). Basal reading series adhere to the learning principle of repeated association in an effort to make this new

vocabulary familiar. To do this, a new word will be repeated a number of times within the book in which it first appears and again, along with the other new words of that book, in the next book of the series. Dolch said,

It makes sure that while adding new words, the old ones won't be lost by disuse. A poor plan is to teach a child ten new words and at the same time let him forget ten words taught previously. This results in no increase in sight-vocabulary. Therefore, as we plan a steady learning of new words, we also plan a continued re-use of old words. These two elements make up what is called vocabulary control, which is absolutely essential for maintaining and increasing sight vocabulary in the most efficient way in school readers.

(Dolch, 1960. p. 265)

Recently, Robert Aukerman (1981) published a book which gives up-to-date information on basal readers in general and reviews several basal reading series that are currently on the market. He describes the basal reading series as having four components:

1. the series of 15 or 16 books starting with the pre-primer and going up through sixth grade (although some continue on through junior high school);

2. the teacher's edition which explains how to teach the lesson;
3. the pupil workbooks which are designed to reinforce what has been presented in the readers; and
4. the management component which involves testing to determine the child's strengths and weaknesses and whether he is ready to proceed to the next level.

In summary, Aukerman says,

A basal series is planned to present very simple, easy-to-master materials and method in the first-grade materials. The second-grade materials are somewhat more advanced, but build on the skills mastered in the first grade. And not until about the third grade does the pupil begin to top off his/her word-recognition and comprehension skills. The content and materials in the intermediate grades (4, 5, and 6) are usually related to the learning of literary skills and the reading of a wide selection from the pupil's literary anthologies at these grade levels.

(Aukerman, 1981, p. 7)

At the end of this book, Aukerman lists twenty disadvantages and nine advantages of the new basal reading series. Among the advantages, two are concerned with vocabulary development: "a sequential program of vocabulary

development" and "a developmental plan of word-analysis techniques" (Aukerman, 1981. p. 333). A cursory examination of the fifteen basal reading series Aukerman describes reveals the fact that over half of these make direct reference to sight words (or some synonym thereof - such as 'foundation words' or 'basic words/vocabulary') and the other programs, which start with a phonic approach, aim for sight mastery. Aukerman states that several of the series have strict vocabulary control and that many of the series base their vocabulary on high-frequency word lists.

Dolch (1955) described how word frequency lists should be used in developing reading programs when he explained the dimensions of a list of the "First Thousand Words in Children's Reading". He pointed out that there were two kinds of words in the list, areas of experience words (i.e. words associated with nature, school, home environment, child's person and clothing) and general words (i.e. words such as 'begin', 'think', 'with', 'these', 'both', 'seven'). General words are those used in a wide range of situations. However, such words are only of value within specific situations since it is impossible to talk about one of these words without putting something, with it (i.e. 'begin' something, 'with' something, 'seven' something) (Dolch, 1955). Since these words are general and will be met over and over again, they should quickly become part of the reader's sight vocabulary (Dolch, 1955). That is, regardless of which of the areas of experience a story happens to be written in, a selection of general words should be made. Dolch published a list of 220 general words and pointed out two important facts:

- 1) grammatically, these words are pronouns, adjectives, prepositions, and

conjunctions - no nouns, and 2) 'physically' these words make up 70 percent of the first grade readers, 66 percent of the second and third grade readers, and over 50 percent of all other reading materials (Dolch, 1955).

Knowing a language may be said to involve a sufficient knowledge of its grammar to enable comprehension and creation of novel sentences in the language, and a knowledge of sufficient vocabulary to permit communication in situations for which the language is required.

(Richards, 1974, p. 69)

He goes on to point out that what is thought to the second language learner is largely a matter of choice and that the selection necessarily implies that some features of the language will not be taught. With respect to vocabulary selection, the choice involves a subjective, objective, or subjective-objective consideration of the contexts in which instruction and use will occur (Richards, 1974).

Richards explains that a subjective approach is based on the instructor's intuition of what vocabulary the learner will need, while an objective approach focuses on word frequency counts which produce word lists of which the most frequent are believed to be the most useful. Subjective-objective approaches use both word frequency and such psychological measures of word utility as

availability (i.e. the ease of recall of words based on how they are structured in memory) and familiarity (i.e. a subjective response to words which is based on the word's meaningfulness and concreteness as well as the frequency of experience with the word) (Richards, 1974).

The purpose of this study is to examine the written language presented to ESL students and determine if the language is adequately representing the target language. ESL students must eventually be able to use effectively. More specifically, written material will be studied so that the basis for selection of words can be described. To do this, the words presented to your learners in an ESL text series will be examined (via frequency counts) and compared with that of the target language.

The ESL text book series YES! (Melgren and Walker, 1977/78) has been chosen to represent the written language young English learners need. The words in this series are examined in terms of development of vocabulary within the series, similarities between the frequency distribution of words in the series and words in the target language, and differences between the words used in the ESL series and those used in basal reading series designed for native speakers. By doing a word frequency count (determining how many different words there are and how many times each of those words occurs). Comparisons and correlations will be made to describe the development of the ESL series YES! and evaluate the selection of words.

The YES! series is designed for children who are learning English. It has been chosen because of its widespread use with young ESL learners. Physically, the series is very similar to basal reading series - there is a series of six books organized according to level of ability, three workbooks which reinforce and expand into the written form what was presented in the books, and a teacher's edition which gives instructions on how to present the material. However, ESL children are in quite a different position when they enter school than native speaker children. ESL children have little experience with the language and so have no oral meaning vocabulary upon which to base reading instruction. Thus, because of this lack of exposure to English, the ESL child must learn many skills at once - listening and speaking meaningfully in the English language as well as reading and writing meaningfully. This means that the books used must take on the added burden of developing a complete language program. The words that are used in such texts cannot be entirely consistent with basal readers which focus on reading because the contexts of language use are much broader. However, such books must be representative of standard English in order to develop ESL children's ability to use the language productively.

The authors of the YES! series, Melgren and Walker, have suggested that teaching English to young ESL students requires that the language used be utilitarian. That is, the choices concerning what language to expose the child to involves asking what language the child needs to communicate. Linguistic analysis involving abstract concepts about English is often useful for older ESL

learners but young children learning the language cannot deal with language on such a level (Melgren and Walker, 1978). However, Melgren and Walker do not suggest any basis upon which to decide what is "meaningful".

For the purpose of their series, Melgren and Walker have created lists of vocabulary words and expressions for each book to represent what should become part of the child's working vocabulary. The only criterion for a word to occur on such a list is that it occurs more than once in a particular book. This qualification ensures that words that must occur once strictly because the context demands it are not overemphasized). However, this also means that while a word may have a high frequency and/or high utility in one book of the series, it may not even occur in any of the other books. Subjectively, we may feel that such words are of high general utility in the child's language development. That is, the children will use such words in other situations. But since these words occur only within one book of the series, they may appear not to have high frequency of general utility when looking at the overall frequency of words for the series. Teachers should know about such words so that they may emphasize them in other language activities.

Since the series offers no word frequency information, it is difficult to determine which words occur in all the books of the series and which words are specific to a single book. Published word frequency lists may be of little value in such a situation simply because the vast number of words in general use are low frequency words and it is impossible to predict which low frequency words will be used in a particular series. (The high frequency words, on the other hand, are

almost pre-determined since they are the structure words that unify the language). Richards (1974) has pointed out that many published word frequency lists omit words that are of high utility (i.e. of useful, practical value in specific contexts). For example, the words 'soap', 'soup', 'dish', 'oven', 'chalk', and 'stomach' do not occur within the first two thousand most frequent words published by Thorndike in 1921 (Richards, 1971). Thus the decision regarding which low frequency words to include in a list may simply be dependent on the language sample taken (i.e. the context) and therefore in no way reflect what is really used or needed to communicate successfully.

While we might expect more low frequency/high utility words in the ESL series because of a concern to expose learners to as many useful words as possible, we cannot foresee which of these types of words will be introduced or to what extent they will be used. Freeman Twaddell (1980) has said that at the time the ESL student reaches the intermediate stage of learning the language, he still has an extremely small vocabulary. However, the decision as to which words should be concentrated upon is a difficult one. Rivers (1981) said that the most frequent words that the learner will encounter should naturally be the basis for decision. However, since these are relatively low information words, we need to ask which words that are of low frequency but of high utility need to be included?

There are certain difficulties in working with words, word lists, and word frequency which should be noted. Barry Richman, one of the authors of the Word Frequency Book, has cited five characteristics of the lexicon that make it such a nebulous system to work with:

- 1) it may be regarded as infinite;
- 2) there is more than one reasonable way to define its elements, and it is not always clear how to distinguish one element from another;
- 3) the structure of the lexicon is interlaced with the grammar of the language;
- 4) the lexicon changes with time; and
- 5) there are important differences in the way the lexicon is used in speech and in writing.

Carroll, Davies, and Richman, 1971. p. v)

There are also a number of idiosyncratic aspects of word lists that should be pointed out because this study involves comparing the frequency of words in a devined text with the frequency of words in the target language (i.e. the language being learned) as defined by word-frequency counts. While there have been several word lists published, they are not necessarily comparable because they may not consistently sample the same data base in the same way. A summary of the dimensions of selection that will affect the structure of the word-frequency lists is outlined below:

- 1) The definition of the 'word' is of upmost concern to the development of any list. Some lists subsume morphological endings under the root word as bounded by space on the right and left (and, thus, the presence of the plural -s makes the word "book" difference from the word books).
- 2) The medium of the language source; as Richman, (1971) pointed out, oral and written material are not exactly alike. Furthermore, what we read is not the same as what we write (i.e. our receptive and productive vocabularies differ).
- 3) If a specific medium is chosen, the particular types of words will reflect certain dimensions of that medium. For example, with respect to reading material, general magazines are composed of quite different words than classical novels or technical reports (i.e. the content of the reading material will affect the distribution of words).
- 4) Recognizing that adults have much different vocabularies than children is important when using and developing word lists. In comparing word frequencies it is necessary to note the age range for which the list represents.
- 5) Finally, because words change over time, it is important to note the publishing date of the word list. There are thousands of common words in our vocabulary today that were not even in existence when some of the early lists came out (i.e. 'television', 'computer', 'jet').

Characteristics of word-frequency lists aside, all lists aim to reflect which words are most frequent within the medium studied. Their stated usefulness ranges from language development (i.e. in the teaching of phonics, spelling, and English to non-native speakers) to research concerns and textbook design (Earle, 1977). However, the value of word frequency lists is only as valuable as the underlying assumptions which prompted its development. The general assumption is that the frequency of a word does make a difference. More specifically, in the domain of readability, the more frequently a word occurs, the easier it is to understand meaningfully and, consequently, the easier it is to read the material. The more words there are that are 'easy' to read in a passage, the easier the passage is to read.

From this basic premise it is easy to see why there might be a great deal of enthusiasm for word-frequency counts. Here we have a quantifiable way of determining what makes a passage easy or difficult to read. Unfortunately, there are other dimensions of readability which are not quantifiable but which are just as important to the reading difficulty of the passage. Dolch (1955) has pointed out that while words are the basic building blocks of the meaning, other dimensions of the language also affect the readability of any material. The reader's span of attention and memory play important roles in the ability to understand the words and the development of meaning. Thus, sentence length, word and phrase order, and experience with the context of the material being read also play important roles in determining the difficulty of a passage (Dolch,

1955). Furthermore, the fact that most words have more than one meaning also affects how difficult a word will be. A word may have one meaning that is very common and a second meaning that is quite rare. Word frequency lists do not take this characteristic of words into account; generally, the meaning aspect of words is ignored when word frequency is studied.

Finally, and perhaps the most important criticism of word frequency lists and related studies is that there is a difference between a meaning vocabulary and a recognition vocabulary (Dolch, 1955). Recognizing a word does not necessarily mean understanding that word. The situation can be likened to recognizing someone's face but having no idea where you have met or seen him before. Recognition without meaningfulness is of little value.

Meaning vocabulary is basic to understanding. An extremely weak meaning vocabulary is characteristic of someone learning English. Therefore, a major concern of an ESL teacher is to determine which words will constitute the meaning-vocabulary and at what point they will be taught. Word frequency lists provide a references to words the ESL student will hear, speak, and read most often. These words are the ones students should be taught early. Having a good grasp of the high frequency words gives the learner a context which can be used to attend to the task of recognizing and understanding the vast number of words that are not known to him. In other words, a word frequency list can help the teacher determine which words need to be taught early so as to ease the memory load in the face of the vast number of unknown words.

The remainder of this paper will deal with the central issue of word frequency. Chapter 2 reviews the frequency structure of words in the English language and looks at studies relating word frequency to recognition and meaning. Chapter 3 describes the data source and procedures for data collection and analysis. Chapter 4 is a presentation of the results and Chapter 5 discusses the findings specifically in terms of the YES! series and generally in terms of the usefulness of word frequency lists to the ESL teacher.

CHAPTER 2

The purpose of this chapter is to examine the distribution of words in the English language and to review related studies that indicate the relationship between word frequency and recognition. Studies concerned with the speed of recognition of individual words and experiments dealing with the relationship between high frequency words and readability are cited in an effort to show why there has been such a great interest in word frequency as one of the indicators of readability. A brief review of the "how and why" of some of the major published word list studies concludes this chapter.

Freeman Twaddell (1972, 1980) describes the frequency structure of the vocabulary as the quantitative aspect of vocabulary which creates difficulties for ESL students learning to read. Twaddell (1980) explains that the frequency-distribution is not what most people would expect. We would expect that graphic representation of the frequency distribution of words in the language would approximate a bell-shaped curve with very few very high-frequency and very low-frequency words and many medium-frequency words. However, quantitative analysis does not support this prediction. The actual distribution can be diagrammatically likened to a ski-jump (Twaddell, 1980). There are very few high-frequency words, a small number of medium-frequency words, and, contrary to expectations, a seemingly infinite number of very low-frequency words. Thus, based on Twaddell's description, the distribution of the vocabulary may be summarized as follows:

1. The highest frequency words are actually very few in number and part of all language used (i.e. articles, conjunctions, prepositions, pronouns).
2. The medium-frequency words are determined by the context in which the language is being used. These words organize the discourse.
3. The very low-frequency words are those which unite the area of interest to the particular situation in which language is being used.

Related to this description is the observation of how quickly the curve tapers off from the few high frequency words to the words which appear only once. In a table of most frequent words based on a corpus of 1,104,235 words put together by Kucera and Francis (1967), Twaddell (1980) shows that the ten most frequent words account for almost one-quarter of the corpus. Furthermore, the table reveals a vast difference between the occurrence of the most frequent word 'the' and the one hundredth most frequent word, 'down'. While 'the' occurs approximately once every fifteen words, 'down' occurs only once every 1133 words. Twaddell points out that this extremely early tapering off of high frequency words means that it is impossible to predict which words a student will encounter in reading.

That word frequency is a good predictor of word difficulty and order of acquisition has been a point of much debate. Wardaugh (1971) questions the importance of the frequency of a stimuli on the basis of evidence drawn from language acquisition studies. For example, he cites studies that show:

1. The telegraphic speech of children omits the most frequent words in the environment.
2. Japanese children acquire a less frequent grammatical form before a more frequently occurring form (McNeil, 1966, 1968).

However, Ingram cites a number of studies in which the relative frequency of a structure is important to the acquisition and use of language by the child:

1. The early learning of the questions "What's that?" and "What doing?" is the result of the high number of 'what' questions presented to the child by older people.
2. Frequency corresponds to simplicity of structure of sentence forms (i.e. the simplest being declarative, active, affirmative). Studies have shown that the least complex sentence forms are learned by young children earlier than the more complex forms.
3. The more concrete references, which are of very high frequency in the speech presented to children, are the ones that children acquire first (Ingram, in press).

