

Did C. S. Lewis write *The Dark Tower*?: An Examination of the Small-Sample Properties of the Thisted-Efron Tests of Authorship

Jeffrey R. Thompson¹ and John Rasp²

¹Dept. of Statistics, North Carolina State University, USA

²Dept. of Decision and Information Sciences, Stetson University, USA

Abstract: *The Dark Tower* is a fragment of a science fiction novel, attributed to C.S. Lewis and published posthumously. Shortly after its publication controversy arose, questioning the work's provenance and authenticity. This controversy still continues. We apply and extend procedures developed by Thisted and Efron (1987) to investigate whether word usage in *The Dark Tower* is similar to that in *Out of the Silent Planet* and *Perelandra*, two works of the same genre and period known to be by Lewis. We further examine the validity and limitations of these procedures in the case at hand. Our results show vocabulary usage in *The Dark Tower* differs from that predicted by the baseline Lewis works.

Zusammenfassung: *The Dark Tower* ist ein Fragment eines Science-Fiction-Romans, das C. S. Lewis zugeschrieben ist und nach dessen Tod veröffentlicht wurde. Kurz nach seiner Publikation kam es zur Kontroverse und man stellte die Herkunft des Buches und die Echtheit in Frage. Diese Kontroverse dauert noch immer an. Wir wenden an und erweitern Verfahren die von Thisted and Efron (1987) entwickelt wurden um zu untersuchen, ob der Wortgebrauch im *The Dark Tower* dem in *Out of the Silent Planet* and *Perelandra* ähnelt, zwei Bücher des gleichen Genre und Zeitraums und bekannterweise von Lewis geschrieben. Wir überprüfen weiters die Gültigkeit und die Beschränktheit dieser Verfahren für den vorliegenden Fall. Unsere Resultate zeigen, dass sich der Wortschatz in *The Dark Tower* von dem durch die Lewis Bücher vorausgesagten unterscheidet.

Keywords: Stylometry, Word Count, Poisson Regression.

1 Introduction

The modern British author C. S. Lewis wrote numerous works on a wide variety of topics, ranging from children's stories to Christian apologetics to literary criticism. Among his books is a trilogy of science fiction novels: *Out of the Silent Planet*, *Perelandra*, and *That Hideous Strength*. After his death, a manuscript was discovered and subsequently published as *The Dark Tower*. The work is a fragment of a science fiction novel, apparently left incomplete, as the extant manuscript breaks off mid-sentence after 64 pages and is missing two leaves. It was purported, based on internal evidence, to have been apparently intended to appear as the sequel to the first book of the trilogy. There has been, however, some question regarding Lewis' authorship of *The Dark Tower*. Lindskoog (1988) has

suggested that the work is spurious - possibly an attempt at a sequel by an unknown author, possibly even a forgery on the part of Lewis' literary executors. Her book has had a mixed reception, some reviewers being intrigued by its claims (Tyson, 1990) while others reject them out of hand (Barker, 1990). She has continued her argument of non-Lewis authorship in a more recent book (Lindskoog, 2001). On the other hand, Poe (2007) recounts anecdotal evidence that the work is genuine, albeit a low-quality unrevised draft. The work's provenance continues to be controversial.

This paper will investigate the issue further, using statistical techniques for studying questions of literary authorship. In particular, we apply and extend procedures due to Efron and Thisted (1976). Their work is based on an earlier paper by Fisher, Corbet, and Williams (1943), which modeled the relationship between the number of species and the number of individuals in a random sample of an animal population, using Malayan butterflies as an illustrative example. Distinct "species" in Fisher, Corbet, and Williams become distinct word types in Efron and Thisted's application. This work was later extended (Thisted and Efron, 1987) to investigate whether Shakespeare wrote a newly discovered poem (the "Taylor poem," after its discoverer) that had been attributed to him. These procedures have been used to investigate other authorship controversies as well. (See, for example, Elliot and Valenza, 1996, who employ these and other tests in an examination of Shakespearean authorship.)

