THE STRUCTURE OF A WORD NETWORK

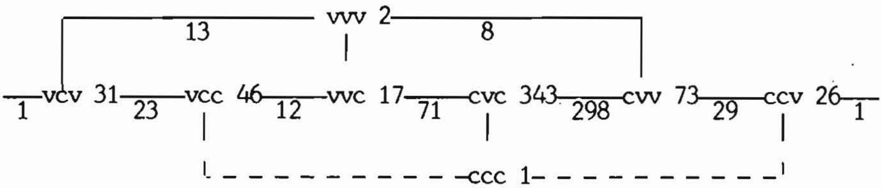
A. ROSS ECKLER
Morristown, New Jersey
LEONARD J. GORDON
Chico, California

A word network consists of a collection of words of the same length, in which any two words that differ by only a single letter in one position (such as EAT and FAT, or STORE and SWORE) are joined by a link. A word ladder, originally invented by Lewis Carroll, joins two words in a network to each other by means of a path traced through successive links (such as EAT-BAT-BAN-BIN-PIN). Ordinarily, there are many such possible paths involving many different intermediate words.

Links involving the change from a vowel to a consonant, or a consonant to a vowel, are usually more difficult to find than consonant-consonant or vowel-vowel links; statistical evidence in support of this assertion appears at the end of this article. In the main part of the article, we partition a word network into subnetworks consisting of words having the same vowel-consonant patterns and look for word tours that visit as many of these subnetworks as possible in as few steps as possible.

Some further definitions are needed here. Tours come in two varieties, open and closed. An open tour is defined as a sequence of words successively joined by links, passing through different subnetworks in turn but never returning to a subnetwork already quitted. A closed tour returns to its starting word and, perforce, the subnetwork in which it is located. Vowels are defined to be AEIOU, and Y if it is located in any position other than the initial one in a word; thus, YACHT has the vowel-consonant pattern cvccc and SKY, ccv. A complete tour (open or closed) visits all possible subnetworks, and an incomplete tour a subset of all subnetworks. The minimum number of steps is achieved in a no-dwell tour, in which each subnetwork is represented by only a single word in the tour. Obviously, complete tours are more difficult to construct than incomplete ones, closed tours more difficult than open ones, and no-dwell tours than ones that spend time in a subnetwork. In general, the shorter the word length and the larger the dictionary size, the better chance one has of upgrading tour quality.

Let us fix ideas by considering the three-letter word network, which divides into eight subnetworks with unique vowel-consonant patterns. If one uses the Merriam-Webster Pocket Dictionary of 539 three-letter words (excluding spelled-out abbreviations such as DDT and TNT), the pattern of subnetwork connections is quite simple to portray:



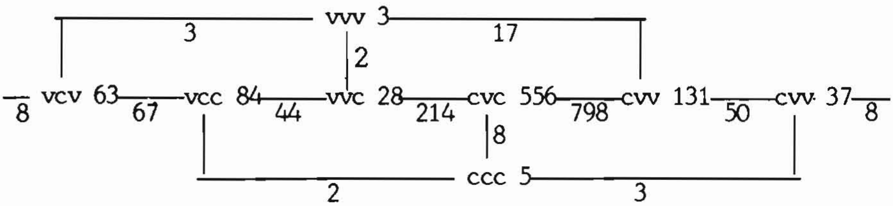
Dotted lines indicate that no actual links involving Pocket Dictionary words exist between the indicated subnetworks, even though such links are theoretically possible. The numbers following the patterns indicate the number of words in each subnetwork, and the numbers below the solid lines indicate the number of internetwork links.

What sort of tours are possible in the Pocket Dictionary network? An incomplete no-dwell closed tour of six subnetworks is the best that one can hope to accomplish; is it, in fact, possible? There are two cases to consider: -vcv-vcc-vvc-cvc-cvv-ccv- and -vvc-vcc-vvc-cvc-cvv-vvv-. In the first case, a little trial and error on the weakest link soon reveals that it cannot be done. The only Pocket Dictionary link between ccv and vcv is BRA-ERA. From BRA, one can construct no-dwell ladders through BOA to BOB, BOG, BOW or BOX, but none of these connects with a vvc word; similarly, ERA goes through ERR and EAR to CAR, FAR, MAR, PAR, TAR and WAR. Although this creates a no-dwell open tour of all six subnetworks, a closed no-dwell tour is impossible. In fact, the shortest closed tour of these six subnetworks is -ERA-ERR-EAR-(BAR-BAG-BOG)-BOA-BRA-. In the second case, the only vvv words are EYE and AYE. One can go from either word to DYE, BYE, RYE or LYE, but there are no links between any of these and a cvc word, so a no-dwell tour is again not possible. The shortest closed tour of these subnetworks appears to be nine: -EYE-(DYE-DUE-CUE)-(CUR-CAR)-EAR-ERR-ERE-.

