BOTTOMS UP!

The August 1992 ***Word Ways*** *reviewed Ted Clarke's startling claim (in Volume 1, Issue 2 of "Wordsworth") that it is more efficient (i.e., quicker) to build word squares from the top down than from the bottom up, as done by formists for more than a century. Two readers, Eric Albert and Leonard Gordon, dispute this conclusion; their rebuttals are given below.*

The evidence provided by the work of over a century of expert human formists, combined with that of several years of computer experiments by me, is unequivocal: all other things being equal, there is an enormous advantage to building large forms from the bottom word up, instead of from the top word down.

As I stated in my **Word Ways** article on finding a 9-square ["The Best 9x9 Square Yet", November 1991], one of the basic reasons for this asymmetry is that English is relatively "ending-poor." In other words, there are many more combinations of letters that begin words than that end words. If you start from the top, you often have to work down deeper before you find you've hit a dead end, and this extra work is part of what makes the top-down approach take more time.

Mr. Clarke knows of this argument (in fact he quotes me on it) and of the historical and computer evidence behind it, so I was quite surprised to see him contradict me based solely on the results of his observations of a few runs of one program on a single base word, using a database that had been artificially seeded to produce a single 10-square.

The speed of a single run depends almost entirely on the base word chosen and the order in which the words in the database are checked to see if they finish off a square. A little thought will show that, given the right base and a suitable ordering of the database, a 10-square could be finished after just nine tries. However, one would be ill-advised to decide, based on this evidence, that it takes only nine tries to finish the <u>average</u> 10-square!

Another flaw with Mr. Clarke's experiment is the program he is using. From the description he gives of his algorithms and data structures, it would seem that his program is unsophisticated and inefficient. I would not argue with the claim that it is possible to write <u>some</u> program that constructs word squares quicker from the top <u>down</u>, but I believe that any well-written, sophisticated program and database package will, in general, work much more quickly from the bottom up.

Oddest of all was Mr. Clarke's claim that he "failed to detect any obvious generally greater frequency of starting combinations."

It sounds like his detective work consisted of a quick (visual?) scan of the output from some of his program runs. It is an easy task to have the computer actually count the number of starting and ending combinations in a database. Twenty minutes of programming could have saved Mr. Clarke from making this peculiar statement.

To summarize: I believe that Mr. Clarke's arguments are ill-founded. Those who wish to attempt building large forms, whether by hand or by computer, should start from the bottom and work up.

--Eric Albert

The August 1992 **Word Ways** contains a review of a magazine article by Ted Clarke in which he claims that the accepted method of building word squares from the bottom up (with reverse words) as was used by Eric Albert is wrong for the computer. He claims it is faster to work from the top down using normal words; the November 1992 **Word Ways** presents his reasoning. My analysis finds that although his observation may be correct in some cases, his conclusion is not general and his reasoning is wrong. Consider an ideal square beginning like this:

```
A B C D E F G H
I J K L M N O P
Q R S T U V W X
```

The existence or number of vertical words beginning with A, B, C, .. is of no real significance. It is not until you choose the second word IJKLMNOP that significant pruning can be used. AI must begin a word, BJ must begin a word, etc. After some IJKLMNOP has been accepted, choose QRSTUVWX and now the truly important pruning enters. AIQ must begin a word, BJR must begin a word, etc. Since there are far fewer ending than beginning trigrams, it is usually better to work from the bottom. But now comes another consideration. For single squares we have:

```
A B C D E F G H
B J M L M N O P
C M S T U V W X
D L
```

Now, trigram pruning only begins with DLT, EMU, etc. Upon examination of my eight-letter word list, I find there are more beginning corner combinations when starting from the bottom. This works against the pruning advantage. After acceptable third words have been placed, there still are fewer combinations to continue from when working with reverse words but there has been a time penalty in getting there. For my particular database, I find that times to exhaust a search are about equal for both procedures. There may be a slight advantage in working in the normal direction.

Here is what happens when working from the bottom up, and here is a possible fix. Many squares end in combinations like:

```
. . N   . . N   . . N   . . E   . . T   . . E
. . E   . . E   . . E   . E S   . . E   . E R
N E R   N E D   N E S   E S S   T E R   E R S
```

Create a separate list (or lists) of beginning words. Examine your data and cull all but one of each set of words that are identical except for the ending(s). Then, as was done in the August **Word Ways** article by Albert and Long, examine the results and introduce the variations. This idea is probably more important when using a database like mine which is derived from the Official Scrabble Players Dictionary, than when using only root words as found in standard dictionaries.

In reply to Ted Clarke, I suggest he back off from 10-by-10 word squares and find some 9-by-9 ones instead (so far, Eric Albert is leading one to nothing). I also suggest that he does not read Frank Rubin's **Word Ways** articles; they may scare him off entirely. *--Leonard Gordon*

THE OXFORD DICTIONARY OF MODERN SLANG

Slang is, according to editors John Simpson and John Ayto, "English with its sleeves rolled up, its shirttails dangling, and its shoes covered with mud." There are more than 5000 such words in the above-mentioned book, concentrating on the slang of the 20th century which has been admitted to the OED (though there are about 500 words or new meanings too recent to have made the Second Edition). Each entry contains the date of the earliest-known printed usage, plus (usually) an illustrative sentence. I scanned the 384 different words (counting the various usages of a word like ***do*** *separately) beginning with D, and found only thirteen first appearing in the 1980s, from* ***dipstick*** *(a quote from Maledicta, referring to the penis) to* ***dweeb****. Still, some of the slang tagged US has sunk into obscurity; how many readers know the slang meanings of* ***ridge-runner*** *1933 (hillbilly),* ***bladder*** *1936 (an inferior newspaper),* ***monkey-man*** *1924 (a servile husband),* ***grid*** *1922 (bicycle), or* ***goop*** *1900 (a stupid person)? It is also a bit surprising that substandard spellings like* ***feelthy, gotta, lotsa*** *or* ***doncha*** *are included; it would be an endless task to document all such dialectal writing. These quibbles do not detract from what is, on balance, a solid work of scholarship and a delightful browse. Who would have thought that* ***outasight*** *dates back to 1893, or* ***screw*** *to 1725? Published by Oxford University Press in 1992, it is available in hardcover for $25.*