Delta Prime?

David L. Hoover
New York University, USA

Abstract

John F. Burrows has proposed Delta, a simple new measure of textual difference, as a tool for authorship attribution, and has shown that it has great potential, especially in attribution problems where the possible authors are numerous and difficult to limit by traditional methods. In tests on prose, Delta has performed nearly as well as for Burrows's verse texts. A series of further tests using automated methods, however, shows that two modified methods of calculating Delta and three alternatives to or transformations of Delta produce results that are even more accurate. Four of these five new measures produce much better results than Delta both on a very diverse group of 104 novels and on a group of forty-four smaller contemporary literary critical texts. Although further testing is needed, Delta and its modifications should prove valuable and effective tools for authorship attribution.

1 Introduction[1]

One of the most interesting developments in authorship attribution in recent years has been John F. Burrows's introduction, in his Busa Award presentation (2001), of Delta, a simple but promising new measure of stylistic and authorial difference. Two recent articles explain and test Delta further (Burrows, 2002, 2003). The new measure is especially welcome because of its success in problems in which traditional authorship attribution methods are not able to limit the possible claimants to a small and manageable group. Like many other statistical measures and techniques used in authorship attribution and statistical stylistics, Delta is based on variations in the frequencies of the most frequent words in a group of texts. In his introduction of the new measure, Burrows uses the frequencies of the 150 most frequent words in verse texts 'by twenty-five poets of the English Restoration period' to calculate Delta (2002, p. 269). These texts form a primary set to which a text of unknown authorship can be compared to determine which, if any, of the authors is likely to have written it. Before any actual authorship problems can be investigated using the new measure, however, it requires testing to demonstrate its effectiveness on texts of known authorship.

Burrows's own initial test of the effectiveness of Delta begins by establishing a word frequency list for the primary set of texts, selecting the 150 most frequent words for investigation and counting the frequen-

Correspondence:
David L. Hoover
Department of English
New York University
19 University Place, 5th Floor
New York, NY 10003
USA.

E-mail:
david.hoover@nyu.edu

1 I would like to thank an anonymous reviewer for correcting some errors and for suggestions that improved the presentation and clarity of my argument.

cies of these words in the works by each of the twenty-five authors in the primary set. He then compares the frequency of each word for each author with its mean frequency in the primary set (2003, p. 10). Because the frequencies decline very rapidly, he converts the difference between the frequency of each word in each text and the mean into a z-score, so that all words contribute equally. The z-scores are calculated by subtracting the mean frequency of the word in the primary set from its frequency in the texts by each author and dividing this difference by the standard deviation of the word in the primary set. The result is a measure, expressed in standard deviations, of how far above or below the mean for the primary set each frequency in the texts of a given author falls. A positive z-score indicates that the word is more frequent than the mean and a negative z-score that the word is less frequent than the mean.

Delta is designed to measure the difference between additional test texts and the texts by the authors of the primary set, and it does so in a simple way: the frequency of each of the 150 most frequent words in the test text is first compared with its mean frequency in the primary set, just as was done with each of the primary texts. The difference between the test text and the mean is then compared with the difference between the texts by each author in the primary set and the mean. For example, assume that A, B, and C are the text samples by three authors in the primary set, that T is a test text, and that the word *the* has the z-scores below for the three texts:

	T	A	B	C
the	1	-1	-2	2

The differences between T and each of the primary samples are then computed as follows:

T - A	T - B	T - C
1 - (-1) = 2	1 - (-2) = 3	1 - (2) = -1

In these examples, it seems clear that the size of the difference is the important fact rather than its sign, which merely indicates whether the z-score of the test text is larger or smaller than the z-score for the author in the main set. For the word *the,* then, text T is less different from sample C than from samples A and B. The final step is to sum the absolute values of the differences between the z-scores for all 150 words and calculate the mean, producing Delta, which Burrows defines as 'the mean of the absolute differences between the z-scores for a set of word-variables in a given text-group and the z-scores for the same set of word-variables in a target text' (2002, p. 271). The primary text that shows the smallest Delta, the smallest mean difference from the test text is 'least unlike' it (Burrows, 2003, p. 15), and the author of that primary text has the best claim, among the authors tested, to be the author of the test text.

Another example will show why taking the absolute values of the differences between the z-scores is such a crucial step in the calculation of Delta. Consider again text T and sample C just discussed and another primary sample D, and assume that the words *the* and *a* have the following z-scores:

	T	C	D	T - C	T - D
the	1	2	0.5	-1	0.5
a	2	1	1.5	1	0.5
Mean Difference:				0	0.5
Mean of Absolute Differences:				1	0.5

It seems obvious that T and D are less different from each other than T and C are from each other: for both *the* and *a*, the frequency in T is closer to that in D than in C. If we simply sum the differences and take their mean, however, T and C show no difference at all.

Delta is a simple unitary measure, but it is very effective when tested on sixteen texts by authors who are members of Burrows's original set and sixteen texts by other authors. In thirty of the thirty-two cases, the Delta scores either correctly identify the author of the test text from among the twenty-five in the primary set or correctly suggest that the author is not among those in the primary set (Burrows, 2003, p. 15). These remarkably accurate results suggest that Delta has great potential in authorship studies with many possible claimants. As I have shown in 'Testing Burrows's Delta', above, Delta is also successful with prose texts, and is even surprisingly resistant to differences in point of view and nationality. Because of Delta's great potential value in authorship attribution, it seems worthwhile to examine it more closely and explore some possible changes in its calculation that may improve its already impressive accuracy.

2 A Baseline Test on Prose

An investigation of the nature and accuracy of Delta, some possible alternatives to Delta, and some transformations of it can best proceed based on a group of texts not too dissimilar in number and size from the one that Burrows uses in his initial presentation. I begin, then, with a group of fifty-six texts, the primary set consisting of samples of pure third-person authorial narration taken from the beginnings of twenty-four novels by twenty-four American authors. These novels, published between 1890 and 1925, are now out of copyright and are available as e-texts (see the Appendix for a list of the texts). After manually removing the dialogue, chapter titles, page numbers, and similar extraneous material, I edited the texts to assure consistent treatment of apostrophes, single quotations, and hyphenation. The samples range in size from about 10,000 to about 39,000 words, with a mean size of about 27,000 words, and the entire primary set consists of slightly more than 650,000 words, roughly 100,000 more than the set upon which Burrows first tested Delta.[2] The secondary set of texts consists of thirty-two similar samples, twenty by authors represented in the primary set and twelve by other

2 All word counts presented here must be considered approximate. Although intuitively *word* seems a simple concept, there is no absolutely 'correct' number of words in a text of any significant length. The number of words depends on a series of decisions about what counts as a word. Different text-analysis programs and different analysts (to say nothing of different word-processing programs) produce different counts. There is no harm in this, as long as words are counted in a consistent way across all of the texts in any one analysis.

authors. The mean size of these samples is about 26,000 words, and the secondary set contains about 840,000 words (see Appendix).

I placed the word-frequency data collected from these texts in an Excel spreadsheet similar to that used by Burrows, but expanded so that it includes the 800 rather than the 150 most frequent words and automated so that a macro calculates and records Delta for various numbers of frequent words, analyzes the results for accuracy, and collects the results for graphing (the spreadsheet and macros are available upon request). As I have shown in 'Testing Burrows's Delta', Delta is usually more accurate when far more than the 150 most frequent words are included, and is also more accurate when personal pronouns and any words for which a single text supplies more that 70% of the occurrences are removed. When contractions are removed, the largest number of correct attributions is sometimes slightly higher. Because the total number of correct attributions is almost always greater when contractions are not deleted, however, the analyses below will retain them. As Burrows found, Delta is quite effective in attributing texts to their correct authors and in suggesting that the true author is not among those in the primary set. For the texts mentioned above, Delta produces the results shown in Fig. 1 (no personal pronouns, culled at 70%). In this case, the best results, correctly attributing eighteen of the twenty texts by authors in the primary set, are achieved when the 100 and 300 most frequent words are included, with the former slightly more accurate overall. The pattern of texts by members of the primary set of authors and by other authors, shown in Fig. 2, is not quite as striking as the one Burrows achieved for Restoration poetry, and there are two errors compared to only one in Burrows's analysis.[3] Yet Delta is still remarkably effective in this difficult open test, failing for at most four of the thirty-two secondary texts.

