

# LOGOLOGY BY COMPUTER

ALAN FRANK

Boston, Massachusetts

Computers have the potential for playing many roles in the field of recreational linguistics. Six general categories are:

- a) making a list of all words with specified orthographic characteristics,
- b) managing logological data bases,
- c) acting as a practice opponent in word games,
- d) solving formal word puzzles such as crosswords, double-crossics, and the types dealt with by the National Puzzlers' League and Games magazine,
- e) composing such puzzles,
- f) typesetting material for magazines of recreational linguistics.

The first five topics will be discussed in the order listed, with the first two sections giving specific examples relating to the 4-set project described by Philip Cohen in the May 1983 *Word Ways*. The last category is more in the field of computers than logology and will not be discussed here. Indeed, the tangential subject of automatic hyphenation could more than double the length of this article. There are other topics, such as use of computers for the structural analysis of English, which I don't consider recreational, so I won't talk about them either.

## A. Making Lists of Words

This sort of dull task is something computers are very good at. Unfortunately, a computer can't just open up a dictionary and start reading; it is necessary for human beings to type all the words in a dictionary into the computer before it can start looking. On the other hand, once this has been done for a particular dictionary it need never be done again; the information can easily and reliably be transferred between computers. There are many dictionaries which now exist in computer-readable form. I mention three to which I have access: Webster's New International Dictionary, Second Edition (NI2); the Official Scrabble Players Dictionary (OSPD); the Webster Vest Pocket Dictionary (VPD). The NI2 was put in machine-readable form as part of a U.S. Air Force project, and is widely available for only the cost of making a copy. The OSPD was typed up clandestinely by a group of employees at a large New England company. This may involve a violation of the dictionary's copyright; Selchow and Righter, the makers of Scrabble, are unhappy about this pirating and have made unsuccessful attempts to suppress it. Because the OSPD contains almost no base words of more than eight letters, I found the smallest dictionary I could (Webster's Vest Pocket Dictionary) and entered into my data base

all words of nine or more letters along with their inflected forms.

Once the data base exists, many lists can be made; for example, the Levine pattern and non-pattern word lists and the Ritchie transpositional list were created from N1 databases. Many commercially available transpositional and positional word lists were also created from computer data bases. Scrabble players have created a plethora of lists for study and reference. For example, given any set of tiles, one can look up all the words that can be made using all the tiles in the set. Given six letters and a blank (usable as any letter), another list will tell you all the letters which the blank can represent to make a seven-letter word. Two other lists, mentioned in Eric Albert's *Kickshaws* in the May 1983 *Word Ways*, are of hooks and of seven-letter words in order of probability.

Many other lists which have been discussed in this journal can also be produced by computer, such as subtransposals (February 1982 *Colloquy* and February 1969), eodermidromes (August 1980), and words which form two six-letter words when the first, eighth and fifteenth letters are removed (May 1983, in Kyle Corbin's article "N-Tile Scrabble Records"). Since many words in the last category are plurals, when I did this search I also looked for fourteen-letter words which split into two sixes upon removal of the first and eighth letters. In constructing a list of sixteen-letter transposals, I attempted to get around the lack of inflected forms in the data base by having the computer process both the simpler form and its inferred plural. This resulted in several absurdities, such as *gasoline electric / geocentricallies*. However, the computer search found *congratulatories (OED) / action regulators* and two other N1 pairs which I'm saving for use as National Puzzlers' League puzzles.

In November 1981, I began using the computer to search for the shortest word containing each possible set of four letters (the 4-set project referred to earlier). At this time I didn't have N12, so the first attempt was made using only the two smaller data bases. A few months later, N12 was added on. This search was done by brute-force methods; all sets of four letters in each word were examined to see if the new word was shorter than anything already entered in the data base.

