Type: Report

Review of Ernest Taves' Book of Mormon Stylometry

Author(s): John L. Hilton
Published: Provo, UT; Foundation for Ancient Research and Mormon Studies, 1986

Abstract: No abstract available.

FOUNDATION FOR
ANCIENT RESEARCH &
MORMON STUDIES

Preliminary Report

REVIEW OF ERNEST TAVES' BOOK OF MORMON STYLOMETRY

HIL-86

John L. Hilton

P.O. BOX 7113
UNIVERSITY STATION
PROVO, UTAH 84602

REVIEW OF ERNEST TAVES'

BOOK OF MORMON STYLOMETRY

by John L. Hilton

Trouble Enough: Joseph Smith and the Book of Mormon. By ERNEST H. TAVES. (Buffalo, New York, Prometheus Books, 1985. xi + 280 pp. $19.95).

[Editor's Introductory Note: The following letter and accompanying critique of the stylometric studies of Ernest Taves have been sent to F.A.R.M.S. by John L. Hilton. Hilton and his colleagues, who have been actively involved in stylometric analyses of the Book of Mormon for several years, plan in the near future to complete their own extensive and thorough stylometric study of Book of Mormon texts. The following general review of Taves' book serves to introduce John Hilton's more detailed remarks.

Most of Ernest Taves' Trouble Enough is not concerned directly with the Book of Mormon, but with the biography of Joseph Smith. Taves writes with an engaging but liberal journalistic style, acknowledging frequently the strong biases of the historical witnesses he quotes, yet invariably following those that fit best into his story line--one which attempts to unify Joseph Smith's story through psychiatry, along the lines of Fawn Brodie's No Man Knows My History (1945), upon which he heavily depends. In Taves' biography, the Book of Mormon figures as a small part of the psychological diagnosis: *inter alia*, Joseph had a fertile, inventive imagination, an inability to tell a story the same way twice, a poor sense of judgment, delusions of power and legal autonomy, a tendency to run away from problems, influential hypnotic powers, and a strong need for female companionship. For Taves, the writing of the Book of Mormon was a joke that got out of hand.

Taves' account is eclectic. For example, he uses Prince (p. 60) but ignores Theodore Schroeder's refutation of Prince in American Journal of Psychology 30 (1919), 66-72. In his historical chapters, Taves variously depicts the Book of Mormon as being written by Joseph Smith, from an unknown Spaulding manuscript, "worked over" by Oliver Cowdery, influenced by Sidney Rigdon, etc. (chs. 4-5). Inconsistently, his stylometric work (chs. 23-25) sets out to prove Joseph Smith the sole author.

Stylometry is the statistical comparison of certain identifiable word-habit frequencies in two texts to determine the probability of their common authorship. This science is still open to skepticism, for several reasons, e.g., writing behaviors may change substantially over time and from subject to subject. Significant results appear obtainable in many cases, but only when the tests are run very carefully. Many interesting projects are now underway.

Using his understanding of stylometry, Taves claims to find similarities between three sections of the Book of Mormon and the Book of Abraham. To the extent these similarities exist, this is perhaps not surprising, since both are works which Joseph Smith claims to have translated. While generously recognizing that his tests are not conclusive, Taves hopes that his stylometric work will be a "considerable advance on previous examinations" (p. 260).

Many problems and errors make it unlikely that this hope will be realized. For example, (1) he excludes the Genesis material from the Book of Abraham but apparently does not exclude the Isaiah material from the book of Mosiah (this can only be presumed, since Taves has not specified which texts from the middle of the book of Mosiah he selected). (2) He seems very willing to inconsistently label divergent phenomena as "anomalous." For example, places where the phrase "the wilderness" occurs at the end of sentences are deleted from the text sample (p. 248), since this happens to be a characteristic of 1 Nephi but not Mosiah. This seems arbitrary; of course, 1 Nephi talks about "the wilderness" (actually the phrase occurs about 171 times in varying frequencies in many parts of the Book of Mormon); and this deletion must to an extent skew the remaining data. (3) Many of his conclusions depend on averaging or homogenizing the differences between his three Book of Mormon sections. Similarities between the Book of Abraham and the average results for Taves' three Book of Mormon sections are noted, but differences among the three Book of Mormon sections are not. (4) Taves has ignored classificatory and purported differences of authorship within the texts. He treats all three sections as if they were all written in one style by one person. (5) Arithmetic errors are apparent, for example in Table 24.8 on p. 240. (6) One is disquieted by the high frequency of tests for which Taves finds that "the occurrences are too few for testing." (7) Many methodological questions remain unaddressed. One wonders, for instance, how one confidently goes about studying the frequency of words according to their preferred sentence positions based on sentence punctuation (p. 236) in a text that was unpunctuated by Joseph Smith and his scribes.

Aspects of these and other problems are discussed below by John Hilton. In addition, a bibliography of source materials on stylometric studies is available; interested readers may request a copy from F.A.R.M.S.

Undoubtedly Taves has had fun enough with his research. It has "pleased" him (p. xi), and his writing manifests this bemusement. But serious use of this kind of historical and statistical hotchpotch is severely limited, if not foreclosed.]

I. Letter of 30 April 1985.

To Dr. Ernest H. Taves, 12 Hubbard Park, Cambridge, MA 02138
From John L. Hilton, 40 Overlook Ct., Walnut Creek, CA 94596
Re Continuing study on stylometry from "Trouble Enough"

Dear Dr. Taves,

I am one of the principals of the "Berkeley Group," a loosely organized inter-disciplinary inter-religious group of scientists who for the last four and a half years have been actively studying computer assisted literary stylometry, finding it a fascinating avocation. Our interest was piqued by the publication of the provocative article published by Larson, Rencher and Layton, "Who Wrote the Book of Mormon," Brigham Young University Studies, vol. 20, no. 3, (Spring 1980). Their purportedly objective finds seemed to stretch our scientific credulity. Once into the study we became fascinated with the whole field and have since used most of our free time studying information theory and statistics, performing literature searches, researching and entering "original" texts, writing computer programs and taking measurements.

We have developed our own set of computer codes and proof texts to study independently and compare alternate stylometric techniques. We see convincing evidence to support at least the possibility of objective stylometry for the works of many writers. As of yet we believe the limits of reliability have not been objectively demonstrated, nor has a generally accepted technique been identified, notwithstanding the claims of the several groups who have been publishing.

It is from this background that I have been asked to review relevant chapters from your book Trouble Enough. I am a serious Mormon, although several of my colleagues in our group are not Mormons. I therefore assume that I am about as biased as you are but in the other direction. Fortunately, in objective stylometry we should both be competent in the use of the scientific method to identify the true analyzable propositions. Questions on stylometry, as Morton reminds us, are resolved by objective measurement, not by the preconceived opinion of the student. Therefore, those of us who attempt to use objective measurements have a scientific obligation to see that all measurements are honestly calculated and reported, and corrected as needed.

I was surprised to read the conclusion of your statistical work supposedly obtained through the use of the "new Morton" series of tests. Our measurements are in strong disagreement with those you present, even though our samples are drawn from the same material and are evaluated with what should be the same technique. Furthermore, I have observed from additional extensive analysis that the statistical distributions found in the texts of the Book of Mormon are significantly different from the noncontroversial writings of Joseph Smith (or Solomon Spaulding, or Sidney Rigdon, or Oliver Cowdery, etc.) Some of our measurements even suggest that multi-authored patterns still exist in the original English manuscript of the Book of Mormon despite the purported much abridging and translating.

