FOUR-LETTER WORDS

JEZEBEL Q. XIXX

We are addicted to using four-letter words.

There are so many of these words that it is impossible to communicate effectively without them. Cryptographic studies show that 31 of the 100 most frequently used words are four-letter ones. A baker's dozen, drawn from the top 60: FROM, HAVE, MORE, SOME, THAN, THEM, THEY, THIS, UPON, WHAT, WHEN, WITH, YOUR. These particular examples are "pure" words, each one using four different letters, and must be distinguished from "contaminated" specimens in which a letter is repeated (THAT, WERE, BEEN and others).

If we take the 26 letters of the alphabet and arrange them in an entirely random linear order, a careful examination of the letter sequence before us will disclose various common words hidden in the sequence: words spelled in correct order, though seldom with consecutive or even nearly consecutive letters of the sequence.

For illustrative purposes, let's consider the most popular of all letter orders, the alphabetic order:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Interested scrutiny of this particular letter sequence uncovers words such as AHoy, AIRy, BEvY, BLOw, CHiN, CRUX, DEFy, DIRT, ECiS, ERST, FILM, FLUX, GIST, GORY, HiNT, HOST, IMPS, KNOT, KNOW, LOPS, LOST, MOPS, MOST, and NOsy. All of these are quite common words. There are some other equally common words that we have failed to point out. How many of them can you spot?

In a typical, random letter arrangement, several dozen common or near-common four-letter words can be discovered. Those of us with a logological passion must inevitably wrestle with two questions: which one of all possible letter sequences will yield the smallest number of four-letter words, and which one the largest such number?

The simple, obvious way of finding out is to test each possible arrangement of the alphabet. Unfortunately, there is an insurmountable roadblock in the way. The 26 letters of the alphabet may be arranged in more than 403 septillion different ways. If we had access to a supercomputer capable of tossing different arrangements to us at the rate of ten million sequences per second, and if we kept this super-computer in continuous operation for one trillion years, the list of possible letter
sequences would still not be complete. Since it may take from one hour to three hours to test even one letter arrangement, the time investment required to solve our two problems is prohibitive. There must be a more practical method!

Before considering ways and means, we must define what we are going to accept as allowable words. The unabridged dictionaries, praiseworthy as they are in many respects, include too many words alien to us both in meaning and in spelling. Since we are interested in reasonably common words, recognizable more or less at sight, let us make The New Merriam-Webster Pocket Dictionary our standard, accepting all regular four-letter words that appear in this dictionary in boldface type in the main alphabetical section (pages 1 through 592). In specifying regular words, we mean to exclude "crippled" words also in the dictionary: hyphenated and apostrophized words, parts only of hyphenated words and of multi-word phrases, multi-word phrases in their entirety, abbreviations, prefixes, suffixes, and acronym-like words. These restrictions will keep out of our lists terms such as CAN'T, CO-ED, SUEY (part of "chop suey"), AFRO (part of "Afro-American"), AS IF, SEMI-, -ENCY, AWOL, and, possibly, XMAS (which may be thought of as an abbreviated form of "Christmas").

The results obtained by using the pocket dictionary should be representative of ordinary English, undistorted by the many strange words included in much larger dictionaries.

The easier of our two problems is minimizing the number of four-letter words in a 26-letter sequence. It ought to be possible to produce an alphabet in which not a single tetragram can be found.

Reflecting on the problem, we realize that all words in which consonants and vowels are mixed can be excluded simply by placing all consonants first and all vowels last. This will prohibit movement back and forth between the two kinds of letters, and will also exclude all words beginning with vowels because there are no all-vowel words in the pocket dictionary.

Our next task is to exclude words beginning with two or more consecutive consonants. We achieve that goal by placing all consonants that normally appear as second consonants first: H, L, M, N, R, W. Odd words such as TZAR are few and far between, must be considered individually, and can quickly be eliminated. One consonant, the letter S, can be followed by other consonants such as C, K, P, and T. We solve that problem by placing S as the very last consonant.

Remaining to be considered are words spelled with one consonant followed by three vowels. There are only 7 such words in the dictionary: BEAU, BUOY, KAYO, LIEU, MOUE, QUAI, and ROUE. By noting the vowel arrangements in these 7 words, we can easily place the six vowels of the alphabet in such an order that none of the 7 words can be spelled out.

The results obtained by using the pocket dictionary should be representative of ordinary English, undistorted by the many strange words included in much larger dictionaries.

