PATTERN WORDLISTS

DAVID SHULMAN
New York, New York

The editor of Word Ways, A. Ross Eckler, has already described and reviewed the pattern wordlists of Professor Jack Levine, Department to Mathematics of North Carolina State University (February 1972, November 1972, and August 1973). These wordlists are so excellent that if you can buy, borrow or steal them, by all means do it.

These lists were derived by computerization from the second and third editions of the unabridged Merriam-Webster dictionary. They were made possible by a grant, and the copies were offered free of charge to various recreational linguists, including members of the American Cryptogram Association and the National Puzzlers' League, as well as Word Ways subscribers. The copies were quickly snapped up, and are no longer available; one person who obtained the last two parts but missed the first has already offered to pay $50 for a copy.

Prior to these lists, Professor Levine started with one in 1957 (also out of print at present) which served as a prelude for his chef-d'oeuvre of later date. For the sake of precision, here is a bibliography of all four works:

A list of words containing no repeated letters (non-pattern words), Raleigh, N.C., 1957. 115 p. 8 1/2 x 11". Wraps. Mimeo.
A list of pattern words of lengths two through nine, Raleigh, N.C., 1971. 384 p. 8 1/2 x 11". Wraps.
A list of pattern words of lengths ten through twelve, Raleigh, N.C., 1972. 360 p. 8 1/2 x 11". Wraps.
A list of pattern words of lengths thirteen to sixteen, Raleigh, N.C., 1973. 269 p. 8 1/2 x 11". Wraps.

So far, the reader who is not familiar with these lists and their primary use in cryptography may want to know the definitions of pattern and isomorph. In my Glossary of Cryptography (Crypto Press, N.Y.C., 1961), they are defined as follows:

Pattern Word: One in which one or more letters are repeated. If enciphered in simple substitution, the substitute word will retain the original pattern.
Isomorph: Identical plain text repetitions differently enciphered.

The latter term is actually applied to the more complex cipher systems rather than to simple substitution, so it might be preferable here to restrict our usage to the former term. But, in the overall field of crypt-

analysis, the solver uses pattern words, isomorphs, idiomorphs and other means to reconstruct the original text of a cipher message.

Back in the 1930s when I first compiled my own lists of pattern and non-pattern words, which I then described in an article in The Cryptogram, some purists of the American Cryptogram Association and the National Puzzlers' League objected that such lists were tantamount to cheating in solving cryptograms. They preferred to use pure mind over matter. Today, one no longer hears of any objection. The pattern wordlist is an accepted tool of cryptanalysis.

For that matter, any kind of wordlist, be it crossword puzzles, synonyms, rhyming dictionary, reversed letters, and so on, has been welcomed by those who wish to deal with words and their ways quickly and effectively. We are not expected to know all the words in all the dictionaries; in the more obscurely-composed cryptograms, anagrams, palindromes, and other types of word play, the composer may have resorted to words we do not know. Unfortunately, finding them in the ordinary dictionary that is structured by alphabetical order may be a time-consuming, often fruitless task. On the other hand, a suitable specialized wordlist can lead us to immediate success.

Professor Levine's pattern wordlists are among the good word lists that deserve to be in one's collection. As a pioneer in that field, I admit it and appreciate such superior work. However, it will not be taken as criticism, if I explain how he has done his listing and indicate the way in which I would have preferred it, for, as A. Ross Eckler has written me, "I doubt if anyone will ever tackle the job of rearranging Levine's three volumes!" I certainly will not, because it is a stupendous task, with or without the aid of a computer.

The method of presentation used by Professor Levine can be summarized as follows. Let n equal the number of letters in a word, and p denote the pattern. For words of two to nine letters:

```
R E O R D E R E D
1 2 3 4 5 6 7 8 9   n = 9, p = 5.9-1.4.7-2.6.8
```

Thus, D is repeated in positions 5 and 9; R, in positions 1, 4 and 7; and E, in positions 2, 6 and 8. For words of ten or more letters:

```
S C H I Z A E A C E A E
1 2 3 4 5 6 7 8 9 0 A B   n = 12, p = 29-68A-70B
```

In order to save space and keep his column presentation regular, Levine used the digit 0 to represent the tenth letter of a word, and the letters A, B, C, ... for the eleventh, twelfth, thirteenth .... letters.

All very good, but I would have preferred the following system that I started with and always used:

```
1 2 3 1 4 2 1 2 4
R E O R D E R E D   n = 9, p = 1 2 3 1 4 2 1 2 4
```

Note that the number pattern actually reflects the word pattern. When it comes to locating such a pattern in the word lists, there is no loss of time. Every pattern is arranged in strict numerical order from 1111 to 1234 in a four-letter list, for example. In the Levine lists, one has to consult a specially prepared table to determine where to locate a particular pattern by page number.

There is still another advantage for the cryptanalyst. Levine's lists do not show a pattern word that has an inflected form, such as a plural where an S has been added, unless one rewrites the word and reorders the pattern. In my method, all one does is drop the final digit, and then uses the list of one letter less with the remaining pattern. Thus, REORDERED can be reduced to REORDERE, where no such word will show up, and then to REORDER, where that word will appear. In other words, one goes from 1 2 3 1 4 2 1 2 4 to 1 2 3 1 4 2 1 2 to 1 2 3 1 4 2 1.

Not only does my method reflect the actual word pattern, but it also reflects other words in the same cipher message when solving. Thus, if there are, say, 50 words of the same pattern as FLORISTS, the question arises as to which is the required word in a particular cipher message. By numbering each of the other words in the message from the pattern of 1 2 3 4 5 6 7 6, suppose we find a four-letter word, 2 5 1 7. After the substitution of the proper letters from FLORISTS, this word becomes LIFT, confirming the choice of the former word.

I have shown how a pattern wordlist more useful to cryptanalysts can be produced. However, I agree with the editor that it is too formidable a task. My own manuscript wordlists, separated into common words and obscure words for solving cryptograms, are far from complete. For the sake of completeness, and because his lists have been published, we can learn how to use them and benefit from the work that Professor Levine has done. For that we certainly owe him our gratitude.