## B. Computer Management of a Logological Database

I knew that there would be many entries in the 4-set data base which were not what I wanted. Spurious words occurred for three reasons: some OSPD words are not N1, there are typographical errors in the version of the N12 data base which I had, and there were no indications of capitalization in that N12 version. Thus, the next step was to have the computer produce an alphabetical list of all the words in the data base. This was checked by human beings against both N12 and N13 and their addenda. Words which could not be found were replaced by the best things I could think of. Subsequent use of the N12 data base was able to refill some of these holes and the use of a cleaner version was able to add some others.

But the computer was only able to supply entries for about 21500 of the 4-sets. The other 1900 would have to be found by hand, in NI and other references. Nevertheless, the computer continued to be of major assistance. It has been used, for example, to produce a collection of want lists. The main want list is a double-spaced listing of all sets for which there is nothing (76 sets) or only a tagged word (1695 sets). There are two subsidiary lists which are extremely useful: the one-bad list contains all sets with exactly one of the hard-to-find letters JQVXZ, and the awful list contains all sets with tag lengths of five or more.

At times, other specialized lists have been produced. For example, a list of sets of tag length four and over containing only letters which can be found in Polish place names is employed when searching a list of such names. As another example, the computer was used to generate a list of words with the property that adding one letter to the word would fill a hole; this was done to identify hyphenated combinations in the OED that could then be examined by hand. Thus, the list indicated that the addition of a W to **proverb** would fill the BPVW hole, and in fact **proverb-wisdom** was found. Although this process was not as lucrative as combining rare-letter sections of the OED such as J, Q, Z and EX, it resulted in many such finds.

Once a potentially useful word is found, it can be typed into a special program which looks at each 4-set the word contains. If the old entry for the set is worse (has a longer tag), it is replaced, but saved on a special backup list in case of error. There is also a printed master list showing what exists for each possible 4-set. Since it runs to about sixty pages, it is not reprinted often, but instead is supplemented by update lists which the computer automatically generates, showing all the sets in alphabetical order which have been improved since the previous update list. There are other miscellaneous tasks for which the computer is used, such as generating statistics showing the progress of the project and producing special lists such as appeared in the May 1983 **Word Ways** article.

Although some logologists are able to put up with inordinate amounts of scutwork to support a project, I'm not. I like the dictionary searching, but the presence of the computer to do everything else makes it much more palatable.

### C. The Computer as a Practice Opponent

If given the right data base, computers are very good at finding all words which satisfy specified orthographic conditions. There are many games in which this ability plays an important part. Boggle is an extreme example in which there are no confounding factors such as luck or strategy. Consequently, it is easy to program a computer to play a perfect game of Boggle. Computers have also been programmed to play one or both sides of Hangman and Jotto, which are more interesting from the theoretical standpoint.

But as most readers of this journal should be aware, I am primarily interested in Scrabble. Indeed, there have been computer

programs written to play this game. One of them was produced with the approval of Selchow and Righter, the holders of the Scrabble trademark, and is commercially available under the name "Monty Plays Scrabble". I know very little about this program (machine, as it can be obtained as a self-contained unit), save that it does not have the entire OSPD available and that, on a low level, it was soundly defeated by an average player. The other computer program for Scrabble was written by a Selchow and Righter employee and does have the entire OSPD. It has no trouble finding the highest-scoring plays in any position. On the other hand, it has very little strategic or complex tactical knowledge programmed in. Fortunately, this does not detract from its ability to teach words and simple tactics. One can play a game against the computer and have it indicate when a better play was available, or use it to post-mortem a game played against another human opponent.

Because the only computer to which I have access has an incompatible operating system, I am unable to use this program regularly. I am working on the development of a program which will run here and which will also have more advanced knowledge. For the meantime, I have developed various practice programs which present me with racks of seven tiles in vitro, from which I try to find the best play. The computer informs me of any good plays which I miss. I also play Boggle with the computer, which is useful for getting the mind going in a different direction and for learning new words. From my experience, all of these programs seem to be of considerable benefit.

In addition to serving as a practice opponent, the program can help determine proper strategy. Given a set of alternative moves in a particular situation, it can play many games, using each of the moves in turn, to see which one results in the highest percentage of wins. One can also compare elements of strategy by using two programs, identical but for slight differences in one parameter (e.g., the value of an S). After playing a few hundred games (overnight on a fast machine), it will be clear which one is doing better and hence which is the better strategy. These two types of test are examples of the Monte Carlo method used by operations researchers and statisticians.

