

# **Some observations on the stylometry of the Pauline epistles**

by

**A. KENNY**

*Balliol College - Oxford - GREAT BRITAIN*

On a number of occasions A.Q. Morton and his associates have argued, on the basis of a study of sentence-lengths and the frequency and position of function words, that only four of the main epistles attributed to St Paul (namely Romans, I and II Corinthians, and Galatians) are genuine, and that the other epistles in the corpus are works of different hands. To the objection that a study of sentence-length must be based on punctuation which is the work of later editors, they have replied that the differences deriving from different editions are statistically insignificant.

The data on sentence-length which have been presented by Morton and his associates are susceptible of a quite different interpretation. Moreover, they themselves contain a refutation of the claim that editorial differences have only a negligible effect on sentence-length distribution. The distributions and constants given for the Pauline Epistles in Table 51 of *Paul, the Man and the Myth* (A.Q. Morton & James McLeman, 1966) differ from those given in Table 14.3 of *Literary Detection* (A.Q. Morton, 1978). The difference between the two tables, Mr Morton has kindly informed me, is due to the difference between the editions of the New Testament used in the two works. An examination of the constants in the two tables shows that there are significant differences in a number of places. Compare, for instance, the mean sentence-lengths for Romans and I and II Corinthians in Morton's two tables.

Epistle	Mean in PMM	Standard Error in PMM	Mean in LD	S.E. in LD
Romans	12.30	0.39	14.33	0.50
I Cor	10.8	0.29	12.23	0.34
II Cor	13.1	0.50	15.99	0.71

It will be seen that in each case the constants differ, between one edition and another, by more than twice the sum of the relevant standard errors. If we are to use a difference between the constants of sentence-length distributions as an indicator of different authorship, there seems a danger that we may have to attribute the Epistles of *Paul, the Man and the Myth* to an author different from the author of the Epistles of *Literary Detection*.

Figure 1 shows the mean sentence-lengths given in *Literary Detection*, with confidence limits corresponding to a margin of two standard errors.

The graph confirms the conclusion that the Pauline epistles do not constitute a population homogeneous in respect of sentence-length. But it does not support the contention that the four major epistles form a uniquely homogeneous group within that population. There is no way in which a single horizontal line could be drawn across the graph intersecting all four of the major Epistles. On the other hand, a number of other homogeneous groups could be formed from the corpus: one, for instance, would consist of Romans, 2 Corinthians, Galatians, Philippians, 1 and 2 Timothy. In particular, there is no reason to separate Philippians from the major Epistles. What is here illustrated with respect to the mean is true also of the other sentence-length constants given in Morton's table.

There is no need to pursue this, however, since Morton has helpfully shown in *Literary Detection* that the sentence length distributions of the Epistles can be fitted to a log normal distribution. Figure 2 shows the log means of the sentence-length distributions of the main epistles, with the 95% confidence limits, based on table 14.5 of *Literary Detection*. It will be seen that once again the four major Epistles do not form a homogeneous population, while other groups (e.g. 2 Corinthians, Philippians, Colossians, 1 and 2 Timothy) do. Morton's conclusion 'Romans, I and II Corinthians and Galatians form one group and the others are separate from it' does not seem to be justified by his own data.

Another hypothesis which has often been suggested to explain the variation between sentence-length in the different Epistles is that St Paul's preference in this matter underwent a change over time. It is difficult to test this hypothesis since the dating of the Epistles is a matter of controversy even among those scholars who regard them as genuine. Suppose, however, that for the sake of experiment we consider the hypothesis that the lengthier Epistles were all written by St Paul, and that they were written in the following order: 1 Corinthians, Galatians, Romans, 2 Corinthians, Philippians, 1 Thessalonians, Ephesians, Colossians. If we then place the log mean sentence-lengths on a graph in order, we find that a straight line can be drawn within the confidence limits for each Epistle. (fig. 3). This shows that even in this very crude form - which takes account only of a hypothetical order of writing, and not at all of the different lengths of time elapsing between one letter and the next - the hypothesis of a preference for longer sentences increasing over time is compatible with the data.

We obtain similar results if we turn from sentence-length to the frequency of common words in the Pauline corpus. Table 1 shows the number of occurrences, proportion (%) and standard error for the conjunction *kai* in the Epistles. Figure 4 shows that the data are consistent with the hypothesis of a gradually increasing preference for *kai* in accordance with the chronological ranking postulated for sentence-length. A similar graph for *de* would show gradual decrease.

