

Reading vocabulary levels in freshman textbooks at the National University and the implications for materials selection in English Language service courses.

By Vivienne Mosback.

The aim of the project which is the subject of this paper was to find out the most useful vocabulary range in English for freshmen at the University, bearing in mind the vocabulary they are exposed to in textbooks in other subjects.

Service English courses should help students to read and understand better the books they have to read and the lectures they have to follow in their specialist fields. Of course in the time allotted for English it is not possible to cover highly technical or specialist terminology in Science, Economics or Business Studies, say, especially as it is current University policy to give only general English courses which must be the same across all faculties. It would seem obvious, however, that the materials devised by the English department should reflect the *general* level of difficulty to be met with in the reading set for other courses.

Guided by intuition and experience, an established English teacher is able to choose passages of a standard demanding enough, but not too difficult for students in a given year at the University. However, if the specialist textbooks are more difficult, or less difficult, than the English teaching materials, is the English teacher justified in keeping to the level of difficulty he considers suitable, or should he rather gear service English courses to the requirements of other subjects.

The writers of the English 111 course used for three semesters in 1972-3 and 1973-4 felt that the 5,000 level provided a reasonable working vocabulary for first year students. They consulted, in particular, the 'Ladder' Series word list for comprehension passages and vocabulary test items. Up to that time, however, there had been no systematic study of vocabulary levels in texts the students had to read in other subjects. This seemed worth doing. If there was an obvious discrepancy between the level expected by the English department and that useful for students working on their own in other fields, then arguably, the English courses should be adjusted either up or down to be of greater service - or possibly recommendations should be made to other departments concerning outstandingly difficult set texts.

The first step was to establish which textbooks were the required reading for a first-year, first-semester student, which sections of them were on the programme, and assess by interview how important a lecturer concerned considered the reading assignments.

Reading vocabulary levels in freshman textbooks at the National University and the implications for materials selection in English Language service courses.

By Vivienne Mosback.

The aim of the project which is the subject of this paper was to find out the most useful vocabulary range in English for freshmen at the University, bearing in mind the vocabulary they are exposed to in textbooks in other subjects.

Service English courses should help students to read and understand better the books they have to read and the lectures they have to follow in their specialist fields. Of course in the time allotted for English it is not possible to cover highly technical or specialist terminology in Science, Economics or Business Studies, say, especially as it is current University policy to give only general English courses which must be the same across all faculties. It would seem obvious, however, that the materials devised by the English department should reflect the *general* level of difficulty to be met with in the reading set for other courses.

Guided by intuition and experience, an established English teacher is able to choose passages of a standard demanding enough, but not too difficult for students in a given year at the University. However, if the specialist textbooks are more difficult, or less difficult, than the English teaching materials, is the English teacher justified in keeping to the level of difficulty he considers suitable, or should he rather gear service English courses to the requirements of other subjects.

The writers of the English 111 course used for three semesters in 1972-3 and 1973-4 felt that the 5,000 level provided a reasonable working vocabulary for first year students. They consulted, in particular, the 'Ladder' Series word list for comprehension passages and vocabulary test items. Up to that time, however, there had been no systematic study of vocabulary levels in texts the students had to read in other subjects. This seemed worth doing. If there was an obvious discrepancy between the level expected by the English department and that useful for students working on their own in other fields, then arguably, the English courses should be adjusted either up or down to be of greater service - or possibly recommendations should be made to other departments concerning outstandingly difficult set texts.

The first step was to establish which textbooks were the required reading for a first-year, first-semester student, which sections of them were on the programme, and assess by interview how important a lecturer concerned considered the reading assignments.

A random sample of pages was taken from the specified sections, the size of the sample varying to some extent according to the length of assigned readings since a disproportionate number of entries from one particular subject might falsify the general level of difficulty. The sample varied between 4% and 10% of the required reading for each subject. In Philosophy: Logic, for example, 4,500 words represented in fact slightly more than 10% of the required reading for the semester; Sociology, at the other extreme, set over 270,000 words of reading from several sources. A ten percent sample of this subject would have considerably overweighted the study. Where a very large amount of reading such as this was set, the sample took sections from each chapter so that *no area* of vocabulary would be omitted.

Every *noun* and every *verb* on each page of the sample was recorded. Nouns and verbs only were recorded since they reflect the subject matter. In other words, the semantic level of difficulty was the main concern. Verbs were recorded as infinitives, nouns, common nouns only, in the singular.