The Thisted and Efron tests show great promise as tools for authorship identification. Yet their use can also be problematic. They are computationally complex, and potentially computationally unstable. They were developed in a context where the authenticating canon is much larger than the work in question, and have difficulties in mathematical convergence when the text being studied is relatively large. Furthermore, relatively little has been done in the way of validation studies. Hence this research also explores and extends application of these techniques.

Section 2 of this paper outlines the principles and mechanics of the Thisted-Efron tests, and introduces a new "uniformity" test for authorship. Section 3 then examines implementation and validity of these procedures in the case at hand. Section 4 presents results from the Lewis authorship tests, and Section 5 draws overall conclusions.

2 Thisted and Efron Procedures

Denote by t the size of the literary work in question, as a fraction of the canon to which it is compared, and by n_x the number of distinct words appearing exactly x times in the canon. Let ν_x be the expected number of word types in the new work which were found exactly x times in the canon, with m_x representing the number actually observed. Thisted and Efron (1987) (henceforth *TE*) show that an unbiased estimator of ν_x is

$$\hat{\nu}_x = \sum_{k=1}^{\infty} (-1)^{k+1} C_k^{x+k} t^k n_{x+k}. \quad (1)$$

In particular, for $x = 0$ (corresponding to the number of words in the new work that did not appear in the canon) this reduces to

$$\hat{\nu}_0 = n_1 t - n_2 t^2 + n_3 t^3 - \dots. \quad (2)$$

Estimation is feasible when the series converges. The sequences used in the *TE* estimators do not necessarily converge, but will do so for sufficiently small values of t . This requirement places a restriction on the usefulness of these procedures; one contribution of this research is to explore the severity of this restriction.

From these basic relationships *TE* suggest three tests for authorship: one based upon the number of new words observed, one based upon the observed number of rare words, and a slope test that uses Poisson regression to combine data. To these we add a fourth: a uniformity test of the various p-values. In each case, the test is of a null hypothesis that the vocabulary usage in the text in question is consistent with that in the canon of known authorship.

The “new words” test compares the observed number of words in the sample which had not appeared in the canon, m_0 , with the number predicted from the *TE*, $\hat{\nu}_0$. Since word frequencies are modeled by a mixed Poisson process, this amounts to testing whether the observed value m_0 is from a Poisson distribution with parameter $\lambda = \hat{\nu}_0$.

The procedure also gives predicted values for the number of words in the sample appearing $x = 1, 2, \dots$ times in the canon. In theory, x is limited only by the size of the canon (or, more precisely, the frequency of its most common word). In practice, however, the relatively small magnitude of the $\hat{\nu}_x$'s when x is large, can make direct use of them problematic. Relatively few word types, after all, will be used (for example) forty-two times in a text. Moreover, as x increases, there tends to be greater instability mathematically in the convergence of the alternating series estimating the $\hat{\nu}_x$'s. Consequently, it makes sense to focus not on the individual $\hat{\nu}_x$ values, but upon them collectively. It is this idea of looking at the $\hat{\nu}_x$'s as a whole that leads to the other three tests.

The “rare words” test totals the number of word types appearing R or fewer times in the canon,

$$m_+ = \sum_{i=1}^R m_i \quad (3)$$

and compares this with the expected value

$$\hat{\nu}_+ = \sum_{i=1}^R \hat{\nu}_i. \quad (4)$$

Note that the choice of R here is somewhat arbitrary, constrained only by the researcher's judgment and potential convergence problems of the estimators. *TE* used $R = 99$; with our smaller canon we employ $R = 40$. The “rare words” test then involves testing whether the observed m_+ is from a Poisson distribution with parameter $\lambda = \hat{\nu}_+$.

The “slope” test takes a different approach to combining the data. For $x = 1, 2, \dots, R$, *TE* modeled the m_x as having independent Poisson distributions with mean

$$\mu_x = \hat{\nu}_x e^{\beta_0} (x + 1)^{\beta_1}. \quad (5)$$

The null hypothesis $H_0 : \beta_0 = \beta_1 = 0$ corresponds with consistency between sample and canon. *TE* focus on testing β_1 , the slope of the log-linear model relating $\log \mu_x$ and $\log \hat{\nu}_x$. This involves a maximum likelihood test, computational details of which are presented in an Appendix.

