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 curiosa. 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 patterns, 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's 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... (Aahhotep)</td>
<td>abab... mementos 78</td>
<td>aabb... opposite 92</td>
</tr>
<tr>
<td>aab.b. (Llanarth)</td>
<td>aba.b. nonmolar*41</td>
<td>abb.a... alliance 40</td>
</tr>
<tr>
<td>aab..b. (Llanfair)</td>
<td>ab..b. titanium 57</td>
<td>abb.a... approach 79</td>
</tr>
<tr>
<td>aab...b. Ilaretas*1</td>
<td>abab..b. memories 66</td>
<td>abb..a. occasion 62</td>
</tr>
<tr>
<td>aab...b. oocystic*2</td>
<td>ab...b. academic 35</td>
<td>ab...a. effusive 27</td>
</tr>
<tr>
<td>aa,bb... (Lllanors)</td>
<td>ab..ab... tortoise 133</td>
<td>ab..ba... capacity 81</td>
</tr>
<tr>
<td>aa,bb... oometers*3</td>
<td>abab..b. property 70</td>
<td>ab..ba... research 29</td>
</tr>
<tr>
<td>aa,bb... oolackan*3</td>
<td>ab..ab... incident 115</td>
<td>abab..b. a. temerity 67</td>
</tr>
<tr>
<td>aa,bb... oomantia*4</td>
<td>ab..ab... advanced 64</td>
<td>abab..b. a. stations 94</td>
</tr>
<tr>
<td>aa,bb... oolemmas*1</td>
<td>ab..ab... churches 186</td>
<td>ab..ba... insanity 68</td>
</tr>
<tr>
<td>aa...b. (Oordegem)</td>
<td>ab..b.. reported 131</td>
<td>ab..a. arterial 75</td>
</tr>
<tr>
<td>aa...b. oothecae*2</td>
<td>ab..a. idolized 72</td>
<td>ab..b. a. segments 108</td>
</tr>
<tr>
<td>aa...b. (Aalsmeer)</td>
<td>ab..ab... national 160</td>
<td>ab..b. a. thoughts 156</td>
</tr>
<tr>
<td>aa...b. oosphere*8</td>
<td>ab..a. b. invasion 43</td>
<td>ab..ba... entrance 91</td>
</tr>
<tr>
<td>aa...b. ooziness*2</td>
<td>ab...ab... educated 105</td>
<td>abab...b. a. students 283</td>
</tr>
<tr>
<td>a.abb... amassing 10</td>
<td>a.bab... proposed 44</td>
<td>a.bba... passport 26</td>
</tr>
</tbody>
</table>
Is grouped AMBOO) is a. In this e 15,054 and four let-

