# DIRECTED WORD CHAIN NETWORKS (PART 1)

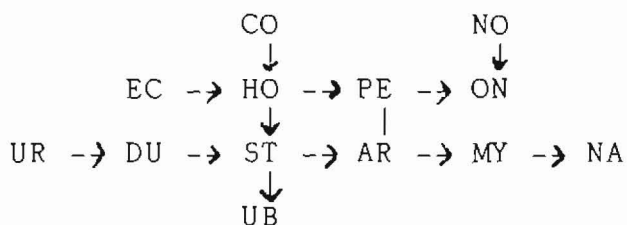A. ROSS ECKLER
Morristown, New Jersey

The February 1991 **Word Ways** introduced the concept of a word string network: a set of n-letter words connected by directed links. Word A is joined to Word B by a directed link if Word B is formed by removing the first letter of Word A and placing a letter (not necessarily the same one) at the end of Word A (such as TEA to EAT, or SAR1 to AR1D). Note that a link does not connect EAT to TEA; one must proceed from EAT to ATE, and by a second link from ATE to TEA. The sequence of words traced out as one moves along the directed links forms a **word string** (also known as a **word** stair) in which any set of n consecutive letters forms a word (such as WASHERAGEMUD with three-letter words).

Word strings are a special case of word chains, in which any number of letters (not just one) are dropped from the start of Word A and added to the end of Word B. In what follows, only **regular** word chains are considered: the number of letters in the word and the number of letters dropped and added remain constant. Define (n, m) **word chains** as those consisting of n-letter words in which the first m letters are removed and a fresh set of m letters are added at the end. Word strings are merely (n, 1) word chains.

**Word chain networks**, then, consist of n-letter words connected by directed links, in which Word A in the network is linked to Word B if Word B can be formed by dropping the first m letters of Word A and adding m letters to the end of Word A. Words in the network can have many links both reaching them and leaving them.

This article examines a special kind of word chain network: one in which m = n/2. Such networks can be readily diagrammed by introducing a closely-related network: one consisting of half-word sequences instead of full words. Leonard Gordon has proposed that these half-word sequences be called frags. In such a network, each word consists of two frags connected by a directed link. For example, here is a small sample of frags and words in the (4,2) word chain network:

```
              CO              NO
              ↓               ↓
    EC  -→  HO  -→  PE  -→  ON
              ↓               |
UR  -→  DU  -→  ST  -→  AR  -→  MY  -→  NA
              ↓
             UB
```
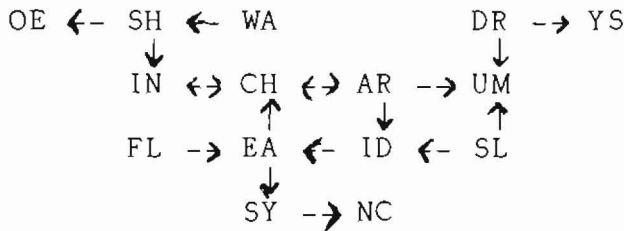
The full network is an extremely complex tangle of directed links

joining several hundred different frags, quite impossible to dia-
gram in full. However, one can gain some understanding of its
characteristics. To aid in this, one must sort frags into various
classes.

In general, a word chain network will contain a core, consist-
ing of all words that can be revisited by a closed chain (such
as CHAR-ARID-IDEA-EACH-CHAR). Frags contained in closed-chain
words are called core frags.