As is almost universally true with pioneering scientific studies, later students take exception to at least part of the initial work. We also take exception to the work of Larson, Rencher and Layton (though not for the reason that your book specifies). Our years of evaluation, verification, and correction have identified many pitfalls that must be avoided before reliable statistical inferences can be drawn. Since the field is so new, it is not surprising that continuing cooperative communication between scholars is needed for the development of correct and generally accepted techniques.

Find enclosed my preliminary critique of your work as seen in the context of our studies. As you will read, at least for now, I believe your work to be completely invalid. I assume that you are interested in scientific accuracy, and we await your response to this critique, so we may as needed both correct our techniques toward objective consensus.

Thank you, sincerely yours,

John L. Hilton

II. Critique of Trouble Enough, chs. 23-25.

A. INTRODUCTION

For the popular audience, Taves' chapters 23-25 give a brief but clear explanation of the nature and development of computer-assisted literary stylometry. To write with such an interesting style and clarity as does Dr. Taves does is most admirable. However, there appears to be little information that is not copied consecutively from the work of the University of Edinburgh Group (see, e.g., A. Q. Morton "Literary Detection: How to Prove Authorship and Fraud in Literature and Documents"). And notwithstanding his admirable literary ability, Taves apparently does not correctly use Morton's author identification stylometric (wordprint) tests. Taves' attempted stylometric analysis of the Book of Mormon is at best superficial, which in this kind of work leads the nonspecialist reader to come to incorrect conclusions. Additionally, the calculations of his fundamentally important test statistic appear to be based on an incomplete (and thus invalid) formula.

I am surprised that anyone would attempt to use stylometry to defend the position of the nonbeliever in the Book of Mormon, since he has everything to lose and almost nothing to gain by such objective testing. To the "believer," even if a conclusive answer had been possible, it would be irrelevant to his faith. Whether the "believer's" multi-authored, much-abridged, much-translated English book should or should not show single- or multi-stylometric patterns is at present unknown. However, to the "nonbeliever" who thinks that he has some theory that "explains" the Book of Mormon as having been written by Joseph Smith (or Solomon Spaulding, Sidney Rigdon, or Oliver Cowdery), this testing, when done correctly, could objectively show that the book was not written by Smith or others (see, e.g., "Wordprint Examples Using the Mann-Whitney and Chi-Square Test Statistics" below). I would imagine that such objective answers to this type of "nonbeliever" would be disquieting. It is a credit to Taves' courage that he tries stylometry testing anyway.

Using three small text samples from the 1830 edition of the Book of Mormon and his variation of stylometry, Dr. Taves tries to measure Joseph Smith's own word-patterns and single authorship. I understand Dr. Taves' conclusion to be that (after he performed desensitized tests) he sees no evidence of multi-authorship above his level of sensitivity. This he contends is consistent with what would be expected if Joseph Smith had written the whole book. He does, however, view his own study as introductory; it is "not brought forward as conclusive" (p. 260).

Dr. Taves is wise in trying to follow closely Morton's stylometry technique, since Morton has probably written more than any researcher on stylometry tests. Morton is generous in sharing his developments with others and has demonstrated that within a carefully defined set of word-pattern tests and text selection rules, valid statistical inferences can be drawn from the writing of many authors.

It appears that as a service, Morton's group prepared the computer-tabulated "concordances" from the text samples chosen by Taves. A spot-comparison of his Nephi word pattern counting against our fully computer-tabulated files for this section of the Book of Mormon manuscript shows only the expected differences of a few counts for both word patterns and important sentence structure. Therefore, Morton's **counting** for Taves is apparently correct.

Without belaboring this critique, I will analyze what appears to me to be the two most glaring faults in Taves' attempt. I will then present results obtained with these errors removed.

B. PROBLEM 1: SELECTION OF TEXT SAMPLES FOR TESTING

A statistical test for author identification is meaningless unless the text samples that are studied at least claim to be the free-flow word patterns of their purported authors. Taves' text samples do not satisfy this requirement.

Taves selected his three Book of Mormon text samples of "approximately 5200 words in length; . . . the text from 1 Nephi was taken from the beginning, of Mosiah from the middle, and of Alma from the end" (p. 242). Taves thus did not select his samples from the writings of purported single authors, but from multi-authored sections of the overall volume. (See Table 1, below.) While he may study sections of the book if he chooses, it will only further desensitize his ability to identify individual authors who, if they exist, are then haphazardly averaged together within the sections of the volume he selected for study. This will increase the statistical uncertainty, thus lowering overall sensitivity.

Taves further selected the Book of Abraham from the Pearl of Great Price as a fourth text for comparison to the Book of Mormon samples on the assumption "that Abraham is the work of Joseph Smith" (p. 241). It is an enigma to me why, if he wished to measure Joseph Smith's personal stylometric writing habits, he did not choose to test any of the noncontroversial samples of Joseph Smith's writing or dictation, or other writings available. For the most recent scholarly compilation of Smith's works, including photostats of the original manuscripts, see Dean Jessee, The Personal Writings of Joseph Smith, Deseret Book, 1984.

Taves further acknowledges an indisputable point, that the last third of the Book of Abraham closely follows the wording of the King James Book of Genesis. But this further means that the Book of Abraham cannot be viewed as a single free-flow word pattern of any unique author, but rather a compilation of word patterns. Although Morton cautions against the use of text samples that "contain passages which may be a rewriting of another source" (Morton, p. 38), Taves attempts stylometric comparison there anyway!!

To test different texts for authorship, the ideal would be to have a battery of word pattern tests that uniquely measure changes in authorship and are completely insensitive to other influences. Morton has selected tests very carefully toward this ideal. He has chosen a battery of about 50 (statistically non-correlated) word pattern tests. Within each text sample of 5000 words, typically there can be tabulated more than the needed minimum of five occurrences for each of 30 to 45 of the patterns from Morton's battery of tests. This then permits "valid" null-hypothesis testing using the chi-square (or other test statistic) of this "valid" portion of Morton's full battery of word pattern tests. Notwithstanding the care taken in choosing his test battery, Morton was perhaps the first to show that some authors at times exhibit improper rejections of certain literary forms. He shows improper null-hypothesis rejections for some cases of dialogue and travelog descriptions. (Noncontroversial 5000-word text from Samuel Johnson's travelog and didactic writings discussed below also show an example where changing literary form [and possibly time] appears to show an incorrect number of null-hypothesis rejections.) This problem can be largely side stepped by limiting comparisons to texts of similar literary form or genre.

Dr. Taves seems to ignore these cautions and selects his Book of Mormon samples as if he were not adequately familiar with the book's composite structure. Each of his samples contains mixtures of various purported authors and literary forms. Table 1 shows the approximate text sample percentages of purported authors and literary forms for the sections of the Book of Mormon used in Taves' samples.

<u>Table 1</u>

Approximate Book of Mormon Samples selected by Dr. Taves, showing percentages of purported author and literary form of each 5200 word sample.

