

Who wrote the Chard Reports? A Stylometric Analysis

By Dr. David I Holmes, The College of New Jersey, USA

1. The Chard Reports

The senior officer at Rorke's Drift on the fateful afternoon of Wednesday 22nd January 1879 was 31-year-old Lieutenant John Chard. Chard was a Royal Engineer who had only arrived in South Africa on January 5th and whose duties at Rorke's Drift were to supervise the construction of a redoubt overlooking the river and to build ponts. His immediate superior, Major Spalding, had gone to seek reinforcements, leaving Chard and Lieutenant Gonville Bromhead, the officer in charge of the garrison's B Company of the second battalion of the 24th Regiment to take care of the station. Chard's peace was disturbed when at about 3.30pm he noticed two horsemen galloping towards the drift from the direction of Isandlwana bringing with them the dreadful news of the massacre. Within an hour, the attack on Rorke's Drift had begun.

Within two days of the battle Chard submitted a perfectly sequential report of the engagement that was complete in extraordinary detail. This first 'Chard Report' became the 'official' report concerning the attack on the mission station and it was this report that Lord Chelmsford forwarded to the War Office. The many suspicions arising from the preparation of this report are covered in detail in Anglo-Zulu War historian Adrian Greaves' book *Rorke's Drift* (2002).

Greaves finds it remarkable that Chard could obtain a sufficient supply of clean undamaged paper when all available paper had been burnt in the hospital fire or destroyed during the fighting. Indeed, there is documented evidence that soldiers had to make do with scraps of scorched paper to pen brief notes home. The fact that amidst the chaos there were incessant downpours of rain makes the preparation of a clean report even more remarkable. Greaves also questions the accurate timings, precise locations and detailed names of the soldiers involved given the fact that Chard had only just arrived at the station and could not have known all the participants. Yet, there is no known record of any NCO assisting Chard with the report. Chard had a reputation for slothfulness and both Captain Walter Parke Jones and Lieutenant Henry Curling, whose contribution to this research is covered in a later section, have provided scathing anecdotes concerning his character.

It would clearly have been in Lord Chelmsford's interest to promulgate a dramatic report of the victory at Rorke's Drift which would deflect those critics seeking to humiliate him for the catastrophic defeat at Isandlwana earlier that day. As Greaves (p. 178) comments, in the minds of the British public "an inglorious defeat could be offset by a glorious victory." Although Chelmsford had not lingered at Rorke's Drift for more than a few hours, there is evidence that he was expecting a report concerning the victory in addition to issuing instructions for a formal enquiry to be conducted into the Isandlwana defeat. Greaves believes that Chelmsford's staff took the initiative in ensuring that Chard signed a suitably impressive report to offset the serious repercussions about to be unleashed following the losses at Isandlwana.

One of Lord Chelmsford's senior staff officers, Major Francis Clery, had remained behind at Rorke's Drift after Chelmsford and his staff departed on 24th January. Chard's 'official' report is dated the following day. Clery was a confidant of Chelmsford and an experienced report writer. He was a staff college graduate, a former Professor of Tactics at Sandhurst Military College, self-assured, observant and prone to making gossipy judgments about his colleagues. Greaves comments that Clery had also been culpable in the unfortunate decision-making process that led to the defeat at Isandlwana and he, of all people, would have realized that a dramatic report from Rorke's Drift might deflect the criticism that would undoubtedly be unleashed upon Chelmsford and his staff. Greaves believes that Clery is a strong candidate for authorship of this first 'Chard Report'.

Later in 1879 and now safely back in England, Chard was required to submit a further report at the request of Queen Victoria. He was required to include greater detail, although nothing is known of the preparation or research that went into this second report other than the fact that apologies for a delay were given to Queen Victoria as Chard claimed to have 'lost his notes'. No notes have ever surfaced. This second 'Chard Report' was submitted to Her Majesty at Windsor Castle on 21st February 1880. Major Clery would not have been present to advise on this report.

The key questions that this research will address are:

Are there any stylistic differences between the first and second Chard Reports?

Are there similarities in style between the first Chard Report and Clery's writings?

2. What is Stylometry?

Stylometry is a body of techniques used for the statistical analysis of literary style. These techniques date back to 1851, when the English logician Augustus de Morgan suggested in a letter to a friend that questions of authorship might be settled by determining if one text 'does not deal in longer words' than another. Since then stylometrists have searched for a reliable way of quantifying the style of a text which may be unique to an author. Growing computer power and the ready availability of machine-readable versions of many literary works have created a plethora of new techniques to examine authorship. Most researchers now believe that common words are the most valuable in characterizing an author's stylometric signature. They study the so-called non-contextual function words – prepositions, conjunctions, articles and adverbs – the 'humble servants' of speech, words that every writer employs subconsciously in a pattern as distinctive as a fingerprint.

Analyses with large sets (between fifty and a hundred) of these words have met with astonishing success. Essentially they pick a certain number (N) of the most common words in the texts under investigation and compute the frequency of these words in each separate text sample, thus converting each text into a multi-dimensional array of numbers. A set of statistical procedures, known as multivariate analyses, are then applied to the data to look for patterns.