Every occurrence was noted, since it is useful to know if the frequency of occurrence of a particular item in *specialist* texts corresponds to the frequency of the item in general word counts such as Thorndike and Lorge (1944) or whether a normally rare word occurred frequently in several subjects in the students' reading. Also it was important to know the actual size of the sample in terms of the total number of occurrences.

Entries were recorded in such a way (colour-coding) that it was possible to recognise frequency by subject as well as total general frequency. In this way it was clear if a low frequency (i.e. difficult) word was common in the students' reading in *one* subject or in several. This also revealed which subjects were more or less demanding as regards difficulty of vocabulary.

Because it was decided to record every noun and verb, an item such as NOTICE was entered twice, under both verb and noun if it so occurred. These were later counted as separate words in the total number of items because of differences in the semantic level of difficulty and in general frequency. In Psychology, for example, the verb TO CONDITION must obviously be of a different level from the noun CONDITION. This factor is not taken into account in Thorndike and Lorge, though it is in West (1953). A check on the sample revealed that counting a noun and verb having the same form as *one* item instead of two would have reduced the total number of items by 8%.

The sample covered the set books in Economics, Geography, History, Logic, Psychology and Sociology - six fields in all. There are no English texts for the freshman Ethiopian Languages courses, and the English language course was, of course, irrelevant to the study. The six fields more than adequately represent the range of texts in English that a freshman has to cope with in *one* semester. A full programme must give a total of 16 credits; the six subjects of the sample give 17, in fact. An Arts range of subjects was taken for the sample, since Maths and Science textbooks rely heavily on formulae, symbols and other non-linguistic features to carry meaning.

The total number of occurrences recorded was 12,509. The number of nouns and verbs represented in this sample was 2,604. Each item was checked against the New Horizon Ladder Dictionary, and given the relevant number - 1,2,3,4, or 5 - assigning it to the 1st, 2nd, 3rd, or 5th thousand commonest words in English. If the Ladder Dictionary did not list the word (i.e. considered it did not fall within the 5 thousand commonest words) it was checked against a word count list with a range greater than 5,000, in this case Thorndike and Lorge.

The Ladder count was chosen as the basis for this piece of research for several reasons. First published in 1970, it is the most recent assessment of words most frequently used. Michael West's General Service List (1953) while more detailed as regards the semantic variations of individual words, presents only 2,000 words - too basic a level for University reading. Thorndike and Lorge, the only other established authority at this level, dates from 1944, and though it is a much more comprehensive list (ranging up to 30,000 in frequency), it has two disadvantages. It was prepared from samples of a very general nature, including fiction and other imaginative writing - not particularly with foreign students in mind. Secondly, in 30 years, some words have acquired a greater frequency, while others once fashionable have become less common. Words of a technical nature are noticeably absent in Thorndike.

The Ladder Dictionary on the other hand (1970, p. vii) is the result of:

reviewing frequency word counts and English-teaching word lists and consulting a number of English teachers. The vocabulary has been used, tested and revised over a period of twelve years in the preparation of books in English at differing levels of reading difficulty, for readers of varied ages and interests.

The English department has a range of these readers available, and for the last two years has encouraged students in the Freshman year to practise reading skills and broaden their background using this series.

All these seemed good reasons for using the Ladder count *initially* to see if a basic 5,000 words was a reasonable vocabulary for a freshman to work with, although the phrase 'basic 5,000 words' inevitably conceals discrepancies of a type already hinted at. In fact, the Ladder Dictionary explains more than 5,000 words. There are 5,000 *basic* words, each of which is given one of the five levels of difficulty. but, for example, under the entry GENERAL (1) is the noun GENERALIZATION, a word which comes in the over 10,000 range in Thorndike and Lorge.

In the current study, the noun GENERAL occurred twice, GENERALIZATION four times, and GENERALITY once. Following the Ladder classification, these three words were all rated as '1st. thousand' items, and this is how they were rated in this project. If Thorndike and Lorge had been used exclusively, the apparent spread of difficulty would have been much wider, and many more items would have come in the 8-10 thousand, or even 10,000-plus range.

