

HOW LIKELY ARE DUPLICATE ANAGRAMS?

A. ROSS ECKLER

Morristown, New Jersey

If different people attempt to form anagrams of the same phrase, how likely are they to come up with the same one? (By "same" I mean the same set of words, perhaps arranged in a different order.) Some light is shed on this question by a corpus of several thousand anagrams constructed in 1993 by Games Magazine readers in the "Mixed Reviews" contest (for examples, see the Word Ways issues of November 1994, February 1995 and May 1995). Certain movie titles were anagrammed by as many as 30 different contestants.

Obviously, the chance of duplication depends upon the number of letters, m , in the title and the number of anagrams, n , that are constructed. Each movie can be plotted as a point on a two-dimensional scatter diagram (m, n), using x to denote one or more duplications and o , none. A curve can be roughly fitted to the (fuzzy) boundary between the x 's and the o 's, representing those values of (m, n) for which there is a 50-50 chance of a duplicated anagram. If $i = 0, 1, 2, \dots$, then the boundary is mathematically expressed by $n = 3(2^i)$ and $m = 9 + 2i$. For example, a phrase of 9 letters with 3 anagrams, or a phrase of 15 letter with 24 anagrams, is likely to have a duplicate anagram. One should beware of using this formula to extrapolate beyond the range of cases actually encountered in "Mixed Reviews"; it is not clear that one can predict that the likelihood is 50-50 that a duplicate pair will arise from 1536 anagrams of a 27-letter phrase.

Four movie titles form a boundary for the x 's--that is, for each of these titles, it is impossible to find another title with duplicate anagrams which simultaneously has at least as many letters and at least as few total anagrams. These titles are:

COMING TO AMERICA (15) 3 anagrams "I am no great comic" (2 people)
 THE GRAPES OF WRATH (16) 5 anagrams "Gasp! Worth the fare" (2 people)
 YOUNG FRANKENSTEIN (17) 11 anagrams "Gene funny? O, it's rank" (2 people) or "OK in art? Gene's funny" (2 people)
 RAIDERS OF THE LOST ARK (19) 26 anagrams "Ford, the real star, is OK" (4 people)

Movie titles such as these, falling on the edge of the scatter plot containing the x 's, generate duplicate anagrams because of the high chance that certain probable words are selected by different constructors. In Young Frankenstein, the word FUNNY appears nine times, accompanied seven times by GENE (or GENE'S). In addition to the two matching anagrams, there were three near-misses: "Gene--rank it so funny", "I rant, funny (Gene's OK)" and "It ran OK--Gene's funny". In Raiders of the Lost Ark, the name FORD (or FORD'S) occurred in 22 of

the 26 anagrams, reducing the problem to 22 anagrams of a fifteen-letter title, which should yield a duplicate anagram about half the time. (Remember, the 50-50 predictor passes through $m, n = 15, 24$.)

Other movies produce probable words as well, but not as dramatically as FUNNY or FORD('S). For instance, Fried Green Tomatoes, an eighteen-letter title with 32 anagrams produced no duplicates because its probable words were much less concentrated: FRIENDS in 5, GREAT in 10 (and GREAT FRIENDS in 3), SEE TO in 7 (and GREAT SEE TO in 3). The Silence of the Lambs is an interesting case. This twenty-letter title had no duplicates among its 26 anagrams, but the phrase THE BEST FILM was formed 7 times, leaving only 9 letters to anagram. A movie title with 9 letters and 7 anagrams is definitely a candidate for a duplication, yet there were none: the letters anagrammed to ECHOS LENA (a misspelling), NO CHELSEA, CHOSEN LEA, CAN HE LOSE, ELAN CHOSE, ON SCALE EH and LONE CHASE. Had there been a duplicate anagram, this film, one letter longer than Raiders of the Lost Ark, would have replaced it in the list above.

With respect to The Silence of the Lambs, there were 25 additional anagrams which incorrectly assumed the title to be Silence of the Lambs. For many of these, it is plausible to add a THE to the anagrammed phrase, and in fact two of the additions to BEST FILM were CHOSE ELAN and CAN HE LOSE. Thus, a case can be made for adding to the above list a film with 20 letters and 35-40 anagrams.