HOW BIG IS ENGLISH?

CHRIS COLE
Newport Beach, California

> *The English language has a complement of somewhere between two million and three million "short" words ...*
>
> Dmitri Borgmann, *Beyond Language* (Scribner's, 1967), p. 226

Since the largest unabridged dictionaries contain far fewer than one million entries, what did Dmitri Borgmann mean by this claim? Could he be referring to the theoretically allowed piling up of prefixes and suffixes (e.g., if "countercountermeasures" is a word, why not "countercounter-counterconterme asures"?). Probably not, since he refers to "short" words, and while he did rail against the omissions of dictionaries, his example words were of a more common sense variety (e.g., "ex-wife"). No, I believe that even without such artificial constructions, Borgmann believed there were over two million words in English. Let's take a look at what the editors of the unabridged dictionaries had to say about the completeness of their word lists.

W.T. Harris, Preface to Webster's New International Dictionary of the English Language, G.& C. Merriam, Springfield, Massachusetts, 1909, p. x:

> Counting together the main words, with their inflections and derivatives in so far as they involve peculiarities of meaning, and the combinations of compound words and phrases, the present vocabulary has more than double the number of entries included in the previous edition, that of 1900. According to an accurate count, the number of words printed in bold-faced type, together with the inflected forms that appear in small capitals, totals more than four hundred thousand…
>
> The general one-volume dictionary appeals to a public universal in its interests, and the vocabulary of such a book must therefore be equally universal in its scope. The selection of the terms for such a vocabulary is one of the most difficult and laborious tasks involved in lexicography. The number available is always far in excess of, and for a one-volume dictionary many times, the number possible to be given. The reviser in making his selection is guided, broadly speaking, by utility; that is, he aims to include the terms most likely to be looked for by the consulter of the dictionary.

Introduction to Webster's New International Dictionary of the English Language, Second Edition, Unabridged, G.& C. Merriam, Springfield, Massachusetts, 1934, p. vii:

> The *comprehensiveness* of the new edition has two aspects—the size of the vocabulary and the fullness of treatment. No other dictionary has approached the new Merriam-Webster in the number of entries [600,000]; and yet the entries comprise only a selection from a much larger collection of terms. Space has not been wasted on words and phrases which are too technical, too rare, too ephemeral, or too local, or which are self-explanatory.

Philip Babcock Gove, Preface to Webster's Third New International Dictionary of the English Language, Unabridged, G.& C. Merriam, Springfield, Massachusetts, 1961, p. 7a:

> This dictionary has a vocabulary of over 450,000 words. It would have been easy to make the vocabulary larger although the book, in the format of the preceding edition, could hardly hold any more

pages or be any thicker. By itself, the number of entries is, however, not of first importance. The number of words available is always far in excess of and for a one volume dictionary many times the number that can possibly be included.

Victoria Neufeldt, editor of the Webster's New World family of dictionaries, quoted in Kenneth F. Kister, Kister's Best Dictionaries for Adults and Young People, A Comparative Guide, The Oryx Press, Phoenix, Arizona, 1992, p. 79:

I hate the word "unabridged." It's stupid and misleading, since it is used for all large dictionaries, regardless of whether an abridged edition of a given dictionary exists; and also, because the word sort of implies the idea of completeness, it encourages the buyer to believe that the dictionary so described contains all the words of the language. No dictionary comes anywhere near doing that.

If we take these editors at their words, then the total lexicon of English words from which they selected the words for inclusion in their "unabridged" dictionaries was many times larger than the included lexicon. What kinds of words are left out? Words that are "too technical, too rare, too ephemeral, or too local, or which are self-explanatory." Each of these exclusions is understandable, except for the last. What does it mean for a word to be self-explanatory?

An example is the word 'airvent,' which does not occur in any dictionary since its meaning is obvious from its parts. Dictionary editors quite reasonably do not want to waste their limited space on such words. This may change in the era of the on-line dictionary, but that is not our topic. The upshot is that even the largest unabridged dictionaries do not contain all of the words in English.

So, how big is English? The editors state that English contains "many times" the number of words in their dictionaries. Since these dictionaries contain over 500,000 words, the range two to three million may well be correct. As a spot check, consider all words starting with the word "air." There are about 250 such words in all unabridged dictionaries combined. A sample of the Internet reveals that there are about 700 closed words beginning with "air," about 900 hyphenated words, and about 1500 open compounds. Since English is pretty fluid about switching between open and closed compounds, the "true" number of English words beginning "air" is somewhere around 1000. Thus four is probably a good estimate for the ratio between the size of English and the lexicon of the unabridged dictionaries. This implies somewhere between two and three million words in the English lexicon, in line with Borgmann's estimate.