FIRST SAMPLE, Book of 1st Nephi [Heading - 7:15]		
author=Nephi	narrative 1st person	80.8%
author=Nephi	didactic	6.2%
author=Lord	didactic	5.0%
author=Lehi	didactic	6.5%
author=Lehi	dialogue	1.5%
SECOND SAMPLE, Book of Mosiah [chapters not reported by Taves]		
author=Isaiah	didactic (close to KJV)	unknown%
author=Abinadi	didactic	unknown%
author=Mormon	narrative 3rd person	unknown%

Table 1 (Continued)

THIRD SAMPLE, Book of Alma 58:28-63:16		
author=Mormon	narrative 3rd person	50.9%
author=Gen.Moroni	didactic	27.2%
author=Pahoran	didactic	13.2%
author=Helaman	narrative 1st person	8.7%

C. PROBLEM 2: CALCULATIONAL DIFFICULTIES

Dr. Taves tries to follow Morton, using exclusively the "chi-square" calculational technique to measure statistical significance. But he does not correctly follow Morton in calculating this important statistic. Taves correctly explains in his Table 23.1 how the binomial coin flip calculation should be made. He takes into account both the "heads" and "tails" of the hypothetical measurement. In his Table 24.4 he tries to apply the technique to the real null-hypothesis testing of different samples purportedly written by different authors. He apparently omits one-half of the required terms of the equation. It seems that he is accounting only for the number of the "heads" and ignoring the "tails" portion of the calculation. When the full equation is used, the value of the chi-square for his example in Table 24.4 should be 0.48, not the 0.37 he alleged. This fundamental mathematical error in his formula apparently was continued throughout his work. For the correct calculation using the chi-square method, see the step-by-step instructions of Morton (or any standard statistical text, e.g., Snedecor and Cochran, Statistical Methods, 7th ed., pp. 120-127). This calculational error makes Taves' chi-square values too small. This further improperly desensitized his answers.

D. RESULTS BASED ON MORE LEGITIMATE METHODOLOGY: WORDPRINT EXAMPLES USING THE MANN-WHITNEY AND CHI-SQUARE TEST STATISTIC

The following analyses are offered in contrast to Taves' studies.

Unlike taking "fingerprints," the present state of the art of taking correct "wordprints" (i.e., a measurement of the unique stylometric writing habits of an author) is not a trivial operation. The technique, if it can be made objective and accurate, as a minimum requires the investigator to complete each step and to justify independently that the assumptions used in the mathematical model adequately match his case under study. With proper assumptions, the objectivity of Morton's technique has been shown for specific cases. A general proof remains yet to be developed. Written English and many of its word patterns are at times neither random enough to assure word block averaged homogeneity (ergodicity), nor to assure an adequate normal distribution for the numbers of events often tabulated in stylometry measurements. The hoped-for identification of an ideal set of word pattern tests that are always completely insensitive to changes of literary form and subject matter has been statistically approached only for modest-sized samples and has not proven in the general case. The assumption of textual ergodicity and statistical normalcy required for standard chi-square testing has also been shown for at least some stylometry cases to be invalid (Morton, p. 140). Ad hoc or subjective spot corrections (as tried by Taves) at best can correct only the most obvious occurrences of the above mentioned problems. Any ad hoc or spot-correcting causes many statisticians to question the objectivity and therefore the value of the overall measurement.

To avoid at least a significant part of these problems, Taves should examine nonparametric statistical models, such as the Mann-Whitney Rank-Sum test. These models do not require any of the normality assumptions intrinsic in the simple chi-square method. Short of future developments of the science, it now appears necessary to verify each model with inter-and intra-control author measurements. The verification control or baseline must be matched to the texts under study in literary style, sample, size, etc.

As discussed above, for rigorous work one must verify the appropriateness of the match of the specific statistical model to the text or author being studied. This is done by first measuring the author's noncontroversial works for the differences within themselves and to each other. One straightforward way to measure the "author stability" is simply to count the number of simultaneous null-hypothesis rejections (at a fixed probability, e.g. $p<.05$) that are obtained when applying Morton's standardized battery of tests to a set of known noncontroversial texts.

We demonstrate this technique by first examining two known authors who come from opposite ends of literary attainment. First is the highly literate Samuel Johnson, author of the first major English dictionary, who seems to write with a working vocabulary of five to eight times that of the second author, Joseph Smith, who was only marginally trained. We sample Samuel Johnson from his didactic newspaper serial, Rambler (1750-51), and his travelog, A Journey to the Western Islands of Scotland (1775). Each work is sampled by ten consecutive 1000-word groups. This permits the dividing of each of his different literary works into two sample texts of 5000 words each for intra-work (and inter-literary form) testing. Selections were made from Joseph Smith's dictated texts in an effort to find noncontroversial samples (see Dean Jessee, *op. cit.*) This yielded ten consecutive 1000-word groups from Joseph Smith's first-person narrative dictated diary (1835-36), five groups from his published newspaper didactic essays (of the same time period), and five groups from Smith's "official history" (1838).

We test for statistical significance by using two different test statistics, the chi-square which Taves attempted to use at a probability of $p<.05$ and the more appropriate (statistically robust) Mann-Whitney, a rank-sum technique at the somewhat lower probability of $p<.0318$. All comparisons we show in this critique are made between texts of 5000 words each. Larger samples would likely show greater statistical separation between writings of truly different authors. For the Book of Mormon, where much larger samples are available, larger statistical separation for authors is obtained by using larger text samples.

Table 2 illustrates what might be expected for unknown authors by showing the number of significant null-hypothesis rejections from the comparisons between two noncontroversial samples from the control author's works.

Table 2

Tabulation of significant null-hypothesis rejections from comparisons between the works of a single known author.
Part A shows comparisons of the same literary form.
Part B shows comparisons of different literary forms.

		Number of significant null-hypothesis rejections from battery of Morton's tests	
Author	Text samples compared	Mann-Whitney ($p<.0318$)	& chi-sq ($p<.05$)
Part A			
Johnson	1st & 2nd parts of Rambler	4	(5)
Johnson	1st & 2nd parts of Travelog	3	(4)
Smith	1st & 2nd parts of Diary	2	(3)
Smith	1st part Diary and History	2	(6)
Smith	2nd part Diary and History	3	(5)
Part B			
Johnson	1st part Ram. & 1st Travelog	4	(9)
Johnson	1st part Ram. & 2nd Travelog	2	(4)
Johnson	2nd part Ram. & 1st Travelog	7	(8)
Johnson	2nd part Ram. & 2nd Travelog	4	(6)
Smith	1st part Diary & Essays	0	(2)
Smith	2nd part Diary & Essays	2	(3)
Smith	History and Essays	2	(4)

As seen in Table 2, Part A, where noncontroversial 5000-word text samples of similar literary forms are compared, the Mann-Whitney ($p<.0318$) calculation of Samuel Johnson works measures three or four rejections, while Joseph Smith shows two or three.[1] This intra-author variation is the summation of the standard predictable statistical uncertainty (which for the approximate 44 valid Morton tests predicts 44 x .0318--about an average of one and one-half rejections). This uncertainty level must be further modified to take into account errors introduced by the

1 The chi-square measurements show a somewhat higher rejection rate, due at least in part to the lower probability of $p<.05$. For simplicity of explanation, the rest of he discussion will be restricted to the Mann-Whitney calculation only.

nontheoretically predictable nonergodicity of the language, plus any small (nontheoretically predictable) component in Morton's test questions that is yet sensitive to changes in vocabulary, subject matter, or author drift. As seen here from the measurements of the above Table 2, Part A, this nonstatistical "noise" component for our control authors is about two and one-half or less (4 - 1.5 = 2.5) for Samuel Johnson and about one and one-half or less (3 - 1.5 = 1.5) for Joseph Smith.

