INDIVIDUAL CONSTANCIES IN WRITTEN LANGUAGE EXPRESSION

Joy Prentice Aked

Thesis submitted for the degree of Doctor of Philosophy in the Department of Psychology, Faculty of Social Science, University of Glasgow, April, 1994.

Abstract

This thesis investigates whether there are individual constancies in writing style, that is, whether there are any features in an individual's writing which remain the same over a large number of samples of their texts. This question is examined first by an analysis of the Cumulative Sum technique; a method which has recently been developed for purposes of authorship identification, and claims to be able to distinguish with 100% accuracy, whether a piece of text has been written by one, or more than one individual. This claim is explored thoroughly, as well as other related claims which concern the technique's application to different genres. The analysis of this specific method, is followed by a systematic analysis of some surface variables in the text, those measured by a stylechecker. The relationships existing among these stylecheck variables are examined. In addition, the relationship between the stylecheck variables and personality is examined using Cattell's 16 P.F.. Finally, the relationships between some more subjective or idiosyncratic features of the text are investigated for any individual differences or constancies.

The results showed that there was no scientific basis to the claims behind the cusum technique. The fundamental assumption of the technique was proved to be wrong. No evidence for the unique separation of individuals using either the stylecheck variables or the subjective variables was found. However, some interesting correlations were detected between the stylecheck variables and the personality factors.

These results implicate certain constancies within individuals but these constancies are not enough to distinguish individuals uniquely.

Contents

PART 3: ANALYSIS OF INDIVIDUAL CONSTANCIES IN SURFACE VARIABLES AND THEIR RELATIONSHIP TO PERSONALITY

CHAPTER 5: Individual Differences in Stylecheck indices

CHAPTER 6: Style Variables and Personality

PART 4: SUBJECTIVE JUDGEMENTS OF WRITING STYLE

CHAPTER 7: Subjective judgement of Style Features

List of Appendices

List of Tables

List of Figures

Acknowledgements

I would first of all like to thank my supervisor, Professor Tony Sanford, for his constant encouragement and guidance throughout my studies and his advice and assistance with the final version of the thesis.

Thanks also go to Linda Moxey for her inspiration with Chapter 6, to Paddy O'Donnell for his constant help with the statistics, and to Jim Mullen for writing the CUSUM programs. In addition, I would like to thank Gwyneth Sneddon and Alison Newlands for putting up with me, especially in the latter weeks of the thesis and Sandy Forsyth for enduring my constant ramblings about my research, and for advising me in the workings of the law.

Finally, I would like to thank my family for their tolerance, and especially my mother, Margot Aked, for her constant proof reading and encouragement throughout my years at university. A special word of thanks goes to my father, Arthur Aked, for his work with the diagrams, and in whose footsteps I was following in undertaking this research. With this in mind, I dedicate the thesis to him.

Declaration

I declare that this thesis is my own work carried out under normal terms of supervision.

INTRODUCTION

Research on Individual Differences

The issue of individual differences is one which has been widely accepted and studied in a variety of areas since close to the conception of the study of psychology. Individuals vary in many psychological characteristics such as their physical and mental ability, knowledge, habits, personality and character traits. It is worth asking then, in what ways are people different, how much they differ, why they differ, and whether or not a scientific study of these differences could prove useful.

There are two main areas into which the study of individual difference has fallen: the study of personality and the study of intelligence. Tests have been specifically designed to measure differing abilities in both of these fields (e.g. the Minnesota Multiphasic Personality Inventory (MMPI), the Stanford-Binet Intelligence Scale) and these are constantly being revised and updated to improve designs and incorporate new findings. Recent research has focused on the reason for these individual differences, and the role that environmental and genetic factors play in their formation.

Comparatively recently, a new study of individual difference has emerged. This has been concerned with individual writing style and the differences which exist between individuals, and the constancies within individuals, in the way that they express themselves through writing.

Individual differences and authorship studies

Once the connection had been made between individual differences and writing style, the field of investigation was readily extrapolated to the forensic context and to studies of authorship in unknown or doubted contexts. Originally these investigations were consigned to studies of literary works such as Shakespeare and Marlowe, or the authors of the Bible, but gradually, the emphasis has moved to the study of the typical individual and the words he or she chooses to use. In particular, it is now being applied in the forensic contexts, where there is a doubt as to the author of a piece of text.

Statistics were calculated on many different indices within writing, in much the same way as they had been on intelligence and personality. The choice of measurements to make were endless, but the emphasis landed on counting words and phrases which were abundant in all writing. The most recent method, the Cusum Technique, at which a large part of this research is directed, also uses frequent words or "habits" in the determination of authorship. Marker words, or words which are used characteristically by one author, have also been used to this end.

Graphology, Personality and Writing Style

The other topic which relates to the issue of writing style, but which is not addressed in this thesis, is that of graphology. Graphology can be defined as the study of handwriting, especially for the purpose of character analysis. It is regarded by some as a signature of an individual, in fact, Loewenthal (1982, p91) even went as far as to say

that " regardless of the limb used to guide the pen .... (it) .... is not so much handwriting but brainwriting". Presumably Loewenthal meant what Michaels et al. state more succinctly; that handwriting is:
" a conditioned response directed by the subconscious part of the brain. The brain creates a symbol before the act of writing is attempted, and the writing is affected by the emotions through the Limbic System" (Michaels, Maze & Hados, 1989, p79) . This is an extreme view however, and there is considerable controversy as to the validity of graphology as a tool for personality assessment. Nevertheless, graphology is still being considered as having some scientific and forensic basis.

It is not difficult to see the connection between graphology, writing style, and the study of personality. Both graphology and the analysis of writing style have been applied to the forensic context since they are both thought to reflect differences across individuals and therefore, be a means of identifying the individual. Graphology looks at *how* something is written, by examining features of the script itself such as height, width , slant, pressure or form of connection, whereas writing style looks at *what* was written, in terms of word usage.

Personality has been much examined in relation to graphology. The objective of graphology is to asses the personality of the writer. However, studies which have addressed the question of the correspondence between personality and handwriting, have generally failed to find a consistent relationship between the two. For example . Bayne et. al. (1988) concluded that the results of a study which attempted to asses personality in terms of the Myers-Briggs Type Indicator (MBTI) using graphological analysis, showed "that none of

the judges were able to judge MBTI preferences or psychological type accurately though they were confident of being able to do so both before and after making their judgements" (Bayne & Rowan, 1988).

A study by Bushnell (1993) related personality assessments via Cattell's 16 P.F., to assessment via handwriting. Using self-assessment and assessment by another person (the social partner), it was found that only the standardised psychometric questionnaire permitted personal identification. The handwriting assessments did not.

The question still remains then, as to whether the relationship between personality and writing style will follow the same pattern as the studies on personality and handwriting. This relationship between writing style and personality using the 16 P.F. is just one of the issues addressed here in the present research.

Writing style as a fingerprint

One interpretation of individual writing style has been applied by Morton[1] which makes it comparable with other studies of individual difference. This is the comparison of writing style with a fingerprint. Fingerprints were first used in the early 20th century when they were discovered to be unique to the individual, and hence a means of identification of that individual. Morton likens an individual's use of words to a fingerprint. He maintains that the proportion of various word types in a sentence is unique to the individual in the same way that his fingerprint is unique, and that this proportion will not change in different circumstances or over genres. It this were true, it would

[1] e.g. First conference for Forensic Linguistics, in Birmingham, November, 1991

make writing a prime candidate for the study of individual differences and would have far-reaching consequences when applied to a forensic context. It is precisely this claim and its consequences, which are the topic of this research.

Conclusion

In conclusion, there are some very complex issues associated with the study of individual difference. How these issues are examined, depends to a great extent on the discipline of the investigators. This thesis examines writing style from a psycholinguistic point of view, in the light of studies on individual difference and constancies.

An examination of writing style is possible using a wide variety of measures. In part 1 of the thesis, the major studies on writing style, and in particular authorship studies, are reviewed since the conception of the field in the mid 1800's. An attempt is then made to relate these studies of forensic stylistics in general, to psychological issues. A brief examination is made on the status of forensic stylistics in the legal field, as well as of law-related problems that have arisen with use of specific stylistic methods.

Part 2 of the thesis examines in detail a method which has only recently received any systematic scientific study. This method is the Cusum Technique. The cusum technique is explained, and the claims made about its power to discriminate authors set out in detail. The basic assumptions behind the technique are then systematically tested using both hypothetical data and real data from real texts. Following this, in chapter 3, the cusum technique is examined in the forensic

context as a tool for authorship discrimination. Again simulations are carried out using hypothetical data and are followed up using data from real text. The probability of correctly identifying when a piece of writing is made up of text from one or two authors is examined, as well as some of the other factors which will affect the interpretation of the cusum curve. Finally in this section, two further studies using cusum analysis are undertaken, the first looking at a comparison of a large number of texts from one and two authors, and the second looking at the difference between spoken and written texts in relation to cusum analysis.

Part 3 of the thesis moves to a different perspective on the study of writing style. Perhaps the most obvious method of analysis is undertaken, which involves the examination of individual differences and constancies in writing using a group of easily counted surface variables. The examination is made using incidences of variables picked up by a computer stylechecker, such as word types and sentence beginnings. Several clustering algorithms are applied to this data in order to determine whether individuals can be separated by a method such as this. This part of the thesis also addresses the relationship between these stylecheck variables and certain personality characteristics. Personality is measured using a well-known personality test (Cattell's 16 P.F.), and scores on this are correlated with scores on the variables measured by the stylechecker. An attempt is then made to explain the correlations which are obtained, in terms of individual differences.

The final section of the thesis uses more subjective measurement of style in the same way as the stylecheck variables were used previously.

The subjective variables, such as use of different scripts for emphasis, abbreviations or numbers, are again examined using several clustering algorithms, to see whether texts can be separated by subject.

This thesis is thus an examination of individual differences and constancies in written language using several techniques.
The basic questions are:
(1) Is the idea of discernible individual differences and constancies tractable?

(2) Do individuals differ from each other sufficiently on measurable dimensions to support the claims that writing style is like a "signature" or fingerprint?

(3) Is the use of style- measurement justified as evidence in legal (forensic) settings?

(4) Is there any evidence for a relation between personality and style variables?

This thesis was begun in the anticipation of finding positive answers to at least some of these questions. Before long, however, it became evident that we are at the beginning of the study of individual differences in writing style. The general optimism which surrounds the area is evident in Part 1 of the thesis. The remainder scrutinises the empirical support for the four issues listed above.

PART 1: Review and Background

CHAPTER 1 : Forensic Linguistics - Review and Background

A. Early Statistical Methods

1. Earliest known Authorship studies

The earliest recorded study of authorship is that of Augustus de Morgan, in 1859, who was one of the first modern exponents of the idea that the data about the content of writing might help identify the author. Semantics were of no interest to him, but simply the physical properties of the writing. De Morgan suggested that the average length of word might vary from person to person, but remain constant within each individual's writing. This did not prove to be the case since word length tends to depend heavily on context and subject matter.

Mendenhall in 1887 published some research on the authorship of some Shakespearean plays, using word length as identification (Mendenhall, 1887). He recorded the number of words with 1 letter, 2 letters, 3 letters and so on, counting about 2 million words. He applied these principles to authors such as Bacon, Jonson, Marlowe and others. The study showed that for most of the authors the highest frequency was of 3 letter words; only Shakespeare and Marlowe were high in their use of 4 letter words. Mendenhall presented his results as a series of frequency graphs showing how many words of each length were found for each author. For comparative purposes, counts for two authors could be placed on the one graph.

Mendenhall showed remarkable foresight in this method, as it appeared again, in a slightly modified form, in the work of Yule in 1939. Yule used the same principles as Mendenhall, but he used them on counts of sentence length instead of word length. Yule (1939) suggested that sentence length distributions could be used as a test of authorship. At the time, his work received little attention, probably because it was not easy at that time to conduct a scientific study using sentence length, as it was very difficult to collect a large amount of data from works of known authorship. Also, there was a problem with Yule's work because of his belief that the variability observed in the known works of a single author was greater than the standard errors of the statistics he employed. (Wake, 1957)

However, William Wake (1957) continued the work which Yule had begun and looked more closely at sentence length distributions, in particular, in classic Greek works. His interest lay in the problems which arose concerning the definition of sentence length, punctuation peculiar to Greek texts, sampling, and the constancy of sentence length distribution for one author over a period of time. He concluded that sentence length distributions could be used as an objective criterion of authorship style as distributions only varied from sample to sample as would be expected for a random sample from a fixed population of sentence length (Wake, 1957). He thought that this was true for the Greek language as well as English, and that provided certain textual uncertainties were avoided, sentence length distributions were useful in authorship studies and prevented the excessive subjectivity usually present in more traditional methods, such as the use of peculiar marker words.