Other frags are defined with respect to the core. If a frag can
only appear at the start of a word, and if that word can be joined
by a word chain to the core, it is called a **beginner**. Similarly,
if a frag can appear only at the end of a word, and that word
can be reached by a word chain from the core, it is called an
**ender**. Intermediate frags required to join beginner frags to core
frags are called **preceders**, and intermediate frags required to join
core frags to ender frags are called **followers**. Finally, frags that
cannot participate in word chains reaching or leaving the core
are called **isolates**. One should think of a cloud of core frags sur-
rounded by an inner ring of preceders and followers, and an outer
ring of beginners and enders. Consider the following network:

```
OE ←- SH ←  WA              DR -→ YS
       ↓                          ↓
      IN ←→ CH ←→ AR -→ UM
             ↑        ↓     ↑
   FL -→ EA ←- ID ←- SL
             ↓
           SY -→ NC
```

In the above, CH, AR, ID, EA and IT are core frags; WA, SL and
FL are beginners; NC and UM are enders; SH is a preceder; SY
is a follower; and OE, DR and YS are isolates. Note the two two-
way directed links (between IN and CH, and CH and AR).

In theory, one can also conceive of bypassers, frags joining
beginner words with ender words by chains not passing through
the core. One can also conceive of separated cores that can be
accessed in sequence by directed links. Such phenomena do not
occur in the (4,2) directed network discussed in this article, but
can be seen in the (3,1) word chain network in the February issue.

The relative abundances of different frag types can vary enor-
mously, depending on the density of the word chain network being
studied. Consider a specific corpus: 1949 words taken from boldface
entries in the 1964 New Merriam-Webster Pocket Dictionary, or in
its successor, the 1974 Merriam-Webster Pocket Dictionary. In the
(4,2) word chains based on this corpus, core frags, beginners and
enders are common, but preceders, followers and isolates are rare,
and bypassers nonexistent. However, there are many words consist-
ing of a beginner and an ender which directly bypass the core
(such as SLUM in the example above).

The table on the next page summarizes the status of each frag
present in the Pocket Dictionary corpus: 133 core frags, 62 begin-
ners, 88 enders, 9 preceders, 6 followers, and 9 isolates.

Second Letter

First Letter

|   | 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A |   | c | b | c | p | c | c | f | c | b | c | c | c | c |   | c |   | c | c | c | c | p | c | c | c |   |
| B | p |   |   | c |   |   |   | b |   |   |   | b |   |   |   | c |   | b |   | e |   |   |   |   | i |   |
| C | c |   |   | c |   |   |   | c | c |   |   | e | b |   |   | c |   | b | e | e | b |   |   | c | b |   |
| D | c |   |   | c |   |   |   | b | c |   |   |   |   | c |   |   |   | b | e |   | c |   |   |   | c |   |
| E | c |   | b | c | f | e | b |   |   | c | c | c | f | e | c |   |   | c | c | c | e | b | c | c | c |   |
| F | c |   |   | c | e |   |   | b |   |   |   | b |   |   |   |   |   | b |   | e | b |   |   |   | e |   |
| G | c |   |   | c |   | e | c | c |   |   |   | b |   | p | c |   |   | b | e |   | b |   |   |   | c |   |
| H | b |   |   | p |   |   |   | b |   |   |   |   | i | c |   |   |   |   | e | p |   |   |   |   | c |   |
| I | f | p | c | c | e | f | e |   |   | b | c | e | c | c | e |   |   | c | c | c |   | e |   |   | i |   |
| J | c |   |   | b |   |   |   | b |   |   |   |   |   | b |   |   |   | b |   |   |   |   |   |   |   |   |
| K | c |   |   | c |   |   |   | c | c |   |   |   | b | b |   |   |   | b |   |   |   |   |   |   | c |   |
| L | c | e | e | e | c | e |   |   | c |   | e | e | e | e | c | e |   | e | e | b |   |   |   |   | c |   |
| M | c | e |   | c |   |   |   | b |   |   |   | e | c | e |   | e |   |   | b |   |   |   |   |   | c |   |
| N | c |   | e | e | c |   | e |   | c |   | e |   | i | c |   |   |   | e | e | c |   |   | e | e |   |   |
| O | b | p | i | c | e | e | c | e | b |   | c | c | c | c | b | c |   | c | e | e | c | b | c | i | e |   |
| P | c |   |   | c |   |   |   | i | c |   |   |   | b |   | c | e |   | b | e | e | b |   |   |   | c |   |
| Q |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   | b |   |   |   |   |   |   |   |   |
| R | c | e |   | e | c | e | e |   | c |   | e | e | e | e | c | e |   | e | e | e | p |   |   |   | e |   |
| S | c | c |   | c |   |   |   | c | b |   | c | b | b | b | c | c |   |   | e | c | b |   | b |   | f |   |
| T | c |   |   | c |   |   |   | c | c |   |   |   |   | c |   |   |   | b | c | e | b |   | b |   | c | b |
| U |   | e |   | e | e | i | c |   |   | c | e | c |   | c |   |   |   | c | c | c |   |   | e |   |   |   |
| V | c |   |   | c |   |   |   | c |   |   |   |   | b |   |   |   |   |   |   |   |   |   |   |   | e |   |
| W | b |   | e | b |   |   |   | b | c |   |   | e | e | e | b | e |   | b | e | e |   |   |   |   | e |   |
| X |   |   |   |   |   |   |   | e |   |   |   |   | b |   |   |   |   |   | e |   |   |   |   |   | e |   |
| Y | c |   |   | c |   |   |   |   | e |   |   |   | c |   |   |   |   | e |   | b |   |   | e |   |   |   |
| Z | i |   |   | c |   |   |   | c |   |   |   |   |   | c |   |   |   | b |   |   |   |   | e | e |   |   |

