

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

Letter	Crossword Frequency	Text Frequency	Average Number of Times Each Letter Occurred in Each Puzzle	Per Cent of Puzzles With Letter
E	.1519	.1250	23.13 (11 to 39)	100
A	.1008	.0805	14.59 (4 to 25)	100
S	.0970	.0655	14.03 (4 to 24)	100
O	.0702	.0760	10.15 (2 to 24)	100
T	.0684	.0925	9.89 (2 to 20)	100
I	.0666	.0729	9.64 (3 to 19)	100
R	.0638	.0613	9.23 (2 to 18)	100
N	.0635	.0710	9.18 (1 to 18)	100
L	.0502	.0414	7.26 (1 to 23)	100
D	.0419	.0397	6.06 (0 to 16)	99.6
U	.0317	.0272	4.59 (0 to 17)	98.6
P	.0301	.0202	4.35 (0 to 17)	97.2
M	.0234	.0254	3.39 (0 to 18)	94.4
C	.0220	.0310	3.19 (0 to 10)	94.8
G	.0205	.0195	2.97 (0 to 10)	94.8
Y	.0176	.0172	2.54 (0 to 8)	92.2
B	.0174	.0154	2.52 (0 to 14)	84.8
H	.0147	.0542	2.13 (0 to 12)	88.8
K	.0091	.0066	1.32 (0 to 6)	71.0
W	.0090	.0188	1.30 (0 to 6)	73.8
F	.0085	.0231	1.25 (0 to 9)	60.0
V	.0075	.0100	1.09 (0 to 8)	63.8
Z	.0024	.0010	0.35 (0 to 3)	27.8
X	.0021	.0020	0.30 (0 to 3)	25.6
Q	.0008	.0011	0.12 (0 to 3)	10.0
J	.0008	.0016	0.11 (0 to 4)	7.6

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 Kučera and Francis (1967).

The vowels A, E, I and O occurred at least once in every cross-word 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 (1960) 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 12th 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 Kučera & 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 & Richman (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 Kučera and Francis study was made by Eckler (1978).

Among the fifty most commonly used words, found in the lists of both Kučera & 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 (1893), 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

- Ball, Walter W. R. 1947. *Mathematical recreations and essays*. 418 pp. Macmillan, New York (see Valerio)
- Berger, Kenneth W. 1977. *The most common 100,000 words in conversations*. 40 pp. Herald Publishing House, Kent, Ohio (review in *Word Ways* 11, 3: 137, 1978)
- Card, Leslie E. & Eckler, A. Ross. 1975. A survey of letter-frequencies. *Word Ways* 8, 2: 81-85
- Carroll, John B., Davies, Peter & Richman, Barry. 1971. *The American Heritage word frequency book*. 856 pp. Houghton Mifflin Co., Boston
- Eckler, A. Ross. 1978. One million English words. *Word Ways* 11, 3: 179-184
- Espy, Willard R. 1973. *The game of words*. 279 pp. Bramhall House, New York
- Francis, Darryl. 1973. An all-vowel crossword (puzzle). *Word Ways* 6, 3: 154-155, 191
- Francis, Darryl. 1975. An all-consonant crossword (puzzle). *Word Ways* 8, 3: 154-155, 192
- Gaines, Helen Fouché. (1939) 1956. *Cryptanalysis*. 237 pp. Dover Publications, New York

- Haldane, R. A. 1976. *The hidden world*. 207 pp. St. Martin's Press, New York
- Hall, Arthur. (1974.) (Frequency of use of letters) (extract from McWhirter & McWhirter, 1974)
- Kahn, David. 1968. *The codebreakers: the story of secret writing*. 1164 pp. Macmillan Company, New York
- Kucera, Henry & Francis, W. Nelson. 1967. *Computational analysis of present-day American English*. 424 pp. Brown University Press, Providence, Rhode Island (review by Eckler, 1978)
- Laffin, John. 1964. *Codes and ciphers: secret writing through the ages*. 152 pp. Abelard-Schuman, New York
- McWhirter, Norris & McWhirter, Ross. 1974. *Guinness book of world records, 1974 edition*. 672 pp. (see p. 201). Sterling Publishing Company, New York (see Hall, 1974)
- Milliken, Donald D. 1942. *Elementary cryptography and cryptanalysis*, 3rd edition. 140 pp. New York University Press, New York
- 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
- Pratt, Fletcher. 1942. *Secret and urgent: the story of codes and ciphers*. 282 pp. Blue Ribbon Books, Garden City, New York
- Randolph, Boris. 1978. *Wright on!* (crossword puzzle without E). *Word Ways* 11, 2: 90-91, 127
- Randolph, Boris. 1978. *Ernest's Revenge* (crossword puzzle with all vowels E). *Word Ways* 11, 4: 224-225, 255
- Shipley, Joseph T. 1960. *Playing with words*. 186 pp. Prentice-Hall, Englewood Cliffs, New Jersey
- Smith, Laurence Dwight. 1943. *Cryptography: the science of secret writing*. 164 pp. W. W. Norton Company, New York
- Valerio, P. 1893. *De la cryptographie*. Paris (extract in Ball, 1947)
- Vizetelly, Frank H. (1946) (Frequency of letters) (Funk & Wagnalls New Standard Dictionary) (cited in World Almanac and Book of Facts, 61st edition, p. 551, New York World-Telegram, New York)
- Walsh, William S. (1946.) (Frequency of letters) (cited in World Almanac and Book of Facts, 61st edition, p. 551, New York World-Telegram, New York)
- Wolfe, James Raymond. 1974. *Secret Writing: the craft of the cryptographer*. 192 pp. McGraw-Hill Book Company, New York
- Zettersten, Arne. 1969. *A statistical study of the graphic system of present-day American English*. 189 pp. Studentlitteratur, Lund, Sweden

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
4. game bird
9. wager
12. unwrought metal
13. mammalian fluid waste
14. chopping tool
15. northern constellation
16. therefore (2 words)
17. do wrong
18. minaret
20. aeries
22. pig pen
24. torn waste cloth
25. at the tip of (poetic)
28. suffix: carbohydrate
30. pierce with a horn
34. barrier
35. wading bird
37. roam about
38. vim
39. more unusual
40. evening (poetic)
41. other
43. long fish
44. sharp
45. remote
47. pronoun
49. rejects
52. frequently
56. poem
57. desire intensely
61. female sheep
62. mental perception
63. lubricator
64. king (Latin)
65. compass direction
(abbreviation)
66. marsupial (pl.)
67. creativeness

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
19. awareness apart from the
senses (abbreviated)
21. ovum
22. over there
24. drives back
25. cutting tool
26. caudal appendage
27. mischievous children
29. edge
31. molding
32. talk with excessive enthusiasm
33. paradise
35. in such a manner
36. adjectival suffix forming the
comparative degree
42. small lizard
44. dreamy tranquility, as from
the use of drugs
46. necktie
48. those who use a cultivating or
weeding tool
49. gag
50. Roman date
51. unaspirated
53. Japanese church
54. water pitcher
55. ensuing
58. inlet
59. fourth caliph of Islam
60. animal doctor (colloquial)

1	2	3		4	5	6	7	8		9	10	11
12				13						14		
15				16						17		
18			19					20	21			
			22		23		24					
25	26	27			28	29			30	31	32	33
34				35				36		37		
38				39						40		
41			42		43				44			
			45	46			47	48				
49	50	51						52		53	54	55
56				57	58	59	60			61		
62				63						64		
65				66						67		