# WORD-LISTS IN ENGLISH

## PROBLEMS OF CONSTRUCTION

HERBERT DENNETT

Missionaries in several parts of the world and Christian workers at home have in recent years become increasingly interested in the use of a simplified form of English for either one or both of the following purposes.

I. The use of a simplified vocabulary in writing English texts for translation into local languages will naturally make things easier for the native helpers who cooperate with the missionary. This will apply whether it is a question of translating the Bible itself or 'helps' of various kinds. The native helper's knowledge of English will normally be limited, and he will, therefore, find it easier to suggest idiomatic equivalents in his own tongue from a simplified than from a normal English text.

2. In countries where English is the normal European language taught, there will be very many local Christians who could read Christian literature in simplified language, but who would be quite unable to grapple with texts in the normal literary idiom. This might be particularly true where people are literate only in English and not in their own tongue.

Any competent writer can, of course, produce a text in a deliberately simplified style. Work of this kind has been done, for example by (a) Mr. A. E. Horton of Angola in the Biblical field and (b) by Prof. C. S. Lewis in the secular field in his 'Narnia' stories for children. Both these writers, however, state that their choice of words was fairly subjective. It is not difficult to see that if a number of such writers, however competent, produced texts in this way, then the total number of words used in their combined writings would soon run into several thousands. This would defeat the very object in view of writing with a small but effective vocabulary, well within the power of new literates to learn easily.

The object of the present investigations is to develop word-lists for simplified English which enable material to be written on any reasonable subject with the minimum 'learning load' on the reader.

It should be made clear at this point that word-lists as here understood are quite distinct from the vocabulary of 'Basic English'. Simplified English is normal in both grammar and style, though in addition to using only common and well-known words, it will naturally employ only the simpler forms of sentence construction. Basic English by contrast is really an artificial language, for it dispenses with all but about eighteen of the tens of thousands of English verbs. As a result it must go round the bush to express some quite common ideas —as a random reading of the Basic New Testament will show. It is not that Basic English is of no use, but it is not the solution to the

problem of producing texts for translation, or for writing for new literates.

**A comparison of versions**

There have been several experiments in recent years in using a simplified vocabulary for translating the New Testament into English. One of the earlier ones was *The New Testament in Plain English* by C. Kingsley Williams (S.P.C.K.). This version is based on the 'Interim Report on Vocabulary Selection' (London, 1936). More recent versions are by Olaf Norlie (the entire New Testament) and Annie Cressman (Mark and John). Neither of these versions gives details of the word-list used. One interesting thing which emerges from a study of these versions is that simplicity is attained at the expense of word economy. Here is a comparison with other versions, taking Mark chapter 1 as a basis:

| Version | Total number of words | Percent 'simple' words |
|---|---|---|
| Cressman | 988 | 95.5 |
| Plain English | 920 | 93.2 |
| Norlie | 911 | 91.9 |
| J. B. Phillips | 962 | 91.5 |
| New English Bible | 870 | 91.4 |
| B.F.B.S. Translators' Translation | 831 | 90.0 |

The 'simple words' on the last column are those on a provisional list of one thousand which is the result of the research described in this article.

**Existing word-lists**

The idea of building a word-list for writing simplified English is far from being new, and for the purpose of the present research the following have been critically examined:

1. *Thorndike's English Dictionary* (The American work revised for Gt. Britain by Dr. P. Ballard—E.U.P.). The approximate frequency of words is here indicated by a code.
2. The *Interim Report on Vocabulary Selection,* (London, 1936), referred to above, attempts to number words in order of their alleged frequency.
3. Dr. F. Laubach in *Teaching the World to Read* (U.S.C.L.) includes a graded list of 1,600 words by Thorndike-Lorge.
4. Charles Duff, a prolific writer on language study, in *English for Adults* (E.U.P.) gives a graded list of 6,000 words.
5. Robert Gunning in *The Technique of Clear Reading* (McGraw-Hill) gives an ungraded word-list by Dr. Edgar Dale of Chicago University of 3,000 words 'known in reading by at least 80 % of the children in grade IV'.