Words can be classifed according to the frags composing them. The commonest varieties are core-core (849 words), core-ender (434 words), beginner-core (298 words) and beginner-ender (209 words). Samples of the rarer varieties include:

```
beginner-preceder: TUBA, BRAE, SLAV, SIGN, SLOB, JEHU, ACHE
beginner-follower: WAIF, WHEN, MIEN, FLEE, GLEE, TREE, OVEN
preceder-ender: BARN, AERY, OBOE, HURT, HERS, RUMP
follower-ender: IAMB, EERY, IFFY, AHOY, ENVY, SYNC
preceder-core: BANE, AEON, AVER, GNAT, OBIT, HUGE, HERO, RUST
core-follower: ARIA, ALEE, COIF, SHAH, AMEN, EASY
isolate-ender: ZANY, OXEN, UFOS
beginner-isolate: BLOC, QUIZ, WHIZ, FINN, JOHN
preceder-isolate: BABY, RUBY
isolate-isolate: PHlZ
```

No examples exist for preceder-follower, preceder-preceder, follower-follower, or isolate-follower.

In two-way word networks such as the one formed from word ladders, the span is defined as the longest irreducible chain between any two words. In directed networks, the span is defined in two ways, the one-way span and the two-way span. The one-way span is defined in the same way as the span. For every pair of words

connected by a set of directed links, there is a minimum number of links achieving this goal (the irreducible span). List the irreducible one-way spans for all word pairs; the largest one of these is the one-way span. For the corpus of four-letter words from Webster's Pocket Dictionary, the one-way span is believed to be 10, achievable in several ways:

```
ST-EP-1C-ON-TO-OL-IO-TA-KA-YO-YO
ST-EP-1C-ED-GE-NE-V1-TA-KA-YO-YO
ST-EP-1C-ED-GY-VE-DA-TA-KA-YO-YO
1C-KY-AT-OP-AL-SO-DA-TA-KA-YO-YO
```

The two-way span is defined for words in the core. For any pair of words in the core, there is a minimum number of links joining one word to the second and return to the first (the irreducible two-way span). List the irreducible two-way spans for every pair of words in the corpus; the largest one of these is the two-way span of the network. For the corpus of four-letter words from Webster's Pocket Dictionary, the two-way span is believed to be 13 from OUZO to KAYO and return:

```
-OU-ZO-NE-AR-1D-OL-1O-TA-KA-YO-KE-P1-TH-
 |                                        |
 |_____|
```