3 Why Z-scores?

It is difficult to argue with success, but it seems worthwhile to consider the nature of Delta more carefully. First, one might wonder whether z-scores are the only or most appropriate measure of difference. Clearly some method of neutralizing the effect of the very rapid decline in frequency as we move down the word frequency hierarchy from the most frequent word to the 800th most frequent is needed, so that the words with lower frequencies can have an effect. One intuitively reasonable alternative to z-scores is the percentage difference between the mean and the test text. If we replace the z-scores with the percentage difference and calculate Delta as before, however, the results are slightly less accurate: percentage change, like Delta, correctly attributes a maximum of eighteen of the twenty texts by members of the primary group (twice), but it yields fewer correct attributions overall when the 100–800 most frequent words are analyzed. Perhaps this is because, unlike the percentage of change, z-scores take into account how variable the frequency of each word is in the primary texts, so that a large difference for a word that does not vary much overall counts more than the same difference for a word that shows

3 This graph is based on the z-scores for Delta, which show how far the Delta score for each author is above or below the mean of all the Delta scores. I will refer to the z-score of Delta as 'Delta-Z' below, and ignore its sign, referring to a Delta-Z of -3 as 'larger' than a Delta-Z of 1 and 'smaller' than a Delta-Z of -4.

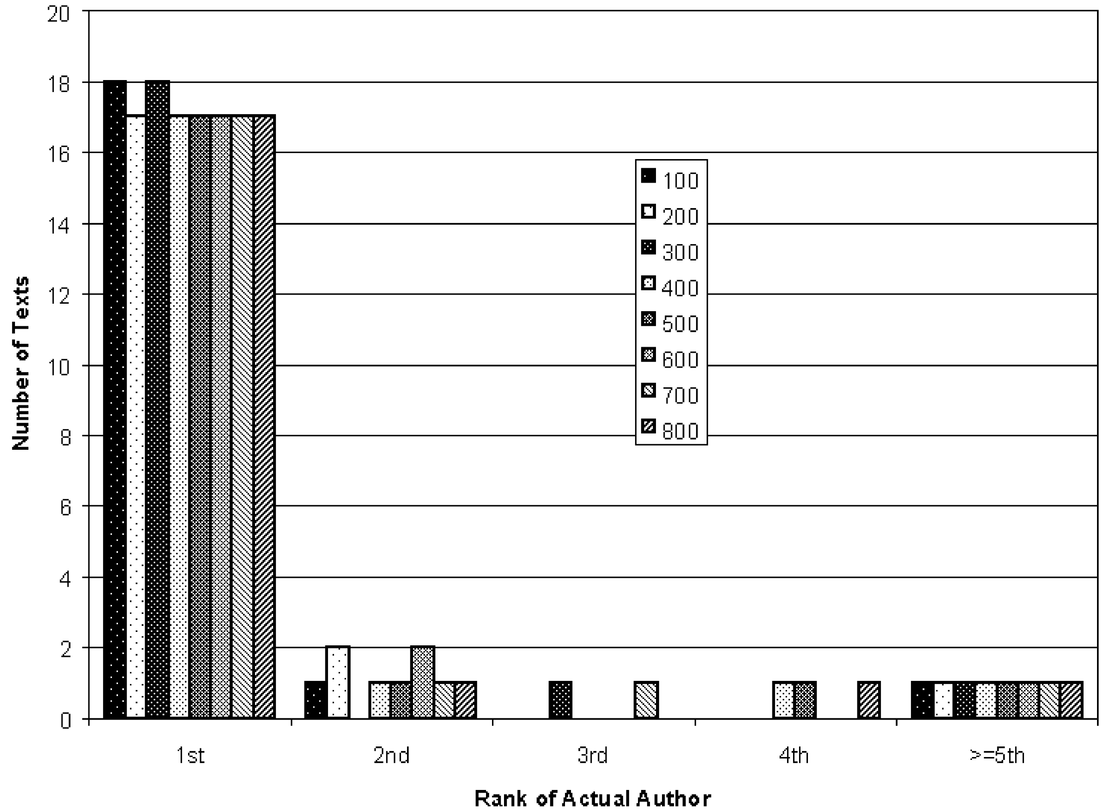


Fig. 1 Results of Delta tests on twenty novels.

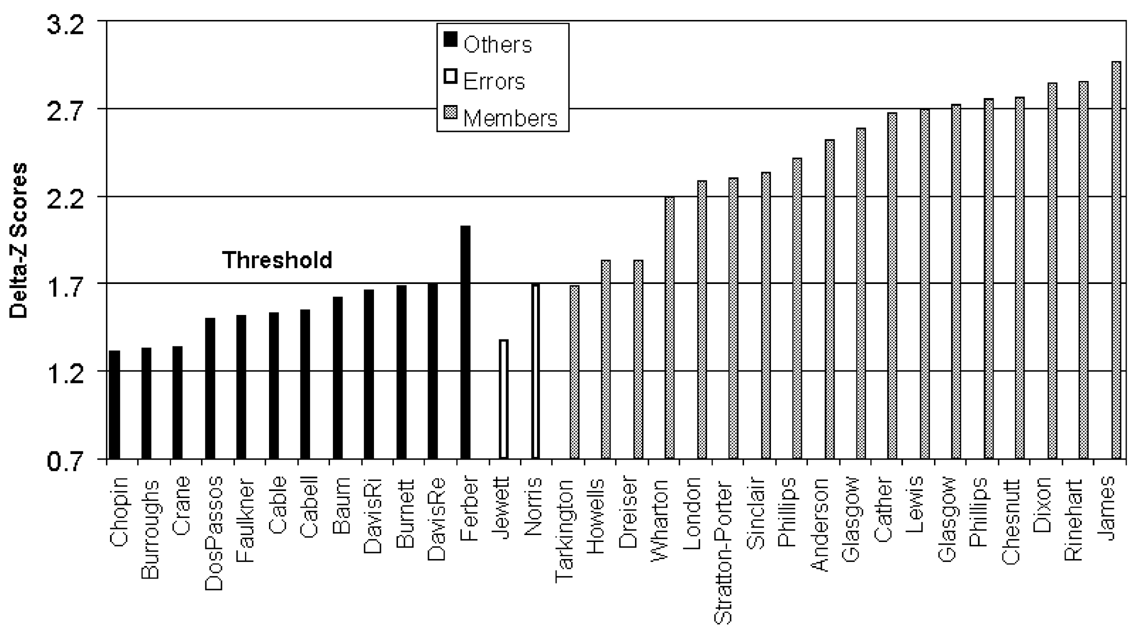


Fig. 2 Thirty-two novels: others, errors, and members (based on the 100 most frequent words).

a great deal of variation among the primary texts. Here, as with the use of the absolute value of the difference, Burrows's decision seems both judicious and effective.

4 Which Z-scores?

Consider next the z-scores themselves. As we have seen, they measure how many standard deviations above or below the mean for the primary texts the frequency of each word in the test text falls. Burrows reasonably collects all such differences from the mean, and then compares them with the differences for the other texts. There seems no a priori reason, however, to assume that words with negative and positive z-scores in the test text are equally important indicators of difference. Once the calculation of Delta has been automated, it is a simple matter to alter the spreadsheet so as to test this question in a straightforward manner. First, we can limit Delta so that it takes into account only those words for which the test text has a positive z-score, concentrating the analysis on words that are more frequent than the mean in the test text. For this test, I alter the formula for the absolute difference between the z-score of the test text and the primary text so that it calculates the absolute difference only if the z-score of the test text is greater than or equal to zero—otherwise, it inserts a zero.[4] Because the formula for Delta remains the same, the total difference in cases where the test z-score is positive is averaged over the entire range of words being analyzed (in this case, the analysis produces identical results when the mean of the non-zero items is used instead of the mean of all words in the analysis). The best result is seventeen correct attributions, compared to eighteen for Delta, and the results are poorer overall. If the formula is altered so that the absolute difference is calculated only when the test z-score is negative and a zero is entered otherwise, concentrating the analysis on words that are less frequent than the mean in the test text, there is one result of nineteen correct, but again the results are poorer overall.[5] These results support Burrows's decision to include words with both positive and negative z-scores in the test text, though it is interesting that negative test z-scores produce a larger maximum number of correct attributions in one instance. Note also that, when this method is used, each test text is compared with the primary texts on the basis of a newly generated and potentially unique set of words.

Another possibility that seems intuitively worth testing is whether test z-scores with large absolute values are more effective indicators of difference than those with small absolute values, or than the whole range of small and large z-scores. Performing a series of analyses in which only words with various limited ranges of z-scores in the test text are included in the calculation of Delta provides an answer. As before, this is accomplished by changing the formula for the absolute difference so that a zero is entered rather than the absolute difference if the condition is not met (the results are similar but less accurate if only the non-zero items are averaged).[6] The logic of this exercise is that words that are much more frequent or much less frequent than the mean in the test text may more

4 The simple formula for the absolute difference is ABS($J11-L11), where cell J11 contains the z-score for the most frequent word in the test text, and L11 contains the z-score for the same word in the first text in the primary set. The altered formula is IF($J11>=0,(ABS($J11-L11)),0).

