CHAPTER - II
CHAPTER II

REVIEW OF RELATED STUDIES

2.1. Introduction

2.2. Studies in English as First Language

2.3. Studies in Indian Languages.

Summary
CHAPTER II

REVIEW OF RELATED STUDIES

2.1. Introduction

The review of studies made in connection with the present study relates to the study of vocabulary in English language and in Indian languages, as mother tongue, for the purpose of teaching learning materials mainly dealing with primary grades. Some of the studies included in this review are traced from the reviews in previous individual research studies and from compilation of studies and some are studies themselves reviewed by the present author. The full names of the authors of the studies referred to in previous studies have been mentioned where as only the titles of the authors of the studies reviewed by the present author are mentioned.

This report of review is prepared under two broad headings. One is 'Studies in English as first language' and the other is 'Studies in Indian languages'. The number of studies in English language being large, they are again placed under four headings like: 'Reading vocabulary', 'Written vocabulary', 'Spoken vocabulary' and 'Other studies on children's vocabulary'. The review of studies in Indian languages has been made mainly under
the names of different languages such as: Hindi, Urdu, Gujrati, Marathi, Kannada, Bengali, Punjabi, Tamil and Oriya. A project on 'Phonemics and morphemics' in six Indian languages has also been included in the review.

2.2. Studies in English as First Language

2.2.1. Reading Vocabulary:

Bongers (1947) seems to be the first historian to give an elaborate account of the 'vocabulary control' as well as 'vocabulary gradation' in the history of English language teaching, particularly as a foreign language. But some of the studies he reviewed are related to the teaching of first language also in so far as principles are concerned.

According to Bongers the Pestalozzian Primers prevailing in 1836 were the embodiment of earliest 'graded vocabularies'. Such gradation was based purely on the assumption of the experienced teachers. Such selected vocabularies meant for gradation were the 'familiar' and useful ones in terms of 'child activity'. This method of selection and gradation of vocabularies meant for early readings of the children could prevail
in the western countries till the advent of the 'objective' method of 'word count' in the field of education during the second decade of the current century.

The objective method of word count for the purpose of vocabulary control and gradation for children's readings first dealt with the children's own school texts, because the 'school primers' and 'first readers' were the earliest as well as the most common reading materials. One of the first studies of objective word count, referred to by Bongers, was conducted in 1918 by E.T. Housh. He had counted and tabulated the words of ten second years readers and found that more than 50% of the words of such readers had the frequency of only three or less.

Besides Housh's study Bongers's history contains review of many such similar studies. The ones which may have some importance for the present study are of Thorndike (1921), Packer (1921), Selke and Selke (1922), Gregory (1923) and Kircher (1925). Rinsland (1945) and McNally & Murray (1968) also reported two similar studies of Wheeler and Howell (1930) and of Hockett (1939) on reading vocabulary. Before detailing Thorndike's study, the one which has had much influence in the field of
research on word count and vocabulary control and which is important even today, a brief account of all the other studies is given below.

J.L. Packer (1921) studied ten first year readers from which he found 3541 different words. Of these different words as much as 58% had frequency of only one to four. In analysing the readers of third grade children C.A. Gregory (1923) found 5190 different words of which 30% had the frequency of only one. As would be observed from the studies of Housh, Packer and Gregory, all of them found out the same truth about the primary readers used then that there was little scope of repetition of most of the words which is generally recommended as a principle for the primary readers for the purpose of effective learning.

The study of Erich Salke and G.A. Salke (1922) is rather different one. They had analysed beginning readers under twelve reading methods and found that several books introduced several times as many words as others, and the amount of repetition in different books varied considerably. Few readers had more than 30% words in common with one another. The average percentage of words common to two books was found to be 28. H.W. Kircher (1925) seems to have gone a little further in finding a basis for a set of graded vocabulary for the
construction of first reader. He worked on 37 books comprising primers, first readers and spellers from which were found 5000 different words. Of these different words only 650 words had sufficient frequency to form the base for the desired set of graded vocabulary. H.E. Wheeler and Ema A. Howell (1930) did a 'First-Grade Vocabulary Study' by analysing two primers and ten first-readers published between 1922 and 1929. As a result of this study they prepared a list of 453 words most frequently used in those texts. John Hockett (1939) studied 'The Vocabularies and Contents of Elementary School Readers' pertaining to Grades I to IV and found that there was wide disagreement in both the number of running words and of different words in elementary readers of these grades.

Coming to Thorndike (1921)'s pioneer work on 'word count' one cannot but be astonished to see the huge bulk of materials that was dealt with for the purpose. The amount of running words counted in this study, according to Bongers, was nearly 4,000,000. Thorndike's original book - THE TEACHER's WORD BOOK (1921) - mentions the exact quantity to be 4,565,000 running words from the following sources:
<table>
<thead>
<tr>
<th>SPECIFICATION OF SOURCE</th>
<th>RUNNING WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature for Children</td>
<td>625,000</td>
</tr>
<tr>
<td>The Bible and English Classics</td>
<td>3,000,000</td>
</tr>
<tr>
<td>Elementary School Text Books</td>
<td>300,000</td>
</tr>
<tr>
<td>Books about Cooking, Sewing, Farming, the Trades and the like</td>
<td>50,000</td>
</tr>
<tr>
<td>Daily Newspapers</td>
<td>90,000</td>
</tr>
<tr>
<td>Correspondence</td>
<td>500,000</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>4,565,000</strong></td>
</tr>
</tbody>
</table>

Out of these running words, Thorndike prepared an alphabetical list of the 10,000 words that were found to occur most widely in the count. He took care of both the 'range' and 'frequency' of each word. Range means, 'How widely is the word used.' Frequency means, 'How often is the word used.' Prior to Thorndike none of word-counters had calculated the 'range' of a word. So, Thorndike's is the first study concerning the teaching of English as the first language in school education, which took care of both 'range' and 'frequency' of the vocabularies for gradation.

Thorndike's TEACHER's WORD BOOK was meant to help teachers in estimating the commonness and importance of words. Such an estimation of a word on the part of the teacher would help to decide quickly for treating the word appropriately while teaching it to particular grade children, because in a particular-grade-reader all the words are
not supposed to get equal treatment in teaching. Some may be taught thoroughly for permanent assimilation by the children, some may simply be explained for understanding the context by the children and some may not be necessary to be taught to them. Hardie (1948) reports that Thorndike himself believed his word-list would, in general, answer the question as to which words are to be taught to which classes in school. This happens to be the most important problem confronted by primary school teachers which is beyond their capacity to solve. The first work of Thorndike was estimated to be the most valuable and reliable one by many critics except Palmer and Dewey who found it to be inadequate as reported by Bongers (1947). But Thorndike (1921) did not forget to confess the inadequacy of his work as is found from his following statement:

This list is not a perfect measure of the importance of words for two reasons. First, a word may be very important for a pupil or graduate to know and yet not figure largely in the world's reading. Second, tens and thousands of hours of further counting would be required to measure the frequency of occurrence of all these words with exactness. If a complete count were made, there would probably be several hundred words found more deserving of a place in the top ten thousand than some of those now included, and the order of the list would be somewhat changed. (pp. III, IV)

Thorndike (1921) claimed his list to be 'far better' than any that had been available then. He was, however,
not content with this list. He had to work continually to improve his first list by adding counts of words amounting to 500,000 from over 200 other sources and thus prepared a list of 20,000 words on the same basis of 'frequency' and 'range' in the year 1931, i.e. a decade after the publication of his first list of 10,000. This time he held the view that the range of a word is the better indicator of the importance of the word than its frequency, i.e. a word which occurs fifty times (frequency) in one book is less important than the word which occurs once in each fifty books (range). Thorndike's work did not end here. In 1936 and 1937 he worked out a word list of about 38,000 on the analysis of 4.5 millions words from 120 books written for children in Grades III to VIII (Rinsland, 1945). Further years of his investigation into the vocabulary and lastly in collaboration with Lorge he produced a greatly improved list of 30,000 words in 1944 (TEACHER'S WORD BOOK OF 30,000 WORDS). This work is the extension of Thorndike's two earlier lists, 10,000 (1921) and 20,000 (1931) plus three other counts of over 4.5 millions words each including 'Lorge magazine count'. This is, says MacGinitie (1970), 'the most extensive word count in English'. In this work also Thorndike followed the same procedure of giving importance to the words in terms of 'frequency' and 'range'. Besides these three publications (1921, 1931 and 1944) Thorndike published many
papers on vocabulary studies relating to books of children in school grades and said a lot about it as has been mentioned by many subsequent investigators in the field. All of them reckoned that Thorndike's prodigious amount of work contributed a lot to the cause of language education, especially at the elementary stage. The most interesting conclusions of Thorndike's work are:

(1) There are slightly more than 1000 words with frequency of more than 100 per million words. These words are recognised as 'AA words'.

There are slightly less than 1000 words with frequency between 50-100 per million. These words are recognised as 'A words'.

(2) Words which the child meets in his reading and everyday life may be divided roughly into two groups. The first group contains words important for the child to learn as part of his permanent intellectual equipment.

The second group contains the remainder; it is probably sufficient if the child is told the meanings of these (by the teacher or by a dictionary) so that he can understand the context in which they occur, but it is not necessary for them to be retained as permanent items in the child's language.

(3) The following rules can settle the finding of the permanent word for the child:

"1. If it is an AA-word (except a few of interest mainly to adults or too hard for a child in grade 1 or 2 to understand—such as 'according')."
"2. If it is a word of great practical service such as 'poison'; or

"3. If it is a word understood by the class when heard and specially useful for learning the phonetic equivalent of certain combinations of letters—such as 'block'."

In spite of such important conclusions, Thorndike's study is not free from some mild as well as severe criticism from those who tried to decide children's vocabulary by some different methods or from those who simply tried to estimate Thorndike's works on vocabulary from logical points of view. Some such criticisms and appreciations are worth mentioning here.

Hardie (1948) says that Thorndike did not specify words of interest mainly to adults or words which are too hard for a child in the first two grades to understand; nor did he offer any information about the words which are supposed to be of great 'practical service'. Pointing to 'THE VOCABULARY OF BOOKS FOR CHILDREN IN GRADES 3 to 8 (Thorndike, 1936-37), Rinsland (1945) says, 'Thorndike (1936 and 1937) has shown something of the very large vocabulary load in elementary books'. Rinsland further says that the 'extensive and ill-chosen' vocabularies for the elementary school children 'conform to no laws of learning, either scientific or philosophical'. More over, adults estimate of children's abilities are
always likely to be overestimates. Burroughs (1957) in reviewing the word lists based on books written for children, states that such lists are open to 'certain objections', because such lists assume some particular validity for the powers of observation of writers of children's books. He concludes that children experience difficulties in reading their books for such assumption. Again words which are easy for adults may not be so easy for children or may be so empty of significance for them that they are difficult to learn. Burroughs, however, appreciates Thorndike's model list of this kind because of its wide use for comparison. Burroughs also believes Thorndike's own words, 'The list tells any one who wishes to know whether to use a word in speaking, writing or teaching, and how common the word is in standard English reading matter'. Bongers's (1947) appreciation of Thorndike's study on word count is observed from his following statement:

Of the many word counters that have been busy since the publication of the Thorndike lists, few have added anything new either to the result or the technique. (p.45)

Pattanayak (1973) says that Thorndike's team working for about twenty years might have found 'chicken' to be a high frequency but not 'hen' on the basis of usage, yet 'hen might be of greater importance for a text on poultry'.

Shanker (1978) is of opinion that it is not justified to decide the usability of a word in the children's text on the basis of its frequency and range of occurrence in the language of the adult world. She, however, admits that Thorndike's word list is of help to language textbooks writers in grading the vocabulary for progressive introduction by instalments.

One of the contemporaries of Thorndike is C.R. Stone who did a great work on "GRADED VOCABULARY FOR PRIMARY CHILDREN (1941)" by word-counts of primary readers. Stone has been quoted by Strickland (1957) and McNally and Murray (1968). Stone analysed 105 textbooks which include twentyone each of pre-primers, primers, first readers, second readers and third readers, published during the period 1931 to 1941. His work was also based mainly on word count. He established 2164 words by selecting those which appeared frequently and graded them for successive stages of primary school reading. He also listed 150 words separately being the 'most important' for 'Beginning Reading'.

Roderick G. Langstone listed a 'core vocabulary' analysing thirtytwo pre-primers and found that only 78 words comprised two-thirds of the total running words. His establishment of 'core vocabulary' was meant for
pre-primer reading only. (McNally and Murray, 1968).

