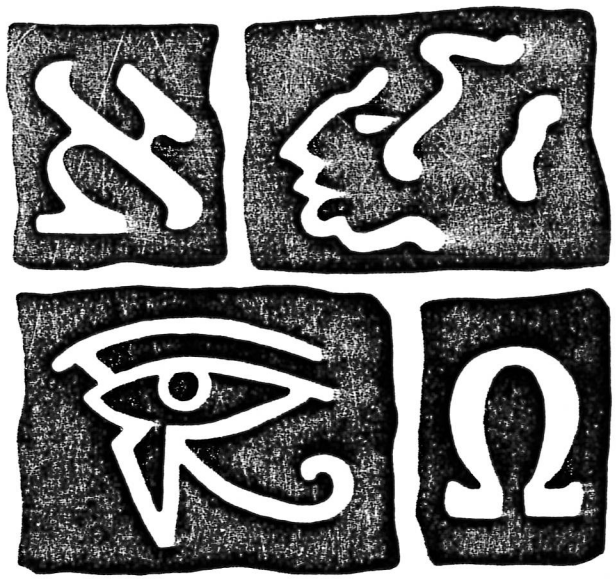
BRIGHAM YOUNG UNIVERSITY • PROVO, UTAH

FARMS Paper



Some Book of Mormon "Wordprint" Measurements Using "Wraparound" Block Counting

John L. Hilton

HIL-88

Paper

FOUNDATION FOR ANCIENT RESEARCH AND MORMON STUDIES

F.A.R.M.S.
P.O. BOX 7113
UNIVERSITY STATION
PROVO, UTAH 84602

SOME BOOK OF MORMON "WORDPRINT" MEASUREMENTS USING "WRAP-AROUND" BLOCK COUNTING

by John L. Hilton
July 1988

Since the advent of modern programmable high speed computers, a new science of computer assisted literary "stylometry" or "wordprinting" has been developing. By statistically analyzing the small non-contextual word patterns of a disputed document, it is now possible to test which suspected authors did not write a given work.[1] Modern wordprinting techniques are based on the somewhat surprising observation that each author studied has shown a consistent near-subconscious habitual tendency to write with complex patterns of often-used words (such as "and," "the," "of," "that," etc.) at statistically significant different rates than do others.

As "stylometry" has been developing, several different analytic techniques have shown promise in specific literary situations. For the wordprint tests reported in this paper, we

---

1 Useful non-contextual word patterns meet the following conditions: they yield a non-ambiguous count, they occur frequently, they have common alternate expressions, and their use rates tend to become habitual. Patterns should be minimally affected by the period of the writer's career, the subject matter, or the genre. Therefore, useful word patterns are typically made up from key words such as common articles, conjunctions, prepositions, and some pronouns. Measurements are calculated from the ratio of the overall key word use rate against the same key word use rate in certain sentence positions, word collocations, proportional pairs, or their use adjacent to certain parts of speech and novel vocabulary words. The significance of such patterns in ascertaining authorship has been explored in recent decades especially by Andrew Q. Morton, Literary Detection (New York: Scribner's Sons, 1979).

used a non-parametric "wrap-around" word-block counting for paired texts.[2] This methodology is explained in detail in a longer technical paper entitled, "On Maximizing Author Identification by Measuring 5000 Word Texts."[3] That paper demonstrates in great detail how this measuring method avoids the main problems found in earlier statistical approaches and how it yields valid results when studying sections of documents approximately 5000 words in length. In particular, it shows how the validity of this type of measurement was tested by calculating 325 diversified wordprint proof tests. These tests used the non-contested writings of 9 control authors. All possible pairs of 26 non-controversial 5000 word texts written by these authors, who wrote under various conditions, were examined. The individual and averaged results rigorously supported the basic wordprint assumption that each author uses many non-contextual word patterns in his writings at a unique, relatively stable rate, irrespective of literary form, time, and subject matter. Among the works studied were texts by Oliver Cowdery and samples of Joseph Smith's autographic and dictated writings.