We introduce a “uniformity” test as a third way to aggregate results. Under the null hypothesis the m_x 's (for $x = 1, 2, \dots, R$) should follow Poisson distributions with parameters $\hat{\nu}_x$. P-values can thus be computed in the usual way. The p-values of these R tests, under the null hypothesis, should be distributed uniformly on the interval $(0, 1)$. This suggests use of a distributional test. While the Kolmogorov-Smirnov test is the best known of these, a procedure due to Anderson and Darling (1954) is generally more powerful (Shapiro, Wilk, and Chen, 1968). The Anderson-Darling test statistic can be used to test the hypothesis that the data are sampled from a population following any specified continuous distribution (normal, uniform, etc.). The statistic for testing whether the R p-values here are from a uniform distribution is

$$AD = -R - \left[\sum_{i=1}^R (2i - 1)(\log p_i + \log(1 - p_{R-i+1})) \right] / R. \quad (6)$$

Test statistics of 2.492 or greater are significant at the $\alpha = 0.05$ level. More complete tables of critical values of the Anderson-Darling test may be found in Pearson and Hartley (1976).

3 Implementation and Validity Issues

The baseline Lewis canon for this research consists of the first two novels of the trilogy, namely, *Out of the Silent Planet* and *Perelandra*. *Out of the Silent Planet* was first published in 1938. It tells of a protagonist Ransom's travels to, and on, the planet Mars, using these adventures as a vehicle for moral and spiritual commentary. For example, reflecting upon his encounter with a sentient Martian, a “hross,” Ransom notes:

It was only many days later that Ransom discovered how to deal with these sudden losses of confidence. They arose when the rationality of the *hross* tempted you to think of it as a man. Then it became abominable - a man seven feet high, with a snaky body, covered, face and all, with thick black animal hair, and whiskered like a cat. But starting from the other end you had an animal with everything an animal ought to have - glossy coat, liquid eye, sweet breath and whitest teeth - and added to all these, as though Paradise had never been lost and earliest dreams were true, the charm of speech and reason. Nothing could be more disgusting than the one impression; nothing more delightful than the other. It all depended on the point of view.

Perelandra was published in 1943 under the title *Voyage to Venus*, with Ransom traveling to that planet and Lewis again addressing religious themes.

The Dark Tower appeared in 1977, fourteen years after Lewis' death. Published by his literary executors, it is a fragmentary manuscript apparently in his handwriting (or resembling it). Here, time- rather than space- travel is the focus.

“Well,” said Orfieu, “time-travelling clearly means going into the future or the past. Now where will the particles that compose your body be five hundred years hence? They'll be all over the place - some in the earth, some in

plants and animals, and some in the bodies of your descendants, if you have any. Thus, to go to the year 3000 AD means going to a time at which your body doesn't exist; and that means, according to one hypothesis, becoming nothing, and, according to the other, 'becoming a disembodied spirit'."

References in the text suggest a date of writing of 1938. Internal evidence also indicates that the work was intended as a sequel to *Out of the Silent Planet*, a place that was eventually taken by *Perelandra*.

Including the third book of the trilogy (*That Hideous Strength*, published 1946) in the canon would significantly expand the size of the baseline corpus (an important consideration). We chose however, to err on the side of caution by using only the first two books, as they should be the closest in word usage to *The Dark Tower* since, in theory, *The Dark Tower* was meant to be a sequel to the first of those two books. Use of Lewis' prolific writings in other genres as part of the baseline reference was not even considered, as genre is known to play an important role in word usage (see, for example, Valenza, 1991).

There are 55411 words in *Out of the Silent Planet* and 83788 in *Perelandra*, with a total of 139199 word tokens of 9858 types appearing in the two works. Table 1 below tabulates n_x , the number of distinct word types appearing x times in the canon, for $x = 1, 2, \dots, 100$. Table layout follows that in *TE*; its compact structure facilitates presentation of a large amount of data in a space-efficient manner. For example, the third entry in the second row of the table indicates that there were 59 different word types that occurred "10+" plus "+3", or thirteen, times in the text. From the table, then, we note that there were 4704 word types appearing once ("0+" plus "+1") in the canon, 1518 appearing twice, etc. (The table stops with word types occurring 100 times. We note that there were also 172 types which appeared more than 100 times.) Only the running text of the works was analyzed. Specifically, words contained in the titles and chapter divisions of the novels were not included in the word counts. Further, text of acronyms and abbreviations were counted as individual words but numerals were not counted as words.

