It will be quite interesting to note the investigations already made in English Vocabulary. E. L. Thorndike¹ published three word-lists, viz., "The teacher's word-book of 10,000 words", "The teacher's word-book of 20,000 words", and "The teacher's word-book of 30,000 words".

In the first list Thorndike used children's story books, classical books, correspondence, newspapers, text-books of Arithmetic, History, Geography and books on cooking, sewing and dressmaking as his sources for collecting words. Thorndike had taken into consideration the frequency of the word, that is, how often the word was used and range of the word, that is, how wide was the word used, or in how many sources it was used. This work was published in 1921.

In the second list, Thorndike had chosen 238 sources for selecting words. He had chosen books like *Pride and Prejudice* by Jane Austen, *Fairy Tales* by Anderson, *Aesop's Fables*, *The Decline and Fall of the Roman Empire* by Gibbon, *Life of Johnson* by Boswell, *The Encyclopaedia Britannica*, *Ladies' Home Journal*, *Saturday Evening Post* and *Literary Digest*. This list was published in 1931.

In the third list prepared by Thorndike and others a large data of the first two lists had also been used as sources for word-counts. They had also used words from other juvenile literature excluding text-books. This list was published in 1944.

Horn\(^2\) was a pioneer in the preparation of the word-list, *A Basic Writing Vocabulary*, which was published in 1926 and is still valuable. It contains 10,000 words commonly written by adults and would therefore be more valuable for older children.

Gates\(^3\) in his book, *A Reading Vocabulary for the Primary Grades*, lists approximately 1,800 words which should ordinarily appear in the basic reading vocabulary of the primary grades. This has been widely used as a check-list in the preparation of text-books for the primary grades.

\(^2\) Ernest Horn, *A Basic Writing Vocabulary* (Iowa city: University of Iowa press, University of Iowa Monographs in Education, No. 4, 1926).

Stone\textsuperscript{4} has also constructed a graded vocabulary of primary reading. It is a list of 2,000 words of greatest value in primary reading based mainly on word-counts of primary readers. It also contains a separate list of the most important 150 words for "Beginning Reading". Stone's graded vocabulary contains primary reading lists, 2,164 words selected from those appearing most frequently in each of twenty-one pre-primer, primer, first readers, second readers, and third readers published between 1931 and 1941. The words are graded into ten reading levels.

Buckingham and Dolch\textsuperscript{5}, in their publication, \textit{A Combined Word List}, combine and show the overlapping of eleven word-count studies. This is one of the most useful of the lists of frequently used words.

Dolch\textsuperscript{6}, in the chapter on "Sight Vocabulary" in his book \textit{Teaching Primary Reading}, gives two word-lists. One is a list of 220 words, omitting nouns, that are highly useful as a basic sight vocabulary; the other is a list of 95 commonly used nouns. This is a useful checklist for work

\begin{itemize}
\item \textsuperscript{4} C. R. Stone, \textit{Stone's Graded Vocabulary for Primary Reading}. (St. Louis: Webster Publishing Co., 1941).
\item \textsuperscript{5} B. R. Buckingham and E. W. Dolch, \textit{A Combined Word List} (Boston: Ginn & Co., 1936).
\item \textsuperscript{6} E. W. Dolch, \textit{Teaching Primary Reading} (Illinois: Garrard Press, 1941).
\end{itemize}
With slow learning children.

Fitzgerald, in an article, "The Vocabulary of children's letters written in life outside the school", gives a list that may serve as a guide in the teaching of spelling.

A number of people have studied the growth of vocabulary in pre-school children. Some have attempted to enumerate the total vocabulary of a child during these early years and their results indicate a wide range of individual differences. For older children, especially children who are beginning to have some experiences apart from their parents, a sampling method is necessary.

The most widely quoted estimation of the probable vocabulary of pre-school children at various ages is that of Smith which was prepared in 1926 and was based on a study of 273 pre-school children. This material is given in table 2.1.


