REFLEXICONS

LEE SALLOWS
Nijmegen, The Netherlands

Reflexiconstruction

A lexicon is a dictionary or a list of words. Hence my use of "reflexive lexicon" or, more crisply, **reflexicon**, for a self-descriptive word list that describes its own letter frequencies:

*

trois a,
trois c, trois d,
neuf e, quatre f, deux h,
neuf i, six n, quatre o, deux p,
cinq q, six r, sept s,
huit t, neuf u,
cinq x

*

Immortal verity sans superfluity. Now that is what I call **belles lettres**! Below we shall look at some English examples. But first, since the answer is far from obvious, how are reflexicons produced?

Imagine a book with pages inscribed as follows. The text on page 1 might be anything--a colophon, an epithet, a dedication to Ross Eckler. For convenience we assume something short. But page 2 and all subsequent pages each comprise a descriptive list, in words of the number of a's, b's, c's, etc., appearing on the previous page. Thus, if page 1 features a single x, then our volume begins like this:

Page:	1	2	3	4	5	6	7
	x	one x	one e	five e's	five e's	.	.
			one n	five n's	three f's	.	.
			one o	five o's	three i's	.	.
			one x	one x	two n's	.	.
					two o's		
					three s's		
					three v's		
					one x		

This may not be the recipe for a best seller, but the plot does have its appeal. You can hardly help wondering how it ends! Will list lengths continue to expand? Clearly, 26 items is the limit. In fact, here none will exceed 16. There will be totals for E,F,G, H,I,L,N,O,R,S,T,U,V,W,X,Y which are the only letters occurring in English cardinals under ONE HUNDRED, a number much higher than feasible list entries, assuming brevity in the opening text. Our example is therefore a lipogram, a work in which A,B,C,D,J, K,M,P,Q,Z will be absent because missing from page 1, the only page on which they could first occur. The end of our story can now be discerned.

Every new page shows a list of at most 16 totals, none of them large. The possible variations are thus finite. Sooner or later, the numbers on one page will recur on another, albeit differently ordered. Suppose the totals on page N are the same as those on page M. Then N is an anagram of M; their letter frequencies agree. But this means that page N+1 will be identical to page M+1, which shows that our book must wind up in a repetitive cycle. And the same will be true whatever the starting text. Call the number of pages occurring in such a cycle its **period**. If the period is p, then we have a closed loop of p sequentially descriptive lists. If p = 2, they will form a mutually descriptive pair. If p = 1, then we have a list whose description is a copy of itself: a reflexicon.

Let distinct letters stand for distinct lists. The onset of a period 1 loop, R, then looks like this: ...L,M,N,O,P,Q,R,R,R... This shows that the reflexicon R not only describes itself, but it describes list Q as well. So Q must be an anagram of R, most probably a different ordering of the same set of totals. Once any of its anagrams turn up, the reflexicon itself follows immediately. No reflexicon is reached **except** via one of its anagrams.

Question: assuming no ABCDJKMPQZ on page 1, how many different loops are there? Using a computer to extend the above shows that its pages converge on a loop of period 155. Extended trials reveal that, provided we stick to priming texts using the 16 cardinal letters only (none to occur more than 99 times), there are just four possible outcomes. One is the loop of period 155, another is of period 14, while the remaining pair are both of period 1, the two basic English reflexicons:

**fifteen e's,
seven f's, four g's,
six h's, eight i's,
four n's, five o's,
six r's, eighteen s's,
eight t's, four u's,
three v's, two w's,
three x's.**

**sixteen e's,
five f's, three g's,
six h's, nine i's,
five n's, four o's,
six r's, eighteen s's,
eight t's, three u's,
three v's, two w's,
four x's.**

The two longer loops are readily reconstructed by extrapolating from any of their constitutent lists, such as the following (condensed into digits):

E	F	G	H	I	L	N	O	R	S	T	U	V	W	X	Y	length
14	7	0	3	6	0	7	6	6	18	8	4	3	3	3	0	155
17	3	1	5	5	0	4	5	5	14	8	1	2	2	2	0	14

Thus, in English, all the multitude of different possible starting positions lead inexorably into one of these four whirlpools or **strange attractors** as mathematicians call them (see Doug Hofstadter's admirably lucid exposition in [1]). This convergence is easy to understand. A 16-element list has 16! = 20,922,789,890,000 permutations (= anagrams), all of them giving rise to a common description which is itself one among 16! new lists having a common successor, and so on. The resultant funnelling effect carries interesting implications.