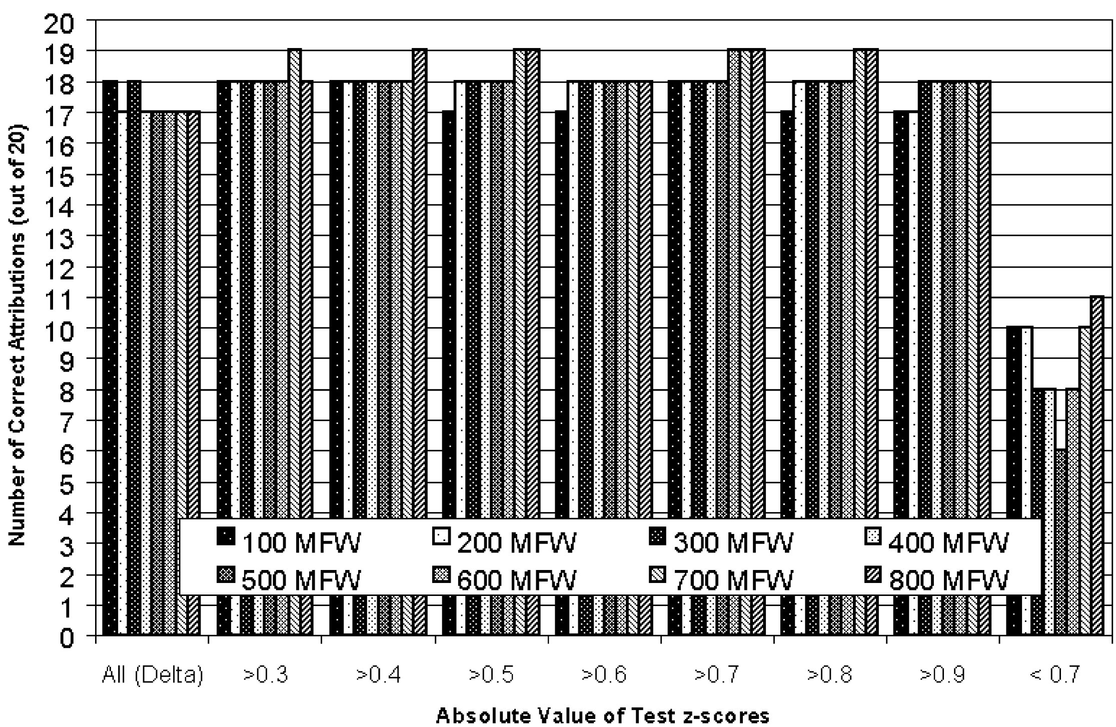
5 Here the altered formula is IF($J11<0,(ABS($J11-L11)),0).

6 For example, to test only those z-scores greater than 0.5 or less than -0.5, the formula is ABS(IF(ABS($J11)>0.5,($J11-L11),0)).

effectively distinguish it from the texts by each of the authors in the primary set. As Fig. 3 shows, when Delta is calculated on the basis of subsets of words with absolute z-scores in the test text greater than 0.3, its accuracy increases, yielding substantial numbers of analyses in which nineteen of the twenty texts by members of the primary set of texts are correctly attributed, rather than Delta's maximum of eighteen, and much better results overall. Note also that the best results using this method occur with the 700 and 800 most frequent words, rather than with the 100 and 300 most frequent using the original method, showing that words from throughout the frequency list are important when they show large z-scores. On the other hand, when Delta is calculated on the basis of words with relatively small z-scores in the test text (between -0.7 and 0.7), its accuracy is greatly reduced. This intuitively reasonable result suggests that further tests are worthwhile, and that it may be possible to 'tune' Delta so that it provides more accurate attribution results. It is important to keep in mind, however, that this method, like the previous two, somewhat counter-intuitively bases the differences between each pair of texts on different numbers of different words, so that it lacks the simplicity and elegance of Delta. For ease of reference, I will designate as Delta-Lz measures that base Delta on relatively large z-scores.

Before turning to some transformations of Delta itself, one additional method of limiting the words upon which the calculation of Delta is based deserves examination: including only those words for which the

Fig. 3 Delta and large and small z-scores.



signs of the z-scores of the test text and the primary text being compared are opposite. The logic of this method is that words for which the test text z-score is negative and the primary text z-score is positive, or vice versa—that is, for which the test text and the primary text are different from the mean in opposite directions—may be especially important. (This is a different question from the one considered above about using the absolute value of the difference between the test text z-score and the primary text z-score.) Consider a hypothetical trio of texts, a test text T and text samples E and F by two primary authors with the following z-scores for the words *the* and *a*:

	T	E	F	T-E	T-F	E-F
the	-1	1	-3	-2	2	4
a	1	-1	3	2	-2	-4
Mean of Abs. Differences (Delta):				2	2	4

Delta is identical for the pairs T-E and T-F, so that the authors of E and F are ranked as equally likely to be the author of T. Simply in terms of the gross differences in z-scores, it is clear that the differences between the texts in the two pairs are the same. If we turn from the arithmetic to the underlying facts, however, it seems reasonable to suggest that the differences between T and E may be more important than those between T and F. The authors of T and F both tend to avoid *the* and favor *a*, trending in the same direction from the mean, though to different degrees. On the other hand, E tends to avoid *a* and favor *the*, so that T and E trend in opposite directions from the mean. Furthermore, it seems counterintuitive that the authors of E and F are ranked equally likely to be the author of T when they are so very different from each other.

As with most questions in statistical approaches to authorship, there is no conclusive theoretical argument for (or against) treating these differences as equivalent, though doing so has the advantage of being simple and straightforward. Fortunately, the matter can be investigated inductively by comparing the accuracy of attribution using original Delta with the results achieved when Delta is calculated only on the basis of those words for which the test text z-score and the primary text z-score have opposite signs. Again, this is accomplished by changing the formula for the absolute difference so that it inserts the absolute difference only if the signs of the two z-scores are opposite, and zero otherwise.[7] In spite of the fact that this procedure again bases each comparison on different words and different numbers of words and reduces the total amount of information involved in the calculation, it dramatically increases the accuracy of Delta, as Fig. 4 shows. Delta achieves eighteen correct attributions only twice when calculated in the original way, but when only those words with opposite z-scores are included, there are seven results with eighteen correct attributions, and one with nineteen (here the results are much less accurate if the mean of the non-zero items is used). There are also three results in which the actual authors of all twenty texts by members of

7 To limit the calculation of Delta to words for which the z-scores of the test text and the primary text are opposite, the formula used is ABS(IF($J11<0,(IF(L11>0, L11-$J11,0)),(IF(L11<0, $J11-L11,0)))).

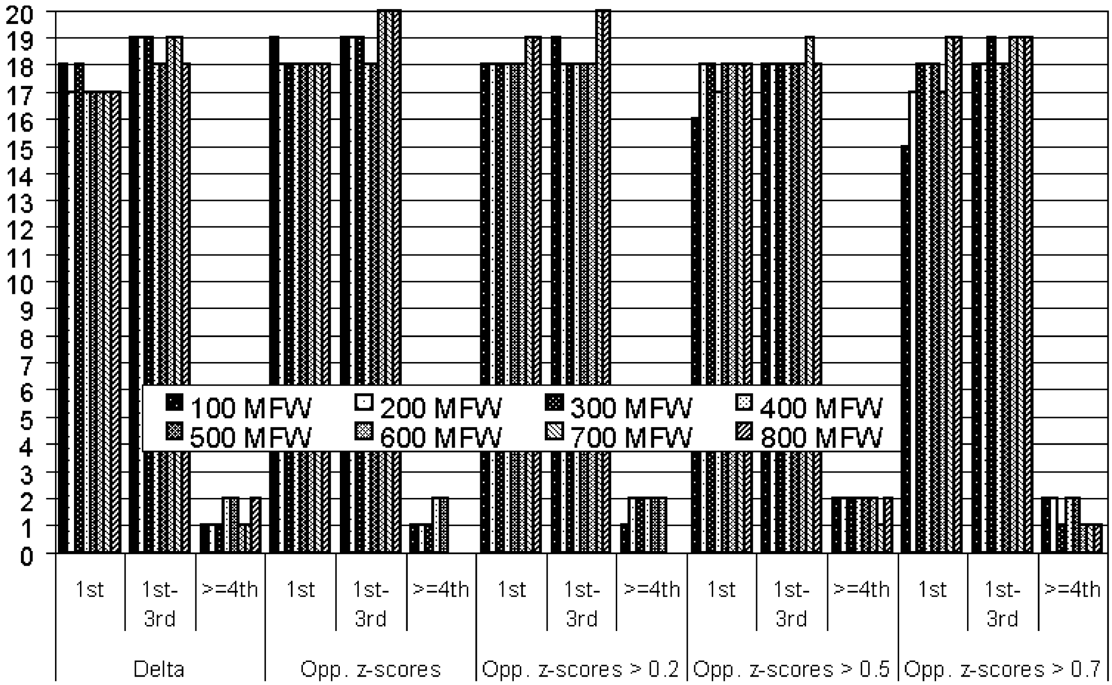


Fig. 4 Original Delta versus Delta based on words with opposite z-scores in test and primary texts.