One type of multivariate analysis is known as 'principal components analysis'. This aims to reduce the dimensionality of the problem by transforming the N variables to a smaller number (usually 2) of new variables. The first two new variables (or principal components) capture a large proportion of the variation in the original data and plotting the text samples in the space of these first two principal components enables the analyst to observe N-dimensional data in just two dimensions. Another type of analysis, known as 'cluster analysis', provides an objective way of identifying any groupings amongst the samples of text. These samples are plotted on a tree-diagram or *dendrogram*. Two samples that have a large degree of similarity in the values of their word-occurrence rates rapidly merge together in the manner of two branches on a tree. For an introduction to stylometry, see Holmes (1998).

3. Sampling and Textual Preparation

A number of control texts are necessary for this analysis, in the same genre and era but by writers who do not feature as contenders to the disputed Chard Reports. Letters from men writing home during the Anglo-Zulu campaign, not just officers, but NCO's and private soldiers whose letters, in the custom of the time, were passed on for publication in local newspapers, are available in *The Red Soldier* (Frank Emory, 1977). Three men in particular, Captain Edward Robert Prevost Woodgate, Captain Walter Parke Jones and Lieutenant William Weallens provided letters of sufficient length to qualify as suitable control texts, with each batch of letters being split into two to facilitate internal comparison.

Another excellent source of textual material is the book *The Curling Letters of the Zulu War* (Ed. Brian Best and Adrian Greaves, 2004). Lieutenant Henry Curling was one of only five officers who survived the slaughter at Isandlwana. A family man and prolific letter writer, his detailed accounts of

the invasion of Zululand make gripping reading. Two textual samples were taken from this book to add to the control texts.

Major Francis Clery was an experienced report writer. His letters are part of the papers of Sir Archibald Alison, who was Chief of Intelligence in the British Army and are published in *Zululand At War 1879: The Conduct of the Anglo-Zulu War* (Sonia Clarke, 1984). Three textual samples were taken from this book.

Both the first and second Chard Reports are available, unabridged, in Appendix A of *Rorke's Drift* (Adrian Greaves, 2002). The second report is reprinted there by kind permission of HM The Queen. The first report at 1,301 words is too small to be split, but the second report at 6,462 words was divided into four approximately equal-sized samples. Regrettably, apart from a brief 312 word letter written in September 1879 to Sir John Stokes (provided by Anglo-Zulu War historian Ian Knight from *The Royal Engineer Journal*) no usable authentic written material from John Chard is known to exist.

All these samples are listed in Table 1, the samples being either typed or scanned into machine-readable form. The choice of text size in stylometric studies is always problematic. Smaller units are too short to provide opportunities for stylistic habits to operate on the arrangement of internal constituents, while larger units are insufficiently frequent to provide enough examples for reliable statistical inference. Forsyth and Holmes (1996) found the median text block size in a selection of stylometric studies to be around 3,500 words. In their study of the *Book of Mormon*, Jockers *et al.* (2008) claim that even the smallest chapters are of adequate size for stylometric analysis, finding no correlation between the correct assignment of an author and the length of text sample. In all the following analyses, the occurrence rates of words are measured as percentages of the total sample size. Thus for this study, differences in sample size are not critical, provided we adhere to a stylometrically desirable minimum threshold of 1,000 words.

Table 1: Textual Samples

Author	Title	Sample	Sample size in words
Henry Curling	<i>Letters</i>	1	3112
		2	3035
Walter Parke Jones	<i>Letters</i>	1	1168
		2	1336
William Weallens	<i>Letters</i>	1	1396
		2	1564
Edward Woodgate	<i>Letters</i>	1	1380
		2	1379
Francis Clery	<i>Letters</i>	1	2559
		2	2324
		3	3047
	<i>First Chard Report</i>	1	1301
	<i>Second Chard Report</i>	1	1687
		2	1632
		3	1698
		4	1445

4. Stylometric Methodology

The pioneering work of Mosteller and Wallace (1964) on the use of function words in authorship attribution was continued by J. F. Burrows (1992). Since then multivariate statistical analyses involving

large sets of non-contextual high-frequency function words have been employed in attributional problems in a wide variety of authors and genres. See, for example, the investigation into the authorship of the so-called 'Pickett Letters' of the American Civil War (Holmes *et al.*, 2001) and the new look at the authorship of the *Book of Mormon* (Jockers *et al.*, 2008). The 'Burrows' approach has become the first port-of-call for attributional problems and will be the technique adopted in this investigation.

The number of function words used (N) varies by application and genre but typically lies between 50 and 75, the implication being that these words should be among the most common in the language and that content words should be avoided. A value of N set at 60 is used as a rule-of-thumb heuristic throughout this analysis, being an appropriate value for these sized text samples. Appendix A lists these sixty most common function words, taken from the corpus of texts in Table 1.

5. Hierarchy of Analyses

5.1 Controls: Curling, Jones, Weallens and Woodgate.

The first phase in this investigation is designed to test the validity of the proposed technique. For the purposes of this study, it is required that known texts can be shown to be internally consistent and separate from each other. The occurrence rates of the sixty words listed in Appendix A were computed for the individual textual samples from the letters of Lieutenants Curling and Weallens and Captains Jones and Woodgate. These were used as input to both a principal components analysis and a cluster analysis. The positions of the samples in the space of the first two principal components, which together explain 52.9% of the variation in the original data, are shown in Figure 1. An alternative analysis of the controls may be provided by conducting a cluster analysis on the textual samples, using the sixty word rates as variables and Ward's method as the clustering algorithm. Figure 2 shows the resulting dendrogram.