Table 1: Number of word types appearing x times in the Lewis canon.

x	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
0+:	4704	1518	823	499	360	254	170	128	107	93
10+:	93	78	59	54	50	44	33	33	33	23
20+:	26	21	21	21	31	21	18	14	13	9
30+:	19	13	10	12	8	6	8	7	8	16
40+:	7	9	8	8	7	5	6	5	8	6
50+:	8	4	9	7	4	6	2	2	4	5
60+:	3	5	5	3	2	1	2	6	5	5
70+:	5	3	2	1	6	3	0	0	4	6
80+:	0	3	2	6	2	0	4	2	6	1
90+:	2	2	1	1	1	0	3	4	1	0

For present purposes, "word" is defined as a unique sequence of letters and symbols delimited by blank spaces. This definition is admittedly somewhat simplistic. For example, it considers homographs (e.g., *saw* as a noun and as a verb) to be the same word, and counts hyphenates, which Lewis used frequently (e.g., *chestnut-tree*), as one word.

However it has the advantage of being easy to implement. More important than ease of use is its utter objectivity. No reader coding needs to be done to decide if (for example) “post-office” is one word or two, or whether somewhat different usages of the same letter sequence should count as one or several words. The inherent subjectivity of such decisions would necessarily introduce additional error variance into the analysis. While this definition of a “word” certainly misses some of the nuance of language, there is ample precedent for it in the literature (for example, Morton, 1986; Lana, 1992), and departure from it would create as many problems as it solves. (Other stylometric research does employ a more complex implementation of the concept of a “word”. See, for example, Rottmann, 2006 or Wilson, 2006.)

The procedures here are computationally involved, requiring not only collation of a large corpus of text but also calculation of some mathematically complex quantities. However, advances in computer technology have made this problem much more tractable. (All computations for this paper were done using a standard spreadsheet program.)

A more telling difficulty is with the mathematical convergence of the TE estimation procedures. Convergence only happens, for the series in question, for very small values of t , that is, when the new work in question is much smaller than the established canon. This was not a problem for TE , as the known Shakespeare corpus is huge (884647 total words) when compared with the length of the Taylor poem (429 words). It does become problematic in the present case. As discussed above, we have conservatively used a relatively small base canon (two books totaling 139199 words) for this research. *The Dark Tower*, at 26702 words, is long as compared with this baseline.

Thus our first work on the problem must address the issue of how large a sample size (n) may be taken from *The Dark Tower* without causing problems with convergence. We further must address the validity of the tests in this context. It has been established (Valenza, 1991) that they work well in Shakespearean usage, both with poems and with plays, but do not perform well between genres. However, no validation studies have been done on works of modern fiction. Hence we examine validity within the known Lewis works.

To this end, twenty-five random samples of $n = 70$ words were selected from both *Out of the Silent Planet* and *Perelandra* - fifty total samples. (A sample of 70 words out of the 139199 total in the two books yields $t = 70/139199 \approx 0.000503$, essentially the same as the $t = 429/884647 \approx 0.000485$ used in the original Efron and Thisted paper.) For each sample, word counts were determined and compared with those predicted by the TE procedure, using as a baseline the balance of the Lewis canon (that is, everything in the two books except the 70-word sample in question).

We would expect the procedure to work well on 70-word samples, as this gives a value of t that is known to work, from the original Efron and Thisted paper. We would like to determine, however, whether the method is feasible with larger sample sizes. Accordingly, this process was repeated for fifty samples of $n = 200$ words, and again for fifty samples of $n = 1000$ words. Results from all these tests are given in Table 2 below.

We would expect, in this situation, that the null hypothesis would be rejected only about as often as the significance level of the test. As can clearly be seen, samples of $n = 200$ and $n = 1000$ produced a sizable number of rejections (thirty to forty percent of the time), far more than would be expected at the $\alpha = 0.05$ and $\alpha = 0.10$ level. Convergence

Table 2: Validation tests: Number (and percentage) of rejections on “New Words” and “Rare Words” tests confirming Lewis authorship of 50 samples of text.