There are a number of prepositions which are favoured by the Pauline corpus. Table two shows the frequency of fourteen common prepositions in Paul and in the rest of the New Testament. ('Paul' here means the Pauline corpus less Hebrews). It will be seen that there are a number of prepositions which are much commoner in the corpus than elsewhere in the N.T. The table gives the distinctiveness ratio for each preposition: i.e. the ratio of its relative frequency in Paul to its relative frequency elsewhere in the N.T. A preposition such as *kata*, with a D.R. of 2.31, is more than twice as popular in the Pauline corpus as elsewhere.

We may group together the Pauline favourite prepositions, and study the frequency of such a group in the several epistles. Table 3 gives the frequencies in the individual epistles of a group of Pauline favourites consisting of *en, dia* with the genitive, *kata* with the accusative, *upo, uper* and *sun*. (The frequencies for the prepositions and for *kai* are based on the word counts in Morgenthaler's *Statistik des Neutestamentlichen Wortschatzes*). It will be seen that the frequency of these favourite prepositions

is higher in the captivity epistles than in the four major epistles. Figure 5 shows how the frequencies relate to the chronological order postulated for sentence-length. It is not quite possible to draw a single line through each of the verticals marking confidence limits, but a very slight departure from linearity, or from the crude equidistant ranking of the epistles would enable a chronological explanation of the data in terms of increasing preference over time for those prepositions which, throughout his career, were favoured by St Paul.

Two concluding remarks are necessary. First, the Pastoral Epistles have been disregarded in the present study. They cannot be assigned a single place in a chronological series on the basis of the several independent phenomena reported here : which may be, added to other criteria, an indication of their inauthenticity. Secondly, I am not proposing, as a serious hypothesis, the chronological ranking used in figures 3 to 5 : it would be premature to do so without much further study both stylometric and historical. I am simply claiming that such a hypothesis would explain the data here discussed just as well as Morton's hypothesis of the inauthenticity of the major part of the Pauline corpus.