The results with these two methods of analysis are mutually supportive, with samples forming clusters on the basis of authorship. Our writers are internally consistent as regards their usage of these sixty words, yet are distinguishable from each other.

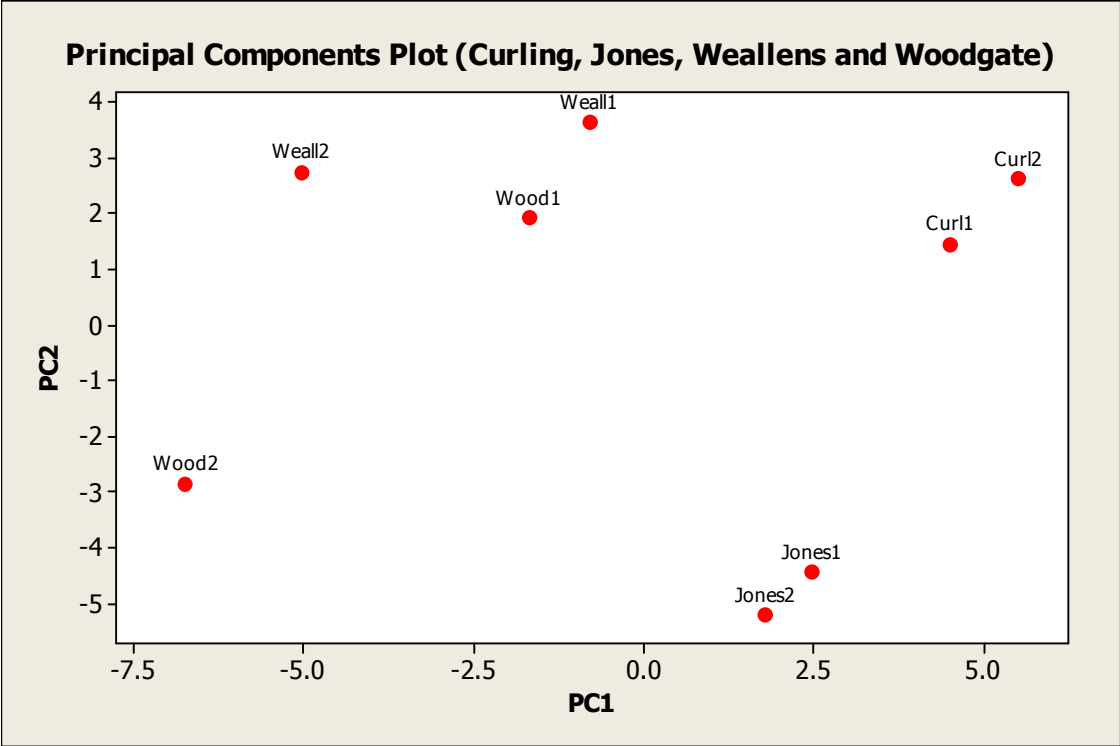


Fig. 1 Principal components plot: Curling, Jones, Weallens and Woodgate.

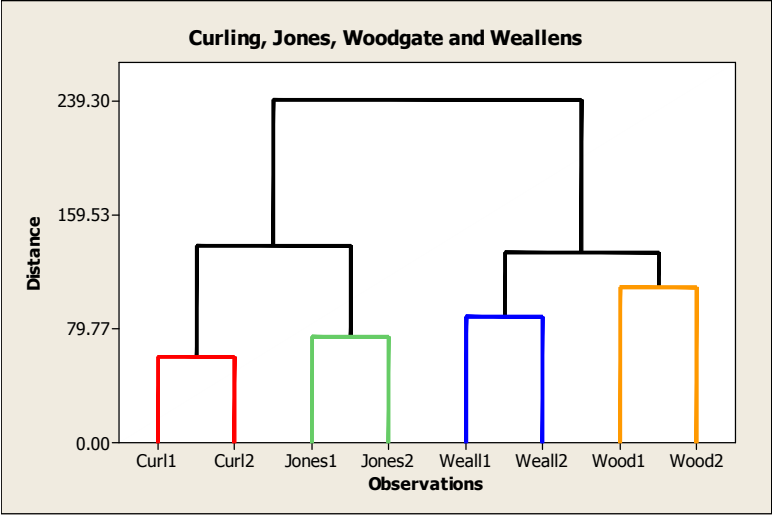


Fig. 2 Dendrogram: Curling, Jones, Weallens and Woodgate.

5.2 Clery, Curling, Jones, Weallens and Woodgate.

We now add the three textual samples from the letters of Major Francis Clery into the stylometric mix. The occurrence rates of the sixty most frequently occurring function words were once again used as

input to both a principal components analysis and a cluster analysis. The positions of the samples in the space of the first two principal components, which together explain 45.6% of the variation in the original data, are shown in Figure 3. An alternative analysis of the controls was provided by conducting a cluster analysis on the textual samples, using the sixty word rates as variables and Ward's method as the clustering algorithm. Figure 4 shows the resulting dendrogram. Both these plots show excellent internal consistency for the three Clery samples. They are also quite distinct from the samples taken from our three controls, providing validation for the 'Burrows' approach on works of known authors.