Kyte (1930) also provided a list of 'core vocabulary' containing 100 words selected from different basic reading schemes. This list, it is said, is a useful one for all phases of early school work pertaining to 'words'.

Vernon (1949) did 'word counts of infant readers' as a part of his investigations conducted for The Scottish Council for Research in Education during 1939-40. This count involved seven sets of Infant Readers, 17 books in all commonly employed in Glasgow schools. On analysis of these books 1495 different words were found out. By examining further Vernon assigned an index of importance to each word as judged by American standard. This revealed the fact that 67% of the words used in those Infant Readers would be unsuitable by American standard. Following are the conclusions arrived at by Vernon's investigations:

- there are gross variations between readers in total vocabulary employed and numbers of repetitions of words in range of ideas or content, and in the use made of phonics;

- the first book to which children are introduced should be, but very seldom is, based mainly on the children's own speech vocabulary, though less familiar words, especially those common in adult speech and reading matter, should be introduced gradually in later readers;

and - the content of readers is often remote from the experiences and interests of children, particularly of urban children, Phonic primers are the most offenders and often present pages of gibberish which discourage rather than stimulate the average child's desire to read.
2.2.2. **Written Vocabulary**

The earliest study on written vocabulary was conducted in the year 1926 by Ernest Horn who is traced from Bongers's history (1947). Horn worked to provide 'A Basic Writing Vocabulary' compiling 10,000 words most commonly used in writing. He could recognise that words used in correspondence might make a more relevant list for 'certain uses'. So, he tabulated about 5 million running words, from both personal and business correspondence to find out 10,000 most commonly used ones on the basis of relative frequency. Horn's successor in this field, Rinsland (1945) remarked that although Horn's is a widely known study having a comprehensive sample, the sample is but of adults' writing and not of children's. Rinsland, on the other hand, claims that children, especially in the elementary school, do not use words with the same frequency as adults, and that adult usage is therefore a more or less invalid criterion. But it is found out from MacGinitie's (1970) statement reproduced below that Horn in fact accepted children's writing for such a study:

Horn also recognised that a word count of fundamental value as a basis for reading instruction and other early educational materials would be a count of the words actually used and understood by children.
It is, however, not understood as to why Horn did not choose a sampling of children's writing vocabulary although in the previous year (1925) he had worked on spoken vocabulary of children, as would be mentioned in this chapter under 'Spoken Vocabulary' and though he had the belief in it as MacGinitie (1970) pointed out above.

Rinsland (1945), however, materialised such a belief in a very comprehensive manner. His work, A BASIC VOCABULARY, is an extensive list from a count of 6 millions running words of children's composition and conversation. Although Rinsland had worked on vocabulary studies since 1926, both singularly and with associates [as would be known from the studies like: 1. THE TECHNICAL VOCABULARY OF A SERIES OF ARITHMETIC TEXTBOOKS (Rinsland, 1926), 2. THE VOCABULARY OF ELEMENTARY SCHOOL CHILDREN (Rinsland and Moore, 1937), 3. THE VOCABULARY OF ELEMENTARY SCHOOL CHILDREN OF THE UNITED STATES (Rinsland and Moore, 1938) and 4. THE VOCABULARY OF ELEMENTARY SCHOOL CHILDREN OF AMERICA (Rinsland, 1939)], all of which Rinsland himself referred to in his study of 1945 his work of 1945 is being referred to very widely even today, by the researchers of his own country, USA and outside as well. But before detailing this work of Rinsland it is necessary to mention
about a combined work of Buckingham and Dolch who seem to be the contemporaries of both Horn and Rinsland and the result of whose vast work on children's writing vocabulary conducted in the year 1935, tried to provide a list of words that was 'as useful as possible for guiding educational work' (MacGinitie, 1970).

B.R. Buckingham and E.W. Dolch (1936) prepared a combined word list with the purpose of determining the vocabulary suitable for school books for the various grades in the elementary stages. Both the investigators carried on this research known as 'Free Association Study'. It involved counting of about 3,700,000 words written by over 20,000 children from grades II to VIII. Those children, at the end of their school session, were asked to write 'as many words as they could think of in fifteen minutes'. The investigators explained in the following statement (Bongers, 1947) as to why such a procedure they did follow. The procedure, according to them,

"eliminated the necessity of counting thousands of running words in order to secure a sufficient number of different words and also eliminated the counting of the articles, prepositions, pronouns and other structural words which make up the bulk of the written language, and which appeared but seldom in the children's lists." (p.45)
The investigators claimed that the lists of words written by the children represented the 'words learned during the course of that year' since the children write the words at the end of school year. To finalise this study the investigators used ten other lists, over and above the words collected from the children. Commenting upon this study Bongers (1947) says:

The only value of this work lies in the light it throws on the methods that were often used and assumptions upon which vocabulary statistics are some times made. (p.45)

Now coming to Rinsland (1945)'s A BASIC VOCABULARY proper it is fit to start with the specific 'problem' that inspired him to engage himself for the study. The problem is how 'to present the actual frequency of the use of words, grade by grade, of many children from all sections of the country' and how 'to give these data in such a way that they may be helpful to the teacher, writer of textbooks and the students of learning'. Such a problem naturally suggests an extensive nationwide comprehensive study. For this study Rinsland planned 'children's writing for children's consumption' and 'many words from many children, widely and wisely sampled' from the whole country which would furnish most of the words needed for the 'average child and his textbooks'. The sample of the study consisted of 'words written by children of grades I to VIII comprising more
than 200,000 individuals representing about 1% of the approximately twenty million elementary school children in the United States. They were drawn from 708 schools which is a 42.7% of 1500 selected schools situated in 416 cities, 236 counties and 56 training schools from six geographical areas. Assuming the limited writing ability and experience of the first grade children, Rinsland included 4630 pages of conversation of first grade children in the sample. Rinsland claims his sampling to be the richest one and comparable to only the largest count from the writings of adults. For the purpose of counting Rinsland used only 100,212 written papers. His counts involved 25,632 different words against 6,012,395 running words, as detailed below, gradewise.

<table>
<thead>
<tr>
<th>GRADES</th>
<th>RUNNING WORDS</th>
<th>DIFFERENT WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>353,874</td>
<td>5,099</td>
</tr>
<tr>
<td>II</td>
<td>408,540</td>
<td>5,821</td>
</tr>
<tr>
<td>III</td>
<td>770,019</td>
<td>8,976</td>
</tr>
<tr>
<td>IV</td>
<td>792,326</td>
<td>9,976</td>
</tr>
<tr>
<td>V</td>
<td>835,130</td>
<td>11,449</td>
</tr>
<tr>
<td>VI</td>
<td>853,409</td>
<td>11,304</td>
</tr>
<tr>
<td>VII</td>
<td>910,754</td>
<td>14,820</td>
</tr>
<tr>
<td>VIII</td>
<td>1,088,343</td>
<td>17,930</td>
</tr>
</tbody>
</table>

TOTAL: 6,012,395

Regarding varying data of running words from grade to grade, Rinsland explains that since the materials from the schools were made available by 'voluntary service' and that since there is a limit to such service a definite number of compositions for each grade could
not have been ordered. Therefore, 'the number of running words sampled from grade to grade was not according to any given formula of sampling.' (By 'formula of sampling' Rinsland means Thorndike's suggestion that for determining the first five thousand words for ordinary purposes a count of five million words would be required.)

Rinsland observed the following rules for counting and tabulating the words to ensure uniformity of treatment to words.

(1) Count all words - roots, derived forms, abbreviations, and contractions- just as they occur.

(2) Tally separately run-together words.

(3) Delete baby talk unless terms are found to be good English words.

(4) Delete illegibles.

(5) Count the correct forms intended when words are spelled unusually or wrongly.

(6) Delete slangs, provincialisms, colloquial expressions, as determined by the dictionary as well as trade names and proper names of persons and places, except very wellknown terms.

(7) Do not tally separately words that may be written in two ways, but consider them the same for purposes of tabulation. Examples are, 'today' and 'to-day', 'tonight' and 'to-night', 'tomorrow' and 'to-morrow'.

(8) Tabulate the correct form intended when errors occur in the use of homonyms. 'Two', 'too' and 'to' are to be tabulated separately.

(9) Tabulate separately words that are compounded if the compounding is incorrect or is used for running composition.
Rinsland was certain that the different words (25,632) found out in his study were more than were needed for a basic or major vocabulary for the average group of children in each grade. Therefore, after illustrating the 'irregular variations in the frequency of occurrence of words from grade to grade' and due to the inevitability of 'uneven sampling', Rinsland decided to list out separately 14,571 words occurring with a frequency of three or more in any one grade, as 'Basic Vocabulary' of elementary schools for grades I to VIII. Nevertheless he also suggested that the remaining 11,061 different words which occurred once or twice are of value. He repeatedly said that the list of words prepared by him should not be used as 'total' vocabulary of learning for the elementary school, but only the 'basic, general, major' vocabulary. The 14,571 basic words are arranged in alphabetical order along with 'actual or raw frequency for the concerned grade', 'total frequency for all grades' and an 'index symbol of the frequency group by hundred, five hundred and thousand' against each word. Rinsland (1945) has high claims for his study some of which are evident from his own statements:

The large number of words given in this list beyond the first few thousand should be useful for the middle and advanced grades. (p.17)
Though there are still many unknowns concerning vocabulary problems, the data on frequency of words for each grade and all grades in this study give a basis for decisions by writers never before available (p. 19).

Not only is there a practical use of the word list in writing textbooks but also in writing stories, pupils' magazines, and reference books... In the field of educational and mental measurements the data in this study should be invaluable. (p. 19)

Commenting upon Rinsland's study of 1945, Burroughs (1957) says that the method of collection adopted by Rinsland selects the vocabulary in an unpredictable way. Rinsland (1945), however, did not forget to admit the unpredictability of his study, for he said, 'At the beginning of the study no one knew just how many words to count in each grade ......' and 'there are other variations not expected by one unacquainted with sampling thousands of different words ......'. The reason why Burroughs (1957) did so comment is that he did not really comment upon Rinsland's methodology in particular but upon the method of study of children's vocabulary by collecting writing vocabulary of children, in general. He advocated that collection of children's 'spoken vocabulary' is the best method for children's vocabulary. In spite of this, Burroughs's own countrymen, Edward and Gibbon (1964), after about seven years of Burroughs's study, chose 'written words, rather than spoken words' because in their opinion it is the written words which
'contribute the vocabulary that children are most likely to recognise in print'. The discussion of their study will follow later. Burroughs (1957), however, appreciates Rinsland's (1945) study so far as its organisation of data is concerned and says that this study meets adequately the comparison test. 'Rinsland's work is a landmark in this field', say Pasricha and Das (1959). Edward and Gibbon (1964) while commenting upon the studies of American origin, including Rinsland's, say that pervading all these surveys is a 'lack of adventure, fun and imagination' and that the lists appear 'pedestrian and staid'. The author of a Canadian study on THE BASIC WRITING VOCABULARY OF ELEMENTARY SCHOOL CHILDREN, Thomas (1972) says the following about Rinsland's (1945) study:

The major research thus far concerning children's own usage was carried out by Henry D.Rinsland in his nationwide study of writing samples. No similar major studies appear to have been carried out in Canada. (p.243)

But in view of the dynamic nature of language that brings new words into speech and writing pattern 'almost daily', Thomas (1972) said that it was not necessary to have faith, 'in studies which could be badly out of date in terms of children's actual usage and linguistic need'
A.F. Watts (1948) studied fifteen children between the ages of six and a half and eleven and half years. These children were persuaded to write diaries on topics of personal interest over a period of about eight weeks spending about 20 minutes daily on such writing. At the end of the eight week period these children produced about one lakh words of which 4079 were different words including 348 proper names. The average size of vocabulary of an individual child was 986 words, the range being 531 to 1338 words. (Pasricha and Das 1959)

Edward and Gibbon (1964) conducted a survey of the words used by children of 5+, 6+ and 7+ in infant schools. The purpose of their survey was 'to arrive at a vocabulary which will help writer and publishers producing books for the youngest readers.' They also hoped that educationists would find it of interest and that teachers would find it to be of practical value. They arranged to collect words from the spontaneous writing of children. Accordingly they selected schools where the children had opportunities to write freely and spontaneously. By this children's writing covering almost every aspect of their daily life could be made available for the study. No topic was set by the surveyor. Schools numbering 45, were selected in every varied district of Leicestershire. Children selected for the purpose were of 'varying ability' with 'different social background' and from 'remote one-class
village schools to large infants' schools in industrial and market towns and dormitory schools'. Altogether 2120 children were sampled with break up into different age-groups, i.e. 820, 794 and 506 children of 5+, 6+ and 7+ years respectively. The concerned schools collected and counted the vocabulary and also summarised the findings. Each word, except proper nouns, written by the children was listed. Over and above, the derivative forms differing greatly in appearance from the root word were also listed separately. After omitting words used by less than 1.5% of the children the total number of different words retained for study for each age-group is:

