EIGHT-LETTER TWO-PAIR WORDS

A. ROSS ECKLER
Morristown, New Jersey

Jack Levine's three volume listing of Websterian words grouped according to their letter-patterns (such as EXCESS and BAMBOO) is an almost inexhaustible treasure-trove of linguistic curiosia. In this article, we focus upon one-thirtieth of his total corpus: the 15,054 eight-letter words in which two letters appear twice each and four letters appear once each. There are 210 different patterns having this distribution of letters; Levine gives examples for all but six. The table below lists the pattern, the commonest known word having that pattern, and the number of Levine examples with that pattern. In most cases, the commonest word is the one having the greatest number of occurrences in Kucera and Francis' Computational Analysis of Present-Day American English (Brown University Press, 1967), a million-word sample of 191 prose from United States sources. However, the frequencies of the rarer words are unreliable, and a few substitutions have been made. For example, if the most frequently-occurring word of a given pattern does not appear in the Merriam-Webster Pocket Dictionary, and another Levine word having the same pattern does, the second word was substituted for the first. The eleven asterisked words appear in Webster's Unabridged but not in the Pocket Dictionary. The six parenthesized words are non-Websterian, and their sources are given below.

<table>
<thead>
<tr>
<th>AABB Patterns</th>
<th>ABAB Patterns</th>
<th>ABBA Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabb...</td>
<td>mementos 78</td>
<td>opposite 92</td>
</tr>
<tr>
<td>aab.b.</td>
<td>nonmolar 41</td>
<td>alliance 40</td>
</tr>
<tr>
<td>aab.b.</td>
<td>titanium 57</td>
<td>approach 79</td>
</tr>
<tr>
<td>aab.b.</td>
<td>memories 66</td>
<td>occasion 62</td>
</tr>
<tr>
<td>aab.b.</td>
<td>academic 35</td>
<td>efficacious 27</td>
</tr>
<tr>
<td>aa,b.b.</td>
<td>tortoise 133</td>
<td>capacity 81</td>
</tr>
<tr>
<td>aa,b.b.</td>
<td>olemes* 3</td>
<td>research 29</td>
</tr>
<tr>
<td>aa,b.b.</td>
<td>ommers* 3</td>
<td>aemetry 67</td>
</tr>
<tr>
<td>aa,b.b.</td>
<td>oomantia* 4</td>
<td>aations 94</td>
</tr>
<tr>
<td>aa,b.b.</td>
<td>oolemas* 1</td>
<td>ainsanity 68</td>
</tr>
<tr>
<td>aa...b.</td>
<td>reported 131</td>
<td>a...b.a.</td>
</tr>
<tr>
<td>aa...b.</td>
<td>idolized 72</td>
<td>a...b.a.</td>
</tr>
<tr>
<td>aa...b.</td>
<td>national 160</td>
<td>a...b.a.</td>
</tr>
<tr>
<td>aa...b.</td>
<td>invasion 43</td>
<td>a...b.a.</td>
</tr>
<tr>
<td>aa...b.</td>
<td>educated 105</td>
<td>a...b.a.</td>
</tr>
<tr>
<td>a...abb.</td>
<td>amassing 10</td>
<td>a...ba.</td>
</tr>
</tbody>
</table>

* oocystic 2 | * academic 35 | * olemes* 3 |
* ommers* 3 | * oomantia* 4 | * oolemas* 1 |
* reported 131 | * idolized 72 | * national 160 |
* invasion 43 | * educated 105 | * amassing 10 |
* oocystic 2 | * ommers* 3 | * oomantia* 4 |
* oolemas* 1 | * reported 131 | * idolized 72 |
* national 160 | * invasion 43 | * educated 105 |
* amassing 10 | * oocystic 2 | * ommers* 3 |
* oomantia* 4 | * oolemas* 1 | * reported 131 |
* idolized 72 | * national 160 | * invasion 43 |
* educated 105 | * amassing 10 | * oocystic 2 |
* ommers* 3 | * oomantia* 4 | * oolemas* 1 |
What are the sources of the non-Websterian words? Four of the six were taken from the Times Atlas of the World: Aalsmeer is a place in the Netherlands, Oordegem is in Belgium, and Llanarth and Llaniair Caereinion are both in Wales. Words for the two remaining patterns were more difficult to locate; Dmitri Borgmann suggested Llannors (towns such as Llannor, in Wales, listed in the Times Index-Gazetteer of the World) and Aahhotep (the consort of an early Egyptian king, listed in William R. Cooper’s Archaic Dictionary).