---

2 In our "wrap-around" approach, we analyze each 4998 word text, in 17 word blocks of 294 words, 49 different times. The first tabulation cycle begins with the first word; the second cycle begins with the 6th word and wraps back around to count the first unused words at the end of the last word block; the third cycle begins with the 12th word and similarly wraps back around again; the fourth cycle begins with the 18th word, etc., through the 49 cycles. A wordprint measurement is then calculated from the mean of these 49 cycles.

3 By John L. Hilton and Kenneth D. Jenkins; it is a technical unpublished working paper, dated September, 1987, available through F.A.R.M.S., P.O. Box 7113, University Station, Provo, UT 84602.

These measurements demonstrate statistically significant differences between (a) paired 5000 word texts where both were written by the same author (i.e. "within-author" samples) and (b) paired writings where the two texts were written by different authors (i.e. "between-author" samples). Most "between-author" pairs can be unambiguously separated from "within-author" pairs to a high degree of certainty, even where skilled authors have consciously tried to make their writings appear to have come from different people.

Also studied were English translations of several semi-classical German texts written by different authors. These academic translations were all carefully done by the same English translator. The wordprint measurement showed that each translated work was consistent within itself, but each was clearly separable from the same translator's rendition of other German authors' works, as well as from his own original English writings. This demonstrates, at least for these cases, that the uniqueness of an original author's wordprint may survive translation.

Wordprinting is an objective measurement. When done correctly, it gives information which is not distorted by the opinions of the investigator. Not all participants in the scientific work that led to the development of the present "wrap-around" block counting technique were believers in the Book of Mormon; they came from a broad spectrum of philosophic backgrounds (i.e. agnostic, Jewish, etc.). That this group has achieved consensus suggests that the wordprinting system was able to achieve the desired objectivity.

Our group has examined the Book of Mormon with wordprint measurements, since the book purports to contain large sections written, at least originally, by different authors. Since some purported authors are represented by several separate 5000 word sections,[4] one may compute the wordprint calculations to almost any degree of statistical certainty desired.

Most people would expect Joseph Smith's limited and somewhat unique 1829 vocabulary to be found throughout the Book of Mormon, since he translated the entire work apparently into his own English language (consistent with D&C 1:24; 9:8). This expectation is objectively supported by a measurement of the average rate at which new vocabulary words are introduced across 500 word sections of the text. Samples of the new-word-use-rate calculated from the Book of Mormon manuscript are presented in the lower four curves of Figure 1. The curves represent the mean measurements made by calculating across all of the 500 word blocks written by the three major Book of Mormon writers for two different genre. As can be seen by each of the end values plotted at the 450 word point, all Book of Mormon authors write with a statistically similar (i.e., less than 95 percent