As mentioned, at times one observes anomalous extra rejections due to changes in literary genre or form. One possible example of such a case appears in Table 2, Part B, where one of the Mann-Whitney calculations shows seven rejections from an inter-literary form comparison of Johnson. If Johnson's different literary forms had no measurable effect, we would have expected but three or four rejections, not seven. At present, one cannot predict when these anomalies will occur. Thus, without complete testing to prove otherwise, it appears unwise to make inferences across literary forms for unknown authors.

Table 3, therefore, compares two known different authors, where all comparisons are made with text samples from <u>similar</u> literary forms.

<u>Table 3</u>

Tabulation of significant null-hypothesis rejections from comparison of known different authors, each writing with the same literary form.

Text samples compared	Number of significant null-hypothesis rejections from Morton's battery of tests: Mann-Whitney (p<.0318)	& Chi-sq (p<.05)
Johnson's 2nd Trav. & Smith's 1st Diary	5	(9)
Johnson's 2nd Trav. & Smith's 2nd Diary	5	(7)
Johnson's 1st Trav. & Smith's 1st Diary	10	(11)
Johnson's 1st Trav. & Smith's 2nd Diary	7	(13)
Johnson's 1st Rambler and Smith's Essays	7	(3)
Johnson's 2nd Rambler and Smith's Essays	7	(5)

The comparison tests tabulated in Table 3 show that two of the total six Mann-Whitney test comparisons measure five rejections each. Here one-third of the comparisons show a rejection count that is too close to the single author expected rejection count of three or four to identify the works of our two known control authors as clearly different (theoretical odds being a modest thirty-to-one favoring the separation). The other four of the six comparisons in Table 3, however, show three test comparisons with seven rejections each and one with ten. These correctly predict overwhelming theoretical odds against the two authors' works having come from a single source [.0318exp(7-4) = about one in 30,000 or more against a single source for the two works, and .0318exp(10-4) = about one in a billion or more against a single source].

In this control author study we have large files. Thus we are able to establish correctly and unambiguously that our two control authors' works, each taken as a whole, statistically do not come from the same source. But if we did not have available such large files and we had been forced to compare two authors that had stylometric patterns as different as Samuel Johnson and Joseph Smith with but a single text sample of 5000 words each by this same technique, we might expect that in perhaps a large portion (i.e., perhaps one-third) of such attempts we may not measure an overwhelming statistical difference even when the compared works are known to come from different authors. Therefore, if on limited data one does not see statistically significant separation, one may not correctly infer that the compared samples are necessarily the work of the same author. On the other hand, however, a single demonstration of a large statistical rejection rate is sufficient to identify the authors' works as statistically different.

With this background information in mind, now turn to the two attached computer printouts labeled "Figures 5 and 6." They show the Mann-Whitney and Chi-square calculations for the first five consecutive 1000-word groups of the Book of Mormon manuscript's didactic writings purportedly first written by Nephi. Nephi's first 5000 words are tested against several other 5000-word didactic text samples both of himself and of other authors. The results are summarized in Table 4.[2]

Table 4

Tabulation of significant null-hypothesis rejections between purported Nephi (part 1), and two of his additional consecutive sample texts (parts 2 and 3), and three different consecutive sample texts purportedly from Alma (parts 1, 2 and 3), and two control authors Joseph Smith and Samuel Johnson. All are written in didactic literary form.

Column from Fig. Six	Test samples compared	Number of significant null-hypothesis rejections from Morton's battery of tests: Mann-Whitney's ($p<.0318$)	& chi-sq ($p<.05$)
1	Nephi part 1 & Nephi part 2	3	(9)
2	Nephi part 1 & Nephi part 3	3	(6)
3	Nephi part 1 & Alma part 1	8	(12)
4	Nephi part 1 & Alma part 2	6	(13)
5	Nephi part 1 & Alma part 3	8	(10)
6	Nephi part 1 & J.Smith essays	10	(14)
7	Nephi part 1 & Johnson Ramb. 1	13	(19)

Table 4 shows that the number of rejections for tests of Nephi's first 5000 didactic words against his two other 5000-word samples (columns 1 and 2 from Figure 6) continue to measure the same intra-author Mann-Whitney value of 3 that was previously measured for the control authors. This consistency leads to an accurate statistical error of <.0318 for these "p<.0318" calculations. This confirms that the calibration of the "unknown" author Nephi can in this case safely be extrapolated from the intra-author measurements of the "known" control authors.

---

2 For those who wish to verify the correctness of the calculational coding, the back side of Figure 5 and 6 Computer printouts shows the raw count tabulation for each of the five 1000-word blocks making up the 5000 test words for the didactic passages of Nephi and Alma with respect to each of the word pattern tests from Morton's battery. Of course the phrase "(and) it came to pass (that)" is rarely used in didactic writing, so no special tabulation is made for it.

The conclusion then to be drawn from the foregoing tests is that the probability that either Joseph Smith or Samuel Johnson wrote the first 5000 of Nephi's didactic words is shown to be insignificantly small. While it is not always to be expected, in this case the 1 Nephi sample tested against Alma (columns 3-5 from Figure 6) also shows a clear separation, with theoretical odds against the compared samples having been written by the same person thus being more than 70 million to one, 30,000 to one, and 70 million to one, respectively.

D. SUMMARY

The textual structure of the Book of Mormon is complex. Wordprint testable questions call for definitive answers, but they will not be answered by simplistic and faulty calculations, as seen in Taves' attempt. For present purposes, we have assumed (as has Taves) that Morton's theories and methods offer valid measurements and techniques. But by applying them in ways we believe to be more accurate and legitimate, as explained above, we have shown with a few preliminary measurements (even at the 5000-word test sample size) that the probability of Joseph Smith having written the first 5000 didactic words of Nephi is unsupportably low. This directly contradicts Taves' results. Furthermore, the samples of didactic Alma are clearly different from this sample of Nephi's comparable writings. Taves did not get these results, for they were below his level of measurement sensitivity, due to improper text sampling, invalid calculation of his test statistic and other causes. I therefore cannot recommend reading Taves' chapters on stylometry, unless one is interested in seeing an example of how it cannot and should not be done.

John L. Hilton

FIGURE FIVE

UNIVARIATE TESTS, MANN-WHITNEY (Rank-Sum) 'T' Statistic (Morton's New Wrd Ratios)
FOR STATISTICALLY SIGNIFICANT REJECTIONS OF THE NULL-HYPOTHESIS (Ho:u1=u2 PROB <.03/8)
TODAYS DATE (AND TIME) IS 04-14-1985 23:55:23
All compared text samples are didactic literary form and 5000 wrds each.

