CROSSWORD PUZZLE LETTER FREQUENCIES

JOHN D. HITCHCOCK
Laramie, Wyoming

In 500 crossword puzzles syndicated by the Newspaper Enterprise Association, and published in the "Laramie Daily Boomerang" six days a week between October 24, 1976 and June 11, 1978 (about 20 consecutive months), the number of times each letter of the alphabet was used was tabulated. The statistical results are summarized in the table on the next page, and discussed in more detail below.

Every puzzle consisted of 14 horizontal rows and 13 vertical columns. The total number of horizontal words in a puzzle varied from 28 to 39, and averaged 34.93; the total number of vertical words in a puzzle varied from 28 to 38, and averaged 34.52. The total words in a puzzle varied from 62 to 75, and averaged 69.35.

The average length of horizontal words was 4.15 letters, and of vertical words, 4.18 letters. Pratt (1942) and Haldane (1976) reported that in general writing the average word consists of 4.5 letters; thus, average word length is slightly shorter in crossword puzzles than in general writing.

The total letters used in a puzzle varied from 131 to 152 and averaged 144.36. Only 6.2 per cent of the puzzles contained more than 150 letters, and only 2.4 per cent contained less than 133 letters. The number of different letters in a puzzle varied from 17 to 26, and averaged 20.81. Less than 1.5 per cent of the puzzles contained more than 23 different letters or less than 18 different letters. Only one of the 500 puzzles contained all 26 letters and none contained 25 different letters.

The regular vowels A, E, I, O and U constituted, on the average, 43.0 per cent of the total letters. Moreover, their percentage was remarkably constant (42.9 to 43.2 per cent) in five subsamples of 100 puzzles apiece. Gaines (1939), Wolfe (1974), and Haldane (1976) reported that, in general writing, vowels average 40 per cent of all letters. Thus, the ratio of vowels to consonants is essentially the same in both crossword puzzles and in writing.

In the crossword puzzles studied, the letter E occurred more frequently than any other letter. It occurred from a minimum of 11 times to a maximum of 39 times in a single puzzle, and an average of 23 times per puzzle. On the average, it constituted 37 per cent of the regular vowels, and 16 per cent of all letters. Cryptologists all agree that in general writing also, E is the most frequently used let-
### Frequency of Use of Letters of the Alphabet in Crossword Puzzles

<table>
<thead>
<tr>
<th>Letter</th>
<th>Crossword Frequency</th>
<th>Text Frequency</th>
<th>Average Number of Times Each Letter Occurred in Each Puzzle</th>
<th>Per Cent of Puzzles With Letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>.1519</td>
<td>.1250</td>
<td>23.13 (11 to 39)</td>
<td>100</td>
</tr>
<tr>
<td>A</td>
<td>.1008</td>
<td>.0805</td>
<td>14.59 (4 to 25)</td>
<td>100</td>
</tr>
<tr>
<td>S</td>
<td>.0970</td>
<td>.0655</td>
<td>14.03 (4 to 24)</td>
<td>100</td>
</tr>
<tr>
<td>O</td>
<td>.0702</td>
<td>.0760</td>
<td>10.15 (2 to 24)</td>
<td>100</td>
</tr>
<tr>
<td>T</td>
<td>.0684</td>
<td>.0925</td>
<td>9.89 (2 to 20)</td>
<td>100</td>
</tr>
<tr>
<td>I</td>
<td>.0666</td>
<td>.0729</td>
<td>9.64 (3 to 19)</td>
<td>100</td>
</tr>
<tr>
<td>R</td>
<td>.0638</td>
<td>.0613</td>
<td>9.23 (2 to 18)</td>
<td>100</td>
</tr>
<tr>
<td>N</td>
<td>.0635</td>
<td>.0710</td>
<td>9.18 (1 to 18)</td>
<td>100</td>
</tr>
<tr>
<td>L</td>
<td>.0502</td>
<td>.0414</td>
<td>7.26 (1 to 23)</td>
<td>100</td>
</tr>
<tr>
<td>D</td>
<td>.0419</td>
<td>.0397</td>
<td>6.06 (0 to 16)</td>
<td>99.6</td>
</tr>
<tr>
<td>U</td>
<td>.0317</td>
<td>.0272</td>
<td>4.59 (0 to 17)</td>
<td>98.6</td>
</tr>
<tr>
<td>P</td>
<td>.0301</td>
<td>.0202</td>
<td>4.35 (0 to 17)</td>
<td>97.2</td>
</tr>
<tr>
<td>M</td>
<td>.0234</td>
<td>.0254</td>
<td>3.39 (0 to 18)</td>
<td>94.4</td>
</tr>
<tr>
<td>C</td>
<td>.0220</td>
<td>.0316</td>
<td>3.19 (0 to 10)</td>
<td>94.8</td>
</tr>
<tr>
<td>G</td>
<td>.0205</td>
<td>.0195</td>
<td>2.97 (0 to 10)</td>
<td>94.8</td>
</tr>
<tr>
<td>Y</td>
<td>.0176</td>
<td>.0172</td>
<td>2.54 (0 to 8)</td>
<td>92.2</td>
</tr>
<tr>
<td>B</td>
<td>.0174</td>
<td>.0154</td>
<td>2.52 (0 to 14)</td>
<td>84.8</td>
</tr>
<tr>
<td>H</td>
<td>.0147</td>
<td>.0542</td>
<td>2.13 (0 to 12)</td>
<td>88.8</td>
</tr>
<tr>
<td>K</td>
<td>.0091</td>
<td>.0066</td>
<td>1.32 (0 to 6)</td>
<td>71.0</td>
</tr>
<tr>
<td>W</td>
<td>.0090</td>
<td>.0188</td>
<td>1.30 (0 to 6)</td>
<td>73.8</td>
</tr>
<tr>
<td>F</td>
<td>.0085</td>
<td>.0231</td>
<td>1.25 (0 to 9)</td>
<td>60.0</td>
</tr>
<tr>
<td>V</td>
<td>.0075</td>
<td>.0100</td>
<td>1.09 (0 to 8)</td>
<td>63.8</td>
</tr>
<tr>
<td>Z</td>
<td>.0024</td>
<td>.0010</td>
<td>0.35 (0 to 3)</td>
<td>27.8</td>
</tr>
<tr>
<td>X</td>
<td>.0021</td>
<td>.0020</td>
<td>0.30 (0 to 3)</td>
<td>25.6</td>
</tr>
<tr>
<td>Q</td>
<td>.0008</td>
<td>.0011</td>
<td>0.12 (0 to 3)</td>
<td>10.0</td>
</tr>
<tr>
<td>J</td>
<td>.0008</td>
<td>.0016</td>
<td>0.11 (0 to 4)</td>
<td>7.6</td>
</tr>
</tbody>
</table>