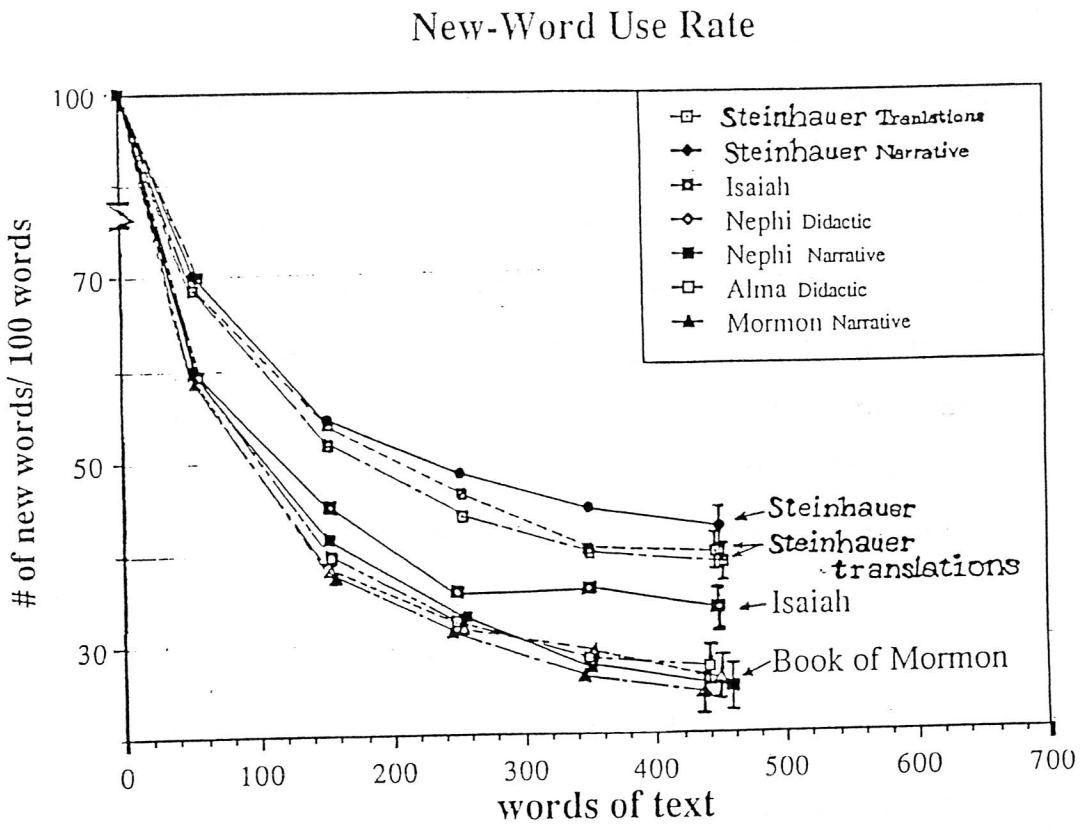
---

4 We selected 3 consecutive 4998 word sections each from the didactic or sermon-like writings of Nephi and Alma, out of their total of 17,982 and 19,382 didactic words repsectively. These texts are taken from the "most primitive" extant sections of the Book of Mormon's handwritten manuscripts. The simple didactic literary form was chosen for two reasons: First for greatest comparability between the largest same-literary-form author pair, and second because it essentially excludes the troublesome phrase "and it came to pass that," the only phrase in the Book of Mormon which repeatedly uses the small "fill words" in a non-independent and non-random sequence.

probability that they are different) low new-word-use rate. This indicates that a consistently small working vocabulary, similar to that of Joseph Smith's, is found throughout the Book of Mormon.

Also shown in Figure 1 are the new-word-use rates of the King James translators' English words for the prophet Isaiah as quoted in the Book of Mormon, and the rates for English writings of Harry Steinhauer, an American man of letters. The average of Steinhauer's personal narrative writings are shown by the highest curve, which demonstrates that he writes with a vocabularly so large that nearly half of the words he uses at the end of his average 500 word block had not been used in the preceding 500 words he had written. The remaining two curves of Figure 1 are found close together immediately below Steinhauer's personal English writings. These are calculated from two of Steinhauer's 1977 academic German to English translations of semi-classical German works. They were originally written in German by Christoph M. Wieland (1733-1813) and Heinrich von Kleist (1777-1811). As can be seen there is no measurable statistical difference (i.e. $p<.05$) in Steinhauer's English new-word-use rates between his translations of these German works. In both of his translations he used a very large new-word-use rate, barely lower than the rate he normally used in his own free English writings.

Figure 1



New-word-use rate, which is a measurement of working vocabulary, is a different matter than stylometric wordprints. Thus, we turn attention to wordprints in the Book of Mormon. Since there are several ways Joseph Smith could have carried out his translation process, there is no reason to say whether Joseph Smith's own wordprint should or should not be found throughout the multi-authored, much abridged English Book of Mormon.

As explained below, however, these preliminary tests strongly show that Joseph Smith's own wordprints are not measurable in any of the tested sections of the Book of Mormon. Furthermore, the

measurements from the extensive didactic writings of Nephi and Alma, while relatively consistent within themselves, differ from each other, as the writings of different control authors differ from one another. No longer can an informed person honestly speculate that the Book of Mormon is nothing but the modern writings of Joseph Smith (or Solomon Spaulding, or Oliver Cowdery) in particular, or of any other single individual in general. Objective wordprint measurements overwhelmingly demonstrate that it is not.