Nephi 1st compared to:	Nephi 2nd		Nephi 3rd		Alma 1st		Alma 2nd		Alma 3rd		Joseph Smith 1st		Samuel Johnson 1st	
Valid Morton tests*	M-W(T)	prob.	M-W(T)	prob.	M-W(T)	prob.	M-W(T)	prob.	M-W(T)	prob.	M-W(T)	prob.	M-W(T)	prob.
AND(fws)/#	19	-	16	p<.02	16	p<.02	17	p<.05	15	p<.01	16	p<.02	15	p<.01
IN(fws)/#	27	small	27	small	25	small	25	small	25	small	27	small	18	-
IT(fws)/#	25	small	27.5	small	27.5	small	22.5	small	27.5	small	25	small	15	p<.01
IT(lws)/#	20	-	26	small	27	small	26	small	27	small	25	small	23	small
OF(2nd lws)/#	25	-	26	-	24	-	19	-	22	-	18	-	24	-
THE(fws)/#	25	small	27	small	25	small	25	small	25	small	27	small	15	p<.01
THE(2nd lws)/#	23	-	25	-	20	-	19	-	22	-	24	-	18	-
A(fb adj)/A	27	-	25.5	-	25.5	-	25.5	-	25	-	28.5	-	27	-
A(fb x AND)/A	20.5	-	20	-	16.5	p<.05	18	-	16.5	p<.05	18.5	-	15.5	p<.02
A(fb x OF)/A	28.5	-	25	-	21.5	-	25	-	25	-	27.5	-	23.5	-
AND(fb ADJ)/AND	25	-	25	-	23	-	22	-	23	-	27	-	25	-
AND(fb THE)/AND	27	-	26	-	23	-	20	-	20	-	24	-	19	-
AND(fb x OF)/AND	24	-	24	-	20	-	27	-	22.5	-	24	-	27	-
BE(pb TO)/BE	18.5	-	25	small	25	-	20.5	-	20.5	-	15	p<.01	15	p<.01
BY(fb THE)/BY	16.5	p<.05	29	-	21.5	-	15.5	p<.02	20	-	16	p<.02	15	p<.01
I(fb HAVE)/I	24	-	16	p<.02	22.5	-	25.5	-	17	p<.05	19.5	-	26	-
IN(fb A)/IN	24.5	-	18	-	18.5	-	27	-	17.5	-	22.5	-	18.5	-
IN(fb THE)/IN	19.5	-	26	-	24.5	-	24	-	25	-	23	-	20.5	-
OF(fb A)/OF	26	small	27	small	24	small	27	small	24	small	24	small	20	-
OF(fb THE)/OF	15	p<.01	21	-	16	p<.02	15	p<.01	15	p<.01	18	-	15	p<.01
OF(fb x and)/OF	26	-	24	-	23	-	23	-	25	-	21	-	18	-
THE(pb AND)/THE	26	-	27	-	24	-	21	-	23	-	25	-	19	-
THE(pb OF)/THE	15	p<.01	19	-	18	-	20	-	16	p<.02	19	-	15	p<.01
THE(pb IN)THE	20	-	27	-	18	-	25	-	23	-	17	p<.05	24	-
THE(pb TO)/THE	24	-	26	-	15	p<.01	15	p<.01	17	p<.05	15.5	p<.02	22	-
THE(fb x AND)/THE	26	-	25	-	19	-	24	-	22	-	24.5	-	19	-
THE(fb x x THE)/THE	20	-	24	-	17	p<.05	17	p<.05	15	p<.01	17	p<.05	15	p<.01
TO(fb BE)/TO	18.5	-	26	small	25	-	22.5	-	20.5	-	16	p<.02	15	p<.01
TO(fb THE)/TO	26	-	27.5	-	25.5	-	25	-	27	-	25	-	22	-
(to-be Verbs)/Verbs	26	-	21	-	27	-	20	-	23	-	24	-	18	-
AN/AN+A	26	-	17.5	-	27	-	23	-	22	-	24	-	26	-
ANY/ANY+ALL	22.5	small	20	-	25	small	22.5	small	20	small	15	p<.01	15	p<.01
NO/NO+NOT	20	-	22	-	22	small	17	p<.05	19	-	16	p<.02	18	-
VERBs/VERBs+ADJs	25	-	17	p<.05	15	p<.01	23	-	20	-	24	-	23	-
A(r)/A(r+1)use only	27.5	small	25	-	27.5	-	25	small	25	-	25.5	-	22.5	-
AND(r)/AND(r+1)only	22.5	-	18	-	20	-	26.5	-	20	-	18	-	18	-
IN(r)/IN(r+1)only	22.5	-	24	-	26.5	-	26.5	-	22	-	22.5	-	27.5	-
IT(r)/IT(r+1)only	27.5	small	24	small	24.5	small	22	-	21.5	-	18	-	15	p<.01
I(r)/I(r+1)use only	26.5	-	27.5	-	15.5	p<.02	29	-	24	-	21	-	27	-
OF(r)/OF(r+1)only	19	-	24	-	15	p<.01	24.5	-	17	p<.05	15	p<.01	15	p<.01
THAT(r)/THAT(r+1)	24	-	20	-	24	-	23	-	29	-	25	-	23	-
THE(r)/THE(r+1)only	24	-	24.5	-	22.5	-	21	-	22.5	-	19	-	22	-
TO(r)/TO(r+1)only	24.5	-	23	-	23	-	20.5	-	22	-	25	-	22	-
Total number of M-W rej.	3		3		8		6		8		10		13	
Total valid Morton tests	36		35		35		36		36		38		42	
Gross Percent Rejections	8%		9%		23%		17%		22%		26%		31%	

* #=number "end of sentence" markers, (fws)=first wrd in sentence, (lws)=last wrd in sentence, (2nd lws)=2nd to last...
(fb)=followed by, (pb)=preceded by, (...x...)=any wrd, (r+1)only=the wrd to the right and left are uniquely used

TODAY'S DATE IS 04-01-1985 THE TIME IS 16:29:35
THIS IS FILE 'WP0-5K1.NIS' WHICH IS THE FIRST 5000 DIDACTIC WRDS OF Nephi
WRDs ARE LISTED IN NUMBER OF WORDS. RATIOS ARE IN PER CENT(e.g. %=100*RATIO).