Significance Level	$n = 70$ $t = 0.000503$		$n = 200$ $t = 0.001437$		$n = 1000$ $t = 0.007184$	
	New Words	Rare Words	New Words	Rare Words	New Words	Rare Words
$p < 0.05$	3 (6%)	8 (16%)	17 (34%)	13 (26%)	17 (34%)	20 (40%)
$p < 0.10$	14 (28%)	14 (28%)	21 (42%)	17 (34%)	21 (42%)	24 (48%)

problems with the estimators when the sample size is this large, relative to the canon, are the culprit here. Therefore the usefulness of the *TE* tests for moderately large t is called into question. Given the convergence problems with larger values, samples of $n = 70$ words are used for the remainder of the paper. With samples of $n = 70$, the “new words” test rejects the null hypothesis six percent of the time, approximately what would be anticipated by the significance level of five percent. However, the number of rejections for the rare words test is noticeably higher than expected. This indicates that the *TE* “new words” test is reliable in the present context - relatively small samples of words from a work of modern fiction. It does, however, indicate potential problems with use of the *TE* “rare words” test.

A related issue is the tests’ ability to distinguish Lewis’ works from similar writings by another author. To investigate this, we use the books of George MacDonald, a Victorian era author of children’s fantasy novels. He was chosen because Lewis viewed him as a literary mentor, and because electronic editions of his works are readily available. Four MacDonald novels were selected: *At the Back of the North Wind*, *Phantastes*, *The Princess and Curdie*, and *The Princess and the Goblin*, all obtained through Project Gutenberg’s library of books in electronic format (<http://www.gutenberg.net> or <http://www.promo.net/pg/>).

Having established previously that 70-word sections serve as appropriate units of analysis, we employ them in this analysis as well. Fifteen 70-word blocks were selected at random from each of the four MacDonald books. We then compare word counts for these 60 samples to that predicted by the baseline Lewis corpus. We would expect to be able to reject the null hypothesis of Lewis authorship, thus validating the *TE* methodology in this context.

Here the “new words” test counts the number of words occurring in the 70-word sample which had not appeared in the Lewis canon, and compares this with the number that would be anticipated from the *TE* model. The “rare words” test similarly counts the total number of words occurring in the 70-word sample which had appeared between one and forty times in the Lewis canon, comparing this with that predicted by *TE*. These forty observed and expected word counts are used in the procedures outlined in the Appendix to give “slope” tests for authorship. These three tests are performed on each of the 60 samples from MacDonald’s writings, with results given in Table 3, below.

Note that the “new words” test rejects the hypothesis of Lewis authorship at the $\alpha = 0.05$ level in fifteen of the 60 samples (25%), and at the $\alpha = 0.10$ level in 24 of the samples (40%), indicating moderate, though not overwhelming power for this procedure.

Table 3: Power confirmation tests: Number (and percentage) of rejections on various tests comparing 60 samples of text by G. MacDonald with known C. S. Lewis works.

Significance Level	New Words Test	Rare Words Test	Slope Test
$p < 0.05$	15 (25%)	6 (10%)	8 (13%)
$p < 0.10$	24 (40%)	15 (25%)	14 (23%)

The “rare words” test is rather less effective – able to reject Lewis authorship at $\alpha = 0.05$ in only six of the 60 samples (10%) and at the $\alpha = 0.10$ level only 15 of 60 times (25%). These numbers are scarcely different from (even slightly lower than) those from Table 2, where the test was validated. Again, these results indicate potential problems with the “rare words” procedure. The “slope” test appears to have moderate power in this context, successfully rejecting eight of the 60 samples (13%) at the $\alpha = 0.05$ level and an additional six samples (23%) at $\alpha = 0.10$. Admittedly none of these percentages is overly large. These results do indicate, however, that the *TE* procedures, and in particular the “new words” test, can be moderately successful in authorship discrimination, even with very small samples (70 words) of text.

