LINE INITIALS IN THE "GEORGICS"

The late Gustav HERDAN published a statistical analysis of the feature to be discussed here in 1962 (1). We have appropriated without change the data supplied by HERDAN as well as the basic mode of reasoning whereby expected values may be established. Our only justification for what would otherwise be an unnecessary replication is that Herdan's treatment was vitiated by an elementary but serious error which brought him to conclusions exactly opposed to what they ought to have been. The unwary and unnumbered classicist will be led astray. In addition, we have added the use of chi-square to measure the significance of differences between actual and expected values.

The feature examined is the interval between recurrences of the alphabetic letter or phoneme which begins a line in the *Georgics*. Thus, if two successive lines begin with the same letter, that is termed a gap or interval of length *0* (XX in Herdan's notation). If a single line with a different initial intervenes, that is a gap of length 1 (X1X), and we may similarly find gaps of length 2, 3, etc. (X2X, X3X, ...). Herdan has supplied the data for gaps up to length 8 (X8X) (2).

The formulation for arriving at expected values in the chance distribution is elegantly presented by Herdan (p. 82). Given the probability p for the occurrence of a particular initial, the probability for its recurrence with a gap of length r is $p(1-p)^r$. For example, 268 out of a total of 2188 lines begin with "a", i.e., 12.25 %, and in this case p = .1225 (3).

Extrait de la Revue (R.E.L.O.)
VIII, 1 à 4, 1972. C.I.P.L. - Université de Liège -

We arrive at an expected value for the number of recurrences of initial "a" with gap of length 0 by the calculation 268 x 0.1225 = 32.83. The actual number of such recurrences is 32. We present the actual and expected values for initial "a" in tabular form :

Gap-length	*Actual value*	*Expected value*
0	32	32.83
1	32	28.81
2	20	25.28
3	19	22.18
4	20	19.47
5	16	17.08
6	16	14.99
7	13	13.15
8	12	11.54

The fit is very good and it requires no great statistical expertise to conclude that Herdan's formulation constitutes a sufficient explanation for the phenomenon observed. It is reasonably safe to conclude that we have here an instance of the chance distribution. We have applied Herdan's formulation to every initial with p greater than .05, and in no case does it seem necessary to abandon the null hypothesis that we deal here with chance distributions (4). Herdan's conclusion to the contrary is based upon an elementary error (5) and should therefore be ignored.

OBERLIN College

Nathan A. GREENBERG

FOOTNOTES

(1) G. HERDAN, *The Calculus of Linguistic Observations* (Mouton, 's-Gravenhage, 1962) pp. 79-85.

(2) G. HERDAN, *Op. cit.*, Table 13a, p. 81.

(3) G. HERDAN, *Op. cit.*, Table 14, p. 83.

(4) None of the 100 chi-square values generated by our computer program extended to the percentile level of significance which we consider reliable for this kind of work, i.e., the level of 99.95 or greater. The following values seem worthy of report in that they are the only ones to extend beyond the modest level of 95.

Letter	*Gap-length*	*Level of significance*
C	4	95
Q	4	95
"	6	99.5
"	0 to 8 combined	99
CQ combined	4	95
"	6	99.5
"	0 to 8 combined	99
S	2	97.5

These values do not, in my opinion, constitute sufficient grounds for rejection of the null hypothesis that these are all chance distributions.

(5) The error consists of comparing incommensurables in Figure 1, p. 84. The percentages recorded in Table 14a, p. 83, are based upon a consideration of gap-lengths 0 through 8, and the existence of larger gap-lengths has been ignored. In Figure 1, the curves for p = .10 and p = .05 have been plotted without regard for this constraint. The proper plottings for these curves, given this constraint, are :

Gap-length	*p* =.*10*	*p* =.*05*
XX	16.3	13.5
X1X	14.7	12.8
X2X	13.2	12.2
X3X	11.9	11.6
X4X	10.7	11.0
X5X	9.6	10.5
X6X	8.7	9.9
X7X	7.8	9.4
X8X	7.0	9.0