the primary set are ranked in the top three, compared to a maximum of nineteen so ranked for Delta. In addition, there are more texts for which the actual author is ranked fourth or below when Delta is calculated in the original way (eleven total) than when only words with opposite z-scores are included (seven total). Finally, limiting Delta to words with opposite z-scores that have absolute values greater than 0.2 yields slightly more accurate results than the opposite z-scores alone, but limiting it to words with still larger z-scores reduces the accuracy (though the results are still more accurate than those based on the original calculation). This seems more likely to be caused by the dwindling number of words left in the analysis as those with smaller z-scores are removed than by any decrease in their significance as the z-scores increase. This second 'improved' version of Delta, based only on words that have z-scores with opposite signs in the test and primary text can be mnemonically labeled Delta-Oz.

5 Alternative Formulas for Delta

We have seen that limiting the calculation of Delta to words with relatively large z-scores in the test text or to words with z-scores of opposite signs in the test and primary text, or both, produces results that are more accurate than Delta itself, at least for this group of texts. Before testing whether these methods are also more accurate for other groups of texts, it seems worthwhile to test whether alternative formulas for or transformations of Delta also yield more accurate results. We can begin by examining the importance of the size of the absolute differences, altering the formula for Delta so that it includes only relatively large absolute differences. The results of a series of such alterations show that redefining Delta as the mean of the absolute differences greater than 0.5 produces slightly more accurate results than Delta, that differences greater than 0.7 are about as accurate as Delta, but those greater than 0.9 or greater than 0.3 are slightly less accurate.

Next, we can return to Burrows's decision to calculate Delta by taking the absolute value of the difference between the z-score of the test text and the z-score of the primary text. We have seen that retaining the sign seems inappropriate because of the possibility that positive and negative differences will cancel each other out, permitting texts with large opposite differences for different words to appear similar. There are, however, other ways to treat the sign of the difference that allow the information about whether the z-score of the test text is larger or smaller than the z-score of the primary text to be taken into consideration. The results of tests on several alternatives to Delta that incorporate such information are shown in Fig. 5.

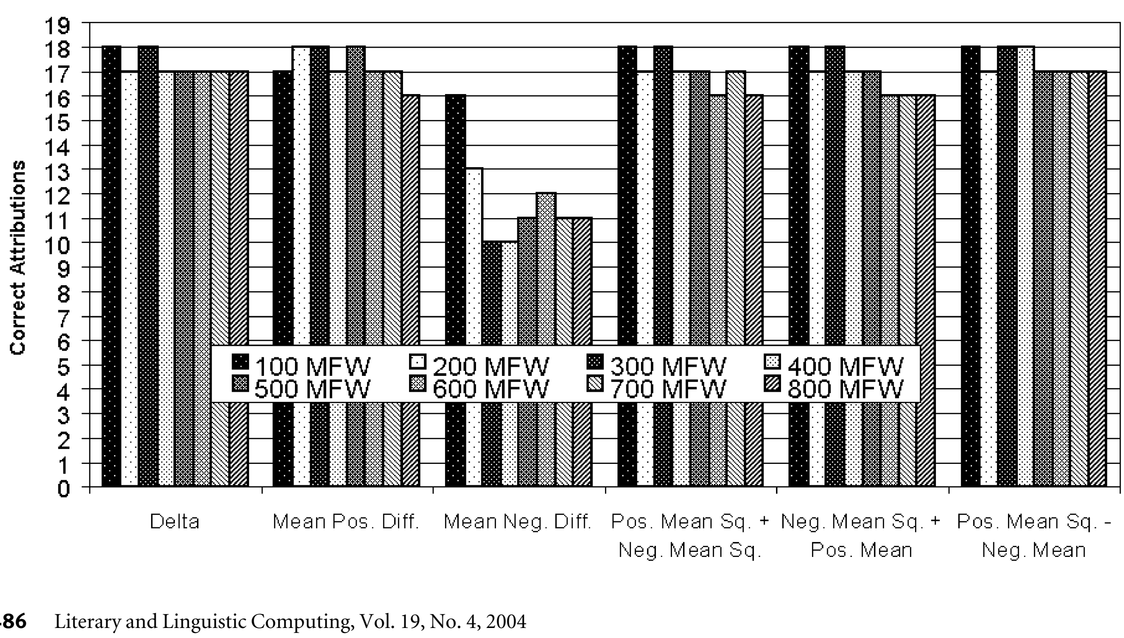
The first alternative in Fig. 5 redefines Delta as the mean of the positive differences between the test text and the primary text, ignoring the negative differences. This method bases Delta only upon words for which the z-score of the test text is greater than that of the primary text.[8] Perhaps surprisingly, this alternative produces results about as accurate as those based on the original Delta. Three analyses correctly attribute eighteen of the twenty texts by authors who are members of the primary set, compared with only two for Delta, but there are the same total number of correct attributions and one more result in which the actual author ranks lower than third (not shown).

8 The formula for Delta is AVERAGE(OFFSET(M$11,0,0,WdsToProcess,1)), where the 800 absolute differences between the test text and the first primary text are in M11-M810 and 'WordsToProcess' is a variable to allow the testing of various different numbers of words. To test the positive mean, I place the positive and negative values of the differences in column M rather than the absolute value and use the formula SUMIF(OFFSET(M$11,0,0,WdsToProcess,1),'>=0')/COUNTIF(OFFSET(M$11,0,0,WdsToProcess,1),'>=0').

9 With positive and negative differences in column M, the formula for the mean negative difference is SUMIF(OFFSET(M$11,0,0,WdsToProcess,1),'<0')/COUNTIF(OFFSET(M$11,0,0,WdsToProcess,1),'<0')

10 With negative and positive

Fig. 5 Delta and some alternatives.



When Delta is redefined as the absolute value of the mean negative difference, however, at most sixteen texts are correctly attributed to their actual authors, and the results are much less accurate overall.[9] Replacing Delta with the sum of the squares of the positive and negative mean differences, or with the square of the negative mean difference plus the positive mean difference produces results that are slightly less accurate than those based on Delta.[10] Replacing Delta with the square of the positive mean difference minus the negative mean difference (so that the negative mean is, in effect, added), however, produces results slightly more accurate than those based on Delta: there is one more result with eighteen correct attributions, one more correct attribution overall, and one fewer result in which the actual author is ranked lower than third (not shown).[11] The logic behind using the squares of the mean positive and negative differences is that the absolute values of the means tend to be in the range of 0.5 to 1.5, so that squaring has the effect of increasing the means greater than 1 and decreasing those less than 1.

Given that positive differences between the test text and the primary text produce results as good as Delta and are much more effective in attributing texts to their correct authors than are the negative differences, transformations that weigh positive differences more heavily than negative ones seem worth exploring further. Two very simple methods are to double or triple the positive mean before subtracting the negative mean (adding its absolute value). A less straightforward method is to add one to the positive mean and square the result. As noted above, the positive mean is typically less than or slightly more than one, so that adding one before squaring assures that squaring will increase all of the means, but will increase the larger ones more than the smaller ones. Finally, the negative mean can be subtracted from this last statistic (adding its absolute value). Although there is no theoretical basis for these transformations, the results presented in Fig. 5 provide a powerful inductive basis—one that can be tested experimentally.[12] The results of such tests on the four transformations just described are displayed in Fig. 6.

None of these transformations produces a larger number of correct attributions than Delta does, but they all produce more results in which eighteen texts are correctly attributed, and all produce fewer results in which the actual author is ranked fifth or higher. Twice the positive mean minus the negative mean, three times the positive mean minus the negative mean, and the square of the positive mean plus 1 minus the negative mean all seem clearly more accurate than Delta. These three transformations, which can be labeled Delta-2X, Delta-3X, and Delta-P1, now join Delta-Lz and Delta-Oz as potential improvements upon Delta.