Consider a computer program able to generate pages for such a book, starting from any text. A basic routine scans TEXTIN = page N, initially page 1, counts its letters and writes their totals in the form of number-words to TEXTOUT = page N+1. TEXTOUT is now substituted for TEXTIN, the routine reiterated, and so on. I like to picture this process as a machine that takes text as input and yields text as output, the latter coupled back to the former via a feedback path. This makes it easier to see that a reflexicon is effectively a **virus:** a code sequence able to subvert the machine so as to get itself prepetually reproduced. One way to hunt for reflexicons is therefore to set such a machine going and just wait for contagion to set in. However, there are still other viruses that may easily usurp it first. These are the loops of longer period, all of them similarly infectious. How are we going to immunize the device against these unwanted invaders? How do we write a book that ends specifically in a loop of period 1?

My answer is to alter the mechanism so as to neutralize cycles. Instead of updating the totals for every letter on every page, let the next page result from correcting the total for a single letter chosen at random each time. The resulting haphazard behaviour is loop-free by definition **except in one case**: when updating a total entails no change in the subsequent list because it is already correct--because the list **is** already a self-descriptor! In this way the program is forever free to keep juggling numbers until it eventually succumbs to a self-reproducer. The only snag is that **anagrams** then pass unheeded, which means 16! chances lost every time. But not if we **alternate** methods: **all** totals updated on one pass, **one** random correction the next, and so on. Now the former will catch any anagram, while the latter prevents latch-up in loops. A few million iterations (mutations) normally suffice to evolve (naturally select) a viable solution (virus)--assuming one exists, of course, failing which the process grinds on forever.

Reflexiconography

Skipping refinements, so much for the basic machinery. What can we do with it? For a start, note that a self-descriptive **sentence** is really a sugar-coated reflexicon, the essential kernel padded out with some palliative dummy text such as "This sentence contains...". Thus, on appending these constant ballast letters to successive counts, our standard process again issues in an associated self-descriptive list, provided it exists. If not, change "contains" to "employs", say, and try again. Passing over the simplest instances, a few special finds made after adapting the mechanism to suit the purpose deserve notice here. These are seen in (a) a (British) letter-totalling reflexicon, (b) an (American) letter-totalling self-descriptive pangram, and (c) a (trans-Atlantic) mutually-descriptive (pangrammatic) pair. Details of the program changes entailed by these special types would occupy us unduly; the basic mechanism remains the same.

This
sentence
contains
one hundred and
ninety-seven letters:
four a's,one b,three c's,
five d's,thirty-four e's,
seven f's,one g, six h's,
twelve i's, three l's,
twenty-six n's, ten o's,
ten r's,twenty-nine s's,
nineteen t's,six u's,
seven v's, four w's,
four x's,five y's,
and one
z.

This
pangram
contains two
hundred nineteen
letters: five a's, one b,
two c's,four d's,thirty-one e's
eight f's, three g's, six h's,
fourteen i's, one j, one k,
two l's,two m's,twenty-six n's
seventeen o's, two p's,one q,
ten r's, twenty-nine s's,
twenty-four t's,six u's,
five v's, nine w's,
four x's, five y's,
and one
z.

The
right-hand
sentence contains
four a's,one b,three c's,
three d's,thirty-nine e 's,
ten f's, one g, eight h's,
eight i's, one j, one k,
four l's,one m,twenty-three n's,
fifteen o's, one p, one q,
nine r's,twenty-three s's,
twenty-one t's,four u's,
seven v's, six w's,
two x's,five y's,
and one
z.

The
left-hand
sentence contains
four a's,one b,three c's,
three d's, thirty-five e's,
seven f's,four g's,eleven h's,
eleven i's, one j, one k,
one l, one m, twenty-six n's
fifteen o's, one p, one q,
ten r's, twenty-three s's,
twenty-two t's,four u's,
three v's, five w's,
two x's, five y's,
and one
z.