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Number of Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>5+</td>
<td>661</td>
</tr>
<tr>
<td>6+</td>
<td>1031</td>
</tr>
<tr>
<td>7+</td>
<td>1349</td>
</tr>
</tbody>
</table>

One new thing that has been added to the 'word study' by Edward and Gibbon (1964) is the 'Popularity Index' of a word. They have worked out a formula to arrive at such an index. It is, 'the product' of 'percentage of children in the sample using the word' and 'the average use per child in the sample' (i.e.

\[
\text{Popularity Index} = \frac{100 \times n}{s} \times \frac{f}{s}
\]

In the formula, the number of children using the word is represented by 'n', the sample by 's' and the number of times the word is used by 'f'; thus the percentage of children using the word comes as '100n/s' and the average use of word per child comes as 'f/s'. The product of
'100n/s' and 'f/s' is the 'Popularity Index'. The range of popularity index of the words so calculated found out to be from 3,667.4 (of the word, 'and') to .025 (of the word, 'tyre') in the 7+ list they prepared. The investigators claimed that giving 'popularity index' for each word in the word-list(s) facilitated the textbook writers to find out the 'commonest' as well as 'most frequently used' words easily. They have arranged the word-list(s) in two chapters. In the first chapter the words for each age group have been listed in groups of 250. Each such '250-group' is arranged alphabetically. The second chapter contains the words age group-wise and also in order of 'popularity index'. The important findings of the study are: (1) almost all the first 500 words in the list are common to all three age-groups; (2) as the children grow older they gradually add new words; (3) the content of the children's vocabulary stress on food, animals and the exciting scientific wonders they see around them. About the content of the children's vocabulary the authors categorically said, these small extroverts show little interest in the parts of their own bodies but a great deal in the activities around them; they pay little attention to common household article but much to their own playthings and possessions. (Chapter-VII)

The authors name their vocabulary lists as 'Graded Vocabulary'. It is published under the title
WORDS YOUR CHILDREN USE'. It is also known as 'Leicestershire List'. Comparing their 'Graded Vocabulary' lists with those of Gates (1935), Thorndike and Lorge (1944), Rinsland (1945) and Dolch (1959) the authors showed that many words of those lists did not find place in their 'Graded Vocabulary' list. They claimed that their 'graded vocabulary' list had 'lively content' contrasting with those American lists. The following statement of the authors would show that they are fully satisfied with their graded vocabulary list:

We set out to discover the words that children themselves use. This we have done. We hope that Words Your Children use will be of value to all those who are interested in children's language in particular in their reading and writing. (Chapter- VII)

This statement of the authors seems to as bold as 'I went, I saw and I conquered'. But about this list- Leicestershire List - McNally and Murray (1968) commented that in using the spontaneous writings of five and six year olds an inevitable bias towards the culturally superior must arise in the word counts because a considerable number of children will be unable to express themselves adequately in writing at this stage. They also said that the 'first 1000' of the said list do not include words like, 'danger', 'slow' which occur in normal reading where as the 'first 500' words of the list include words like 'diesel', 'tiger' and 'bunglow'.

The most recent study on writing vocabulary of elementary school children available to the present
In counting and tabulating the words Thomas (1972) used the lexical units employed by Rinsland (1945). In comparing the frequency of word usage of his study with that of Rinsland, Thomas found that the first 50 words of his own list comprise 48% of his total word count which corresponds to 40% of Rinsland’s study. On the other hand the first 100 words of Thomas’s list while comprising 58% of his total word count is a little less than the 60% of Rinsland’s. Again, the corresponding figures in case of the first 2000 words are 89% and 97% respectively. This comparison, however, does not correspond to grade reference. It was further found that there was only 80% correspondence between the 100 words being used in Alberta then and the 100 words used by American students in the year 1937. Besides compiling a non-graded list of 2000 most frequently used words, which meant the 'basic-core-vocabulary', Thomas wanted to determine the 100 most frequently used words in each grade. But due to heavy overlap of words from one grade to the next he found a cumulative list of 170 words for all the six grades. After keeping the grade one list of 100 - the 100 most common words in order of frequency - in its entirety, Thomas listed all the new words of the
it would be interesting to determine whether there is an identifiable writing vocabulary which is distinctive to particular regions or whether there is a common core which is largely Canadian. (p.248)

Shanker (1978) mentioned the following earlier foreign studies on written vocabulary of children:
(The author has, however, not mentioned the dates of these studies.)

Fitzerald tabulated approximately 6,95,000 running words from letters written by elementary school children outside schools.

Ashbaugh tabulated about 200,000 words from letters written by junior and senior high school students.

Mckee compiled a list of more than a million running words both from children's letters and from their compositions or writings done in the school.

Dillion and Wheeler each tabulated approximately 2,89,000 running words from children's writings on various themes.
Smith tabulated approximately 2,40,000 running words taken from the writings of children in various subjects of the school curriculum.

2.2.3. Spoken Vocabulary:

With the literature available in the hand of present investigator the earliest word list prepared from the counting of 'spoken vocabulary' of children is that of E. Horn in the year 1925 (MacGinitie, 1970)*. As per the statement of MacGinitie (1970), this count of Horn was an extremely valuable one except that it was too old to be accepted.

The next study in the line is Vernon's (1940) study of vocabulary of Scottish children. It was, 'A Preliminary Investigation of Vocabulary of Scottish Children Entering School'. This study was undertaken at the instance of decision of the Reading Committee of the Scottish Council for Research in Education in 1940 to collect a representative vocabulary of five year old Scottish children. The sample was the speech vocabulary of two hundred children all aged 4½ to 5½.

years. This sample included 100 boys and 100 girls, who were fairly well distributed geographically, of course excluding the Highlands and Islands. The objectives of the study were to compile a 'Scottish Word List' and with the help of such list either to assess the existing infant readers or to prepare new ones. Collection of speech vocabulary was mostly done by teachers during play time or other odd moments at school. The recording of speech was done through a record form which was prepared and distributed (along with instructions) by the author of the study. The following page shows the record form and instructions prepared by Vernon. This count of Vernon involved approximately 28,000 words with an average of 140 different words per child. The count yielded 2948 different words of which 1038 occurred only once or used by one child only. Excluding these words the author made a list of 1910 words which were used by at least two children. This list of 1910 words, according to the author, is List I. He also prepared another list of 491 words used by 15 children and more. This second list was named as List II. The words of List-II were classified according to parts of speech. The following is the distribution of number of words against different parts of speech.
STUDIES IN READING
THE SCOTTISH COUNCIL FOR RESEARCH
IN EDUCATION
Reading Committee

VOCABULARY OF CHILD ENTERING SCHOOL—
RECORD SHEET

Boy or girl......................................... Parental occupation...........................
Name ............................................... Christian name..................................
Education Authority ..................................................
School ............................................. Burgh or Parish ..............................
Date of birth..................................................................
Dates of observation ........................................................................
Hour of day........................................................................
Circumstances in which observation was made, e.g. home, travelling, school.

INSTRUCTIONS

It is a generally accepted principle that the content of the first reading book should consist to a large extent, if not wholly, of terms used by pupils entering school, the meanings of which are known to them. The Reading Committee have consequently decided to compile a speech vocabulary of pupils aged 4½ to 5½ years, and would welcome your cooperation. The procedure to be adopted is detailed in the instructions below. It is not considered advisable to restrict records to those taken in school; if possible, observations should be made in as extensive a variety of circumstances as possible. Your records, when completed, should be returned to—

The Convener, Reading Committee,
Scottish Council for Research in Education,
46 Moray Place, Edinburgh 3.

1. Listen to child's conversation with other children, or with adults, or engage the child in conversation, or induce him to talk, either by your own conversation or through the use of pictures or objects.

VOCABULARY OF SCOTTISH CHILDREN

2. Do not restrict your stimulus to classroom situations. Remember that the child may know many words relating to current events: the street, cinema, wireless, home (eating, housework, relatives), church, occupations, games, toys, weather, holidays, etc. (Omit rhymes, jingles and other learned repetition.)

3. Write down the words actually used by child, including ejaculations, slang, contractions, single words and sentences.

4. Conversations should be recorded during ten-minute periods; from two to ten such periods should be spent with each child according to the time available to the observer.

5. It would facilitate the investigation if the observer would also arrange in alphabetic order the words of each child studied. It is, however, essential that the original records of the child's vocabulary as spoken should be forwarded.

6. In arranging words in alphabetic order, tabulate as separate words inflections of verbs and contractions such as I'll, I'd, We'll, Don't. Plurals of nouns and verbs which are not regular—for example, Mouse, Mice; War, Were—should also be included as separate words. Possessive nouns should not be counted separately.
as calculated from Vernon's List -II.*

<table>
<thead>
<tr>
<th>PARTS OF SPEECH</th>
<th>NO OF WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>231</td>
</tr>
<tr>
<td>Verb</td>
<td>117</td>
</tr>
<tr>
<td>Adjective</td>
<td>46</td>
</tr>
<tr>
<td>Pronoun</td>
<td>36</td>
</tr>
<tr>
<td>Adverb</td>
<td>32</td>
</tr>
<tr>
<td>Preposition</td>
<td>19</td>
</tr>
<tr>
<td>Conjunction</td>
<td>6</td>
</tr>
<tr>
<td>Interjection</td>
<td>4</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>491</strong></td>
</tr>
</tbody>
</table>

These 491 words in List II, according to Vernon, were the most common words used by Scottish children then.

The author was confident enough that within limits, the size of his 'sample' of children did not seem to affect the relative distribution of word frequencies. He further stated that had the sample been two or more times as large, some 30% of the words might have been used by a single child and that some 10 to 15% of the

*Calculated by the present investigator from Vernon's List II which appeared in STUDIES IN READING, Vol. I, University of London Press Ltd., 1949, pp.119-123.
of the words by 10% or more children.*

The author was quite conscious of the reliability of his List II - the 491 commonest words. He verified the reliability in two ways. After calculating the probable error (PE) statistically and after comparing 476 commonest words of first 100 children with List II he concluded the following with certainty.

- his list contains 450 out of 500 words most commonly occurring among five year old Scottish children (Subject to the conditions under which the records were made);

- there is, however, no guarantee about the words with frequencies below 25; and

- some 40 to 50 words would alter in another sample and it is not possible to identify those.

This study of Vernon was subsequently used to verify that viability of INFANT READERS of which word-count and analysis were done by him (as has been discussed under 'Reading Vocabulary' in this chapter just before).

Following Vernon's approach, the other main study on speech vocabulary of 5 to 6½ year old children was

* The corresponding figures being about 35% and 12.66% respectively in the result of the present sample.
conducted by Burroughs (1957). Burroughs had the objective to produce a vocabulary of spoken English which could be taken as representative of the vocabulary of children who were about to learn to read. In other words the main aim of this study was to construct a word list for use in 'making materials for early reading experience'. This study was conducted on behalf of the University of Birmingham Institute of Education and therefore is named as 'Birmingham Study'. A sample of 330 children with an equal division of numbers of boys and girls, i.e. 165 boys and 165 girls was drawn for the study. This sample of both boys and girls was further distributed in three 'six monthly age groupings', e.g. 5-5½ year, 5½-6 year and 6-6½ year. So the number in each group of sample was 55 children. The children were drawn from a number of Infant Schools, both urban and rural areas. No more than 4 children were selected from any single school. The data were collected through student-teachers. These student-teachers were also selected randomly. Again, each student-teacher had to select randomly one boy and one girl from the whole group of children who fell within the age limit in the concerned practising school of the student-teacher. The student-teacher had to devote eleven ten-minute periods for each child to record the words of the particular child. The first one of the
eleven ten-minute periods was to be devoted for establishing rapport with the child. These eleven ten-minute periods were to be taken with a span of two weeks time. Burroughs's decision to sample 'conversation in school' was mainly on two grounds. One is the 'practical consideration' and the other was the support from one, Davis, who, intended 'to capitalise the common factors of school experience....thus gaining better standardised working conditions'. About the time span for collection of data, Burroughs did capitalise Loomis's experience that, 'after two hours' observation there was no increase of efficacy. About the correctness of his time decision Burroughs says:

The pattern of the present study allowed eleven ten-minute periods of recording for each child, which is rather longer than most studies of this type though a little short of two hours. (p.8)

But in the long run of data collection Burroughs felt that by the end of the tenth session the yield was beginning to be small. To facilitate conversation or just to talk something around, pictures and scrapbooks, though not specific, were used when necessary. Each fresh word used by the child was to be noted down. For the purpose of counting and tabulation of words or analysis of data the following principles are followed by Burroughs (1957)
(1) Plurals of nouns formed regularly have not been separately listed.