In a. a. b. \textit{electric} 45 a. a. b. \textit{totalled} 16

a. a. b. \textit{awakened} 50 a. a. b. \textit{emerging} 35

a. a. b. \textit{analysis} 44 a. a. b. \textit{odorous} 16

a. a. b. \textit{roommate} 11 a. a. b. \textit{addition} 51

a. a. b. \textit{accepted} 65 a. a. b. \textit{assemble} 57

a. a. b. \textit{occurred} 16 a. a. b. \textit{forms} 19

a. a. b. \textit{loosened} 37 a. a. b. \textit{alleging} 20

a. a. b. \textit{weevely} 21 a. a. b. \textit{afforded} 24

a. a. b. \textit{football} 52 a. a. b. \textit{reproofs} 30

a. a. b. \textit{textbook} 50 a. a. b. \textit{abstracts} 45

a. a. b. \textit{purses} 45 a. a. b. \textit{highball} 30

a. a. b. \textit{harassed} 53 a. a. b. \textit{honorary} 46

a. a. b. \textit{militant} 67 a. a. b. \textit{mutually} 34

a. a. b. \textit{Japanese} 92 a. a. b. \textit{pitiless} 16

a. a. b. \textit{balloons} 26 a. a. b. \textit{happened} 239

a. a. b. \textit{caféine} 52 a. a. b. \textit{ballroom} 59

a. a. b. \textit{collided} 64 a. a. b. \textit{fullness} 37

a. a. b. \textit{number} 33 a. a. b. \textit{manually} 73

a. a. b. \textit{requests} 97 a. a. b. \textit{progress} 35

a. a. b. \textit{opera} 36 a. a. b. \textit{provost} 62

a. a. b. \textit{fortress} 34 a. a. b. \textit{standard} 54

a. a. b. \textit{spittoon} 13 a. a. b. \textit{circumspection} 38

a. b. a. \textit{tortured} 72 a. b. a. \textit{concerns} 83

a. b. a. \textit{circular} 56 a. b. a. \textit{coincide} 63

a. b. a. \textit{monument} 85 a. b. a. \textit{amenable} 47

a. b. a. \textit{adequate} 58 a. b. a. \textit{endorsed} 67

a. b. a. \textit{sorority} 42 a. b. a. \textit{suburban} 43

a. b. a. \textit{honoring} 87 a. b. a. \textit{visitors} 62

a. b. a. \textit{tendency} 157 a. b. a. \textit{rotation} 78

a. b. a. \textit{cultural} 123 a. b. a. \textit{vitality} 220

a. b. a. \textit{fighting} 119 a. b. a. \textit{flexible} 229

a. b. a. \textit{movement} 76 a. b. a. \textit{together} 76

a. b. a. \textit{landlord} 70 a. b. a. \textit{contacts} 51

a. b. a. \textit{abstract} 62 a. b. a. \textit{contract} 72

a. b. a. \textit{flailing} 48 a. b. a. \textit{arrests} 89

a. b. a. \textit{district} 90 a. b. a. \textit{language} 121

a. b. a. \textit{district} 160 a. b. a. \textit{measures} 95

a. b. a. \textit{petition} 30 a. b. a. \textit{texts} 31

a. b. a. \textit{presents} 57 a. b. a. \textit{diaper} 45

a. b. a. \textit{thinking} 148 a. b. a. \textit{weighing} 91

a. b. a. \textit{tensions} 165 a. b. a. \textit{idol} 72

a. b. a. \textit{interior} 36 a. b. a. \textit{embodied} 63

a. b. a. \textit{athletes} 80 a. b. a. \textit{traverse} 76

a. b. a. \textit{governor} 109 a. b. a. \textit{cather} 56

a. b. a. \textit{standard} 54 a. b. a. \textit{constant} 38

a. b. a. \textit{training} 105 a. b. a. \textit{redwood} 27

a. b. a. \textit{impressed} 177
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 *aab* (a, b) and *aba* (a, b) among all eight-letter one-pair words, one can predict that the number of occurrences of the pattern *aab* agrees in Levine's list with expectation, and the deviation is due to the general relative size of the list.

Unfortunately, this does not tell us that the patterns for the two-letter words are randomly distributed. It merely shows that the patterns for the two-letter words are not more likely to be found in Levine's list than random patterns.

Patterns...

* * * *

The preceding two-letter patterns, for example, would be aabb and aabb; the pattern aab...aab would not be expected to be the pattern in Levine's list.

Is there any general rule governing the occurrence of the twoletter-patterns? The answer is yes. The general rule is that the twoletter-patterns are included in Levine's list if their letter-pairs are not of the form (sonant); sonants are consonants, and the letter-pairs are not sonants.

The problem is now clear. The aabb and aabb...aabb are included in Levine's list if their letter-pairs are not of the form (sonant).
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 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. pat-
tern 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 occur-
cences, 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 occur-
rences 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:

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>.aa.....</td>
<td>0.0006</td>
</tr>
<tr>
<td>.a.a.....</td>
<td>0.0210</td>
</tr>
<tr>
<td>a..a.....</td>
<td>0.0340</td>
</tr>
<tr>
<td>..a.a.....</td>
<td>0.0298</td>
</tr>
<tr>
<td>a...a.....</td>
<td>0.0250</td>
</tr>
<tr>
<td>a....a.....</td>
<td>0.0368</td>
</tr>
<tr>
<td>a.....a.....</td>
<td>0.0494</td>
</tr>
</tbody>
</table>

aa........ 0.0251
.a.a..... 0.0404
.a.a..... 0.0372
.a.a..... 0.0438
.a.a..... 0.0576
.a..a..... 0.0711
.a..a..... 0.0377
.a..a..... 0.0558

TRAFFIC

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

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

Once the graveside fiddlers gleed: "I peel an idyl!".
Pegleg idiots, sidereally fevered, danced in Hell.
Ages reeled. Needy devil, drill and police the figs.
Peddler -- I'll give her fledged in seedy escalation.
Verily fallen, I dig deep St. Charlie Dodge's needle.

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