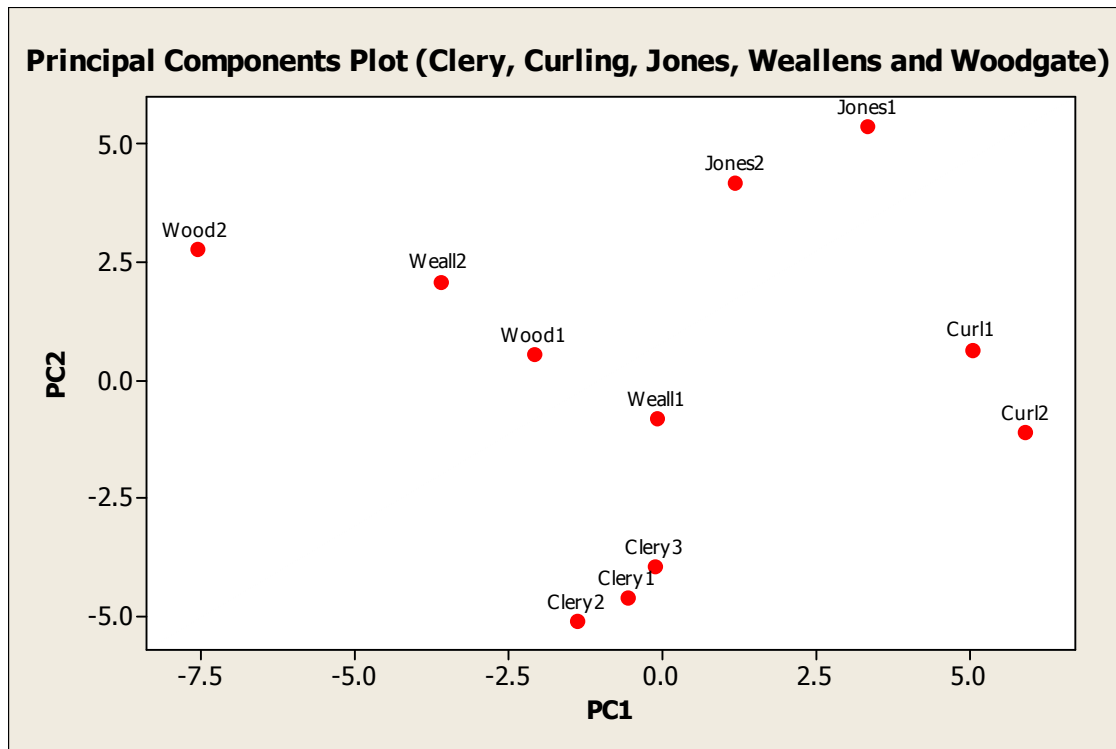


Fig. 3 Principal components plot: Clery, Curling, Jones, Weallens and Woodgate.

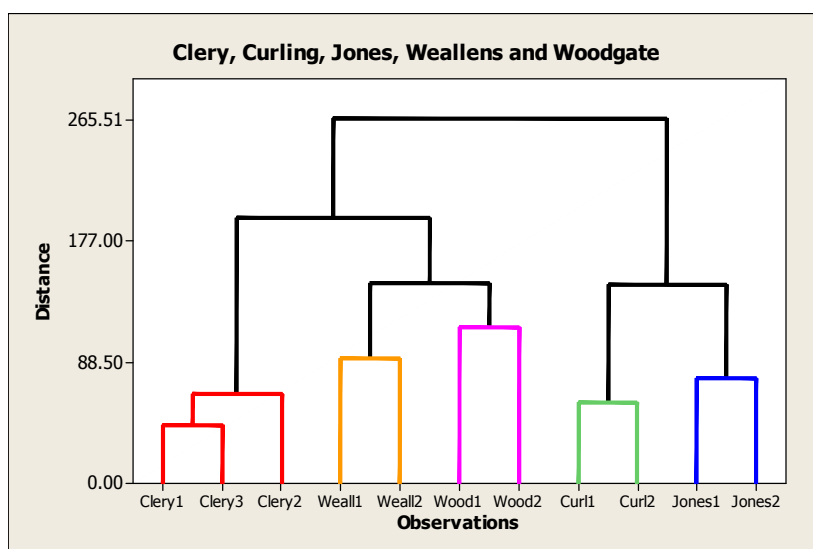


Fig. 4 Dendrogram: Clery, Curling, Jones, Weallens and Woodgate.
5.3 The Chard Reports and Clery.

Having successfully established the internally consistent style in the letters of Major Francis Clery concerning the usage of the sixty function words it is now time to focus on both Chard Reports. Discarding Jones, Weallens and Woodgate, who have done their duty, we now take the first Chard Report and the four samples from the second Chard Report, and add them into the mix with the three Clery samples. Figures 5 and 6 show the principal components plot and the dendrogram, respectively, from similar multivariate statistical analyses on the sixty function words. The former plot explains 50.4% of the variation in the original data.

Two important conclusions may be drawn from these clear and mutually supportive plots. First, there appears to be no difference in style between the first and second Chard Reports, suggesting single authorship. Secondly, the Chard samples cluster quite distinctly and separately from the samples of the writings of Clery. Our strong contender for authorship of at least the first Chard Report appears not to be a match. Might there be another contender amongst the British soldiers serving in the campaign?

