

THE POLYGLOT TEN-SQUARE

A. ROSS ECKLER

Morristown, New Jersey

In "Hunting the Ten-Square" in the May 2004 Word Ways, Rex Gooch mentions that Graham Toal recently succeeded in constructing by computer ten-squares based on 592,361 words drawn from various European languages. For the record, here are the two squares that he referred to:

A A N G E H A R D E Dutch
 A P E R N A S E I S Spanish
 N E C E L I S T V I Czech
 G R E N A D E R E N Norwegian
 E N L A G U N A R E Spanish
 H A I D U C E S T E Romanian
 A S S E N E R A I S French
 R E T R A S A R S E Spanish
 D I V E R T I S S E French
 E S I N E E S E E N Finnish

A A N G E V A R E N Dutch
 A T E R M A N E L E Romanian
 N E Z A B R A N E N Czech
 G R A D U A D O R A Spanish
 E M B U S T I M O S Spanish
 V A R A T T O M A N Finnish
 A N A D I O M E N E Italian
 R E N O M M E R E Z French
 E L E R O A N E L E Romanian
 N E N A S N E Z E N Czech

Since then, he has posted 776 such squares on his website, with initial words AANGEHARDE through ACHISPASES (www.gtoal.com/wordgames/wordsquare/tensquare.results). This represents about 1/176th of Stephen Marshall's *Complete Ten Letter Word Book* (1950) based on Webster's Second, and 1/213th of a French dictionary. Taking the more conservative extrapolation, one can expect about 135,000 polyglot ten-squares from Toal's program! Here are the results of an email colloquy with Toal (my comments in italics):

A small increase in vocabulary can lead to a dramatic increase in the number of squares. Clearly one cannot hope to publish even a tiny fraction of these?

Indeed, and since there are so many there is little point in even generating them. As it was putting too much of a load on my home PC, I killed the program. The partial results are there on the web page and are quite sufficient for proof of concept.

The chance of a monolingual square is, of course, infinitesimal, but perhaps squares with only three or four languages may exist. How about a square with ten different languages? Are some languages more likely?

It looked that way to me...I have a suspicion that later in the alphabet where you get languages that have a strong CVCV repeating pattern such as Spanish, they might work better with words of the form VCVC... My first impression was that some of the Eastern European languages (Polish? Russian?) have so many more words than English due to more cases and declensions (i.e., fewer base words but more variation in endings) that they might be more fruitful.

What dictionaries did you use? How many words from each language?

I am working on a project to write software to critique Scrabble games, and I wanted to do this for every foreign language that is supported by a Spears' Games (now Mattel) Scrabble set. Most

of the words are reasonably good quality lists received from the Scrabble community in those countries. If an official word list did not exist, I used the best-quality spelling-checker file I could find on the net. The word sources are not perfect, but they are not full of the usual cruft that accumulates in the majority of public-domain wordlists. I extracted the words rather crudely and am not in a position to duplicate the same lists exactly to tell you the numbers [in each language]. There are lots of factors such as incompatible character sets and accented words, and convergence to a single case where you have languages such as German with esszet, which exists only in lower case.

It would be helpful to label on the website the language represented by each line of each square.

It shouldn't be hard to do a post-processing stage where the word squares are read back in and tagged with the source of each word.

Among your 776 squares there exist clusters of closely-related ones, some with only a single letter-swap on the diagonal. What words appear in the most squares? Do these words have certain characteristics? Do some words gravitate to certain square positions (especially the tenth)? I suspect that there exist squares with no words in common with any other square.

Quite likely, but not especially interesting to me; my interest is more in the computational side. Now that we know that we get a high return from a polyglot wordlist, I'd like to see how the code fares with 11-squares. My code is algorithmically about as good as it can get now; low-level tweaking might get a factor-of-two improvement, but we need orders of magnitude improvements to take this to the next stage. There's a little interest among word game programmers to throw some parallel computing power at the problem and perhaps reduce the runtime from years to days. Jean-Charles Meyrignac is probably the person who will be doing the next interesting stuff in this arena. He has a distributed computing project underway, and has promised us some time on his array in a few months when his own project is ready.

A Book of Serbian Word Squares

Miroslav Lazarevni is the author of ЕНИГМАТСКИ МАГИЧНИ КВАДРАТИ (Enigmatical Magic Squares), published in Belgrade in 1999, which displays 314 eight-squares (accompanied by one-line definitions of all words therein), as well as 1112 eight-squares, 40 nine-squares and 5 ten-squares that have appeared in various Serbian publications between 1934 and 1998. Here are the most recent ten-squares, published in 1987 and 1996; can anyone evaluate their word quality?

С А М О Н А М Е Н А
А Н И М А Л И З А М
М И Л А Д И Н О В О
О М А Њ И С И Т А Р
Н А Д И З А Т Е Љ И
А Л И С А М А Р И Ћ
М И Н И Т А Љ И В А
Е З О Т Е Р И Ч А Н
Н А В А Љ И В А Т И
А М О Р И Ћ А Н И Н

К А М А Т А Р И Н А
А Ђ У С А Л А Т И Н
М У Н А Р А Н И Н И
А С А Л А М И Н И Ј
Т А Р А С И Ј Е В А
А Л А М И Н А Р И Н
Р А Н И Ј А Њ А Н И
И Т И Н Е Р А Р И Ј
Н И Н И В И Н И Ц А
А Н И Ј А Н И Ј А Ц