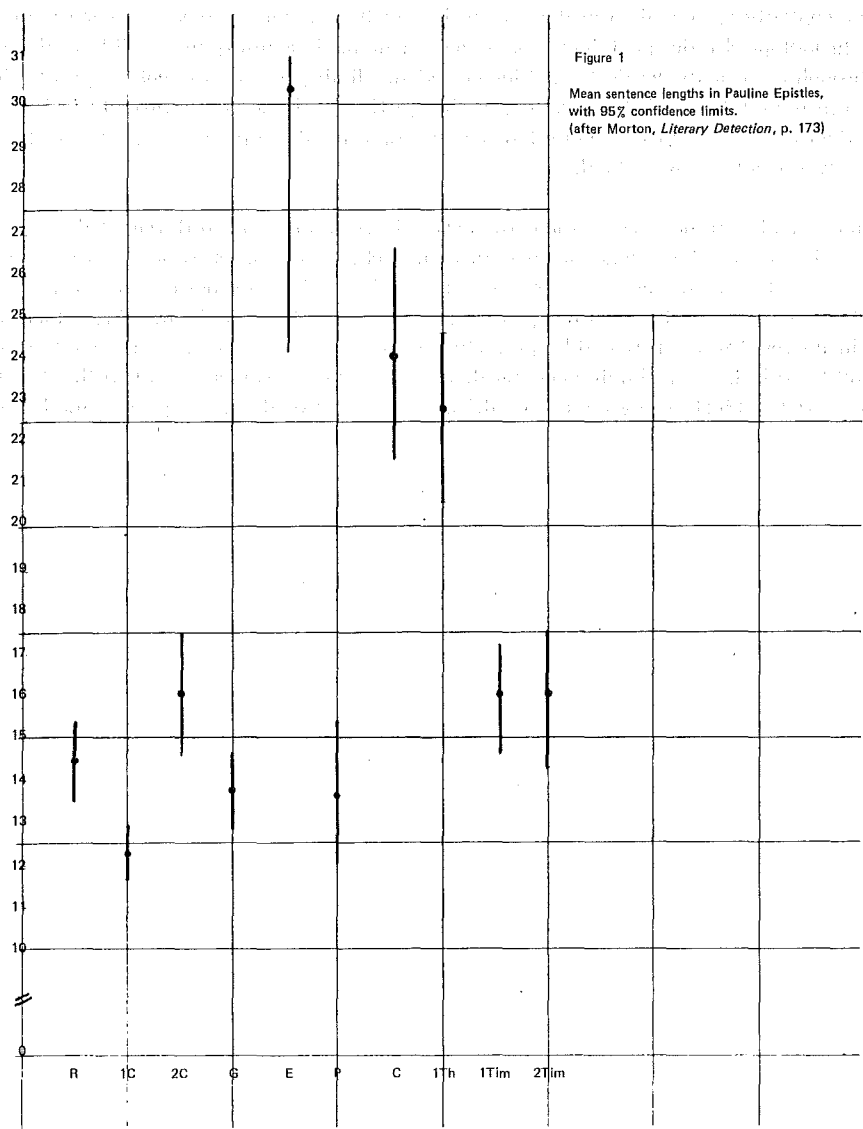
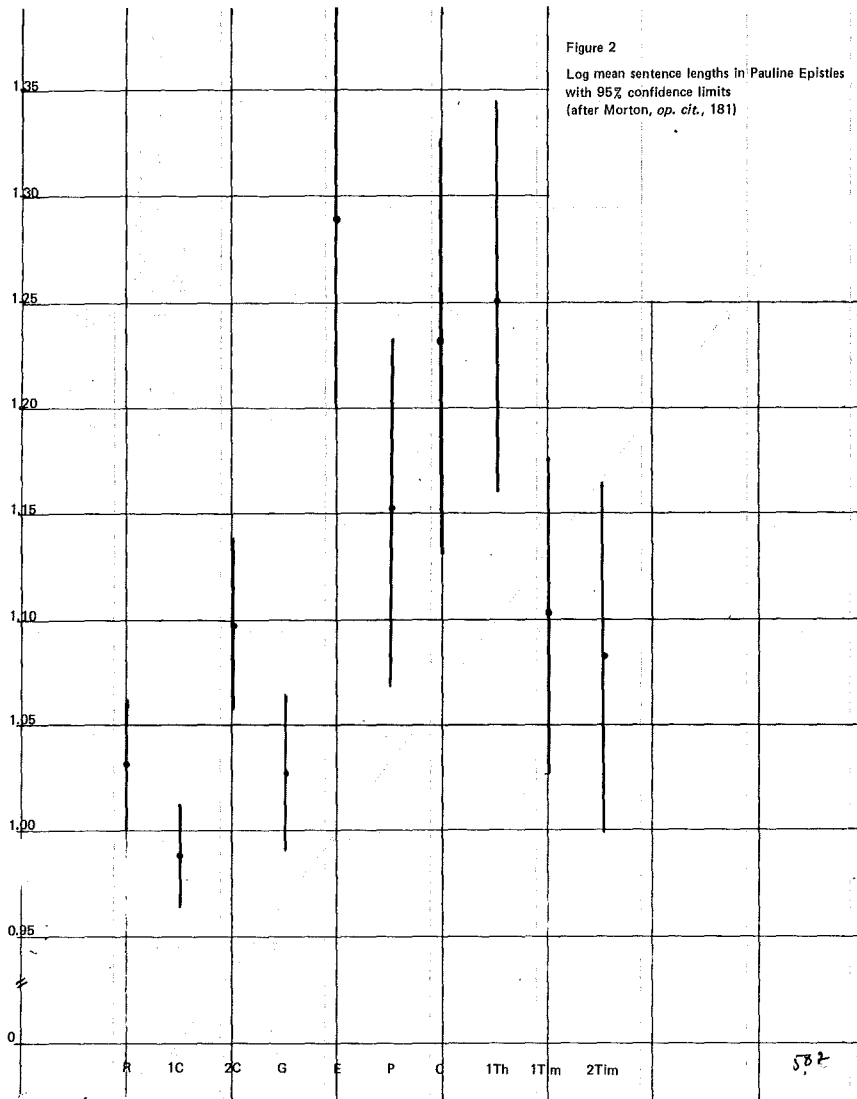


Figure 1  
 Mean sentence lengths in Pauline Epistles,  
 with 95% confidence limits.  
 (after Morton, *Literary Detection*, p. 173)