6. *Key Words to Literacy,* Mc.Nally & Murray (Schoolmaster Publishing Company) gives a list of 300 words only, which are claimed to be the most suitable for teaching children and foreigners.
7. *The 300 Words* and *300 more Words* published in Hong Kong by the Council on Christian Literature for Overseas Chinese.
8. *Basic Vocabulary of Elementary School Children,* H. Rinsland. An analysis of some six *million* words written by American school children.
9. *A study of the vocabulary of young children,* G. E. R. Burroughs, Birmingham 1957. A much smaller scale oral investigation.

It is natural that these published word-lists should vary in size and purpose, but the disturbing thing about them is that no two agree on the importance of any but the most common words. For instance, the smallest list of all (McNally & Murray) includes 'apple', 'café', 'exit', 'police', 'station' to the exclusion of such words as 'also', 'begin', 'both', 'each', 'enough'—words of considerable operative value. Even greater contradictions in choice may be found in the larger lists. Furthermore one list will give its highest rating to words not included in another list at all.

From this it is evident that none of these lists is suitable as a basis for translation texts, or for writing simplified English for use in countries where English is a second language. A fresh approach to the problem is clearly needed.

**Problems of constructing a word-list**

The construction of an English Word-List is a surprisingly difficult task. It is no mere question of counting the frequency with which words are used in ordinary speech and writing, though naturally this must be taken into account. The 'area of meaning' of each word, its usefulness in word building by derivatives, [1] the existence of homonyms, [2] special meanings in the plural or in suffixed forms not

[1] The fundamental question raised is 'What is a word?'. The single word-form 'love' can be noun, verb or adjective, but the equivalent in other languages would be separate word-forms in most instances. Then what of 'loves', 'loving', 'loved'? Most would include these as inflexions of 'love'. Fair enough, but then how should 'lovely', 'beloved', 'loveable' be treated? The problem is where to stop, for in many other words the 'tails' are still longer: 'unaccountability', 'realization' etc. Not that these cumbersome forms would themselves appear on any word-list, but they are logical outgrowths of the root words. Many derivatives have a meaning quite distinct from that of the root word: 'good' ('goods', merchandise) 'new' ('news', information) 'feel' ('feeler', antenna) 'rub' ('rubber', eraser) 'wait' ('waiter', hotel servant) and so on.

[2] Homonyms are also a problem—one that is ignored in some word-lists. The word 'fine' will appear without any indication whether it is an adjective meaning 'high quality' or a noun meaning 'an indemnity'. It seems essential to give warning of homonyms of such words as: long, may, mean, rest, case, fair, stay, box, list, lock, fly, etc. Whether the less common homonyms should be indicated is another question. If so it would be for the benefit of the foreign reader. Examples are 'last' (shoe), 'bill' (of bird) 'painter' (boat-rope) 'game' (crippled).

implicit in the root word (e.g. odd/odds, stretch/stretcher), the possible synonyms of the word—all these questions affect the claim of a word for admission to the list of the most useful simple words in the language. Add to this the fact, peculiar to English, certainly among European languages, that a single word-form may be used in anything up to five different parts of speech (e.g. 'down', 'past', 'round'), and the complexities of English word-list building become evident.

**Methods of construction**

There are two ways in which word-frequency may be measured. The usual one is to register the total number of times a word occurs in a large number of diverse passages of writing. The result is called the 'raw frequency' of the words. To this plan there is the objection that a few relatively uncommon words may occur a considerable number of times in a particular passage, and so receive an unwarrantably high rating. Some of these words may not be found at all in any of the other passages used in the research.

The alternative and safer method is to register the number of *passages* in which a word occurs, irrespective of whether it occurs one or many times in a particular passage. Obviously the more common words will be found in almost every passage; others only in a few; borderline claimants for admission only in one or two passages. A passage of 1,000 words has been taken throughout as the unit of the present investigation, first for uniformity, secondly for ease of calculating percentages of various frequencies. Twenty-five such passages of 1,000 words each were taken as a start distributed as follows:

| | | | |
|---|---|---|---|
| Biography | 2 | History | 2 |
| Technical | 2 | Fiction | 5 |
| Commercial | 3 | General | 2 |
| Christian-general | 3 | Children's writings | 4 |
| Christian-technical | 2 | | |