Displayed in Figure 2 are bar graphs showing the results of tests of 325 paired texts. These tests compare non-controversial 5000 word writings of known authorship. These texts were taken from a broad cross section of different control authors and texts. Of these tests, 33 compared a text by one author with another text by the same author, while 292 of the tests compared one author's writing against another's. The 33 "within-author" measurements peak at about only 2 "rejections,"[5] as is theoretically to be expected, since the texts being compared in these cases are known to have been written by the same author. In taking these measurements, we typically find slightly over 40 (i.e., between 40 to 47) of Morton's 65 testable word patterns present in the texts often enough to yield statistically valid results (95 percent statistical confidence). We therefore expect (and have measured) that true "within-author" comparisons show an

---

5 A "rejection" results whenever the two texts being compared measure a null-hypothesis rejection for any one of Morton's word patterns that are used often enough to be valid at a statistically significant rate ($p<.05$).

average "rejections" rate at slightly over 5 percent (i.e., .05 x 40 = 2).

The 293 "between-author" tests plotted in Figure 2, however, show the average or peak at about 7 "rejections,"[6] which is higher than the extreme "within-author" cases. Therefore any other wordprint measurement comparing two texts which are comparable to any of the control author's works that measures 7 or more "rejections" is very likely to have <u>not</u> been written by the same author. The lower the number of rejections, the greater the likelihood that the two texts were written by the same author; the higher the number of "rejections," the more likely the texts are from different authors. Unfortunately, measured "rejections" in the 1 to 6 range cannot be used to assign authorship unambiguously, because of the overlap between the "within-author" and "between-author" distributions in this range. For tests measuring zero "rejections" there is a high probability that the compared texts were written by the same author. Figure 3 illustrates these "rejection" ranges.

---

6 The confidence of separation for any single "between-author" test that measures a total of 7 to 10 "rejections" calculated against the full "within-author" distribution is obtained by a one tailed Student "t" test from $\bar{x}$=2.58, s=1.60, df=32 as:
7 "rejections" (t=2.76) giving >99.5% confidence that the two texts were written by different authors,
8 "rejections" (t=3.39) giving >99.9% confidence that the two texts were written by different authors,
9 "rejections" (t=4.02) giving >99.99% confidence that the two text were written by different authors,
10 "rejections (t=4.64) giving > 99.997% confidence that the two texts were written by different authors.