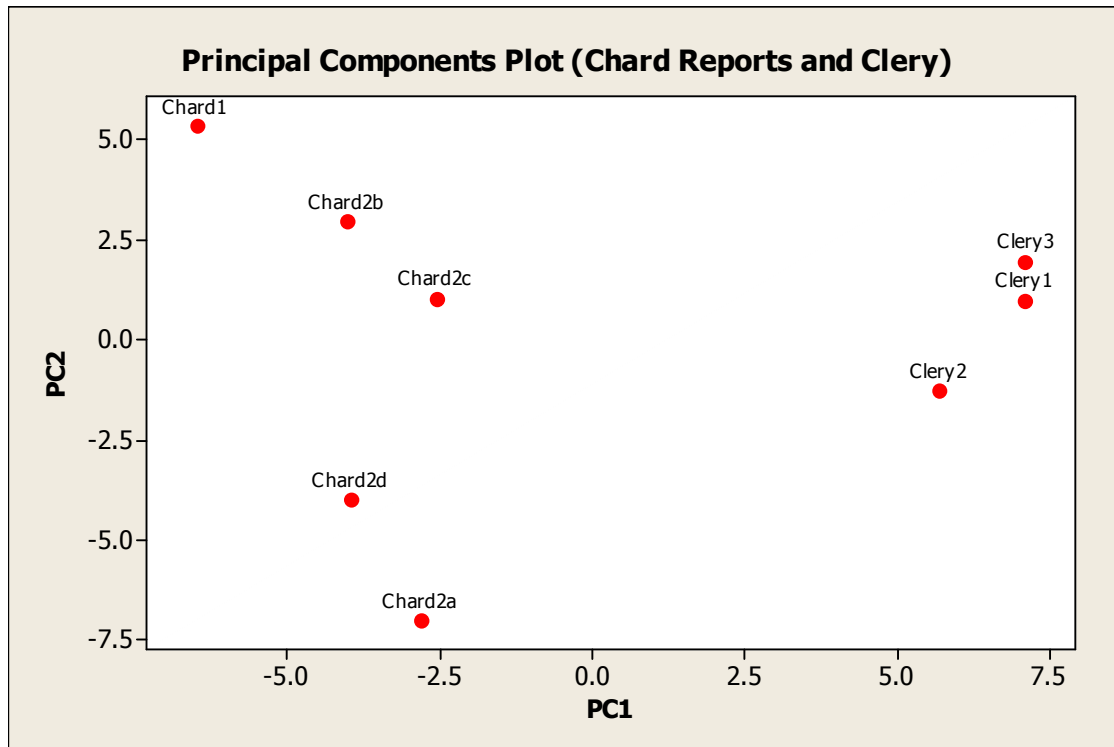


Fig. 5 Principal components plot: Chard Reports and Clery.

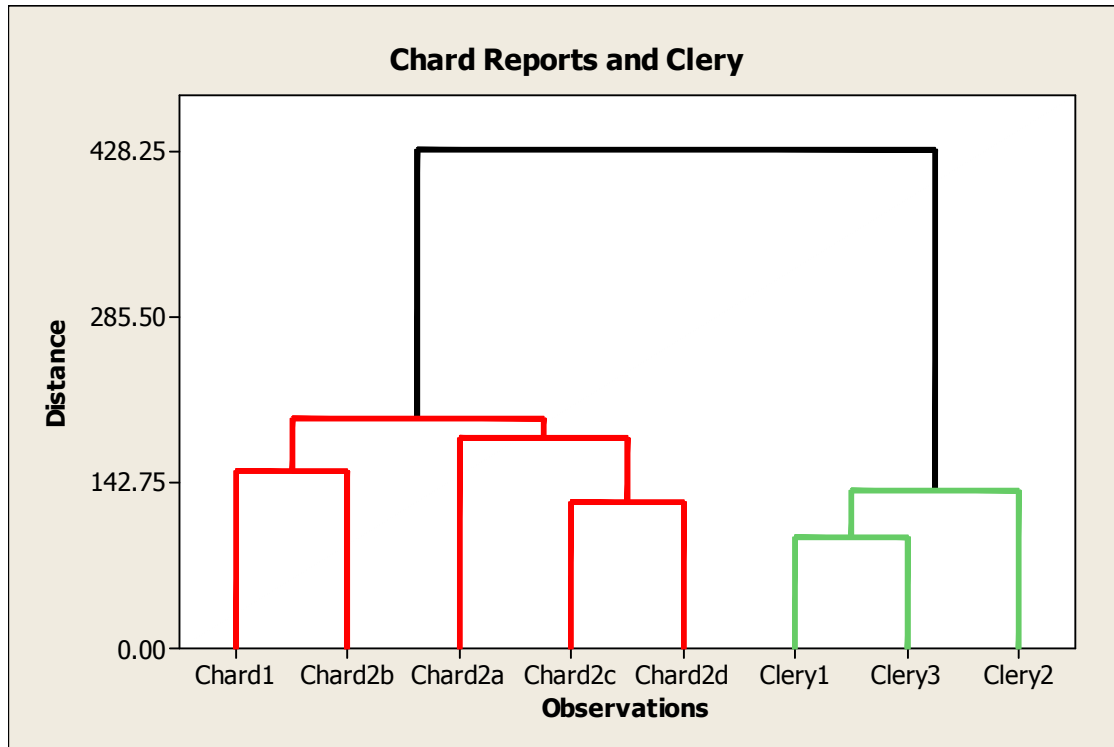


Fig. 6 Dendrogram: Chard Reports and Clery.