Returning to reflexicons proper, in line with French practice above, the plural S is dispensible. Two instances are then found, one trivial (FIVE F, FIVE I, FIVE V, FIVE E), the other (see below) less dull. This is compact, but logologists like their alphabet soup thick. Now, there are 14 words and 12+6+3+7+2+2+5+5+5+6+3+6+4+4 = 70 letters in the text, an average of 5 letters per word, or 0.2 words per letter. Plural S has been dropped. ls there another way to increase the semantic density through discarding still further redundant symbols? One approach (see next page) leaves a certain amount up to the viewer; letters

TWELVE E,	FIVE R,
SIX F,	FIVE S,
THREE H,	SIX T,
SEVEN I,	THREE U,
TWO L,	SIX V,
TWO N,	FOUR W,
FIVE O,	FOUR X.

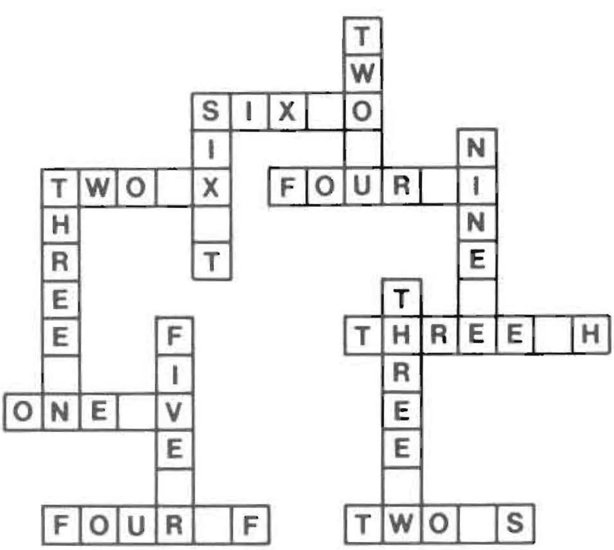
under the line are properly invisible.

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

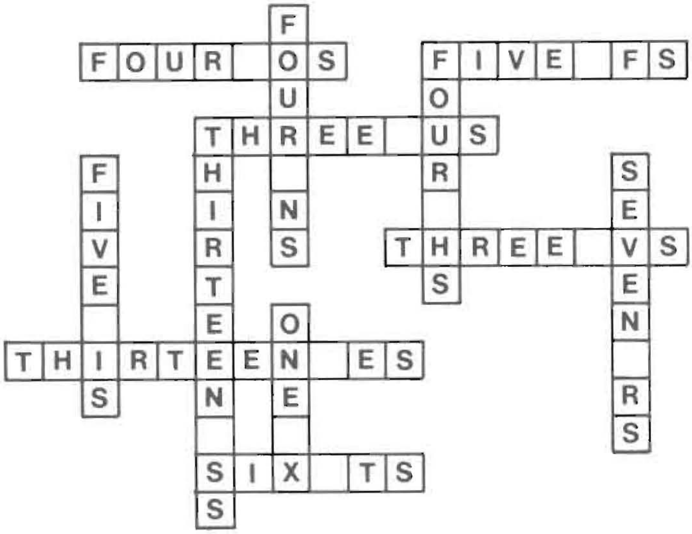
Here we find 12 words using 41 letters = 3.416 lpw or 0.29 wpl, a stiffer broth, although the 12 referent letters are now to be inferred (partly through word positions), a dubious device, perhaps. However, a breakthrough comes in seeing that the same letters can be used twice as in a crossword. Writing out TEN E, ONE F, ... on twelve squared strips of card, at length trial-and-error shuffling led to my first self-intersecting reflexicon (at left) which solves the problem of the referent letters.

This was a good start, but OTT, NWW, EOO, etc. struck me as pseudoword blemishes. A different layout wouldn't help, either, since ONE H and ONE R always remain bonded together with HR in THREE. No, to escape this problem called for a new set of items involving fewer intersections per strip so as to win elbow-room. This brings us to a key insight.