Lefevre (1962) does not see the single word as a major language unit because of its relative semantic and syntactic instability, its meaninglessness in isolation, and its insignificance when considered as part of the larger units of language (i.e. the sentence or the paragraph). However, a great deal of research

shows results suggesting that the word is important - especially when considered as part of the larger unit. Before looking at these studies, it must be recognized that word frequency is used as a predicator of word familiarity. That is, the most familiar words are those words which occur most frequently in the language. Dolch (1960) refers to these very familiar words as 'sight-words' which are instantly recognized and do not cause hesitation. Expert reading implies a large sight vocabulary (Dolch, 1960). Many word count lists have been developed to indicate which words are the most familiar (i.e. Thorndike's Teacher's Word Book (1921); Dolch's Basic Sight Vocabulary (1936); Rinsland's A Basic Vocabulary of Elementary School Children (1945); Dale's List of 3,000 Familiar Word (1948); Carroll, Davies, and Richman's Word Frequency Book (1971); Harris and Jacobson's Basic Elementary Reading Vocabulary (1972).

Studies dealing with word frequency can be divided into those concerned with the individual word and those concerned with the word within larger units of language such as sentences and paragraphs (i.e. those dealing with words within a context). To understand the theoretical basis for the value of word familiarity (as measured by word frequency), attention will first be given to those studies which deal with our ability to identify individual words. These studies typically deal with the concept of familiarity by measuring the response time for tachistoscopically presented material. The pioneers in this area were L. Postman, R. Solomon, D. Howes, and C.E. Noble (1950, 1951, 1953, 1954).

A series of studies has shown that high-frequency words are recognized quicker than low-frequency words. Howes and Solomon (1951) correlated the speed of recognition (the length of time the word was in the subject's visual field) of a word with frequency of occurrence of that word (based on three published word frequency lists) by flashing words on a screen for identification by subjects. They found that almost 51 per cent of the total variance was accounted for by the log word frequency being correlated with the duration threshold. In other words, they found a very clear inverse relationship between word frequency and duration threshold: the more frequent the word, the shorter the duration of stimulus necessary for identification.

Postman and Solomon (1950) showed experimentally that the recency of the stimulus (i.e. how recently the stimulus was last seen by the subject) was associated with the duration of presentation. Solomon and Postman (1952) point out the intimate relationship between recency and word frequency: the more often a word occurs, the more likely it is to have occurred recently. In their 1952 study, Solomon and Postman showed that even the learning of nonsense words was a function of recency and familiarity. By using nonsense words they could control the frequency (or familiarity) variable. After having subjects read and pronounce a series of nonsense words which ranged in frequency from one to twenty-five, the recognition thresholds (the speed of recognition) for each word was determined by tachistoscopic presentation. Once again, it was shown that the speed of recognition was positively correlated with familiarity.

Noble (1954) noted the results of the Solomon and Postman (1952) data and proposed to evaluate the functional relationship between familiarity and frequency of stimulation. He found an index of relationship between frequency and familiarity of .998. This extremely high correlation, along with the Solomon-Postman results, strongly suggests that "familiarity is a learnable attribute of the stimuli" (Noble, 1954, p. 14).

More recently, Mason (1976) has looked at how orthographic, phonological, and word frequency variables affect the speed of word recognition. She used letter sequences of the form CVCC to determine how vowel regularity, initial consonant frequency, final consonant frequency, and word familiarity affected word-nonword decisions made by children and adults. Her major finding was that word familiarity was the primary factor in such decisions.

The model developing out of this perspective is that of LaBerge and Samuels' (1974). They suggest that common words are processed differently from uncommon words. Very common, or familiar, words are automatically coded into a visual word code which excites meaning. Uncommon, or unfamiliar, words, on the other hand, are coded into orthographic spelling and phonological patterns before the reader can obtain meaning. The result of this is that it takes longer to process unfamiliar words. She cites LaBerge and Samuels (1974) when she says that her results support the belief that the major influence on automaticity of recognition is word frequency of usage: the influence of orthographic and phonological variables seems only to be through their interactions with word familiarity. She explains:

Shorter decision time for common words can be understood in terms of the set of words that must be searched, because, by definition, the number of stored words containing high frequency letters exceeds the number of words containing low frequency letters. The effect is reversed for uncommon words since these often begin with low frequency initial consonants.

(Mason, 1976, p. 205)

Finn (1978) reanalyzed data collected by Bormuth in 1966 to explain the positive relationship between word frequency and the likelihood of its being supplied in a cloze passage. He concluded that extremely common words are supplied very readily because they are expected by the reader and, therefore, carry very little information. The reader does not spend the same amount of time or effort on each word. Depending upon prior choices made in the reading of the passage and language/reading experience, words that occur more frequently will be given less time and attention (Goodman, 1967).

Marks, Doctorow, and Wittrock (1974) examined the relationship between word frequency and reading comprehension. Using subjects between the ages of 10 and 12, they found that reading comprehension could be significantly increased by substituting 15 per cent of the low-frequency words with higher frequency words. Their findings suggested that gains in reading comprehension could be obtained by manipulation of these words (nouns, verbs, adjectives, and adverbs) on the basis of their relative frequencies.

A more recent study by Graves, Boettcher, Peacock, and Ryder (1980) looks at the relationship between words and reading comprehension from a different perspective. They investigated how well students' reading vocabularies could be predicted by word frequency lists. After testing 432 seventh to twelfth grade students by administering two 43-item multiple choice vocabulary tests, they found that there was a positive relationship between the frequency of a word (as determined by Carroll, Davies, and Richman's 1971 word frequency list) and the students' response to that word: correct responses declined as the word became less frequent. Their results also suggested that other factors such as meaningfulness, pronunciability, letter frequency, and sequential probability of letters also played an important role in determining whether a student knew a particular word. Thus, they concluded that word frequency is really a rather crude measure of familiarity. However, the ability to pronounce a word can be argued to be of little consequence since the most frequent words follow few phonics rules (Otto, Rude, Spiegel, 1979). Furthermore, because of the way the 'word' was defined, several words which appeared to be less frequent were really merely derivations of more common words.

Thus, there is a great deal of empirical research showing that word frequency is an important factor in word recognition. The significant implication here is that word frequency affects the readability of materials. George Klare (1968) examined the role of word frequency in readability by reviewing a number of studies in the areas of word and sentence difficulty, reading

efficiency, word familiarity and recognition, and reading preference. He points out that word frequency studies were first undertaken in response to the fact that the common words were more comprehensible than the less common words. This means that passages which contain higher frequency words are typically evaluated by the reader to be 'easier' to read. In an effort to determine exactly which words were the most frequent, several studies have resulted in lists which define the frequency of words for a particular corpus of data. The remainder of this chapter will cite the reasons some of these lists have been compiled.

Howes and Solomon (1951) describe how word-frequency counts are made:

Word-frequency counts are made by selecting a sample of language behaviour (usually written) that contains a given number of words, and then tabulating the number of times that each particular word occurs.

(Howes and Solomon, 1951, p. 401)

The results of multiple sampling and tabulation are written up as lists which reflect the number of times a particular word occurs in the language. Richman (Carroll, Davies, and Richman, 1971) has pointed out that the more diverse the contexts and the broader the source of the corpus (i.e. the greater the range), the more reflective the resultant word list is of the distribution of words exposed to and used by the general population.

There have been many purposes cited for developing word frequency lists. Edward Thorndike, the author of several word lists, suggested three ways that his 1921 word list could be of service to teachers:

- 1) it helps the teacher decide what teaching strategies to use by telling him/her the relative frequencies of words;
- 2) it helps the novice teacher identify the important words and the words that are likely to cause difficulty, and;
- 3) it can act as a guide on how to teach certain words if notations are made by the teacher.

(Thorndike, 1921)

In his expanded version, The Teacher's Word Book of 30,000 Words (in collaboration with Irving Lorge), Thorndike adds that the list also allows teachers to know the importance of each word with respect to popular reading for adults and approved reading for children (Thorndike and Lorge, 1944).

Gates (1935) felt that his list, which consisted of data from speech and texts, had pedagogical value in facilitating reading if the words of all subjects were limited to one vocabulary. He also felt that these words would probably be the most widely used across the curriculum and were therefore worthy of inclusion on spelling lists. Furthermore, Gates suggested that the words of the list could be used in developing tests in the areas of reading and writing. He explained,

Tests of ability to recognize and pronounce the words singly, and especially to read with understanding various types of passages based entirely on words from different levels of the list, would indicate the range of the basal vocabulary and the degree of independent reading ability a pupil has achieved, and consequently, the extent to which he may be entrusted, without danger of practicing errors, with reading miscellaneous children's materials in the school or home.

(Gates, 1935. p. 3-4)

(Of course, a major assumption made here is that a child will not understand what he reads if he has not mastered the listed words. This is somewhat questionable since there are many other cues to consider then reading a particular selection).

Rinsland (1945) compiled A Basic Vocabulary of Elementary School Children because he felt there was a need to examine the words children in grades one to eight in use their own writing. He suggested the major use of such a list was in the area of writing books for children in these grades. Strothers, Jackson, and Minkler (1947) constructed A Canadian Word List: Grades I - VI as a first step in studying language development in Canada. They, too, saw the list as a basis for constructing reading materials.

More recently, Carroll, Davies, and Richman (1971) published the Word Frequency Book which used the resultant word list in the construction of The American Heritage School Dictionary for use by students in grades three through nine. Harris and Jacobson (1972) give practical and theoretical reasons for developing Basic Elementary Reading Vocabularies. The study was undertaken to reveal the words being used in the elementary textbooks at each grade level in 1970. They reasoned that most other lists were far too outdated to be of much pedagogical use. Practically, the advent of computer technology has allowed far quicker and more thorough studies of this nature to take place.

Harris and Jacobson also cited a number of additional purposes and uses for word lists. Among them area:

- 1) Studies comparing the content of this word list with other word lists,
- 2) determination of words which have risen or fallen in use over a period of time,
- 3) comparisions of the vocabulary content of specific books or series,
- 4) comparison of grade placement in the list with the measured difficulty of specific words,
- 5) studies involving cross-cultural comparisons
- 6) development of new variables for use in measuring readability.

(Harris and Jacobson, 1972. p. 3)

There are also many 'short' lists that have been drawn from the longer lists in an effort to produce basic, or core, vocabularies. The authors of these lists determine which words are common to several lists and then compile a summary list based on their findings. Such lists are formed with the intention of outlining the absolutely necessary words that a student must know in order to read. For example, Dolch (1936) published a list of 220 words which he said made up a least fifty per cent of the running words in elementary school reading materials (Harris and Jacobson, 1972). This list contains no nouns and so in 1950 Dolch compiled a list of 95 common nouns as well as a list of "The First Thousand Words for Children's Reading" (1948).

For a summary of general and core vocabularies published between 1930 and 1961, Harris and Jacobson (1972) offer a good overview and bibliography. Furthermore, Harris and Jacobson themselves have published a core list. They looked at fourteen series of textbooks written for students in grades one to six. Their core list is made up of words which appear in three or more of the basal readers. In 1979, Charles Walker published a word list of the one thousand words of highest frequency in the 1971 Carroll, Davies, and Richman study.

This chapter has cited evidence to show the importance of considering word frequency as an important variable in the recognition of words and the readability of materials. In knowing which words are most frequent, we know which words should become part of the sight vocabulary of the reader. There

have been no studies dealing exclusively with ESL texts. The lack of research in second language reading and vocabulary control is the result of a traditional focus on an oral approach to teaching language.

The Audiolingual method focuses on the oral skills of speaking and listening and sees reading and writing merely as a reinforcer of the oral skills. The skills involved in reading and, consequently, vocabulary development, are not among the goals of the Audiolingual method and so are ignored.

The second reason for the lack of research can be found in the general view of language and the methods used to teach the language. Twaddell (1972) points out that language consists of two elements - that of choice and that of habit. Choice is what creates the meaning and is under the control of the language user. Habit, on the other hand, is represented by the conventions (phonology and syntax) of the language. The user does not control this aspect of language because it is what orders the meaning for anyone using the language (i.e. the "universal" element) (Twaddell, 1972). The important difference between choice and habit is that we, as native speakers, do not normally pay much attention to the 'habit' part of language (since we are so familiar with its relatively predictable structure). We attend to the meaning of the language because we cannot predetermine the choices. However, the second language learner does not have the familiarity with the language so both choice and habit are novel and become noticeable. Twaddell (1972) points out that anything we notice is meaningful. The second language learner notices everything about the language because it is all novel to him. He's unable to focus on the appropriate meaningful cues because everything he sees and hears is meaningful.

As noted in Chapter 1, Goodman (1967) has described reading as a "psycholinguistic guessing game." However, the second language learner is in no position to deal with the reading material in this way. Carlos Yorio (1971) sums up the differences between the native and second language learner's situation with respect to the psycholinguistic view of reading:

1. The second language reader's knowledge of the foreign language is not like that of the native speaker.
2. The guessing or predicting ability needed to pick up the correct cues is hindered by the imperfect knowledge of the language.
3. The wrong choice of cues of the uncertainty of the choice makes associations more difficult.
4. Due to the unfamiliarity with the material and lack of training, the memory span in a foreign language in the early stages of its acquisition is usually shorter than in our native language.
5. At all levels, and at all times, there is interference of the native language.

Yorio, 1971. p. 108)

Thus, before a beginner reader can embark on a "psycholinguistic guessing game" he must know the 'rules'. That is, as Joyce Morris (1968) explains, beginning reading instruction must focus on helping the learner to break the code

in order to recognize that certain aspects of the language are essentially out of his control. Twaddell (1972) explains that these rules are really explications of the habits: they explain the 'How', not the 'Why'. However, a consequence of this initial focus on the rules of the language is that it often leads second language readers to conclude that these rules really are meaningful in and of themselves. The learner does not recognize that the explicit rule focus is the means of forming habits that are taken for granted (as redundant cues) by the fluent reader.

This explicit focus on the habits of the language has affected how the vocabulary is developed. In focusing on the structures - the most habitual part of the language - ESL programs have viewed the vocabulary aspect of English as a potential hazard. At the introductory and beginner levels of second language learning, the vocabulary is rigorously controlled so that there is no distraction from the structures. It is not until the intermediate stages, when the learner is believed to have experience with many of the frequently used grammatical patterns, that vocabulary expansion is even considered (Twaddell, 1972).

Lack of research in ESL reading vocabularies may be due to the fact that ESL methods are based in the psycholinguistic view of reading which has never seen the word unit as very important in the quest for meaning. Lefevre (1962) has argued that there are a number of good reasons to relegate the word unit to a secondary role:

1. Semantically and structurally, the word is an unstable element.
2. Analyzing and speaking single words in isolation may give the learner a false impression that reading is a fragmentary process.
3. In isolating words, the intonational patterns of words in context are lost.
4. The essence of the sentence as a meaningful unitary pattern made up of syntactic and grammatical forms would be lost if there was a focus on the single word.

While Lefevre does make some valid criticisms of a focus on the single word, he, like most proponents of the psycholinguistic approach to reading, forgets that the situation is rather different for the ESL learner who is learning how to read. Both Yorio (1971) and Twaddell (1980) state that the lexicon is a problem for the non-native learner because he has no previous exposure to the language. This fact is probably responsible for another reason teachers and theoreticians downplay the vocabulary aspect of language learning. Language learners are constantly searching for a one-to-one relationship between their first language and the language they are learning (Twaddell, 1980). This leads the learners to focus on the word as a definitive unit of meaning and results in made dashes for dictionaries every time a new word is encountered. Obviously, such activity is inadequate and far too time-consuming to develop any kind of vocabulary so the language programs have started by using a highly controlled vocabulary which quickly becomes familiar to the students. Vocabulary expansion does not begin until the learner has had some experiences in his new environment so that the words have some experiential meaning associated with them.