2. Studies of the 60's

By the 1960's, the development of statistical techniques and the increasing availability of data for comparison was gradually beginning to build the study of stylistics into a discipline in its own right, and the measurement of so-called "stylistic variables" caught the attention of many scholars of literature. The most notable investigations of this time include Alvar Ellegard's (1962) study of The Junius Letters, Frederick Mosteller and David Wallace (1964) who studied the Federalist Papers and the work of Andrew Morton on the authorship of the Pauline corpus in the New Testament.

The Junius Letters are a series of political pamphlets written in 1769-72 whose authorship was one of the mysteries of English literary history. Forty people had been suggested as possible authors. In 1962, Ellegard began to investigate the problem from a stylistic point of view by taking 500 words and expressions which he found to be characteristic of Junius and used it, plus an examination of the choices made by Junius of 50 pairs or triplets of approximate synonyms, to calculate the characteristics which Junius shared with each of the other candidates for authorship. It turned out that there was a remarkable agreement between the habits of Junius and a Sir Philip Francis. Through these constancies within the two texts, Ellegard concluded that Sir Philip Francis was most likely to be Junius.

Mosteller and Wallace (1964), used a more elaborate method than Ellegard, to determine authorship of the Federalist papers. These were documents written in 1787-8 to persuade the people of New York to ratify the Constitution. There were two possible authors, Hamilton

and Madison. As the two were using the same pen name, it was not unlikely that they may have consciously tried to imitate each other. Therefore, Mosteller and Wallace used features which could not easily be copied, such as sentence length. It seemed more probable that the writer would be unaware of small but consistent changes in his use of words which have relatively high frequencies. Mosteller and Wallace had access to works which were known to be written by each of the two possible authors and from these, scored words depending on the number of works in which they appeared. They located words such as "enough", which appeared in nearly all of Hamilton's work but in none of Madison's. Their technique involved the use of Bayes theorem (which enables initially assessed odds on an outcome to be constantly revised as more evidence is added) to calculate the final odds on the authorship of each essay of the Federalist Papers. From such evidence, it was concluded that Madison had written the disputed papers, with high frequency function words being the best style markers.

3. Move to Study of Frequent words

It was at approximately this time that stylistic studies underwent a shift in alignment. Scholars had concentrated more on what were termed "marker" words, that is, words which made one author stand out by his comparatively frequent use of them. It had been discovered that such words were few and far between, thus stylistic analysis began to look for frequent words which were used by most authors, but at different rates. Words such as prepositions, conjunctions, articles - 'function words' whose frequency is almost independent of context - were deemed suitable. Mosteller and Wallace (1964) were pioneers in this approach, which allowed far more sophisticated

statistical techniques than the simpler marker word method and enabled conclusions which were "beyond reasonable doubt" on certain authorships. The introduction of computers at this time also enabled more complex quantitative procedures to be performed.

4. Quantitative analyses of text

Following the invention of the computer, more elaborate quantitative techniques could be employed than the early statistical methods which though relatively simple, required laborious counts of variables within a text. These techniques involved the same basic ideas as previously employed; frequency distributions of words of different lengths, sentence length and combinations of various mathematical analyses with content analysis, but manipulations could be extended further and be made to envelope for instance, more complicated patterns of word positions and more intricate word relations due to the increase in speed and efficiency gained by the use of the computer.

A study which shows a wide range of methods of this sort was undertaken by Stratil and Oakley (1987). In this study 16 different computer analyses and 8 clustering methods were applied to three plays attributed to the Spanish playwright Tirso de Molina. Attribution of authorship to Tirso was only certain in one of the plays. The methods they used are grouped into 6 streams of investigation which serve to illustrate some of the methods used under this heading. These are: a) sentence length and form analysis, b) word length and distribution analysis, c) total word count analyses, d) word frequency analyses, e) word content and contextual analyses, and f) cluster analyses.

Their first step was to calculate the average sentence length for each of the three plays. When these averages were compared with plays by another author (Calderon), they were consistently lower. Also, the two disputed plays of Tirso were closer in average to the undisputed play than to the other author. When the sentence length tables were plotted in graph form their trajectories were very similar in two of the suspected Tirso plays apart from differing levels at one point, which the authors claim may support a the theory that Tirso originally wrote the third play, but that it had been doctored. The same path was taken by all three plays with the same pattern of peaking but at a different level. This, according to the authors, would indicate the same author but a different period.

The second method used word length as a distinguishing variable. The authors were interested in the positioning of words of a certain length. A computer program noted for each word, its length, its position in the sentence and the length of the sentence. When a position analysis was done on all words of the same length, no regular pattern emerged, but when this was recalculated for the total number of **running** words (*Average length of word appearing in a particular position over the text examined)* and not the total number of words of a particular length, a strong pattern emerged. A graph of word length in a particular position in the sentence was plotted and this proved to be very similar for all three plays. The authors felt that the slope of this graph for a particular position may indicate the structure of the author's vocabulary, and the syntactic properties of his stock of words. This result, they say, would seem to indicate that the same person had written all three texts.

The total word count analysis would seem to back this up. It produced a very standard pool of vocabulary and showed stability in the range of word length over the three plays and also in the proportion established between the length of the text and the size of vocabulary. A word frequency analysis revealed that the top thirty positions were taken up with only 44 words over the three plays. When these were compared with the frequencies of words in 55 different Spanish plays (taken from a dictionary of this), only just over half of these 44 words were in the top 30 most frequent dictionary words. Thirty percent of these high frequency words were not included in the high frequency band in the dictionary. The authors concluded that there was a very individual use of words within the three plays.

A programme devised to examine word content calculated the difference between the word use in the three plays, taking the undisputed play as a basis for calculation. The authors found significant differences between the plays for two words in particular. These could be partially explained in terms of tense of the plays, and also by the supposed doctoring of one of them, but the authors admit that these explanations are not entirely satisfactory.

The final analysis was a series of 8 cluster analyses which were applied to 8 different Spanish plays, including the disputed texts. The attributes tested pertained to variables such as sentence length, word position and repetition ratio. Results were varied but evidenced similarities between one of the disputed plays and another of the 8 tested. This study despite having statistical drawbacks, does however throw up a variety of methods for text analysis.

Burrows (1986) makes use of a similar method. He took six novels of Jane Austen and counted the 30 most common word-types listing them in descending order of frequency. The text was divided into pure narrative, character narrative and dialogue, giving 3 groups with 3 respective frequencies. Any words which were associated with the differences in speech associated with the three forms were removed. Eighteen words remained, which made up more than a quarter of the words used in Jane Austen's novels. The correlation between the number of words for each book in each class with the classes in every other book. It was found that pure narrative correlated best with pure narrative, dialogue with dialogue and so on. These results were represented as a graph of eigen vectors whose respective eigen values indicated their relative contribution to the shape of the whole matrix. The 3 groups differed sharply on this graph suggesting that Jane Austen usually makes it possible to distinguish her two main types of narrative.

Burrows then compared the frequency values for pure narrative alone for Jane Austen with some other authors. The narrative was divided into sections of 2000 words and frequencies were compared within books and between corresponding segments of different books. Eigen vectors were again drawn for these results and also for "rolling" (overlapping) values, which override the arbitrariness of the boundaries. The eigen graphs varied in their degree of clustering, suggesting that the separation of the clusters in graphs like these is an expression of authorial differences in narrative style. Finally, he also found statistically significant differences between the frequency patterns of modal auxiliary verbs among Jane Austen's characters, dialogue and narrative and between differing modes of narrative,

within a single literary work and between similar constituents of a different literary work. This evidence was supported when comparisons were made with another author, Virginia Woolf, and also with Another Lady who completed Jane Austen's novel "Sandition".

Damerau (1975) used function word frequency as a measure of writing style, on the basis that these would be a comparatively context-independent set. He assumed a word to be independent of context by using a modified mean test (Lewis, 1965) which transforms data which will be normally distributed if it followed a Poisson distribution. For a sample of 60,000 words and 5 samples from different authors, Damerau made several observations. Firstly, he found considerable diversity between authors on whether a particular word followed a Poisson Distribution. Some authors had many words which corresponded to the Poisson distribution whilst others had very few. However, he did not show that these were characteristic of a particular author. Very few of the words used were randomly distributed in all the samples. On the highest criterion, (Poisson distribution greater than 0.01) only 4 words qualified as random, (*make, other, the* and *with*) whereas on the lowest criterion (probability of greater than 0.001 in three out of four authors) this could only be increased by 19 words (*ever, first, still, though, went, all, an, any, away, been, before, but, into, much, now, another, came, off,* and *these.*. Damerau concluded that many of these words will not be useful for authorship studies and since no within author comparisons have been made " the search for minor encoding habits as indicators of style and authorship should apparently be pursued in other places" (Damerau, 1975).

5. Collocational and Positional Methods

The drawbacks encountered in the quantitative methods thus far, meant that authorship research took another turn. This time in the direction of methods which involved word position. The bulk of research on collocational and positional stylometry has been carried out by Andrew Morton and Sydney Michaelson. Their interest in these methods originated in an interest in Greek prose, when much of their work was focused on books of the New Testament. Based on Greek, they put forward several ideas for authorship tests such as the class of the last word in a sentence (Michaelson and Morton 1972) and simple binomial classification of key words (Morton 1986).

In the early 70's, Morton and Michaelson began to examine "positional stylometry" (Michaelson and Morton 1973). The authors began to look at the advantages of using word position as opposed to vocabulary as a criterion for authorship discrimination. Also, the additional feature of word mobility was examined, mobility being the whole set of positions in a sentence that a word can take. Word Position in this study was defined by "counting the ordinal position of a word as the first, second and third word, and so on, measuring from the start and end of a sentence, and to measure mobility in terms of isotrophy" (Michaelson and Morton 1973). They stated that sentence length distributions were isotrophic - it does not matter whether they are counted from the first word or the last word in the sentence, since if a word is completely mobile, it will occupy each of the positions in a sentence and have a mean position at the mid-point of the sentence. The mid-point is thus isotrophic. Conversely, if a word seems to occur in a preferred position it is termed anisotrophic.

The authors point out that once words are classed into these two groups, isotrophic and anisotrophic, and these two categories within positional stylometry can be used to demonstrate certain features about the text. Isotrophic distributions, according to the authors, are useful to display the periodicity of a number of occurrences and the heterogeneity of the text. Also, positional distributions will have a comparatively constant variance, due to the fact that it will have a mean of approximately half that of the selected sentence-length distribution, but a variance close to the parent sentence-length distribution.

Anisotrophic distributions are further divided into mildly anisotrophic or acutely anisotrophic. Mildly anisotrophic words show "relatively few occurrences in the preferred position compared to the number of sentences in which these are recorded" (Michaelson and Morton, 1973, p 3), whereas acutely anisotrophic obviously show much more. If the frequency of acutely anisotrophic words is high enough, the authors advocate the use of a cumulative sum chart for the complete inspection of the text. (This is an early use of a technique which Morton takes much further in later studies.) Isotrophic distributions will reveal special features of the writer, while anisotrophic ones, by comparison of their cumulative sum charts will allow the text to be examined for authorship.

In 1972 Michaelson and Morton published "Last Words: A Test of Authorship for Greek Writers". In this they examined different classes of words being used in the last position in the sentence in the writings of Paul on the supposition that word classes rather than word tokens would not be affected by subject or context. The chi-squared test is

used to distinguish observed and expected frequencies of last words in the Pauline Epistles and this procedure is explained in detail. However, there was much criticism on the use of this method and also of the conclusions reached about authorship (see Johnson, 1973). This initial examination of positional stylometry was taken a step further by Morton in 1977 when he began to look at collocational information within a text (Morton, Michaelson and Hamilton-Smith, 1977). Morton et al. assert that measurements of word mobility are not as useful in English as in Greek because of the phrase structure of English making words further apart. He therefore advocates the use of collocational tests so that a word can be placed in its immediate context. Collocations such as the occurrence of AND with THE should vary much more between writers than within a writer. In addition to this, resulting from a study of Greek adjectives, came the notion of proportional pairs of words "in which the occurrences of one word is measured by recording the occurrences of the other" (Morton, Michaelson and Hamilton-Smith 1977), without any linking of any literary quality such as function or meaning. The proportional pairs were therefore purely observational. The pair could consist of two particular words or of one word and a class of word. These word pairs and collocations were termed by the authors as 'habits'.

The collocational and proportional pair method was demonstrated using examples from three authors; Sir Walter Scott, Henry James and John Fowles, who have experienced different changes over their lifetimes, thus providing a basis for examination. Using statistical comparisons neither Scott nor James showed any significant difference over these changes in any of the collocations or proportional pairs used, while Fowles seemed to have some success in changing his style

on one habit. This would imply that comparisons within the same forms are straightforward, but comparison across different ones should be used only with habits shown to be unaffected by changes in form.

Morton et al. explain why the method used here is advocated as opposed to the original positional theory; it allows a very large number of combinations of words and thus habits to be available, and secondly, by the nature of the English language, the positioning of words is restricted and thus it makes more sense to look at the immediate context of a word rather than its absolute position.
A final test of this method is to see how it fares with samples of text which are of doubtful authorship, and for this, the authors look at some Shakespeare, Austen, Conan Doyle, and the case of an apparently fabricated police statement. The investigations drew conclusions using this method which were similar to those drawn before. This they say, statistically upholds their theory.