(2) Adjectives formed simply from nouns have been listed under the root word.

(3) Comparative and superlative forms have not been separately listed.

(4) Participles formed in 'ing' and 'ed' have been listed under the root word.

(5) Negatives of the form 'isn't' have not been separately listed.

(6) Words formed from others by internal inflection have been shown separately.

(7) Purely children's expression such as 'ch-ch' have not been listed.

(8) Proper nouns, with a few exception such as 'England' have not been listed.

(9) Certain compound words have been listed under simpler forms and other words grouped when the meaning was the same.

When he listed the words with the above conditions the total number came to be 90,040 with an average of 273 words per child, the range being 56 to 578 per child. The total number of different words are arranged in two lists. List-1 contains 1909 words used by 5 or more children and list-2 shows the remaining 1595 words used by 4 or fewer children. It was the intention of the investigator to list out first 2000 words in list-1, but due to difficulty in breaking through the middle of a particular frequency group he had to take 1909 words in list-1. List-1 has further been divided into four groups.
of 500 words. Again, due to the same difficulty the actual number of words are 501, 499, 502, and 407 in the 1st, 500, 2nd, 500, 3rd 500 and 4th 500 groups respectively.

In order to be confirmed about the reliability of his word list Burroughs had to repeat the experiment at Weymouth keeping in view the geographical differences in mind. The result of the repeated experiment is the 'Weymouth List'. Comparison of both the lists (Birmingham and Weymouth) revealed approximately 55-60 doubtful words in the first 500 of the Birmingham list. Burroughs attributed this doubtfulness to the 'unreliability of the experimental procedure and geographical variation'. This, however, revealed that about 90% of words of Birmingham list are secured.

The Birmingham list was also compared with five other lists out of which one is the Scottish list and the four others are American lists (Thorndike– the revised one, Dolch, Rinsland and Gates). The first 1000 words of each of the four lists were taken for comparison. For the purpose of comparison Burroughs prepared nine lists more (List-3 to List-11). The comparison with the Scottish list offered ground for reasonable confidence in the 'new' Birmingham list. On comparison with the four American lists,
Burroughs concluded, 'not all the American lists are entirely suitable; this applies particularly to the Thorndike list and least to the Rinsland list'. Burroughs's view about comparison with other lists is, however, not positive, because he says:

......though for practical purposes comparison with other lists may be useful, we are comparing vocabularies which are not strictly comparable and that, in fact, each list must be considered in terms of its usefulness for particular purposes. (p.8).

In the final analysis of his data Burroughs arrived at four conclusions. They are:

(1) Comparing with Dolch's list of 220 basic sight words which contains no nouns. Burroughs concluded that inspite of greater variability of nouns from list to list their constancy is too large to suppress them in a basic sight vocabulary.

(2) Some interesting things in the groupings and relative frequencies of 'sets of words' were found about numbers, the days of the week, colours and foodstuffs. They are: (a) after the number 'one' each odd number shows a slightly lower frequency than the even number which immediately follows it, (b) Saturday and Sunday are the highlights of the child's week, and (c) the belief that red is the dominant colour for children of this age is fully substantiated.

(3) The size of 'core vocabulary' - 'one used by 50% or more of the children' - is found to be 99 words of which 47 are nouns or pronouns and 52 are other parts of speech in the Birmingham list. This number happened to be a third on average, i.e. from an average
list of 273 words per child. This 'superimposed' that each child has a larger fringe vocabulary which he shares with some half (or less) of his colleagues.

(4) About 'sex and age difference' nothing categorically was found except that there are very few sex difference at this age and also that few words which are uncommon at 5-5½ become common at 6-6½.

MacGinitie (1970) mentioned about a similar study which was conducted by H.A. Murphy in the year 1957. The study was on 'The Spontaneous Speaking Vocabulary of Children in Primary Grade'. About this study MacGinitie only mentioned that Murphy analysed a large sample of words that were elicited from the children by a variety of techniques in classroom settings. MacGinitie (1970) also mentioned another study of D. Howes (1966), 'A Word Count of Spoken English'. This count involves 2,50,000 running words. These words, however, relate to adults' usage during interview.

2.2.4. Other Studies on Children's Vocabulary:

Some of the investigators of children's vocabulary have prepared vocabulary lists for the purpose of early readings basing upon the results of earlier studies or previous word-list(s). Some of them collected and analysed
their own data in addition to the analysis and selection of vocabulary from previous list(s) in order to prepare a more viable vocabulary list. Such studies either relate exclusively to any one of vocabularies ('speech/spoken vocabulary', 'writing vocabulary', 'reading/printed vocabulary') or a mixture of these. All the same all of them have a similar purpose. It is, therefore, planned to discuss them separately.

One such basic list, originating in America, was presented in the year 1928 by the 'Child Study Committee of the International Kindergarten Union'. Summarising, it is said, several earlier studies the Committee prepared a list of 2596 words considered to be the most frequently found in the speeches of children entering the first grade. This study, therefore, exclusively deals with 'Speech/spoken vocabulary'. This study is known as 'Kindergarten Union List'. The Kindergarten Union List has been used by the later researchers as would be found from the discussion that would follow in this section (McNally & Murray 1968).

Edgar Dale (1931) prepared a list of 755 words resulting from a comparison of the first 1000 words of Thorndike's list with the Kindergarten Union List. (Chall, 1958 and McNally & Murray 1968). This seems
to have been published under the title, 'A Comparison of Two Word Lists' in Educational Research Bulletin, 1931, 10, 484-489 (MacGinitie, 1970). As stated by MacGinitie Dale's original list of 755 words was revised to 769 words, by Stone, in the year 1956.

A.I. Gates (1937) produced a list of 1811 words by using Thorndike's first 1000 words and the Kindergarten Union List along with his own word-counts of written vocabulary collected from primary school children. It is said that Gates's list served as a 'basic reading vocabulary' for primary grades and that it influenced the vocabulary content of subsequent reading schemes to a considerable extent. The title of Gates's study is, "A Reading Vocabulary for the Primary Grades (rev. and enlarged)". (It is a publication of the Bureau of Publication, Teacher's College, Columbia, 1937). Russel described this list as 'probably the best single source for checking words that should ordinarily appear in the basic reading vocabulary of the primary grades' (McNally & Murray, 1968). Rawat (1971) seems to have quoted the same Gates's list which was published in the year 1935 under the same title from the same source of publication and it contains 'approximately 1500 words based on six previous studies'. From the available information and in the absence of the original studies, it may be
said that Gates originally prepared the list in 1935 with approximately 1500 words and later, in 1937, revised and enlarged the list to 1811 words and that Gates used four more previous lists over and above the two mentioned by McNally & Murray (1968).

E.W. Dolch compiled a list of 1000 words taking the 755 words of Dale's original list (which was the product of the two great lists referred to above) of 1931 and by adding to it his own 'Interview list' of speech vocabulary obtained by confronting children with 'spoken Dale's words', arranged topicwise and by listing the additional words spoken by them in response (Edward and Gibbon, 1964). Later, in 1945, Dolch prepared another list of 220 words popularly known as 'basic sight vocabulary'. This 'basic sight vocabulary' was the product of comparison of the following number of words from the following lists:

(1) Most frequent 510 words of the Kindergarten List (1928)

(2) The first 500 words of the Gates List (Gates, 1926)

(3) 453 words of the list of Wheeler and Howell (1930)

Dolch found that there were 315 words common to all the above three lists out of which 95 words were nouns and
the rest 220 were other words - 'non-nouns'. He, therefore, retained only the 220 'non-nouns' as the 'basic sight vocabulary' meant for 'remedial reading' and omitted the nouns with the feeling that nouns represented the most unreliable part of vocabulary.

Hardie (1948) prepared six lists of 'Minimum Vocabulary' basing upon three basic lists:

(1) Ogden's Basic English;
(2) Rinsland's Basic Vocabulary
(3) Thorndike's 'AA' and 'A' words.

The source of Hardie's idea of 'Minimum Vocabulary' is Thorndike's view that there are certain words which the child must learn as a part of his 'permanent intellectual equipment' which must be represented by a 'Minimum Vocabulary'. Further, according to Hardie, this minimum vocabulary should enable an individual to do two things. First, it should enable him to say anything he wants to say with ease and confidence. Secondly, it should enable him to understand, with occasional use of the dictionary, most of what he hears and reads in the process of communication. Hardie concluded that Ogden's Basic English (minus the 'basic syntactical rules' can help the first objective and Thorndike's 'AA' and 'A' words the second. Hardie took the published information, i.e. the raw data from Rinsland's 'Basic Vocabulary' as his 'rough guide' to
prepare the minimum vocabulary lists. Of the three lists, Hardie used for his purpose, no discussion has yet been made of Ogden's Basic English in the present study. It is because this Basic English is a discovery or an invention of Ogden to provide a 'universal secondary language', and to provide a simplified method of introducing foreigners to ordinary English. But Hardie believed that the 'Basic' makes a permanent contribution to education in English speaking countries by providing a vocabulary which is sufficient for the ordinary purposes of human beings when they communicate with one another. It also gives answer to the difficulty about Thorndike's second rule as to which words are of most practical service. With these strong beliefs Hardie accepted Ogden's Basic English for his use. A point of difficulty which Hardie faced in using the above three lists was that they lack uniformity in 'Meaning for the word 'word'. As for example, while Rinsland listed LIGHT, LIGHTS, LIGHTING, LIGHTED, LIGHTER as different words, Ogden counted only the root meanings as words with elimination of verbs and Thorndike followed a third procedure; 'regular plurals, comparative and superlatives, verb forms in s,d, ed and ing..........are ordinarily counted in under the main word'. This, Hardie observed, is somewhat arbitrary. So he also adopted his own definition
of 'word' - 'only root words', of course 'allowing
all the modification which it is possible for them to
have in ordinary English'. Hardie, therefore, said:

I have considered that a word appears in the
first 2000 words of the Rinsland list in
either grade 1 or grade 2 if ANY form of it so
appears; and I have considered that a words
does not appear in the first 2000 words
occurring in either grade 1 or grade 2 if NO
form it so appears. (p.6)

Adopting the above procedure Hardie prepared the
following six lists:

List-1: Common words in Basic list vrs Rinsland's.
List-2: Common words in Thorndike's 'AA'
words vrs Rinsland's, except the
proper names and adjectives formed
from proper names and also numerical
adjectives.
List-3: Common words in Thorndike's 'A'
words vrs Rinsland's and except as
in List-2.
List-4: Thorndike's 'AA' words not appearing
in Rinsland's.
List-5: Thorndike's 'A' words not appearing
in Rinsland's.
List-6: Basic words not appeared in the
above five lists and also the adje-
tive INTERNATIONAL, and the 50
international nouns and 12 interna-
tional titles recognised by 'Basic'
which have not already appeared in
lists 1-5.

Besides, the 'days of the week' and 'months of the
year' have been inserted in List-1. Although the number
system and numerical adjectives have not been included in any of the above lists, Hardie suggested that these should be regarded as starting in List-1 and continuing through all lists. Hardie has high claims for his lists which are summarised as under:

(1) the Minimum Vocabulary prepared with the available published information, have the best chance of realising it and given these tools the average child will be able to express himself and to understand more easily;

(2) these lists will provide means of assessing the value of school readers;

(3) the first three lists ought certainly to be mastered by grade 2 and the remaining lists as soon as possible thereafter;

(4) the early reading books and the teaching associates with them should concentrate particularly on these lists.

Hardie however warns that the teachers and parents must not attempt to exclude other words and the child should live in as rich a verbal environment as possible so that by the time the child leaves the Primary School, 'his vocabulary should be about four times as large as that contained in six word lists'.

The latest work in this group, as available to the present investigator, is McNally and Murray's 'KEY WORDS TO LITERACY' (1968) which they meant to be a 'basic word list for developing early reading and
writing skills'. This is a very small list of 200 'Key Words' including only 20 nouns. This list has been prepared by a very refined and sophisticated method of selection of words from a number of earlier lists. The investigators were quite choosy in picking up the words from the various lists. In addition to this they also made two different small counts of 500 running words from 'eight news-papers' and from 'eight popular school readers of recent date'. Given below are the main sources of their data:


(2) The most important 500 words from A.I. Gates's list of 1811 words: "A Reading Vocabulary for the Primary Grades", Bureau of Publications, Teachers' College, Columbia University, New York, 1937.