This paper focuses on how language experience is developed through written materials presented to the ESL learner. As Richards (1974) pointed out, part of knowing a language is having a sufficient vocabulary for use in a given situation. This tenet implies two things: 1) that the use of vocabulary is dependent upon the particular situation (or context) and, consequently, 2) learners must be exposed to a wide variety of contexts in order to develop a large vocabulary which will lead to competency in the language. By determining the frequency distribution of words in the YES! series and examining both the high and low frequency words used, conclusions will be drawn with respect to the method of selection of words, the control of vocabulary (as a means of achieving an adequate sight vocabulary), and the pedagogical consequences of such selection and control.

CHAPTER 3

The major focus of this study is to examine the written language being presented to ESL students. The text series YES! will be described in terms of the word frequency distribution. As Dolch (1960) pointed out, it is vocabulary control - composed of the introduction of new words and the re-use of "old" words - that maintains and increases sight vocabulary. The primary criterion upon which vocabulary control will be studied is word frequency. By observing the frequency distribution of the words as the series progresses from level A to level F the following hypotheses will be tested:

Hypothesis I: Since the YES! series aims to teach language that is meaningful and that has immediate value, the word frequency distribution will reflect that of the target language in three important ways:

- a) The word frequency distribution will be similar to that predicted by Twaddell (1980), i.e. a "ski slope".
- b) The frequency distribution of the words in the YES! series will be similar to frequency distribution of published lists.
- c) The most frequent words found in the YES! series will be significantly correlated with the most frequent words found in the published word lists.

Hypothesis 2: Since the YES! series aims to teach language for communication, the development of the series will recognize the need for the ESL child to become familiar with a vast number of different words. To enable maximum exposure to the vocabulary of the target language, the YES! series should be developed in the following manner:

- a) As the series progresses from Book A to Book F, there will be an increase in the number of different words. The result of this increase will be increase in the type-token ratio.
- b) There will be little correlation between the words used in the earlier books of the series (A to C) and those used in the later books of the series (D to F) because Books A-C are introductory while Books D-F expand on what was taught in the earlier books.
- c) The number of low frequency words will increase as the series progresses from Book A to Book F because the number of words for each book increases.

Hypothesis 3: Since the ESL children are unlike their native counterparts in that they are not familiar with oral language before they are introduced to the written form of the language, the textbook series ESL children use will be different from the basal reading series used by native speaking children in the following ways:

- a) The first two books of the ESL series YES! will have a higher type-token ratio (fewer repetitions) than specified levels of the Ginn 720 Series (Levels 2, 3, 4, and 5) and the MacMillan series (Levels 4, 5, 6 and 7) because a greater variety of words are introduced to the ESL learner who has a very small oral vocabulary.
- b) The majority of word types found in the first two books of the ESL series YES! will not be found in either of the basal reading series under investigation because the ESL learner does not start with the same stock of oral vocabulary.
- c) The "typical" vocabulary load of the first five levels of basal reading series (Dolch, 1960), as outlined in Chapter 1, will not be found in the YES! series because ESL students need to be exposed to so many more word types as quickly as possible in order to communicate effectively. Instead, many more words will be found in the ESL books than in the first five levels of a typical basal reading series.

These three hypotheses will be tested by looking at three dimensions of the words in the YES! series:

- 1) the most frequently occurring words in the series;
- 2) the least frequently occurring words in the series;
- 3) the level at which the words are introduced.

This chapter will describe the sample, the procedure for the collection of data, and the methods used to analyze the data. Testing these hypothesis will permit the YES! series to be described in terms of vocabulary control and will allow conclusions to be drawn regarding the ability of the series to prepare the child for regular classroom activities. Knowing about the frequency distribution of words will enable the development of teaching strategies and materials design by pointing out the contexts in which language occurs and the effects of the contexts upon the vocabulary selection.

The Sample

The source of the data is Lars Mellgren and Michael Walker's Young English Series: YES! (1977/78). YES! was chosen because it is a standard, widely used text for young ESL learners. The series consists of six books, A - F, with accompanying workbooks for D - F. Each of the Teachers' Guide lists structures and vocabulary for the designated book and for the previous books. New 'structures/structural words' and 'vocabulary expressions' for each book are listed by the page on which they are first presented. The number of times each word or expression is presented is not given.

Presentation of the series should, ideally, progress from Book A to Book F. However, books A to C are considered "entry level" texts and are described by the authors as follows:

1. Book A has no printed matter. This level is probably suitable for students who have had very little or no reading experience in their own language. They may be from six to eight years old, depending on when they started school.
- 2) Book B introduces the printed word. This level reviews and reinforces Book A, and offers a limited amount of new vocabulary. The students may be from seven to nine years old.
3. Book C introduces writing skills. It reviews, reinforces and expands on Books A and B. The students may be from eight to ten years old.

(Mellgren and Walker, Book C:
Teachers' Guide, 1977, p. v)

They stress that each of these three books introduces a new linguistic skill, starting with listening and speaking in Book A and going to reading Book B and writing in Book C. The authors suggest that a student may be entered at any one of these three levels.

Mastery of the contents of the first three levels is considered a necessary prerequisite to progress to the higher three levels in the series. This is because D, E, and F assume competency in the listening, speaking, reading, and writing skills presented in the earlier books. Book D emphasizes reading and writing while Book E stresses grammar and Book F expands on previously learned material in an effort to stimulate production of the language. The workbooks for these three levels reinforced the material learned in the text by emphasizing writing skills.

Procedures for Collection of Data

The data consists of all words in the YES! described above. The definition of the word, for the purpose of this study is:

A word is defined as consisting of any number of letters bounded by space on both the right and the left.

In other words, words with plural endings or other morphological derivations were counted as words and not subsumed under the base word (as many word frequency studies do). Furthermore, words that were part of pictures were also included in the data. The only stipulation for these words was that they had to be clearly visible.

The words of each book in the series were typed, in lower case letters, into separate files in the computer. There was no punctuation used except to define abbreviations (i.e. to distinguish 'us' from 'u.s') and denote the possessive case, "'s". After this was done, two additional files were created; one containing all the words from all the books and workbooks and another containing only the words from the books, levels A - F.

In order to determine the frequency and rank order of the words in each of these files, a program was run to count the words in a given file, give a frequency and rank listing, and print the results. The limitations to the program

were that no word over forty characters would be counted and that the total maximum number of words to be counted would not exceed 30,000 (Miller, 1975).

Analysis of Data

Analysis of the data was accomplished by using computer programs, comparison by percentages, and the Pearson correlation in order to describe and draw conclusions about the development of sight words in the YES! series.

A) Computer Programs

Because the word count program itself was already described, the information given by running the program will be focused upon here. The program gives the total number of words in the sample file (the number of tokens) and the total number of different words in the sample file (types). Two words lists are produced. The first list, an alphabetical ordering of words, gives the number of times a particular word occurs in the sample file and the percent frequency with which it occurs in comparison to all other words. The second list is a rank ordering of the words and again gives the number of times a particular word occurs. In addition to this, the second list also shows the cumulative frequency of the words up to any given ranking. For example, the rank list showed that the first ten words accounted for almost 23 percent of the total number of words (tokens) in the file containing the words from the six books in the series

(designated the Total Word Count). The program also gives a type-token ratio which indicates the repetitiveness of the words in the books (A-F) and the series as a whole. The larger the type-token ratio, the less repetition of words.

The second computer program that was used was one which computed the Pearson correlation coefficient. This correlation was used to compare the frequencies of the 321 most frequent words of the YES! series as a whole (i.e. the six books added together to make up the Total Word Count) with the frequencies of the same words as counted in each of the levels (A - F) of the series. The correlation was also used to compare the 321 most frequent words of the Total Word Count with published word lists which also give frequency counts (Durr's "188 Words of More than 88 Frequencies", 1973; Carroll et al. The Word Book, 1971; and Rinsland's Basic Vocabulary of Elementary School Children, 1945).

B) Comparison by percentages

Comparison could not be done between the word frequency data collected from the YES! series and word lists that gave no frequency information. However, many such lists are of interest because they propose to describe the word distribution of one or more of the language skills of speech, writing, or reading. It was decided that comparing a cross-section of such lists in order to determine how the YES! series controls vocabulary in these language skill areas would aid in describing the words of the series. A brief description of each of the lists used for comparison follows.

Hillerich's "240 Starter Words" (1974) which constitutes a basic language arts vocabulary for reading and writing. The list is based on five previously published lists (Carroll et al., Hillerich, Kucera-Francis, Rinsland, and Thorndike). This list was selected because of its application to the language arts.

Durr's "188 Words of more than 88 Frequencies" (1973). This list allowed for a comparison to be made between the ESL data and words of high frequency in popular library books children select. Proper names were omitted and only base words of common inflected and compound words were counted. Durr points out that the top ten words on his list account for twenty-five percent of the total running words in his sample. He suggests that instant recognition of these few words would greatly facilitate reading. The list of 188 words he gives accounts for only six percent of the word types, but almost seventy-two percent of the word tokens.

Rinsland's Basic Vocabulary of Elementary School Children (1945) offers a quantitative analysis of words written by children in grades 1 through 8. All words are counted just as they occur (i.e. roots, inflectional forms, abbreviations, and contractions are counted separately). This list has been chosen for comparison in an attempt to determine whether the YES! series is offering a vocabulary that aids in written production of the language.

Dale's list of '769 Easy Words' resulting from "A Comparison of Two Word Lists" (1931). This list was developed by determining the commonality between the International Kindergarten List and the first one thousand words of the Thorndike list. The importance of this list lies on the fact that a number of readability formulas used it in their calculations.

Johnson's A Basic Sight Vocabulary for Beginning Reading (1971). This lists gives the oral and printed frequencies for 306 words which combine the Kucera-Francis list based on printed English (1967) and Murphy's list of oral words (1957). This list was used because frequencies are given for each word in both the oral and printed genre.

Comparisons by percentages will be done to show the relationships between the Total Word Count for the series and other published lists (as described above) of high frequency words. Percentages are also used to describe the development of types and tokens throughout the series itself. In this way, the percentage of new words presented in each book, the relative importance of a given word from level to level, and the frequency distribution of the most common words from level to level may be examined.

More specifically, the three hypothesis stated at the beginning of this chapter will be tested in the following ways:

- 1) Hypothesis 1, which deals with the similarity between the frequency distribution of the words in YES! as compared to 'standard' English, will be tested by first expanding upon the information given by the computer program with respect to the number of types and tokens. This information will be used to graphically compare the frequency distribution of the words in the YES! series to that of the 'standard' language and other published word lists. The hypothesis will be further tested by correlating the most frequent words found in the YES! series with high frequency words found in published word lists.
- 2) Hypothesis 2 is concerned with the development of the series from Book A to Book F. This hypothesis will be tested by using the information obtained from the Word Count program and by using the Pearson Correlation.
- 3) Hypothesis 3 aims to show the differences in word distribution between the YES! series and two basal reading series (the Ginn 720 and the MacMillan series). To do this, comparable levels of readability had to be determined for the YES! series. This was done by using the Spache Readability Formula (Covell, 19). Then comparisons were made on the basis of the type-token ratios, the word types occurring in the basal series versus the ESL series, and the number of "new" words introduced in each book.

The purpose of testing the three hypothesis is not merely to conclude that the vocabulary control exhibited by the YES! series is different from that of basal reading series. Intuitively, we already know this to be so. The ultimate goals of this study are to determine if ESL series such as YES! are supplying vocabulary that will be necessary in a regular classroom situation and what teacher and designers of materials can do to help familiarize the ESL student with the vast number of words necessary to communicate effectively. By looking at word frequency, we will be able to determine which words occur most often in the series, how they are developed throughout the series, and how they compare with published word frequency lists which reflect the types of words the child is most likely to frequently encounter in his experiences in the school and community environment. Given this information, teachers and authors can design programs to accentuate and supplement was already exists.

All words were entered into the computer in lower case lettering since any change in visual cues led to the word being counted as a different word even if the change was merely capitalization to begin a sentence (i.e. 'the' and 'The' would be counted as different words). This characteristic of the computer is particularly problematic when a word has more than one meaning. A word may have an inflated word frequency merely because it has two or more meanings and is subsumed under one orthographic type. Furthermore, a proper noun may be subsumed under a word type that has a totally different meaning simply because it becomes orthographically identical to that other word. For example, if a story

happens to deal with the 'Green' family every instance of reference to one of its members by their last name will be counted as an instance of 'green' along with the color 'green'.

Thus, semantics cannot be dealt with using this method of word collection and organization. What is being examined is the words, in their visual form, that ESL children must recognize and ultimately understand to word with the YES! series successfully. The series claims to develop reading, writing, and oral skills. By comparing the frequency of the words used in the series to those used by the general speaking population (as defined by published word lists) conclusions will be drawn as to how well these goals are being met with respect to the vocabulary control exhibited.

CHAPTER 4

The hypotheses stated in this paper are concerned with how well the YES! series reflect the language it is proposing to teach. Word frequency analysis is used to make this evaluation because it is an objective method and it allows comparison with studies done on the target language, as well as permitting descriptive analysis. The first hypothesis, directly compares the number of types and tokens found in YES! with Twaddell's (1980) prediction concerning the shape of the frequency distribution curve. This hypothesis also predicts a similarity between the frequency distribution of the YES! series and that of published word lists that have been used to guide and predict the content of materials for reading, writing, and speaking.

As was pointed out in Chapter 3, the computer program WORDCOUNT gave alphabetical listing of the words, a rank order listing of the words (with their respective frequencies and cumulative frequencies), and statistical information. Table I gives a summary of the types and tokens found in the series as a whole.

TABLE I

A Summary of All Types and Tokens Found in the YES! Series

Total number of words	(Tokens) = 33,549
Total number of sorted words	(Types) = 2,799

Thus, Table 1 shows that YES! series includes a running total of 33, 549 words of which 2,799 are different. However, for the purpose of testing Hypothesis 1 more detailed information about the relationship between the types and tokens is needed. Therefore, the percentage of types and tokens accounted for by the first 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, ... 1000 most frequent word types was calculated for the series as a whole. Table 2 shows the results of this tabulation.

The 2 important features of this table are that:

1. There are a few high frequency words that account for most of the running words in the series, i.e.
 - a) the 10 most frequent words account for almost 23% of the running words in the series;
 - b) the 70 most frequent words account for almost 50% of the running words in the series;
 - c) the 400 most frequent words account for nearly 75% of the running words in the series; and,
 - d) the 1000 most frequent words account for over 89% of the running words in the series.
2. The few high frequency words that account for most of the running words actually make up an extremely small portion of the number of different words accounted for, i.e.