Morton (1986) returned again to the notion that the position of words can be useful in authorship studies. This time it was the position of words that occur only once in the sample of text which supply the information. The argument is that "the number of once-occurring words varies in a complex manner, the number placed in different positions is characteristic of the author" (Morton 1986 p 1). Morton carried out comparisons of once occurring words in the first position in the sentence with those in either the first or last position. He also used the method which abandons punctuation and uses marker words instead as references for position. His samples are Greek prose, namely the Epistles of the Pauline corpus from the New Testament of

the Bible. Morton claims that these samples cover a wide range of subject matter, and have a large variety of literary form. He also held that this method offers a reliable testing method for authorship. This opinion is much disputed, however, and criticisms of this method primarily by Wilfred Smith (1985), will appear in the next section.

Smith himself (1982; 1987; 1988) was one of the scholars who has studied the play "Pericles" on the assumption that it was not entirely written by Shakespeare. Smith (1988) comes to the conclusion that the first two acts were written by somebody else, the most likely contender being Wilkins. He developed a method in which he used the selection of a pool of words, based on the first word of every speech to discriminate suspected authors. His method involved an analysis of first words of speeches, all other words of dialogue, and pairs of consecutive words such as "all the" and "like a". Smith, using a series of pre-determined rules, employed a series of paired comparisons to select, by means of chi-squared tests, those words or collocations that through their contrasting frequencies, discriminated one dramatist's work from another. He validated his method by applying it to known authors and then applied it to *Pericles*. He showed that acts III-V are more akin to Shakespeare than to Chapman, Middleton, Jonson, Webster, Tourneur or Wilkins. Acts I and II on the other hand, proved to be more akin to each of the other dramatists except Webster on the First words of speeches, while all three analyses established a greater affinity of Pericles I and II with Wilkins. Smith concludes that Wilkins is more likely to be the original author of the first two acts of Pericles than Shakespeare, but that Shakespeare may well have added to them some parts of his own, as well as writing the last three acts.

Further evidence that the first two acts of *Pericles* were written by Wilkins comes from Jackson (1991). He used counts of 16 function words which were found in dialogue passages of 20 known dramatists. He firstly showed that the variation between authors, in the rates for the selected function words, was greater than the variation within author. His next analysis was based on the rank ordering of function word rates of *Pericles* as compared by chi-square to rates of other plays. Samples were ranked according to their likeness in use of function words to either section of *Pericles*. Simply, his evidence pointed to the fact that Wilkins was more likely to have written the first two acts of *Pericles* than Shakespeare according to the chi-squared values. Despite this accumulating amount of evidence which supports the theory that Wilkins did write this section of *Pericles* (see also Taylor, 1987), there are still studies which appear to show that Pericles in its entirety is by Shakespeare (Morton, 1978; Merriam, 1982).

6. Criticisms of Collocational and Positional Methods

Much criticism has been aimed over the years at the methods put forward by Andrew Morton. Johnson (1973) takes a critical look at the method used by Michaelson and Morton (1972) when they consider the last words in the sentence of Greeks text. His criticisms lie mainly in the authors application of the Chi squared test; their lack of specification of a hypothesis, misuse of the Yule correction factor, over consideration of quotations and inconsistent calculations of the expected values for the "Pauline expectation" (Michaelson and Morton, 1972). Johnson finds that if the figures are recalculated without these fundamental misconceptions, then the conclusion with this method should be "that II Corinthians was homogeneous (against Michaelson

and Morton) and that Romans was not homogeneous" (Johnson, 1973, p99). Johnson points out the this latter conclusion would have been as evident from their figures as from his own and therefore it is mysterious that the authors failed to draw it. However, Michaelson and Morton, unable to accept the conclusion that Romans was not written by Paul, make a special plea that 'the epistle is not a free composition of the authors', implying that Paul had perhaps incorporated someone else's work. Finally, Johnson suggests that there may be four things to learn from these findings: that the sentence may not be a good unit of measurement, that context may be affecting the words that appear at the end of Paul's sentences, that it is invalid to treat the passages of writing as samples from an infinite population, and that word categories have not been sufficiently well defined.

One of Morton's primary critics, Wilfred Smith, (1985) provides a fair criticism of some of the methods employed by Morton in his book "Literary Detection" (Morton 1978). Smith examines in great detail the three sections of the book, looking at the samples chosen by Morton and the conclusions he has come to using particularly, the chi-squared test. Smith mainly concerns himself with the type of tests used by Morton, his lack of adequate documentation of the samples he has used and the variation in the size of samples he uses. In short he summarises by saying that " scientific standards have not been adopted either by stating with precision the aims and objectives for each section of his work or by analysing critically the success or otherwise of their fulfilment" (Smith, 1985, p355).

Smith (1987) investigated Morton's claims about the discriminatory powers of words that appear once in a sample. (Morton, 1986).

Morton looked at once occurring words at both the beginning and the end of a sentence. He also investigated once occurring words preceding or following selected high frequency words. Smith's first criticism is one which has been consistently pointed out not only about Morton's work, but about many authorship studies and that is the problem with the use of the Chi-squared test as a test of homogeneity between two samples. He points out that from the data available, a Type II error is not calculable *(not rejecting the null hypothesis when it should be rejected)* therefore there is no evidence for the identity of authorship.

Secondly, Morton tries to establish the minimum actual and/or expected number of occurrences of a particular word for testing. He says that the minimum number expected is accepted to be not less than 5 for the size of contingency table that he uses (2x2), but Smith points out that there is some confusion in the testing as to whether the value 5 is being used as the actual or the expected value. Morton does in fact appear to use this value for both. Such small values for the actual values, he points out, are not really reliable to reveal authors' tendencies.

Thirdly, Morton appears to have a problem in that he has only examined works by one author. There should therefore, be no reason to believe that for any sized sample, the numbers of once occurring words in a particular position can distinguish **between** authors. Finally, Morton receives criticism for his choice of positions in which to examine words. Smith points out that only a small number of sentences include a once occurring word which can occur in the first

and the last position. Sentences which do have them are usually quite far apart and therefore cannot be very good candidates for comparison. On the whole, Smith claims that Morton makes some very far-reaching conclusions based on very little data and that Morton "did not develop his proposals from a statistical foundation but instead relied on practical verification" (Smith 1987, p145).

Research by Totty, Hardcastle and Pearson (1987) also examines another aspect of Morton's use of his methods - in the testing of validity of claims made by convicted persons. These methods have been used by Morton on written transcripts of confession statements to compare disputed utterances with accepted ones. The authors express some concern over Morton's choices of collocations when examining material of this type, and warn about the effects of context.

In particular, the authors examine a case tested by Morton in 1978 : the "St. Germain Case". Certain utterances supposedly made by Germain were denied. Morton therefore examined the questioned statement along with 2 personal letters of the accused. He found consistent habits within the accepted samples and compared these with the disputed parts of the statement. Forty habits were examined and 25 of these were found to be different in the disputed statement, giving a significant result.

Totty, Hardcastle and Pearson questioned whether these results were meaningful and examined further written utterances produced by St. Germain. They examined the habits determined by Morton for the accepted utterances and compared them with a later document. Only 2 habits differed significantly - not enough to conclude any more than

a chance difference between the documents. There was the additional benefit in that there was no way that St. Germain could have known that the documents would be compared, and therefore it was not possible that he could have attempted to write with the same habits.

Totty et al. say that this supports Morton's methods to some extent, as St. Germain seemed to use the same habit three years later when the second document was written. They do point out however that other habits which were not examined may have changed. On the basis of this case, it is recommend that stylometry of this sort is useful only for the comparison of texts which are approximately the same size.

Using these constraints, the next test which Totty et al applied was to compare St. Germain with the writer of another statement using 16 common habits. They found 3 of these to be significantly different. Finally, they compared common habits of St. Germain and Smith and a third writer "Brown". Taking part of Smith's statement to be disputed, they found differences between St. Germain and Smith which were not picked up before. When comparing all three authors they again found only two or three differences which could have occurred by chance. The three authors were therefore not distinguishable.

7. The Cusum Technique

The Cusum technique was first used as a method to determine authorship by Morton and Michaelson in their paper "The Q-sum Plot" (1990). The technique was revised in "Proper words in Proper Places (1991). The technique was originally used as a method of industrial quality control in oil production (Woodward & Goldsmith,

1964). It involves plotting "the cumulative sum of the differences between a series of observations and the average of all the observations" (Morton and Michaelson, 1990). When applied to oil production, it enabled a check to be kept on the oil output to see whether the fluctuations were maintaining an output around the average, or whether there was any overall fall or rise.

When applied to text, these observations can be almost any set of countable and frequent variables within the text. The basic measurement which Morton uses, to which all other observations are compared, is the number of words in each sentence, or sentence length. Other variables Morton uses include number of 2 or 3 letter words in a sentence, number of 3 or 4 letter words in a sentence, numbers of initial vowel words (or words beginning with a vowel) numbers of nouns, verbs or prepositions. In fact he maintains that practically any variable within a sentence can be used, as long as it occurs frequently within the text.

Once these variables have been chosen, the cumulative sum process is as follows:

(a) Choose a block of sentences of a set length, say 25 sentences. (Morton maintains that 15 sentences is about the minimum usable for the technique, but he himself tends to use blocks of 50).

(b) Record the sentence length or number of words in each sentence.

(c) Find the average length of sentence for the block by dividing the total number of words by the number of sentences.

(d) Subtract this average, in turn, from the number of words in each sentence.

(e) Add these differences in succession to obtain cusum values.

The same procedure would be carried out on a second variable chosen from the text, for example 2 or 3 letter words. Once the cusum plot has been obtained for the second variable, the y-axis of one of the two graphs is adjusted so that they are comparable. This is done by simply dividing the range of the sentence length graph by the range of the other variable and multiplying the second variable by this number. The two graphs produced can then be superimposed. Morton (1991) also advocates that the incidences of two of the chosen habits can be added together and compared with sentence length in order to enhance the use of the habit. The cusum technique as a whole is therefore a comparison of two cusum plots of sentence length and one or more habit variables.

The claim made by Morton is that if the two variables are chosen from one piece of text, the two graphs will superimpose exactly. If however, the cusum values of two variables of one author are inserted into the cusum plot from an extract of another author, then the graph obtained will diverge, for a certain number of observations, at, or close to, the point of insertion.

The cusum technique is the most recent method put forward for use in authorship studies. It is to this technique that a large part of this research is directed, and thus it will be explained more thoroughly later.

8. Miscellaneous Authorship Studies

An interesting and original method which at first looks like it may have implications for authorship studies was tested by Baker (1988).

He examined the frequency with which an author generates new words. This frequency is based on vocabulary richness as a measure of authorship 'pace'. A sorting program was devised which allowed the text to be counted and new words set apart giving a count of the total number of words in the text plus a count of the rate of new words and their vocabulary types. The "PACE" index was then calculated as actual vocabulary/total word number or in numerical expression:
Pace = 1/Type-token ratio

It was possible then to calculate this ratio for many texts and arrange them in order from highest to lowest pace value. Baker maintains at first that pace value is extremely characteristic of an authors style, and that the same author will have the same pace value. Some authors in his study did however turn out to have very variable pace values (e.g. Marlowe), thus implying that this is not as characteristic as first asserted. Baker says then that the pace value would perhaps be more useful in measuring the maturity within an author, than making distinctions across authors.

B. Psychology and Writing Style

How should the psychologist address the problem of writing style and how does it help the study of the human mind? If it is possible to identify an individual by means of his writing style, in the same way in which a fingerprint identifies him, then the psychologist may have an external handle in the workings of the mind.

The value of the many statistical techniques quoted can be seen from two angles. On the one hand, it is undeniably useful to be able to tell the difference between two pieces of text merely by looking at a selection of variables. By measuring and counting the so-called stylistic traits such as frequency of words or vocabulary items, it may be possible, it seems, to discover some regularities and general characteristics. A statistical theory will also allow a more compact description of the quantifiable features of style, and will give some idea of how much confidence we can place in these generalisations, when presented with only a small portion of text. A statistical theory will also assist in making systematic comparisons between the stylistic habits of two authors, and may contribute to the solution of traditional problems concerning authenticity and integrity.

Despite the above, there is only so much information available through these variables. Many features of writing style such as the lengths of words or sentences, or the number of words in a poet's line, *are* quantifiable and straightforward in an unproblematic way. However, as Kenny (1982) points out, it is possible to study other features such as vocabulary, or preferences for different parts of speech, but intuitively these do not seem as likely to show individual differences, nor are they

as easy to quantify. What do these features tell us, and what vital information about the mind of the writer are we missing if we ignore these deeper aspects in favour of a more superficial, yet statistical measurement of style?

