WORD TREES IN RUNNING TEXT

A. ROSS ECKLER
Morristown, New Jersey

In "Mysterious Precognitions" in the August 1998 Word Ways, Martin Gardner showed how to generate a sequence of words in running text. Start with any word, and use the following (iterative) rule: if the ith word in the sequence has n(i) letters, then the (i+1)st word is located n(i) words further in the running text. For example, take the first sentence of this paragraph: pick IN as the first word, count two words to PRECOGNITIONS, count thirteen more words to A, count one more word to SEQUENCE, and so on.

Using this technique, Gardner showed that if any of the ten words of the first verse of Genesis (King James Bible) is selected, every sequence so generated includes GOD, the second word in the third verse. In other words, all the various sequences have by then converged, forming a single sequence thereafter.

One can visualize the collection of different sequences as a word tree inbedded in running text. The tree has a trunk sequence starting with the first word, which is joined by various branch sequences, each one headed by words unreachable by any earlier words in the text (call these words leaves).Most branch sequences join the trunk after a few steps. To aid the reader in visualizing the component parts of a tree, the text below capitalizes all words in the trunk sequence and underlines all leaves. Whenever a branch sequence joins the trunk sequence, the number of steps in the branch sequence is given (in parenthesis).

IN the BEGINNING God created the heaven and the earth. And THE(2,2) earth was WITHOUT(1) form, and void; and darkness(1) was UPON(1,4) the face of THE(1) deep. And(2) THE spirit of GOD moved(2) upon THE face of THE(1,13*) waters. And GOD...

For example, the (2,2) after THE indicates that this is the third word of the two sequences (the the THE), and (the created THE). The asterisk after 13 indicates that this is not a simple branch sequence joining the trunk, but instead one which has already amalgamated three branch sequences of its own: (form darkness), (was face and), (of deep and).

As can be seen from this example, there are usually a few earlier words in the text contained in sequences that have not yet joined the trunk (spirit, face, of, waters, and).However, it can happen that all earlier words converge to a single word in the text. This occurs if and only if the last few words before the convergence-word form a reverse rhopalic (snowball) or better. A reverse rhopalic is a sentence or phrase in which each word has one less letter than its predecessor, ending with a one-letter word. "Or better" is shorthand for the possibility that a word may be even shorter than its required length; for example, the fourth-to-last word in the rhopalic can be four or fewer letters long.

How likely is full-text convergence? Assuming that the length of a word in running text is not influenced by the length of adjacent words, one can use word-length statistics to calculate the chance of a one-letter word preceded by a one- or two-letter word, preceded by a one-, two- or three-letter word, etc. The probability of a word of n letters is tabulated on p 366 of Kucera and Francis's *Computational Analysis of Present-Day American English* (1967):

1	2	3	4	5	6	7	8	9	10	11	12
.03	.17	.21	.16	.11	.09	.08	.05	.04	.03	.02	.01

The probability that a word in running text will have all words converge to it is approximately 0.006, or an average of once every 1600 words.

One can check the reasonableness of this theore tical estimate by sampling text. The first three chapters of A. Conan Doyle's *A Study in Scarlet* contain about 11,000 words. There are three occurrences in the first chapter and two in the second:

- "You don't know Sherlock Holmes yet," he said; "perhaps you would not care for him as a [constant]..."
- "--an enthusiast in some branch of science. As far as I know he is a [decent]..."
- "The proportion of blood cannot be more than one in a [million]..."
- "...faculties of observation, and teaches one where to look and what to look for. By a [man's]..."
- 'Here is a gentleman of a medical type, but with the air of a [military]...'

This would suggest that the 1-in-1600 estimate is slightly optimistic, but the sample is small. There is a sixth example very early in Chapter 4.

The reader might want to use text-convergence to win bar bets. Tell your victim to select a text and a random word in a sentence therein, and to generate a sequence by the rule given earlier. You then tell him the first word of his sequence when it reaches a later sentence. Although there is no guarantee of convergence, the probability is high that you, selecting (say) the first word in his sentence, will arrive at the same target as he does. In the first four verses of Genesis, the tardiest convergence occurs if the victim picks the word SPIRIT in the second verse; the sequence that this heads doesn't join the trunk until the THE after "divided", 33 words later. So, to be reasonably safe, select a final word at least 40 words beyond the end of his sentence!