Twelve strips bearing 12 excess letters imply 12 intersections. Yet, N strips can cross at most N-1 times **unless linked to include a closed chain.** (Look at ONE X, SIX N, SEVEN O and TWO S in the upper diagram.) Contriving such a loop is the major constraint in devising solution layouts. Thus, a new list requiring fewer intersections than strips makes for a big gain in layout flexibility (and vice versa), although two or more fewer will imply a non-connected pattern. To avoid this, the obvious course then is to seek an N item list involving N-1 excess letters (= intersections); an example is given at the left.

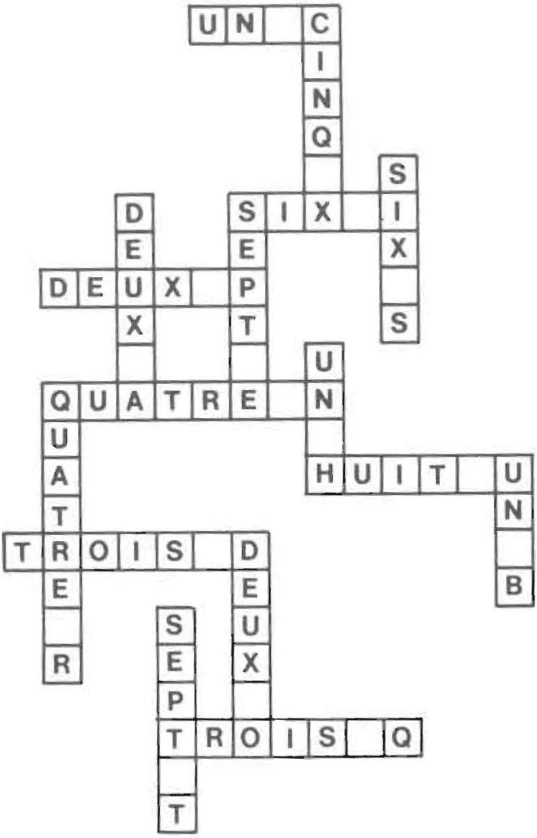


This is more like it: no pseudowords and 3.846 lpw or 0.26 wpl, which, with the words now spatially interlocking, is virtually alphabet jelly! The trouble is that now one letter, F, is alone in not occupying an intersection, a niggling asymmetry. At some loss in semantic density, however, restoring plural S is another way to win room for maneuver, as in the diagram below.