It is very important, therefore to consider further why the Ladder count is more suitable for our purposes. Our interest lies in the *comprehensibility* of items rather than with the precise numerical frequency with which they occur. In the Thorndike count, the items GENERALIZATION and GENERALITY occurred once in a million and twice in a million words respectively. This corresponds to a frequency between 10,000 and 30,000. However, their real level of difficulty for our students cannot be compared with, say GLUTTON or GOSSAMER - words which have the same frequency level. The Ladder count scores in this respect in that it lists prefix and suffix frequencies *separately*, and most of them fall within the first thousand.

An English course intending to use the 5,000 word range could usefully incorporate units on prefixes and suffixes, which can obviously modify the commonest of words. With -IZE, -IZATION, -ITY, etc., a much wider range of vocabulary can be covered than the Thorndike and Lorge 5,000 would admit. In Thorndike and Lorge CLARIFY occurs once in a million (i.e. over 10,000 in frequency) whereas the Ladder count has it at 5,000. Examples of this kind of 'discrepancy' are many, but most teachers in the Ethiopian University context would probably agree that the Ladder assessment is more in keeping with what students could be expected to *assimilate* with little or no difficulty.

There are some words in the Thorndike and Lorge count which have a higher frequency than in the Ladder. In Thorndike and Lorge for example, CATHEDRAL appears at 4,000, in the Ladder at 5,000. Our students' needs are perhaps better reflected by this sort of displacement.

RESULTS OF THE STUDY

According to the Ladder classification, entries were tabulated 1-5 thousand in frequency to produce the following data, from which certain conclusions clearly emerge.

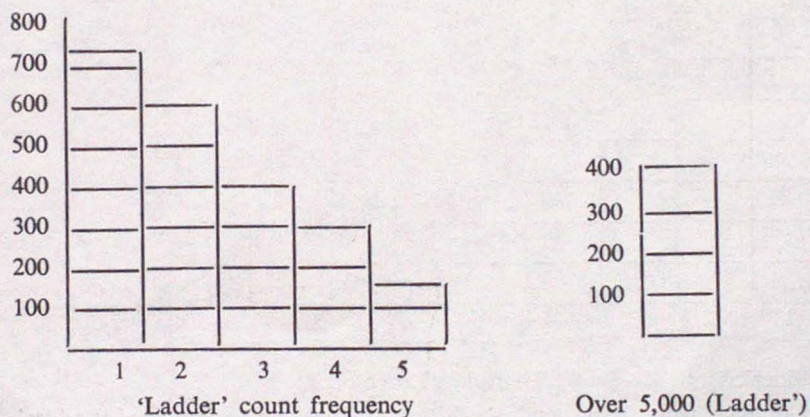


Fig. 1. Word frequency in freshmen text books based upon the 'Ladder' count only.

Each square in the diagram represents 100 individual items (words) in the count. The number of *occurrences* of each word is not shown. It can be seen that the overwhelming majority of the entries fell within the first 3,000 frequency of the Ladder count (730 in the first thousand, 600 in the second thousand, 400 in the third thousand). There was a steady fall-off to 300 in the fourth thousand and 190 in the fifth thousand. This left a residue of some 400 entries over the 5,000 level of frequency, i.e. not to be found in the Ladder count. The next stage of the project was to plot these over-5,000 items according to the Thorndike and Lorge count.

Of the 400 items classified as over-5,000 according to the Ladder count, 22 were, in fact included in the Thorndike and Lorge range from 2-5,000, 16 of them being in the 4-5,000 bracket. Thus there was not a particularly significant discrepancy between the two counts in this area. Of the rest, Thorndike and Lorge had:

17	items	at	6,000
23	"	"	7,000
29	"	"	8,000
29	"	"	9,000
25	"	"	10,000

The Thorndike and Lorge levels are approximations based on occurrences per million words, but nevertheless, the totals per range for this study were all significantly lower than the numbers of items in the 4 and 5 thousand Ladder categories (cf. 300 for 4 thousand, 200 for 5). The total for the Thorndike and Lorge 6-10 thousand was thus 123 items, still leaving 282 items unaccounted for. That is, 13% of the total number of items in the count came in the over 10,000 level. This 13% (282 items) was further split. Approximately 8% came within the Thorndike and Lorge 10 - 30,000; just over 4% were above even the 30,000. The total spread of items can be shown diagrammatically as follows:

