10-BY-10 SQUARES: ANOTHER UPDATE

FRANK RUBIN
Wappingers Falls, New York

It has now been one year since I began my computer search for a 10x10 double word square. My efforts have been directed against three main obstacles: the need to obtain a massive number of words, the limited speed of a personal computer, and its limited storage.

Since my first report in August 1988, Eric Albert has responded with three important additions to the word stock. He has supplied me with a list of words from the Official Scrabble Player's Dictionary, a list of multi-word phrases, and a list of pseudo-plurals, all on diskettes that I could read directly into the PC. The OSPD list and the phrases could be merged directly into the word stock. I accomplished each of these merges in a single evening.

The pseudo-plurals took far more effort. This list consisted of words of 8 to 10 letters that had been pluralized as though they were all nouns, following the normal rules about word endings. For example, words ending in Y became IES, those ending in MAN became MEN, those ending in H, S, X, or Z added ES. The problem was to sort through these lists removing the spurious words. I spent several hours each night for over a month combing through this list to obtain not only the genuine noun plurals, but verb tenses and plausible possessive forms. I also changed many spurious plural adjectives into valid comparative forms, added past tenses to parallel the present tense verbs that had been formed, and made other changes to word endings to stretch out the list of new words.

With the addition of these three lists, the word stock grew to over 93,000 ten-letter words.

Throughout this period I kept typing in words from Jack Levine's pattern word lists. I own a print copy of this list, and I also have a computer tape of the word stock from which it was originally compiled. Unfortunately, however, I do not have a tape drive on my PC capable of reading the tape.

The speed limitation of the PC has caused a rollback of my original goals. My intention was to perform an exhaustive search for all possible 10x10 double word squares. After the word stock passed the 50,000 word level, an exhaustive search became infeasible. I began adding heuristics to the program to attempt to limit the size of the search. For example, when the program considers the words for the top two lines of the word square, it checks that each of the ten vertical bigrams begins a valid word. To reduce

the search, I changed this to require that each bigram begin 10 valid words, later 25. Similarly for trigrams and tetragrams.

Unfortunately, this approach may miss valid squares if even one of the words begins with a rare bigram or trigram. Consequently, I abandoned the search for a double word square, and began concentrating entirely on the single square. Since that decision, the computer speed has ceased to be a serious limiting factor.

The restricted storage of the PC posed a more difficult challenge. I have rewritten the program several times to increase the capacity. The first rewrite was described in my previous progress report in the February 1989 **Word Ways**. Instead of storing the entire 10-letter word, I had 26x26 separate lists, one for each of the 26x26 possible initial bigrams. This allowed me to store only the last 8 letters of the word, and increased the capacity to 54,000 words. The second major rewrite divided each of those 676 lists into 26 sections, one for each possible third letter. This let me keep only the last 7 letters of each word, and increased the capacity to 64,000 words.

At that point I thought I had hit a dead end. To record the start and end of those 26x26x26 sublists required 2 characters of computer storage per sublist, or over 35,000 characters of computer storage. To progress to 26x26x26x26 lists was impossible, since that would take up more than the computer's total storage with the pointers to those lists.

It took about 3 months before I came up with a more complex scheme that solved the dilemma. I used two separate 26x26 lists. The first list pointed to a list of 3rd and 4th letters of the words, and the second pointed to a list of the last 6 letters of the words. The list of letters 3 and 4 also contained an index into the list of 6-letter endings showing where the endings for each set of 4 initial letters began and ended. This scheme proved very difficult to program, and took over a month to get working.

This scheme increased the program capacity to 74,000 words. I used a scheme of letter probabilities to select the 74,000 best words from the 93,000 word dictionary. This does not significantly reduce the number of squares produced, since low-frequency letters rarely appear in the partial squares, and even medium-frequency letters like B, F, and Y are uncommon.

I have created these ever-more-complex storage schemes with an awareness of the inherent irony. Five years from now, PC memories will no doubt be several times larger, and all this effort will be unneeded. Ten years from now few will even remember how programmers had to struggle to fit into the 640K memories of today. (I well remember how I struggled with the 32K storage of the 709 computer, then the world's largest, around 1960, and with the 256K limit of the early 360 computers circa 1965. When old programmers tell these tales, today's programmers wear a pained expression, like I wore for my uncles' description of pre-Depression prices.)

The version with the 3-letter prefixes had taken 33 days with

70,000 words at the start, and 84,300 words when it finished. The version with the 4-letter prefixes amazingly took only 14 days, with 93,300 words in the dictionary at its conclusion.