TABLE 2

% of Tokens and Types Accounted for by 1st
1, 10, 20, 40, 50, 60, 70, 80, 90, 100, 150, 200
250, 300, 350, 400, ... 1,000 Words in the
Count (i.e. Books A - F)

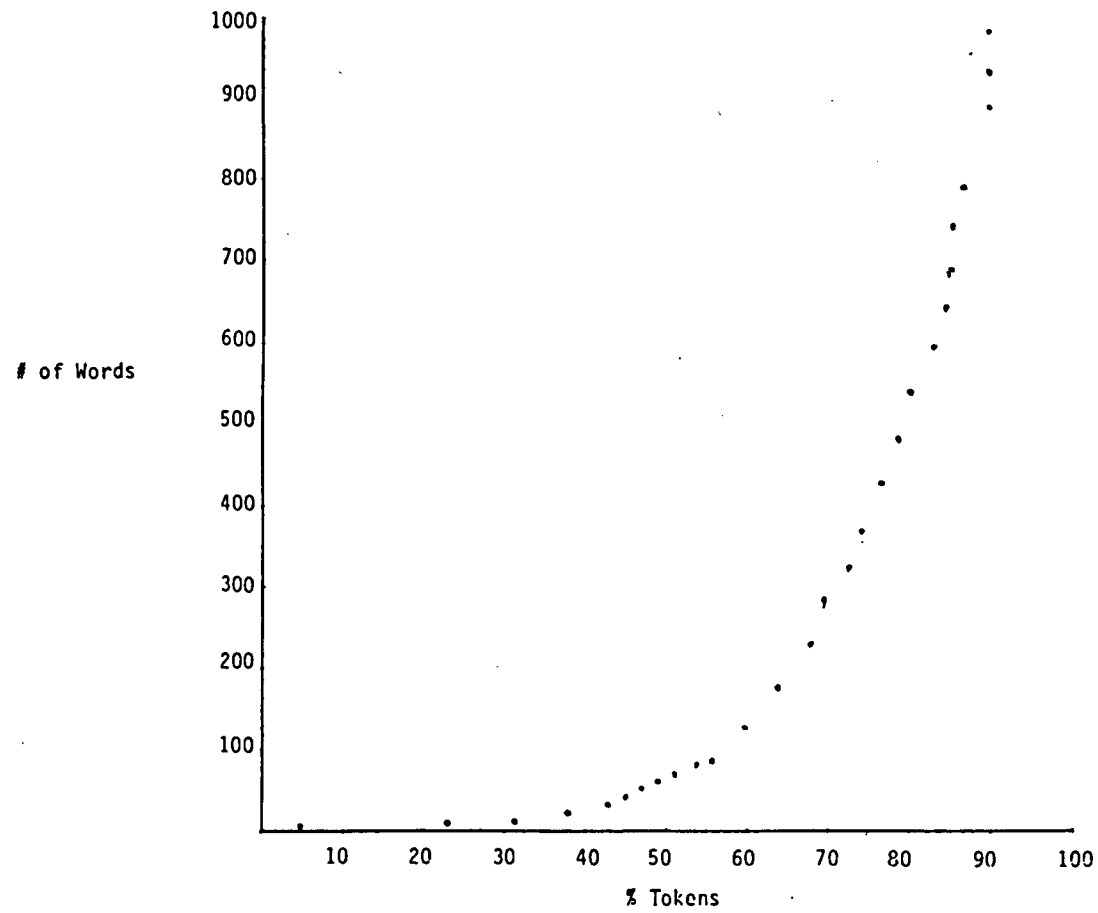
<u># of Words</u>	<u>*% of tokens (running words) accounted for</u>	<u>% types (diff. words) accounted for</u>
1	5.6514	.0357
10	22.9306	.3573
20	32.4540	.7145
30	38.5168	1.0718
40	42.5646	1.4291
50	45.5811	1.7863
60	47.8584	2.1436
70	49.5872	2.5009
80	51.4755	2.8582
90	53.0791	3.2154
100	54.7736	3.5727
150	60.1328	5.3591
200	64.4833	7.1454
250	68.2733	8.9318
300	71.3285	10.7181
350	73.5357	12.5045
400	74.8905	14.2908
450	77.5403	16.0772
500	79.1991	17.8635
550	80.8996	19.6499
600	82.4153	21.4362
650	84.0442	23.2226
700	84.0442	25.0090
750	85.3215	26.7953
850	89.3857	30.3680
900	89.3857	32.1543
950	89.3857	33.9407
1,000	89.3857	33.7270

- a) The 10 most frequent words that account for almost 23% of the running words do not even make up 1% of the different words that will be encountered in the series;
- b) the 70 most frequent words that accounted for almost 50% of the running words, make up only 2.5% of the types found in the series
- c) the 400 most frequent words that accounted for nearly 75% of the running words, make up only 14% of the word types that will be encountered; and,
- d) the 1000 most frequent words (accounting for 89% of the running words) makes up less than 36% of the different words that will be encountered.

The two features point out the nonlinear relationship between the word types and the word tokens found in the series. Figure 1 depicts the relationship that exists between the types and tokens. Graphically, such a relationship looks like the "ski slope" predicted by Twaddell (1980).

Thus, part (a) of Hypothesis 1 has been proven correct. The word frequency distribution is quantitatively similar to Twaddell's prediction. Table 2 and Figure 1 combine to show that there are very few high-frequency words, a slightly larger group of medium frequency words, and an infinite number of low frequency words.

Figure 1 - Percentage of Tokens Accounted for by 35.7% of Types (1000 Different Words)



Part (b) of Hypothesis I further supports the findings of part (a). Part (b) states that the frequency distribution of the words in the YES! series will be similar to frequency distributions of published word lists. Since published word lists reflect the target language by collecting printed, spoken, and/or written samples of the language, the frequency distribution displayed by their lists will reflect the target language. Thus, comparisons were made between the word frequency distribution of the YES! series and those of published word lists produced by Johnson (1971), Durr (1973) and Walker (1979).

Table 3 shows the relationships existing between the types and tokens for each of the three published word lists and the list compiled for the YES! series.

Table 3 shows that there is a high degree of consistency between the percentage of tokens accounted for by the first through the two hundredth word for the four lists. After the two hundredth word, the YES! series tends to have slightly higher percentages. This is probably due to the fact that the sample for the YES! list (i.e. the number of tokens) is much smaller than for the other three lists. This fact is also reflected in the much larger percentage of types accounted for by any given number of words. (In this respect, Durr's list is quite similar to the YES! list.)

The similarity between published lists and the YES! series is further revealed in graphing the information of Table 3. Figure 2 compares the percentage of tokens accounted for by the published lists and the YES! series.

Figure 2 - Percentage of tokens accounted for by the 500 most frequent words in the YES! series and the Published Lists of Walker, Johnson, and Durr.

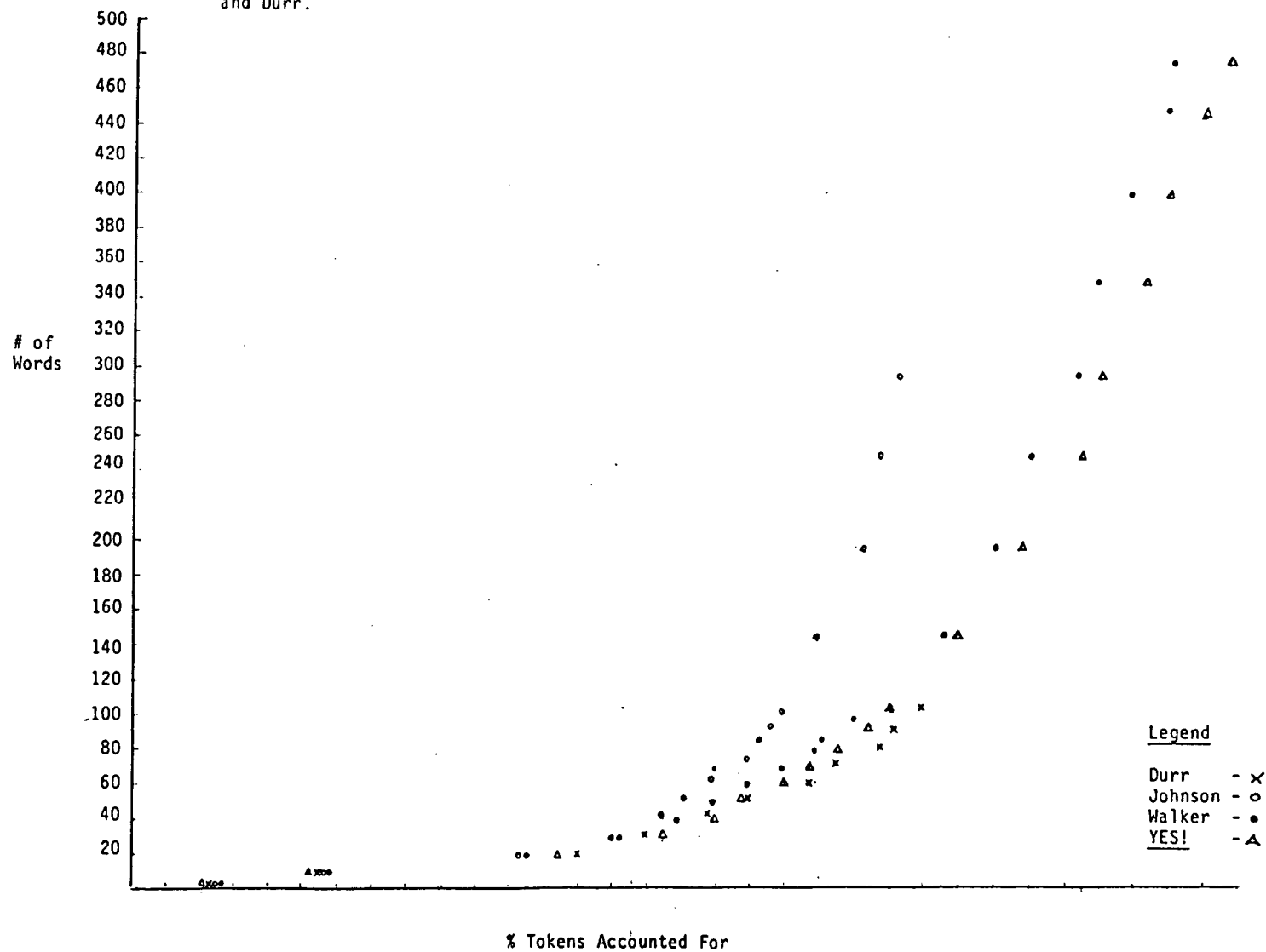


TABLE 3
Type Relationship for Token
for 3 Published Word Lists and the YES! Series List

Number of Words	Johnson's (Printed)		Durr		Walker		YES! Series	
	Tokens = 1,014,232 Type = 50,406		Tokens = 105,280 Types = 3,220		Tokens = 5,088,721 Types = 86,741		Tokens = 23,549 Types = 2,799	
	% Tokens Accounted For	% Types Accounted For	% Tokens Accounted For	% Types Accounted For	% Tokens Accounted For	% Types Accounted For	% Tokens Accounted For	% Types Accounted For
1	6.8989	.0020	5.9622	.0311	7.3507	.0012	5.6514	.0357
10	24.2562	.0200	23.7557	.3106	24.5397	.0115	22.9306	.3573
20	31.0353	.0397	32.5247	.6211	31.5341	.0230	32.4540	.7145
30	35.3276	.0595	37.8390	.9317	35.9627	.0346	38.5168	1.0718
40	38.3278	.0794	42.2122	1.2422	39.5186	.0461	42.5646	1.4291
50	40.6573	.0992	45.9273	1.5528	42.5376	.0576	45.8111	1.7863
60	42.5194	.1190	48.9723	1.8630	45.0823	.0692	47.8584	2.1436
70	44.1024	.1389	51.7620	2.1739	47.3567	.0807	49.5872	2.5009
80	45.3904	.1587	54.1799	2.4845	49.3452	.0922	51.4755	2.8582
90	46.4532	.1786	56.1598	2.7950	51.0077	.1038	53.0791	3.2154
100	47.3589	.1984	57.8894	3.1056	52.5268	.1153	54.7736	3.5727
150	50.8420	.2976	64.7426	4.6584	58.5238	.1729	60.1329	5.3591
200	53.0180	.3968			62.7030	.2306	64.4833	7.1454
250	54.5503	.4560			68.3959	.3459	71.3285	10.7181
300	55.6361	.5952			68.3959	.3459	71.3285	10.7181
350					70.5410	.4035	73.5357	12.5045
400					72.3245	.4611	74.8905	14.2908
450					73.8752	.5188	77.5403	16.0772
500					75.2512	.5764	79.1991	17.8635
550							80.8996	19.6499
600							82.4153	21.4362
650							84.0442	23.2226
700							84.0442	25.0090
750							85.3215	26.7953
800							87.4810	29.5816
850							89.3857	30.3680
900							89.3957	32.1543
950							89.3857	33.9407
1000							89.3857	35.7270

Part (c) of Hypothesis I correlates the most frequent words found in the YES! series with high frequency words found in published word lists. Published word lists are compiled in an effort to describe which words of the language will be encountered most often. It is argued that knowing the most frequent words is important to learning language in general (since the most frequent words are the structural basis for coherence), to decoding new words (Walker, 1979 suggests that high frequency words set up the context by which new words can be deciphered), and to fluent reading (Durr, 1973). Thus, if the YES! series is highly correlated with published word lists which reflect the frequency distribution of the words that will be encountered in learning the language, the series will be doing its job of exemplifying standard language.

In order to make correlations between the YES! series and published word lists a decision had to be reached on what could be considered a high frequency word with the series. Two important factors had to be considered with respect to forming a high frequency word list. First, such a list cannot be too large because it would be of no practical value to teachers, students, or material writers. Second, the list should not be too small. A list that is too small merely reflects the necessary very high frequency words that are found on all lists because they are the structure words (i.e. prepositions, conjunctions, pronouns). Thus, in order to determine if the YES! series was unique in the frequent use of certain words, enough words had to be selected in order to reveal the characteristics of the series.

It was decided that high frequency words would be defined as the words that made up 70% of the running words in the series. This meant that if the child knew the words that made up 70% of the tokens, he would, on average, know seven out of every ten words. Walker (1979) felt that this was an acceptable level for elementary school children. However, because there are many words with the same frequency in the series, the high frequency words totalled 321 and accounted for nearly 72% (71.8859%) of the running words. The 321 most frequent words of the YES! series is shown in Table 4.

The list, along with its accompanying frequencies, formed the basis for comparison with the published word lists of Walker (1979), Durr (1973), Johnson (1945). For each word on the YES! list, a frequency of occurrence was obtained from each of the published word lists. Then all of the data was put into the computer which statistically determined the correlation between each of the lists. Table 5 shows the results of the relationships in the form of a Pearson Correlation Coefficients table.

Table 5 shows that the YES! list correlates highly with all the published word lists. Walker's list has the highest correlation with the YES! list (.8574) while Johnson's Oral vocabulary rates as the lowest correlation (.7007). This means that the ranking of the words of the YES! series that are common to the other lists is more similar to the Walker list than the Johnson Oral Vocabulary list with the other three lists falling somewhere between these two extremes.

