ON WRITING LIPOGRAMS

ROBERT CASS KELLER

Why is it that writers usually omit E, the most common letter in English text, when creating a literary lipogram? I believe that this tendency has been reinforced by the existence of the only book-length English-language lipogram, Ernest V. Wright's E-less novel, Gadsby (Wetzel, Los Angeles, 1939). In fact, its influence may have extended even to France. OuLiPo member Georges Perec constructed La Disparation (Denoel, Paris, 1969) without using the letter E; some of the book's reviewers did not even realize that it was a lipogram!

Obviously, omitting any other letter should make the lipogrammatic task easier. To restore balance, I propose that a lipogram ought to be constructed omitting as many letters of the alphabet as possible. Throwing away the rarest letters first, how many can be eliminated before the task of writing with the remainder is comparable to a missing E? A first answer to this question can be given by removing letters which collectively have the frequency of E in English text -- the eleven letters ZQJXKVBYGWP.

When constructing a lipogram, short words are more valuable than long ones, for it is likely that most long words will be eliminated no matter what letters are thrown away. If a given alphabetic letter is more prevalent in short words than long ones, this letter ought not to be removed. For example, the rather rare letter W is concentrated in several common short words, such as WAS, WITH, WHERE, WERE, WHEN and WILL; F occurs in the very common OF, FOR and FROM.

The value of a set of rare letters can be assessed by comparing the probability that a randomly-chosen i-letter word is eliminated with the probability that a randomly-chosen i-letter word has the letter E in it. If these probabilities are matched as well as possible for small values of i (say, two through six), one can reasonably conclude that the task of writing a lipogram is the same as it was for Wright.

One point, however, needs clarification -- how are words to be randomly chosen? One can argue that one needs as large a stockpile of words as possible, and therefore each word in the dictionary ought to have an equal chance of being selected. On the other hand, one can argue that some words (like FOR) are vastly more useful in writing than others (like FIG) and therefore ought to be sampled with probability proportional to English-text usage. Rather than adjudicate this difference of opinion, I present the best set of rare letters to eliminate for both alternatives.

To avoid getting bogged down in lengthy calculation, I took from H. Kucera and H. N. Francis' Computational Analysis of Present-Day American English (Brown University Press, 1967) the 25 commonest two-letter words, and the 100 commonest three-letter, four-letter, five-letter and six-letter words, together with their numerical frequencies in a sample of a million words of English text. I weighted each word by its relative frequency, obtaining the following table.

Probability of Drawing a Word at Random
Without the Indicated Letters

	E	ZQXJVK LCPBG	ZQXJVK LCPBGU
2-letter	.88	.92	.90
3-letter	.55(.84)	.92	.88
4-letter	.58	.65	.62
5-letter	.43	.47	.38
6-letter	.35	.32	.26

For three-letter words, THE overwhelms the others in frequency, accounting for more than one-third of the usage; the quantity in parenthesis gives the probability with THE removed.

When counting words without regard to English-language text frequency, I enlarged the sample with 100 additional four-letter and five-letter words, for a total of 625 words of all lengths.

Probability of Drawing a Word at Random
Without the Indicated Letters

	E	ZQXJVK CBPFU	ZQXJVK CBPFUG
2-letter	.88	.76	.72
3-letter	.72	.69	.63
4-letter	.47	.51	.48
5-letter	.44	.46	.36
6-letter	.43	.36	.29

Although the choice of letters is not absolutely clear, it is evident that eleven rare letters are about equivalent to E when writing lipograms. It should be a considerably more difficult task to write a lipogram after throwing away half the alphabet. Who will be the first to write a half-alphabet lipogram (pick your own letters) for Word Ways, at least a page long (500 words), that sounds so natural that readers will not suspect anything is amiss?