Note that it is possible to construct a no-dwell open tour of seven subnetworks: EYE-ERE-ERR-EAR-PAR-PAY-PLY.

Since there are obviously no possible closed tours visiting just four subnetworks, one concludes that a no-dwell closed tour is impossible for three-letter words in the Pocket Webster Dictionary.

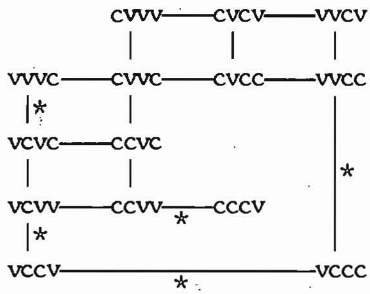
If the 907 three-letter words of the Official Scrabble Players Dictionary are used, the subnetworks are much more connected:



The five all-consonant words are CWM, NTH, PHT, SHH and TSK, and the three all-vowel words, AYE, EYE and EAU.

A closed complete no-dwell tour is impossible, but a closed no-dwell tour of six vowel-consonant subnetworks can be achieved by the highly-symmetric -BRA-BAA-BAR-EAR-ERR-ERA-. Furthermore, an open complete no-dwell tour is now possible: POI-PHI-PHT-PIT-AIT-ALT-ALE-AYE.

The study of four-letter networks is more interesting. To begin with, there are sixteen subnetworks instead of eight; it is difficult to present the full network in a comprehensible form if all possible connections are included. Fourteen of the subnetworks contain Pocket Webster words.



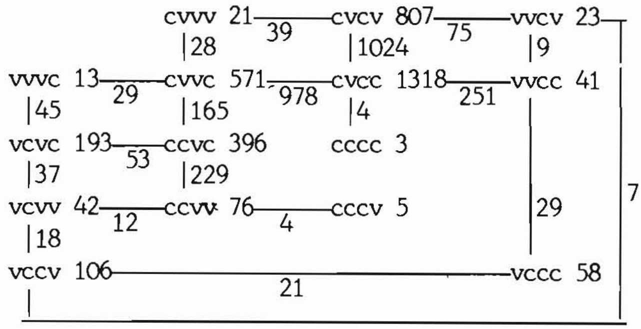
Examining this network, we see that many closed tours of length six or eight are theoretically possible through the subnetworks. However, the links labeled with asterisks have only one word-pair apiece: AMAH-AYAH, EAST-ERST, INCH-INCA, AWAY-AWRY, and SPRY-SPAY. The one remaining tour of length six is not possible, either, because of the scarcity of cvvv words (QUAI, QUAY, MOUE, ROUE, KAYO, LIEU) and the limited number of links for these. The longest incomplete no-dwell close tour visits only four subnetworks: -EAST-EASE-CASE-CAST-. The shortest closed tour of six subnetworks with dwell allowed appears to be 12: -ROUE-(ROVE-COVE-CAVE-CASE)-EASE-EAST-(CAST-COST)-(COOT-ROOT-ROUT)-.

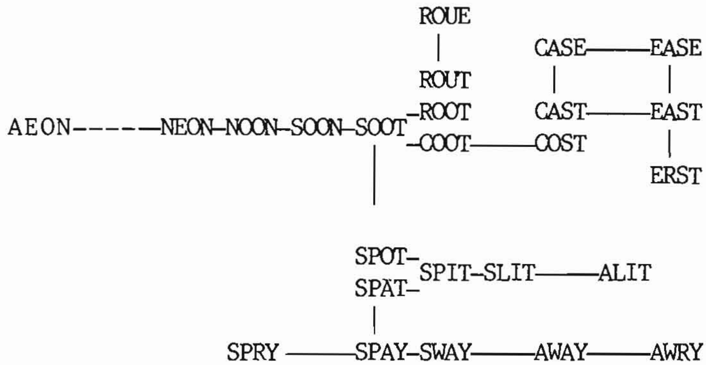
Further examination of the network reveals that it is theoretically possible to find an open tour which visits 10 subnetworks. One cannot expect this to be a no-dwell tour, and in fact the minimum-length tour appears to take 14 words: AWRY-AWAY-(SWAY-SPAY)-(SPAT-SPOT)-(SOOT-COOT)-(COST-CAST)-CASE-EASE-EAST-ERST.