Crossword frequencies are based on 500 crossword puzzles containing a total of 72,180 letters (J. D. Hitchcock, 1978)

Text frequencies are based on 4,744,003 letters of English-language text first printed in 1961 (H. Kučera and W. N. Francis, 1967, tabulated by A. Zettersten, 1969)
ter, and calculations from their data show that it constitutes about 12 to 13 per cent of all letters. The letter matrices for linotype, which reflect letter frequency, also show E as the most frequent (Kahn, 1968). Gaines (1939), Smith (1943) and Haldane (1976) summarize letter frequencies, not only in English, but also in French, German, Italian, and Spanish. In all these languages E is the most frequently used letter, but the frequencies of other letters vary considerably in the different languages. Gaines (1939) also gave letter frequencies in Portuguese, in which A slightly exceeded E in frequency, but the data were based on "probably military" text, rather than general literature. Zettersten (1969) gives letter frequencies based on the 4.5-million letter corpus of Kucera and Francis (1967).

The vowels A, E, I and O occurred at least once in every crossword puzzle, but U was absent from 1.4 per cent of the puzzles. The consonants S, T, N, R and L also occurred in every crossword puzzle.

In crossword puzzles, S was the most frequent consonant and the third most frequent letter (after E and A). By contrast, most cryptologists agree that in general writing T is the most frequent consonant, and the second most frequent letter; see, for example, Valerio (1893), Ohaver (1933), Gaines (1939), Pratt (1942), Smith (1943), Walsh (1946), Moore & Waller (1962), Laffin (1964), Zettersten (1969), Espy (1973), Hall (1974), Wolfe (1974) and Haldane (1976). In crossword puzzles, T was the fifth most frequent letter. In general writing S is usually said to be the seventh or eighth most frequent letter, exceeded in frequency by T, N and R.

In general writing, T is usually considered the most frequent initial letter, according to Ohaver (1933), Pratt (1942), Smith (1943), Laffin (1964) and Card & Eckler (1975), although both Vizetelly (1946) and Shipley (1946) reported that S is the most frequent initial letter. The final letter of a word is most frequently E, according to Ohaver (1933), Pratt (1942), Smith (1943), Laffin (1964) and Card & Eckler (1975). The latter article presents an excellent critical review of letter frequencies.

In crossword puzzles, the letter H was only the 18th in frequency, and occurred an average of only two times in each puzzle. It did not occur at all in 11.2 per cent of the puzzles. In general writing, H is usually stated to be the 9th in frequency, according to Valerio (1893), Ohaver (1933), Gaines (1939), Pratt (1942), Smith (1943), Moore & Waller (1962), Laffin (1964), Wolfe (1974) and Haldane (1976). Miliken (1942) reported H as 11th in frequency. Espy (1973) quoted the World Book Encyclopedia as claiming H to be 26th in frequency. The word THE, which was recorded for only the last 200 crossword puzzles studied, occurred only five times among 13,818 words, or less than 0.04 per cent, whereas Kucera & Francis (1967) determined that in present-day American English writing THE occurred nearly 7.0 per cent of the time. A similar computer study by Carroll, Davies & Richardson (1971) showed that THE constituted about 7.3 per cent of over five million words. (Incidentally, Berger (1977) determined that in
spoken English, THE is the third most common word, and constitutes about 3 per cent of all words.) A critical review of the Kucera and Francis study was made by Eckler (1978).