TABLE 2.1

INCREASE IN SIZE OF VOCABULARY IN RELATION TO AGE

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Number of Children</th>
<th>Average I.Q.</th>
<th>Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>Months</td>
<td></td>
<td>No. of words</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>52</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>14</td>
<td>22</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>14</td>
<td>118</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>25</td>
<td>272</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>14</td>
<td>446</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>20</td>
<td>896</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>26</td>
<td>1222</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>26</td>
<td>1540</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>32</td>
<td>1870</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>20</td>
<td>2072</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>27</td>
<td>2289</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>9</td>
<td>2562</td>
</tr>
</tbody>
</table>
Madeline Darrough Horn and the Child Study Committee of the International Kindergarten Union\(^9\) studied the vocabulary of normal children before entering first grade. The words listed were obtained from three sources: words used by children in kindergarten, words children used when stimulated by pictures, and words children used in the home. This study, reported in 1928, had been used in various ways by teachers, text-book writers, makers of tests, and mothers in the home guidance of their children during the more than twenty years since its publication. It lists 2,596 words in the order of frequency of use in the records gathered from the three sources. This study is still valuable though there is little doubt that in the intervening years children of this age have added many other words to their vocabularies through contact with radio, motion pictures, television, and other more recent sources of stimulation and experience.

Rinsland\(^{10}\) studied children's writing vocabularies. He analyzed over 100,000 papers written by school children in grades one through eight, a total of more than 6,000,000 running words, and found the writing vocabularies at the various grade levels as follows.

---


Grades: 1 2 3 4 5 6 7 8
Different words: 5,099 5,821 8,976 9,976 11,449 11,304 14,820 17,930

Roots, derived forms, abbreviations, and contractions were all listed in this study as separate words. Only baby-talk, slang, provincialisms, colloquial expressions, trade names, and proper names, except very well-known terms, were deleted. Letters, expositions, original stories, poems, examination papers, and other materials were sent in from schools all over the United States, supplemented by records of conversations of first grade children.

Mary Katherine Smith used a test devised by Seashore and Eckerson, administering the test to the first twelve grades. The test used a systematic sample taken from an unabridged dictionary, in this case the third word down from the top of the left-hand column of every eighth page. This gave a total of 331 basic words. The results of the test are given in table 2.2 in terms of basic words and derivative words.

The average growth in total vocabulary is approximately 5,000 words per year from the first through the twelfth grades. The proportion of derivative words in the entire vocabulary increases greatly through the years, from about 30 per cent in first grade to about 60 per cent in college. Seashore and Eckerson carried the study on to the college level and found in three universities an
average vocabulary on the part of students of 61,000 basic words, plus 96,000 derivative words, or a total of 157,000 words.

Lorge and Thorndike's semantic frequency count\(^\text{12}\) (the selection of words for the teaching of English to foreigners), to a large degree, has been based on the frequency of the occurrence of words in printed English. Such facts about word frequency have been useful in the choice of basic vocabularies of specified size. The teaching of these words, however, left much to the judgment of the teacher, for he had to choose which meaning or meanings of these words to impart. In general, the more frequent the occurrence of a word, the greater is the variety of meaning in which it is used. A brief description of the basic procedure involved in semantic counts will illustrate the entire process.

First, it was necessary to select a sample of material to be read. For this purpose many sources, such as encyclopaedias, magazines, text-books, novels, essays, biographies, books about science, poetry and the like were sampled. Altogether, the original sample included about two-and-a-half million words. This sample was later supplemented by another sample of the same size.

The Semantic Count is based on the differentiation of the meaning in the *Oxford English Dictionary*. The thirteen-volume dictionary was split into folios of 32 pages. Each person selected to do the semanticizing was of high intelligence and especially trained for the task. He proceeded to read the materials. Whenever he found a word in his section of the dictionary he made a record of it, giving its location by unit, page and column in the sample. He then studied its apparent meaning in context to assign its proper meaning, number and letter as given in the *Oxford Dictionary*. The semanticist usually read all of the five million words for this purpose.

Two major problems confront the teacher of English as a second language: how to teach and what to teach. English is so complex a language that, in order to simplify the learner's task, selection from it becomes a necessity, and the teacher must decide which word meanings and phrases are most worthy of inclusion in this simplified vocabulary.