What is the minimum number of words needed to provide an opportunity to travel from any subnetwork to any other subnetwork? The resultant structure can be a tour (as we have seen in the three-letter network), but it need not be restricted to such a format--in general, it will resemble a tree. For the four-letter network based on the Merriam-Webster Pocket Dictionary, at least 25 words are necessary to visit all 14 subnetworks; the diagram is given on the next page.

The four-letter OSPD network contains words in fifteen out of the sixteen possible subnetworks, all linked to each other (in the cccc subnetwork, CWMS via CAMS, PSST via PAST, PEST, or POST, and TSKS via ASKS). There is a much higher degree of connectivity than for the Pocket Dictionary words; in the interest of simplicity, one cccc-vccc link, seven cccv-cvcv links, and two ccvv-cvvv links have been left out.

The longest closed no-dwell tour that can be found visits eight





different subnetworks: -EYES-ERES-ERNS-ERNE-EYNE-TYNE-TYEE-TYES- The longest open no-dwell tour visits 13 subnetworks: FLEX-FLEA-OLEA-OLES-OYES-TYES-TYEE-TYNE-EYNE-ERNE-ERNS-EONS-TONS.

If one does not insist on the no-dwell restriction, can all 15 subnetworks be toured -- that is, visited in turn without repeating a visit to a subnetwork after departure from it? This is an example of a Hamiltonian line in graph theory; no general solution to the problem of finding a Hamiltonian line for an arbitrary graph is known, but trial and error suggests that it is impossible for the OSPD network. In other words, the largest number of subnetworks that can be toured is 14. The minimum number of words needed to accomplish this is unknown; readers are invited to improve on the following example:

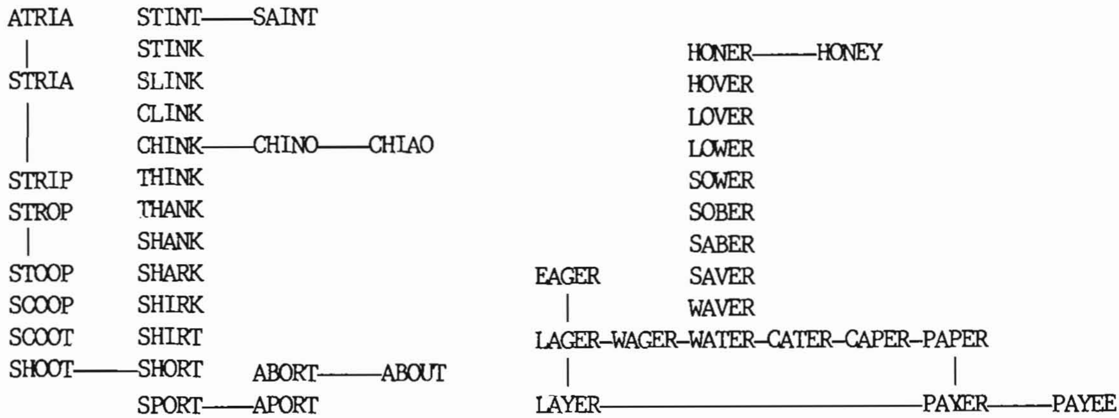
CWMS-(CAMS-CANS-CONS-TONS)-EONS-ERNS-ERNE-EYNE-TYNE-TYEE-TYES-OYES-OLES-OLEA-FLEA-FLEX

If one does not insist on a tour, all fifteen subnetworks can, of course, be connected. What is the fewest number of words needed? It is possible to do it with 20, by attaching the words THRO-TYRO-TYRE to TYNE in the tour given above.

The study of five-letter networks is less complete. To begin with, the diagram of subnetwork interconnections is exceedingly complex. Even for the Merriam-Webster Pocket Dictionary, 26 of the 32 vowel-consonant patterns are represented. The OSPD contains 29, with only the patterns vvvvc, vvvcv and vvvvv left out (CRWTH is one of the two ccccc words). The table on the next page summarizes the subnetworks and their connections.