TABLE 4

A Rank Order List of the 321 Most
Frequent Words in the YES! Series

<u>Frequency</u>			<u>Frequency</u>			<u>Frequency</u>		
1.	the	1897	23.	do	229	45.	yes	105
2.	is	1061	24.	at	221	46.	as	104
3.	a	916	25.	how	216	47.	car	96
4.	he	655	26.	your	195	48.	old	90
5.	you	618	27.	where	177	49.	going	86
6.	to	554	28.	no	176	50.	name	82
7.	in	544	29.	does	172	51.	like	81
8.	and	509	30.	there	165	52.	not	81
9.	what	486	31.	for	163	53.	me	80
10.	she	454	32.	her	152	54.	Oh	80
11.	I	431	33.	when	139	55.	an	79
12.	are	369	34.	but	136	56.	why	75
13.	it	365	35.	that	134	57.	doing	73
14.	did	353	36.	it's	131	58.	mary	73
15.	was	324	37.	one	128	59.	up	72
16.	they	304	38.	were	127	60.	tom	70
17.	my	283	39.	go	126	61.	said	69
18.	can	257	40.	have	122	62.	see	68
19.	of	256	41.	with	118	63.	or	66
20.	this	253	42.	two	114	64.	am	64
21.	his	250	43.	what's	110	65.	school	64
22.	on	233	44.	we	107	66.	from	63

Table 4 Cont'd

<u>Frequency</u>		<u>Frequency</u>		<u>Frequency</u>	
67.	him	63	91.	about	49
68.	all	62	92.	good	49
69.	had	62	93.	home	48
70.	now	62	94.	play	48
71.	bus	61	95.	brother	47
72.	can't	61	96.	friend	47
73.	I'm	61	97.	only	47
74.	book	60	98.	ten	47
75.	many	60	99.	dog	46
76.	out	60	100.	house	46
77.	very	60	101.	into	46
78.	who	60	102.	make	46
79.	day	59	103.	mother	46
80.	don't	59	104.	then	46
81.	big	57	105.	man	45
82.	color	55	106.	nine	45
83.	get	55	107.	too	45
84.	help	55	108.	wearing	45
85.	fly	53	109.	father	44
86.	world	53	110.	little	44
87.	she's	52	111.	these	44
88.	their	52	112.	three	44
89.	right	51	113.	sister	43
90.	than	51	114.	take	43
			115.	five	42
			116.	Mr.	42
			117.	know	41
			118.	number	41
			119.	long	40
			120.	much	40
			121.	by	39
			122.	didn't	39
			123.	eight	39
			124.	say	39
			125.	white	39
			126.	he's	38
			127.	yourself	38
			128.	first	37
			129.	cat	36
			130.	dan	36
			131.	monday	36
			132.	new	36
			133.	our	36
			134.	people	36
			135.	work	36
			136.	wrong	36
			137.	years	36
			138.	yesterday	36

Table 4 Cont'd

<u>Frequency</u>			<u>Frequency</u>			<u>Frequency</u>		
139.	green	35	163.	please	31	187.	brown	26
140.	looking	35	164.	seven	31	188.	come	26
141.	red	35	165.	because	30	189.	find	26
142.	time	35	166.	last	30	190.	here	26
143.	bird	34	167.	morning	30	191.	maria	26
144.	blue	34	168.	next	30	192.	start	26
145.	four	34	169.	over	30	193.	stop	26
146.	has	34	170.	reading	30	194.	could	25
147.	hat	34	171.	understand	30	195.	other	25
148.	lion	34	172.	write	30	196.	read	25
149.	look	34	173.	yellow	30	197.	river	25
150.	them	34	174.	children	29	198.	store	25
151.	tree	34	175.	every	29	199.	street	25
152.	twenty	34	176.	eat	28	200.	word	25
153.	buy	33	177.	fill	28	201.	zoo	25
154.	eating	33	178.	live	28	202.	answer	24
155.	train	33	179.	walk	28	203.	ask	24
156.	want	33	180.	ball	27	204.	best	24
157.	six	32	181.	black	27	205.	boy	24
158.	some	32	182.	bob	27	206.	english	24
159.	tall	32	183.	page	37	207.	james	24
160.	way	32	184.	peter	27	208.	most	24
161.	doesn't	31	185.	sally	27	209.	saw	24
162.	listening	31	186.	be	26	210.	swimming	24

Table 4 Cont'd

<u>Frequency</u>			<u>Frequency</u>			<u>Frequency</u>		
211.	test	24	235.	put	22	259.	bag	19
212.	under	24	236.	sweater	22	260.	eleven	19
213.	well	24	237.	tennis	22	261.	garage	19
214.	yard	24	238.	took	22	262.	gas	19
215.	again	23	239.	carrying	21	263.	isn't	19
216.	elephant	23	240.	door	21	264.	its	19
217.	giant	23	241.	girl	21	265.	just	19
218.	later	23	242.	hair	21	266.	lunch	19
219.	minutes	23	243.	listen	21	267.	meany	19
220.	more	23	244.	living	21	268.	mexico	19
221.	off	23	245.	open	21	269.	miss	19
222.	snow	23	246.	pen	21	270.	mrs.	19
223.	so	23	247.	plane	21	271.	paper	19
224.	they're	23	248.	show	21	272.	piece	19
225.	thirty	23	249.	washing	21	273.	police	19
226.	use	23	250.	came	20	274.	says	19
227.	whole	23	251.	captain	20	275.	susan	19
228.	after	22	252.	down	20	276.	thank	19
229.	bed	22	253.	sir	20	277.	waiter	19
230.	carry	22	254.	small	20	278.	water	19
231.	forty	22	255.	soup	20	279.	window	19
232.	garden	22	256.	tv	20	280.	bank	18
233.	jack	22	257.	twelve	20	281.	chair	18
234.	o'clock	22	258.	week	20	282.	circus	18

Table 4 Cont'd

<u>Frequency</u>		<u>Frequency</u>			
283.	england	18	307.	horse	17
284.	friday	18	308.	left	17
285.	happy	18	309.	miles	17
286.	hello	18	310.	milk	17
287.	if	18	311.	practice	17
288.	morris	18	312.	ride	17
289.	must	18	313.	short	17
290.	park	18	314.	sitting	17
291.	pat	18	315.	swim	17
292.	purple	18	316.	trash	17
293.	sam	18	317.	try	17
294.	shoes	18	318.	us	17
295.	table	18	319.	wanted	17
296.	that's	18	320.	wednesday	17
297.	top	18	321.	which	17
298.	watch	18			
299.	animals	17			
300.	any	17			
301.	baseball	17			
302.	bike	17			
303.	bill	17			
304.	friends	17			
305.	full	17			
306.	full	17			

TABLE 5

Pearson Correlation Coefficients for Published
Word Lists of Walker, Durr, Johnson (Printed and Oral),
Rinsland and 321 Most Frequent Words of the YES! Series

	YES	WALK	DURR	JP	JO	RINS
YES		0.8574 (232) p=0.000	0.8480 (147) p=0.000	0.8021 (175) p=0.000	0.7007 (175) p=0.000	0.8234 (303) p=0.000
WALK			0.9247 (145) p=0.000	0.9867 (163) p=0.000	0.6143 (163) p=0.000	0.8545 (232) p=0.000
DURR				0.6954 (137) p=0.000	0.8874 (137) p=0.000	(147) p=0.000
JP					0.5698 (175) p=0.000	0.8228 (175) p=0.000
JO						0.8666 (175) p=0.000
RINS						

Note also that the degree of correlation has nothing to do with how many words are common to the two lists being correlated. The correlation is concerned with the ranking of the words that are common. Thus, while the Rinsland list has 303 words that are common to the YES! list, it also has a lower correlation than the Walker list which has only 232 words that are common to the YES! list.

A more revealing statistic that can be derived from the correlations is obtained by squaring the correlation (i.e. r^2). This statistic gives a percentage which reflects the likelihood of a word in a given list having the same ranking as that word in the YES! list. Table 6 shows the percentages for the likelihood of finding this match.

TABLE 6
Likelihood of Words in Published
Word Lists Having the Same Ranking
as Words in the YES! List

	<u>Walker</u>	<u>Durr</u>	<u>Johnson- Printed</u>	<u>Johnson- Oral</u>	<u>Rinsland</u>
<u>YES!</u>	73.5%	71.9%	64.3%	49.1%	67.8%

Table 6 tends to more clearly reveal which lists are most like YES! in their ranking of common words. The Walker and Durr lists are better suited for reflecting the words in the YES! series than are the other lists.

This table, along with Table 5, also shows that the oral vocabulary ranking presented in Johnson's list is quite unlike most of the other lists which reflect printed and written materials. The Johnson Oral and Printed Vocabularies are made up of the same words. However, the ranking of these words is quite different. Table 5 shows that the 175 words that both lists have in common with the YES! list correlate quite differently. The Johnson-printed has a correlation of 0.8021 while the Johnson-Oral is .7007. This difference is further demonstrated in Table 6 where the likelihood of the words having the same ranking as YES is 64.3% for the printed list and 49.1% for the oral list. Thus, there is a difference between the ranking of words used orally and those used in print and writing. Given the same words (as Johnson does), the YES list will be reflected in the printed ranking far better than in the oral ranking.

It has been shown in testing Hypothesis 1 that a) the word frequency distribution of the words in the YES! series show the typical "ski slope" shaped described by Twaddell (1980); b) the frequency distribution of the words in the YES! series is similar to that of a sample published word lists; and c) the correlation between the words of published word lists and the YES! series is high for lists which reflect printed and/or written vocabulary.

Hypothesis 2 is concerned with the vast number of different words the ESL child will need to become familiar with in order to be able to communicate effectively in the target language. The testing of this hypothesis involves looking at how the series develops from Book A to Book F.

To test the part (a) of this hypothesis, the number of running words (tokens), the number of different words (types), and the ratio between the types and tokens were calculated for each book and for the series as a whole (the Total Word Count). Table 7 summarizes this information:

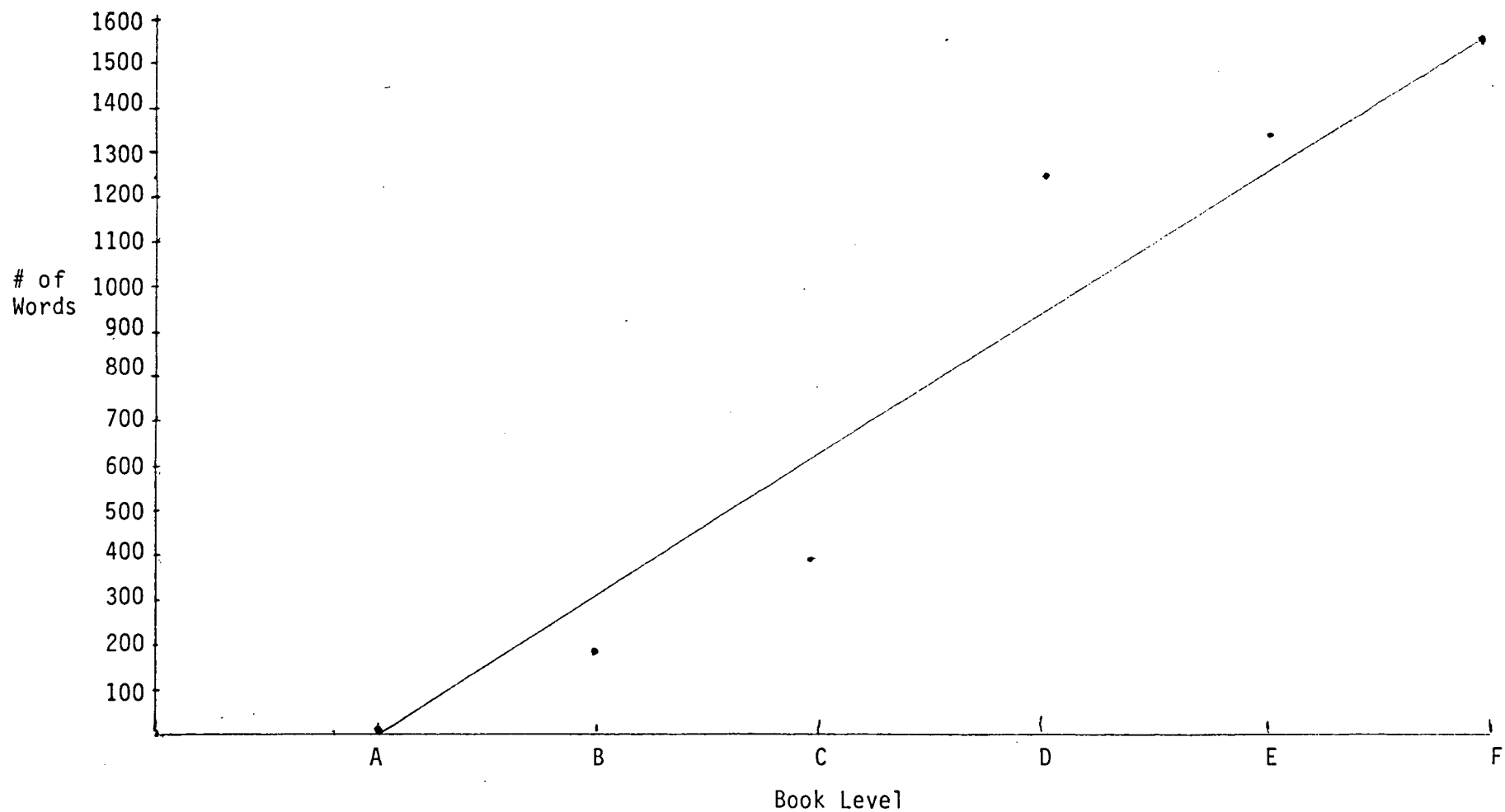
TABLE 7
A Summary of the Word Distributions
for the Six Books in the YES! Series
and for the Series as a Whole (Total Word Count)

	<u>Book A</u>	<u>Book B</u>	<u>Book C</u>	<u>Book D</u>	<u>Book E</u>	<u>Book F</u>	<u>Total Word Count</u>
Total Number of Words (Tokens)	29	2,133	4,437	8,846	8,858	9,546	33,549
Total Number of Different Words (Types)	10	214	518	1,262	1,481	1,675	2,799
Type-Token Ratio	.3448	.1003	.1167	.1427	.1731	.1755	.0834

Table 7 shows that there is a steady increase, from Book A to Book F, in the number of different words as a linear one. This information is reflected in Figure 3.

The table also shows that, except for Book A, there is an increase in the type-token ratio as the series progresses from Book A to Book F. A high type-token ratio means a lower rate of repetition. Thus, Book B, an early book in the

Figure 3 - A Graph of the Number of Different Words in Each of the 6 Books of the YES! Series.



The coefficients given in Table 8 clearly point out the lack of relationship between Book A and all the other books in the series. There is a negative correlation for each of the books when correlated with Book A. The information in the table also points out how alike Books D, E and F are in their rankings of words that are common to each book. The correlation between these books never drops below .92 and therefore means that there is a better than 84% chance that the ranking of a word in one of these three books will be the same in either of the other two books. However, between any one of these three books (D, E or F) and any of the first three books, there is less than a 55% chance of common words matching in ranking.

Overall, the table shows that there is a developmental relationship between the books of the series if Book A is not considered. Books A and C are highly correlated with one another and Book C has a higher correlation with Book B than any other book. Books D, E, and F show an even clearer developmental relationship as each book is more clearly correlated to the Book it preceedes than any other book in the series.

Another developmental aspect that can be derived from looking at the 321 most frequent words in the series is that as the series progresses, more and more of these 321 words will be included. Table 9 shows how many of these words are included in each book.

TABLE 8

Pearson Correlation Coefficients for each
of the Books in the Series

	A	B	C	D	E	F
A		-0.0250 (231) p=0.328	-0.0129 (321) p=0.409	-0.0084 (321) p=0.441	-0.0385 (321) p=0.246	-0.0369 (321) p=0.255
B			0.8048 (321) p=0.000	0.4850 (321) p=0.000	0.4210 (321) p=0.000	0.3748 (321) p=0.000
C				0.7262 (321) p=0.000	0.6671 (321) p=0.000	0.5899 (321) p=0.000
D					0.9333 (321) p=0.000	0.9262 (321) p=0.000
E						0.9421 (321) p=0.000
F						

TABLE 9
The Number of High Frequency
Words* in Each Book

<u>Book</u>	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	<u>F</u>
Number of words	5	122	234	298	302	300
% of the 321 words included	1.6%	38.0%	72.9%	92.8%	94.1%	93.4%

*The 321 most frequent words in the YES! series.

Once again we see that the series develops sequentially and that Book D, E, and F are highly similar and Book C trends to be a transition point.

Thus, the number of types in each series increases as the series progresses from Book A to Book F and the type-token ratio increases during this progression. Correlation show that the 321 most frequent words of the series correlate more and more highly as the series progresses. However, it should also be true that the least frequent words follow a similar developmental pattern as the most frequent words. That is, as the series progresses, there should be more and more low frequency words. To test this, Table 10 was developed to summarize the recurrence of the least frequent words in the series and their contribution to the Total Word Count.

TABLE 10

**Recurrence of the Least Frequent Words* and
Their Contribution to the Total Word Count**

<u># of Types</u>	<u>Total % Types</u>	<u>Recurrence</u>	<u>Total % Tokens</u>	<u>Cumulative Totals</u>	
				<u>% Total Type</u>	<u>% Total Tokens</u>
911	32.5%	1	2.7%	32.5%	2.7%
416	14.9%	2	2.5%	47.4%	5.2%
277	9.9%	3	2.5%	57.3%	7.7%
165	5.9%	4	2.0%	63.2%	9.7%
131	4.7%	5	2.0%	67.9%	11.7%
104	3.7%	6	1.9%	71.6%	13.6%

* The words occurring 6 times or less in the series.

The most significant information that this table reveals is:

- a) Words occurring only once in the series (of which there are 911) account for over 32% of the word types found in the series, and
- b) words occurring 6 times or less in the series account for 71.6% of the types and 13.6% of the tokens found in the series.