6 Results of Delta and Five Alternatives on Two Additional Sets of Texts

Given the novelty of Delta itself and the lack of any transparent theoretical arguments for the possible improvements proposed here, further

differences in column M, the formula for the sum of the squares of the negative and positive differences is as follows: (SUMIF(OFFSET(M$11,0,0, WdsToProcess,1),‘<0’)/ COUNTIF(OFFSET(M$11,0, 0,WdsToProcess,1),‘<0’)) ˆ2+((SUMIF(OFFSET(M$11,0,0,WdsToProcess,1),‘>= 0’)/COUNTIF(OFFSET(M$11,0,0,WdsToProcess,1),‘>= 0’))) ˆ2.
For the square of the negative mean difference plus the positive mean difference, the formula is as follows: (SUMIF(OFFSET(M$11,0,0, WdsToProcess,1),‘<0’)/ COUNTIF(OFFSET(M$11,0, 0,WdsToProcess,1),‘<0’)) ˆ2+(SUMIF(OFFSET(M$11, 0,0,WdsToProcess,1),‘>=0’)/ COUNTIF(OFFSET(M$11,0, 0,WdsToProcess,1),‘>=0’)).

11 The formula for the square of the positive mean difference minus the negative mean difference is (SUMIF(OFFSET(M$11,0,0, WdsToProcess,1),‘>=0’)/ COUNTIF(OFFSET(M$11,0, 0,WdsToProcess,1),‘>=0’)) ˆ2-(SUMIF(OFFSET(M$11, 0,0,WdsToProcess,1),‘<0’)/ COUNTIF(OFFSET(M$11, 0,0,WdsToProcess,1),‘<0’)).

12 The formula for twice the positive mean minus the negative mean is (SUMIF(OFFSET(M$11,0,0, WdsToProcess,1),‘>=0’)/ COUNTIF(OFFSET(M$11, 0,0,WdsToProcess,1),‘> =0’)*2)-(SUMIF(OFFSET(M$11,0,0,WdsToProcess, 1),‘<0’)/COUNTIF (OFFSET(M$11,0,0,WdsToP rocess,1),‘<0’)). For three times the positive mean minus the negative mean, the

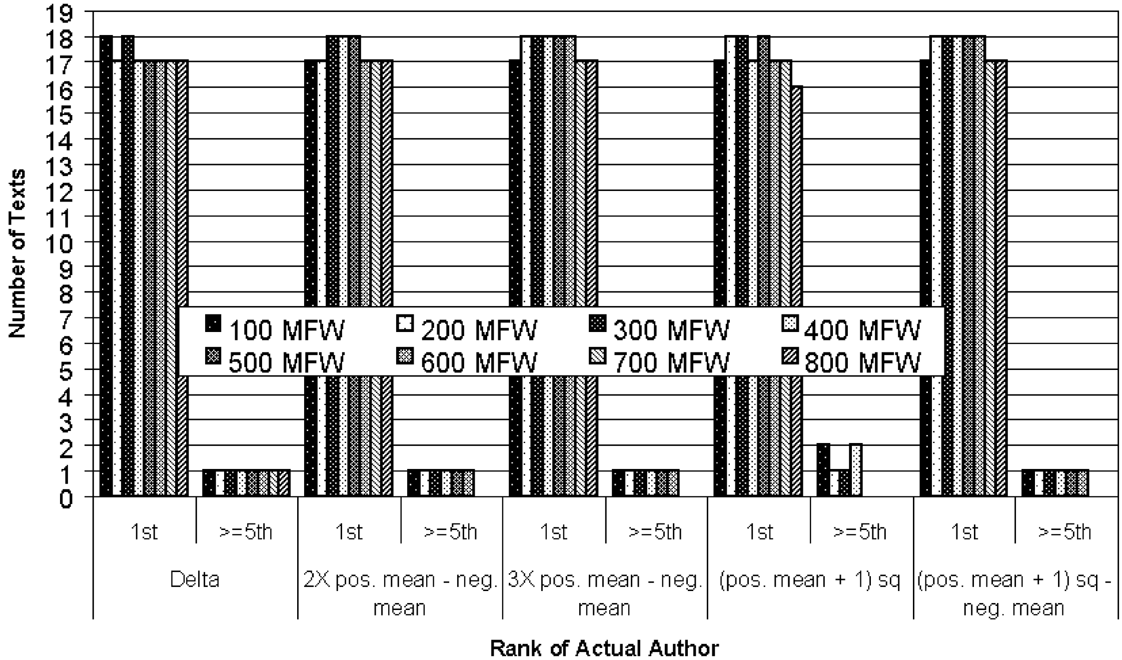


Fig. 6 Delta and four transformations.

testing on other kinds and sizes of texts will be required to determine whether any of these measures is consistently better than the others in authorship attribution. As a beginning, I have tested the five measures mentioned above as producing consistently more accurate results than Delta on two additional and quite different groups of texts. The first consists of sections of approximately 25,000 words of pure narrative (first- and third-person) from the beginnings of 104 novels by sixty-one authors, including the fifty-six novels included in the first group examined above. There are forty-five texts by forty-five authors in the primary set and fifty-nine texts in the second set, forty-three texts by members of the primary set, and sixteen by other authors (see Appendix). Such a large and diverse group can be expected to generate a large number of errors, so that any differences in the relative effectiveness of Delta and the new measures should be more apparent. The second new group of texts consists of forty-four sections of literary criticism by thirty authors, written in the past ten years, and ranging in size from about 4,200 words to about 14,300 words. There are twenty authors in the primary set, and twenty-four in the secondary set—fourteen texts by members of the primary set and ten by other authors.[13] This group tests Delta and the possible alternatives to it on smaller contemporary texts of a very different genre.

As Fig. 7 and Fig. 8 show, four of the measures, Delta-Lz, Delta-2X, Delta-3X, and Delta-P1, produce more accurate results than Delta on both of the new sets of texts. The fifth measure, Delta-Oz, is less effective than Delta for the critical texts, though it produces very good results for

formula is the same except for the substitution of '3' for '2'. For the square of the sum of the positive mean and one, the formula is (SUMIF(OFFSET(M$11,0,0, WdsToProcess,1),'>=0')/ COUNTIF(OFFSET(M$11, 0,0,WdsToProcess,1),'>=0') +1)^2. Finally, for this same statistic minus the negative mean, the formula is ((SUMIF(OFFSET(M$11,0,0 ,WdsToProcess,1),'>=0')/C OUNTIF(OFFSET(M$11, 0,0,WdsToProcess,1), '>=0')+1)^2)-(SUMIF (OFFSET(M$11,0,0, WdsToProcess,1),'<0')/ COUNTIF(OFFSET (M$11,0,0,WdsToProcess, 1),'<0')).

13 The twenty texts in the primary set are Altieri

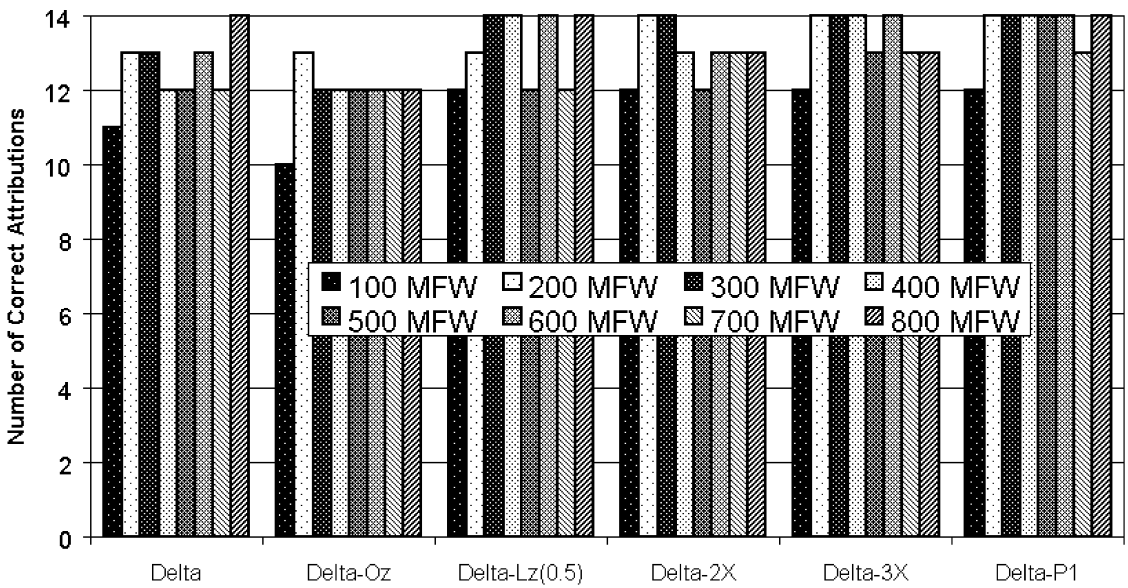


Fig. 7 Delta and five alternatives: results for forty-four sections of literary criticism.