It must also be noted that some methods of statistical analysis, such as the Chi-squared test favoured by Morton, take no account of the serial nature of language. Sentence length distributions or word frequency lists do not preserve the order of words or sentences. It is intuitive that the authors choice of one word affects the next, since indeed some words can only be followed by particular others. The fact that the great majority of research published on stylometry does not take into account this order effect does not invalidate it, however. It simply means that there is still much to explore in the statistical study of writing style.

The crux of the matter for the psychologist lies in the question of *why* the writer writes in the way that he does. In other words, the aim of a psychological study of style, is not *what* is written, but *why* it is written in that particular way. Why does the author use habits which are so individual to himself? Many questions inevitably stem from this. For example, can people consciously change style, or is it impossible to disguise your own individual style once it has been established? Could particular life events alter the style with which you write? Morton (1990) would maintain that an individual cannot change his style, nor would any life events be able to disguise it. Does this mean that once a person's style is established, it is with him for life? What, therefore, constitutes the actual evolution of stylistic habits?

The least the psychologist can do is to go some way into exploring these questions. Iris Murdoch was quoted saying as in an interview that her philosophical works were radically different from her novels, yet when both were analysed using Morton's Cusum technique, (Morton 1990) they yielded a "fitting cusum". Nor did there seem to be any change longitudinally using this method. Moreover, when Morton examines Sir Walter Scott's works between 1816 and 1832, he found no appreciable difference despite the fact that Scott suffered a number of major life changes within this period. The ultimate question then, is whether or not there is a level of individual constancy within writing which is on a par with a fingerprint .

C. Application of Techniques within the Legal System.

Of what practical use are the applications of these techniques to determine authorship? The main application of stylometry over the years has been in either the detection of the author of a piece of work whose identity is unknown, or merely suspected, or, a comparison of two pieces of text to determine whether they were written by the same or different persons. When stylometry is used in this way in a legal context, it is termed "forensic linguistics". Black (1990) defines forensic linguistics as "a technique concerned with in-depth evaluation of linguistic characteristics of text, including grammar, syntax, spelling, vocabulary and phraseology, which is accomplished through a comparison of textual material of known and unknown authorship, in an attempt to disclose idiosyncrasies peculiar to authorship to determine whether the authors could be identical" (Black, 1990:648). It is obviously invaluable to be able to do this, both from the linguistic point of view, and also in integrity studies.

The potential value of these procedures, is obvious within legal systems. In Scots law, in order to establish the guilt of the accused, the Crown must prove facts crucial to the charge by " the direct evidence of two witnesses or two or more evidential facts spoken to by separate witnesses from which the crucial fact may be inferred, or of a combination of the direct evidence of one witness and of one or more evidential facts spoken to of other witnesses which support it " (McPhail, 1987; 23.01). This is the doctrine of corroboration without which the sufficiency of proof, beyond reasonable doubt, cannot be established. Due to the nature of the investigation of crime, the investigation of crime often rests on the evidence of two or more policemen who assert that the accused has said something which he denies ever having said. In cases like these, stylometric techniques have been presented before courts to be used as a method of discriminating between genuine and fabricated confessions to the police. There are obvious problems with this however.
In the so-called "Judges' Rules" it states that "in writing down a statement, the words used should not be translated into 'official' vocabulary; this may give a misleading impression of the genuineness of the statement." In other words what is to be recorded is the actual words of the accused. The psychologist knows that how an individual speaks is seldom how he writes, since speech is full of hesitations, repetitions and incomplete sentences. Research in this area includes that on the Map Task corpus (Anderson et al, 1991). A statement must usually be an edited summary of what took place.

Psychologists have maintained that it is not possible for one person to make an accurate record of what another has said. There is usually a conflict of interest between the law and the accused, leading to an almost certain misunderstanding. Even at best, the two parties

will have different backgrounds, experiences and interests, enabling quite different interpretations. With respect to these problems, it may be the case that a police statement could be of no use whatsoever when it comes to stylometric analysis, as it could well reflect the style of the policeman rather than the accused.

It is maintained however that it is the one rule "use the words of the accused", which swings the balance. It is the trivial habits which no police officer thinks to be of any significance which are important in stylometry, and on which the stylometric tests are based. If the policeman is honest, then he will write down what the accused has said, whereas if he is dishonest, he will write down something which has not been said. However, the policeman may see no harm in changing some words in order to make a statement grammatical. Other foreseeable problems could arise, for instance, when it is necessary to read back a statement to an illiterate man.

Not all law courts therefore, will accept evidence based on such techniques, indeed, even some of those most familiar with statistical authorship studies would have misgivings about using statistical stylometry, at the present stage of its development, for forensic purposes.

1. Use of Forensic Linguistics in Civil and Criminal Contexts.

Despite the drawbacks of using linguistic evidence within the criminal law context, there are many documented cases of its use in both Britain and the USA.

In the USA, Rule 901 of the Federal Rules of evidence is " the Requirement of Authentication or Identification" of text that is to be used as evidence. Section (b) of this rule provides for expert witness comparison of writings plus illustration in conformance with the requirements of the rule which allows for authentication based on stylistic analysis. (McMenamin 1993).

In Scotland, the accused is entitled to apply to the court for a warrant to inspect any "productions" which are to be used at his trial. This may include evidence such as police statements, and when there is any dubiety about these productions, the accused is entitled to submit them to an expert for scientific examination. This was first applied in the case of William Turner Davis (Petitioner) in 1973 and quoted in the Scots Law Times 1973 (p36).

The courts have admitted stylistic evidence on document format, spelling, capitalisation, abbreviations, punctuation, word choice, syntax and content. Distinctions have been made between linguistic evidence of authorship and "mechanical" evidence of typist identification but there is considerable overlap between the two. Authorship evidence can include cases where a certain typing technique is used, but more often it has referred to distinctive characteristics of the author's use of language itself.

For instance, the case of Kane's Estate (1933) in the USA used authorship evidence on typing technique . This used the fact that a particular typist used the capital "I" for the numeral "1". In an action to contest a will it was alleged that the typist had typed a second will, but other documents by the same typist showed that the typist used the lower case letter "l" to produce the numeral "1". The action was subsequently quashed.

One of the earliest cases to use distinctive authorship characteristics in the USA is quoted by McMenamin (1993) in the case of Pate vs. People (1846). The prosecution witness "Randall" established that " he did not execute a receipt and contract by demonstrating that their style of writing and spelling differed from his own". The accused attempted to contradict this by introducing to the court documents which had been written by Randall and which contained writing and spelling similar to that in the questioned writing. The court allowed this evidence and it was established later that the documents introduced by the accused had been doctored to resemble the questioned writings of Randall.

Thus there have therefore been cases based on forensic linguistic evidence since as the early 19th century. In the early 1990's in Britain, the Cusum technique, which is dealt with thoroughly in Part 2, began to be used extensively, with varying degrees of success, on the side of the defence in appeal court cases. It had been used prior to this in Australia and is now also being used in Ireland (Expert Evidence, 1992, vol. 1). Both Morton, the principal exponent of this technique, and his colleague Farringdon, have appeared consistently as expert witnesses on the side of the defence, in an attempt to establish reasonable doubt on the authorship of confession statements. In the

case of Regina vs. Beck which took place in the Crown court Leicester in November 1991, Morton prepared Cumulative Sum charts of several witnesses which were to be called by the prosecution and claimed that their statements came from a number of sources rather than from a single witness.

In R vs. McCrossen, in July 1991, the Court of Appeal in London admitted expert evidence based on the cusum analysis. Cusum analysis was used to pour doubt on the author/s of the alleged confession by comparing samples of writing of the accused with samples from the police officers who interviewed the accused. Farringdon appeared in the case of R vs. Bourke, in December 1992, were he subjected four statements from a witness to cusum analysis. He concluded that there was more than one author in one of these statements and not all the statements were produced by the same individual. In this case however, the evidence of Farringdon was refuted by Professor David Canter, who also analysed the statements from this witness and concluded that the statements could not conclusively be said to be written by more than one author. Ultimately, the accused was charged and sentenced to 6 years.

The Role of Expert Evidence

In addition to the many questions of the scientific validity of the cusum technique itself, there has been much debate on the legal issues surrounding the cusum technique as expert evidence. Firstly, according to British law, the opinion of an expert witness like Morton or Farringdon on the Cusum technique, should carry very little weight. The following explains the function of an expert witness when giving evidence in Scotland.

"Their duty is to furnish the judge or jury with the necessary scientific criteria for testing the accuracy of their conclusions, so as to enable the judge or jury to form their own independent judgement by the application of these criteria to the facts proved in evidence. The scientific opinion evidence, if intelligible, convincing and tested, becomes a factor (and often an important factor) for the consideration along with the whole other evidence in the case, but the decision is for the judge or jury. In particular, the bare *ipse dixit* of a scientist, however eminent, upon the issue in controversy, will normally carry very little weight, for it cannot be tested by cross-examination, nor independently appraised, and the parties have invited the decision of the judicial tribunal and not an oracular pronouncement by an expert" (McPhail, 1987, S17.10B).

The relevant points here, are that the evidence from an expert witness should have undergone testing and be comprehensible and it should not be relied upon, since it cannot be thoroughly examined through cross-examination.

The second question lies in the fact that cusum analysis has only been admitted as part of the case for the defence. This sort of analysis is not deemed reliable enough to convict a defendant, despite the very high claims of reliability that the cusum technique has. The appropriateness of allowing expert evidence to create reasonable doubt, but not allowing the same evidence to assist the prosecution case, is still being questioned (Expert Evidence, 1992).

The British courts also seem to have adopted the fact that if a scientific test is deemed to be relevant, then it should be admissible in court. There has been doubt as to whether the relevance factor outweighs factors of prejudice on the part of the defence.
The use of cusum charts in the court setting is also debatable. The "expert" who has prepared the cusum chart will have had time to prepare an interpretation of it. On the other hand, the jury who are shown the cusum charts "cold", could be open to manipulation by the expert witness and require more than just the visual evidence of the separation of the cusum curves on which to base an opinion on the authentication of a document.

2. Conclusions

Kenny (1980) quotes an unspecified professor at the University of Michigan, who is familiar with the study of style and statistics, who outlined the difficulties facing the would-be expert witness in stylometry:

"In my view, there are at least three views which define the circumstances necessary for forensic authorship attribution.
that the number of putative authors constitute a well defined set.
that there must be sufficient quantity of attested and disputed samples to reflect the linguistic habits of each candidate.
that the compared texts be comeasurable."

Number 3 is obviously the most difficult to comply with when it comes to courtroom cases, since the differences between, for example, a diary and an oral confession by the same person, are bound to be greater

than those between two persons writing in the same mode of composition.

In writing for a law journal on the subject of whether expert evidence should be based on scientific research, Kenny submitted a four-part test of scientific procedure in the courtroom (Expert Evidence, 1992, vol. 1). He said that findings should be consistent and experts should only disagree on cases that they regarded as borderline. The discipline should also be methodical so that different experts would be able to come to the same conclusions when examining the same evidence. The findings should also be based on earlier work and finally the discipline must be predictive and falsifiable.

If the above guidelines could be met, stylometry could undeniably be of great importance in the legal field. No doubt in future occasions there will be times when the accused in criminal trials will seek to use stylometry to dispute confessions which have been attributed to them, or when those conducting a lawsuit will use similar methods of textual study to contest, for instance, contracts or wills. The discipline may develop far enough in the future to allow this, but at the moment it remains debatable whether stylometric analysis can be held as a high enough criterion for absolute personal identification in legal issues. What remains is that perhaps the existence of the science of stylometry will be enough to prevent the abuse of the confession and statement.

D. Overview of studies of style to the present day

Studies of authorship began in the mid 19th century when they focused initially on individual differences in sentence length and then incidences of "marker" words in disputed texts. Several key studies were made from this time up until the 1960's when a move was made to the study of frequent words which were used by most authors, but at different rates. The introduction of computers made this task easier and techniques were introduced which involved counts words of different lengths and complicated patterns of word position. Large samples of text were analysed using every conceivable statistical method, however, the success rate in the determination of authorship was low.

The poor success rate of these initial quantitative methods encouraged another direction to be pursued. This time, the focus was on collocational and positional methods of investigation. Many studies were carried out on word position, primarily by Morton and Michaelson, but these methods received criticism from many quarters, mainly due to the statistical techniques applied.

In the early 1990's, another method arose, again developed by Morton and Michaelson. It was called the Cusum Technique, and claimed to have a near 100% success rate in author identification. This technique was snapped up by the courts for use in cases of disputed statements and for two years was used without opposition. However, in the last two years this technique too has received a great deal of criticism. The first part of the research in this thesis in directed at the claims of the cusum technique.

E. Chapter Conclusions

The basic notion behind many of the stylistic studies up to the present day, has been that there are features within an individual's writing which add up to something akin to a stylistic "fingerprint". Indeed Morton himself uses this simile when explaining the "habits" peculiar to an individual in the cusum technique. A hypothesis which has been less thoroughly explored is one which would identify *style clusters,* or groups of people who use the same set of stylistic variables. A particular set of variables may be unique to a sample of the population but are not capable of identifying an specific individual.
The next section of the thesis looks at the cusum technique in the light of this theory, and attempts to establish whether it can determine authorship uniquely or can merely distinguish groups of people who use a particular habit.