Considerable work has been done on this problem by studying the frequency with which various words occur in written and spoken English. In 1934 a conference was held under the auspices of the Carnegie Corporation, and its work was published as *An Interim Report on Vocabulary Selection*. Which included a General Service List of 2,000 English words compiled by a sub-committee consisting of Faucett, Palmer and West. This report rapidly became a most valuable aid to teachers all over the world, and when it
went out of print, arrangements were made to revise and expand it to include the subsequent research carried out in this field.

The result is the volume, *A General Service List of English words*, in which the various meanings and uses of selected words have been rearranged in the light of the Lorge-Thorndike semantic frequency count, and the comparative frequency values of the various meanings added against each word to indicate their relative importance. The book also contains a supplementary scientific and technical vocabulary.

It is fitting that Michael West, a member of this distinguished group of workers on vocabulary selection, who has done so much to forward its study, should have been responsible for the compilation of a work which will prove invaluable to all teachers and students of the English language.

STUDIES IN INDIA

Koenig, a pioneer in the field of vocabulary studies in this country, attempted at an analysis of Hindi words used in juvenile literature. Pursuing Thorndike's technique he patiently collected, out of a million "running" words from 153 sources of juvenile literature, 12,500 "different" words which pupils in the primary classes were

---

expected to absorb. Out of the latter he selected 4,000 words with a frequency of ten or more and these were held as basic vocabulary in Hindi for grade I to IV. Koenig's interest mainly lay in drawing up grade vocabulary and not age vocabulary; use of the scientific criterion of age those days posed quite a few practical problems.

Comparing the arrangement of new words employed in Hindi and English primers (American) Koenig, came across a striking difference. Whereas the American readers employed on the average one new word in seventeen, Hindi books used new words much more frequently - one new word after every second to eleventh word. This hardly offered an opportunity for drill work that is important for word absorption. Pertinently he posed the query: 'Does not these excessive proportions of new words explain the fact that many pupils of Indian primary schools leave the school without comprehension? At the same time the vocabulary burden of Hindi primers which is 11.8 fell far short of its counterpart in American primers which is 20.

Bhat\(^{14}\) was one of the earliest Indian investigators in this field. He attempted at a longitudinal survey of vocabulary growth from 5 to 9 years and his work bears the drawbacks that have to be faced by a pioneer, mostly in

\(^{14}\) S. R. Bhat, "An Inquiry into the Marathi Vocabulary Attainments of Children Five to Nine years Old in the city of Bombay" (Unpublished M.Ed. dissertation, Bombay University, 1939).
matters of methodology and technique. He made it clear that 'it is the qualitative aspect of the matter that I am concerned with'; consequently depiction of numerical data is conspicuous by its absence.

Adopting the technique of Piaget, Dewey, Drever, Gessell and others, Bhat took recourse to the 'peep hole' method of studying short sample behaviour commonly employed in child and comparative psychology. He talked to a limited number of children under each age group (not the same number in each group though) on a few selected topics and here he has left a gap to guess about his detailed programme. "I analysed their responses and came to know that a child of three years knows about 250 words which were, more or less, concrete in nature denoting things, animals and actions. The vocabulary of a four-year old child was about 400 words - an extension of the vocabulary for three years". He does not explain why he used children of the 3-5 year age group when his work concerned the group 5 - 9 years. Pursuing further he proceeded to determine the vocabulary of the relevant age groups, the results, with the figures rounded off, are shown below.

<table>
<thead>
<tr>
<th>Complete years</th>
<th>Words known</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>500</td>
</tr>
<tr>
<td>6</td>
<td>640 - 1100</td>
</tr>
<tr>
<td>7</td>
<td>850 - 1,990</td>
</tr>
<tr>
<td>8</td>
<td>920 - 2,910</td>
</tr>
<tr>
<td>9</td>
<td>1,190 - 4,100</td>
</tr>
</tbody>
</table>
In Conclusion Bhat observes 'the vocabulary attainments of Marathi speaking children from the 6th to the 10th year are less than those of children in the western countries' and that 'children of different (social) classes use different words to denote one and the same thing, specially in the early stages.' He thus opened the doors ajar for crosscultural studies which in future may prove to contain a rich field if pursued more systematically. Bhat's work can certainly be held as a milestone in the field of developmental studies with reference to language.