6. Colour Sergeant Frank Bourne

For many people, the outstanding character in the iconic 1964 film “Zulu” was Colour Sergeant Bourne, as portrayed by Nigel Green. The real Frank Bourne was born in 1854 and volunteered for the army in 1872. He was short and in his own words painfully thin, but upon arrival in South Africa rapidly ascended the promotion ladder. Bourne’s first duties at Rorke’s Drift on 22nd January 1879 would have been to supervise the taking down of the bell tents to give a clear field of fire, to post lookouts on the higher ground and then to lead a skirmishing line to intercept the advancing Zulu force.

In penning his first report after the action, John Chard would certainly have elicited help from NCO’s present at the defence of the post since he would not have seen everything nor known everyone’s names. Frank Bourne was literate and acted as ‘unpaid private secretary’ to those soldiers who could barely read or write, deciphering and answering their letters home. Might Bourne be the author of the first Chard report and might Chard have drawn on Bourne’s notes for the second report?

Bourne died in May 1945 and it is believed that he was the last surviving member of the Rorke’s Drift garrison. Yet in December 1936 he made a BBC radio broadcast concerning the battle for a series entitled ‘I was there’. It generated enough interest for 350 people to write to him. Regrettably the BBC scrapped the recording during the 1950’s as being of insufficient interest but a transcript of this broadcast is available in Appendix E of *Rorke’s Drift* (Adrian Greaves, 2002). This ‘Bourne Report’ has 2,246 words and was divided into two approximately equal-sized samples for analysis. The Bourne Report samples were added to those from both Chard Reports and the Clery letters and the occurrence rates of the sixty most frequently occurring function words were once again used as input to both a principal components analysis and a cluster analysis. The positions of the samples in the space of the first two principal components, which together explain 49.7% of the variation in the original data, are shown in Figure 7. Figure 8 shows the associated loadings plot from the principal components analysis,

which helps to explain the groupings in the main plot. One can imagine superimposing this graph on top of the principal components plot. Figure 9 shows the dendrogram, again using Ward's method as the clustering algorithm.

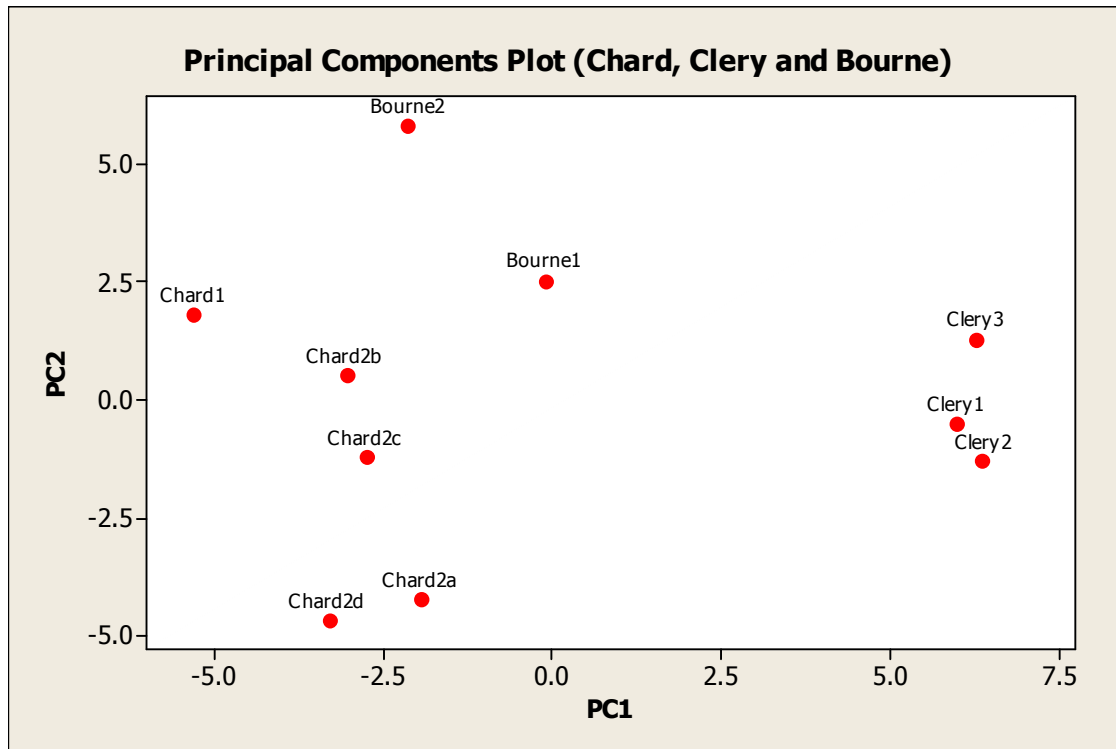


Fig. 7 Principal components plot: Bourne, Chard Reports and Clery.