These improved programs have obtained several partial squares at levels not previously achieved. I had obtained some 6,10 squares during the double square search, but the symmetric search had not produced any. The improved program has turned up two such squares (given at the left, below).

No 9,9 square had ever been found. By removing all tests on the rightmost column of the square, I was able to produce the following pair of 9,9 squares (given at the right, below).

HORBACHITE	MALACODERM	FLORISSANT	TRIPTOMARS
OVERLOOSEN	ARACHNOPIA	LOMENTARIA	RESORCINOL
RECALLMENT	LACCAINICS	OMNIVALENT	ISOPIESTIC
BRASILETTO	ACCLINATES	REIMAGINED	POPULARISM
ALLICESHAD	CHAINSTORE	INVAGINATE	TRILINEATE
COLLEGEICE	ONINSTINCT	STAGIRITES	OCEANLANES
HOMESEEKER	DONATISTIC	SALINITIES	MISREADERS
ISETHIONIC	EPITONIONS	ARENATIONS	ANTIANEMIC
TENTACULAE	RICERCATAS	NINETEENTH	ROISTERING
ENTODERMAL	MASSETERIC	TATDESSSH	SLCMESSCG

Some simple variations are possible, for example REIMAGINES or POPULARIST.

Following the 9,9 breakthrough, I required all squares to be 4,10 before I would continue the search to the 5th row or beyond. With this restriction, the program was still able to find 113 squares at the 8,8 level, and seven squares at the 7,9 level (that is, the squares are both 4,10 and 7,9). Since no 7,9 square whatever had been achieved by the earlier searches, I present four below.

AUTOBLASTS	HEMISTATER	METEORITIC	STRUCTURAL
UNECLECTIC	EPICYEMATE	ETHYLAMINE	TROPHONEMA
TELEOSTOME	MISENTERED	THREETIMES	RONSARDIAN
OCEANLINER	ICEDEALERS	EYEDOTTERS	UPSETMENTS
BLONDENESS	SYNEDRIANS	OLEOPLASTS	CHATTERTON
LESLEYITES	TETARCONID	RATTLETRAP	TORMENTERS
ACTINIFORM	AMELIORATE	IMITATIONS	UNDERTUNIC
STONETOTER	TAREANAGES	TIMESROMAN	REINTENDED
TIMESERIES	ETERNITIES	INERTANCES	AMATORIANS
SCERSSM	REDSSDE	CESSSPS	LANSNSC

The remaining three squares are variations of the first, substituting TELEOSTOMI, ACTINIZOAN, and TIMESEALED.

At this stage, I have solved the problems of speed and capacity. The biggest problem remains obtaining an adequate word stock. Although I continue to type words from the pattern word list, it will take years to reach an adequate vocabulary to complete a 10x10 square. Just progressing from the current 7,9 squares to a 7,10 square may require 10,000 or more new words. Yet, with 94,200 words in my list today, there is always the chance that 9 of the required 10 words are already there, and the word I am typing

this moment is the missing tenth word.

The yield from the typing is much lower now. For each 100 words I type, 70 are already in the computer file. Only 30 per cent are new. Also, my eyes have worsened considerably since I began the project.

There appear to be three main sources of new words. First, there may be specialized word lists in computer form -- for example, geographic terms, Bible names, medical terms, etc. If anyone has access to such lists, or knows where they can be obtained, please contact me.

Second, somebody may have a tape drive capable of reading the tape I have with the original Levine word stock--assuming it is still readable. It is a half-inch, 9-track, 800-BPI, odd-parity tape. Possibly, someone has a tape backup unit that reads such tapes. If not, a more complex process may be needed: read the tape on a mainframe or minicomputer, download it to a PC, and write it to disk. If anyone has such a capability, please let me know.

The third hope is to use a scanner. There are computer scanners that can read a page of text and turn it into a computer text file. (Image scanners that read a page and turn it into a picture file are not useful here.) Unfortunately, I do not know anyone who has one. The most likely place is a university. If anyone has university connections, and could obtain the use of a text scanner, again please contact me, Frank Rubin, at 59 DeGarmo Hills Road, Wappingers Falls NY 12590.

QUERY

"A picture is worth a thousand words" is of course cryptarithmetically impossible, since WORD x 1000 = WORD000, not translatable to PICTURE. Is the cryptarithm possible for some other number than 1000? What is the smallest (or largest) number of words that a picture can be worth? And how close to 1000 can one actually get? This is a problem ideally suited for the PC, for many cases must be considered. Note that all of the digits (0 through 9) must be used in the solution.