Recall that the 321 most frequent words account for only 11.5% of the types and almost 72% of the tokens. The fact that the distribution of the least frequent words appears to be a mirror image of the most frequent words makes these low frequency words worthy of closer inspection. However, as Table 10 indicates, there is an extremely large number of these low frequency words. Thus, it was decided that rather than look at the low frequency words on the basis of the whole series, the low frequency words in each book, A through F, would be examined in an effort to determine whether the low frequency words

followed the same pattern as the high frequency words (i.e. an increase as the series progresses from Book A to Book F). Table II shows the distribution of the least frequent words in each of the books.

Table II clearly shows that as the series progresses from Book A to Book F the number of low frequency words increases. This table also shows that the percentage of low frequency word types increases as the series progresses from Book A to Book F. Furthermore, the greatest proportion of the low frequency words are accounted for by words that occur only once. The implications of these findings will be discussed in Chapter 5.

In testing Hypothesis 2 it has been found that the YES! series is structured to expose the ESL child to the vast number of words in the English language. As the series develops, more and more word types are introduced and less and less repetition occurs. Once the child has a basic foundation from which to work (Books B and C) the series develops more consistently (i.e. Books D, E, and F) so as to fully reflect the series as a whole and the target language. The vast number of low frequency words which occur in the series (and which increase as the series progresses from Book A to Book F) again show the concern with exposing ESL children to many words.

Both Hypothesis 1 and 2 point out the special needs of the ESL child that must be fulfilled by the ESL series YES! These children come to the classroom with little or no English and therefore need to be supplied with language that will

TABLE II

The Distribution of the Least Frequent Words in Each Book of the YES! Series

<u>Book</u>	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	<u>F</u>
Number of words occurring once	3	67	153	405	598	692
% of types accounted for	33.3%	31.3%	29.5%	32.1%	40.4%	41.3%
Number of words occurring two times or less	4	97	254	631	879	1012
% of types accounted for	40.0%	45.3%	49.0%	50.0%	59.4%	60.4%
Number of words occurring three times or less	4	11	301	778	1017	1198
% of types accounted for	40.0%	51.9%	58.1%	61.6%	68.7%	71.5%
Number of words occurring four times or less	10	124	329	879	1106	1307
% of types	100%	57.9%	63.5%	69.7%	74.7%	78.0%
Number of words occurring five times or less		137	348	951	1176	1371
% of types accounted for		64.0%	67.2%	75.3%	79.4%	81.9%
Number of words occurring six times or less		141	366	999	1234	1412
% of types accounted for/book		65.9%	70.7%	79.2%	83.3%	84.3%

enable them to communicate effectively early in their language experience. Because the ESL children lack the oral language that forms the basis for basal reading series used by native speakers, special series must be developed to enhance opportunity for both oral and written language practice. Hypothesis 3 is concerned with this very special role that an ESL series (particularly the early books in the series) must fulfill.

To test Hypothesis 3 comparable data had to be obtained for the ESL and basal reading series. The two basal reading series that were chosen were the Ginn 720 series and the MacMillan series. Because the concern was with the early books in the ESL series, only lower levels of the basal reading series were chosen to be used in the comparison. The choice of the specific levels was determined on the basis of age equivalency because no readability formula could be found to estimate pre-primers and primers. Thus, Levels 2, 3, 4, and 5 were chosen from the Ginn 720 series. (Levels 2 and 3 make up the pre-primers, Level 4 is the primer, and Level 5 is the first reader). Levels 4, 5, 6, and 7 were selected from the MacMillan series where Levels 4, 5, and 6 make up the pre-primers and Level 7 makes up the primer. From the YES! series, Books A and B (which are suggested for six to nine year olds) were estimated to be of pre-primer and primer status since neither book met maximum readability on the Spache Table for calculation of grade level readability (Spache, 1975). Table 12 has been developed to compare the distribution of words within the three series.

The table shows information for individual books in the series and for the books tallied together in each of the three series. The information given in Table 12 can be summarized as follows:

- a) There are many more tokens in the basal reading series than in the YES! series (i.e. the Ginn 720 series has over five times as many words token and the MacMillan series has more than double the number of word tokens.
- b) While the YES! series has the least number of word types of the three series, it has only slightly less than half of the number of word types as the MacMillan series which has the most word types of the three series.
- c) The YES! seires has a higher type-token ratio than either of the other two series. This indicates that the two basal reading series are more repetitive than the YES! series.

Thus, part (a) of hypothesis 3 has been shown to be true. The YES! series does have fewer repetitions of words in te early books of the series than the two basal series with which it was compared.

Since there are fewer repetitions of word types in the ESL series, the word types that do exist should be quite different from the word types found in the basal reading series. Part (b) of Hypothesis 3 states that,

TABLE 12

A Comparison of Tokens, Types, and Type-Token Ratios
for the YES! Series, the Ginn 720 Series, and the MacMillan Series

	<u>A</u>	<u>YES!</u> <u>B</u>	<u>A & B</u>	<u>2</u>	<u>3</u>	<u>Ginn 720</u> <u>4</u>	<u>5</u>	<u>2+3+4+5</u>	<u>4</u>	<u>5</u>	<u>MacMillan</u> <u>6</u>	<u>7</u>	<u>4+5+6+7</u>
Total no. of words (tokens)	29	2133	2162	1792	1300	1781	6277	11,150	641	640	869	2474	4624
Total no. of different words (types)	10	214	217	38	81	142	506	550	80	104	113	245	326
Type-Token Ratio	.3448	.1003	.1004	.0212	.0623	.0797	.0806	.0493	.1248	.1625	.1300	.0990	.0705

The majority of the word types found in the first two books of the ESL series YES! will not be found in any of the four levels of either of the two basal reading series under investigation.

Testing this involved comparing the 217 word types found in the YES! A & B with word types found in either of the series. Simply stated, commonality of words on the YES! list with words in the basal reading series was being determined. The findings are summarized in Table 13.

TABLE 13

**Commonality of Words found on the YES! List
with Words found in the Ginn 720 (Levels 2+3+4+5)
and MacMillan (Levels 4+5+6+7) Lists**

	<u># of Words in Common with YES! A & B out of a possible 217 words)</u>	<u>% of Commonality</u>
Ginn 720	89	40%
MacMillan	68	31.3%

Table 13 clearly shows that the majority of the words found in the early books of YES! A & B are not found in the beginner reading books of the basal reading series used in this study. Naturally, many of the words found only in the YES! A & B series are low frequency words. However, investigation of the

percentage of low frequency words found in the YES! (A & B) and each of the basal reading series showed that there is a fair amount of consistency with respect to the percent of word types occurring only once. Table 14 shows this comparison:

TABLE 14
Percentage of Word Types Occurring
Only Once in Each of the Series

	<u>% of Word Types Occurring Only Once</u>
<u>YES!</u> (A & B)	32.3%
Ginn 720 (2+3+4+5)	27.5%
MacMillan (4+5+6+7)	28.5%

While the YES! (A & B) has slightly more word types occurring only once, the difference is not enough to explain the low percentage of common words between the basal reading series and YES! However, even among the fifty most frequent words found in YES! (A & B) there are several words not found in the basal reading series. Among the fifty most frequent words of YES! (A & B) there are fifteen words not anywhere on the Ginn 720 (Levels 2 - 5) list (i.e. 30% of the words are different from any of Ginn 720 words regardless of frequency) and nineteen words not found anywhere on the MacMillan list (Levels 4 - 7) (i.e. 38%). Furthermore, if the entire 217 words of the YES! (A & B) are examined we find that there are 106 words that are not found in either of the basal series lists (i.e. 48% of the 217 words types are unique to the YES! (A & B) series). A list of these words is found in Table 15.

TABLE 15

A Rank List of the 106 Words that
are Unique to the YES! A & B Books

1. color	31. pencil	61. bench	91. room
2. wearing	32. seven	62. cage	92. sam
3. understand	33. short	63. cookie	93. seesaw
4. train	34. shorts	64. ears	94. seventy
5. white	35. cup	65. england	95. sixty
6. nancy	36. gloves	66. fifty	96. slide
7. jane	37. milk	67. grandfather	97. sorry
8. sweater	38. sandwich	68. grandmother	98. start
9. drinking	39. six	69. hole	99. stockings
10. number	40. socks	70. horse	100. swing
11. crayon	41. u.s.a.	71. hundred	101. tea
12. desk	42. hands	72. jim	102. teresa
13. peter	43. hot	73. judy	103. toronto
14. washing	44. julia	74. kangaroo	104. twenty
15. listen	45. lemonade	75. kitchen	105. u.s.
16. susan	46. over	76. legs	106. (new) york
17. blouse	47. pat	77. living	
18. skirt	48. ramon	78. mail	
19. table	49. thirty	79. maria	
20. tie	50. tom	80. marta	
21. belt	51. twelve	81. mary	
22. brother	52. very	82. merry-go-round	
23. shirt	53. los angeles	83. montreal	
24. ten	54. australia	84. nigeria	
25. eight	55. banana	85. ninety	
26. hair	56. barn	86. orange	
27. nine	57. basket	87. pear	
28. pants	58. bathroom	98. pedro	
29. glass	59. bedroom	89. picture	
30. morning	60. bee	90. ricardo	

It is important to note at this time that half of these words do occur more than once in the YES! (A & B) books. Furthermore, these words are naming nouns. This suggests that the series, as Hypothesis 3 suggests, is attempting to expose the ESL child to the many things he/she sees in his/her environment.

The third part of Hypothesis 3, part (c), was designed to clearly show that the ESL series should introduce many new words much earlier in the series than the typical basal reading series. As mentioned earlier, Dolch (1960) outlined the typical basal reading series. He said that the pre-primer has 50 new words; the primer, 100 new words; the first year books, 150 new words; the second year books, 400 new words, and the 3rd year books, 600 new words. To test the distribution of new words in the YES! series, a readability score for each book had to be calculated so that the distribution of words could be compared with Dolch's claim. To do this, the Spache Readability Formula (1953) was used because it is suitable for estimating grade levels for reading materials for younger children. Following Spache specifications and recommendations for applying the Spache Formula, the following information was determined.

- a) Books A and B did not reach the minimum requirements for placement on the "Table for Quick Computation of the Readability of a Selection...." Thus, these books were considered pre-primer and primer, respectively.

- b) Book C consisted of passages that ranged from the pre-grade one level up to a grade level of 2.5. However, on average, the readability was within the grade one area and so was considered as such.
- c) Book D had passages that consistently fell within the mid-grade two level of reading on the "Table for Quick Computation" and thus was designated a grade two reading level.
- d) Book E had very broad range of readability scores for passages (ranging from 2.1 to 3.6). However, on average, the samples indicated that the book was at a grade three level.
- e) Book F had reading passages that ranged in readability from 3.0 to above the maximum score (4.1) on the readability chart. However, on average, the readability was that of a near-end-of-year grade three (3.8). Thus, this book, too, was considered a grade three level text.

In summary, the following grade level designations were given the six books in the YES! series.

Book A - pre-primer

Book B - primer

Book C - 1st year book (comparable to grade 1)

Book D - 2nd year book (comparable to grade 2)

Book E - 3rd year book (comparable to grade 3)

Book F - 3rd year book (comparable to grade 3)

Next to compare Dolch's (1960) claim of the distribution of words in basal series with that of the YES! series, the number of new words introduced in each of the six books had to be tabulated. This was done by using the Total Word Count list and designating the book in which each word was introduced. Then the number of new words introduced in each book was totalled.

Table 16 shows the results of this tabulation and compares these results with Dolch's numbers.

TABLE 16

**The Number of New Words Introduced in Each Book
of the YES! Series and the "Typical" Number of Words
Introduced by a Basal Reading Series (as outlined by Dolch, 1960)**

<u>YES! Series</u> <u>Book</u>	<u>Number of Words</u> <u>Introduced</u>	<u>Basal Reading</u> <u>Series</u>	<u>Number of Words</u> <u>Introduced</u>
A	10	Pre-primer	50
B	197	Primer	100
C	361	1st year book	150
D	880	2nd year book	400
E	699	3rd year book	600
F	653		

The results shown in Table 15 are very interesting. Books B, C and D have double the number of newly introduced words as their basal reading series counterparts, the primer, the first, and the second year books. Books E and F, on the other hand, do not appear to be so very different from their basal series counterparts. Speculation as to the reason for this distribution will be discussed in Chapter 5. However, it is important to note at this time that the total

number of new words introduced in the five levels (be it Books A to E or pre-primer to 3rd year books) is much higher for the ESL series than it is for the "typical" basal reading series. Therefore, part (c) of Hypotheses has been proven to be correct in that far more words are introduced in the ESL series than in the basal reading series.

The purpose of Chapter 4 has been to describe the results obtained from testing the three Hypotheses stated in Chapter 3. The hypotheses tested have been concerned with examining the YES! series, word distribution in order to determine if the series recognizes the special needs of the ESL child with respect to word distribution and whether the series is reflecting the standard language so that the ESL student's final accomplishment is in being able to communicate meaningfully and effectively in the English language. The results have shown that, 1) the words in the YES! series do reflect the target language; 2) the series develops to allow exposure to a large number of different words; and 3) the YES! series is not particularly similar to the basal reading series with which it was compared. The importance of these results with respect to the using frequency distribution information in teaching strategy and materials development will be discussed in Chapter 5.

CHAPTER 5

The purpose of this study has been to determine the extent to which the written language presented to ESL children represents the target language (English). More specifically, how and what vocabulary is presented to the ESL child has been of major concern. As was pointed out, the ESL child comes to the reading task with the serious disadvantage of having little oral meaning vocabulary upon which sight vocabulary and reading skills can be built. A decision regarding which words a child will need in order to communicate meaningfully can only be regarded as educated guessing. Predictions for word choice are typically based on foreseeing the kinds of situations children will most likely find themselves.

An ESL series such as the YES! series must recognize the needs and abilities of its users. As Melgren and Walker (1977) point out, the activities must be meaningful and have immediate value. The words that are used must be within the learner's experiences. However, the words that will be used are often determined by the topic. Beyond the most frequent words which help to structure all language, vocabulary use is largely a reflection of the context. The greater the number of different contexts, the greater the number of different words needed. Thus, vocabulary control becomes more and more difficult when a great many contexts are used.

The hypotheses stated in this study are concerned with the frequency distribution of the words in the YES! series. By looking at this dimension of the words occurring in the series, information comparing the distribution of words in the series with that of the target language and basal reading series was obtained. More specifically, the three hypotheses were concerned that the vocabulary load of the YES! series should achieve two goals:

- 1) The distribution of words should represent the target language so learners gain experience using real language and not some contrived simplified version.
- 2) The series should aim to quickly expand the number of words the learners come in contact with in an effort to expose them to the vast number of words that must become part of their sight vocabulary in order to read and communicate successfully.

These two goals reflect the needs of all young learners who will eventually be working within regular school system.

It is the purpose of this chapter to review the findings of the three hypotheses and reflect upon how these findings can be useful in helping teachers and materials writers expose learners to the vast number of words that must become meaningful. The first section of this paper will, therefore, deal with interpreting the data presented in the last chapter. Then, conclusions will be drawn and recommendations for further studies in this area will be made.

Interpretations

I. Hypothesis I

Hypothesis I stated that the word frequency distribution of the YES! series should be similar to the target language. This was tested by comparing distribution of words of the YES! series with published word lists that claimed to reflect 'standard' English in one or more of the language skills (oral, written, or read). A graph of the distribution of word types in the series (Graph 1) showed that the words used in the series did reflect the "ski slope" curve described by Twaddell (1980) who pointed out that the most frequent words were really very few in number while the low frequency words were what really made up the bulk of the words used. A second graph (Graph 2) revealed that published word lists also showed this characteristic "ski slope" shape and that the distribution of words in the YES! series was similar to those of the published lists.