#### D. Solving Word Puzzles

This subtopic was inspired by the article "Wordplay by Computer: Two Early Views", the February 1983 *Word Ways* reprint of two long-ago articles from the *Enigma* on how a computer might (or might not) help solve word puzzles such as transpositions or word forms. Unfortunately for National Puzzlers' League members looking for easy routes to "completes" (all puzzles solved in an issue), computers are not very helpful in solving most types of puzzles though they are able to provide some assistance. The main reason for this is that none of the data bases contain any semantic information. Thus, if one wishes to solve a seven-letter transposition, the computer can only provide a list of all the thousands of such pairs, requiring a human search for the right one. (Of course, if one knows one word, the computer can find the other

word quickly, but a competent human with anagram tiles can usually do that too.) However, there are some unusual cases. For example, if one is trying to solve a transposition in which one term has the letter-pattern 4 2 6 (that is, a three-word phrase of the form xxxx xx xxxxxx) and the other has the pattern 4-8, the complete list will be very short and the solution may be found at once. I intend to create a small on-line dictionary with semantic information to examine the feasibility of unassisted computer solution of this type of puzzle.

Computers are useful for solving letter-substitution ciphers, the type of puzzle which appears in many newspapers and puzzle magazines. They may be used in two ways. Without needing a dictionary, the computer can perform frequency counts and other numerical tests which enable experienced human cryptographers to identify certain letters or letter categories (such as vowels). With a dictionary, a computer can solve almost all letter-substitution ciphers by brute-force methods. There are more sophisticated methods discussed in various technical journals.

Some progress has been made on computer programs for solving crossword puzzles, but again the difficulty lies in providing the computer with the necessary semantic information. For details, the reader is referred to "Crosswords and the Computer" in the November 1980 *Word Ways*.

Many of the objective contests in *Games* magazine are very well suited to computer solution. The ones best suited are generally not logological but combinatorial. The reason for this is that *Games* uses N13 as its dictionary of record, whereas N12 is more accessible in computer-readable form. Some solvers have been working around this problem and have accumulated large sets of N13 information on computers. Michael S. Wolfberg, who has done a lot of work on this, has won several contests with the help of hours of machine time on a powerful computer. The *Games* staff does not publicize such information, but most serious contestants are well aware that they may be competing with computers.

#### E. Computer Composition of Puzzles

This subject is somewhat less controversial than the previous one. Because puzzles are meant to entertain, it usually makes little difference to the solver whether a particular challenge was formulated by a human being, a computer program, or a troupe of monkeys seated in front of typewriters. Actually, computer construction of puzzles is currently limited to the discovery of bases (filled in crossword grids, pairs of transposable words, etc.); the clueing must still be done by a human being. Construction of crossword grids has been discussed in "Crosswords and the Computer" in the November 1980 *Word Ways*, in "Crossword Construction by Computer" in the August 1981 *Word Ways*, and in the February 1983 *Colloquy*. Existing programs have generally been based on a single dictionary, which is very different from the way human crossword composers work. To a human, any combination of letters is a possible puzzle entry, though some combinations are more ac-

ceptable than others. For example, NYTO may be used, in case of desperation, clued as "\_\_\_\_ L.A. (busy air route)" or RRRR may be clued as "curriculum at advanced school?". The inclusion of such oddities will often make possible the simplification of other parts of the grid or the inclusion of a greater number of thematic entries. I have, nonetheless, found computers useful in helping find words to fill specific spots in a crossword grid; if I need an eight-letter word with the third letter one of BCFGPS, the fifth letter a vowel, and the seventh letter one of DEIMPT, it can quickly give me a complete list.