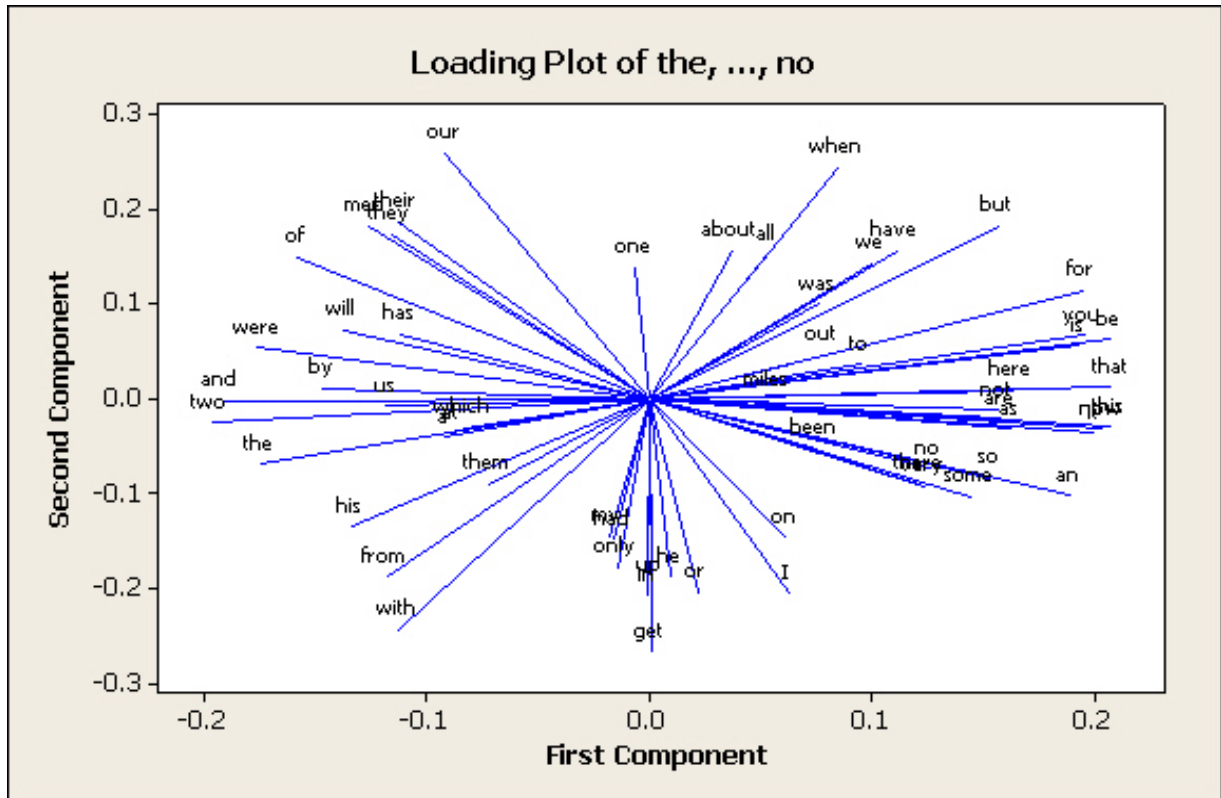


Fig. 8 Loadings plot: Bourne, Chard Reports and Clery.

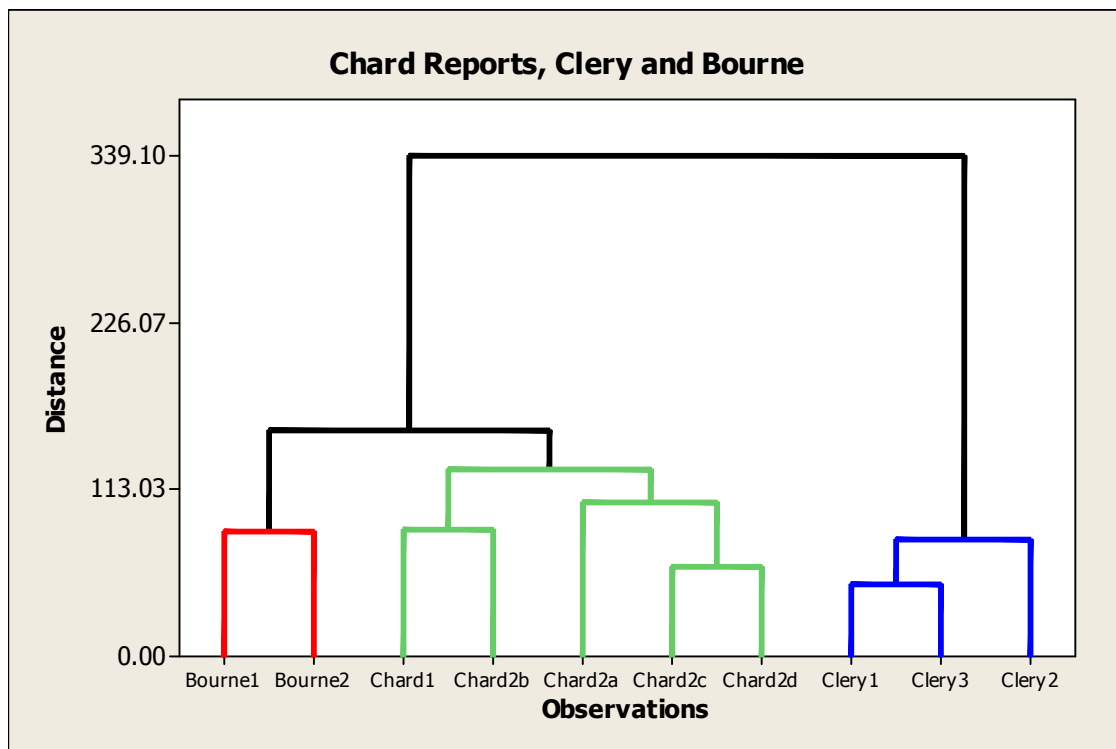


Fig. 9 Dendrogram: Bourne, Chard Reports and Clery.