Wrds & wrd patterns	(Wrd Groups)					Total COUNTS	Mean /1k	Sig /1k
A	6.000	9.000	6.000	5.000	5.000	31	6.20	1.64
AN	3.000	1.000	1.000	0.000	2.000	7	1.40	1.14
AND	46.000	60.000	105.000	100.000	58.000	369	73.80	26.80
ANY	0.000	0.000	0.000	0.000	0.000	0	0.00	0.00
ALL	7.000	8.000	5.000	4.000	12.000	36	7.20	3.11
BUT	3.000	0.000	0.000	0.000	2.000	5	1.00	1.41
BE	13.000	11.000	0.000	0.000	7.000	31	6.20	6.06
BY	2.000	5.000	3.000	3.000	4.000	17	3.40	1.14
I	6.000	30.000	53.000	41.000	29.000	159	31.80	17.40
IN	19.000	21.000	9.000	9.000	11.000	69	13.80	5.76
IT	11.000	12.000	17.000	28.000	12.000	80	16.00	7.11
NO	0.000	1.000	0.000	0.000	0.000	1	0.20	0.45
NOT	4.000	3.000	3.000	1.000	12.000	23	4.60	4.28
OF	49.000	44.000	57.000	60.000	78.000	288	57.60	13.05
THAT	39.000	31.000	36.000	40.000	31.000	177	35.40	4.28
THE	75.000	85.000	91.000	98.000	115.000	464	92.80	15.01
TO	15.000	17.000	18.000	24.000	18.000	92	18.40	3.36
WITH	2.000	7.000	3.000	2.000	2.000	16	3.20	2.17
FORMS OF 'TO-BE'	22.000	17.000	29.000	20.000	30.000	118	23.60	5.68
HAVE	15.000	3.000	0.000	0.000	4.000	22	4.40	6.19
VERBs	223.000	218.000	216.000	180.000	193.000	1030	206.00	18.56
ADJs	145.000	164.000	133.000	165.000	192.000	799	159.80	22.47
End of SENTENCE	43.000	43.000	67.000	66.000	40.000	259	51.80	13.48
A(fws)/#	0.000	0.000	0.000	0.000	0.000	0/ 259	0.00	0.00
AN(fws)/#	0.000	0.000	0.000	0.000	0.000	0/ 259	0.00	0.00
AND(fws)/#	39.535	76.744	91.045	98.485	62.500	201/ 259	73.66	23.54
IN(fws)/#	0.000	0.000	0.000	0.000	2.500	1/ 259	0.50	1.12
IT(fws)/#	0.000	0.000	0.000	0.000	0.000	0/ 259	0.00	0.00
IT(lws)/#	0.000	0.000	0.000	3.030	0.000	2/ 259	0.61	1.36
OF(fws)/#	0.000	0.000	0.000	0.000	0.000	0/ 259	0.00	0.00
OF(2nd lws)/#	11.628	9.302	14.925	6.061	17.500	30/ 259	11.88	4.51
THE(fws)/#	2.326	0.000	0.000	0.000	0.000	1/ 259	0.47	1.0[illegible]
THE(2nd lws)/#	32.558	11.628	13.433	7.576	17.500	40/ 259	16.54	9.6[illegible]
WITH(2nd lws)/#	0.000	0.000	0.000	0.000	0.000	0/ 259	0.00	0.00
A(fb adj)/A	16.667	11.111	16.667	60.000	60.000	9/ 31	32.89	24.85
A(fb x AND)/A	16.667	33.333	33.333	40.000	20.000	9/ 31	28.67	9.89
A(fb x OF)/A	16.667	0.000	50.000	0.000	20.000	5/ 31	17.33	20.47
AND(fb ADJ)/AND	10.870	23.333	22.857	22.000	12.069	72/ 369	18.23	6.20
AND(fb THE)/AND	2.174	13.333	14.286	14.000	5.172	41/ 369	9.79	5.70
AND(fb x OF)/AND	4.348	1.667	0.952	3.000	6.897	11/ 369	3.37	2.36
BE(fb A)/BE	0.000	0.000	0.000	0.000	0.000	0/ 31	0.00	0.00
BE(pb TO)/BE	7.692	0.000	0.000	0.000	14.286	2/ 31	4.40	6.45
BUT(fb A)/BUT	0.000	0.000	0.000	0.000	50.000	1/ 5	10.00	22.36
BY(fb THE)/BY	100.000	80.000	100.000	100.000	50.000	14/ 17	86.00	21.91
I(fb AN)/I	0.000	0.000	0.000	0.000	0.000	0/ 159	0.00	0.00
I(fb HAVE)/I	0.000	6.667	0.000	0.000	10.345	5/ 159	3.40	4.84
IN(fb A)/IN	10.526	9.524	11.111	0.000	9.091	6/ 69	8.05	4.57
IN(fb THE)/IN	26.316	33.333	55.556	22.222	27.273	22/ 69	32.94	13.25
OF(fb A)/OF	0.000	2.273	1.754	1.667	0.000	3/ 288	1.14	1.07
OF(fb THE)/OF	55.102	40.909	36.842	41.667	55.128	134/ 288	45.93	8.58
OF(fb x and)/OF	10.204	11.364	26.316	8.333	7.692	36/ 288	12.78	7.71
THE(pb AND)/THE	1.333	9.412	16.484	14.286	2.609	41/ 464	8.82	6.77
THE(pb OF)/THE	36.000	21.176	23.077	25.510	37.391	134/ 464	28.63	7.54
THE(pb IN)THE	6.667	8.235	5.495	2.041	2.609	22/ 464	5.01	2.64
THE(pb TO)/THE	5.333	1.176	3.297	2.041	5.217	16/ 464	3.41	1.86
THE(fb x AND)/THE	9.333	7.059	16.484	10.204	12.174	52/ 464	11.05	3.55
THE(fb x THE)/THE	0.000	1.176	0.000	1.020	0.000	2/ 464	0.44	0.60
THE(fb x x THE)/THE	20.000	20.000	20.879	26.531	23.478	104/ 464	22.18	2.82
TO(fb BE)/TO	6.667	0.000	0.000	0.000	5.556	2/ 92	2.44	3.37
TO(fb THE)/TO	26.667	5.882	16.667	8.333	33.333	16/ 92	18.18	11.75
(to-be Verbs)/Verbs	9.865	7.798	13.426	11.111	15.544	118/1030	11.55	3.02
AN/AN+A	33.333	10.000	14.286	0.000	28.571	7/ 38	17.24	13.65
ANY/ANY+ALL	0.000	0.000	0.000	0.000	0.000	0/ 36	0.00	0.00
NO/NO+NOT	0.000	25.000	0.000	0.000	0.000	1/ 24	5.00	11.18
VERBs/VERBs+ADJs	60.598	57.068	61.891	52.174	50.130	1030/1829	56.37	5.13
A(r)/A(r+1)use only	0.000	0.000	33.333	33.333	33.333	4/ 21	20.00	18.2[illegible]
AND(r)/AND(r+1)only	58.621	71.429	66.667	66.667	63.636	99/ 150	65.40	4.7[illegible]
IN(r)/IN(r+1)only	77.778	70.000	75.000	50.000	63.636	25/ 36	67.28	11.06
IT(r)/IT(r+1)only	0.000	0.000	50.000	0.000	50.000	2/ 5	20.00	27.39
I(r)/I(r+1)use only	0.000	42.857	33.333	0.000	50.000	7/ 21	25.24	23.79
OF(r)/OF(r+1)only	73.913	80.769	84.000	75.000	94.444	88/ 109	81.63	8.28
THAT(r)/THAT(r+1)	100.000	100.000	100.000	0.000	83.333	23/ 25	76.67	43.46
THE(r)/THE(r+1)only	23.077	11.765	10.345	25.000	35.000	25/ 130	21.04	10.19
TO(r)/TO(r+1)only	33.333	80.000	50.000	100.000	83.333	16/ 27	69.33	27.02

FIGURE [illegible]X

UNIVER[illegible] TESTS 2 x 2 CHI SQUARE Statistic, (Morton's New Wrd R[illegible]os)
FOR STATISTICALLY SIGNIFICANT REJECTIONS OF THE NULL-HYPOTHESIS(Ho:u1=u2 PROB <.05)
TODAYS DATE (AND TIME) IS 04-14-1985 16:03:44
All compared text samples are didatic literary form and 5000 wrds each.