The 60 samples of MacDonald text produce 60 sets of observed frequencies of words appearing $i = 1, \dots, 40$ times in the canon. Most of these values will be quite small. Therefore, to enhance the mathematical stability of the procedure we sum across the 60 samples to produce a single total set of observed frequencies of words appearing 1 through 40 times in the canon. These observed totals can be compared with the expected word frequencies, again using the standard Poisson distribution test. The forty resulting p-values should be uniformly distributed. The uniformity test produces an Anderson-Darling statistic of 4.809. The large magnitude of this statistic, with its corresponding low p-value, indicates that the uniformity test is effective in differentiating authorship here.

Taken together, these results show that the *TE* procedures require a canon much larger than the work under investigation – that is, a small value of t . For present purposes, samples of 70 words seem most reliable and create no convergence problems. In general, it appears that values of t much larger than 0.0005 lead to difficulties. Usage of the *TE* “rare words” test appears problematic, as the procedure showed scant discriminatory ability. However, results indicate that the *TE* “new words” and (to a lesser extent) “slope” tests are valid in distinguishing between authors of modern fiction, and suggest that word samples as small as $n = 70$ produce tests of moderate power in this context.

4 Results

Having thus validated the procedures to be used, we turn to the central issue: investigating whether C. S. Lewis wrote *The Dark Tower* by comparing word usage there with works known to be by Lewis.

Each of the seven chapters of *The Dark Tower* was divided into successive 70-word blocks of text, giving a total of 379 sample passages. Observed word frequencies were

counted for each sample. Expected frequencies were computed, using the procedures outlined in Section 2 and the word counts from the Lewis canon given in Table 1. For example, the expected number of words appearing in the 70-word section which had previously not appeared in the canon is (for $t = 70/139199$)

$$\hat{\nu}_0 = n_1t - n_2t^2 + n_3t^3 - \dots = 4704t - 1518t^2 + 823t^3 - \dots = 2.365. \quad (7)$$

Comparison of observed with expected word frequencies serves to test the hypothesis of Lewis authorship, as detailed below.

Under the null hypothesis of Lewis authorship of *The Dark Tower*, from the *TE* model we would expect a new 70-word sample of text to contain 2.365 new words, that is, words not appearing previously in the canon. For each of the 379 (seventy-word) samples we counted the number of observed new words. A test of Lewis authorship is then a test of $H_0 : \lambda = 2.365$, that is, that the mean of the observed Poisson process for each sample is indeed that predicted from the Lewis canon. P-values for the test are computed directly from the Poisson probability distribution. For example, the very first sample observed ten new words. Our p-value is thus $\Pr\{\text{Poisson}(2.365) \geq 10\} \approx 0.000109$. (As a continuity correction only half of the atom of the probability at $m_0 = 10$ is counted). Results from these tests are summarized in Table 4.

Table 4: Test results: Number (and percentage) of rejections on various tests comparing 379 seventy-word sections of *The Dark Tower* with Lewis canon.

Significance Level	New Words Test	Rare Words Test	Slope Test
$p < 0.05$	110 (29%)	45 (12%)	93 (25%)
$p < 0.10$	172 (45%)	75 (20%)	123 (32%)

The number of new words observed in the 70-word samples ranges from a low of 0 to a high of 15, averaging 4.507 new words per sample. The null hypothesis of Lewis authorship is rejected at the $\alpha = 0.05$ level for 110 of the 379 samples, or 29% of the time. Rejection at the $\alpha = 0.10$ level occurs in 172 (45%) cases.

Similarly we can compute the expected number of “rare” words in each sample. We define “rare” in this context as words in the sample that appeared between one and forty times, inclusive, in the canon. The expected number of rare words per 70-word sample under the *TE* model is 15.53. For each sample, a test of the null hypothesis of Lewis authorship is a test of $H_0 : \lambda = 15.53$. P-values are computed similarly to those from the previous test. These results are likewise summarized in Table 4.

The number of rare words per 70-word samples ranges from 5 to 24, with an average of 14.19. Rejection of the null hypothesis of Lewis authorship at $\alpha = 0.05$ occurs in 45 of the 379 samples (12%); rejection at $\alpha = 0.10$ in a total of 75 samples (20%).

