motivated by the same old-fashioned determinism: Rojas, as a converso, was bound to do *thus*; Dunn, being British, must surely think *so*. As a matter of fact, I have learned a lot from the late Américo Castro; I also remember Wagner's warning: "Don't imitate anyone, least of all, me."

Gilman's letter may be a lesson in karate, but it's scarcely cricket. More to the point, it isn't critic, either. This has been a disagreeable exercise, and by your leave I'll hurry back to the much more delicious pleasures of torturin' me masterpieces.

PETER N. DUNN
University of Rochester

Frequency of Progressive Forms in Poetry in English

To the Editor:

G. T. Wright has advanced a compelling argument for the importance of present-tense verbs in English lyric poems of the past four centuries in "The Lyric Present: Simple Present Verbs in English Poems" (*PMLA*, 89, 1974, 563–79). In the course of his argument the issue of changes in present and past progressive verb tense frequencies is raised, and an appendix of frequencies is given, along with a narrative account of the changes. Wright finds the increase over the centuries to be surprisingly irregular, and cites a "strange restraint" between Jonson (b. 1574) and Thomson (b. 1700), a timid reappearance at the end of the eighteenth century, and a "general swing to the progressive" since the Romantic movement. Statistical analysis of Wright's data, however, suggests that the frequency of these verb forms changes quite uniformly over time. Further, this kind of uniformity seems more likely to reflect changes in the English language, rather than "complex psychological and philosophical considerations" by poets (p. 577).

The statistical analysis was as follows. For the sake of uniformity, I changed the data to a ratio of progressive forms per 1,000 lines of verse. I graphed these ratios against time, and two trends emerged: The frequency clearly increased, and, as it increased, the range of values got wider. The general pattern of increase appeared to be exponential, with "Gaussian" dispersion. A standard procedure to describe data of this sort computes a curve which goes down the "weighted" middle of all points, while balancing the extremes at each point. The weighting includes the number of lines in each sample. The data points and the curve are shown on the graph (Figure 1).

The mathematical description of the curve is: $\ln y = -11.5 + 0.0074t$, where y is the estimated ratio of progressive forms (per thousand lines) at a given date (t), expressed as a natural logarithm. The first number (-11.5) is a theoretical starting point at the "0th" year. The second number ($0.0074t$) is the annual rate of increase, which might be compared to $\frac{3}{4}\%$ yearly simple (not compound) interest. If one "invested" in progressive verbs in 1500, the "return" by 1900 would be forty-fold.

The next step is to estimate how well the curve describes both the trend and the range of the data. Using analysis of variance the "goodness of fit" (R^2) is 0.875, a value which is statistically very significant. This accuracy suggests that the model deserves to be tested in other situations, and that it can be used for interpolation and careful extrapolation. The difference from the curve for a given author is accounted for by the statistical model.

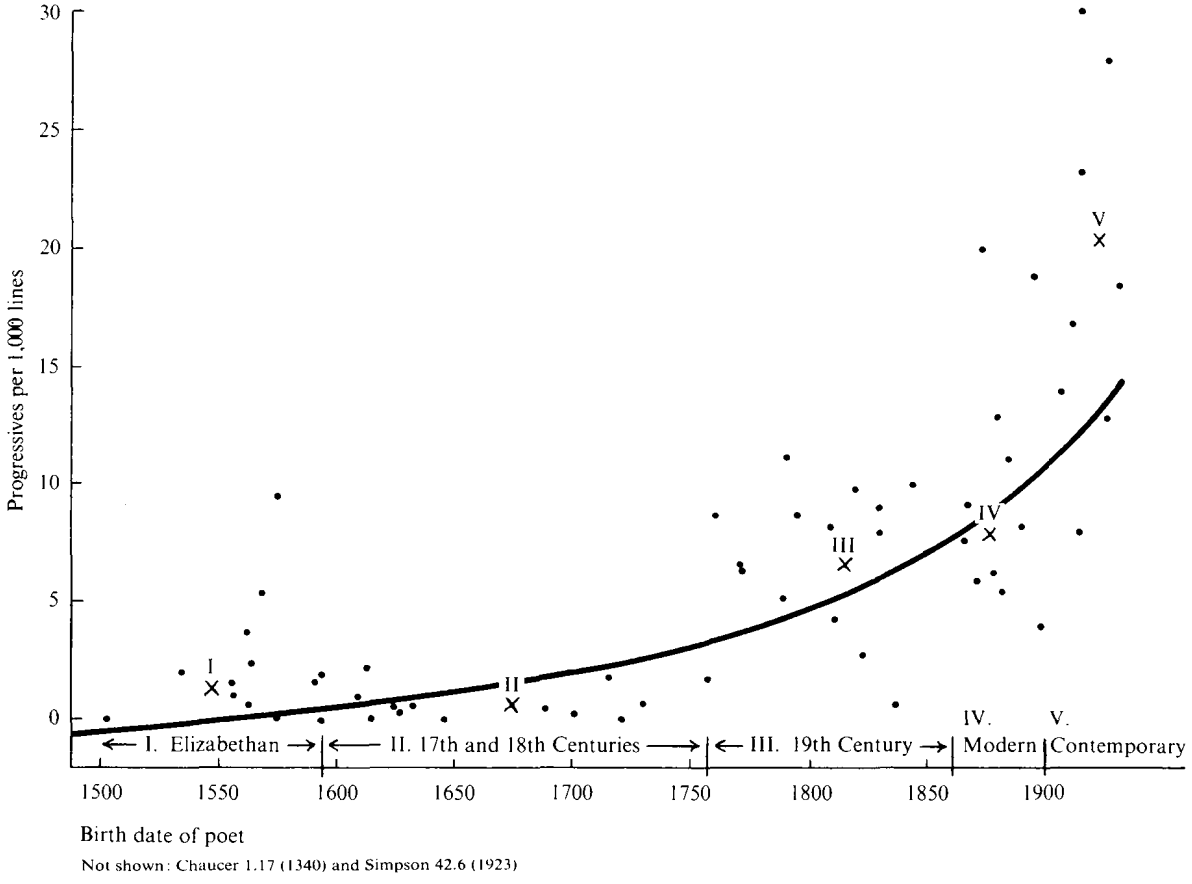
Although Wright looked at 326,000 lines of verse in his search for progressive verbs, perhaps that sample isn't big enough. The survey includes some odd-looking numbers: 13,197 lines from Milton, 553 from Sackville, exactly 1,000 for eight poets, exactly 10,000 for five. Do the numbers accurately represent the authors? Should more have been sampled? Or could the counting have been more efficient? Josephine Miles in several studies used the first thousand lines from either a collected edition or long poem written by the many poets whom she surveys; is hers a better tactic?