(1998), Arac (1997), Choi (1996), Cohn (1995), Culbertson (1995), DeBruyn (2001), Donoghue (1996), Emerson (1996), Flieger (1997), Fluck (1996), Gans (1999), Gumbrecht (1995), Hartman (1995), Lerner (1995), McGann (2003), Morson (1996+1998), Nussbaum (1995), Ree (2001), Stock (1995), Tarlton (1996). The fourteen texts by members of the primary set are Altieri (2001), Arac (1999), Donoghue (1999), Emerson (1998), Fluck (2000), Gans (2000), Gumbrecht (1999), Lerner (1998), McGann (2001), Morson (1998), Nussbaum (2001), Ree (1995), Stock (1998), Tarlton (1999). The ten texts by other authors are Hollywood (1996), Hult (1997), Knight (1995),

the 104-text group and for the original group. More testing will be required to discover whether the weaker results for Delta-Oz on the critical texts is related to their smaller size or their genre. It seems unlikely that the contemporary date of these texts is the cause, for Delta-Oz produces results for contemporary poetry at least as good as the very strong results using original Delta that are reported in 'Testing Burrows's Delta' (n. 3). (The other four measures produce even stronger results.)

As a final way of assessing the relative accuracy of Delta and the new measures investigated above, we can return to the first group of texts examined above and compare Fig. 9, a new graph of others, errors, and members based on Delta-Lz(0.7), which produces the most accurate results overall, with Fig. 2, above, the graph based on Delta. Not only does Delta-Lz correctly attribute nineteen rather than eighteen texts, separation of the others from the members is clearer and the Delta-z scores for the members of the primary set are much higher overall, suggesting that the identifications are very unlikely to occur by chance.

7 Conclusion

At the very least, the results presented here show that simple unitary measures like Delta, Delta-Lz, Delta-Oz, Delta-2X, Delta-3x, and Delta-P1 are capturing authentic authorship characteristics. These measures produce results in difficult open authorship tests that are remarkably accurate, even on very large and diverse groups of first-person and third-person texts by authors of different nationalities. They do so even though each selection in both the primary and secondary sets is taken from a single text in the analyses above. Real authorship problems of such a level

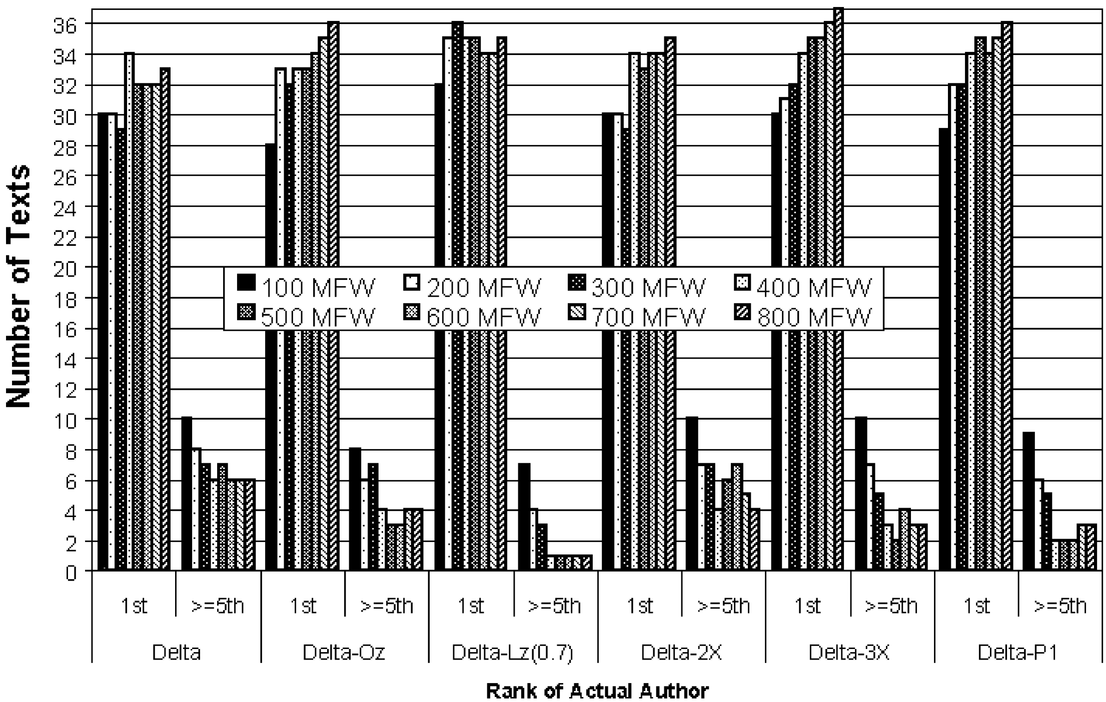


Fig. 8 Delta and five alternatives: results for 104 novels.

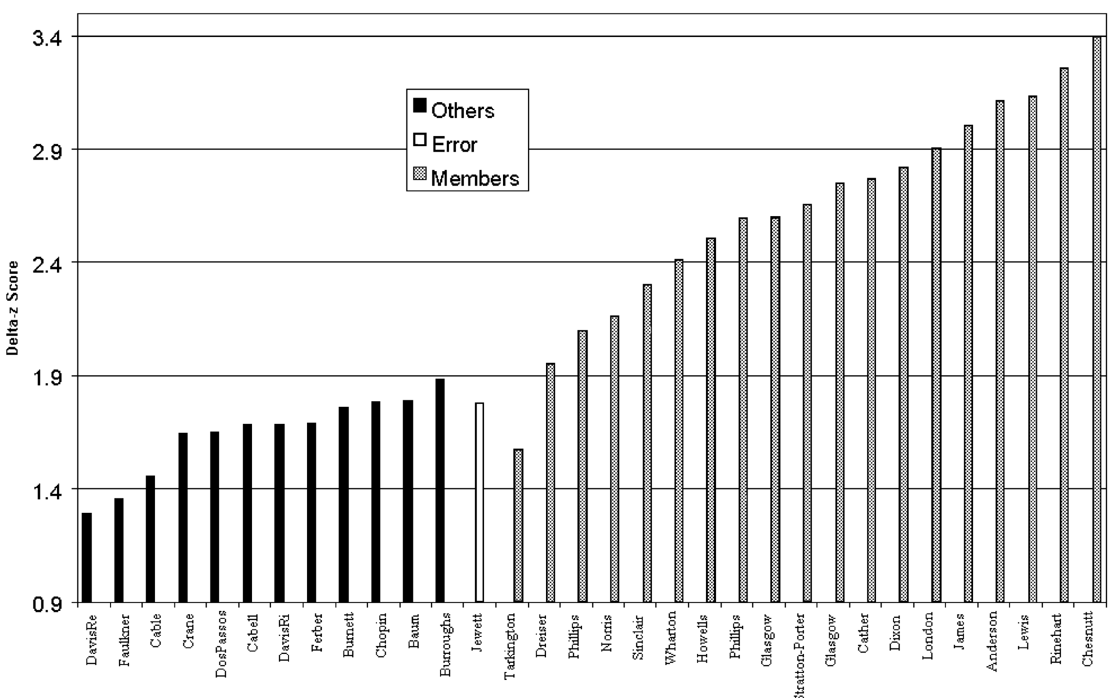


Fig. 9 Delta Lz(0.7) and thirty-two novels: others, error, and members (based on the 800 most frequent words).

O'Loughlin (2000), Sidorsky (2001), Stanbury (1997), Steele (1996), Willmott (2001), Wolff (1996), Zamir (2000). Note that the primary text by Morson combines two articles.

of difficulty are unlikely, a fact that strongly suggests that these measures will yield compelling results in real-world problems. The fact that Delta and four of the alternatives above produce excellent results when tested on both large sections of novels and smaller literary critical texts shows that they are versatile as well as robust. Preliminary results for contemporary poetry, along with those for contemporary literary criticism described here, suggest that Delta and its variants are also generally effective across long spans of time as well across various genres and sizes of texts. Further testing with additional kinds and sizes of texts will be needed to confirm their accuracy and to determine if one or another of them is consistently more reliable than the others or if their accuracy varies in any systematic way with the genre, size, historical period, or point of view of the texts. Additional transformations that further increase the accuracy of Burrows's basic technique may also be discovered. Nevertheless, Delta and its variants clearly seem poised to become important tools for non-traditional authorship attribution.