Slope tests for the 379 samples give mixed results. The null hypothesis of Lewis authorship is rejected at the $\alpha = 0.05$ level on 93 of the samples (25%), and at the $\alpha = 0.10$ level in 123 cases (32%).

The 379 separate samples give us 379 independent values for m_1, \dots, m_{40} , the number of observed words appearing 1, \dots , 40 times in the canon. Aggregating these across the

Table 5: Uniformity test: Comparing *The Dark Tower* with the Lewis canon.

Word	Word	Word	Word
Frequency	p-value	Frequency	p-value
1	0.0287	11	0.2923
2	0.0001	12	0.0045
3	0.8005	13	0.0402
4	0.0010	14	0.1224
5	0.2376	15	0.2704
6	0.1722	16	0.0002
7	0.7856	17	0.0000
8	0.2331	18	0.4589
9	0.3572	19	0.8179
10	0.0001	20	0.0000

Anderson-Darling test statistic $AD = 52.47$.

379 samples removes much of the difficulty presented by the small sample size and gives us forty separate observed and expected counts and hence forty p-values. Under the null hypothesis of Lewis authorship, these p-values follow a uniform distribution on $(0, 1)$. The uniformity test is conducted on these data, with results presented in Table 5. For our data we compute $AD = 52.47$ which is significant at the $\ll 0.01$ level.

5 Conclusions

We have observed that the *TE* procedures employed in this paper require t to be small enough to allow convergence – that is, a sample text much smaller than the baseline canon. In the present case, this results in quite small sample sizes and accordingly with moderately low power for tests. This limits the usefulness of the *TE* procedures in resolving some authorship issues. We note, however, that repeating tests of small samples from the work in question can be effective in overcoming this limitation.

We further explored the validity of the *TE* procedures by applying them to a genre other than their initial use, by examining how successful the tests were in distinguishing George MacDonald from C. S. Lewis. Here we saw that the *TE* “rare words” test did not perform well, rejecting Lewis authorship barely more than would be expected by chance alone. The *TE* “slope” test fared somewhat better. The “new words” test was most successful in this regard. The procedure has relatively low power, for samples this small, but is indicated to be useful as an authorship test, even with very short texts.

Ultimately, we saw moderately strong evidence differentiating the vocabulary in *The Dark Tower* from that of *Out of The Silent Planet* and *Perelandra*. While the “rare words” test again showed little discriminatory ability, the “slope” test and, to a greater extent, the “new words” test, rejected the hypothesis of Lewis authorship consistently more than would be expected.

Overall then, it does appear that word usage in *The Dark Tower* is inconsistent with that found in *Out of The Silent Planet* and *Perelandra*.

This may be attributable to the fact that a manuscript in rough draft form (such as *The*

Dark Tower) would inherently display word usage different from a final, polished work or perhaps, as Poe (2007, p. 45) suggests, that *The Dark Tower* was written when Lewis had "... a bad day ... [or simply] ... committed a flawed plotline to paper". However, it is also consistent with the claim that C. S. Lewis did not, in fact, write *The Dark Tower*.

A Computing the Thisted-Efron Slope Test

For a set of observations $\{m_x\}$, $x = 1, \dots, R$, the Poisson likelihood under the *TE* model is

$$L = \prod_{x=1}^R \exp \left\{ -\hat{\nu}_x e^{\beta_0} (x+1)^{\beta_1} \right\} \left[\hat{\nu}_x e^{\beta_0} (x+1)^{\beta_1} \right]^{m_x} / (m_x)! .$$

Here, the m_x 's and $\hat{\nu}_x$'s are known; the maximum likelihood estimates (MLE's) $\hat{\beta}_0$ and $\hat{\beta}_1$ are those values of β_0 and β_1 which maximize L . This is most easily accomplished by taking logarithms of both sides of the equation:

$$\log L = \sum_{x=1}^R \left\{ -\hat{\nu}_x e^{\beta_0} (x+1)^{\beta_1} + m_x [\log \hat{\nu}_x + \beta_0 + \beta_1 \log(x+1)] - \log(m_x!) \right\} .$$

Various numerical techniques exist for solving for the maximum of this function. Modern commercial spreadsheet packages (Excel, Lotus, Quattro Pro) include built-in routines for solving such problems, greatly simplifying the computational task.