The usual statistical procedure would be to survey lines not in sequence, but to pick them at random. A table of "random numbers" would be used to select lines, and a different section of the table would be used for each poet. The number of lines could vary with the size of the poet's corpus, or a fixed number of lines could be picked for every poet.

The sample-size problem, put briefly, is that the smaller the relative frequency (the percentage) of the item one looks for, the larger the sample needed. One version is a formula which sets a "relative error" (δ) at 10%, and a "Gaussian deviate" (u) for a specified "significance level" ($\alpha = 0.05$) which is 2.0[1] The sample size (N) becomes: $N = (u/\delta)^2 \cdot p_i = 400/p_i$, where p_i is the proportion. For something that happens 3% of the time, 13,333 items need to be counted. The largest value among Wright's data is 43 progressives per thousand lines (Simpson); the average value for all the poets is 5.1 per thousand lines, which would ideally be based on samples of about 80,000 lines. One solution would have been the horror of looking for progressives among $4\frac{3}{4}$ million lines of verse. A more rational approach would be to compute progressives as proportions of a less frequent event, such as all verb tenses. This numerical strategy, coupled with a random selection of lines, would have given a more representative sample.

Further study would be needed to determine (1) whether the minor poets, not included in Wright's sample, would alter the distribution, (2) whether progressive forms have shown the same changes in prose or nonlyric poetry, (3) whether progressive forms have

Figure 1. Progressive tenses as a function of time: observed values and estimating curve



Not shown: Chaucer 1.17 (1340) and Simpson 42.6 (1923)

"replaced" other tenses, and (4) whether progressive forms correlate with changes in the overall frequency of verbs per line, say as part of a growing tendency for lines of poetry to be one grammatical clause long.[2]

DONALD ROSS
University of Minnesota

Notes

[1] R. M. Frumkina, "The Application of Statistical Methods in Linguistic Research," in O. S. Ashmanova et al., *Exact Methods in Linguistics Research*, trans. David G. Hays and Dolores V. Mohr (Berkeley: Univ. of California Press, 1963), pp. 89–96.

[2] The statistical analysis was done with the aid of Donald Berry and Gordon Mikkelson of the Univ. of Minnesota's School of Statistics. The comments here pertain only to the data in Wright's appendix and have nothing to do with the validity or persuasiveness of the article itself.

Mr. Wright replies:

It would appear to be incumbent on Donald Ross to explain why he chose this curve as appropriate to "data of this sort." He had evidently noted that "the frequency [of progressive forms] clearly increased, and, as it increased, the range of values got wider." But, although this is true for the whole period, it does not hold for the period of 150 years in which, as I claim, the poetic use of progressives declined. Inevitably, a curve chosen to exhibit long-range tendencies will exhibit them; if it does not take account of short-term anomalies, that is not necessarily because the anomalies do not exist but rather because the curve is not the right sort of curve to bring them out. The Ross curve and my table are designed to serve different purposes, the table to acknowledge divergences, the curve to suppress them. The curve is designed to show the overall uniformity of increase in progressive forms; the table and text call attention to a century-and-a-half-long lapse in the continuity of increase. While the Ross curve flows over the centuries undismayed by erratic poetic practices, the table shows the remarkable reluctance on the part of poets born between 1594 and 1757 to use progressives as much as we might expect them to, as much even as their Renaissance predecessors had done.

Still, as I read the Ross graph, it bears out my findings. Almost all the Renaissance dots lie above the