One can define various tours of the core. What is the maximum number of frags with no repeats that can be strung together to form a word chain? What if the chain must be closed, as in the two-way span? What if frags can be repeated, but not words? Or if this chain is closed? Optimum solutions appear to be extraordinarily difficult' to find, even if a computer is programmed to do so; the search rules for the optimum appear to be complex and subtle. Here is a hand-generated solution (perhaps close to the maximum?) for the first problem, using 112 of the 133 core frags. Note that it is mathematically impossible to find a chain using all core frags:

* KH must be preceded and followed by AN
* UP and ZI can only be followed by ON
* OW, EK, EY can only be followed by ED
* YA and G1 can only be preceded by MA
* KI and R1 can only be preceded by SA
* JA and N1 can only be preceded by RA
* GH can only be preceded by N1
* GH and KY can only be followed by AT
* LY and PY can only be followed by RE
* EY and AW can only be preceded by TH

```
KI-WI-SP-UR-DU-MA-GI-BE-AU-TO-OK-RA-JA-1L-EA-ST-YE-T1-DY-KE-
P1-NE-VI-SA-R1-SK-ID-OL-1O-TA-KA-YO-GA-ME-NU-TS-AR-AB-ET-CH-
AP-EX-AM-MO-NO-UN-DO-PA-TH-OU-ZO-OM-1T-EM-1R-1S-LE-AF-RO-DE-
LI-MY-NA-Z1-ON-CE-DI-VA-SE-EK-ED-GY-VE-EP-1C-KY-AT-OP-US-ER-
GO-OD-OR-AL-SO-LO-CI-TY-PE-AS-HY-PO-CO-HO-AX-ES-PY-RE-AD-ZE-
BU-BO-LA-CY-AN-TE-AK-1N-CA-SH-AG-OG
```

In a directed word chain network, each frag has i frags that can reach it (its reverse fan) and j frags that it can reach (its

fan); some frags can be simultaneously in both. All frags in the core have the same fan and the same reverse fan. The fan consists of all other core frags, the followers, and the enders, whereas the reverse fan consists of all other core frags, the preceders, and the beginners. For the (4,2) word chain network, the fan size of a core frag is 132+6+88 = 226, and the reverse fan size is 132+62 +9 = 203. Preceders and beginners have slightly larger fans than core frags do, but very small or non-existent reverse fans; in contrast, followers and enders have slightly larger reverse fans than core frags do, but very small or non-existent fans. Isolates ahve very small fans and reverse fans. In the Pocket Webster corpus, all preceders have fans of 227 and reverse fans of 1; all beginners have fans of 227 or 228, and no reverse fan. Similarly, all followers have reverse fans of 204 and fans of 1; all enders have reverse fans of 204 or 205 and no fan.

Similar definitions can be made for the fan and reverse fan of ancestor and descendant words associated with any word in the network. For example, for any core word of four letters in the Pocket Webster corpus, the fan is 1374 and the reverse fan is 1252 in size. The word with the largest fan in TUBA, with 1406; the word with the largest reverse fan is ENVY, with 1268. At the other extreme, PHIZ, QUIZ, ZANY, BLOC, FINN, OXEN, RUBY, BABY, JOHN, WHIZ and UFOS are words which have neither a fan nor a reverse fan.

As mentioned previously, the (4,2) word chain network is an exceedingly complex tangle of directed link joining frags. Can one abstract from it a subset of frags and links that form a regular pattern? To show what is possible, here is a rectangular 5x4 mesh of 20 different frags joined by 31 directed links, 16 left to right and 15 up to down. Can larger ones be found?

```
HA SP UN DO PA
LO IN TO NE ST
CO CA RE AR AB
MA SH AM ID LY
```

Various substitutions are possible, especially at the corners: VE,LE, TE for PA; LE for LY; DA, LA for MA. As a variant exercise, one can ask for the largest free-form mesh that can be constructed.