(4) The first 395 words from the list of over 2000 words graded by C.R.Stone: Graded Vocabulary for Primary Reading, Webster Publishing Co., St Louis, 1941.

(5) 201 words, very common in adult reading, from Lorge's magazine count (Readers' Digest etc.) and the Semantic Count which appear over 3000 times per 4½ million running words in either count: Thorndike, E.L. and Lorge, I, Teachers' Word Book of 30,000 Words, Bureau of Publications, Teachers' College, Columbia University, New York, 1944.
They first prepared an initial list of 773 words gathered in the following manner:

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>NO. OF WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Vernon's list of Speech Vocabulary</td>
<td>491</td>
</tr>
<tr>
<td>(2) All words in Gates's first 500 words not listed by Vernon</td>
<td>192</td>
</tr>
<tr>
<td>(3) Vernon's &quot;Words Counts of Infant Readers&quot; (42 words which appeared in all the series but not in Vernon's list of 491 words. Out of these 42 already covered in Gates's list above. So, the rest ..........)</td>
<td>16</td>
</tr>
<tr>
<td>(4) All new words from Dolch's 220 words (i.e. not already covered by the above sources)</td>
<td>2</td>
</tr>
<tr>
<td>(5) New Words from Stone's 283 words (not already covered by the above sources)</td>
<td>33</td>
</tr>
<tr>
<td>(6) New Words from Kyte's 100 words (not already covered by the above sources)</td>
<td>1</td>
</tr>
<tr>
<td>(7) New Words from the Thorndike-Lorge list of 201 words (not already covered by Gates's list)</td>
<td>26</td>
</tr>
</tbody>
</table>
(8) New words occurring in
News-Paper count (occurred at least in three newspapers out of eight). 6

(9) a) New words occurring in the
Popular School Readers count
(occurred at least in three readers out of eight) 4

b) New words occurring in
earliest books of four
Junior School Stage;
(occurred at least in two
of the series). 2

TOTAL 773

These 773 words were arranged alphabetically and illustrated in a table showing the relative importance of each and every word under its respective source(s). Keeping in view the relative importance of every word from every source, a 'grade marking' or 'ranking' was assigned to it like "A", "B", "C", "D" plus and "O". Time and again the investigators checked each and every word's grading and thus marked, placed it in their 'Key Words' list. Finally they chose 200 key words which appeared to them to be of fundamental importance in the development of reading skill. Later, they also enlarged the list upto 300 words although they categorically said:

After the 200 word stage diminishing returns become more and more evident so that there would seem to be little
to gain in listing specific words much beyond the 200 mark. As a first list we have, therefore, limited the selection to 200 thus ensuring that we are presenting words very commonly used in reading and the ready recognition of which is essential for the development of basic reading skill. (McNally & Murray, 1968, p.8)

Taking the instances of Watts and Dolch, the investigators are of opinion that an individual's approximate 'sight vocabulary' is 20,000. But all these words are not of equal value. Only 12 words account for about 25% of the running words used in everyday reading and writing and so also that only 100 words account, on the average, for 50% or more of everyday reading material. The investigators illustrated such of their findings in the following table:

<table>
<thead>
<tr>
<th>Parts of Word List used to assess samples</th>
<th>Percentage of samples covered by parts of the Word list.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Juvenile Reading</td>
</tr>
<tr>
<td>First 12 words</td>
<td>25.2</td>
</tr>
<tr>
<td>First 32 words</td>
<td>38.3</td>
</tr>
<tr>
<td>First 100 words</td>
<td>56.9</td>
</tr>
<tr>
<td>First 100 words plus 20 basic nouns</td>
<td>62.3</td>
</tr>
<tr>
<td>First 150 words including 20 nouns</td>
<td>67.0</td>
</tr>
<tr>
<td>First 200 words - including 20 nouns (The Key Words)</td>
<td>70.8</td>
</tr>
<tr>
<td>The &quot;Key Word List&quot; plus 50 additional nouns</td>
<td>73.3</td>
</tr>
<tr>
<td>&quot;Key Word List&quot; plus 50 nouns (above) plus 50 &quot;reserve&quot; words (i.e. the Enlarged Word List of 300 words)</td>
<td>76.1</td>
</tr>
</tbody>
</table>

(McNally & Murray, 1968, p.37)
From the above table it is clear that the first 100 words of the "Key Words List" are all non-nouns and that the "Key Words List" of 200 contains only 20 'basic nouns'. It is also clear that the "Key Words List" covers 70.8% of words in the entire 'Juvenile Reading' which itself speaks of the soundness of the list. The investigators started the work with the genuine feeling that there are difficulties in compiling a general word list suitable for early reading. They examined the competing claims from vocabulary studies based on 'reading content', on 'spoken word' and on 'free writings' of children. They concluded that all 'one-way methods' are unreliable. They discussed the pros and cons of all such methods. Their data encompassed a task which is a difficult one.

2.3. Studies in Indian Languages

2.3.1 Hindi

Koenig is a pioneer in the field of vocabulary studies in India, especially in the context of primary school pupils. As early as 1931, Koenig attempted an analysis of Hindi words used in juvenile literature including Hindi primers and readers. He followed Thorndike's technique and counted a million running
words from 153 sources. His count resulted in 12,500 different words which, according to him, pupils in primary classes were expected to absorb. Koenig, however, finally selected only 4000 different words out of the above 12,500 different words and held them as 'Basic Vocabulary' in Hindi for standards I, II, III and IV.

Naseem (1978) claims that one Tarini Charan Mittra happened to be the pioneer to prepare a Hindi Glossary of words of a 'textbook' from the educational point of view more than a century before (near about 1823) Koenig's attempt in this field of work. Naseem's source further reveals that "PREMSAGAR" was the above 'textbook' which was being studied at 'Fort William College' by 'Civilian Ministry personnel'. Naseem's source does not, however, speak whether PREMSAGAR was being taught to Hindi-speakers or Non-Hindi-speakers. With the information available from Naseem's source this much can be said that Mittra's PREMSAGAR was a Hindi Glossary of textbook for adults and might be meant broadly for non-Hindi-speakers. Mittra's work does not, therefore, come within the purview of the present study. Naseem however acknowledged Koenig's work as being 'on broad and scientific lines'.
For counting purpose Koenig took the simplest form of the word as a different word which includes all its inflections caused by number, gender, case, tense etc. But in case any such inflection modifies the root itself it is taken as a separate word. He also counted feminine nouns as separate words. Koenig, however, excluded all proper-nouns and all poetry. The first 4000 words selected were alphabetically arranged in four groups of 1000 words, each in order of frequency from higher to lower, i.e. the first one thousand words are of the highest frequency and then the next and the next and the next. Such arrangement was thought of for the benefit of teachers and authors of textbooks for primary grades (Naseem, 1978).

Koenig's work is published under the title TEACHERS' AND AUTHORS' LIST OF 4000 IMPORTANT HINDI WORDS by Mission Press, JABBOLPORE (1931). (This publication could not be available to this investigator in spite of all efforts made to collect it). Besides counting the word frequency, Koenig compared the Hindi Primers and readers with those of American Origin of the time. The unhappy fact, the comparison revealed, was that Hindi books used new words much more frequently, i.e. one new word after every second to eleventh word. This ultimately speaks of a greater 'vocabulary burden' on
the primary school pupils of India than on American pupils. By vocabulary burden is meant the proportion of new words to the total number of running words in the book. According to Koenig the satisfactory vocabulary burden should be 1:15. But Koenig's count revealed 1:11.8 and 1:20 respectively in respect of Hindi and American primers. Koenig also observed that most modern readers aim at 1:20 as the vocabulary burden. From his findings Koenig concluded that such heavy vocabulary burden compels the Indian primary school pupils to end primary education without comprehension of the matter supposed to be comprehended.

Critics of 'word-frequency-count' do criticise Koenig also. For example Shanker (1978) referring to Koenig's study concludes that the study suffers from the same drawbacks of 'frequency' studies since it is based on frequency counts. Still, perhaps, it would be no exaggeration in naming Koenig as the 'Thorndike of India' as a word-counter for the purpose of primary school textbooks.

(2) The Central Institute of Education, Delhi, took up a study of the Hindi vocabulary of children the report of which is published as A STUDY OF THE VOCABULARY OF HINDI PRIMERS (CIE Publication No.10, 1956). The study was conducted by A.N. Basu, M.K. Malhotra and Bir Bahadur. The main aim of this study
was to find out whether the textbooks were written in accordance with psychological and educational principles. In order to find out this the investigators fixed up some criteria such as:

- the vocabulary load at each grade level;
- the difficulty value of words;
- suitability of words according to age needs in children's social environment;
- intellectual difficulties of children in reading criteria.

This study involved eight textbooks supposed to be representative of various types used in Delhi. Word count of all the books was held yielding 10,984 running words and 2311 different words. Out of these different words 75 had the frequency of 20 or more and 924 words occurred only once. Only 33 words were found to be common words in all the eight textbooks. From their findings the investigators observed the lack of observance of scientific laws of learning in these books; textbooks authors were not guided by any basic or core vocabulary which would indicate the average vocabulary of the grade; the vocabulary was quite arbitrary, etcetra. In suggestions the investigators mentioned that 'reading vocabulary', being 'the very flesh and blood' of teaching material, should be constructed on scientific and educational foundations and that grading and
standardisation of Hindi Vocabulary were of 'immediate and essential needs'. Quoting this study Pasricha and Das (1959) say, 'The very first study of the Hindi vocabulary of the children was designed at the Central Institute of Education, Delhi'. (SOURCE: Pasricha and Das, 1959)

(3) It is reported that AN INVESTIGATION INTO THE REPRODUCTION VOCABULARY IN HINDI was conducted by J.P. Tiwari in 1957 in Madhya Pradesh. This was a study at M.Ed. level. This study involved 'written vocabulary' of children of the fifth class, who are usually of 10+ age, numbering 330. The children were asked to write three compositions each, one descriptive, another narrative and the third a reflective on three different days. The time allowed for writing for each composition was 40 minutes. Finally the compositions of 200 pupils were taken for count. Tiwari followed Koenig's procedure of counting and classifying words. (SOURCE: Shanker, 1978)

(4) R.C. Rukmini is reported to have conducted a study of children's vocabulary in Hindi as a Ph.D. study under Rajsthan University in the year 1960. This was a survey of the Hindi vocabulary of 7+ age group children reading in the second grade. The study also included assessment of vocabulary used in the nationalised textbooks for grade 2 of Rajsthan state. To collect spoken vocabulary a sample consisting of 20 children of 7+ and 10 children
of 8+, reading in grades 2 and 3 respectively were drawn from 5 schools in and around Udaipur. Taking 8+ children in the sample has been explained as finding out 'words falling within the appreciation level of 7+ children'. Each sample child was observed for 2 hours with 5 time spans such as: first 15 minutes, second, third and fourth 30 minutes each and last 15 minutes. The first 15 minute span was devoted for rapport establishment with the child. The three 30 minute spans were devoted to conversation with the child by presenting the child pictures of 'home situation', 'school situation' and 'play situation' one after the other, respectively. In the last 15 minutes the child was required to narrate a story. Words spoken by the child were noted down. The different words so collected were counted to be 1232 and 1647 from 7+ and 8+ children respectively. The 1232 words of 7+ children were classified into four frequency groups, such as:

<table>
<thead>
<tr>
<th>FREQUENCY GROUP (In percentage)</th>
<th>NO. OF WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>75-100</td>
<td>65</td>
</tr>
<tr>
<td>50- 74</td>
<td>123</td>
</tr>
<tr>
<td>25-49</td>
<td>424</td>
</tr>
<tr>
<td>1- 24</td>
<td>620</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>1232</strong></td>
</tr>
</tbody>
</table>

These 1232 words comprised 771 nouns, 241 verbs, 112 adjectives, 24 pronouns and 84 other words. Out of 1647
words of 8+ children the common words used by both 7+ and 8+ children were excluded. Thus the balance number of words, exclusively used by 8+ children came to be 430 words. From these 430 words again the words used by 1% of 8+ children were taken out. So, 208 words of 8+ children were finally retained which were supposed to be within the appreciation level of 7+ children and were considered suitable for inclusion in the textbook for Grade 2. The analysis of textbooks of Grade 2, on the other hand, yielded 821 different words including 424 nouns, 149 verbs, 108 adjectives, 27 pronouns and 113 other words. The important findings of the study are: (i) the spoken vocabulary of children are much more than the vocabulary used in the nationalised textbooks for the grade; (ii) 436 words were commonly used by the children of both the groups and in the textbooks for the grade; and (iii) in both, spoken vocabulary of children and the textbook vocabulary, the nouns dominated. So, the investigator suggested the increase of vocabulary in the textbook of grade 2 and words for the textbook of the grade should be selected from the Basic Word List prepared through this study. (SOURCE : Adaval, 1968).