The easier of our two problems is minimizing the number of four-letter words in a 26-letter sequence. It ought to be possible to produce an alphabet in which not a single tetragram can be found.

Reflecting on the problem, we realize that all words in which consonants and vowels are mixed can be excluded simply by placing all consonants first and all vowels last. This will prohibit movement back and forth between the two kinds of letters, and will also exclude all words beginning with vowels because there are no all-vowel words in the pocket dictionary.

Our next task is to exclude words beginning with two or more consecutive consonants. We achieve that goal by placing all consonants that normally appear as second consonants first: H, L, M, N, R, W. Odd words such as TZAR are few and far between, must be considered individually, and can quickly be eliminated. One consonant, the letter S, can be followed by other consonants such as C, K, P, and T. We solve that problem by placing S as the very last consonant.

Remaining to be considered are words spelled with one consonant followed by three vowels. There are only 7 such words in the dictionary: BEAU, BUOY, KAYO, LIEU, MOUE, QUAI, and ROUE. By noting the vowel arrangements in these 7 words, we can easily place the six vowels of the alphabet in such an order that none of the 7 words can be spelled out.

The results obtained by using the pocket dictionary should be representative of ordinary English, undistorted by the many strange words included in much larger dictionaries.

The easier of our two problems is minimizing the number of four-letter words in a 26-letter sequence. It ought to be possible to produce an alphabet in which not a single tetragram can be found.

Reflecting on the problem, we realize that all words in which consonants and vowels are mixed can be excluded simply by placing all consonants first and all vowels last. This will prohibit movement back and forth between the two kinds of letters, and will also exclude all words beginning with vowels because there are no all-vowel words in the pocket dictionary.

Our next task is to exclude words beginning with two or more consecutive consonants. We achieve that goal by placing all consonants that normally appear as second consonants first: H, L, M, N, R, W. Odd words such as TZAR are few and far between, must be considered individually, and can quickly be eliminated. One consonant, the letter S, can be followed by other consonants such as C, K, P, and T. We solve that problem by placing S as the very last consonant.

Remaining to be considered are words spelled with one consonant followed by three vowels. There are only 7 such words in the dictionary: BEAU, BUOY, KAYO, LIEU, MOUE, QUAI, and ROUE. By noting the vowel arrangements in these 7 words, we can easily place the six vowels of the alphabet in such an order that none of the 7 words can be spelled out.
from one the time in
the. There
what we are
onaries,
any words
interested in
sight. let us
standard. ac-
parts only
cronym-like
such as
part of 11 Afro-
ll ) •
ould be re-
number of four-
ible to pro-

h which con-
placing all
also exclude
-vowel words

or more con-
sons N, R, W.
considered
ment, the letter
T. We

the consonant
the diction-
UE. By not-
7 words

These reflections permit us to construct a letter sequence such as

H L M N R W B C D F G J K P Q T V X Z S Y E I U A O

Study this arrangement diligently. If you can find any four-letter word
in it, allowable under our rules, telegph the editor of Word Ways
immediately, requesting a plaque certifying you as a member of the
Logological Hall of Fame!

The second problem, that of maximizing the number of four-letter
words that may be read in an alphabetic sequence, is much the more
difficult one. To begin with, we do not know what word total is to be
achieved, except for the vague feeling that there must be some letter
sequences permitting "hundreds" of tetragrams to be formed.

One way of attacking the problem would seem to be arranging the
letters in order according to the relative frequency with which they
appear as first letters of four-letter words. Another way would seem
to be arranging them in reverse order according to the relative fre-
quency with which they appear as fourth letters of such words. Using
published positional word lists compiled from the pocket dictionary,
it is a simple matter to count the number of eligible words that begin
or end with each letter of the alphabet. Such counts establish that S,
B, and C are far and away the most common initial letters, while E,
D, and T are conspicuously the most common terminal letters. This
suggests that our letter sequence should begin with S B C ... and end
with ... D T E.

These two approaches are both one-sided ones. To close in on our
objective, it might be best to combine them, particularly as regards
the 20 letters in between our tentative start and finish. Thus, if a let-
ter ranks 7th on our initial list and 21st on our reverse terminal list,
perhaps it ought to be assigned the 14th position on a combined tentative
list. The rare letters of the alphabet (J, Q, X, Z, K, V) can be stud-
ied as a separate phenomenon, and shunted to the front end or the tail
end of the sequence depending on whether there are more words begin-
ning or ending with them.