PART 2: THE CUSUM TECHNIQUE

CHAPTER 2: The Cusum Technique: Introduction and Study of Habit Variability

A. Introduction

Very little is known about individual differences in writing style. This is true for the idea that an individual has his or her own personal writing style and that different individuals have different writing styles. The lack of research on the normal everyday styles of literate individuals and constancies in their writing habits is surprising since an understanding of the nature of styles of writing has several practical and theoretical consequences.

Perhaps the most important issue with writing style has been the work on authorship studies, many of which have been described in Chapter 1. (e.g., Kenny, 1986; Morton, 1978; Morton and Michaelson, 1990, 1991). The aim has been either to show one or many hands at work, or to decide amongst a small set of contenders for authorship.

Possibly the prime force in this area is Morton who, by use of the Cumulative Sum Technique, has made the only existing claim that the method he advocates is capable of separating individuals by their writing styles with forensic accuracy (Morton and Michaelson 1990; Morton 1991). The practical importance of this claim is manifest and is therefore the subject of this chapter. Morton's services in cases of legal disputed authorship are currently being sought on a regular basis. Examples of these are the Carl Bridgewater case and the appeal of the Birmingham Six. Since the acceptance of an invalid technique by the courts would inevitably lead to miscarriages of justice, it is imperative

that this method receives systematic, scientific appraisal. Chapters 2 through 4 report an attempt at just such an appraisal.

1. Aims

The aim of this chapter is to provide a detailed explanation of the Cusum process. These claims made about this technique will be investigated firstly using hypothetical data and then using real data from real texts. The fundamental assumption of constant ratio will be examined within individuals, in a study using 5 of the habits proposed by Morton (1990).

2. What is the Cusum Technique?

Morton's cusum technique requires plotting the cumulative sum of the differences between a series of observations, and the average of all the observations (Morton; 1990). In principal, these observations may be any set of variables within the text. The basic measurement which Morton uses, to which all other observations are compared, is the number of words in each sentence, or sentence length (sl). Morton claims that this is because the sentence is the basic unit of language. Variables with which sl is compared include **number of 2 or 3 letter words** in a sentence **(23lw), number of 3 or 4 letter words** in a sentence **(34lw), number of initial vowel words -ivw-**(or words beginning with a vowel), **number of 23lw which are not ivw (ivwx), number of nouns, verbs or prepositions**.[2]

In fact he maintains that practically any variable within a sentence can be used as long as it occurs frequently within a piece of text. The

---

[2] Some of these may even be taken in combination such as 23lw and ivw (Morton, 1991)

technique thus relies upon regularities within an individual occurring with these simple variables.

Once a variable has been chosen, the Cusum calculation is as follows:

1. Chose a block of sentences of a set length, say 25 sentences. (Morton maintains that 15 sentences is about the minimum that can be used for the technique, but he himself, tends to use blocks of 50).
2. Record the sentence length or number of words in each sentence.
3. Find the average length of sentence for the block by dividing the total number of words by the number of sentences.
4. Subtract this average, in turn, from the number of words in each sentence.
5. Add these differences in succession to obtain the cusum values.

These cusum values can then be plotted against sentence number. The plot should end on the horizontal axis, since the last point will be the difference between the number of units in all observations and the product of the number of observations and the average value.

The first cusum plot would therefore be for sentence length. The same calculation is then carried out for one of the habit variables chosen from the text, for example, 23lw. Once the cusum plot has been obtained for the second variable, the y-axis of the second graph is adjusted so that it is comparable with the sentence length plot. This is done by simply dividing the range of the sl graph by the range of the other variable and multiplying the second variable by this number. The two graphs thus produced can then be superimposed. The cusum technique as a whole is therefore a comparison of two cusum plots of different variables. An example of a cusum plot for the comparison of two variables, sl and

23lw is shown in Figure 2.1. An example of a cusum calculation for sentence length can be seen in Appendix 1.

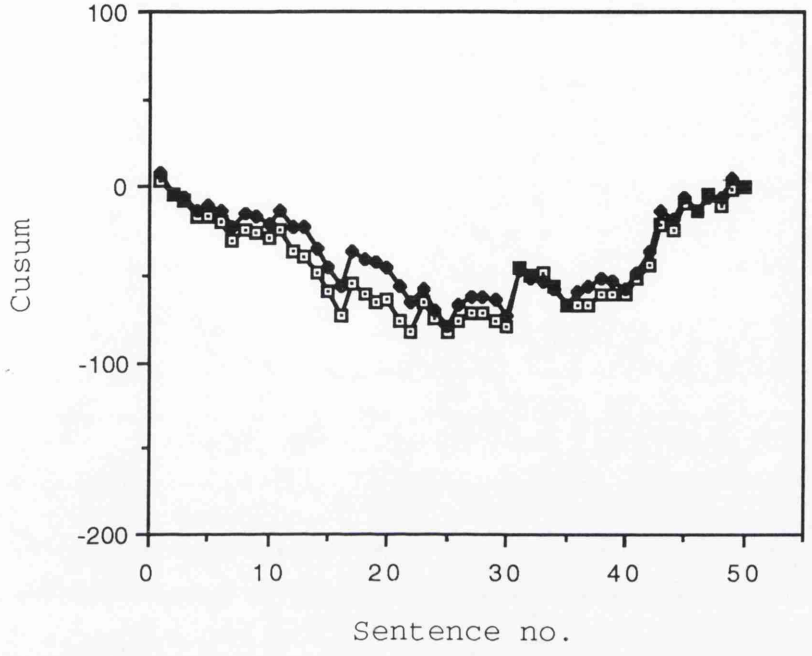


Figure 2.1: Cusum plot showing an apparently good fit for sentence length and 23lw

3. Claims made for the Cusum Technique

Morton claims that it is the constancy of the ratio of sl to the other variable in an individual which gives graphs, such as Figure 2.1, such a close fit. If on the other hand, a piece of text is written by **two** people who have ratios which differ from one another, a single cusum plot like Figure 2.1 will become distorted with the two curves separating, as in Figure 2.2. In this case the first 25 sentences are from one author and the second 25 from another.

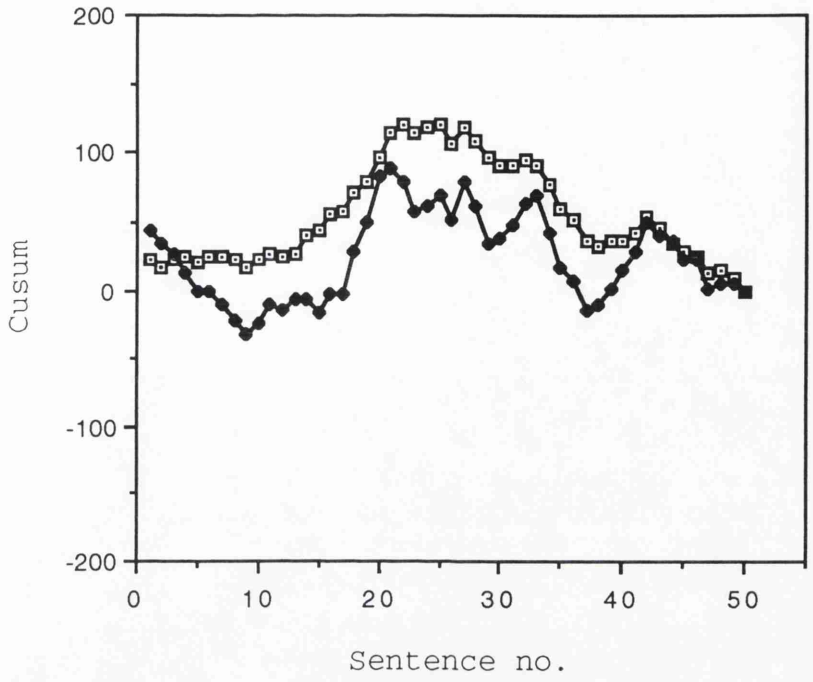


Figure 2.2: Cusum plot showing separation for sentence length and 23lw

So the basic claim behind Morton's technique is that a distortion such as that in Figure 2.2, in a cusum chart is evidence that two or more individuals have been involved in the production of the text. He claims that this has forensic accuracy and likens it to fingerprinting. However , to say the least, this view is merely an assertion, not a demonstrated fact.

Morton's views embody several other claims:

(a) The "habits" must be fixed in an individual. In other words, the ratio of 23lw (or any other variable) to sentence length must be stable and constant for an individual. Morton produces considerable but anecdotal evidence in apparent support of this claim including analyses of literary items from, for example Walter Scott and Thomas Hardy. Examples of some graphs produced by Morton can be seen in Figures 2.3 and 2.4.

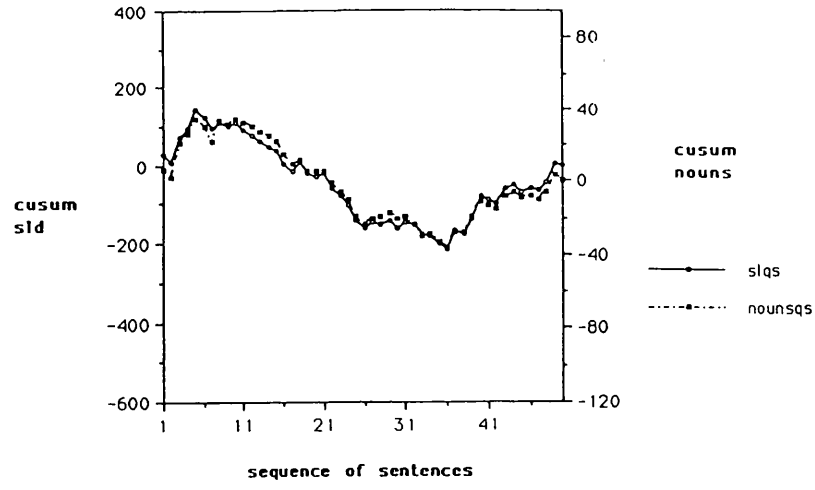


Figure 2.3: Cusum of Sentence length and Nouns in two successive sequences from Walter Scott. From Morton & Michaelson (1990)

This constant ratio assumption is therefore of prime importance. If this the ratio of sl to habit does not hold within an individual, the technique will loose any discriminating power.

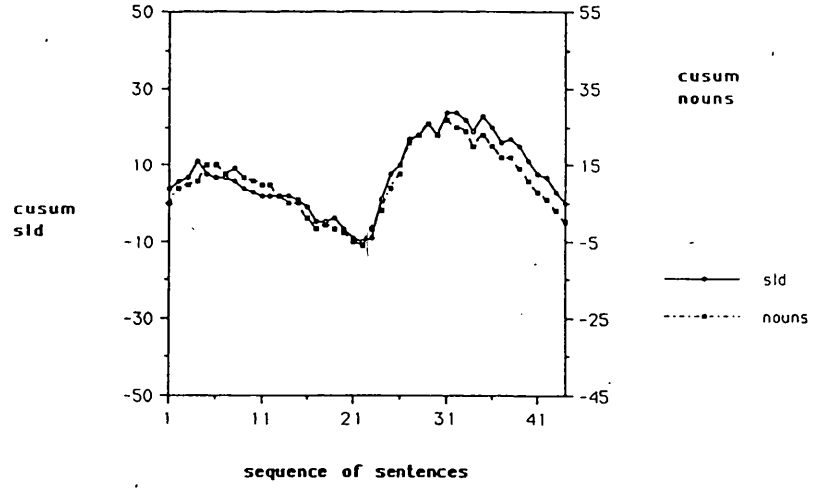


Figure 2.4: Cusum of Sentence length and Nouns in two successive sequences from Thomas Hardy. From Morton & Michaelson (1990)

(b) The ratio holds over genre and modality. Morton claims that the ratios are fixed for individuals, regardless of whether the data are derived from spoken or written material, and regardless of topic. This is important because for forensic applications, data may be produced which is of "verbatim speech" as well as writing from many years previously or recent writing. Morton presents only a limited amount of anecdotal evidence to support this, mainly from an interview with one individual (Iris Murdoch) in which he compares her speech and writing (Morton & Michaelson, 1990).

(c) The cusum technique has discriminating power. The ratios must therefore differ from person to person sufficiently frequently to be useful. Two people may of course "share the same habit" (Morton & Michaelson, 1990) in which case they should produce indiscriminable cusums. The power of the test therefore relies on the idea that if the fit of the two cusum curves is poor, then multiple authorship is implied.

(d) The technique is resistant to plagiarism. Morton claims that if individual A tries to write in the style of individual B, this will still be recognisable as writing from individual A, because the basic habits are ingrained, and this will also be reflected in the Cusum graph.
The Cusum technique would be a very powerful tool if all of these claims held, but a priori, it seems unlikely.