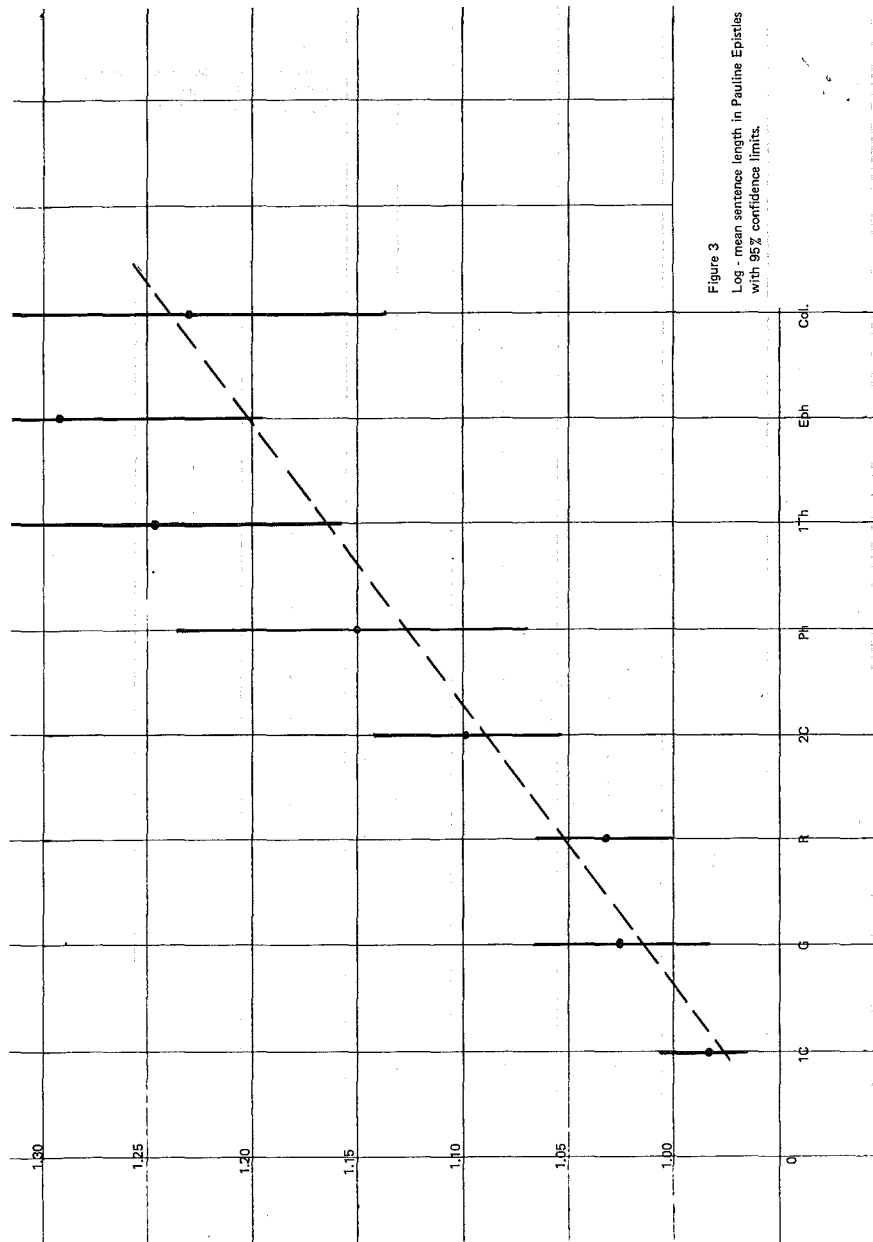


Figure 3  
Log - mean sentence length in Pauline Epistles  
with 95% confidence limits.