References

Burrows, J. F. (2001). *Questions of Authorship: Attribution and Beyond. Association for Computers and the Humanities and Association for Literary and Linguistic Computing, Joint International Conference*, New York, 14 June 2001.

Burrows, J. F. (2002). 'Delta': a measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, **17**: 267–87.

Burrows, J. F. 2003. Questions of authorship: attribution and beyond. *Computers and the Humanities*, **37**(1): 5–32.

Literary Critical Texts Analyzed

Altieri, Charles. (1998). Poetics as 'untruth': revising modern claims for literary truths. *New Literary History,* **20**(2): 305–28.

Altieri, Charles. (2001). Taking lyrics literally: teaching poetry in a prose culture. *New Literary History,* **32**(2): 259–281.

Arac, Jonathan. (1997). Postmodernism and postmodernity in China: an agenda for inquiry. *New Literary History*, **28**(1): 135–45.

Arac, Jonathan. (1999). Why does no one care about the aesthetic value of *Huckleberry Finn? New Literary History*, **30**(4): 769–84.

Choi, Julie. (1996). Feminine authority? Common sense and the question of voice in the novel. *New Literary History,* **27**(4): 641–62.

Cohn, Dorri. (1995). Optics and power in the novel. *New Literary History*, **26**(1): 3–20.

Culbertson, Roberta. (1995). Embodied memory, transcendence, and telling: recounting trauma, re-establishing the self. *New Literary History,* **26**(1): 169–95.

De Bruyn, Frans. (2001). The classical silva and the generic development of scientific writing in seventeenth-century England. *New Literary History,* **32**(2): 347–73.

Donoghue, Denis. (1996). Orality, literacy, and their discontents. *New Literary History*, **27**(1): 145–59.

Donoghue, Denis. (1999). Teaching literature: the force of form. *New Literary History*, **30**(1): 5–24.

Emerson, Caryl. (1996). Keeping the self intact during the culture wars: a centennial essay for Mikhail Bakhtin. *New Literary History*, **27**(1): 107–26.

Emerson, Caryl. (1998). Pushkin, literary criticism, and creativity in closed places. *New Literary History*, **29**(4): 653–72.

Flieger, Jerry Aline. (1997). Postmodern perspective: the paranoid eye. *New Literary History*, **28**(1): 87–109.

Fluck, Winfried. (1996). 'The American Romance' and the changing functions of the imaginary. *New Literary History*, **27**(3): 415–57.

Fluck, Winfried. (2000). The search for distance: negation and negativity in Wolfgang Iser's literary theory. *New Literary History*, **31**(1): 175–210.

Gans, Eric Lawrence. (1999). Staging as an anthropological category. *New Literary History*, **31**(1): 45–56.

Gans, Eric Lawrence. (2000). The last word in lyric: Mallarmé's Silent Siren. *New Literary History*, **30**(4): 785–81.

Gumbrecht, Hans Ulrich. (1995). Epiphany of form: on the beauty of team sports. *New Literary History*, **30**(2): 351–72.

Gumbrecht, Hans Ulrich. (1999). The future of literary studies? *New Literary History*, 26(3): 499–518.

Hartman, Geoffrey H. (1995). On traumatic knowledge and literary studies. *New Literary History*, **26**(3): 537–63.

Hollywood, Amy M. (1996). On the materiality of air: Janet Kauffman's Bodyfictions. *New Literary History,* **27**(3): 503–25.

Hult, David F. (1997). Words and deeds: Jean de Meun's *Romance of the Rose* and the hermeneutics of censorship. *New Literary History,* **28**(2): 345–66.

Knight, Deborah. (1995). Women, subjectivity, and the rhetoric of anti-humanism in feminist film theory. *New Literary History*, **26**(1): 39–56.

Lerner, Laurence. (1995). Peace studies: a proposal. *New Literary History*, **26**(3): 641–65.

Lerner, Laurence. (1998). Poet versus abstract noun: an agon. *New Literary History*, **29**(4): 769–84.

McGann, Jerome. (2003). Herbert Horne's *Diversi Colores* (1891): incarnating the religion of beauty. *New Literary History,* **34**(3): 535–52.

McGann, Jerome. (2001). Visible and invisible books: hermetic images in n-dimensional space. *New Literary History,* **32**(2): 283–300.

Morson, Gary Saul. (1996). Misanthropology. *New Literary History*, **27**(1): 57–72.

Morson, Gary Saul. (1998a). Contingency and freedom, prosaics and process. *New Literary History*, **29**(4): 673–86.

Morson, Gary Saul. (1998b). Sideshadowing and tempics. *New Literary History*, **29**(4): 599–624.

Nussbaum, Martha C. (1995). Political objectivity. *New Literary History,* **32**(4): 883–906.

Nussbaum, Martha C. (2001). The window: knowledge of other minds in Virginia Woolf's *To The Lighthouse. New Literary History,* **26**(4): 731–53.

O'Loughlin, Jim. (2000). Articulating *Uncle Tom's Cabin*. *New Literary History,* **31**(3): 573–97.

Ree, Jonathan. (2001). Subjectivity in the twentieth century. *New Literary History,* **26**(1): 205–17.

Ree, Jonathan. (1995). The translation of philosophy. *New Literary History,* **32**(2): 223–57.

Sidorsky, David. (2001). The historical novel as the denial of history: from 'Nestor' via the 'Vico Road' to the Commodius Vicus of Recirculation. *New Literary History,* **32**(2): 301–26.

Stanbury, Sarah. (1997). Regimes of the visual in premodern England: gaze, body, and Chaucer's *Clerk's Tale*. *New Literary History,* **28**(2): 261–89.

Steele, Miele. (1996). Metatheory and the subject of democracy in the work of Ralph Ellison. *New Literary History,* **27**(3): 473–502.

Stock, Brian. (1995). Ethical values and the literary imagination in the later ancient world. *New Literary History*, **29**(1): 1–13.

Stock, Brian. (1998). Reading, writing, and the self: Petrarch and his forerunners. *New Literary History*, **26**(4): 717–30.

Tarlton, Charles D. (1996). Fortuna and the landscape of action in Machiavelli's *Prince. New Literary History*, **30**(4): 737–55.

Tarlton, Charles D. (1999). 'The Word for the Deed': Hobbes's two versions of *Leviathan. New Literary History*, **27**(4): 785–802.

Willmott, Glenn. (2001). Canadian ressentiment. *New Literary History,* **32**(1): 133–56.

Wolff, Larry. (1996). 'The Boys are Pickpockets, and the Girl is a Prostitute': gender and juvenile criminality in early Victorian England from *Oliver Twist* to *London Labour. New Literary History,* **27**(2): 227–49.

Zamir, Tzachi. (2000). Upon one bank and shoal of time: literature, nihilism, and moral philosophy. *New Literary History,* **31**(3): 529–51.