(5) The National Council of Educational Research and Training (NCERT), New Delhi in the Department of Social Sciences and Humanities conducted a Hindi vocabulary study to list out the most useful Hindi words (Hindi ke adhik upyog shabdon) in the year 1971 for use in the textbooks
for pre-primary, primary and middle grades in schools. No report of the study is available. The present investigator, however, personally gathered a little information about the study as well as a mimeographed copy of an alphabetical word list meant for primary grades III, IV and V from the department. What has been gathered is that the study was originally designed to prepare 4 lists as detailed below:

1. sisû súcî (list meant for pre-primary class)
2. bâl súcî (ka) (list meant for primary class I, II)
3. bâl súcî (kha) (list meant for classes III, IV, & V)
4. kišor súcî (list meant for middle classes)

By now, List-3 is out in mimeograph form (the list collected by the present investigator). This list contains about 3200 words in alphabetical order with 'parts of speech' indicated against each word. About the procedure of preparation of such list, it is said that words were first picked up from dictionary by groups of experienced teachers of different levels. Thus tentative lists of different levels were made ready in this manner. These tentative lists were checked by different groups of teachers of different levels. Necessary addition, alteration, deletion were made as a result of this check. After this, the lists were re-checked by professors of Education. In
this way the final list was approved for the appropriate grade levels. (SOURCE: NCERT, 1971).

(6) Sharma (1972) conducted a study to find out 'Basic Hindi Vocabulary' of children of class IV. This is a study for the Ph.D. degree in Education under Kurukshetra University. A stratified representative sample of 1400 children of class IV was drawn up for the study. The investigator collected words supposed to be usable by class IV children. The sources of collection of words were textbooks and other children's literatures. A tentative list of such words was made. After some sort of screening the author retained 1627 words which were alphabetically arranged and divided into 12 check lists. These check-lists were administered to the sample 1400 children to decide the difficulty value of each word. Such value was determined by the percentage of correct responses of children. The words which elicited correct responses from 30% or less children were taken as 'difficult words'. Similarly those words which elicited correct responses from more than 70% of children were fixed as 'easy words'. The remaining words in between, i.e. correctly responded by 31% to 70% of children were fixed as 'suitable and usable' words for class IV children. The 'difficult', 'suitable', 'easy' words so worked out come to be 146 and 1250 and 231 words respectively. The author,
however, prepared a glossary of all the 1627 words showing the difficulty value of each word and suggested the list of words to be the 'Basic Hindi Vocabulary' for use in the textbooks for class IV in the State of Haryana. (SOURCE: Sharma, 1972 and Adaval, 1968)

(7,8 and 9*) Three other Ph.D. level studies similar to that of Sharma (1972) were conducted under the Kurukshetra University by Kalra (1977), Naseem (1978) and Radhakishan (1979) on 'Basic Hindi Vocabulary' pertaining to 'Third Class (usually of 8+)', 'V-Class' and 'Classes I and II' respectively. All the studies were limited to the Haryana State. The method and procedure of these studies are the same as that of Sharma (1972). All of them prepared glossaries of 'Basic Hindi Vocabulary' for use in the textbooks of the respective classes in Haryana State. The numbers of 'suitable' words in the glossaries of the three studies are 1632, 2025 and 1465 respectively. (SOURCES: Kalra, 1977, Adaval, 1978, Naseem, 1978 & Radhakisan, 1979)

(10 and 11*) Two other studies on 'reading vocabulary' are reported to be the contribution of Madhya Pradesh Government. One is 'Word Counting' of Primary schools readers of classes II, III and IV involving 61,471 running words and 4554 different words. The other is a contribution
of the Government Central Pedagogical Institute, Allahabad. This is a work of sorting out different words used in Hindi books from class I to class VIII in Basic Schools. Eight mimeographed lists of different words, used in different class-textbooks were prepared containing 146,337,249,252,718,497, 561, and 341 different words respectively. The years of such studies are not found in the source.

(SOURCE: Shanker, 1978)

(12.) Rostogi (1982)'s study on 'Hindi Vocabulary' pertaining to primary grades education seems to be the latest one in the line. This study has been conducted on behalf of the NCERT, New Delhi. This is a very big project of a comprehensive study on 'spoken', 'written' and 'reading' vocabularies of children of classes I to V.

This study was conducted by a band of researchers, drawn up from the Linguistics and Hindi Department of Delhi University, Central Hindi Directorate, Delhi, Central Institute of Hindi, Agra, Directorate of Education and the Education Department of M.C.D and N.D.M.C. The study took five years. The report of the study, a mimeographed bound book of half-fulscape size paper containing 138 pages (84 pages of main report and 54 pages containing 15 Appendices), written in Hindi language, is prepared under

*M.C.D. - Municipal Corporation of Delhi
N.D.M.C. - New Delhi Municipal Committee
the authority of NCERT, New Delhi. The title of this book is 'Hindi sabda bhandar ka sankalan tatha bislesan' (kaksya ek se panc), i.e. Compilation and Linguistic Analysis of Hindi Vocabulary (classes One to Five).

The study covered 22 Hindi dialects of 7 Hindi speaking States (Uttar Pradesh, Madhya Pradesh, Rajsthan, Bihar, Haryana, Delhi and Chandigarh). The samples were drawn up from 96 schools under 24 dialect-centres (3 centres of 'KHARI BOLI' and 21 centres of other dialects) on fifty-fifty bases of urban and rural areas as well as boys' and girls' schools. So, out of 96 schools 48 were from urban area of which 24 were boys' schools and 24 were girls' schools and the same was the case with rural areas. The sample consisted of 1920 school children (960 boys and 960 girls) reading in classes One to Five at the rate of 4 children from each class of every school. Out of the 4 children of a class one was 'advanced', two 'mediocre' and one 'poor'. The selection of such children was based upon school-examination achievement as well as teachers' opinion. On the whole from each school 20 (5x4) children and a total of 1920 (5x4x96) children or from each class 384 (4x96) and a total of 1920 (4x96x5) children were drawn up as the sample.

Collection of vocabulary was done through the selected teachers of the concerned schools who were given training for the purpose. The following is the brief and broad
account of procedure of collection of different types of vocabulary, such as: spoken, written and reading.

**Spoken**: The selected children of all the five classes were individually asked some questions about 'home', 'neighbours', 'school' and 'society' by the concerned teacher. The children's spoken responses were noted down by the teachers.

**Written**: For the purpose of written vocabulary the children of class One were exempted because of the fact that they were not expected to express in writing. So, the children of classes Two to Five were asked to write compositions on four different topics. Classwise topics were centrally set. More than four topics were set for each class and those topics were other than the topics on which the children of the concerned class had already written compositions in their classes. Each child of every class of all the schools wrote on four topics as per his/her choice. Thus totally 1536 (4x4x96) written compositions of the children were collected for written vocabulary.

**Reading**: The reading vocabulary was collected from all the textbooks on 'language', 'social science', 'science', 'arithmetic' and 'homescience' used in classes One to Five in all the Hindi-speaking States referred to
analysed structurally, grammatically and semantically.
The data processing was done through computer.

It is reported that altogether 15,197 words were compiled for the consumption of the teaching learning situation in primary education in the concerned Hindi-speaking States. Moreover, it is hoped that the data presented in this study-report would be helpful in taking up several types of pedagogic, linguistic, psycholinguistic and sociological research projects.

The study seems to be a very much sophisticated one and is capable of providing immense guidance to similar researches in the other Modern Indian Languages of the country.

For the purpose of comparison of the finding on 'reading vocabulary' of the present investigator's study the data of Hindi textbooks presented in Rostogi's study, relating to classes One to Three, excluding the subject, 'home-science', are reproduced below:
<table>
<thead>
<tr>
<th>CLASS</th>
<th>lang. no. of words</th>
<th>mean- 'f'</th>
<th>Sc no. of words</th>
<th>mean- 'f'</th>
<th>s.s. no. of words</th>
<th>mean- 'f'</th>
<th>arith. no. of words</th>
<th>mean- 'f'</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>1473</td>
<td>6.46</td>
<td>440</td>
<td>3.95</td>
<td>31</td>
<td>0.55</td>
<td>346</td>
<td>5.77</td>
</tr>
<tr>
<td>Two</td>
<td>1892</td>
<td>5.45</td>
<td>643</td>
<td>3.05</td>
<td>788</td>
<td>3.97</td>
<td>819</td>
<td>7.31</td>
</tr>
<tr>
<td>Three</td>
<td>1732</td>
<td>9.74</td>
<td>995</td>
<td>6.75</td>
<td>720</td>
<td>7.01</td>
<td>569</td>
<td>17.75</td>
</tr>
</tbody>
</table>

Note: lang. = language, sc = science, s.s. = social science, arith. = arithmetic and 'f' = frequency.

2.3.2. URDU

One study of Ahmed (1973), 'URDU WORD COUNT - A Study of Urdu word values statistically determined from extensive word count' is available to the present investigator. Although Ahmed's study is a 'word count' of a 'general' nature, the following reasons tempted the present investigator to include the same in the review here.

The author has full confidence in 'word count' which, in his view, would be of immense help in (a) the teaching of the language to the young, (b) the teaching of literacy to adult speakers of the language and (c) the teaching of the language to speakers of the other languages.

The corpus of data of his study includes a cross section of the whole Urdu literature ranging from middle of 19th century up to the present day, including children's
literature, spoken language and language of the movies. Of course scientific and technical writings as well as literature produced outside India were not taken into account in the count. The corpus of the count involves 4609 Urdu books yielding 5 lakhs of running words. For the purpose of counting one card for each word was used and different colour cards were used for different categories of the corpus. To find out the different or 'specific' words the 'word-cards' were alphabetically arranged. Different colours were used to write the figures of frequency more than 'one'. The following principles were adopted for the counting of words:

1. proper-nouns, i.e. names of persons and places were not counted;
2. certain compound words were listed as separate word;
3. certain Persian and Arabic prepositions closely knit to suffixing words are listed as separate words;
4. nouns indicating case are counted under nominative singular forms;
5. adjectives are taken under the masculine nominative forms;
6. pronouns, demonstratives and a few adverbs are listed in respective forms;
7. all the verbal forms are shown under the respective infinitives;
and 8. homophones written alike but with different meanings are shown as separate words.
The author very clearly stated that his study was not to draw any conclusion but only to present the basic data in terms of frequency of use of the Urdu words so as to facilitate regular study or studies separately in selected specialised fields. The following classification of first 1000 words with frequency ranging from 50 to 36739 have been presented by the author:

<table>
<thead>
<tr>
<th>CLASSIFICATION OF WORDS</th>
<th>PERCENTAGE OF RUNNING WORDS COVERED</th>
</tr>
</thead>
<tbody>
<tr>
<td>First 10 words in the list</td>
<td>34.64%</td>
</tr>
<tr>
<td>First 25 words in the list</td>
<td>46.68%</td>
</tr>
<tr>
<td>First 100 words in the list</td>
<td>62.89%</td>
</tr>
<tr>
<td>First 500 words in the list</td>
<td>78.16%</td>
</tr>
<tr>
<td>First 1000 words in the list</td>
<td>85.09%</td>
</tr>
</tbody>
</table>

According to the author the first 1012 words are 'indispensable' words in the Urdu language. The next block of words with frequency ranging from 35 to 49 are 'essential'; the next with frequency ranging from 25 to 34 are 'useful' and the next with frequency ranging from 15 to 24 are 'special' words.

2.3.3. GUJRATI

Information about 5 studies on Gujrati vocabulary count, for the purpose of children's education, is available
from the sources of Adaval (1968), Buch (1974), Shanker (1978) and Buch (1979). All these are studies at Ph.D. level under different universities. Only two of these studies are related to the present study because they deal with the vocabulary of primary children. The other three deal with the vocabulary of children beyond primary level, i.e. children of '12+' (R.T. Raval, 1959, Bombay University), '13+' (U.T. Lakadwalla, 1959-1960, Bombay University), 'standard VI' (J.D. Bhal, 1975, Saurastra University). So, the facts about only two studies are reproduced below.

(1.) K.S. Vakil studied the 'Basic Vocabulary of Gujarati children of the age '11+' in the year 1954-55 under Bombay University. Adaval (1968), Buch (1974), and Shanker (1978) - all the three authors have mentioned this study of Vakil. Buch (1974) abstracted the study as it is, with no comment. Adaval (1968) abstracted it with some mild comments on the methodological discrepancies. Shanker (1978) reviewed the study with comments on some 'serious drawbacks' of the study.