4. The relationship of habit ratios to cusum: Studies with idealised data

We now examine how the kind of relationship between a habit variable and sentence length influences cusum deviations as part of an

understanding of the basic logic of Morton's assumptions. Using idealised (invented) data, the following factors and predictions are examined.

a) The linearity assumption. Morton assumes that a good fit comes from a fixed ratio. A fixed ratio entails a linear relationship between sl and habit variables. Non-linear relationships should therefore produce poorly fitting cusums.

b) The high correlation assumption. If the linearity assumption is met, the cusum fit could still be poor if the linear correlation is poor through noisy data. This is also examined.

4. 1 The linearity assumption

In order to test this assumption, a set of three lines were used which had the same zero intercept on the y-axis, but differing slopes or ratios. An illustration of the lines used is shown in Figure 2.5 below.

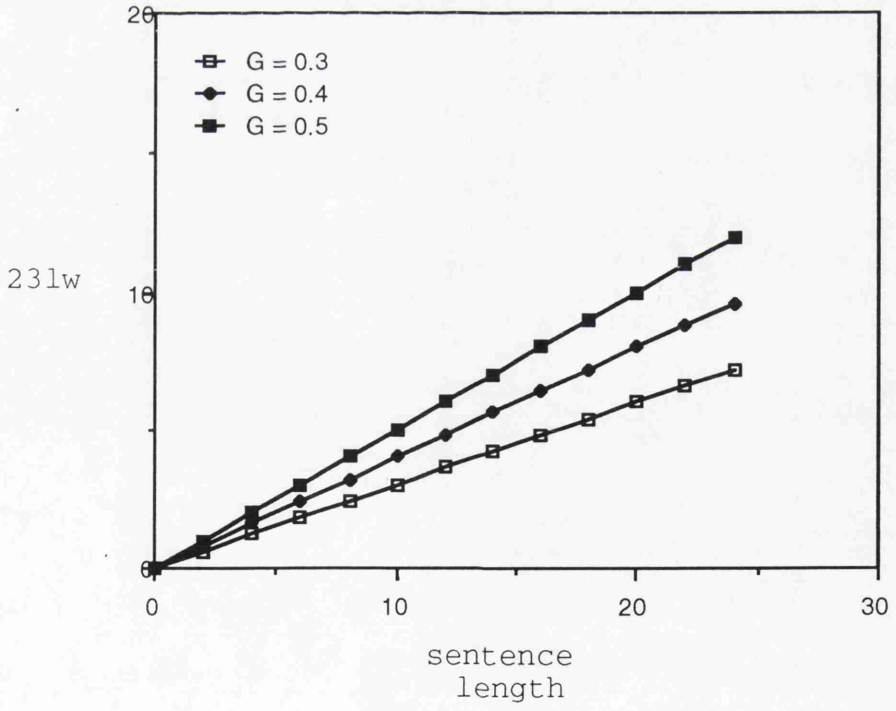


Figure 2.5: Three ratios (Gradients, G) used to plot hypothetical cusums

Lines with ratios of 0.3, 0.4, and 0.5, were used. They were formed using 15 points for each line. These lines represented 3 different sentence length to habit ratios. The data points were, for example, for the text with a ratio of 0.4, a sentence length would be 5 and the corresponding habit would be 2, since 2 is 4/10ths of 5. All three lines passed through the origin.

Three different non-linear equations were also used: $y=x^2$, $y=x^3$, y=root x were also used. These are shown overleaf in Figure 2.6.

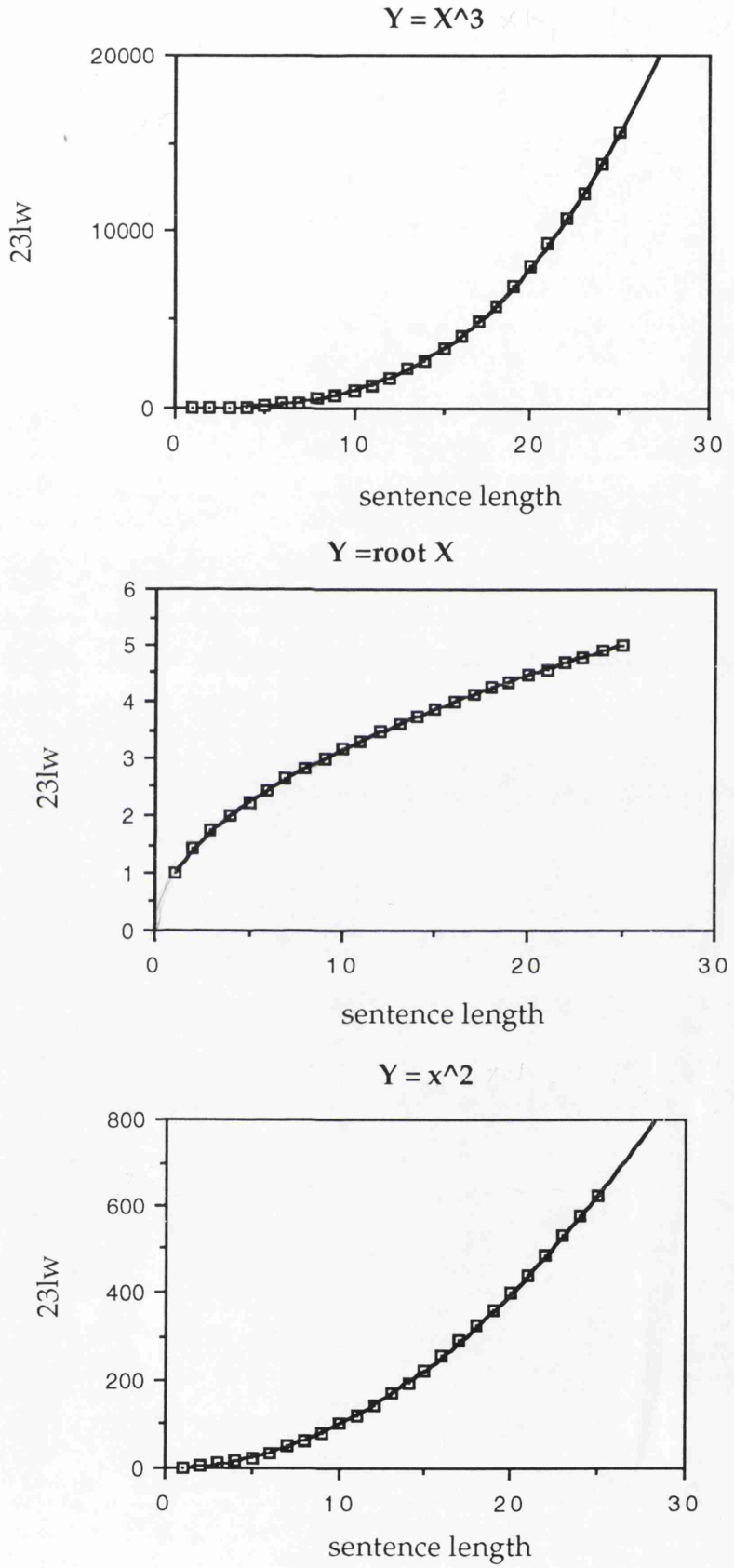


Figure 2.6 : Plots of three non-linear equations

The linear cusums are not shown since, analytically, they all fit exactly. The cusum plot for the non-linear equations yields a systematic deviation between the two cusums. The deviation between the two curves was calculated by summing the difference between the absolute values at each point on the curve. The deviation measures are shown in Table 2.1 below.

Non Linear Equation	**Area Deviation**
$y = x^3$	260.1
y = x2	154.07
y = root x	101.96
linear equation	0.00 (by definition)

Table 2.1: Total Deviations for Linear and Non Linear Data

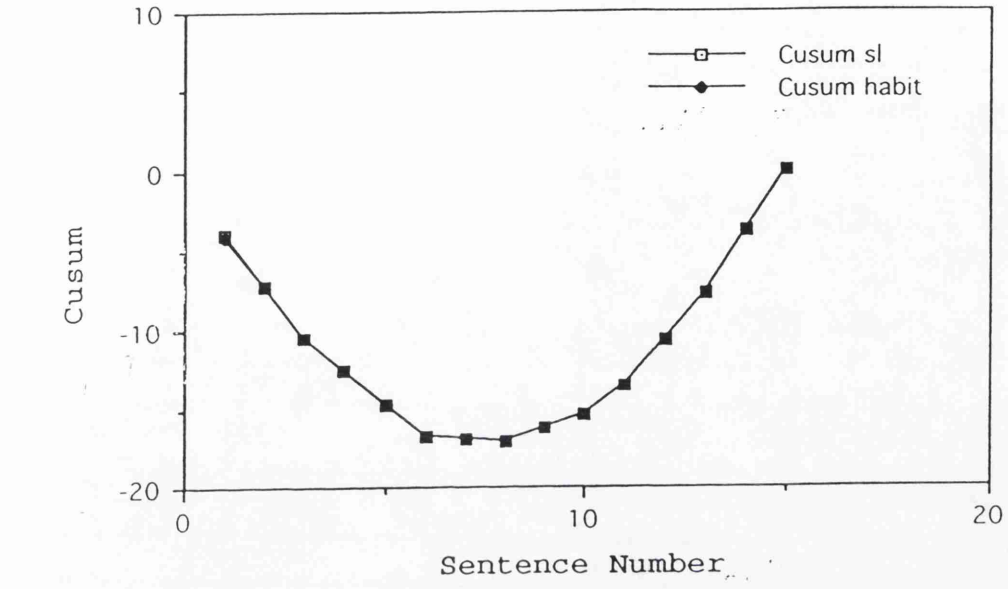
The extent to which non-linear functions produce deviations in cusum depends on how well approximated the functions are by straight lines.

4. 2 The Correlation assumption

Deviations in cusum were dependent on how well a linear function described the data. It should be the case that linear fits based on poor correlations will be a poorer approximation to a linear function than if there is a high correlation, and therefore the fit of the resultant cusum curves will be poor.

Data producing straight line fits with correlations of r = 0.7, r = 0.8, r = 0.9 and r = 1.0 for 15 data points per line, were invented in the same way as the linear equations in the previous experiment. The points were put into ascending order, since it can (and will) be shown that the area measure of deviation from this sort of plot is a monotonic function of the differences in the 23lw-to-sentence length ratios of the constituent

discourse, and so it gets around some of the problems inherent in natural orderings. The cusum charts of these correlations were plotted. They can be seen in Figures 2.7 (a) and (b)



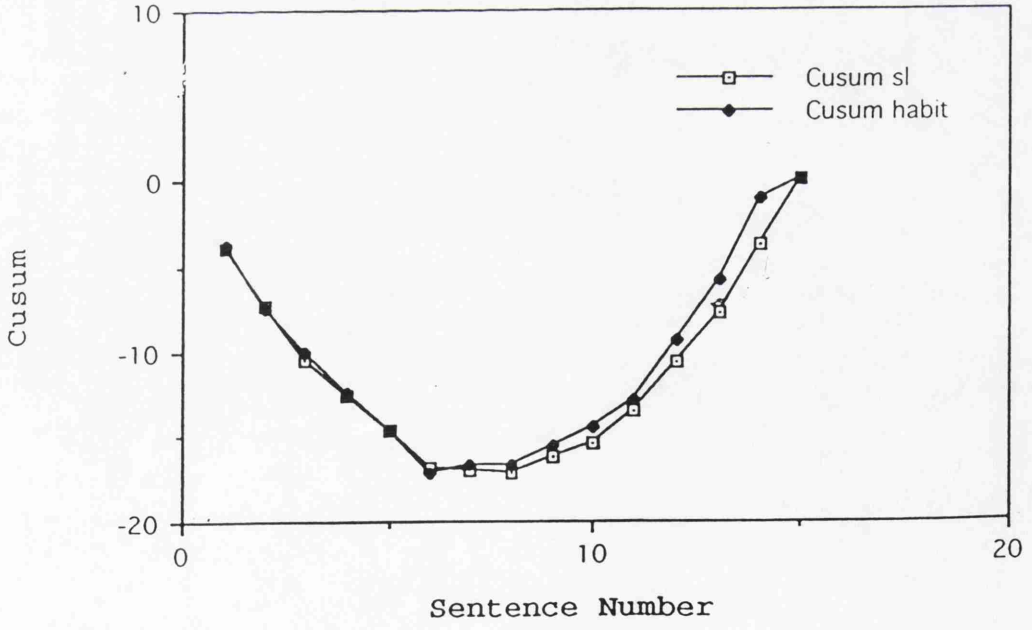
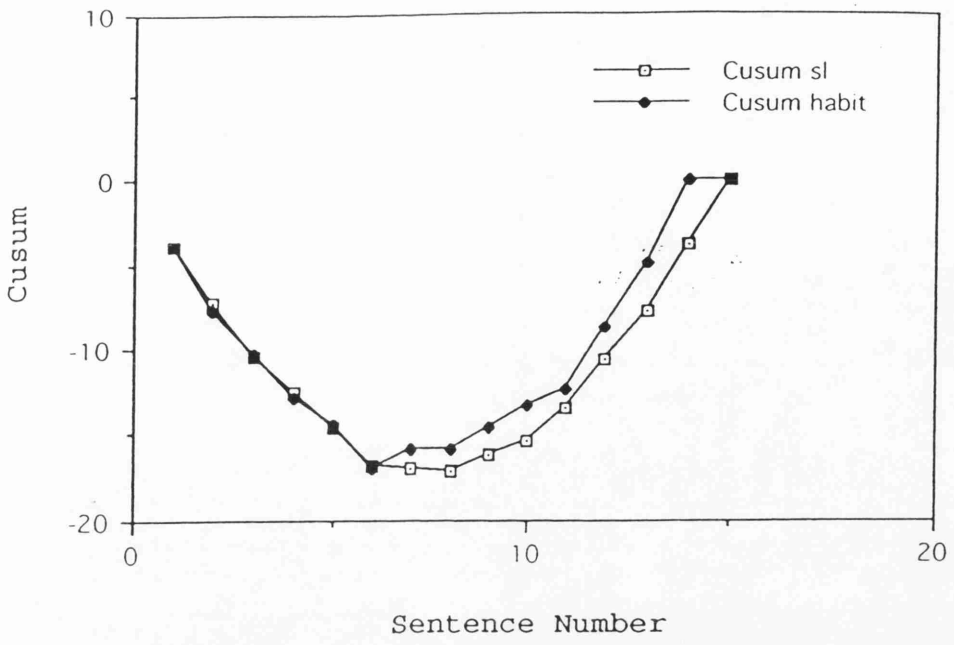