* * * *

One of the most striking features of this list is the extreme variation in the number of examples given by Levine -- the mean number is about 71.5, but individual lists of isomorphs range from 0 to 317. Is such a variation a reflection of the fact that certain word-patterns are easier to speak or spell, leading to preferred combinations of letters? Or, at the other extreme, can one claim that whenever a new word is formed, each of the 210 different patterns is equally likely to be selected? One would, of course, anticipate some variation in the numbers even if the latter explanation were the correct one; after all, if one rolls a die 600 times one does not expect to see exactly 100 ones, 100 twos, etc. It is not hard to show by statistical means, however, that the variation in the list sizes is considerably larger than that which would occur by chance. For example, slightly more than half of the 210 numbers should lie between 66 and 77 (inclusive), yet only 24 actually do.

Is the observed variability purely a property of the individual letter-patterns, or can a more general theory of variation be found? It is quite obvious that the 15 patterns in which the first two letters are the same generate relatively few words. Turning to that part of Levine listing eight-letter words in which only one letter appears twice, one quickly discovers that, among these 28 different patterns, by far the rarest one is that in which the first two letters are the same (16 words of the pattern displayed by aardwolf and oometric). These observations suggest the following question: can the number of occurrences of an eight-letter two-pair word of a specified pattern be explained in terms of the number of occurrences of the two corresponding eight-letter one-pair words?

The answer to this question is yes, at least in a large number of cases. To illustrate: looking at the frequency-of-occurrence of the two letter-patterns a...a, and bb... among all eight-letter one-pair words, one can predict that the number of occurrences of the pattern...abb... agrees with Levine's data on the general rule.

Unfortunately, occurrences of words with too few letters for the patterns found above fall too low or are missing from the list, and such a check is not possible for the patterns of the form aabb.

Patterns

...aab, b vamooses 22
...aba, b optimism 60
...abb, a unsettle 30
...aa, bb dumbbell 25
...ab, ab changing 10
...abba specific 103
...aabb weakness 76
...bab expanded 72
...bab reminded 126
...bba trapdoor 39
...aabb meatball 42
...a...bab premises 81
...a...bab joyfully 66
...a...bab farewell 123
...a...bab shielded 97
...a...bab banshees 31
...aabb Caribbee* 7
...aab wanderer 104
...aab roulette 109

* * * *
unsettle 30
specific 103
sheriff's 44
trapdoor 39
joyfully 66
bannees 31
roulette 109

Our of the six is a place in the Llanfair patterns. Llannors ex-Gazetteer of king, is-

Supreme variation a number is to 317. Is patterns are s of letters? A word is to be select-

Is there any pattern to these exceptions? The under-predicted letter-patterns fall largely into two groups: 4 patterns having the internal palindrome abba, and 6 patterns in which the repeated group ab is separated by other letters. The 5 other letter-patterns are dominated by specific combinations of letters. For example, 230 of the 271 one-pair letter-patterns are words of the form sCVCCers, and 18 more are words of the form gCVCCing (where V denotes a vowel and C, a consonant); similarly, 103 of the 123 ...a.abb letter-patterns are words of the form...eness, ...eress or ...eless.

The picture of the over-predicted letter patterns is much less clear. There is some evidence that internal patterns of the form aabb and a.bba inhibit the formation of words, but the majority of the patterns cannot be easily classified.

The predicted number of occurrences of a given eight-letter two-pair letter pattern is proportional to the product of the relative frequencies of the two corresponding one-pair letter patterns. For example,
the predicted number of occurrences of .abb..a. is proportional to the product of 0.0711, the fraction of words having the .a....a. pattern among all eight-letter one-pair words, and 0.0558, the fraction of words having the ..bb.... pattern among all eight-letter one-pair words. To scale up this product to the predicted number of occurrences, multiply it by the factor $N / \sum P$, where $N$ is the total number of eight-letter two-pair words (15,054) and $\sum P$ is the summation of the 210 individual products, one for each of the eight-letter two-pair patterns (0.268). Putting it all together, the predicted number of occurrences of .abb..a. is $(0.0711)(0.0558)(15,054)/(0.268) = 223$, as given earlier. For those who wish to make other predictions, the individual fractions (calculated from Levine) are given below:

- aa........ 0.0006
- a.a....... 0.0210
- a.a....... 0.0340
- a.a....... 0.0298
- a.a....... 0.0250
- a.a....... 0.0368
- a.a....... 0.0494

TRAFFIC

This is the title of an as-yet-incomplete transposition poem by Tom Smith, a teacher at Castleton State College in Vermont. It will eventually consist of five major sections (entitled Complaints, Blessings, Speculations, Alphabets, and Fables). Each section consists of five subsections labeled with Roman numerals, and these in turn consist of five lines each. Each of the 125 lines in the poem is transposed into five other lines, as for example:

"Fiddlededee's not English," replied Alice Gravely.


Forty per cent of the poem can be purchased from the author in an autographed booklet for three dollars (Box 223, Castleton, Vermont 05735).