

WORD LISTS WITH ALL STARTING LETTERS

A. ROSS ECKLER

Morristown, New Jersey

How large must a word list be in order to have a reasonable chance of including words starting with every letter of the alphabet? Because words starting with X are far less likely to occur than others, this question essentially reduces to the probability that a word list of length n contains one or more x-words. If one draws at random n words from a much larger list in which a fraction p are x-words, then this probability is equal to $1 - (1-p)^n$, where p^n is p raised to the n th power.

To get some idea of the list size needed, consider the 186,574 words in the *Word Game Winning Dictionary* (1984) compiled by Tom Pulliam and Gorton Carruth. This includes 217 x-words, for a p of 0.00116. If one raises 0.99884 to the 608th power, one obtains 0.5, suggesting that a list of this size has a fifty-fifty chance of including an x-word.

This number must be taken only as an approximate guide, for the probability of an x-word varies considerably according to the population under consideration. For example, the US Post Office has compiled a state-by-state directory of towns with zip code numbers. Only three states—Kansas (808), Ohio (1375) and Illinois (1521)—have lists containing all starting letters. Pennsylvania, with 2396 names, is missing an x-word, a probability with a value of only 0.06 according to the Pulliam-Carruth model; many other states of similar size are also bereft.

How short a list can one find which contains all starting letters? This is an example of textual logology, as discussed in the November 2000 Word Ways. The most well-known example of textual logology is the search for a pangrammatic window (a textual sequence containing all letters of the alphabet). In a small search in his library, the author discovered that *Admiral of the Ocean Sea: A Life of Christopher Columbus* (1942) by Samuel Eliot Morison, has an inclusive index of 461 items. *Secret and Urgent: The Story of Codes and Ciphers* (1942), by Fletcher Pratt, has an index of only 314 items including an x-word (Xerxes), but—alas!—none beginning with the letter I.