

HOW LARGE IS THE ENGLISH LEXICON?

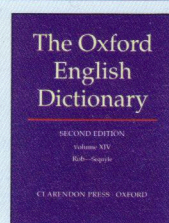
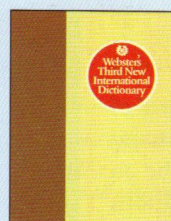
The two biggest dictionaries suggest around half a million lexemes – a total approached by the unabridged *Webster's Third New International* (which claimed over 450,000 entries in 1961) and by the integrated edition of the *Oxford English Dictionary* (which claimed over 500,000 entries in 1992). The true figure is undoubtedly a great deal higher.

A comparison of these two dictionaries – or of any other group of dictionaries of comparable size – shows a remarkable lack of identity between headword lists. In the sample analysed here (see right), the *Webster* and *Oxford* dictionaries have only 21 headwords in common out of a possible 57 – less than two-fifths. If this pattern were continued, their combined lexicon would exceed three-quarters of a million.

Discrepancies are usually caused by differing editorial emphases. The *Oxford* has far more historical references and British dialect items than does the *Webster*, which in turn has far more local American items. On the other hand, neither work would claim to be comprehensive in its coverage of the vocabulary of the 'new Englishes' (Part V) in such parts of the world as India, Singapore, and Nigeria, where thousands of new lexemes are coming into the language. And because the tradition in lexicography is to use the written language as the test for inclusion (p. 442), much local spoken nonstandard vocabulary will be omitted. There must be thousands of slang expressions currently in common use which have never been recorded, such as all the lexemes which express the concept of 'being drunk' – *canned*, *blotto*, *squiffy*, *jagged*, *paralytic*, *smashed*, etc.

Even if we restrict the issue to standard vocabulary, there are many items which could be included as part of the lexicon, but which are not usually found in a dictionary. There are some half a million abbreviated forms in English (p. 120), many of which have a clear lexical status (*BA*, *FBI*, *NATO*, etc.); and fauna and flora also provide a vast lexical resource. For example, there are apparently some million insects already described, with several million more awaiting description. This means that there must be at least a million designations enabling English-speaking entomologists to talk about their subject. Should all of these be allowed into the word-count as well?

It is difficult to see how even a conservative estimate of English vocabulary could go much below a million lexemes. More radical accounts, allowing in all of scientific nomenclature, could easily double this figure. Only a small fraction of these totals, of course, is learned by any one of us (p. 123).



| | | |
|-----------------------|---|---|
| saba | • | |
| sabadilla | • | • |
| sabadillia | | • |
| sabadilline | | • |
| sabadine | • | • |
| sabadinine | • | |
| sabaeen 1 | • | • |
| sabaeen 2 | • | • |
| sabahdaur | | • |
| sabai grass | • | |
| sabaism | | • |
| sabakha | • | |
| sabal | • | • |
| sabalo | • | • |
| sabalote | • | |
| sabal palmetto | • | |
| sabana | • | |
| sabaoth | | • |
| sabarcane | | • |
| sabate | | • |
| sabathé's cycle | • | |
| sabatia | | • |
| sabatier | | • |
| sabatille | | • |
| sabatine 1 | | • |
| sabatine 2 | | • |
| sabaton | • | • |
| sabayon | • | • |
| sabba-day | | • |
| sabbat | • | • |
| sabbatarial | | • |
| sabbatarian 1 | • | • |
| sabbatarian 2 | • | |
| sabbatarianism | • | • |
| sabbatary | | • |
| sabbath | • | • |
| sabbathaism | | • |
| sabbatharian | • | • |
| sabbatharie | | • |
| sabbath day | • | • |
| sabbath-day house | • | |
| sabbath-day's journey | • | |
| sabbathine | | • |
| sabbathize | | • |
| sabbathless | • | • |
| sabbathly 1 | • | • |
| sabbathly 2 | • | • |
| sabbath school | • | |
| sabbatia | • | • |
| sabbatian 1 | • | • |
| sabbatian 2 | • | • |
| sabbatianism | | • |
| sabbatic | | • |
| sabbatical 1 | • | • |
| sabbatical 2 | • | |
| sabbatically | | • |
| sabbaticalness | | • |

THE SUPERDICTIONARY

This comparison of the boldface items listed in two unabridged dictionaries shows the surprisingly limited extent of their overlap; and if we looked in addition for correspondence between senses, the extent of the discrepancy would be even greater. Nor is this the whole of the English lexicon at this point in the alphabet. Reference to *Chambers English Dictionary* (a much shorter work) brings to light another five items – *saba* (in a different sense from the one given in *Webster*), *sabahan*, *sabbath-break*, *sabbath-breaker*, and *sabbath-breaking*. Reference to Willis's *Dictionary of the Flowering Plants and Ferns* (8th edn) gives *sabalaceae*, *sabalacineae*, *sabaudia*, *sabaudiella*, *sabazia*, *sabbata*, and three senses of *sabbatia*. We have reached over 70 items now, with many other specialist dictionaries left to consult – but will anyone ever have enough time and motivation to consult them all, for the entire alphabet, and thus arrive at a truly complete superdictionary? Until someone (or, we must suppose, an electronic something) does, estimates about the size of the English lexicon will remain pure guesswork.

The dictionaries handle the capitalization of lexemes in different ways, and several items are variable in their use of capital letters. To avoid complicating the issue, no capitals are shown in the table. Alternative spellings, likewise (e.g. *sabaeen* and *sabeen*) have been ignored, as have optional hyphenations (e.g. in *sabbath day*).