Figure 2.7(a): Cusum charts of correlations of 1.0 & 0.9



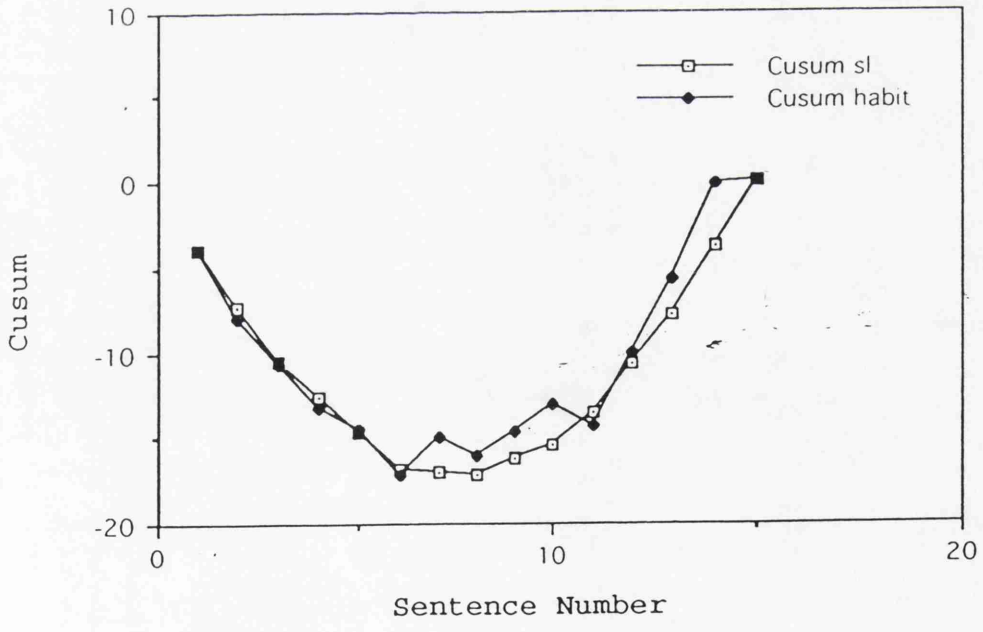


Figure 2.7(b): Cusum charts of correlations of 0.8 & 0.7

The deviations for each graph in 2.7(a) and (b) are shown in Table 2.2.

Correlation	Total Deviation
1.0	0.00
0.9	19.15
0.8	25.74
0.7	42.47

Table 2.2: Area Deviations for Correlations of 0.7, 0.8, 0.9 and 1.0 in ascending order

So, not only deviation from linearity, but poor linear correlations produce a predictable monotonic increase in deviation.

Therefore, if there is a good linear relationship, or constant ratio between the data from the sl of the text and the chosen habit, then the cusum plot will fit exactly after scaling. Non-linearity produces a disruption in the cusum curves and so does poor linear correlation. In the next investigation, we look at whether a good linear fit with high correlation holds for samples of real text. The remainder of this chapter deals with question of the constancy of ratio within individuals, not the cusum per se.

5. Constant Ratios and High Correlation in Real Data

The critical question is, do good linear relationships exist within real data from real text? Excerpts of 15 sentences were taken from a 6 texts written by 4 different authors. Plots of the linear relationship $y = mx+c$ between sl and habit are shown in Figure 2.8 The fits were made using a least-squares procedure.

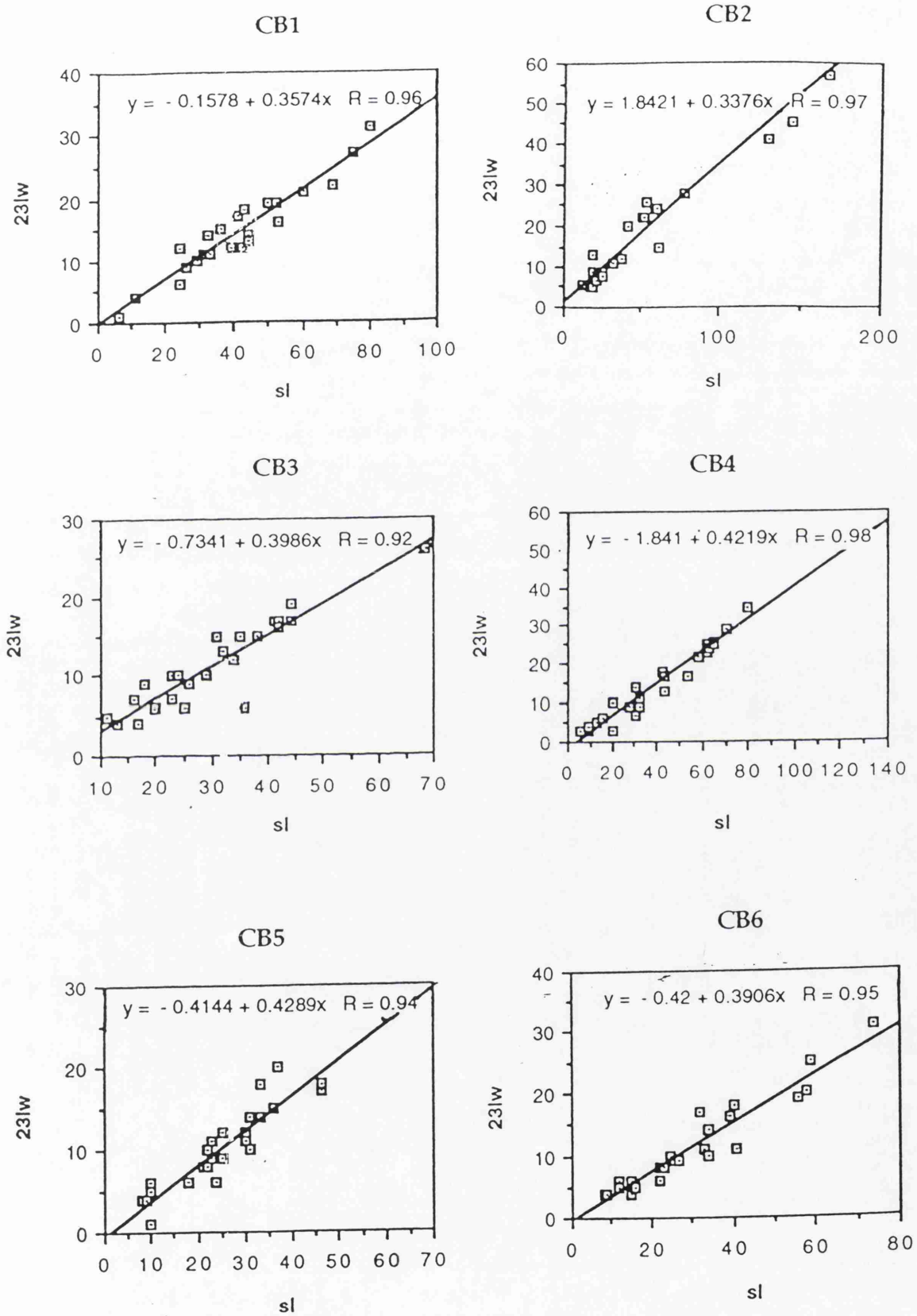


Figure 2.8(a) : Plots of linear relationship between sl and habit (with intercept) for data from real text (author 1)

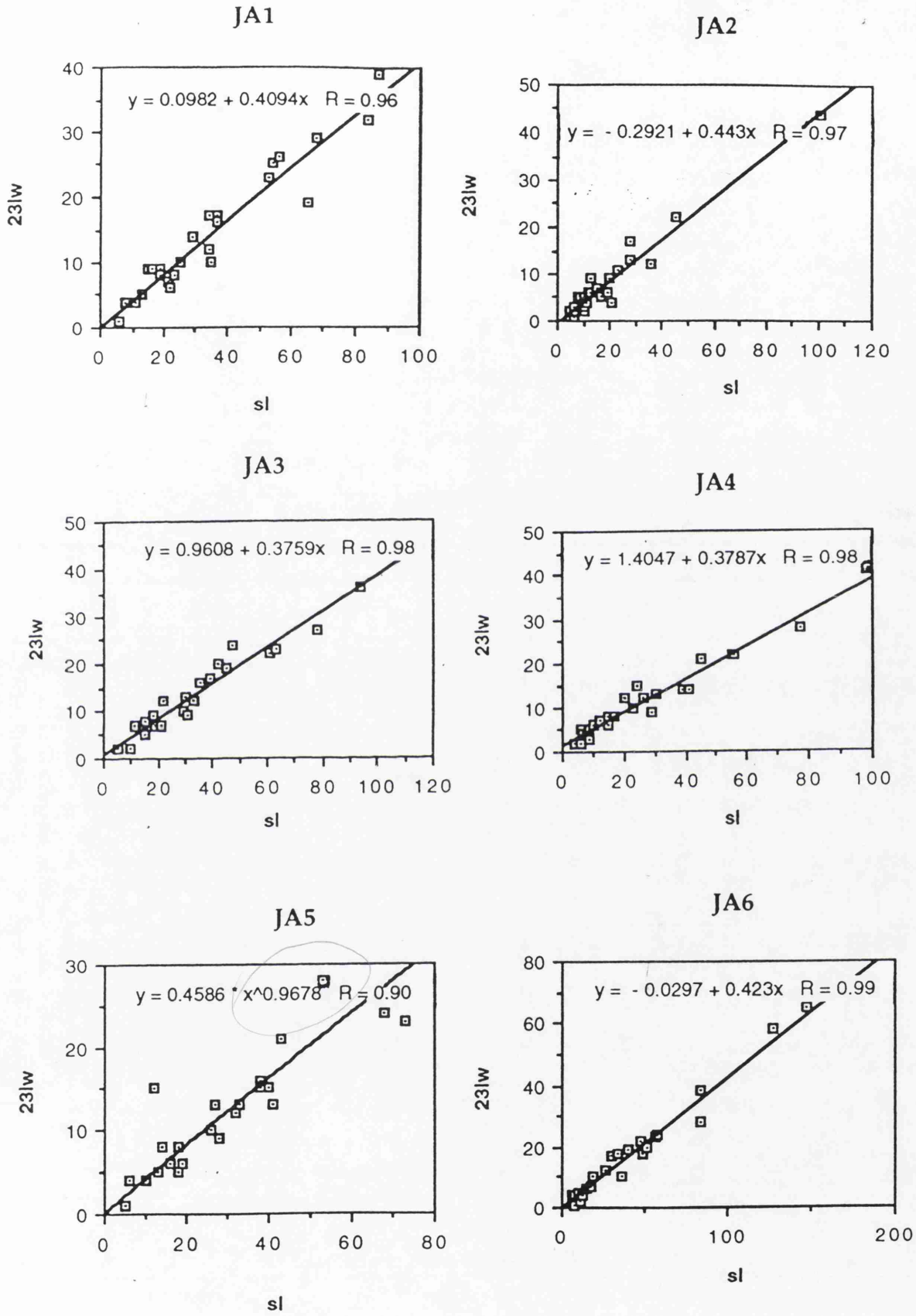


Figure 2.8(b) : Plots of linear relationship between sl and habit (with intercept) for data from real text (author 2)