Obtaining standard errors for the MLE's $\hat{\beta}_0$ and $\hat{\beta}_1$ involves computing three second partial derivatives:

$$\begin{aligned} A &= \frac{\partial^2 \log L}{\partial \beta_0^2} = \sum_{x=1}^R -\hat{\nu}_x e^{\beta_0} (x+1)^{\beta_1} , \\ B &= \frac{\partial^2 \log L}{\partial \beta_0 \partial \beta_1} = \sum_{x=1}^R -\hat{\nu}_x e^{\beta_0} (x+1)^{\beta_1} \log(x+1) , \\ C &= \frac{\partial^2 \log L}{\partial \beta_1^2} = \sum_{x=1}^R -\hat{\nu}_x e^{\beta_0} (x+1)^{\beta_1} [\log(x+1)]^2 . \end{aligned}$$

Evaluate these quantities at the computed MLE values $\hat{\beta}_0$ and $\hat{\beta}_1$. The standard error of $\hat{\beta}_1$ is then

$$s.e.(\hat{\beta}_1) = \sqrt{-C / (AC - B^2)} .$$

The slope test statistic $z = \hat{\beta}_1 / s.e.(\hat{\beta}_1)$ is asymptotically standard normal.

References

- Anderson, T. W., and Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, 49, 765-769.
- Barker, N. (1990). C. S. Lewis, Darkly. *Essays in Criticism*, 40, 358-367.

- Efron, B., and Thisted, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika*, 63, 437-447.
- Elliot, W. E. Y., and Valenza, R. J. (1996). And then there were none: Winnowing the Shakespeare claimants. *Computers and the Humanities*, 30, 191-245.
- Fisher, R. A., Corbet, A. S., and Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, 12, 42-58.
- Lana, M. (1992). Xenophon's *Athenaion Politeia*: A study by correspondence analysis. *Literary and Linguistic Computing*, 7, 17-26.
- Lindskoog, K. (1988). *The C. S. Lewis Hoax*. Portland, Oregon: Multnomah Press.
- Lindskoog, K. (2001). *Sleuthing C. S. Lewis: More Light in the Shadowlands*. Macon, Georgia: Mercer University Press.
- Morton, A. Q. (1986). Once. A test of authorship based on words which are not repeated in the sample. *Literary and Linguistic Computing*, 1, 1-8.
- Pearson, E. S., and Hartley, H. O. (1976). *Biometrika Tables for Statisticians* (Vol. 2). London: Biometrika Trust.
- Poe, H. L. (2007, February). Shedding light on The Dark Tower: A. C.S. Lewis mystery is solved. *Christianity Today*, 44-45.
- Rottmann, O. A. (2006). Aspects of the typology of Slavic languages. Exemplified on word length. In P. Grzybek (Ed.), *Contributions to the Science of Text and Language. Word Length Studies and Related Issues* (p. 241-258). Dordrecht, The Netherlands: Springer Publishing.
- Shapiro, S. S., Wilk, M. B., and Chen, H. J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63, 1343-1372.
- Thisted, R., and Efron, B. (1987). Did Shakespeare write a newly discovered poem? *Biometrika*, 74, 445-455.
- Tyson, N. J. (1990). The C.S. Lewis Hoax (book review). *Extrapolation*, 31, 284-286.
- Valenza, R. J. (1991). Are the Thisted-Efron authorship test valid? *Computers and the Humanities*, 25, 27-46.
- Wilson, A. (2006). Word-length distribution in present-day lower Sorbian newspaper texts. In P. Grzybek (Ed.), *Contributions to the Science of Text and Language. Word Length Studies and Related Issues* (p. 319-327). Dordrecht, The Netherlands: Springer Publishing.

Authors' Addresses

Jeffrey R. Thompson
 Department of Statistics
 North Carolina State University
 Campus Box 8203
 Raleigh, NC 27695-8203, USA
 E-mail: thompson@stat.ncsu.edu

John Rasp
 Department of Decision and Information
 Sciences
 Stetson University
 Campus Box 8398
 DeLand, FL 32720, USA