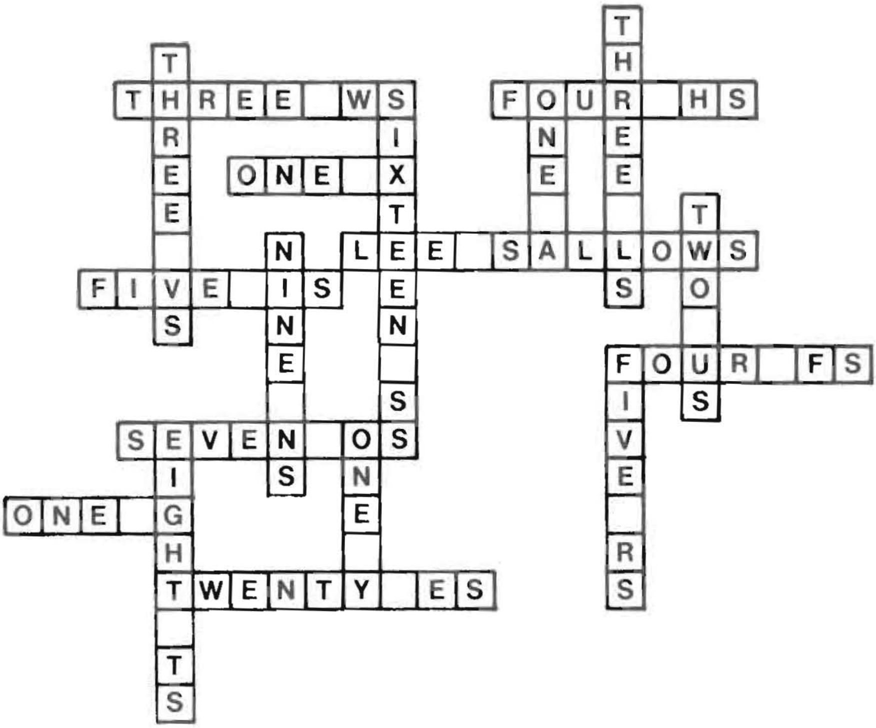


Here we are back to 12 strips and 12 intersections (necessitating a loop), each occupied by one of the 12 letters occurring. On consideration. this is a remarkable property, more so than first sight suggests. since it depends on finding a list in which the letters outnumber their totals by one exactly, the excess then vanishing in intersects. The list used in the above crossword is thus exceptional. For example. no French or Italian equivalent exists. Unusually, however, English enjoys two such lists, the second comprising 13 words, although its internal peculiarities impede the construction of elegant self-intersecting layouts. Some readers may like to try their hand at constructing the crossword; the totals are E:15, F:8, G:1, H:3, I:5, L:1, N 4, O:5, R:5, S:11, T:4, U:3 and V:4.

Of course. in general there is nothing against letters appearing on intersects more or less than once, as with the French and Italian reflexicons given at the top of the next page (neither language calls for plural S). Both of these illustrate a further trick in the reflexiconographer's repertoire: the use of ONE # (here, UN B, UNO A, UNO B) as unobtrusively appended dummy text. This is a useful stratagem when "pure" solutions cannot otherwise be found, although the arbitrariness of letter used (UN B could equally be UN Z, say) detracts from their logological elegance, a point to bear in mind when assessing the merits of different specimens. Dummy text may take more conventional forms, of course, as in the reflexicon at the bottom of the next page, where intersections outnumber strips, a fact reflected in multiple loops. However, the construction of such specialties is demanding, to say the least.

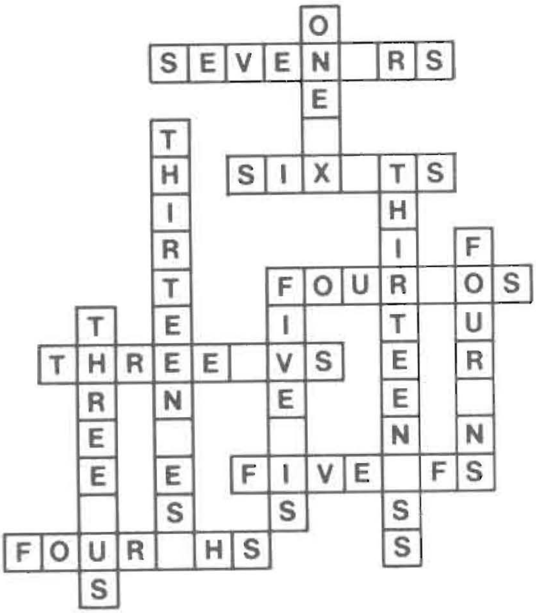

UN C
SIX
DEUX
QUATRE
HUIT
TROIS
TROIS Q


TRE S
UNO A
SEI
DUE R
CINQUE
OTTO E


THREE WS
FOUR HS
ONE
LEE SALLOWS
FIVE IS
FOUR FS
SEVEN OS
ONE
TWENTY ES

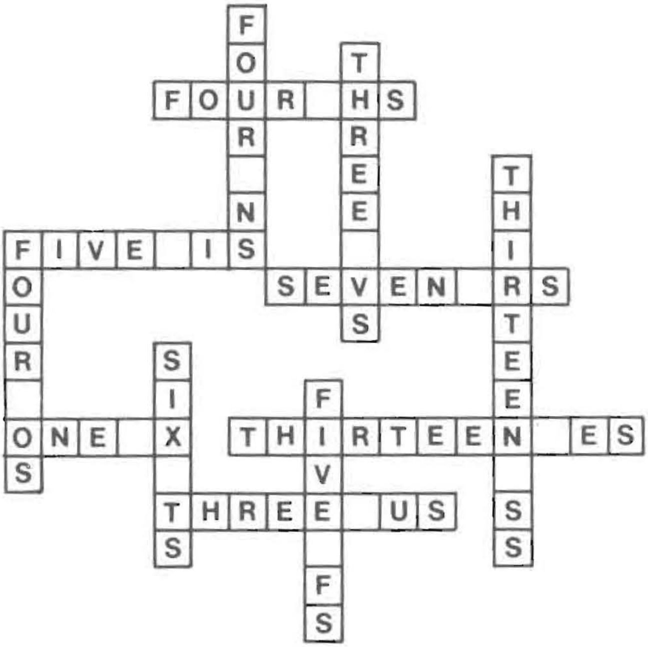
Some loops are not what they seem. The reflexicon below exhibits **pseudoloops** and the two ways they arise: via intersection on a blank, viz. THIRTEEN SS and FIVE FS, and via abutment onto a blank, viz. THIRTEEN ES and FOUR HS. The single real loop here is formed by FOUR OS, FOUR NS, FIVE FS, and FIVE IS.