Figures 2 and 3

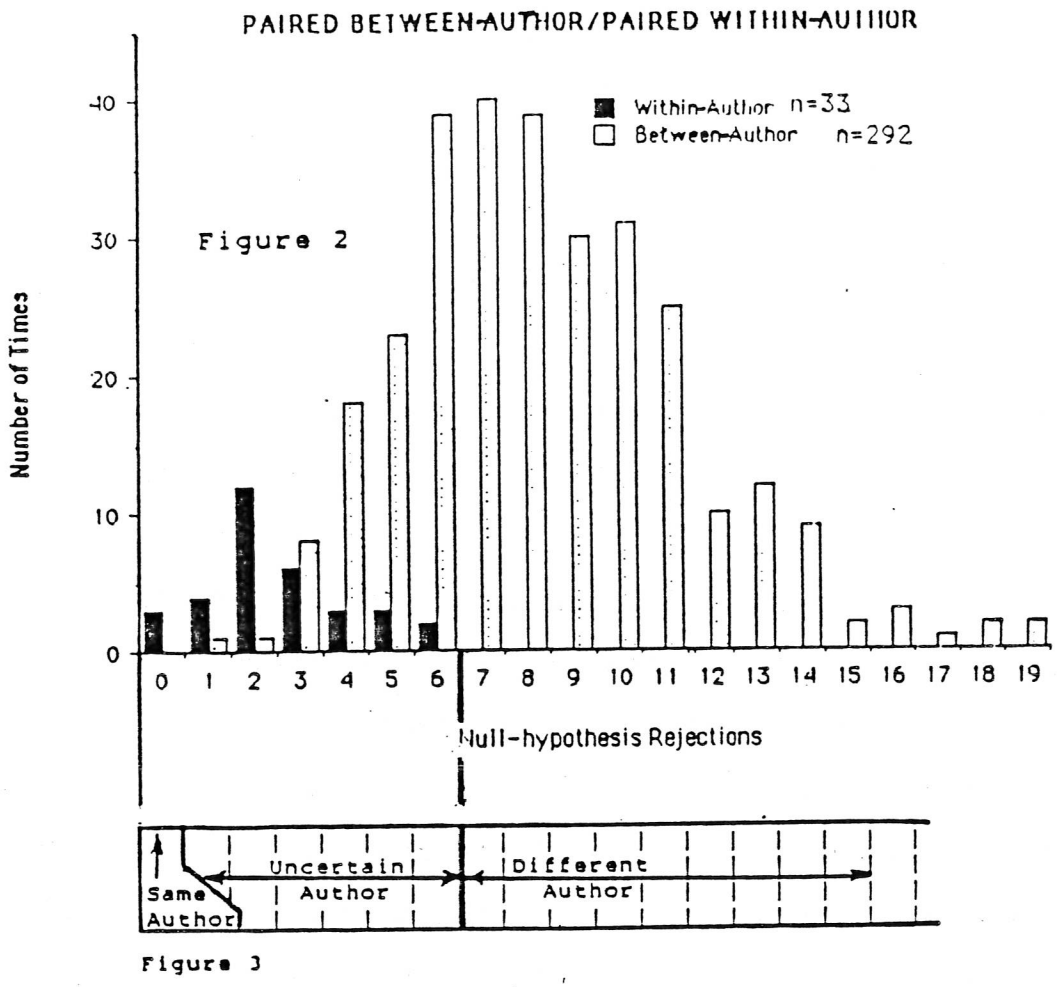


Figure 3

Figure 4 shows the distribution of the wordprints for the six comparisons of the independent 5000 word "within-author" tests for the purported Book of Mormon authors of Nephi and Alma. These tests were calculated in an identical manner to those of the control authors shown in Figure 2. When each was tested against himself, Nephi and Alma showed the same somewhat scattered distribution seen for the "within-author" control tests of Figure 2, ranging from 0 to at most 5 "rejections," peaking at 2. Similarly, the "within-author" proof tests of Figure 4 show a tight internal consistency between the two Oliver Cowdery, two Solomon Spaulding, and three Joseph Smith texts.

Figure 4

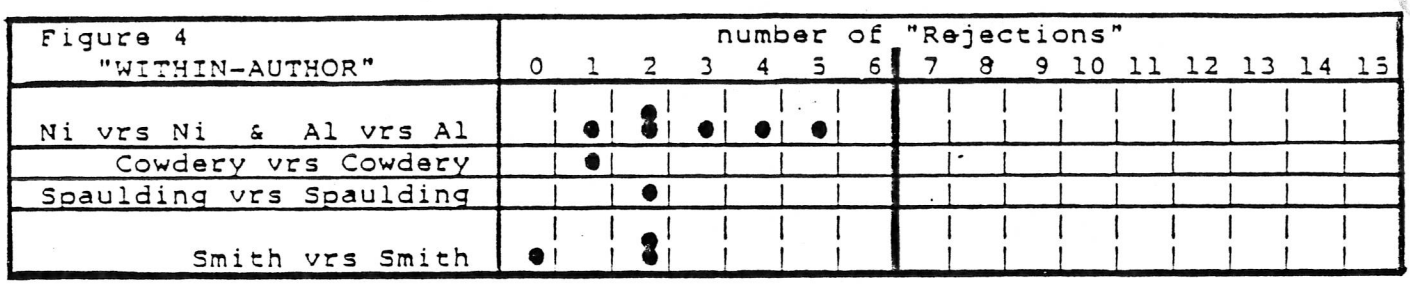


Figure 5 then shows the results from the "between-author" tests of direct interest to the Book of Mormon authorship question. As can be seen, the nine comparisons of Nephi against Alma show the same relatively large number of "rejections" found in the "between-author" distribution from the control authors of Figure 2. In 8 of these 9 tests, 5 or more "rejections" resulted. Four of these tests produced 7, 8, 9, and 10 "rejections." Taken individually, these four high "rejections" tests independently measure a statistical confidence of greater than 99.5%, 99.9%, 99.99%, and 99.997% that these texts from the books of Nephi and Alma were written by different authors. Furthermore, the probability of simultaneously measuring all four of these high "rejections" tests, and yet have all Nephi and all Alma texts written by the same writer, can be calculated by multiplying together each of the four individual same-author probabilities. The same-author probability is equal to one minus the probability for different authorships (reported above), which therefore is .005, .001, .0001, and .00003 respectively. Thus, the overall probability that the same author wrote all these texts, which is computed by multiplying these four measurements