Nephi 1st compared to:	Nephi 2nd	Nephi 3rd	Alma 1st	Alma 2nd	Alma 3rd	Joseph Smith 1st	Samuel Johnson 1st
Valid Morton tests*	X-sq(prob.)	X-sq(prob.)	X-sq(prob.)	X-sq(prob.)	X-sq(prob.)	X-sq(prob.)	X-sq(prob.)
AND(fws)/#	55.8(p<.005)	79.2(p<.005)	98.9(p<.005)	109.3(p<.005)	103.9(p<.005)	51.1(p<.005)	174.4(p<.005)
IN(fws)/#	(small)	(small)	(small)	(small)	(small)	(small)	(small)
IT(fws)/#	(small)	(small)	(small)	(small)	(small)	(small)	22.3(p<.005)
IT(lws)/#	(small)	(small)	(small)	(small)	(small)	(small)	(small)
OF(2nd lws)/#	.6(-)	.0(-)	1.4(-)	3.4(-)	1.3(-)	(small)	1.3(-)
THE(fws)/#	(small)	(small)	(small)	(small)	(small)	(small)	20.8(p<.005)
THE(2nd lws)/#	2.5(-)	.4(-)	2.5(-)	8.8(p<.005)	3.4(-)	(small)	8.5(p<.005)
A(fb adj)/A	.0(-)	.0(-)	.0(-)	.2(-)	.0(-)	.2(-)	.6(-)
A(fb x AND)/A	.9(-)	2.3(-)	5.3(p<.025)	(small)	2.6(-)	2.8(-)	(small)
A(fb x OF)/A	(small)	.0(-)	.8(-)	(small)	(small)	(small)	(small)
AND(fb ADJ)/AND	.1(-)	.3(-)	2.0(-)	6.4(p<.025)	2.2(-)	.1(-)	.0(-)
AND(fb THE)/AND	.0(-)	.0(-)	3.5(-)	11.9(p<.005)	8.6(p<.005)	1.7(-)	9.5(p<.005)
AND(fb x OF)/AND	.2(-)	.4(-)	1.2(-)	.0(-)	1.1(-)	.5(-)	.6(-)
BE(pb TO)/BE	(small)	(small)	(small)	3.6(-)	(small)	8.3(p<.005)	8.0(p<.005)
BY(fb THE)/BY	9.9(p<.005)	.5(-)	1.6(-)	11.7(p<.005)	6.5(p<.025)	21.2(p<.005)	(small)
I(fb HAVE)/I	4.8(p<.050)	(small)	6.9(p<.010)	(small)	13.4(p<.005)	(small)	(small)
IN(fb A)/IN	(small)	(small)	(small)	.0(-)	(small)	(small)	(small)
IN(fb THE)/IN	2.5(-)	.0(-)	.1(-)	.1(-)	1.2(-)	.9(-)	1.6(-)
OF(fb A)/OF	(small)	(small)	(small)	(small)	(small)	(small)	5.9(p<.025)
OF(fb THE)/OF	20.8(p<.005)	5.2(p<.025)	22.8(p<.005)	14.7(p<.005)	39.6(p<.005)	9.4(p<.005)	44.9(p<.005)
OF(fb x and)/OF	.1(-)	.1(-)	1.0(-)	3.1(-)	.1(-)	3.6(-)	5.5(p<.025)
THE(pb AND)/THE	.4(-)	.0(-)	2.9(-)	8.4(p<.005)	3.3(-)	7.7(p<.010)	16.9(p<.005)
THE(pb OF)/THE	12.2(p<.005)	8.6(p<.005)	10.2(p<.005)	5.8(p<.025)	19.8(p<.005)	9.7(p<.005)	15.6(p<.005)
THE(pb IN)THE	5.3(p<.025)	.2(-)	8.7(p<.005)	1.9(-)	2.4(-)	9.9(p<.005)	.9(-)
THE(pb TO)/THE	2.9(-)	.0(-)	5.0(p<.025)	6.9(p<.010)	6.6(p<.010)	4.2(p<.050)	3.8(-)
THE(fb x AND)/THE	.0(-)	.4(-)	4.8(p<.050)	2.2(-)	1.1(-)	.6(-)	4.8(p<.050)
THE(fb x x THE)/THE	1.9(-)	.8(-)	6.0(p<.025)	6.1(p<.025)	10.8(p<.005)	9.0(p<.005)	20.6(p<.005)
TO(fb BE)/TO	6.2(p<.025)	(small)	(small)	(small)	3.4(-)	6.7(p<.010)	10.3(p<.005)
TO(fb THE)/TO	.3(-)	.0(-)	.5(-)	.5(-)	.0(-)	1.1(-)	3.0(-)
(to-be Verbs)/Verbs	.2(-)	4.2(p<.050)	.1(-)	3.6(-)	1.7(-)	1.3(-)	9.2(p<.005)
AN/AN+A	.0(-)	(small)	.2(-)	.8(-)	(small)	.4(-)	.0(-)
ANY/ANY+ALL	(small)	(small)	(small)	(small)	(small)	12.5(p<.005)	27.5(p<.005)
NO/NO+NOT	(small)	(small)	(small)	(small)	(small)	5.8(p<.025)	(small)
VERBs/VERBs+ADJs	3.3(-)	14.4(p<.005)	28.6(p<.005)	16.5(p<.005)	10.2(p<.005)	1.9(-)	4.3(p<.050)
A(r)/A(r+1)use only	(small)	(small)	(small)	(small)	.4(-)	.3(-)	.9(-)
AND(r)/AND(r+1)only	.1(-)	1.9(-)	1.7(-)	.1(-)	.4(-)	1.4(-)	2.3(-)
IN(r)/IN(r+1)only	.0(-)	.0(-)	.0(-)	.0(-)	.0(-)	.4(-)	.0(-)
IT(r)/IT(r+1)only	(small)	(small)	(small)	(small)	(small)	(small)	(small)
I(r)/I(r+1)use only	.0(-)	(small)	4.4(p<.050)	(small)	.0(-)	1.7(-)	(small)
OF(r)/OF(r+1)only	5.0(p<.050)	.4(-)	6.2(p<.025)	2.8(-)	4.9(p<.050)	7.2(p<.010)	8.7(p<.005)
THAT(r)/THAT(r+1)	4.1(p<.050)	6.3(p<.025)	1.8(-)	4.6(p<.050)	.5(-)	1.1(-)	4.9(p<.050)
THE(r)/THE(r+1)only	1.4(-)	1.6(-)	1.1(-)	4.2(p<.050)	2.3(-)	6.1(p<.025)	3.0(-)
TO(r)/TO(r+1)only	.0(-)	.0(-)	.5(-)	2.0(-)	.0(-)	.0(-)	.7(-)
Total number of Chi-sq Rej	9	6	12	13	10	14	19
Total valid Morton tests	31	28	31	29	31	32	33
Gross Percent Rejections	29%	21%	39%	45%	32%	44%	58%

* #=number "end of sentence" markers, (fws)=first wrd in sentence, (lws)=last wrd in sentence, (2nd lws)=2nd to last...
(fb)=followed by, (pb)=preceeded by, (...x...)=any word, (r+1)only=the wrd to the right and left are uniquely used

TODAY'S DATE IS 03-31-1985 THE TIME IS 12:19:03
THIS IS FILE 'WP0-5K1.AAS' WHICH IS THE FIRST 5000 WRDS OF ALMA
WRDs ARE LISTED IN NUMBER OF WORDS. RATIOS ARE IN PER CENT(e.g. %=100*RATIO).