Subnetwork	Number	Internal Connections	Connections to Other Subnetworks
A cvcvc	2101	7258	D-530 N-208 J-196 O-189 Q-7
B cvccv	1116	1862	D-364 J-74 I-93 U-5
C ccvcc	874	2189	F-280 G-327 E-156 L-55
D cvccc	871	3257	A-530 B-364 E-351 S-90
E cvvcc	647	1890	C-156 D-351 I-146 N-31 Y-10
F ccvcv	430	848	C-280 I-47 K-61 W-11
G ccvvc	420	884	C-327 N-12 W-18 Q-8
H vccvc	368	317	P-20 M-34 O-35 Q-4 &-1
I cvvcv	296	354	B-93 E-146 F-47 T-16
J cvcvv	231	157	A-196 B-74 T-8 V-2 X-4
K vcvcv	192	144	F-61 L-43 R-9 Z-1
L vcvcc	183	142	C-55 K-43 M-15 Y-10
M vcvvc	110	53	H-34 L-15
N cvvvc	100	90	A-208 E-31 G-12 T-11
O vvcvc	66	58	A-189 H-35 S-3 V-3
P vccvv	58	25	H-20 R-3 X-1
Q cccvc	48	61	A-7 G-8 H-4 X-14
R vcccv	42	15	K-9 P-3 &-1
S vvccc	19	6	D-90 O-3 U-4
T cvvvv	14	12	I-16 J-8 N-11
U vvccv	13	2	B-5 S-4
V vvcvv	12	3	J-2 O-3
W ccvvv	11	4	F-11 G-18
X cccvv	11	6	J-4 P-1 Q-14
Y vvvcc	6	1	E-10 L-10
Z vcvvv	4	0	K-1
& vcccc	3	0	H-1 R-1
# ccccv	2	0	
$ ccccc	2	0	

It is difficult to connect very many of the Pocket Dictionary subnetworks, even when the requirements of no-dwell and tour are eliminated. At the left, a set of 27 words joins ten subnetworks; at the right, a set of 20 words joins five other subnetworks.



Can these two sections be joined? The fragment

FOYER----FLYER-FLIER-FLEER-FLEET-SLEET-SHEEP-SWEEP-SWEET
SHEET SWEAT

nearly does the job; if SHEAT or SWOAT existed, it would connect via SHOAT to SHOOT, and if FAYER or POYER existed, it would connect to PAYER.

Of course it is impossible to join all 26 subnetworks, for some of them (notably cvccc, with --GHT words) contain no words in the main network. And, even if two subnetworks both contain words in the main network, there is no guarantee that there exists a path between them which traverses intermediate subnetworks seriatim (with no exits and re-entries).

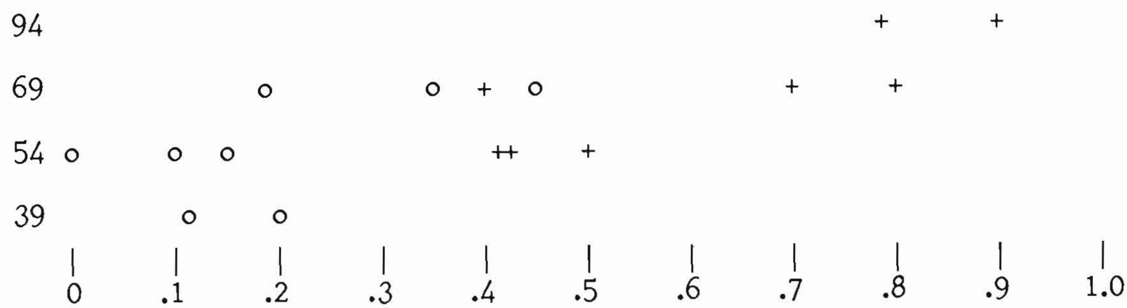
How many subnetworks can be connected by OSPD five-letter words and what is the minimum number of words needed to do the job?

* * *

We conclude this article by presenting some statistical evidence for the assertion that it is more difficult to find a link changing a vowel to a consonant (or vice versa), than it is to find a link changing a vowel to a vowel, or a consonant to a consonant. In other words, we exhibit some justification for looking at the word network in terms of cohesive vowel-consonant subnetworks joined somewhat more tenuously to each other.

The measure of connectivity we use is a simple one. If one selects a word at random in a subnetwork having a specific vowel-consonant pattern, what is the probability that there exists one or more links from this word to other words in the subnetwork? to other words in other subnetworks? If the assertion is true, the first probability should, for most subnetworks, be larger than the second probability.

Samples of 20 to 30 five-letter Pocket Dictionary words were drawn from eight different subnetworks, and the corresponding pairs of probabilities calculated. These are plotted in the graph below,



Probability of link to another word
+ within subnetwork o outside subnetwork

with + indicating within-network probabilities, and o, between-network probabilities. The latter probabilities are almost invariably smaller than the former, lending substance to the assertion.