Appendix Literary texts Included in Group 1 and Group 2

Author	Title and Date of Publication	Words	Narr.	Nation.	Group 1	Group 2
Cather	*My Antonia* (1918)	23,841	1st	A		Mem.
Cummings	*The Enormous Room* (1922)	28,184	1st	A		Pri.
Ferber	*Dawn O'Hara, the Girl Who Laughed* (1911)	24,750	1st	A		Mem.
Freeman	*The Heart's Highway* (1900)	25,780	1st	A		Pri.
Harris	*The Bomb* (1908)	22,753	1st	A		Oth.
Jewett	*The Country of the Pointed Firs* (1896)	18,935	1st	A		Mem.
London	*The Sea Wolf* (1904)	22,016	1st	A		Mem.
Rinehart	*The Case of Jennie Brice* (1913)	18,034	1st	A		Mem.
Wister	*The Virginian* (1902)	22,470	1st	A		Oth.
Anderson	*Marching Men* (1917)	28,756	3rd	A	Mem.	Pri.
Anderson	*Winesburg, Ohio* (1919)	38,051	3rd	A	Pri.	Mem.
Baum	*The Wonderful Wizard of Oz* (1900)	17,425	3rd	A	Oth.	Pri.
Burnett	*The Shuttle* (1907)	26,123	3rd	A	Oth.	Pri.
Burroughs	*Tarzan of the Apes* (1914)	38,849	3rd	A	Oth.	Pri.
Cabell	*Jurgen: a Comedy of Justice* (1919)	22,077	3rd	A	Oth.	Pri.
Cable	*John March, Southerner* (1899)	22,661	3rd	A	Oth.	Pri.
Cather	*The Professor's House* (1925)	21,871	3rd	A	Mem.	Pri.
Cather	*Song of the Lark* (1915)	29,394	3rd	A	Pri.	Mem.
Chesnutt	*The House Behind the Cedars* (1900)	27,049	3rd	A	Mem.	Pri.
Chesnutt	*The Marrow of Tradition* (1901)	33,447	3rd	A	Pri.	Mem.
Chopin	*The Awakening* (1899)	21,429	3rd	A	Oth.	Pri.
Crane	*The Red Badge of Courage* (1895)	33,898	3rd	A	Oth.	Pri.
Davis, Re.	*Frances Waldeaux* (1897)	10,311	3rd	A	Oth.	Pri.
Davis, Ri.	*Soldiers of Fortune* (1905)	34,834	3rd	A	Oth.	Pri.
Dixon	*The Clansman* (1905)	26,543	3rd	A	Mem.	Pri.
Dixon	*The Leopard's Spots* (1902)	20,972	3rd	A	Pri.	Mem.
Dos Passos	*Three Soldiers* (1921)	27,532	3rd	A	Oth.	Pri.
Dreiser	*Sister Carrie* (1900)	36,269	3rd	A	Mem.	Pri.
Dreiser	*The Titan* (1914)	27,082	3rd	A	Pri.	Mem.
Faulkner	*Light in August* (1932)	32,639	3rd	A	Oth.	Pri.
Ferber	*Emma McChesney & Co.* (1915)	16,969	3rd	A	Oth.	Pri.
Fitzgerald	*This Side of Paradise* (1920)	17,524	3rd	A	Pri.	Pri.
Frederic	*The Damnation of Theron Ware* (1896)	35,562	3rd	A	Pri.	Pri.
Glasgow	*The Battle-Ground* (1902)	27,359	3rd	A	Mem.	Pri.
Glasgow	*The Deliverance* (1904)	24,014	3rd	A	Mem.	Mem.
Glasgow	*Virginia* (1913)	23,763	3rd	A		Mem.
Glasgow	*Voice of the People* (1900)	33,236	3rd	A	Pri.	Mem.
Grey	*The Man of the Forest* (1919)	29,319	3rd	A	Pri.	Oth.
Howells	*A Hazard of New Fortunes* (1890)	27,187	3rd	A	Mem.	Pri.
Howells	*The Kentons* (1902)	34,282	3rd	A	Pri.	Mem.
James	*The Ambassadors* (1909a)	32,664	3rd	A	Mem.	Pri.
James	*The Europeans* (1878)	24,511	3rd	A		Mem.
James	*The Wings of the Dove* (1909)	34,526	3rd	A	Pri.	Mem.
Jewett	*A Country Doctor* (1884)	25,551	3rd	A	Mem.	Pri.
Jewett	*The Tory Lover* (1901)	22,337	3rd	A	Pri.	Mem.
Lewis	*Babbitt* (1922)	26,033	3rd	A	Mem.	Pri.
Lewis	*Main Street* (1920)	31,006	3rd	A	Pri.	Mem.
Lewis	*Our Mr. Wrenn* (1914)	19,236	3rd	A		Mem.
London	*Burning Daylight* (1910)	28,298	3rd	A		Pri.
London	*The Call of the Wild* (1903)	27,602	3rd	A	Mem.	Mem.
London	*White Fang* (1906)	25,877	3rd	A	Pri.	Mem.
Norris	*McTeague* (1899)	30,794	3rd	A	Mem.	Pri.
Norris	*Octopus* (1901)	32,314	3rd	A	Pri.	Mem.
Phillips	*The Conflict* (1911a)	29,612	3rd	A	Mem.	Pri.

Phillips	*The Fashionable Adventures of Joshua Craig* (1909)	21,954	3rd	A	Mem.	Mem.
Phillips	*The Grain of* Dust (1911b)	23,492	3rd	A	Pri.	Mem.
Phillips	*Susan Lenox* (1917)	27,739	3rd	A		Mem.
Rinehart	*The Breaking Point* (1922)	22,558	3rd	A	Mem.	Pri.
Rinehart	*Dangerous Days* (1919)	21,991	3rd	A	Pri.	Mem.
Robins	*The Mills of the Gods* (1920)	10,562	3rd	A	Pri.	Oth.
Sinclair	*The Jungle* (1906)	28,656	3rd	A	Mem.	Pri.
Sinclair	*The Metropolis* (1908)	27,514	3rd	A	Pri.	Mem.
Stratton- Porter	*Freckles* (1904)	24,536	3rd	A	Mem.	Pri.
Stratton-Porter	*A Girl of the Limberlost* (1909)	25,027	3rd	A	Pri.	Mem.
Tarkington	*Alice Adams* (1921)	20,763	3rd	A	Mem.	Pri.
Tarkington	*The Conquest of Canaan* (1905)	14,821	3rd	A	Pri.	Mem.
Twain	*The Tragedy of Pudd'nhead Wilson* (1894)	23,489	3rd	A	Pri.	Oth.
Wharton	*The Age of Innocence* (1920)	26,967	3rd	A	Mem.	Pri.
Wharton	*The House of Mirth* (1905)	36,826	3rd	A	Pri.	Mem.
White	*The Silent Places* (1904)	21,509	3rd	A	Pri.	Oth.
Conrad	*Lord Jim* (1900)	28,171	1st	B		Pri.
Conrad	*The Nigger of the Narcissus* (1897)	24,094	1st	B		Mem.
Doyle	*The Hound of the Baskervilles* (1901–02)	12,073	1st	B		Pri.
Doyle	*The Valley of Fear* (1914–15)	16,775	1st	B		Mem.
Ford	*The Good Soldier* (1915)	34,443	1st	B		Pri.
Murdoch	*Under the Net* (1954)	12,746	1st	B		Oth.
Stevenson	*Treasure Island* (1883)	11,055	1st	B		Mem.
Stoker	*Dracula* (1897)	17,242	1st	B		Oth.
Wells	*The War of the Worlds* (1898)	26,621	1st	B		Mem.
West	*The Return of the Soldier* (1918)	15,706	1st	B		Oth.
Chesterton	*The Man Who was Thursday* (1908)	24,612	3rd	B		Pri.
Forster	*Howards End* (1910)	18,744	3rd	B		Pri.
Forster	*Room with a View* (1908)	24,357	3rd	B		Mem.
Golding	*The Inheritors* (1954)	17,457	3rd	B		Oth.
Hardy	*Jude the Obscure* (1896)	20,616	3rd	B		Pri.
Hardy	*The Mayor of Casterbridge* (1922)	17,064	3rd	B		Mem.
Hardy	*Tess of the d'Urbervilles* (1891)	15,175	3rd	B		Mem.
Huxley	*Brave New World* (1932)	9,951	3rd	B		Oth.
Isherwood	*A Single Man* (1964)	11,282	3rd	B		Oth.
Joyce	*A Portrait of the Artist as a Young Man* (1916)	13,122	3rd	B		Oth.
Kipling	*The Jungle Book* (1893)	10,002	3rd	B		Mem.
Kipling	*Kim* (1901)	16,407	3rd	B		Pri.
Kipling	*The Light That Failed* (1890)	19,428	3rd	B		Mem.
Kipling	*Stalky & Co* (1899)	14,398	3rd	B		Mem.
Lawrence	*Lady Chatterley's Lover* (1928)	23,931	3rd	B		Pri.
Lawrence	*Sons and Lovers* (1913)	17,404	3rd	B		Mem.
Maugham	*Of Human Bondage* (1915)	22,088	3rd	B		Oth.
Orwell	*Nineteen Eighty-Four* (1949)	25,638	3rd	B		Oth.
Stevenson	*The Black Arrow* (1888)	22,499	3rd	B		Pri.
Wells	*The Invisible Man* (1897)	14,072	3rd	B		Pri.
Wells	*When the Sleeper Wakes* (1899)	25,065	3rd	B		Mem.
Wilde	*The Picture of Dorian Gray* (1891)	17,724	3rd	B		Oth.
Woolf	*To the Lighthouse* (1927)	15,054	3rd	B		Pri.
Woolf	*The Voyage Out* (1915)	26,386	3rd	B		Mem.