Wrds & wrd patterns (Wrd Groups)						Total COUNTS	Mean /1k	Sigm /1k
A	12.000	10.000	3.000	3.000	5.000	33	6.60	4.16
AN	3.000	7.000	0.000	1.000	0.000	11	2.20	2.95
AND	47.000	35.000	48.000	49.000	51.000	230	46.00	6.32
ANY	0.000	1.000	0.000	0.000	0.000	1	0.20	0.45
ALL	4.000	7.000	6.000	7.000	8.000	32	6.40	1.52
BUT	1.000	3.000	2.000	3.000	4.000	13	2.60	1.14
BE	7.000	2.000	12.000	11.000	15.000	47	9.40	5.03
BY	12.000	6.000	6.000	3.000	6.000	33	6.60	3.29
I	21.000	29.000	33.000	28.000	7.000	118	23.60	10.24
IN	22.000	8.000	20.000	17.000	18.000	85	17.00	5.39
IT	3.000	2.000	5.000	3.000	7.000	20	4.00	2.00
NO	1.000	1.000	1.000	1.000	0.000	4	0.80	0.45
NOT	5.000	18.000	14.000	12.000	15.000	64	12.80	4.87
OF	47.000	45.000	36.000	31.000	50.000	209	41.80	7.98
THAT	17.000	29.000	41.000	40.000	29.000	156	31.20	9.81
THE	54.000	51.000	65.000	55.000	64.000	289	57.80	6.30
TO	22.000	15.000	18.000	23.000	16.000	94	18.80	3.56
WITH	4.000	4.000	1.000	3.000	3.000	15	3.00	1.22
FORMS OF 'TO-BE'	31.000	46.000	21.000	25.000	19.000	142	28.40	10.85
HAVE	12.000	16.000	9.000	16.000	9.000	62	12.40	3.51
VERBs	235.000	256.000	232.000	217.000	248.000	1188	237.60	15.08
ADJs	129.000	132.000	131.000	129.000	119.000	640	128.00	5.20
End of SENTENCE	50.000	52.000	41.000	46.000	42.000	231	46.20	4.82
A(fws)/#	0.000	0.000	0.000	0.000	0.000	0/ 231	0.00	0.00
AN(fws)/#	0.000	0.000	0.000	0.000	0.000	0/ 231	0.00	0.00
AND(fws)/#	32.000	23.077	48.780	32.609	30.952	76/ 231	33.48	9.38
IN(fws)/#	0.000	0.000	0.000	0.000	0.000	0/ 231	0.00	0.00
IT(fws)/#	0.000	0.000	0.000	0.000	0.000	0/ 231	0.00	0.00
IT(lws)/#	0.000	1.923	0.000	0.000	0.000	1/ 231	0.38	0.86
OF(fws)/#	0.000	0.000	0.000	0.000	0.000	0/ 231	0.00	0.00
OF(2nd lws)/#	18.000	25.000	4.878	13.043	14.286	36/ 231	15.04	7.34
THE(fws)/#	0.000	0.000	0.000	0.000	0.000	0/ 231	0.00	0.00
THE(2nd lws)/#	10.000	5.769	7.317	2.174	28.571	24/ 231	10.77	10.3
WITH(2nd lws)/#	0.000	0.000	0.000	0.000	0.000	0/ 231	0.00	0.00
A(fb adj)/A	58.333	0.000	33.333	0.000	20.000	9/ 33	22.33	24.60
A(fb x AND)/A	0.000	10.000	0.000	0.000	20.000	2/ 33	6.00	8.94
A(fb x OF)/A	8.333	40.000	33.333	66.667	20.000	9/ 33	33.67	22.12
AND(fb ADJ)/AND	14.894	14.286	10.417	18.367	15.686	34/ 230	14.73	2.87
AND(fb THE)/AND	4.255	2.857	6.250	8.163	9.804	15/ 230	6.27	2.82
AND(fb x OF)/AND	2.128	0.000	0.000	0.000	3.922	3/ 230	1.21	1.77
BE(fb A)/BE	14.286	0.000	0.000	0.000	0.000	1/ 47	2.86	6.39
BE(pb TO)/BE	42.857	0.000	0.000	9.091	6.667	5/ 47	11.72	17.87
BUT(fb A)/BUT	0.000	0.000	0.000	0.000	0.000	0/ 13	0.00	0.00
BY(fb THE)/BY	66.667	33.333	50.000	66.667	83.333	20/ 33	60.00	19.00
I(fb AN)/I	0.000	0.000	0.000	0.000	0.000	0/ 118	0.00	0.00
I(fb HAVE)/I	0.000	10.345	15.152	21.429	0.000	14/ 118	9.38	9.43
IN(fb A)/IN	0.000	0.000	5.000	0.000	5.556	2/ 85	2.11	2.90
IN(fb THE)/IN	40.909	50.000	40.000	29.412	22.222	30/ 85	36.51	10.82
OF(fb A)/OF	4.255	0.000	0.000	0.000	0.000	2/ 209	0.85	1.90
OF(fb THE)/OF	17.021	22.222	30.556	16.129	38.000	53/ 209	24.79	9.35
OF(fb x and)/OF	14.894	6.667	5.556	6.452	12.000	20/ 209	9.11	4.11
THE(pb AND)/THE	3.704	1.961	4.615	7.273	7.813	15/ 289	5.07	2.46
THE(pb OF)/THE	14.815	19.608	16.923	9.091	29.688	53/ 289	18.02	7.58
THE(pb IN)THE	16.667	7.843	12.308	9.091	6.250	30/ 289	10.43	4.13
THE(pb TO)/THE	5.556	7.843	6.154	10.909	6.250	21/ 289	7.34	2.17
THE(fb x AND)/THE	7.407	1.961	9.231	1.818	9.375	18/ 289	5.96	3.79
THE(fb x THE)/THE	0.000	0.000	0.000	0.000	0.000	0/ 289	0.00	0.00
THE(fb x x THE)/THE	12.963	13.725	16.923	9.091	20.313	43/ 289	14.60	4.24
TO(fb BE)/TO	13.636	0.000	0.000	4.348	6.250	5/ 94	4.85	5.62
TO(fb THE)/TO	13.636	26.667	22.222	26.087	25.000	21/ 94	22.72	5.36
(to-be Verbs)/Verbs	13.191	17.969	9.052	11.521	7.661	142/1188	11.88	4.02
AN/AN+A	20.000	41.176	0.000	25.000	0.000	11/ 44	17.24	17.57
ANY/ANY+ALL	0.000	12.500	0.000	0.000	0.000	1/ 33	2.50	5.59
NO/NO+NOT	16.667	5.263	6.667	7.692	0.000	4/ 68	7.26	6.04
VERBs/VERBs+ADJs	64.560	65.979	63.912	62.717	67.575	1188/1828	64.95	1.88
A(r)/A(r+1)use only	0.000	57.143	33.333	0.000	25.000	6/ 23	23.10	24.1
AND(r)/AND(r+1)only	47.368	53.846	60.714	53.333	67.742	80/ 138	56.60	7.[illegible]
IN(r)/IN(r+1)only	50.000	100.000	62.500	66.667	83.333	28/ 40	72.50	19.45
IT(r)/IT(r+1)only	100.000	0.000	0.000	100.000	100.000	4/ 6	60.00	54.77
I(r)/I(r+1)use only	50.000	85.714	57.143	66.667	100.000	16/ 23	71.90	20.65
OF(r)/OF(r+1)only	73.684	59.091	73.684	72.222	42.105	63/ 98	64.16	13.77
THAT(r)/THAT(r+1)	75.000	75.000	70.000	100.000	76.923	33/ 42	79.38	11.81
THE(r)/THE(r+1)only	30.769	25.000	10.714	27.586	40.000	29/ 115	26.81	10.64
TO(r)/TO(r+1)only	45.455	40.000	54.545	60.000	60.000	24/ 50	52.00	8.96