The vertical axis of the graph is an attempt to make the comparison a fairer one, by showing the theoretical number of ways that a word in one subnetwork can be transformed to another word in the same subnetwork, or to a word in a different subnetwork. For example, if the word being sampled is HOVER (pattern cvcvc), it can be theoretically changed nineteen ways in the first position (BOVER, COVER, DOVER, FOVER, ..., YOVER, ZOVER), five ways in the second position (HAVER, HEVER, HIVER, HUVER, HYVER), twenty ways in the third and fifth positions, and five ways in the fourth position (HOVAR, HOVIR, HOVOR, HOVUR, HOVYR) -- a total of 69 changes. However, there are only 54 ways that HOVER can be altered to a potential word in another subnetwork, which suggests that the observed probability of the second event ought to be judged a bit more leniently than the observed probability of the first event. The possibilities are

	within	between
words with 2 consonants	54	69
words with 3 consonants	69	54
words with 4 consonants	94	39

Not surprisingly, the observed probabilities in the graph do appear to depend somewhat on the number of links theoretically possible.

Does the pattern of this graph hold for the more comprehensive statistics on the five-letter OSPD words? The actual numbers of words having links with other words, either within or outside subnetworks, was not tabulated, but these can be approximately estimated from the statistics that were collected:

prob (random word in subnetwork i is linked to one or more other words in the same subnetwork) =
1 - exp[-2(links in i)/(words in i)]

prob (random word in subnetwork i is linked to one or more words in other subnetworks) =
1 - exp[-(sum of links between i and other networks)/(words in i)]

For example, for subnetwork A (cvcvc pattern) the inside probability is estimated to be 1 - exp[-2(7258)/2101] = 0.999, and the outside probability, 1 - exp[-(530+208+196+189+7)/2101] = 0.416. In other words, there are only about two of the 2101 cvcvc words that do not link with other cvcvc words, but some 60 per cent do not have outside links.

For a handful of subnetworks, the estimated probabilities reverse: the estimated probability that a random word has one or more inside links is <u>less</u> than the estimated probability that a random word has one or more outside links, suggesting that these vowel-consonant groupings may be less than apparent. Two subnetworks, vvccc and vvvcc, show extreme behavior in this regard. The estimated inside probability for vvccc words is 0.47, and the estimated outside probability is 0.99. Examining the nineteen vvccc words in the OSPD, one finds, in fact, that ten have inside links and fifteen have outside links, leading to actual probabilities of 0.53

and 0.79. (The estimate 0.99 markedly exceeds the actual probability, because some words like EIGHT have an unusually large number of links; the idealized assumptions underlying the above formulas do not hold.) For the record, the ten linked vvccc words are AIRNS-AIRTS-AIRTH, OUGHT-AUGHT, EASTS-OASTS-OUSTS and EARLS-EARNS; the only words without outside links are AUGHT, OUGHT, OOMPH, and OUPHS. The estimated inside probability for vvvcc words is 0.28, and the estimated outside probability, 0.96. For the four vvvcc words found, the actual inside and outside probabilities are 0.50 and 1.00, respectively.

MUSHFAKERS, BAWCOCKS AND GONGOOZLERS

These and ninety-two other unusual words are defined and, in most cases, illustrated with droll cartoons in Lost Words of the English Language (Bob Adams, Inc., Holbrook MA; $3.95 in paperback). The authors, Robert Schachner and John Whited, have selected words that are obscure, describe obsolete concepts or actions, and sound funny. Some choices (amanuensis, billingsgate, corvine, bildungsroman) are rather tame, but others (taghairm, whipjack, agrostographer) are right on target. You haven't heard of these? They are, respectively, prognostication while wearing a bullock's hide under a waterfall, a panhandler pretending to be a shipwrecked sailor, and a writer whose subject is grass.

The closest analogue in **Word Ways** *is Eric Albert's "A Lode for Logastelli" in the November 1988 issue; eight of his words appear in Lost Words. Mrs. Byrne's Dictionary also comes to mind; 58 of her words are used. Some definitions are not well-supported by dictionaries. A* **fossarian** *was a minor clerical order whose main (sole?) job was gravedigging;* **owling** *is usually defined as smuggling sheep or wool out of a country, not specifically England to France; and* **tyromancy** *is divination by means of cheese, not solely by its coagulation.* **Taghairm** *and* **capric** *are misspelled. The authors missed the most striking definition of* **merkin**, *false female pubic hair (see the discussion in Maledicta 8, 1984-5). Each word "has been thoroughly researched" and "appears in at least one major published dictionary" but no specific references are given. However, it's probably unrealistic to expect a high level of scholarship in a book specifically written to amuse the reader.*

A final note: the middle (47th) word is **famulus**, *earlier in the alphabet than all but one of 50 alphabetical lists surveyed by Alan Frank in the August 1984* **Word Ways**. *Did the authors get tired when they reached the last half of the dictionary, or are funny words heavily concentrated in the early letters?*