together, is vanishingly small, at approximately 1.3×10^{-13}. This simultaneous calculation is an enormous statistical overkill, demonstrating statistical independence between the didactic writings of Nephi and Alma. In Figure 5 we also see the increasingly higher statistical improbability that these sections of the Book of Mormon were written by Joseph Smith, Oliver Cowdery, or Solomon Spaulding. Table 1 shows the individual "wrap-around" wordprint tests used in producing Figure 5.

Figure 5 and Table 1

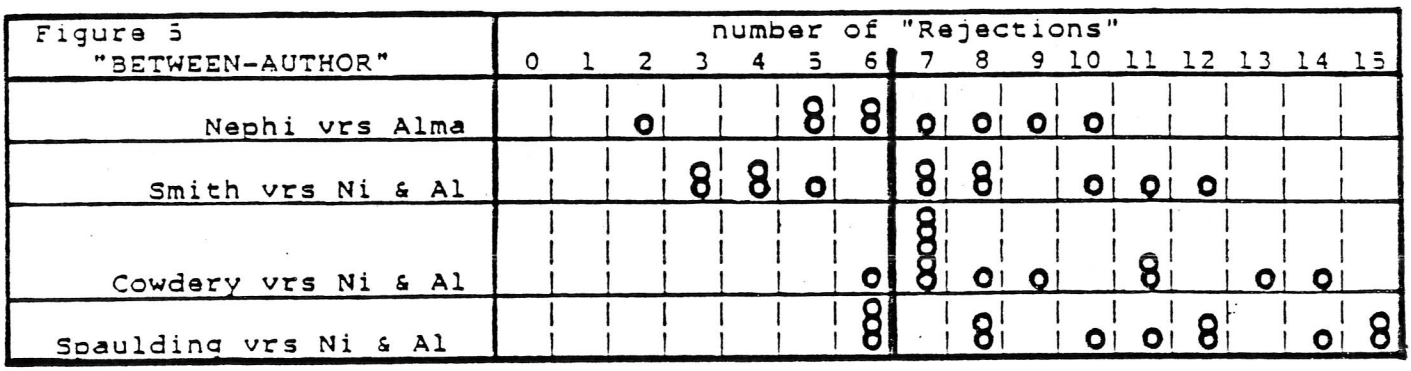


TABLE 1

The Number of Independent Wordprint Tests Tabulated Under the Number of "Rejections" Measured for Each Test. Tests Were Calculated Between All Possible Pairs of Three Each Book of Mormon Texts Attributed to Nephi and Alma Along with Intercomparisons with Seven Other Non-contested Texts Authored by Known Writers of Interest.

TEXT vrs TEXT	NUMBER of "REJECTIONS" 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	NO. OF TESTS
Nephi vrs Nephi			1		1	1											3
Alma vrs Alma		1	1	1													3
Smith vrs Smith	1		2														3
Cowdery vrs Cowdery		1															1
Spaulding vs Spaulding			1														1
Nephi vrs Alma			1			2	2	1	1	1	1						9
Smith vrs Nephi					1				2		1	1	1				6
Smith vrs Alma				2	1	1		2									6
Cowdery vrs Nephi							1	1				2		1	1		6
Cowdery vrs Alma								4	1	1							6
Spaulding vrs Nephi											1	1	1		1	2	6
Spaulding vrs Alma							3		2				1				6

In conclusion, the proposition that Joseph Smith, or Oliver Cowdery, or Solomon Spaulding was the author of the Book of Mormon didactic writings of Nephi or Alma is statistically indefensible. The writings of Nephi and Alma are statistically independent of each other, much as are the writings of our independent non-contested control authors. However, the rate at which new words are introduced in the Book of Mormon texts consistently measures a low single value through out the whole book. A simple (if not the simplest) consistent explanation for these two objectively measured phenomena is that the Book of Mormon is a continuous literal translation of non-English writings by different original authors, expressed by a literal translator using a restricted English vocabulary.