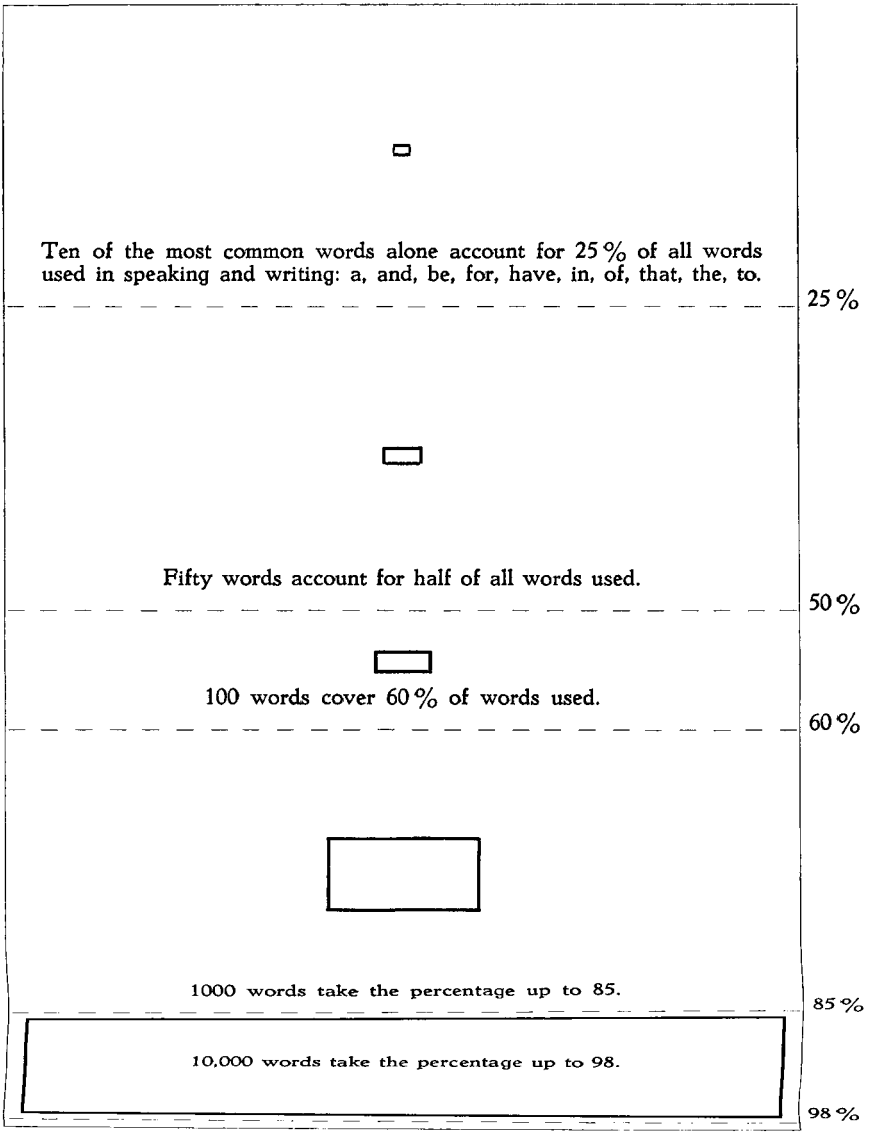
Children's writings included *The Wind in the Willows,* which is far from being childish in style. Christian Technical included an extract from *The Bible Translator.* A further 25 passages of 1,000 words each, all dealing with Christian subject-matter, were subsequently investigated. Up to the moment no oral matter has been included in the investigation, but it is planned to do so.

The present list, however, is not based solely on 'frequency counts'; the basis of selection is rather what a word can *do.* An attempt has been made to include all the 'operative' words needed to express ideas used in normal spoken and written communication.

**The English Word Pyramid**

It is common knowledge that about ten or so words account for about a quarter of all words used in writing and speaking English.

## ENGLISH VOCABULARY—THE WORD PYRAMID




In the above diagram each boldly printed rectangle represents the number of different words forming the respective percentage of total words used in normal speech and writing. All the remaining 400,000 or so words in English account for only 2 % of the total used.