This hypothesis was also tested by looking at the most frequent words of the YES! series and correlating them with the most frequent words of five published word lists. Four out of the five published words lists were highly correlated with the YES! list (i.e greater than .8). However, as was pointed out, correlations only deal with ranking, not with the more basic fact of the raw number of words that are similar between the YES! list and other lists. For example, the Walker (1979) list consists of one thousand words yet only 232 of those are found in the first 321 of the YES! series. On the other hand, of the 188 words on the Durr list, 147 of these are also on the YES! list. Recognizing that

the YES! list of 321 most frequent words was different from all other published high frequency lists suggested that the 321 words should be examined more closely. By simply forming a checklist and observing whether a word from the YES! list occurred on a published word list certain characteristics were observed. In making this comparison between the 321 words of YES! and four published word lists (the Dolch lists of "Basic Sight Vocabulary" and "95 Common Nouns" (1960); Hillerich's "240 Starter Words" (19740); Dale's "769 Easy Words" (1931); and Walker's list of 1000 base words of the Word Frequency Book (1979)) the following facts regarding the specific nature of the most frequent words were found:

- 1) All of the most frequent 27 words of the YES! list were found in the other four lists.
- 2) Words that were found on one or fewer of the published lists were typically,
 - a) proper nouns such as 'Mary', 'Tom', 'Dan', 'Sally' and 'Mexico'.
 - b) Nouns (particularly beyond the 200th most frequent word) such as 'zoo', 'elephant', 'lion', 'giant', 'circus', 'baseball', 'bike', 'trash', 'soup', and days of the week.
 - c) Contracted forms of verbs such as 'it's', 'what's', 'can't', 'I'm', 'don't', 'she's', 'didn't', 'he's', 'they're', 'doesn't', 'isn't', and 'that's'.
 - d) Numbers above ten (i.e. 'eleven', 'twelve', 'forty').
 - e) The '-ing' form of the verb (i.e. 'doing', 'wearing', 'looking', 'eating', 'listening', 'swimming', 'carrying', and 'washing').

The fact that the most frequent twenty-seven words of the YES! list are found on all lists is really not surprising in light of the evidence showing that the correlation of words is high between the YES! and the four published lists. However, the specific nature of many of the 321 most frequent words reveals some important information. Most important here is the contracted and '-ing' forms of verbs. While it must be recognized that most lists simply choose to cite only the base or root forms of words, the fact that the YES! series uses the contracted and '-ing' form so frequently cannot be ignored for two reasons. First, the contracted form of the verb is typically associated with oral, not printed language. The frequent occurrence of such forms suggests an attempt to expose learners to 'natural' speech.

Second, the existence of the '-ing' form of the verb occurring so frequently is highly interesting. This morphological ending appears more frequently in the list of 321 words than any other ending. Moreover, the '-ing' form of the verb often occurs before the root form of the verb. This suggests that much of the language being used is in the form of continuous action. Since the verb 'was' is not introduced until book D, much of the action must be in the present continuous form. This use of verbs seems to be highly consistent with the authors' wishes to focus on the 'here and now'. A quick look through Books B and C show this to be the case with such examples as,

He/She is wearing...

Book B, p. 40-44)

I/He/She/They am/is/are eating/drinking

Book B, p. 56)

What are/is you/he/she doing? I/She/He am/is
_____ing...

(Book B, p. 58 - 63)

(Book C, p. 42 - 47)

What is she/he looking for?

(Book C, p. 31)

When is/are he/she/you coming/playing/going...?

(Book C, p. 59, 63)

Why are/is you/he/she going...?

(Book C, p. 73)

The predominance of the progressive form is not insignificant. Brown (1973) studied the acquisition of fourteen grammatical morphemes by native English speaking children. (He described a grammatical morphemes as morphemes that either modified the meaning or clarified the relationship of the content words). Results showed that there was consistency in the order in which there were fourteen grammatical morphemes acquired. Most important, however, is the fact that the present progressive was found to be acquired first.

Furthermore, Dulay and Burt (1974) gathered evidence to show that the order of acquisition for the grammatical morphemes was universal. That is, regardless of the first language, all children acquire the English grammatical morphemes in the same order. Thus, this research supports the extensive use of the present progressive in the YES! series.

Larsen-Freeman (1978) has attempted to explain this universal order for the oral production of morphemes in terms of frequency rather than syntactic, semantic, or phonological complexity. Using already tabulated frequency counts for morpheme occurrences in the speech of English-speaking parents, Larsen-Freeman found significant correlation between the morpheme acquisition order of the second language learners. She concluded that more attention should be paid to the language environment to which the ESL learner is immersed:

Since grammatical morphemes have limited semantic weight, perhaps it is not in morpheme acquisition where the learner's cognitive involvement is evident in the second language task. Perhaps the creative talent of the second language learner is reserved for more complicated structures, while the learner concentrates on simply matching native-speaker input for structures at the morpheme level.

(Larsen-Freeman, 1978, p. 379)

Larsen-Freeman implies that certain aspects of the language are simply acquired by exposure and mimicry. While the assumption that morphemes are 'simple structures' is debateable, there is little doubt that frequency does play some role in the learning of language (Dale, 1976). The fact that the YES! series displays many repetitions of a grammatical morpheme that is of high frequency in the target language shows that the series does indeed represent the target language.

In examining the results of testing Hypotheses 1, the language ESL children are exposed to has been shown to represent the target language. The frequency distribution of words and the predominance of an early acquired grammatical morpheme suggests that learners are being exposed to realistic language samples.

II. Hypothesis 2

Hypothesis 2 examined how the frequency distribution of words was developed through the six books of the series in order to give learners ample opportunity to be exposed to many different words.

The results showed that, as expected, the number of different words (types) increased as the series progressed from Book A to Book F. This was partly due to the simple fact that there was an increase in the number of running words

(tokens) in the series as it progressed from Book A to Book F, but was also partly the result of there being less repetition of word types in the later books of the series. This result was reflected in the low correlations obtained when comparing the early books in the series with the more advanced books in the series when looking at the 321 most frequent words occurring in the series as a whole.

However, both these results, the decrease in the ratio of number of repetitions of words and the lack of correlation between the first and second half of the series, can be explained by the occurrence of lower frequency words. An examination of low frequency words (defined as occurring six times or less in the series as a whole) showed that they account for almost seventy-two percent of the total number of word types qualified to be called "low frequency words" (Table 10). This fact alone suggested that these words are going to dramatically affect any word distribution results.

It was necessary to consider how these low frequency words were distributed across the series. The results indicated that the number of low frequency words did increase as the series progressed. Furthermore, the words occurring only once were shown to account for over 40% of the low frequency words in Books B to F. This, then, is the key to the earlier results of decreasing word type repetitions and lack of correlation. The vast number of low frequency words affects the overall distribution of words in two ways. First of all, it decreases the number of times any word other than the extremely high frequency

words occur. Different contexts demand different words. The later books in the series cover many more different contexts and so have many more words which are important only within the given context. This leads to the second effect that the low frequency words have on the overall distribution of words. Because there are many more different contexts, and therefore many more different words, in the later books of the series the ranking of the most frequent words is affected. This, along with the fact that 77 of the 321 most frequent words in the series were not even introduced until Book D or later, is obviously going to alter the ranking of the most frequent words. When the early books are correlated with the later books the effects of the newly introduced words are evident.

The fact that the lower frequency words are so numerous and do account for such a large portion of the words the learner will encounter suggests that it is important to be aware of even the words occurring only once. Melgren and Walker have made lists of words that occur more than once in the series. However, considering there are 911 different words (accounting for over 32% of the word types in the series as a whole) that do occur only once it is worthwhile to know more about these words.

The 911 words that occur only once in the YES! series include derivations of more common words. Since the least common words are of greater importance in the later books of the series it can be assumed that common derived forms (i.e. -ed, -s, -ing, 's or contractions) would not pose much of a

familiarity problem for learners and therefore could be deleted from the list. This left 512 words that truly occur only once. Of these 512 words, well over 300 are nouns. Dolch (1960) pointed out that nouns are of little universal value because they are so context specific: since different contexts require different nouns, little can be gained from knowing these words. Many of the nouns occurring here are typical of Dolch's description. There are many nouns that simply name people and places and are therefore very context specific. However, a knowledge of the countries named may prove useful in constructing supplementary materials. For example, knowing that countries such as Germany, Finland, Sweden, and Poland are mentioned may initiate development of a unit on these countries.

The remainder of the singularly occurring words also offer starting points for supplementary materials. For example, 'ears', 'elbow', 'hips', 'knee', 'lungs', and 'toes' all occur only once. This suggests that while certain body parts are frequently referred to (i.e. head, arm, face) many other parts of the body are not named. Knowing that the naming of most parts of the body is not included in this series allows the teacher to develop additional material to expose learners to more words. Another area that could be developed is that of food. The words 'groceries' and 'menu' occur only once. Furthermore, there are many food words that occur only once (i.e. 'vegetables', 'corn', 'pear', 'muffin'). Using the concepts of grocery shopping or ordering off a menu could increase learners' exposure to these words and to related words that do not occur at all (i.e. meats such as chicken, beef, steak, hamburger, pork, roast, vegetables such as corn, peas, lettuce, and fruits such as pears, plums, grapes, cherries). This plan of taking a

low frequency word and expanding upon it can be also done for animals (words such as 'alligator', 'bull', 'leopard', and 'otter' occur only once and therefore suggest that there are many animals not mentioned at all in the series i.e. 'cougar', 'raccoon', 'skunk', 'porcupine', 'beaver', 'salmon' and 'robins'), occupations, and transportation vehicles (i.e. methods of getting from one place to another via motorcycle, subway, or submarine are all low frequency words).

Thus, the point of knowing which words occur infrequently offers the possibility of developing whole supplementary units that will expand the learners' experience. The above examples show that there is a great deal of opportunity for teachers and materials developers to create interesting and valuable activities to increase learners' familiarity with words they will see and use in the regular classroom and the community. The goal here is not to duplicate what series already provides but complement it with additional materials.

In testing Hypotheses 2 it was found that low frequency words are a very important part in the distribution of words. High frequency words are really very few in number. The words that are not repeated over and over really make up the bulk of the words in the language presented to ESL students. Like the target language, context controls how often all but the very high frequency words occur. The results suggest that rather than repeating a few words over and over, an approach that presents more variety is used. The basis for limiting the repetitiveness of words in favor of wider variety (via different contexts) encourages learners to relate what they already know in order to understand new

concepts. This, needless to say, is an essential step for ESL students. No ESL curriculum, no matter how extensive, can predict every word and context that will be encountered. Thus, offering a wide variety of contexts, and an opportunity to use and expand these contexts, is essential. In presenting the language to ESL learners in as natural a form as possible, the series is offering its users practical and realistic language experiences.

III. Hypothesis 3

Hypothesis 3 stated that the YES! series' word frequency distribution would not be similar to basal reading series that were designed to teach native speakers to read. In light of the results of the first two hypotheses, the findings that showed the YES! series has unique characteristics when compared with the basal reading series was not surprising. Table 12 showed that the words in the basal reading series were repeated far more often than the words in the first two books of the YES! series. A closer examination of the word types used in the two types of series (ESL and basal) was undertaken to determine the effect the differing type-token ratios had upon the words used in each of the series. Table 13 showed that the highest degree of commonality of words between YES! and the two basal reading series examined was 40%. In specifying the nature of the words unique to the YES! (A & B), a list of 106 words was created (Table 15). This list is predominantly made up of nouns (i.e. words for clothing, objects, people and places).

Finally, it was shown that the basal reading series introduces far fewer words per level combined. An explanation for these differences may be found in the differing assumptions authors of basal and ESL reading series make about their readers. The basal reading series writers can assume that the users already have a fairly large oral vocabulary. The focus is on repeating words many times so that they quickly become part of the child's meaningful sight vocabulary. The ESL series writers, on the other hand, are more likely to assume that its users have little oral vocabulary upon which to base reading instruction. Rather than repeating a few words over and over again, the author chooses to offer a wide variety of words that are, on average, repeated fewer times than those words found in the basal reading series. The ESL series must be concerned with total language experience while the basal reading series focuses on one language skill -reading. (This is not, however, to suggest that the basal reading series totally ignores other aspects of language. It only points out that the main concern, especially at the lower levels, is to teach children to read about things for which they already have an oral vocabulary.) The ESL series YES! recognizes the need for learners to build up a vocabulary base from which they can work and thus offers exposure to many different words.

Conclusions

This study has examined the frequency distribution of words in order to evaluate the language being presented to young ESL learners. Using the YES! series as an example of the language presented to ESL learners, word frequency

counts were used to compare the occurrence of words presented to learners with those in the English language which is being learned.

The results of this study have shown that the language being presented to ESL learners is highly representative of the target language. Word frequency distribution information has shown that the language in the YES! series is not contrived or abnormally repetitive. In recognizing the lack of experience ESL learners have with the language, the YES! series has sacrificed repetition in order to broaden experience by presenting a great variety of contexts. By creating many different language situations, the learner is exposed to the language naturally.

Context, however, does not guarantee the occurrence of words we may intuitively feel to be useful. Too often a context uses what are considered to be the most common words. This results in great gaps in the ESL learner's knowledge of names for things. (As was pointed out, a topic such as food may only use the most common words such as 'milk', 'bread', 'tea', and 'coffee'.)

Educators and materials writers need to recognize what the ESL text does in order to be able to use it effectively. The YES! series uses an increasing number of contexts to expose the ESL learner to many different words (in as 'natural' situations as possible) in an effort to build a vocabulary for oral language as well as for sight vocabulary. Thus, it is the lower frequency words that should be of interest to those wishing to utilize this series to its full extent.

A series such as YES! can be considered a base from which to develop additional activities in which the low frequency words and additional vocabulary (words not in the YES! series) can be used.

This study has shown that the YES! series is unique in the nature of specific words used. Many words found in the YES! series are not found in published lists or basal reading series because the YES! series focuses on vocabulary expansion which means lower frequency words are typically used. Vocabulary control in this series only occurs insofar as the contexts allow. The authors have not been afraid of exposing the learner to the many words he/she will soon encounter in the regular classroom. Teachers using this series should follow Melgren and Walker's example. There should be no fear in introducing new vocabulary if it is done in such a way as to supplement what is already familiar to the learner. The YES! series offers a great many situations and contexts which can be used as starting points for activities that re-use the low frequency words of the series and introduce new words. Since there is so little material for young ESL learners, careful examination of what does exist is essential for curriculum planning. Teachers and materials writers would do well to base new materials upon what already exists in the field. Not only is this pedagogically reasonable but also economically sound.

REFERENCES

- Aukerman, Robert A. The Basal Reader Approach to Reading. Toronto, John Wiley & Sons, 1981.
- Bloom, Lois (Ed.). Readings in Language Development. Toronto, John Wiley & Sons, 1978.
- Bormuth, John R. (Ed.) Readability in 1968. National Conference on Research in English, 1968.
- Brown, Roger. A First Language: The Early Stages. Cambridge, Mass., Harvard University Press, 1973.
- Buckingham, B.R. and Dolch E.W. A Combined Word List. Boston, Ginn & Co., 1936.
- Carroll, John B, Davies, Peter and Richmand, Barry. The American Heritage Word Frequency Book. New York, American Heritage Publishing Co. Inc., 1971.
- Causey, Oscar (Ed.). The Reading Teacher's Reader. New York, The Ronald Press Co., 1958.
- Chastain, Kenneth. Developing Second Language Skills: Theory to Practice, 2nd Edition. Philadelphia, Rand McNally College Publishing Co., 1976.
- Covell, H.M. "Worksheet for Application of the Spache Readability Formula." Vancouver, UBC, 1979.
- Croft, Kenneth (Ed.). Readings on English As A Second Language. Chicago, Winthrop Publishing, Inc., 1972.
- Croft, Kenneth (Ed.). Readings on English As A Second Language, 2nd Edition. Chicago, Winthrop Publishers, Inc., 1980.
- Dale, Philip S. Language Development: 2nd Edition. New York, Holt Rinehart and Winston, 1976.
- Dale, E. and Chall, J.S. "A Formula for Predicting Readability". Educational Research Bulletin (Ohio State U.), 27, 1, 1948.
- Dauzat, JoAnn and Dauzat, Sam V. Reading: The Teacher and the Learner. Toronto, John Wiley and Sons, 1981.
- Dolch, Edward W. "A Basic Sight Vocabulary". Elementary School Journal, 36, 6, pp. 456 - 460, 1936.