Seolekar's work\textsuperscript{15} represents an attempt to determine the reproduction vocabulary in Marathi language of children of the fifth standard (age about 12 years). The pupils were required to compose essays which were scrutinized against the criteria of the extent of vocabulary, frequency of words used, grammatical classifications of the words, and content classification of the vocabulary. The composition was also judged for an 'analytic study of words in the various kinds of lists prepared to find the functional as well as the general nature of the reproduction vocabulary'. Compositions numbering 810, comprising the work of 270 pupils and resulting in 100,139 words, were scrutinized and the total number of different words was found to be 2,883. These words were classified according to the four criteria

\textsuperscript{15} U. R. Seolekar, "The Reproduction Vocabulary in Marathi of Pupils Who have Completed the Primary Course" (Unpublished M.Ed. dissertation, Bombay University, 1946).
mentioned earlier and the data arranged in tables, though
use of statistical devices was not made. The author drew up
a list of 1,500 words against the criterion of frequency.

Javli's work\(^\text{16}\) bears the influence of the Work-play
Books by Gates-Huber-Ayer and the new Work-play Books by
Gates and Bartett. After consulting 162 source-books in the
Kannada language meant for the beginners in primary schools
the author has drawn up a list of 72,000 running words, out
of which he has derived a list of 2,000 words. He selected
1,000 words with a frequency of 11 or more and also 1,000
words bearing the frequency of four to ten though he has
failed to account for the criteria employed.

Vakil\(^\text{17}\) introduced a more sophisticated tone in the
methodology of his investigation which claimed "to prepare
a vocabulary list in Gujarati for pupils who have completed
the first year of the senior basic or the secondary school
(11+) stage based on a study of text-books in common use
and to find out how many of the words in the list form
part of the children's active recognition vocabulary, to
ascertain the reproduction vocabulary of children at that
stage and to compare it with the recognition vocabulary". Vakil selected the 11+ year group for his investigation

\(^{16}\) V. K. Javli, "Teaching Beginners to Read Kannad:
An Experimental study" (unpublished Ph.D. Thesis, Bombay
University, 1948).

\(^{17}\) A. S. Vakil, "The Basic Vocabulary of Gujarati
Children at the age of 11+ (Bombay: Law Publishers, Princess
Street, 1967).
because it was mostly at this age that the pupil completed primary education and a substantial number of the pupils moved out in order to earn livelihood. His objective was to acquire information about the vocabulary assets of these pupils.

Vakil used four popular text-books from which words were to be finally selected. The books gave him 95,000 running words and 14,200 different words. These words were tabulated under eight categories of frequency, the maximum being 13 and the minimum 1. Three hundred sixty words from the group of different words (excepting the first group of highest frequency) were sampled out (how?) for using these in his tests. As for the subjects, three groups of fifteen male pupils from standard VI (age not given) were selected, with the logic "Those who had secured more than 65 marks at the last examination, those who had secured between 50 and 65 and those who obtained only 40-49 marks". It remains to be explained as to why pupils below the last category were exempted from the test.

The fifteen pupils could only recognize 52 words out of 362 and this result is somewhat surprising since the 362 words represented a most widely recognized group of words. After reaching this stage Vakil found his method 'though very reliable not very convenient' and therefore switched on to another scheme showing thereby that the technique of prior designing of experiments had not yet grown in significance in India.
Expecting better results, Vakil now drew up a list of 429 words from 'a list of first five thousand words', his samples based on difficulty of the words, frequency of its appearance and 'personal experience'. By way of trial test ten male subjects of grade V were used. Eight of the ten boys were found to recognize 202 words. The method of selecting words hereafter becomes slightly confusing. A list containing 1,075 unrecognized words was now drawn up and used in constructing the tests. Four tests of the usual recognition type were prepared to determine the number of words that the 460 boys could recognize and this number came to be 854.