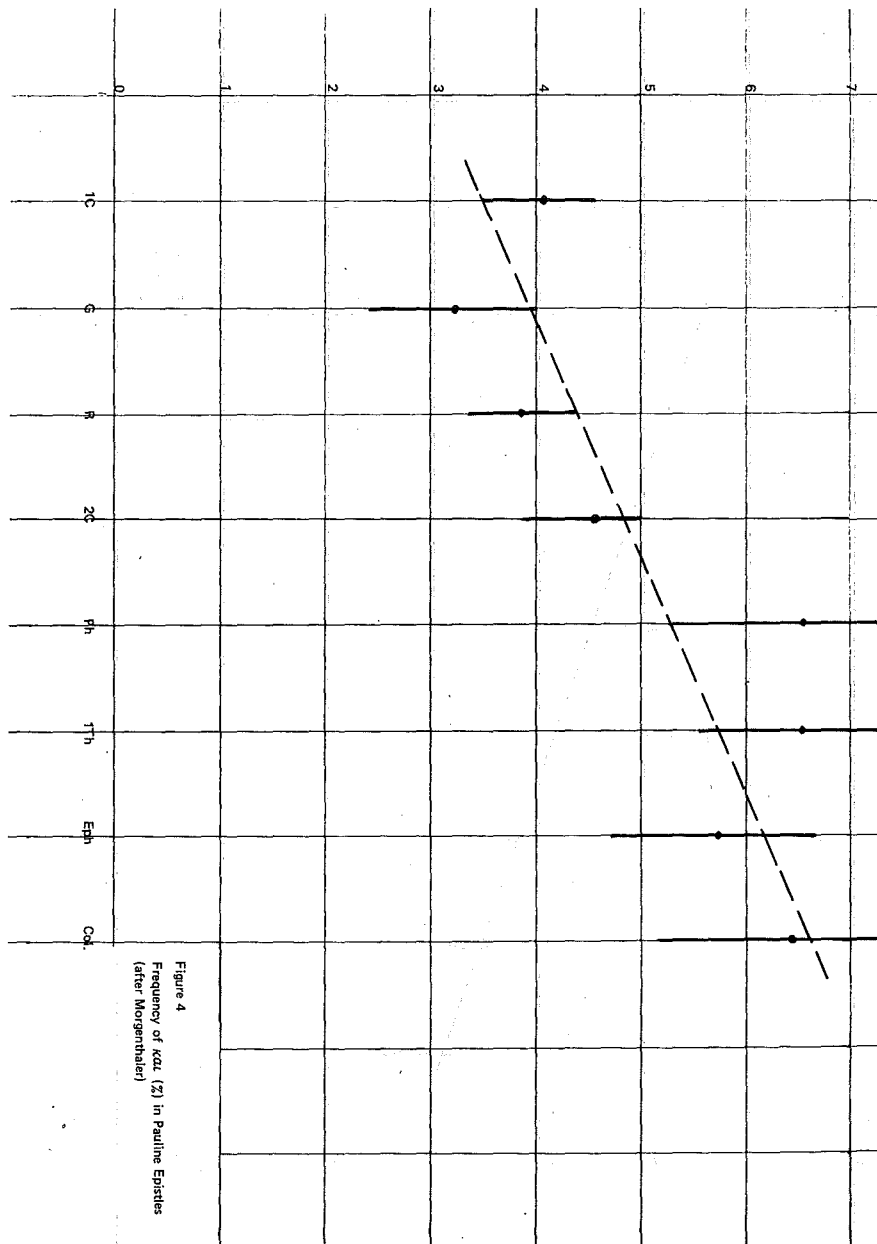


Figure 4  
Frequency of 'ka' (%) in Pauline Epistles  
(after Morgenstaler)