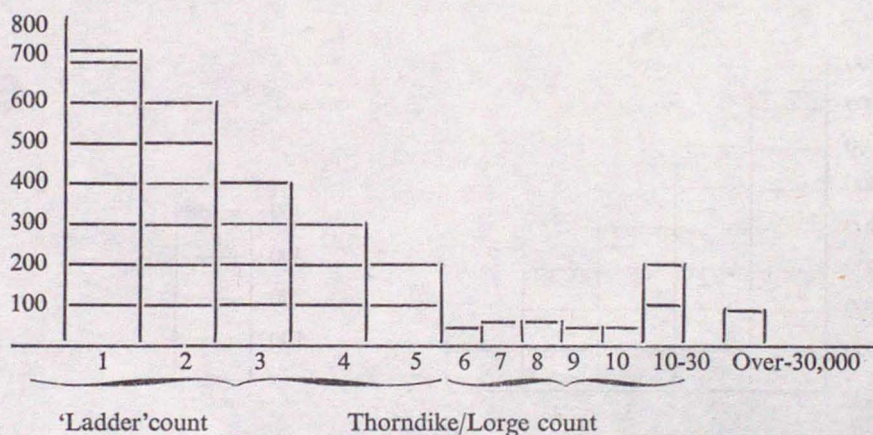


Fig 2. Complete word frequency in freshmen textbooks using combined 'Lodder' and Thorndike and Lorge counts

It is clear that the Ladder classification of the first five thousand words with their allied forms covers the vast majority of items in the sample.

A check of words coming *outside* the Thorndike and Lorge count (i.e. over 30,000) showed that the words were either highly specialist ones like EGO, ID, ECOLOGIST, SOCIOGRAM, GENE, etc., words occurring in a specialist text-book, usually several times - or words of more general application which occurred in the sample only once.

CONCLUSIONS

It is worthwhile in the freshman English programmes to concentrate on the words within the Ladder 5,000 range, as this covered nearly 85% of the words recorded in the sample.

Because the Ladder Dictionary includes, under headwords, derivations such as SUPPOSITION under SUPPOSE, STATISTICIAN under STATISTICS, derivations which fall outside the 5,000 of Thorndike and Lorge, but which can easily be grasped, English courses could usefully include more work on the modification of known roots.

A selection of the over-5,000 useful specialist words for *each subject* could be compiled, in liaison with the departments concerned, and distributed at the beginning of each semester to students taking those subjects. The data obtained in this study might be helpful in this, though particular words may have been missed in the sample and some further work in listing words of specialist difficulty might be carried out in the near future.

Data from this study and the follow-up suggested above could be consulted when examination passages and vocabulary test items are being devised. If an item occurred in more than one subject, once or several times, it would appear to be worth testing even if it were a low-frequency item in the Thorndike and Lorge count.

The Ladder series of readers should be promoted as useful reading material for improving vocabulary within an appropriate range, and the Ladder Dictionary would be a useful adjunct to such a reading programme.

BIBLIOGRAPHY

- Shaw, John with Janet Shaw. *The New Horizon Ladder Dictionary of the English Language*. New York: Signet Classics, The New American Library, 1970.
- Thorndike, Edward L. and Irving Lorge. *The Teacher's Word Book of 30,000 Words*. New York: Teachers' College, Columbia University, 1944.
- West, Michael. *A General Service List of English Words*. London: Longmans, 1953.

Data from the following textbooks:

—Broom, Leonard and Philip Selznick. *Sociology. A Text with Adapted Readings*. Fourth Edition. New York: Harper & Row, 1968.

—Jones, A.H.M. and Elizabeth Monroe. *A History of Ethiopia*. Oxford: Clarendon Press, 1935.

—Mesfin Wolde-Mariam. *An Introductory Geography of Ethiopia*. Addis Abeba: H.S.I.U. Press, 1972.

—Morgan, Clifford T. and Richard A. King. *Introduction to Psychology*. Third Edition. New York: McGraw Hill Book Company, 1956.

—Samuelson, Paul A. *Economics. An Introductory Analysis*. Fifth Edition. Tokyo: McGraw Hill Book Company Inc., Kogakusha Co-Ltd., 1961.

—Ullendorff, Edward. *The Ethiopians. An Introduction to Country and People*. Second Edition. London: O.U.P., 1965.

—Walsh, Joseph B., S.J. *Logic*. New York: Fordham University Press, 1940.