The first thing about the study is to decide whether or not this study relates to primary class children, since according to the title of the study it relates to the children of '11+' which normally means students of class VI in this country. According to
Adaval (1968)'s abstract, 'Vakil selected the 11-year-group for his investigation because it was mostly at this age that the pupils completed primary education. Shanker takes it as a serious drawback of the study in the sense that Vakil did not ascertain the ages of children of standard V but assumed it to be 11+ which is not warranted by fact. So, she suggested that the vocabulary furnished by Vakil should, therefore, be for standard V and not for children of 11+.

Vakil wanted to arrive at the 'Basic Vocabulary' of the highest class of primary level by way of finding out the 'recognition' as well as 'reproduction' vocabularies of the children of that grade. For finding out the recognition vocabulary he chose four popular textbooks of standard-V. He counted the words of these books and found out 95,000 running words and 14,200 different words with a frequency range from 1 to 13. Then the different words were classified into four different groups of frequency. From this 362 words of most widely recognised group were selected for preparation of a test which was administered to 15 pupils of three learning groups (advanced, mediocre, poor) of standard VI. But the result of this test was a failure in the sense that only 52 words were found to be recognised by the pupils. He then prepared another test by selecting 1075 words out of the first 5000 of his list. This test was administered
on 460 students who could recognise 854 words. For the purpose of reproduction vocabulary Vakil put 906 students to writing on a selected topic from a given list of topics. Thus he collected 906 compositions which yielded about one lakh running words and about 4000 new/different words. These different words were further analysed into parts of speech, the findings of which are as under:

<table>
<thead>
<tr>
<th>Part of Speech</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>66.70%</td>
</tr>
<tr>
<td>Pronouns</td>
<td>1.10%</td>
</tr>
<tr>
<td>Adjective</td>
<td>15.40%</td>
</tr>
<tr>
<td>Verb</td>
<td>11.27%</td>
</tr>
<tr>
<td>Adverb/Preposition</td>
<td>4.29%</td>
</tr>
<tr>
<td>Participle</td>
<td>1.29%</td>
</tr>
</tbody>
</table>

(2) The other study on Gujarati Vocabulary is that of S.S. Sukla conducted in the year 1976 under M.S. University, Baroda. This is a study of Gujarati Vocabulary of students in standards I to V in the age-group 6 to 11, confined to Surat district. This study dealt with 'written', 'spoken' and 'reading' vocabularies of the children of the concerned age-groups. The objectives of the study may be seen to be four, such as:

(i) to collect and study the 'active' vocabulary of the concerned age-group/class children and also
to ascertain the range thereof with 'urban' and 'rural' variables;

(ii) to classify the word list into various parts of speech;

(iii) to compare the reproduction vocabulary of the students with the word-list of the nationalised textbooks (Gujrati Readers) used in standards I-V;

and (iv) to study the impact of foreign languages on Gujrati language.

The sample of the study consisted of 1100 students at the rate of 10 each from classes I and II and 30 each from classes III, IV and V from 5 urban schools and 5 rural schools. These students were asked to write the answers of a given questionnaire. Interview was also used to collect data. The students' writing and interview with them yielded 1,11,869 running words and 1973 new words. The classwise distribution of new words is 434,238,556,408, and 337 in classes I, II, III, IV and V respectively which shows an uneven growth of vocabulary in students. The 1973 words were further analysed into different parts of speech with findings that 1412 were nouns, 33 pronouns, 184 adjectives, 229 verbs, 94 adverbs and 21 participles. It was further found that no word was used by more than 886 students and that the highest frequency of a word was 13,169. This range included only 85 words.*

(SOURCE: Buch, 1979)

*Although in the above abstract of the study one of the objectives was supposed to be a comparison of the reproduction vocabulary of the students with the word lists of the nationalised Gujrati Readers, the abstract is silent on this issue.
Three studies are traced out on Marathi vocabulary of primary school children from the sources of Adaval (1968), Buch (1974) and Shanker (1978). They are noted below:

(1) 'An enquiry into the Marathi Vocabulary Attainments of Children Five to Nine Year Olds in the City of Bombay' was conducted by S.R. Bhat in the year 1939 in his M.Ed. level study. In other words this was a longitudinal survey of vocabulary growth in 5 to 9 year old children. Bhat seems to be one of the earliest Indian investigators in this field. It was, therefore, natural on the part of the investigator to have faced difficulties in methodology and techniques. But the investigator is said to have been concerned with the qualitative aspect of the matter and not with sophisticated statistics. It is said that the investigator used the 'peep hole' method of studying short sample behaviour commonly employed in child as well as comparative psychology. Talking to a limited number of children on a few selected topics the investigator elicited the children's vocabulary from their responses. It is also said that Bhat used a list of 6000 words, collected from Apte's Marathi Dictionary of 26,000 words, while talking to the children with the assumption that 6000 words are the maximum vocabulary of a 10 year old
child. The findings on vocabulary size of different age group children are given below:

<table>
<thead>
<tr>
<th>AGE GROUP</th>
<th>NO. OF WORDS KNOWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>(complete year)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.....</td>
</tr>
<tr>
<td>6</td>
<td>.....</td>
</tr>
<tr>
<td>7</td>
<td>.....</td>
</tr>
<tr>
<td>8</td>
<td>.....</td>
</tr>
<tr>
<td>9</td>
<td>.....</td>
</tr>
</tbody>
</table>

The investigator concluded that the vocabulary attainment of Marathi speaking children of 6-10 year old was less than that of their counterparts in western countries. Bhat's study is held as a milestone in the field of developmental studies with reference to language.


(2) U.R. Soelker studied the reproduction vocabulary in Marathi of the pupils on the completion of primary course. This was his M.Ed. level study in 1946 under Bombay University. It was an attempt to determine the size of reproduction vocabulary in Marathi language of the 5th Class children in school. 270 students wrote 810 compositions yielding 1,00,139 running words and 2883 different words. After scrutiny a list of 1500 words, meant to be functional as well as general nature of
reproduction vocabulary, was drawn up. Shanker says that this list does not encompass the comprehensive vocabulary of the concerned students/children.


(3) S.N. Tambhane took up a project to study the Basic Vocabulary of Marathi speaking children of primary schools, in the year 1965. This project was financed by the NCERT, New Delhi. A huge sample of 8000 school students of classes I to IV in the age-groups 6-8 and 8-10 was involved in this study.

The main purpose of the study was to determine the list of most common and essential words in Marathi language which could be considered the 'formation vocabulary' for the students in the age-groups 6-8 and 8-10. The samples were drawn from 8 districts of Vidarbha region. Out of the total sample of 8000 students 3000 were from 6-8 age-group and 5000 were from 8-10 age-group. These students were observed under controlled conditions, free-play situations as well as specially organised situations. Oral and written sheets were used to identify reproductive and recognition vocabularies of the students/children.

For the purpose of counting of words the investigator followed the principles suggested by Koenig (1931). After counting, only those words having frequency of 70 or more were retained. The main findings of the study were as under:
(i) the total general vocabulary of both the groups or the entire group (6-10 years) of children was 4550 words;

(ii) the total reproductive vocabularies of 6-8 age-group and 8-10 age-group were 1705 words and 1057 words respectively;

(iii) the total recognition vocabulary of 6-8 age-group and 8-10 age-group were 755 words and 2090 words respectively;

(iv) the reproductive vocabulary of higher age-group (8-10) is a part of general vocabulary of lower age-group (6-8) but the recognition vocabulary of higher age-group is exclusive of the general vocabulary of the lower one;

and (v) out of the tentative basic word list of 2043 words 1135 were determined as the real basic vocabulary of the entire age-group of 6-10.

(SOURCE: Buch, 1974)

2.3.5. KANNADA

The two studies, (1) 'Teaching Reading to Beginners' and (2) 'Basic Vocabulary' in Kannada language are discussed below.

(1) V.K. Javali took up an experimental study at the Ph.D. level in 1949 under Bombay University. The study is on 'Teaching Reading to Beginners'. For the purpose of this study the author listed 72,000 running words from
162 source-books (31 readers, 10 magazines and 121 miscellaneous books) in the Kannada language meant for the beginners in the primary schools. Lastly 2000 words, having frequency of 4 and more were selected by the author as 'Kannada Primary Reading Vocabulary'. The author finally divided the 2000 words into two lists of 1000 words each, the first 1000 having frequency of 11 and more and the second/last 1000 words having frequency 4 to 10. The author, however, failed to lay down the criteria employed to arrive at such a decision.


(2) The second one is an investigation into the basic vocabulary, in Kannada, of elementary school children belonging to standards I to VII of Mysore State conducted by B.K. Chandrasekhariah in 1968 on behalf of the Educational Research Bureau, Bangalore, financed by the grant-in-aid scheme of the NCERT, New Delhi. The preparation of a comprehensive basic graded vocabulary of about 4000 to 5000 words supposed to be understood by almost all children as well as to be the basis for all reading materials for children of primary standards I to VII in the Mysore State, was the objective of the study.
Picking up 5757 words from different sources, comprising departmental readers—standards I to VII, word-lists published earlier in the state and other reading materials a preliminary list was prepared by the author. Two hundred primary school teachers at the rate of 10 teachers from each of 20 districts of Mysore were selected who were required to judge and select and to place the words in proper grades and also to find out the difficult words. The teachers were supplied a list of criteria basing on which such selection and grade placement of words would be made. Even though the criteria were fixed there was wide disagreement of opinion among the teachers with regard to grade-placement of words as well as deciding difficult words. About the divergent opinions the author says that it was not the judgement of the teachers but the regional variations in the usage of words that counted for much of the divergence.

Finally a list of 5000 words was prepared as the 'comprehensive basic vocabulary' with the following grade distribution.

<table>
<thead>
<tr>
<th>GRADES</th>
<th>WORD LOAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>700</td>
</tr>
<tr>
<td>II</td>
<td>700</td>
</tr>
<tr>
<td>III</td>
<td>700</td>
</tr>
<tr>
<td>IV</td>
<td>750</td>
</tr>
<tr>
<td>V</td>
<td>800</td>
</tr>
<tr>
<td>VI</td>
<td>700</td>
</tr>
<tr>
<td>VII</td>
<td>650</td>
</tr>
<tr>
<td>TOTAL:</td>
<td>5000</td>
</tr>
</tbody>
</table>

(SOURCES: Chandrasekhariah, 1968, Buch, 1974)
2.3.6. **BENGALI**

(1) The earliest study on Bengali vocabulary was conducted by Roy and Roy (1946) on behalf of the 'Committee of Bengali Vocabulary Studies'. The study was published as 'Easy Bengali Vocabulary' in the form of a 'Research Monograph'.

This vocabulary study was meant for use in writing textbooks for the children between 7 to 12 years of age. The study was initiated with the feeling that the existing textbooks contained difficult words beyond the comprehension ability and interest of the children for whom they were meant. The authors collected words from 20 first-grade-readers and counted the frequency thereof. The authors finally prepared a list of 3500 Easy Bengali Vocabulary of high frequency and arranged them alphabetically. It was recommended by the authors that the textbook-writers and writers of children's literature should use such words which the beginners could understand without exerting much strain on their minds.

It is said about the study that although the difficulty value of different listed words was not given still the study was an eye opener to the textbook writers (Naseem, 1978). (SOURCES: Roy and Roy, 1946 and Naseem, 1978)
Another study was conducted in the year 1978 under the name of 'Basic Bengali Vocabulary' by T. Dasgupta, as his Ph.D. study under Gauhati University. The objectives of this study were:

(i) to construct a glossary of Basic Bengali Vocabulary for classes I, II, III and IV;

(ii) to assign difficulty value of all words in the glossary in respect of class IV only so as to ascertain whether more than one year exposure has only significance on the vocabulary performance of the pupils;

(iii) to identify the best, worst and medium scores on the vocabulary test batteries;

and (iv) to compare the performance of class IV students of Shillong and that of two districts of West Bengal to ascertain whether the glossary has wider applicability.