Turning to National Puzzlers' League puzzles, computers are very good at churning out long lists of words having specified properties, such as transposition bases. However, it must be noted that among long words most transposition bases are exceedingly dull (e.g., photomicrograph and microphotograph). This is also true of puzzle types such as beheadments (f-rightfulnesses), charades (anti plus militarism makes antimilitarism), and so on. Thus, a fair amount of human time is still required to plow through such computer-generated lists to find puzzles which are worth posing.

Surprisingly, computers may also be useful partners in composition of the letter rebus. One use is to have the computer locate words which have a high density of substrings which turn into fewer characters in the rebus display. I fed the computer a list of common rebus shorthand and it told me about misatone which can be indicated with the display MI (M is at [Roman] one) and other such simple bases. Sometimes a human being will come up with a clever rebus idea but will not know a proper answer word to use to best express this idea. For example, I came up with the idea of having an answer word such as picture and a rebus display TRUCE, to be interpreted as "PI (scramble the letters of) CTURE". I figured that this would be more impressive with a longer word and had the computer search for candidates. I could have done this myself, looking through the PI section of the dictionary and unscrambling everything myself, but the computer can do it faster and more thoroughly. It came up with pillowcase ("PI COLESLAW") among others, which I wrote up and submitted as a puzzle to the National Puzzlers' League.

Several years ago, a rebus was done with display T1 T I UM and answer titanotherium (TI, another I, UM). This takes advantage of the long word **another** fortuitously nested inside an even longer one. A computer could search for such pairs, but the percentage of uninteresting hits (unPRAISEworthy, etc.) can be expected to be so high as to make searching for the good ones almost fruitless. However, the National Puzzlers' League has another puzzle, the **suber**, or reversed rebus, in which the display represents the answer word backwards. Here the computer can be useful, because almost all backward embeddings of words are interesting. I have, in fact, obtained a list of word pairs of this type which are suitable for being puzzle bases. An example (which has already appeared at least once) is K , H as a display for hammock (K comma H).

Computers are also useful in the construction of cryptograms. One way to use the computer is to have a program which attempts to decide which symbols represent vowels. Revisions of a crypt can be entered into the program until it identifies the vowels as consonants and vice versa. Eric Albert has come up with some extremely difficult puzzles by this method. Sometimes the composer will set a particular goal, such as a lipogrammatic crypt, one with lots of vowels, or a similar technical constraint. The computer can then be used to print a list of words fitting such a constraint, from which the composer can pick those which make a reasonable English sentence.

#### F. Other Uses

There are many other uses of computers of possible interest to Word Ways readers. These include, but are certainly not limited to, checking of puzzles such as cryptograms for conformity to technical rules, lexicography, drawing crossword diagrams, and proof-reading (this system is able to check for words which are not in a standard dictionary, but more advanced systems might be able to detect usage errors as well). The editor has pointed out that computers are potentially of great use in constructing multi-word structures such as symmetric crash groups, discussed in the November 1978 Word Ways (find eight seven-letter words with a pattern analogous to hated horny fitly fauns would wires).

Computers will not revolutionize recreational linguistics, at least not in the near future, nor will they eliminate the large amounts of dictionary searching which are often necessary to the preparation of articles for this journal. However, they will reduce the searching time and make possible the discovery of results which were previously out of reach.

#### *FUTURE MEETINGS*

*The 1984 WHIM (Western Humor and Irony Membership) Conference will be held from March 28 to April 1, 1984 at the Phoenix Townhouse Hotel, Phoenix, Arizona. In addition to keynote talks, there will be approximately 300 presentations of 15 minutes apiece on all aspects of contemporary humor: academic humor, bilingual humor, children's and adolescent humor, ethnic humor, feminist humor, French humor, German humor, humor in American literature, humor in British literature, humor in the classroom, humor in the mass media, humor in popular culture, humorous poetry, linguistics and humor, humor and philosophy, political humor and Newspeak (Orwell's 1984), humor and psychology, religious humor, Russian humor, scientific humor, social laws, Spanish humor, etc. (Paper proposals accepted until January 1.) Registration of \$20 includes one luncheon plus a copy of the Proceedings; send it to 1984 WHIM Conference, English Department, Arizona State University, Tempe AZ 85287.*