The author distinguished reproduction vocabulary from recognition vocabulary and treated them separately. Male and female pupils were required to write essays on a selected list of topics. The 906 compositions contained over a lakh of running words and four thousand new words which were further analysed against parts of speech. Distribution of the reproduction vocabulary according to the parts of speech was as follows:

- **Nouns**: 66.70%  
- **Pronouns**: 1.10%  
- **Adjectives**: 15.40%  
- **Verbs**: 11.27%  
- **Adverbs and prepositions**: 4.29%  
- **Participles**: 1.29%
Dave's work\(^{18}\) falls on slightly different lines as his object mainly was to construct and standardize arithmetic ability test, and also to draw up vocabulary list and reading comprehension list to help test construction and standardization. Introducing experimental sophistication appropriate to the period, Dave determined reliability and validity of the tests and their inter-correlations. It is rather curious that employing elementary statistical devices so essential to the present field of inquiry should have been so long to come. However, separate norms were drawn for boys and girls and he found that scores in vocabulary tests show high correlation with the scores in intelligence tests. According to him the Gujarati child of twelve has on the average a vocabulary of 12,390 different words; a comparison with other linguistic groups would have no doubt proved interesting.

Using the test devised by him he arrived at the following results in scores for boys and girls.

<table>
<thead>
<tr>
<th>Age</th>
<th>Mean Score for Boys (N = 3,792)</th>
<th>Mean Score for Girls (N = 2,155)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 +</td>
<td>12.75</td>
<td>11.20</td>
</tr>
<tr>
<td>8 +</td>
<td>15.40</td>
<td>14.50</td>
</tr>
<tr>
<td>9 +</td>
<td>18.95</td>
<td>17.39</td>
</tr>
<tr>
<td>10 +</td>
<td>24.20</td>
<td>24.35</td>
</tr>
<tr>
<td>11 +</td>
<td>30.20</td>
<td>20.05</td>
</tr>
<tr>
<td>12 +</td>
<td>33.50</td>
<td>32.95</td>
</tr>
<tr>
<td>13 +</td>
<td>35.10</td>
<td>34.50</td>
</tr>
<tr>
<td>14 +</td>
<td>37.70</td>
<td>38.75</td>
</tr>
<tr>
<td>15 +</td>
<td>41.30</td>
<td>41.95</td>
</tr>
<tr>
<td>16 +</td>
<td>45.10</td>
<td>45.50</td>
</tr>
</tbody>
</table>

It may be noticed that there is little sex difference in the scores and this finds support in the work of western workers in this field. McCarthy, for example, observes "the magnitude of the sex differences usually found in young children's language is not large enough to yield statistically significant differences when the usual criterion is employed". The developmental trends also follow the pattern of results obtained by Heiders, Stormzand, O'shea and others. Anyway, Dave's work represents a corner-stone in sophisticated educational research in this country.

Raval's investigation\textsuperscript{20} is of the usual snapshot type compiling vocabulary from five popular text-books for recognition and reproduction purposes. For finding out the reproduction vocabulary he gave composition tests to 967 pupils and derived about 1.5 lakh of running words and a list of 6,676 different words. He classified the words of both recognition and reproduction vocabulary according to grammatical analysis, origin, etc. To confirm his previous findings Raval also tested the pupils against seven tests of recall and recognition types.

Lakdawalla's investigation\textsuperscript{21} in the year 1959 follows the pattern of her previous study; more than half of the thesis is devoted to historical introduction. She collected over two lakh running words from six current text-books (it is assumed that their contents must be to the taste of the children for whom they are prescribed) and arrived at nearly 15,000 different words. Out of these, about 6,700 words were selected by ten teachers as words which the children might recognize. Finally, a list of 385 words was drawn up and these words were used for three recognition tests employed


by her. A large number of pupils of the 13+ age group were tested and it revealed that 54 per cent of the words were recognized, which were classified, as it was customary, according to the parts of speech, etc.