Among the fifty most commonly used words, found in the lists of both Kucera & Francis and of Carroll et al. (1971), other words besides THE containing the letter H are THAT, HE, WITH, HIS, THIS, HAD, HAVE, THEY, WHICH, THERE, THEIR and WHEN. Yet these twelve words, which are very frequent in writing, are rare in crossword puzzles. Their infrequency, and the almost total absence of THE, help explain why the letter H occurs far less frequently in crossword puzzles than in general writing.

Surprisingly, J was the least frequently used letter in crossword puzzles, being used even less often than Q, X or Z. Authorities differ as to the comparative use of these infrequent letters in general writing. According to Gaines (1939), Smith (1943) and Wolfe (1974), only Z is used less frequently than J, but according to Valerio (1983), Ohaver (1933), Pratt (1942), Moore & Waller (1962), Laffin (1964) and Haldane (1976), both Z and Q are used less often than J.

It should be kept in mind that in crossword puzzles the letter combinations very frequently include abbreviations, initials (of organizations, etc.), foreign language words (especially Latin, French, German, Spanish and Hebrew), proper names (of countries, cities, rivers, famous persons, etc.), nicknames, slang, prefixes, suffixes, two or three words together, or alternate spellings of words. These undoubtedly affect the frequency of letters, as compared with their frequency in general writing.

REFERENCES

Francis, Darryl. 1975. An all-consonant crossword (puzzle). Word Ways 8, 3: 154-155, 192

Kahn, Merle. 1971, 11
Kucera, H. & Francis, D. 1977. The most frequent 100,000 words in general prose (1913-1967). Word Ways 10, 1: 5-45
Randolph, William. 1977. The most common 100,000 words in conversations. 40 pp. Herald Publishing House, Kent, Ohio
Hall, Arthur. (1974.) (Frequency of use of letters) (extract from McWhirter & McWhirter, 1974)
Moore, Dan Tyler & Waller, Martha. 1962. Cloak and cipher. 256 pp. Bobbs-Merrill Company, Indianapolis, Indiana
Ohaver, M. E. 1933. Cryptogram solving. 32 pp. Author, Columbus, Ohio
Valerio, P. 1893. De la cryptographie. Paris (extract in Ball, 1947)
Various specialized crossword puzzles have appeared in Word Ways. Francis (1973) devised a crossword puzzle consisting solely of three-letter all-vowel words, and later (1975) devised a remarkable crossword puzzle of three-letter all-consonant words. Randolph (1978) composed a crossword puzzle without the letter E, and followed this up (1978) with a crossword puzzle in which every vowel was E. The reader is invited to determine the unusual feature in the crossword puzzle given below.

**HORIZONTAL**

1. congestion
2. game bird
3. wager
4. unwrought metal
5. mammalian fluid waste
6. chopping tool
7. northern constellation
8. therefor (2 words)
9. do wrong
10. minaret
11. aeries
12. pig pen
13. torn waste cloth
14. at the tip of (poetic)
15. suffix: carbohydrate
16. pierce with a horn
17. barrier
18. wading bird
19. roam about
20. aeries
21. torn waste cloth
22. at the tip of (poetic)
23. evening (poetic)
24. other
25. long fish
26. sharp
27. remote
28. barrier
29. wading bird
30. roam about
31. evening (poetic)
32. other
33. long fish
34. sharp
35. remote
36. barrier
37. wading bird
38. roam about
39. evening (poetic)
40. other
41. long fish
42. sharp
43. remote
44. barrier
45. wading bird
46. roam about
47. evening (poetic)
48. other
49. long fish
50. sharp
51. remote
52. barrier
53. wading bird
54. roam about
55. evening (poetic)
56. other
57. long fish
58. sharp
59. remote
60. barrier
61. wading bird
62. roam about
63. evening (poetic)
64. other
65. long fish
66. sharp
67. remote

**VERTICAL**

1. sudden shake
2. prefix: of Mars
3. cat sound
4. liquid measure
5. footed vase
6. assist
7. those holding elective office
8. feminine name
9. singing voice
10. departure
11. decades
12. awareness apart from the senses (abbreviated)
13. ovum
14. over there
15. drives back
16. cutting tool
17. caudal appendage
18. mischievous children
19. edge
20. molding
21. talk with excessive enthusiasm
22. paradise
23. in such a manner
24. adjectival suffix forming the comparative degree
25. small lizard
26. dreamy tranquility, as from the use of drugs
27. necktie
28. those who use a cultivating or weeding tool
29. gag
30. Roman date
31. unaspirated
32. Japanese church
33. water pitcher
34. ensuing
35. inlet
36. fourth caliph of Islam
37. animal doctor (colloquial)