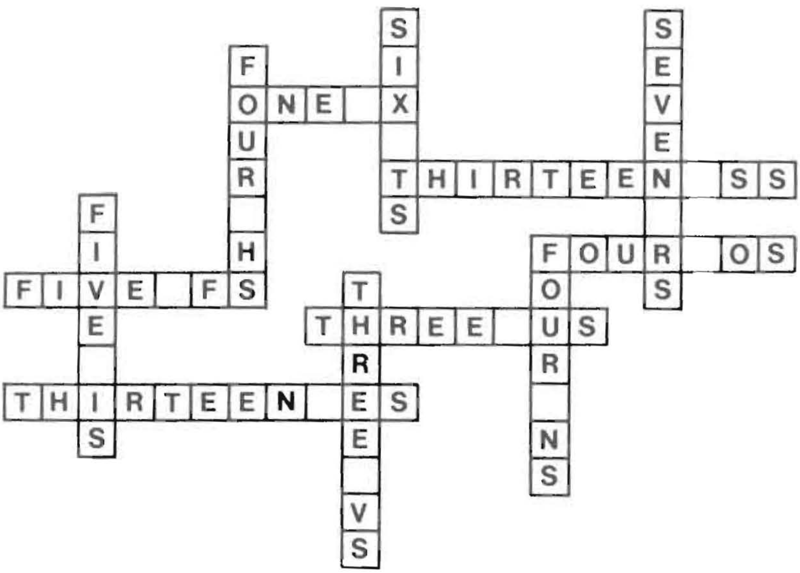
Pseudoloops can make for compacter layouts, a fact seen in comparing the diagram below with the one two pages earlier, both of which, be it noted, use the **same entries** (the special set of strips), whereas the two patterns occupy rectangles of 14x16 and 14x18, respectively.



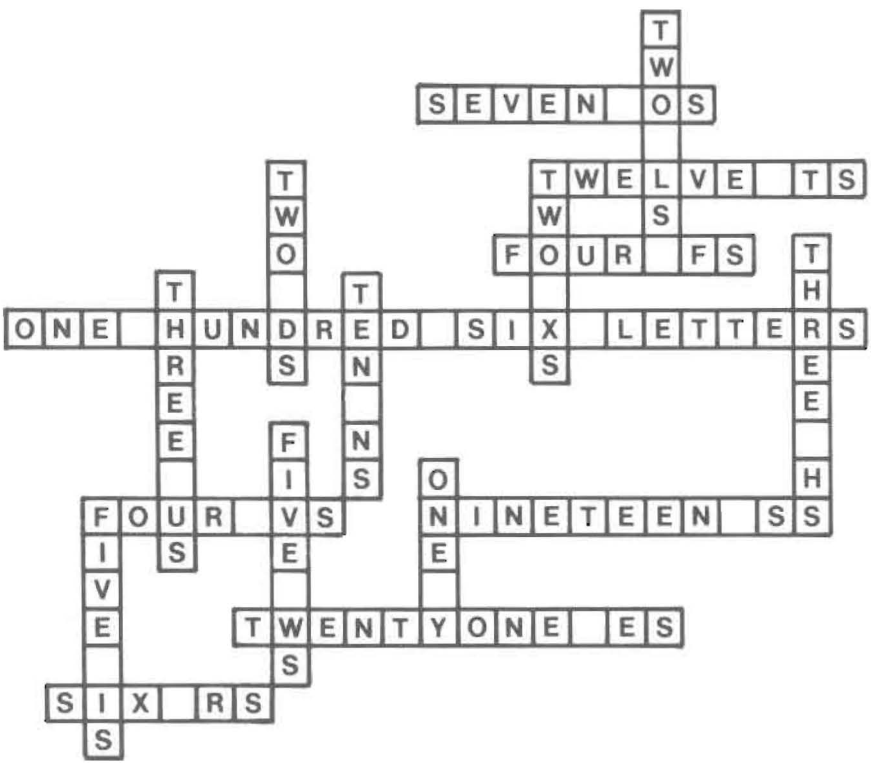
Two natural questions now arise: (a) how many distinct (fully connected) self-intersecting reflexicons can be formed from this set of strips? (b) which of them is the most compact? To seek answers, Victor Eijkhout, a mathematical friend, wrote a recursive strip-shuffling computer program able to scan for solutions. However, although several days running on a mainframe computer produced **thousands** of alternative solution layouts, it became clear there was no chance of the job terminating within any feasible time-scale. The two questions thus remain unanswered. The reflexicon given above, which was hand-produced, remains the most compact specimen known.