7. Conclusion

From Figures 7 and 9 we see that the two halves of the Bourne Report show excellent internal consistency but, although slightly closer in style to the Chard Reports than Clery's writings, cannot really be considered to be a stylometric match. Figure 8 shows how words on the right such as 'for', 'you', 'is', 'be' and 'that' have high usage by Clery, whereas words on the left such as 'were', 'and', 'the' and 'his' are words of high usage in the Chard Reports. Favoured words in the Bourne Report are words such as 'our' and 'their'. We are left then with the following conclusions:

Both the first and second Chard Reports appear to be by the same hand.

That hand is not that of Francis Clery, despite Clery being mooted as a strong contender for author. Neither do the reports appear to be by Frank Bourne. Bourne may well have added details but the slightly closer similarity here may reflect the fact that Bourne was very familiar with Chard's first report by the time he gave his own account.

Anglo-Zulu War historians Adrian Greaves and Ian Knight, in private correspondence, have both sent interesting comments pertaining to these conclusions. Greaves still believes that Chard could not have written the first 'official' report, it being too accurate and of too academic a nature for him. Greaves remains in firm belief that Chard was neither clever nor erudite enough to pen this report and, if it is not by Clery, then the author could have been any erudite officer who had been able to piece together the sequence of events.

Knight agrees that, if some other officer was the author of the 'official' report then Clery was by far the obvious suspect since he was always complaining that everything was left for him to deal with. But, while being in no doubt that Chard was 'leaned upon' to provide a glowing report, Knight stresses that this is not the same thing as letting someone else write it. Knight believes that, although the ordinary soldier might not have had access to paper immediately after the Rorke's Drift battle, Clery could easily have given Chard some sheets when he passed on Chelmsford's request to write the report. Given Lieutenant Gonville Bromhead's reluctance to write a report himself, Knight thinks it more than likely that he would have offered up names and incidents for Chard to include, so the fact that Chard himself did not know the names of the private soldiers at Rorke's Drift is of little relevance. In short, the above results confirm Knight's belief that Chard himself was the hand behind his reports.

This research project now lies somewhat tantalizingly on hold, awaiting any future discovery by Anglo-Zulu War historians of new textual material from officers who were present at Rorke's Drift.

Acknowledgements

We wish to thank Anglo-Zulu War historians Dr. Adrian Greaves and Ian Knight for their invaluable advice throughout this project and for their help in providing textual material. We also wish to thank Dr. Richard Forsyth of the University of Leeds, UK, for the specialist computer software used and The College of New Jersey students Chris Millelot and Claudia Beard for their help in textual preparation.

References

- Best, B. and Greaves, A. (Ed.)** (2004). *The Curling Letters of the Zulu War*. England: Pen and Sword.
- Burrows, J.F.** (1992). Not unless you ask nicely: the interpretive nexus between analysis and information. *Literary and Linguistic Computing*, 7: 91-109.
- Clarke, S.** (1984). *Zululand At War 1879: The Conduct of the Anglo-Zulu War*. England: Brenthurst Press.

Emory, F. (1977). *The Red Soldier*. London: Hodder and Stoughton.

Forsyth, R.S. and Holmes, D.I. (1996). Feature-finding for text classification. *Literary and Linguistic Computing*, 11: 163-74.

Greaves, A. (2002). *Rorke's Drift*. London: Cassell.

Holmes, D.I. (1998). The Evolution of Stylometry in Humanities Scholarship. *Literary and Linguistic Computing*, 13: 111-117.

Holmes, D.I., Gordon, L.J. and Wilson, C. (2001). A widow and her soldier: stylometry and the American Civil War. *Literary and Linguistic Computing*, 16: 403-20.

Jockers, M.L., Witten, D.M. and Criddle, C.S. (2008). Reassessing authorship of the 'Book of Mormon' using delta and nearest shrunken centroid classification. *Literary and Linguistic Computing*, 23: 465-91.

Mosteller, F. and Wallace, D.L. (1964). *Applied Bayesian and Classical Inference: The Case of The Federalist Papers*. Reading, MA: Addison-Wesley.

Appendix A. Sixty most frequently occurring function words used in the analyses.

the	and	to	of	a	we	in	I	is	it
have	on	as	was	at	are	with	that	not	they
all	be	for	were	he	very	from	but	had	our
there	about	so	this	will	one	men	been	which	them
out	now	by	get	my	up	no	two	or	when
miles	here	you	their	his	an	some	only	has	us

Editor's footnote.

Whilst I totally accept the accuracy and findings of the above research, I remain uncertain that the report could have been written at Rorke's Drift due to the well-known lack of writing paper, the severe weather conditions prevailing at the time and the lack of shelter at Rorke's Drift. Had Clery or anyone else had paper, I would expect that they would have made it available for much more important documents such as Daily Orders and records of events and incidents. For example, the original message sent to Colonel Glyn with the names of the six praise-worthy soldiers was written by Bromhead on a small piece of paper.

My original guess was that Clery wrote the original 'Chard report' because he was at Rorke's Drift when the report was written, and he was one of the few individuals present who was in a position to carefully collate all the information before compiling the remarkably accurate and mistake-free account. So, if it was not Clery, who wrote it?

When later requested to re-write the report for Queen Victoria, Chard was unable to comply until he obtained a copy of the first 'Chard Report', claiming his original notes were 'lost'. I wonder if he had them in the first place.

This mystery is typical of what makes the subject of the Anglo Zulu War so intriguing; we just don't know who wrote the report.

Adrian Greaves