To achieve the above objectives four textbooks, one from each class, were chosen and analysed in terms of total and different words for each class. The different words were alphabetically arranged in four progressive glossaries. The total different words found out were 1209 out of which 26.96%, 33.01%, 16.04% and 23.57% of words were of classes I, II, III and IV respectively. One of the findings of the study was that 8.33%, 32.01% and 59.64% words of the glossary prepared by the author were of high, medium and low difficulty level value respectively. (SOURCE: Buch, 1979)
2.3.7. **PUNJABI**

'An Investigation into the Basic Punjabi Vocabulary of Fifth Class students (10+) in the State of Punjab' was conducted by H.S. Soch in the year 1974 under Punjab University. The main aim of the study was to prepare a glossary for fifth graders comprising 'general words of common usage' as distinct from 'specific or technical terminology'. For this 1229 words were collected from different sources. These words were divided into 17 check-lists giving five alternatives for each word in the lists. These check-lists were administered on 2000 fifth grade children drawn up almost randomly from 12 districts of Punjab. Analysis was made of the data in terms of (1) difficulty level of words, (2) standard error of population and (3) difficulty level corrected for chance. (SOURCE: Buch, 1979)

2.3.8. **TAMIL**

A study of 'Functional Vocabulary' of pre-school Tamil children was conducted by V. Arunajatai and G. Srinivasachari on behalf of SITU*, Council of Educational Research, Madras in the year 1968. Children of 30 to 60 month age-group, numbering 700, were involved as the sample in the study. Four supervisors and 30

*SITU : South India Teachers' Union.
trained observers collected data relating to thought and language of the sample children through study-sheets as well as tape-recording of children's speeches. Seventy such study-sheets and 74 recorded speeches supposed to be related to children's most vital experience were used for the interim report of the study. In the whole study the functional vocabulary of pre-school age Tamil children was found to range from 1500 to 2000 words. It was also found that the children had considerable command over use of the three persons and singular and plural numbers in ordinary speech. Over and above they were capable of using basic sentence patterns and about 50% of them were capable of constructing elementary complex sentences. The mean number of words used in their sentences was found to be 3.7 and the mean number of syllables was 9.7. It is said that the children compared favourably with American and British children of the age-group. (SOURCE: Buch, 1974)

2.3.9. ORIYA

It seems that no attempt has yet been made to study the language of Oriya speaking children, whether 'spoken' or 'written' or 'reading' for the purpose of using it in the children's textbooks or other children's literature. The few works on Oriya vocabulary done till today are related to learning of Oriya language by adult learners.
as a second language. Since the present study is on Oriya vocabulary the investigator wishes to mention all these works in brief.

(1) Matby (1874) had prepared a 'Practical Handbook of Oriya Language'. This is a glossary of some Oriya words essentially useful for ordinary communication with English equivalents. This glossary was meant for the non-Oriya officers who were serving in Oriya speaking districts under the then Madras Presidency. No information is available about the procedure of collecting of such essential Oriya vocabulary by Matby. Subsequently the Government of Orissa thought this glossary to be useful to the students of Oriya language and, therefore, got the glossary revised by Rai Sahib Jadunath Mohapatra and also got it published under its authority in the year 1945.

(2) Matson (1970) and his Oriya associates (the project staff) did a huge 'word count' for preparation of Graded Readings in Oriya' meant for foreign students. This research project was executed pursuant to a contract between Cornell University and the Office of Education, United States Department of Health, Education and Welfare, under the authority of Title VI of the National Defence Education Act.
A 'sizeable staff' including Mrs. Matson, as the coordinator, was engaged to work out the project.

This 'word count' involved a huge amount of materials consisting of 13 Oriya stories, 10 Oriya short stories and 3 Oriya novels. The aim of this count was to determine the order in which these reading materials were to be organised as 'graded readings' on the basis of frequency of words in different stories and novels as well as frequency of words on the whole count.

The detailed report of the work is not available. Only a page of 'Foreword' by the author (Matson) is available in some of the volumes of the mimeographed publication entitled, 'The Oriya Language Textbook Series'. As it seems, the original plan of the work was to publish these series in 8 volumes, but the first two volumes (series 1 and 2) are yet to see the light. The eighth volume of the series is 'Oriya Word Count'. Volumes 3 to 7 contain the Oriya stories, short stories and novels in transliterated forms along with their glossaries.

The 'Oriya Word Count' (vol. 8) contains 249 pages of handwritten grid forms listing 5683 transliterated different Oriya words with their meanings in English, as they occur
in the sampled reading materials. Against each word frequencies of occurrence in different stories, short stories and novels as well as the total frequency in the whole count have been noted. The range of frequency in the total/whole count is found to be 1 to 2404. There are, as counted by the present investigator from this volume, 242 words having the total frequency of 50 and above. Of these 98 are nouns, 48 verbs, 51 indeclinables, 27 pronouns and 18 adjectives.

About the use of this word count the author states that this volume provides for teacher and student alike on account of the utility in terms of reading efficiency within the series of texts.

(3) Pattnayak and Das (1972) prepared a course of 'Conversational Oriya' meant for the foreign students to learn Oriya language. While designing the course the authors have taken into consideration the fact, 'language learning is basically a process of habit formation.' So the course has been designed in the forms of conversations under different situations of day to day common life with gradation of Oriya structures befitting to the situation of conversation and with control of vocabulary as well, providing ample scope for repetition of the structures and the vocabulary. The authors admit that the control of vocabulary
they have put in the course is based mostly on their 'intuition as native speakers'. About vocabulary frequency the authors have clearly stated that it is useful only upto a certain point and not after that because after that what is more important is 'frequency of items in the text and not frequency of items in language as a whole'.

2.3.10. **Phonemic and Morphemic Frequencies in Indian Languages**

A project on 'phonemic and morphemic frequency counting' in six Indian languages, such as Hindi, Gujarati, Marathi, Kannada, Malayalam, and Oriya was sponsored by the Ministry of Education, Government of India. The purpose of this project was to help 'evolution of a system of shorthand' or 'divising speedwriting, typewriter key words etc.' This project was taken up by Poona University and Decan College. Information as to when exactly this project was started is not available to the present investigator. The earliest publication under this project seems to be the work on Marathi language by Bhagwat (1961). The other three publications are on Hindi by Ghatage (1964), Gujarati by Pandit (1965) and Kannada by Ranganathan (1981). The present investigator has, however,
had a chance to see and handle the manuscript of the work on Oriya language which was lying in the Central Institute of Indian Languages (CIIL), Mysore and was waiting to be sent to the press (in January '82). The manuscript contains only the alphabetical lists of words, phonemes, morphemes along with their frequencies. The report of the study (on Oriya) was yet to come from its author (Ashoke Kelkar). It was also learnt then from the source of CIIL that the work in Malayalam language was already in the press. But the latest information from the CIIL source reveals that both the works (in Oriya and Malayalam) are still unpublished.

All the above six counts involved about one lakh running words in each of the languages which were collected from randomly selected reading materials (literature from all levels, newspapers, periodicals etcetra) as well as from radio programmes of the respective six languages. In all such works there are the lists of different words alphabetically arranged with frequencies in descending order. There are also lists of phonemes, morphemes, syllables in the same manner. Over and above these lists, the words with frequency of 10 and above have been classified into percentages of coverage of the total corpus of the language. The numbers of such different words are 1526* in Hindi, 1430* in Gujarati and about 1857* in Oriya which cover 33%, 58%* and about 69%* of the corpuses of the

* Calculated by the present author.
counts of respective languages.

Although the main purpose of these counts was to devise 'speed writing' and 'type-writer key-boards' in the concerned languages some of the authors (Pandit, 1965 and Ranganathan, 1981) claim that the findings of their work would also help in preparation of 'textbooks with graded vocabularies' and 'preparation of efficient primers for instruction at the elementary level'. Such claims of the authors prompted the present investigator to include these studies in this review.

Summary

Studies under Reading Vocabulary in English date back to 1836 known as 'graded vocabulary' provided in Pestalozzian Primers based on assumption of the experienced teachers. The objective method of 'word count' was started early in the current century, mostly in nineteen twenties, thirties and forties. All the studies aimed at studying primers and readers through frequency count of words and to prepare 'core vocabulary', 'basic vocabulary', 'graded vocabulary' for the primary children/early reading materials. The major studies in the line were done by Thorndike (1921, 1931, 1936, 1937, 1944) in USA who presented many interesting conclusions about the importance and use of vocabulary in the reading materials for the elementary grades. Till today
his studies are being referred to by the researchers both in his country as well as outside including India. The other great study on reading vocabulary is of Vernon in UK during 1939-40 which is published under the title STUDIES IN READING- Vol. I by the University of London, 1949. This study of Vernon found out the futility of Infant Readers commonly employed in Glasgow schools then.

The earliest study under 'Written Vocabulary' is traced back to 1926. There are two vast studies under written vocabulary. One is 'A BASIC VOCABULARY' by Rinsland (1945) and GRADED VOCABULARY (Leicestershire List) by Edward and Gibbon (1964) in USA and UK respectively. The most recent study on this is a Canadian study known as 'Basic Core Vocabulary' by Thomas (1972). All these studies are vast word counts of written vocabulary collected from the writings of the children.

The 'Spoken Vocabulary' studies date back to 1925 by Ernest Horn. Another great study is known as 'Scottish Word List' by Vernon (1940) who collected the spoken Vocabulary of children aged 4½ to 5½ years. Vernon made two lists of words of which one contains 1910 words used by at least 2 children and the other contains 491 words used by 15 children and more. The author claimed his second list to have contained the commonest words. The latest
study on spoken vocabulary is of Burroughs (1957) in U.K. The study is known as 'Birmingham List'. It consists of two lists. List 1 contains 1909 words used by 5 or more children and List 2 contains 1595 words used by 4 or fewer children. These were further divided into 9 sub-lists in order to compare them with different previous lists separately so as to conclude the reliability of the Birmingham List.

The other studies on children's vocabulary either relate exclusively to any one of the vocabularies such as: 'speech/spoken', 'writing', 'reading-printed' vocabularies or mixture of those. Of major importance are the studies of Dale (1931), Gates (1937), Hardie (1948) and McNally and Murray (1968). All these studies encompass original study of vocabulary as well as the use of previous word lists so as to prepare improved word lists for the purpose of preparation of viable reading materials for the children of early grades.

The major studies in Indian languages have been conducted in Hindi language. The 12 studies reviewed here include 7 individual studies including one M.Ed level study and 5 Ph.D. level studies. The earliest study dates back to Koenig (1931) whose study, it is said, is comparable to Thorndike's study. The latest comprehensive study in
Hindi is that of Rostogi (1982) which at once includes 'spoken', 'written' and 'reading' vocabularies of the first five grades of primary stage.

The one study in Urdu language reviewed here is known as 'Urdu Word Count' by Ahmed (1973). It is a word count of general nature. This count resulted in sorting out four categories of words in Urdu like: 'indispensable', 'essential', 'useful' and 'special' words with frequency ranges 50-36739, 35-49, 25-34 and 15-24 respectively.

Five Ph.D. studies were traced in Gujrati language from compilation sources (Adaval-1968, Buch-1974) out of which only two are related to the present study. One of them dealt with 'Basic Vocabulary of the highest primary class children (Vakil, 1954-55) and the other is by Sukla (1976) dealing with 'written', 'spoken' and 'reading' vocabularies of 6 to 11 age group children.

Out of the three studies in Marathi language traced from the compilation sources (Adaval-1968 and Buch-1974) two are M.Ed. level studies on vocabulary attainments of children of five to nine years old and on reproduction vocabulary of primary course. The third one is a project on basic vocabulary of Marathi children of primary schools, financed by NCERT.
Two studies of Kannada language are (1) Teaching Reading to Beginners and (2) Basic Vocabulary traced from compilation source (Adaval-1968).

Out of the two studies in Bengali language available for review the earliest one was conducted in 1946 by Roy and Roy on behalf of the Committee of Bengali Vocabulary Studies. This study prepared Easy Bengali Vocabulary for writing text books for the children 7 to 12 years age. The other one is a Ph.D. study by Dasgupta (1976) under Gauhati University mainly to construct a Glossary of Bengali Vocabulary for classes I to IV.

One study in Panjabi language was conducted by Soch (1974) on Basic Panjabi Vocabulary for 5th class students.

One study in Tamil Language was conducted by Arunjatai and Srinivasachari (1968) on functional vocabulary of pre-school children.

No related study has so far been conducted in Oriya language. The only word count made by Matson (1970) is from adult reading materials and meant for teaching Oriya language to foreign students. The preparation of 'Conversational Oriya' by Pattanayak (1972) is also meant for the foreign students.
The project of 'Phonemic and Morphemic Frequencies' in six Indian languages (Hindi, Gujrati, Marathi, Kannada, Malayalam and Oriya) was taken by Poona University and Decan College for the purpose of 'speed writing' and 'type writer key board'. But some of the authors of this study (Pandit-1965 and Ranganathan-1981) claim that this study would also help in preparation of text books with graded vocabularies.

*****