

### 30. Worldwide Words

#### How Many Words?

Common questions for the Q&A section:

How many words there are in the English language?

What is the average size of a person's vocabulary? These questions sound easy; they're indeed easy to ask. But they're almost impossible to answer satisfactorily, because it all depends on what you mean by 'word' and by 'vocabulary' (or even English).

What we mean by word sounds obvious, but it's not. The verb 'climb' for example, has many forms such as climbs, climbed, climbable, and climbing, climber (noun) climb-down climbing frame, climb on, climb over and climb down. A question: are all these distinct words, or all come together under 'climb'?

The entry for 'set' in The Oxford English Dictionary runs to 60,000 words. The noun alone has 47 separate senses listed. Are all these distinct words?

What do you include in your list of words? All the variations of English, Or slang, Dialect, Family or private language, Proper names and names of places, abbreviations, informal and formal names for living things. The wood louse is known in Britain by many local names- tiggy-hog, cheeselog, pill bug, chiggy pig and rolypoly among others. Are these all to be counted as

separate words? *Saccharomyces cereviae* is the formal name for bread yeast, is it counted as a word or two?

The other difficult term is vocabulary. What counts as a word that somebody knows? Is it one that a person uses regularly and accurately? Or perhaps one that will be correctly recognized-say in written text-but not used? Or perhaps one that will be understood in context but which the person may not easily be able to define? This distinction between is called active and passive vocabularies.

English speakers not only know words, they know word forming elements, such as the ending -phobia for some irrational fear. A journalist rushing to meet a deadline might take a word he knows, like *serb*, and tack on the ending to make *serbophobia*. He's just added a word to the language (probably only temporarily), but can he easily be said to have that word in his vocabulary? If nobody ever uses it again, can we legitimately count it? The total number of words in the English language and in person's vocabulary are so difficult to make, and they vary so much from one another. David Crystal, in the Cambridge Encyclopedia of the English Language, suggests that there must be at least a million words in the language. Tom McArthur, in the Oxford Companion to the English Language, comes up with a similar figure. David Crystal further says that if you allow all scientific terms the total could easily reach two million (this does not count the formal names for organisms).

Assessing the size of the vocabulary of an individual is at least as problematical. Take Shakespeare: you'd think it would be easy to assess his vocabulary. We have the plays and sonnets and we just have to count the words in them (according to the American Heritage Dictionary, there are 884,647 of them, made up of 29,066 distinct forms, including proper names). But estimates of Shakespeare's vocabulary vary from about 18,000 to 25,000 in various books, because writers have different views about what constitutes a distinct word.

It's common to see figures for vocabulary quoted such as 10,000-12,000 words for a 16-year old, and 20,000-25,000 for a college graduate. These seem not to have much research to back them up. Usually they don't make clear whether active or passive vocabulary is being quoted, and they don't account for differences in lifestyle, profession and hobby interests between individuals.

David Crystal describes a simple research project- using random pages from a dictionary- that suggests these figures are severe underestimates. He concludes that a better average for a college graduate might be 60,000 active words and 75,000 passive ones. But this method of vocabulary counts dictionary headwords only; it would be possible to multiply it several-fold to include different senses, inflected forms, and compounds. Another assessment- of a million-word collection of American texts- identified about 38,000 headwords. Bearing in mind this was all

general writing, this does not sound so different from David Crystal's estimates for graduate vocabularies.