- Dolch, Edward W. Teaching Primary Reading. Champaign, Ill., Garrard Press, 1960.
- Dolch, Edward W. Methods in Reading. Champaign, Ill., Garrard Press, 1955.
- Dulay, Heidi and Burt, Marjna. "Natural Sequences in Child Second Language Acquisition." Language Learning, 24, 1, pp. 37-53, 1974.
- Durr, William K. "A Computer Study of High Frequency Words in Popular Trade Juveniles". The Reading Teacher, 27, 14, 1973.
- Earle, Richard A. Classroom Practices in Reading, Newark, I.R.A., 1977.
- Finn, Patrick J. "Word Frequency Information Theory, and Cloze Performance: A Transfer Feature Theory of Processing in Reading". Reading Research Quarterly. 13, 4, p. 508-537, 1977-78.
- Fry, Edward. "Developing A Word List for Remedial Reading". Elem. Eng., 34, 7, pp. 456 - 458, 1957.
- Fry, Edward. "Teaching A Basic Vocabulary". Elementary English, 37, 1, pp. 38 - 42, 1960.
- Gates, Arthur Irving. A Reading Vocabulary for the Primary Grades. New York, Bureau of Publications Teachers College, Columbia University, 1935.
- Gates, Arthur I. A Reading Vocabulary for the Primary Grades: Revised and Enlarged. New York, Bureau of Publications, Teachers' College, Columbia University, 1935.
- Ginn & Company. Rainbow Edition, Reading 720 Series. Lexington, Ginn & Co., 1980.
- Gleason, Jean Berko. "The Child's Learning of English Morphology", in Lois Bloom (Ed.) pp. 39 - 59, 1978.
- Goodman, Kenneth S. "Reading: A Psycholinguistic Guessing Game" in Singer and Ruddell, pp. 497-508, 1967.
- Goodman, Kenneth S. and Goodman, Yetta M. "Learning About Psycholinguistic Processes by Analyzing Oral Reading" in C.M. McCullough, pp. 179-201, 1980.
- Graves, Michael; Boettcher, Judith A.; Peacock, Judith L.; Ryder, Randall J. "Word Frequency as a Predictor of Students' Reading Vocabularies." Journal of Reading Behaviour, 12, 2, 1980.
- Harris, Albert J. and Jacobson, Milton. Basic Elementary Reading Vocabularies. New York, The MacMillan Co., 1972.

- Hatch, Evelyn (Ed.). Second Language Acquisition. Rowley, Mass., Newburg House Publishers, Inc., 1978.
- Higa, Masanori. "The Psycholinguistic Concept of 'Difficulty' and the Teaching of Foreign Language Vocabulary" in Kenneth Croft, 1972, pp. 292 -303.
- Hildreth, Gertrude. Teaching Reading: A Guide to Basic Principles and Modern Practices. New York, Henry Holt and Co., 1958.
- Hillerich, Robert L. "Word Lists - Getting It All Together." The Reading Teacher, 27, 4, pp. 353 - 360, 1974.
- Horn, E. "The Commonest Words in the Spoken Vocabulary of Children Up to and Including Six Years of Age." Report of the National Committee on Reading. 24th yr. book of the Society for the Study of Educ'n, Part I, Chapt. 7, 1925.
- Howes, Davis H. and Solomon, Richard L. "Visual Duration Threshold as a Function of Word-Probability." Journal of Experimental Psychology, 41, 6, pp. 401-410, 1951.
- Ingram, Elizabeth. "Psychology and Language Learning." In Press.
- International Kindergarten Union, Child Study Committee. A Study of the Vocabulary of Children Before Entering First Grade. Washington, D.C., Internation Kindergarten Union, 1928.
- Johnson, Dale. "A Basic Vocabulary for Beginning Reading." Elem. School Journal, 72, 1, pp. 29 - 34, 1971.
- Judd, Elliot. "Vocabulary Teaching: A Need for Re-evaluation of Existing Assumptions." TESOL Quarterly, 12, 1, pp. 71 - 76, 1978.
- Klare, George, R. "The Role of Word Frequency in Readability." in John Burmouth (Ed.), pp. 7-17, 1968.
- Kucera, H. and Francis, W. Computational Analysis of Present-Day American English. Providence, Rhode Island, Brown University Press, 1967.
- Larsen-Freeman, Diane. "An Explanation for the Morpheme Accuracy Order of Learners of English as a Second Language." in Hatch, pp. 371-379, 1978.
- LeBerge, D. and Samuels, S. "Toward a Theory of Automatic Information Processing in Reading." Cognitive Psychology, 6, 2, pp. 293 - 323. 1974.
- Lefevre, Carl A. "Reading Our Language Patterns: A Linguistic View - Contributions to a Theory of Reading." Challenge and Experiment in Reading. New York, IRA Conference Proceedings, Vol. 7, N.Y. Scholastic Magazine, pp. 66 - 69, 1962.

- Maclatchy, Josephine and Wardwell, Frances R. "A list of Common Words for 1st Grade." O.S.U. Educational Research Bulletin, 30, pp. 151 -159, 1951.
- MacMillan Publishing, Series r. MacMillan Reading. New York, The MacMillan Publishing Co., 1980.
- McCullough, C.M. (Ed.). Inchworm, Inchworm: Persistent Problems in Reading. Newark, IRA, 1980.
- Marks, Carolyn B.; Doctorow, Marleen J.; and Wittrock, M.C. "Word Frequency and Reading Comprehension." The Journal of Educational Research, 67, 6, pp. 259 - 262, 1974.
- Mason, Jana M. "The Roles of Orthographic, Phonological, and Word Frequency Variables on Word-NonWord Decision." American Educational Research Journal, 13, 3, pp. 199 - 206, 1976.
- Mellgren, Lars and Walker, Michael. YES! English for Children. Philippines, Addison-Wesley Publishing Co., 1977/1978.
- Miller, Alan. A Word Counting and Frequency Analysis Program. Vancouver, UBC Computing Centre, 1975.
- Morris, Joyce. "Barriers to Reading for Second Language Students at the Secondary Level." TESOL Quarterly, 10, 1, pp. 99 - 103, 1976.
- Murphy, Helen A. "The Spontaneous Speaking Vocabulary of Children in Primary Grades." J. of Educ'n (Boston), 140, 2, pp. 1 - 105, 1957.
- Nilsen, Don L.F. "Contrastive Semantics in Vocabulary Instruction." TESOL Quarterly, 10, 1, pp. 99 - 103, 1976.
- Noble, Clyde E. "The Familiarity - Frequency Relationship." Journal of Exptal Psychology, 47, 1, pp. 13 - 16, 1954.
- Otto, Wayne; Rude, Robert; and Spiegel, Dixie Lee. How to Teach Reading. Philippines, Addison-Wesley Publishing Co., 1979.
- Pearson, P. David and Studt, Alice. "Effects of Word Frequency and Contextual Richness on Children's Word Identification Abilities." Journal of Educational Psychology, 67, 1, pp. 89 - 95, 1975.
- Postman, L. and Solomon, R.L. "Perceptual Sensitivity to Completed and Incompleted Tasks." Journal of Personality, 18, pp. 347 - 357, 1950.
- Richards, Jack C. "A Psycholinguistic Measure of Vocabulary Selection." IRAL, 8, pp. 87 - 102, 1971.

- Richards, Jack C. "Word Lists: Problems and Prospects." RELC Journal, 5, 2, pp. 69 - 84, 1974.
- Rinsland, Henry D. A Basic Vocabulary of Elementary School Children. New York, MacMillan Co., 1945.
- Rivers, Wilga. Teaching Foreign Language Skills: 2nd Edition. Chicago, U. of Chicago Press, 1981.
- Rinsland, Henry D. A Basic Vocabulary of Elementary School Children. New York, MacMillan Co., 1945.
- Saville-Troike, Muriel. "Reading and the Audiolingual Method." TESOL Quarterly, 7, 4, pp. 395 - 406, 1973.
- Singer, Harry and Ruddell, Robert (Eds.) Theoretical Models and Processes of Reading: 2nd Edition. Newark, IRA, Inc., 1976.
- Smith, Frank. Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Read: 2nd Edition. Toronto, Holt, Rinehart, and Winston, 1978.
- Solomon, R.L. and Howes, J. "Word Frequency, Personal Values and Visual Duration Thresholds." Psychological Review, 58, 4, pp. 256 - 270, 1951.
- Solomon, R.L. and Postman, L. "Frequency of Usage as a Determinant of Recognition Thresholds for Words." J. of Experimental Psychology, 43, pp. 195 - 201, 1952.
- Spache, George. "A New Readability Formula for Primary Grade Reading Materials." Elementary School Journal, 53, 7, pp. 410-413, 1953.
- Stone, David R. "A Sound-Symbol Frequency Count." The Reading Teacher, 19, 7, pp. 498 - 504, 1966.
- Strothers, C.E.; Jackson, R.W.B.; Minkler, F.W. A Canadian Word List: Grades I - VI. Toronto, The Ryerson Press, 1947.
- Thorndike, Edward L. The Teacher's Word Book. Columbia University, Bureau of Publications Teachers College, 1921.
- Thorndike, Edward L. and Lorge, Irving. The Teachers Word Book of 30,000 Words. Columbia University, Bureau of Publications Teachers College, 1944.
- Twaddell, Freeman. "Linguistics and Language Teachers." in Kenneth Croft (Ed.), pp. 268-276, 1972.

Twaddell, Freeman. "Meanings, Habits, and Rules," in Kenneth Croft (Ed.), pp. 15 - 22, 1972.

Twaddell, Freeman. "Vocabulary Expansion in the TESOL Classroom," in Kenneth Croft (Ed.), pp. 439-457, 1980.

Walker, Chalres Munroe. "High Frequency Word List for Grades 3 thru 9." The Reading Teacher, 4, pp. 803 - 811, 1979.

Wardhaugh, Ronald. "Theories of Language Acquisition in Relation to Beginning Reading Instruction." Language Learning, 21, 1, pp. 1 - 14, 1971-72.

Yorio, Carlos A. "Some Sources of Reading Problems for Foreign - Language Learners." Language Learning, 21, 1, pp. 107 - 115, 1971-72.

APPENDIX

Spache Readability Formula

CLARENCE E. STONE'S REVISION OF THE
DALE LIST OF 769 EASY WORDS

a	bath	building	corner	everything
about	be	bump	could	eye
acrossbear	bunny	count		
afraid	beautiful	bus	country	face
after	became	busy	cover	fall
afternoon	because	but	cow	family
again	bed	butter	cried	far
air	bedroom	buy	cross	farm
airplane	bee	buzz	crumb	farmer
all	been	by	cry	fast
almost	before		cup	fat
alone	began	cabbage	cut	father
along	begin	cage		feather
already	behind	cake	dance	feed
also	being	calf	dark	feel
always	believe	call	day	feet
am	bell	came	dear	fell
an	belong	can	deep	felt
and	beside	candy	deer	fence
animal	best	cap	did	few
another	better	car	dig	field
answer	between	care	dinner	fill
any	big	careful	dish	find
anyone	bigger	carry	do	fine
anything	bill	cat	does	finish
apple	bird	catch	dog	fire
are	birthday	caught	doll	first
arm	bit	cent	done	fish
around	black	chair	don't	fit
arrow	blew	chick	door	five
as	blow	chicken	down	flag
ask	blue	child	draw	flew
asleep	board	children	dress	floor
at	boat	circus	drink	flower
ate	book	Christmas	drive	fly
away	both	city	drop	follow
automobile	bottom	clap	dry	food
	bow	clean	dock	foot
baa	bowl	climb		for
baby	bow-wow	close	each	found
back	box	clothes	ear	four
bad	boy	clown	early	fox
bag	branch	cluck	east	fresh
bake	bread	coat	eat	friend
baker	break	cock-a	egg	frog
ball	breakfast	doodle-do	else	from
balloon	bright	cold	elephant	front
band	bring	color	end	fruit
bang	brother	come	engine	full
bark	brought	coming	enough	fun

barn	brown	cook	even	funny
barnyard	bug	cooky	ever	
basket	build	corn	every	game
	himself	left	Mrs.	peanut
garden	his	leg	much	peep
gate	hit	let	mud	pennies
gave	hold	let's	music	people
get	hole	letter	must	pet
girl	home	lie	my	pick
give	honey	light		picnic
glad	hop	like	nail	picture
go	horn	line	name	pie
goat	horse	lion	near	piece
God	hot	listen	neck	pig
going	house	little	need	pink
gold	how	live	nest	place
*good	hunt	long	new	plant
good-bye	hurry	look	next	play
got	hurt	lost	nice	please
grandfather		lot	night	pocket
grandmother	I	loud	no	point
grass	ice	love	noise	policeman
gray	if	lunch	north	pond
great	I'll	made	nose	pony
green	in	mail	not	pop
grew	Indian	make	note	poor
ground	inside	man	nothing	post
grow	into	many	now	present
guess	is	march	nut	press
	it	matter		pretty
had	its	may	of	puff
hair		me	off	pull
hall	jar	meat	often	push
hand	joke	meet	oh	put
happen	jump	men	old	puppy
happy	just	meow	on	
hard		met	once	quick
has	keep	mew	one	quiet
hat	kept	mice	only	quite
have	kill	might	open	
hay	kind	mile	or	rabbit
he	kitchen	milk	orange	race
head	kitten	milkman	other	rain
hear	knew	mill	our	rake
heard	knock	minute	out	ran
heavy	know	miss	outside	read
held		Miss	over	ready
hello	lady	money	own	real
help	laid	monkey		red
hen	lamb	moo	paint	rest
her	land	more	pan	ride
here	large	morning	paper	right
herself	last	most	park	ring
hid	late	mother	part	river
hide	laught	mouse	party	road
high	lay	mouth	pat	

hill	learn	move	paw	robin
him	leaves	Mr.	pay	rock
rode	six	summer	today	wear
roll	skate	sun	toe	wee
roof	skin	sunshine	together	weed
room	skip	sure	told	week
rooster	sky	surprise	tomorrow	well
root	sled	swam	too	went
rope	sleep	sweet	took	were
round	sleepy	supper	top	west
row	slide	swim	town	wet
rub	slow	swing	toy	what
run	small		train	wheat
	smell	table	tree	wheel
said	smile	tail	trick	when
same	smoke	take	tried	where
sand	sniff	talk	trunk	which
sang	snow	tall	try	while
sat	so	tap	turkey	white
save	soft	teach	turn	who
saw	sold	teacher	turtle	why
say	some	teeth	two	wide
school	something	tell		wild
sea	sometime	ten	uncle	will
seat	song	tent	under	win
see	soon	than	umbrella	wind
seed	sound	thank	until	window
seem	soup	that	up	wing
seen	splash	the	upon	winter
sell	spot	their	us	wish
send	spring	them	use	with
sent	squirrel	then		without
set	stand	there	vegetable	woman
seven	star	these	very	wonder
shake	start	they	visit	wood
shall	station	thin	voice	woke
she	stay	thing		wolf
shell	step	think	wagon	word
sheep	stick	this	wait	work
shine	still	those	wake	world
shoe	stone	though	walk	worm
shop	stood	thought	want	would
short	stop	three	war	write
should	store	threw	warm	
show	story	throw	was	yard
shut	straight	ticket	wash	year
sick	street	tie	watch	yellow
side	string	tiger	water	yes
sign	strong	time	wave	you
sing	such	tired	way	your
sister	suit	to	we	
sit				zoo