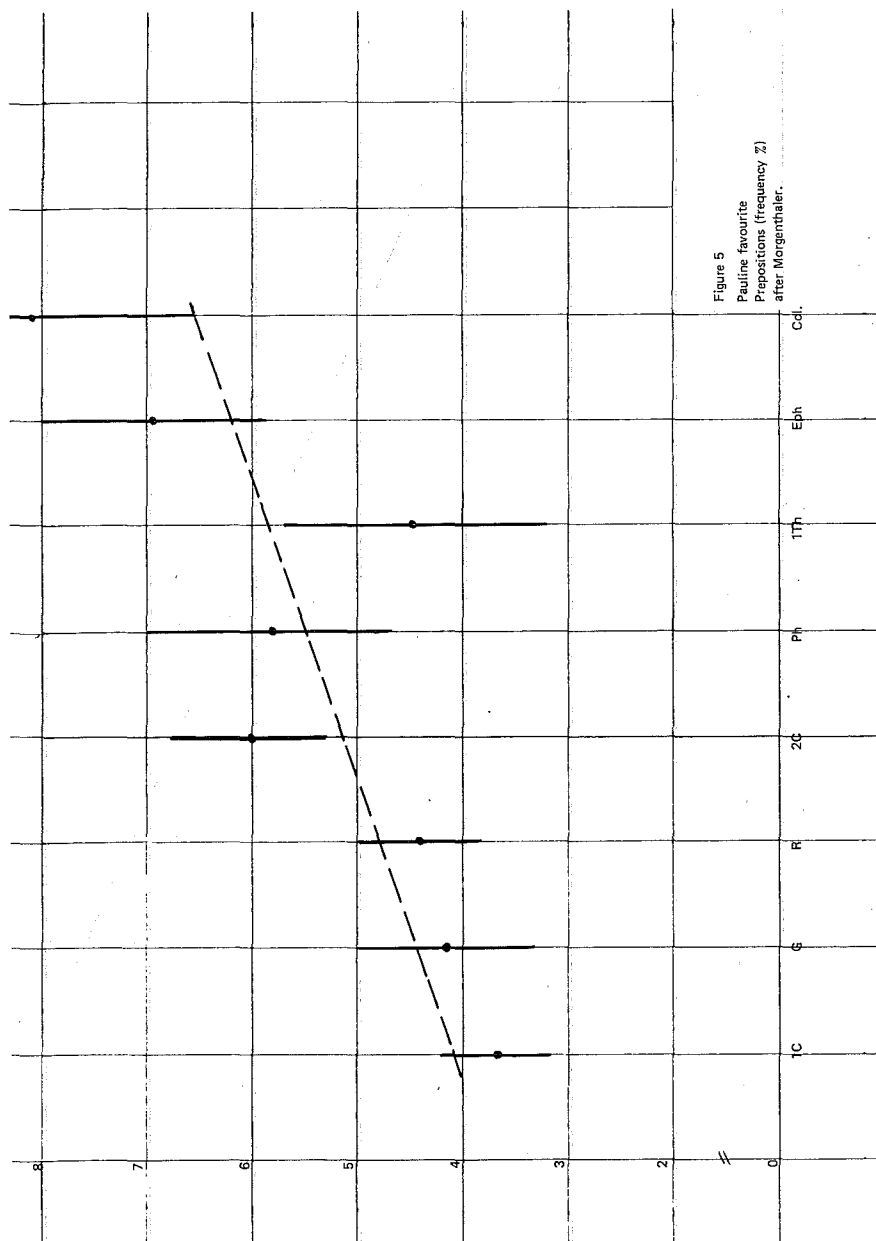


Figure 5  
Pauline favourite  
Prepositions (frequency %) after Morgenthaler.

TABLE 1

*kai* in the Pauline epistles

Epistle	Occurrences	Proportion	S.E.
Romans	274	3.86%	0.23%
1 Cor	277	4.07	0.24
2 Cor	197	4.43	0.30
Gal	72	3.24	0.38
Eph	137	5.65	0.48
Phil	107	6.59	0.64
Col	101	6.40	0.64
1 Th	101	6.86	0.68
2 Th	50	6.07	0.86
1 Tim	92	5.80	0.60
2 Tim	68	5.52	0.67
Tit	36	5.43	0.90

TABLE 2

## Fourteen Common Prepositions in Paul and the rest of the NT.

Prep	Paul (%)	Rest of NT (%)	DR
en	3.06	1.64	1.87
eis	1.32	1.26	1.05
ek	0.65	0.67	0.97
epi	0.41	0.71	0.58
pros	0.45	0.52	0.87
dia	0.90	0.36	2.50
apo	0.33	0.51	0.65
kata	0.60	0.26	2.31
meta	0.25	0.38	0.68
peri	0.16	0.27	0.59
upo	0.23	0.14	1.64
para	0.13	0.14	0.93
uper	0.31	0.05	6.20
sun	0.12	0.08	1.50
TOTAL	8.93	6.99	1.28

TABLE 3

Prepositions : Pauline plus-words

(en + diaG + kataA + upo + uper + sun)

Epistle	Occurrences	Proportion (%)	Standard Error
Romans	319	4.44	0.25
1 Cor	250	3.67	0.23
2 Cor	267	6.00	0.37
Gal	93	4.19	0.43
Eph	168	6.92	0.53
Phi	94	5.79	0.60
Col	122	8.05	0.71
1 The	65	4.42	0.55
2 The	43	5.22	0.80
1 Tim	56	3.54	0.47
2 Tim	53	4.29	0.59
Tit	24	3.62	0.74
Paul	1554	4.81	
Rest of NT	2312	2.20	

D.R. = 2.19