Nevertheless, at my suggestion Eijkhout set his (slightly modified) program to work on a new but related search which was to bear fantastic fruit. The two reflexicons on the next page embody two jewels of logology (we seem to have reached alphabet ice). Here are the classic strips again, the loop now realized as the entire set holding hands in a single twelve-linked bracelet! The pair shown are among 18 such specimens found by the program, not counting rotations and reflections, but including trivial variations such as when FIVE IS is switched with FIVE FS.





Marvellous as Eijkhout's finds are, further collector's pieces probably await discovery. For example, could there exist a reflexicon with a **symmetrical** layout? A **congruent** pair showing distinct solution entries? A **3-dimensional** bracelet that forms a knot? A (possibly interlacing) **co-descriptive** pair? A **pangrammatic** reflexicon (without dummy text)? The list is easily extended. In the meantime, one classic specimen has passed unmentioned. The reflexicon below again features 12 intersections, each occupied by one of the 12 letters occurring, although now there is no plural S. It is a relative of the first self-intersector examined, where the number of excess letters also matches the number of items, but which cannot be solved without pseudowords. As with the list used earlier, a second list with the same property (but minus plural S) has been found. (A third trivial case is FOUR [F], FOUR [O], FOUR [U], FOUR [R].) The analogous question then arises: how many distinct solutions can be formed from the entries in the reflexicon below?



Lastly, to conclude this brief review, I offer at the bottom of the preceding page a final example of the state-of-the-art, a reflexicon that incorporates its own letter-total. Can one be found using still fewer than 106 letters? Another tough challenge for the computational logologist! However, the seemingly-insuperable problem that keeps me awake at nights is how to produce a self-descriptor that will tell us what letters it uses **where**?

REFERENCE

[1] Hofstadter, D. Metamagical Themas (Basic Books, 1985), p. 364.

METAMORPHOSES OF COMMONPLACES

Willard Espy subtitles The Word's Gotten Out (Clarkson Potter, 1989) "a commonplace book" -- that is, a book of memorabilia, youthful adventures, passages encountered in reading, clippings sent in by readers, and the like. Yet, in Espy's skilled hands, this often-mundane material suffers a sea-change, illuminated by clever turns of phrase, satirized by bits of light verse. The English language is the unifying theme: its logical or orthographic inconsistencies, its long-forgotten words, its playfulness as exemplified by rebus or palindrome (but not by word square, which he detests). One can sample some of the book's flavor by referring to Espy's Kickshaws in the May 1987 ***Word Ways****, where about half of the material is recycled in his book; he also prints a set of atrocious literary-name puns by Derek Pell that appeared in February 1983, and a Polish pangram and 1903 rebus cited by Will Shortz in the May 1986 Kickshaws. A sampler of other material:*

Punctured poetry: I heard a fly buzz when I died (Emily Dickinson) / Of fruit sprayed with insecticide (Espy)
Revised Etymologies: CONDOM originally came from the Latin **contra Domine***, which means contrary to the will of God*
Clement Wood's expandible palindrome: Di, Al, Togo, Böll, Edna, Todd, Adolf ... Flo, Dad,Dot and El Lobo got laid
Presidential Mini-biographies: GERALD FORD Proved, which much astounded some / That he could walk and still chewgum
John Culkin: Recently, I realized that the word **dollop** *reads the same way upside-down. Would you call it an invertogram?*