The results of our combined approach will tend to place most of the
vowels (A, I, O, U) near the middle of the sequence. Their relative
order is determined by inspecting the common vowel digraphs in En-
lish: AI, AU, OA, OI, OU, UI. Each of these occurs with far greater
frequency than its reversal. There is only one relative vowel order
that incorporates all 6 of these digraphs: O A U I. Accordingly, this
particular order must be maintained, irrespective of whether or not
any consonants are eventually interspersed among these vowels.

Some relative letter orders cannot be predicted solely on the basis
of a priori considerations. Are there more words with the digraph IM
or the digraph MI? The only way is to count the number of each that
appear in pocket dictionary four-letter words. In this case, IM gets
Very careful analysis of all pertinent considerations bearing on letter order finally produces the following suggested alphabet sequence:

BSFPWCHJQMOAVUIRNGKLTDZEXY

There are many possible switches that can be made in this sequence without affecting the number of four-letter words that it will yield. For instance, the first three letters can be juggled in any of the mathematically possible 6 orders with no effect on the result. Similarly, the four letters H J Q M can be arranged in 24 different ways, and the three letters K T D can be arranged in 6 different ways.

How many eligible words can be read in this sequence? There are 398 given in the list below:

bade bore silt swag fund pore whir char
bail born sing swan funk pork whit chat
bait bout sink swat furl port whiz chid
bake bulk sire sway fury pour wide chin
bald bung site swig fuze pout wild chit
bale bunt size fade page prey wile chon
balk bunt smog fail paid puke wilt chug
band burn smug fain pail pule wily cite
bane bury smut fair pain punk wind city
bang sage soak fake pair punt wine coal
bank said soar fane pale puny wing coat
bard sail soil fang pane pure winkle coaX
bare sake sold fare pang purpIe wire code
bark sale sole fate pant wade wiry col1
barn salt song faun pard wage woad coin
bate sand sore faze pare wall woke coke
blde sune sort file park wain wold cold
bike sang soul find part wait wont colt
bile sank sour fine pate wake word cone
bilk sate span fire pave wale wore conk
blind save spar flex phiz walk work cony
bird scan spat foal pike wand worn cord
bite scar spay foxy pile wane wort core
boar Scot spin foil pine want wove cork
boat scud spit fold cling ward cage corn
bode shad spot folk pink ware cagy cote
body shag spry fond pnt warn cake cove
bogy shin spud font pitty want calk crazy
boll shod spun forT pixy wary cane cult
bold shoe spur fore poke wave cant curd
hale shot suit forT poky wavy card cure
bolt shun sulk furt pole waxy care curl
bond shut sung foul pond what cart curt
bone side sunk forT pone whey cave cute
bony silk sure foxy pony Whig cavy hail

It is impossible, however, to list any of the words of length five or more. It should be noted that the above list includes the substantive letter sequences for all the four-letter words in the dictionary.

Readers are invited to extend the above search through the volumes of Augustus H. Blaine's Dictionary of English Language and the pocket edition of the Webster Standard Dictionary, which contain over 3,000,000 words.

U I A

Can a letter be added to the list above? How about a whole word? A whole word is admitted of course.
It is interesting to note that not a single word in the list begins with any of the last 11 letters in our arrangement, and that 380 of the 398 words begin with one of the first 10 letters in the sequence: the next 5 letters contribute only another 18 words. It is also of interest to note that none of the 398 four-letter words are spelled out by consecutive letters of the sequence; MORN and MAIN, each interrupted by four extraneous letters, are the most compact.

Readers are invited to try altering our letter sequence so as to increase the number of words of four letters it yields.

COMMENT (A. Ross Eckler): Ms. Xixx has done a fine piece of research on a problem originally introduced by Dave Silverman in the August 1971 Kickshaws. He reported there that Darryl Francis, using an unspecified word-source (but one larger than the Merriam-Webster Pocket Dictionary), had constructed a letter sequence in which only the word LAOS appeared:

UIAOEYSDBPMHTGNLRKCFWVJQXZ

Can a letter sequence by constructed which bars all three-letter words from the Pocket Dictionary? The answer appears to be no, but one can come surprisingly close. The sequence

UAOEYJQXVGMHKTPSZDBFCNRLW1

admits only the word OBI, which appears in the current edition of the Pocket Dictionary, but not in some older editions.