OMITTED LETTERS IN CROSSWORDS

JEREMY MORSE
London, England

Sixty or more years ago I read a humorous story—possibly by A.A. Milne—in which the two main characters were called ETAION and SHRDLU. Later I came to realise that these names were taken from a letter frequency list There are a number of these lists with small differences between them reflecting the context of the letter count. For *Language on Vacation* Dmitri Borgmann used one starting ETAISONHRDLU, differing from my author's only in the positioning of the S. This article is concerned with letters in crosswords, but it starts from the other end by looking at letters not present in their solutions.

One of the oldest crossword series is in the London Daily Telegraph; it started in 1925, and puzzle no. 25,000 appears in 2006. Nowadays the main puzzle has a 15x15 blocked diagram which yields about 160 letter-squares; the clues are cryptic but not too difficult. Printed alongside it is a smaller Quick Crossword (13x13 with about 120 letter-squares and definitional clues). Over the past five years I have examined the solutions of 1,000 pairs of these crosswords, recording omitted letters for the larger puzzle and for the two taken together—but not for the smaller puzzle by itself, except to note if it was pangrammatic, i.e. used all 26 letters.

Although it is a finer point of crossword composition to use the whole alphabet, cryptic composers concentrate more on clue-writing, and only 14 of the 1,000 larger puzzles were in fact pangrams. By contrast, 101 of the smaller and simper puzzles were pangrams. On six occasions both puzzles were pangrams, and there were eleven other occasions when neither puzzle was a pangram, but between them they used every letter.

There was a similar imbalance in puzzles using 25 out of the 26 letters. Only four of the larger puzzles had this feature (with three omitting Q and one X), whereas there were 61 occasions when the two puzzles together omitted only one letter (F3, J16, Q18, V1, W1, X4, Y2 and Z16). At the opposite end, the biggest number of different letters omitted from the larger puzzle was eight, and this occurred thirteen times; the two puzzles together omitted at most six different letters, also thirteen times.

In total, the 1,000 larger puzzles omitted 4,556 letters (averaging 4½), and the two puzzles together omitted 2,706; the latter figure reflects the greater tendency to pangrammatism in the smaller puzzle. As many as sixteen different letters (nearly two-thirds of the alphabet) were omitted at least once in the larger puzzle, as follows (with figures for the two puzzles together in brackets): B 79(15), C 2(0), F 222(80), G 26(2), H 20(1), J 826(648), K 283(93), M 12(0), P 6(0), Q 875(684), U 20(0), V 214(52), W 253(79), X 638(405), Y 212(37), Z 868(610).

Unsurprisingly, the single most often omitted combination was JQXZ 91 times (105 times in the two puzzles together), followed by JQZ 45(100), JKQXZ 43(12), JQWXZ 37(12), FJQXZ 28(14), JQVXZ 28(7), JQXYZ 22(5), JKQZ 17(20), and QZ 17(39). It can be seen that whereas the numbers for combinations of five letters decline when the smaller puzzle is added in, the numbers for smaller combinations increase, such as JQZ, JKQZ and QZ. Indeed, the number of puzzles omitting JQXZ or some smaller set of these four letters was as high as 200 (20 per cent) for the larger puzzle and 523 (over 50 per cent) for the two puzzles together.

The omitted letters group themselves in descending order as follows: QZJ, X, K, WFVY, B, GHUMPC. Putting this order alongside the inversion of Borgmann's frequency list, we get:

Borgmann Z X Q J K B V G P Y W F M C U L D R H ...
Crossword Q Z J X K W F V Y B G H U M P C...

Since most language use is peppered with small words like THE, OF, WHAT, etc., it is not surprising that W, F and H are less rare in Borgmann's list. More puzzling is the relative rarity of X and B in the Borgmann list.

One might have expected a closer correlation between my results and the letter-values of Scrabble, a game based on the crossword and (like the crossword) a comparatively recent invention. However, the two sets of groupings show some notable anomalies:

Crossword QZJ, X, K, WFVY, B, GHUMPC
Scrabble QZ 10, JX 8, K 5, HWFVY 4, BCMP 3, DG 2

The crossword statistics suggest that H should be revalued from 4 to 2, CMP from 3 to 2, D from 2 to 1, J from 8 to 10, and U from 1 to 2.

Plainly the research behind this article is limited, and there may be differences between frequency and infrequency studies which I have not thought through. Those more expert than myself in these matters may be able to offer explanations for the anomalies I have mentioned.

The editor called my attention to an earlier Word Ways article on crossword letter frequencies, "Crossword Puzzle Letter Frequencies" by John Hitchcock in February 1979. He examined 500 14x13 puzzles from the Laramie Daily Boomerang (Oct 24 1976-Jun 11 1978), noting the percentage of puzzles containing a given letter. His corresponding infrequency list was JQXZFVKWBHYMCGPUD; the remaining nine letters appeared in all puzzles. Sometimes the percentages differed substantially:

	Q	Z	J	X	K	W	F	V	Y
Telegraph	12.5	13.2	17.4	36.2	71.7	74.7	77.8	78.6	78.8
Boomerang	10.0	27.8	7.6	25.6	71.0	73.8	60.0	63.8	92.2

Boomerang puzzles contained 144 letter squares, compared with the 160 (noted earlier) for Telegraph puzzles, which may account for some of the difference; however, the differences for J and Z, and to a lesser extent X, are puzzling. Hitchcock noted the reduced frequency of H, F and W in crosswords compared to running text; contrariwise, Z showed up much more frequently. The greater frequency of Z in Boomerang may be explained by US preference for –IZE against British preference for –ISE. Only one of the 500 Boomerang puzzles contained all 26 letters, compared with 14 (noted earlier) for the 1000 Telegraph puzzles.