Fifty of the most common words will push the coverage up to nearly 50 %, but another 50, making 100 in all will raise the percentage only to about 60. The most common 1,000 words will cover about 85 % of all words used, and 10,000 will reach up to 98 %. All the remaining 400,000 or so English words account for only the remaining 2 % of word-usage. In view of this severe operation of the law of diminishing returns, it would not seem to be economic to build up a word-list for simplified English beyond the one thousand mark.

**Few difficult words needed**

The size of the word-list must also be considered from the viewpoint of the reader for whom English is not the first language. A vocabulary of one thousand words, mostly short ones, is not an excessive learning-load. But what of the fifteen words or so in every hundred (if the 1,000 word list covers 85 % of all words used) not covered by the word-list? Of course many can be avoided or paraphrased in simplified English. For the rest, the answer is that they will not all be *different* words. In an article on the Resurrection, for instance, the words 'tomb', 'angel', 'disciple', 'earthquake' would probably not appear among the 1,000 most common words, but together they might well occur forty or fifty times in the 1,000 word article, so leaving quite a small number of new words to be tackled by the reader.

Simplified English has one thing in common with the many native languages with which missionaries have to do. An original text written directly in it by a 'native-speaker' is likely to be more effective than a 'translation' of a text from normal literary English. That is not to say that translation from normal to simplified English cannot be of any use. The Oxford Press issue a series of secular classics re-told in simple language, though unfortunately not based on a uniform vocabulary it would seem.

As an experiment a passage from a decidedly 'heavy' account of the French Revolution was taken. In it no less than 20 % of the words were not on the 1,000-word provisional list used in this research. In the 'translation' to simplified English only six words outside the list were used, and there was little loss of precision of meaning.

It would be foolish indeed to attempt to produce texts containing no words outside the special list, but it should certainly be possible to keep their number down to one or two percent of total words, and to explain the meaning of each strange word at its first occurrence.

In an early issue of *The Bible Translator* Mrs. F. N. Phair gives an example of what can be done to avoid difficult words. 'Why does this man blaspheme' is paraphrased 'Why does this man speak so; he is talking against God'.

It is only in recent years that the Science of Communication has arisen to teach us that it is the *reader* who matters more than the

writer, and that it is only *what goes in* that really counts in the passage of idea from one person to another.

The extent to which a simplified form of English can help in increasing the proportion of 'information conveyed' will be the measure of its usefulness in Christian communication.

---

*Since the above article was written the author has completed a very tentative draft of his word-list. Readers who are interested in putting this tentative list to a practical test are invited to write to: Mr. H. Dennett, 51 Oakwood Avenue, Boreham Wood, Herts., England. It should be noted that in compiling his list, the author has not selected the thousand English words that are most likely already known to literates with a limited knowledge of English; rather he has sought to select the thousand words which are of the highest 'operative' value, and which may be relatively easily learned. To this end he also plans to produce reader materials which teach this small but highly operative vocabulary. He also hopes to deal later with questions of syntax in simplified English.*

---

## THE STORY OF
# A POPULAR VERSION IN SWEDISH

ÅGE HOLTER

In 1962 a new translation of Mark into Swedish became the great sensation of the year. The newspapers and the weekly press, not only in Sweden but all over Scandinavia, had big headings such as: "The Shock Bible", "The Beatnik Bible" or "The Slang Bible". Ten thousand copies were sold in a few days, and a new edition appeared very shortly. A best-seller was born. And inquiries from distant parts of the world indicate that rumours of it spread like wild-fire.

In spite of this, the history of this new translation is an unhappy one. The sensation it created seems to have served very well the reaction against any attempt at adapting the Bible text to modern readers. Any church and any Bible Society would have declined with thanks this sort of publicity. To Christian people in Sweden it has been an embarrassing experience. How are we to explain these seemingly incompatible facts?

It all started with the excellent idea of pastor Grönlund to translate the gospel of Mark into the spoken every-day language of Swedish teenagers. He had especially in mind a certain group of very bad repute ("raggarne", known for alcoholism, sexual promiscuity and aggressive behaviour). He felt that something radical had to be done in order to make the message relevant to these young people, whose world is the very opposite of the Biblical reality. Here is one of his statements: "Take for instance a key-word like 'gospel'. Does that word convey anything at all to a fourteen year old boy in a Stockholm suburb? If it means anything to him, it would rather be associated with something dull. But probably it means less than nothing. Yet this