COMMENTS ON SOME COMMON WORDS

ROBERT J. WILSON
Yonkers, New York

Quick now, what's the most common non-Germanic word in the English language? The most common non-Germanic noun? Verb? Preposition? The most common word from Greek? From Celtic? From Indic? From a non-Indo-European language? While you are thinking, I decided to examine the Word Frequency Book by John B. Carroll and others to discover the answers to some of these questions. Keep in mind that Carroll's book is based on writing, not speech, in American English. Since Carroll does not distinguish among homographs, I also consulted W. Nelson Francis's and Henry Kucera's Frequency Analysis of English Usage to ascertain the commonest verbs, nouns, etc. This book is derived from Kucera and Francis's Present-Day American English, which can be compared with Carroll for overall word-frequencies (not distinguishing among parts of speech). Not surprisingly, these two references disagree to some extent as to what the commonest words in English may be.

The most commonly-used words are overwhelmingly Germanic; many have been in the language since its very beginnings. The commonest ten are:

Carroll: the, of, and, a, to, in, is, you, that, it
Kucera & Francis: the, of, and, to, a, in, that, is, was, he

According to Carroll, the most frequent seventy-eight words are Germanic. Seventy-five of them are native to English with three borrowings from Old Norse (they, their, them). Carroll claims that ninety-five of the top one hundred words are Germanic; Kucera and Francis, that one hundred are.

The frequency of the makes a somewhat uncommon sound common, namely the voiced th represented phonemically by /ð/ (compare the voiced th in either with the unvoiced th in ether). The most frequent hundred words contain most of the initial voiced /ð/ words in English:

the, that, they, this, there, their, them, then, these

Those outside the top hundred:

those, though, thus, thence, thee, thou, thy, thine

The five most frequent nouns of all are:

Carroll: time, people, way, water, words
Francis & Kucera: man, time, year, state, day

Carroll counts singular and plural forms of nouns separately, but Francis and Kucera lumps them together into a single count. "Words,
words, words" says Hamlet; word is almost unique among English
nouns in having its plural form at least as common as its singular
(year is another).

The five most frequent verbs, excluding be and the modal auxili­
aries (have, do, can, will, would, could, shall, should, may,
must) are:

Carroll: said, see, make, made, find
Francis & Kucera: say, make, go, take, come

Carroll counts as different verbs separate forms such as say and
said, whereas Francis and Kucera lumps these together as a single
verb.

The five most frequent non-Germanic words are:

Carroll: people, use, very, just, used
Kucera & Francis: because, just, people, very, used

People is most solidly a noun, though not always; use is far more
often subject to functional change or conversion. The inclusion of
used points to use as the most frequent non-Germanic verb. Not
to be overlooked is its occasional use as a modal auxiliary in
the form used to. Both very and just share adjective and adverb
functions, although very as an adjective meaning "true, exact,
actual" (the only way Chaucer used the word) might strike one
as used only in formal contexts. Conversely, just as an adverb
meaning "barely" might strike one as somewhat informal. We place
a heavy syntactic burden even on our borrowed words. Because
rounds out this section, showing that some borrowings have made
deep inroads into what is now considered a closed system of prepo­
sitions and conjunctions, usually reserved for native (that is, Ger­
manic) words.

Breaking down the most frequent non-Germanic words into parts
of speech, we get the five most frequent nouns:

Carroll: people, part, place, number, air
Francis & Kucera: state, people, school, number, part

The five most frequent non-Germanic verbs:

Carroll: use, used, form?, study, try
Francis & Kucera: use, turn unite, try, move

School and air are probably the commonest English words derived
from the Greek, and change appears to be the commonest from Indo­
European's Celtic branch (which has given us whiskey among other
fine and dangerous words). The most frequent word from Hebrew
may well be the given name John. Arabic gives us the common words
coffee, sugar and cotton, and Chinese, tea. Oddly, the most fre­
quently-encountered word from Indic, the oldest branch of Indo­
European, is Indian!

The difference between Carroll and Kucera & Francis is most dra­
matically illustrated with hyphenated words:

Carroll: good-by, twenty-five, well-known, man-made, twenty-four,
grown-up, old-fashioned, far-off

Kucera & Francis: anti-trust

Only one word does not appear in the other hyphenation: hallelujah.
The five ending forms are: hallel, hallelujah, hallelujas, hallelu­
jahs, hallelujahing. The inclusion of these forms is due to mis­
spelling.

The compounding forms are: United States, Orleans, North Carolina.

This category of words is relatively larger than others. The word
United States is probably the commonest. One must be careful not
to overlook the frequent occurrence of the hyphen in the names of
place names.

Hyphemes (oddly present in some present day dictionaries) are hy­
phemas widely used in the English language. Proper nouns, a
word that is not hyphenated, is the one most often seen in
the English language. The word is hyphenated in many
frequent words, such as street probably.
Kucera & Francis: long-range, over-all, long-term, so-called, anti-trust, part-time, Bang-Jensen, twenty-five

Only one word, twenty-five, is common to both lists. Note that goodbye does not have the usual stress pattern of compounds that the other hyphemes show: stress on the first element with strong secondary stress on the second element. The OED comments that goodbye (the OED spelling) is a "contraction of the phrase "God be with you (or ye)." Hence the second element is a blend itself. Later forms are god be wy you and god b'(o)y you (my insertion of stress throughout). Perhaps the weakening of initial stress can be explained by the (linguistic crutch of) analogy of good-b'y with halló, helló; these latter two earlier underwent stress shift themselves. Perhaps they have also influenced similar phatic expressions as good night and good day.

The compilers of the Word Frequency Book wisely included only place names as open compounds. The most frequent ten:

United States, New York, New York City, San Francisco, New Orleans, Los Angeles, Great Britain, New Jersey, South Carolina, North Carolina

This category simply points to American sources for the corpus. Similarly, Kucera and Francis encountered Providence more frequently than Boston, Philadelphia, Los Angeles, and Detroit, and almost as often as Chicago (they are professors at Brown University). One must take frequency-counts for hyphemes and placenames with a grain of salt.

Hyphemes and open compounds present orthographic (but not decisive) evidence for compounding. Solidemes or closed compounds present difficulties since no such evidence is adduced; one must rely also on phonological and semantic evidence. Moreover, compounding, adding a base to a base, is difficult at times to distinguish from affixation, i.e., adding a prefix, infix, or suffix to a base. Yet the most frequent historical compounds (such as another) are perceived as unbounded morphemes by the average guy in the street. I think I will continue looking for the least frequent word in the English language. That average guy in the street probably knows it.