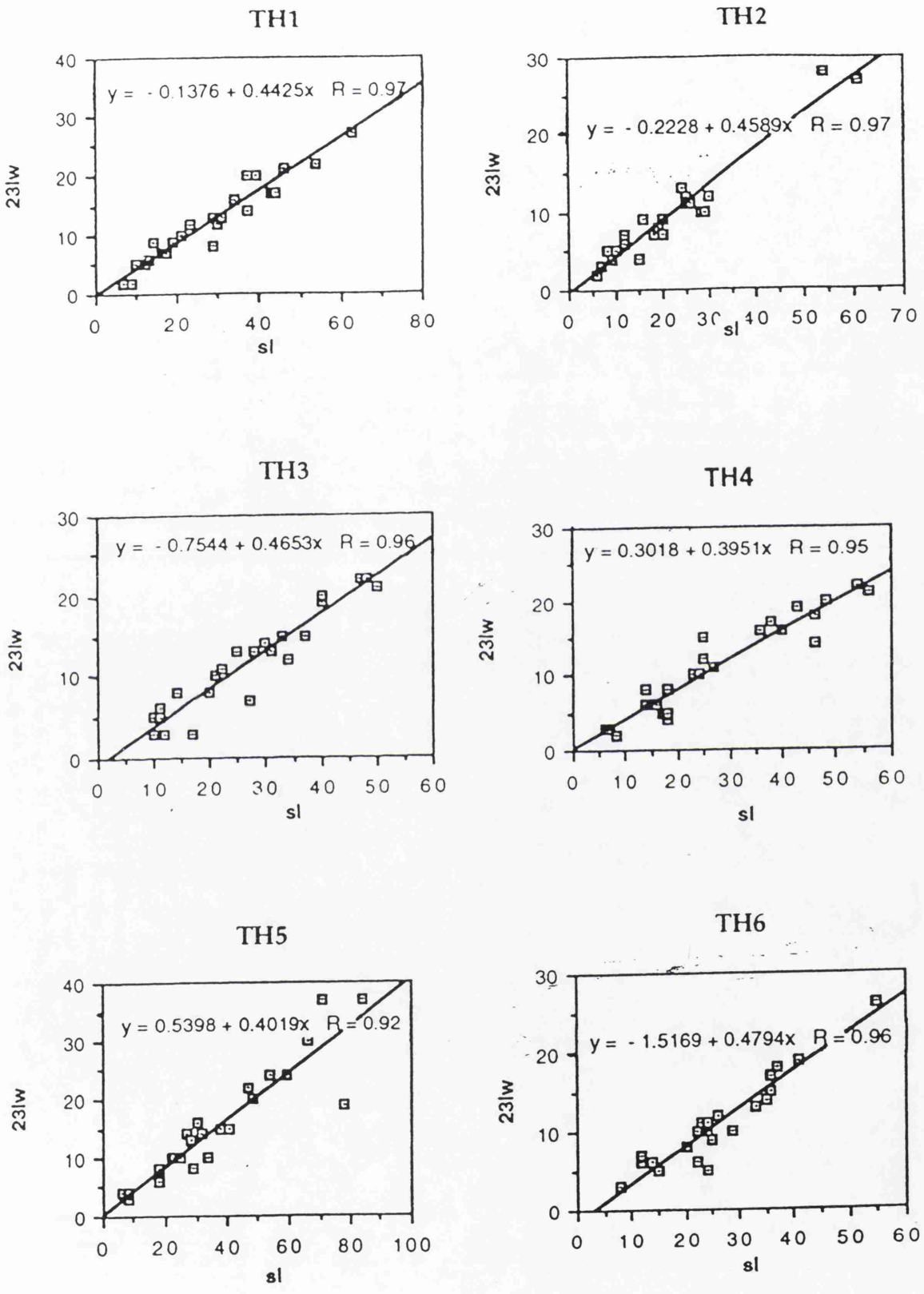


Figure 2.8(c) : Plots of linear relationship between sl and habit (with intercept) for data from real text (author 3)

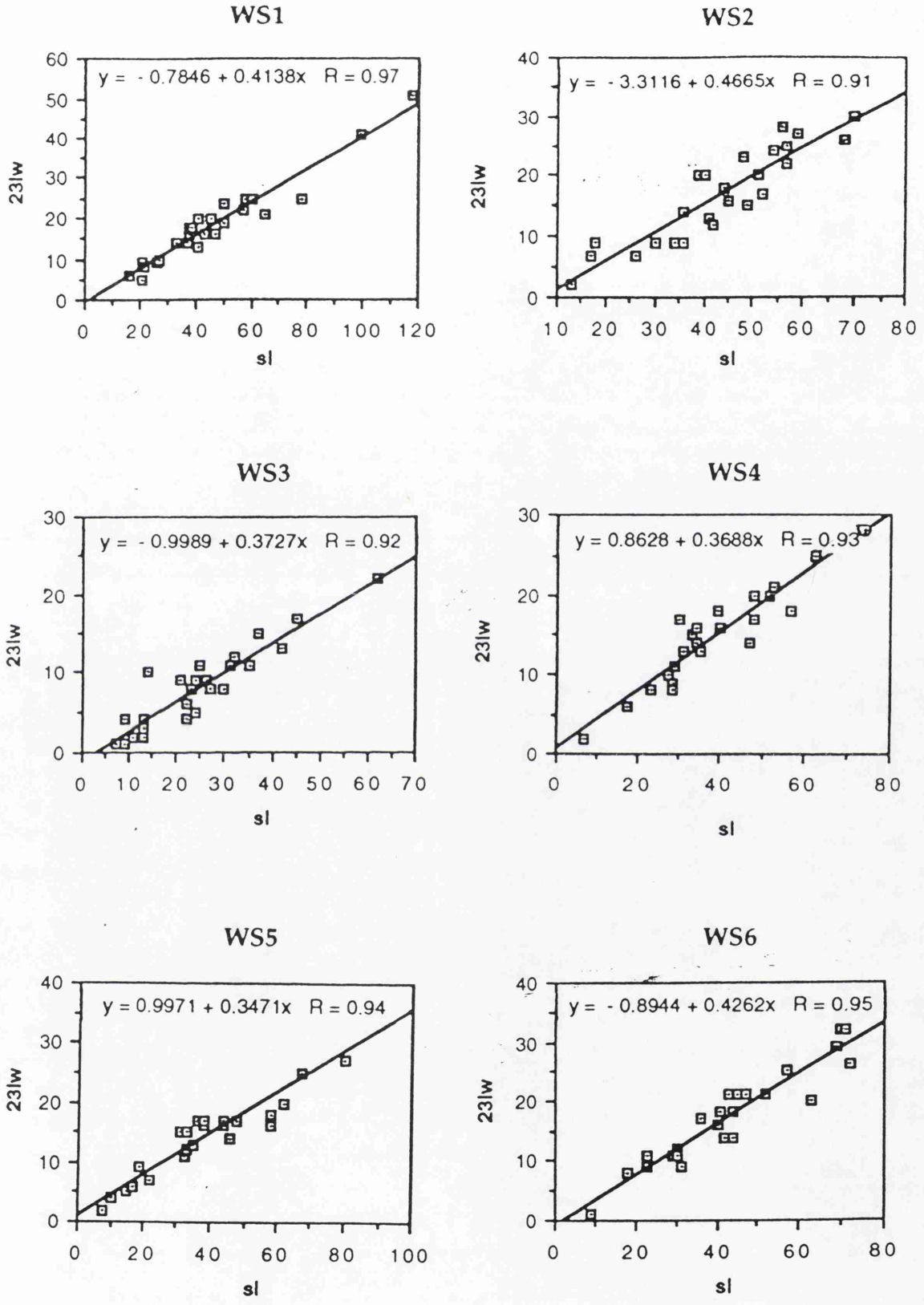


Figure 2.8(d) : Plots of linear relationship between sl and habit (with intercept) for data from real text (author 4)

It is obvious that there are very high correlations in real data between sentence length and the chosen habit. The proportional linear model y = mx was also fitted to this data as well as polynomial fit. A table of linear versus polynomial equations and fits is shown for the 5 different authors in Table 2.3.

Author	Fit of y = mx	Fit of y = mx + c	polynomial fit[3]
TH1	0.99	0.97	0.97
TH2	0.99	0.97	0.97
TH3	0.99	0.96	0.97
TH4	0.99	0.95	0.96
TH5	0.99	0.92	0.93
TH6	0.98	0.96	0.96
WS1	0.98	0.97	0.97
WS2	0.99	0.91	0.92
WS3	0.98	0.92	0.92
WS4	0.99	0.93	0.93
WS5	0.99	0.94	0.96
WS6	0.99	0.95	0.95
JA1	0.99	0.96	0.96
JA2	0.98	0.97	0.97
JA3	0.99	0.98	0.98
JA4	0.99	0.98	0.98
JA5	0.98	0.90	0.92
JA6	0.99	0.99	0.99
CB1	0.98	0.96	0.96
CB2	0.99	0.97	0.97
CB3	0.98	0.92	0.92
CB4	0.99	0.98	0.97
CB5	0.99	0.94	0.99
CB6	0.99	0.95	0.96

Table 2.3: Linear and Polynomial Fits for 4 authors and 24 texts

On the basis of this sample it can be seen that the best fit is given by the simple, proportional linear equation. There is no advantage in using a polynomial equation or a linear equation which allows a non-zero intercept on the y axis.

[3] polynomial fit based on 3rd order polynomials: $y = ax + bx^2 + cx^3 + \text{constant}$

6. Variation in "Habit" within and between author

Morton's central claim (Morton, 1991; Morton and Michaelson, 1990) is that very simple habits which emerge in discourse production can be used to differentiate individuals in order to make claims about authorship. In particular, he deals with:

(a) the ratio of 2 and 3 letter words to the total number of words in successive sentences of a piece of discourse.

(b) the ratio of nouns to the total number of words in successive sentences and

(c) the ratio of initial-vowel words to the total number of words in successive sentences.

A major part of Morton's claim is that these ratios will be more or less constant within an individual, regardless of whether the material is written or spoken, regardless of whether it was produced at the same or different times in the person's life, regardless of genre, and regardless of whether or not the individual concerned was trying to write in the style of someone else. His evidence is however entirely anecdotal and he gives examples mainly from his own writing and from that of Walter Scott. The key issue addressed below is whether or not the claim that the ratio of 5 of the habits used by Morton, (2 and 3 letter words (23lw), 3 and 4 letter words (34lw), initial vowel words (ivw), initial vowel words which are not 23lw, (ivwx) and nouns) to number of words in a sentence (sl) could be called constant within an individual (relative to between individuals).

B. Study 1: Variation within Individuals for 5 habit ratios

1. Subjects

There were 20 subjects in the experiment. 10 of these were staff at Glasgow University and were therefore termed the **staff** group, and 10 were undergraduates and postgraduates at the university and were therefore termed the **student** group.

2. Design

The subjects were required to watch a video which showed a number of episodes which revolved round a series of people carrying out various activities and hobbies. There was no dialogue in the video and it was set to music. It lasted about 7 minutes from start to finish. After the video, the subjects were required to either write in as much detail as possible what had occurred in the video, or to talk about it, in as much detail as possible. The next day, they were required to do the opposite of what they had done immediately after the video, that is, if they had spoken first, they would write about the video, and if they had written first, they would talk about the video. To avoid any order or memory effects, half the subjects were in the speak first condition and half were in the write first condition. All the spoken monologues were recorded and transcribed. Each subject therefore had a comparable written and spoken account of the video they had seen.

3. Analysis

Twenty written texts and 20 spoken transcriptions resulted from the experiment.

All the transcripts were put through a purpose written program (COUNT). This program counted the number of each length of word per sentence, starting at 1 letter words and going up to 10 letter words. A distinction was also made between words which started with a consonant and ones which started with a vowel. A sample output from COUNT can be seen in Appendix 2.
For each of the 20 subjects 15 sentences were chosen. Counts were made of:
(a) sentence length
(b) the number of 23lw per sentence
(c) the number of 34lw per sentence
(d) number of ivw per sentence and
(e) number of ivw which were not 23lw (ivwx) per sentence.

Number of nouns per sentence was also taken though this had to be counted by hand.

These measures were also counted for blocks of 15 words rather than for sentences, in order to get around the problem of punctuation which arises during transcription.

The M value for the equation (a fixed intercept through the origin) was then calculated where habit rate = M * sentence length. A measure of the Rho value (R) which measures the goodness of fit of the correlation was

also taken. For the block data, instead of a ratio, the mean number of habit items over the fifteen, 15 word blocks was calculated.

4. Additional Writing Data

Following these analyses on the spoken and written texts, the subjects were asked to provide a further 3 examples of their own writing. Two of the students subjects dropped out at this point and one of the staff left the university and was therefore uncontactable. This left a total of 17 subjects with four texts for analysis. The same measures were taken as before for the same habits on these additional texts. Comparisons were made of writing between and within subject primarily for the ratio of sl to habit (**M values**) but also for the two additional measures; Rho value and Total Deviation. The comparison of spoken and written data was made separately using the same measures and will be discussed in a different chapter. Figure 2.9 shows the M values within subject columns for 23lw in both sentence and block conditions.

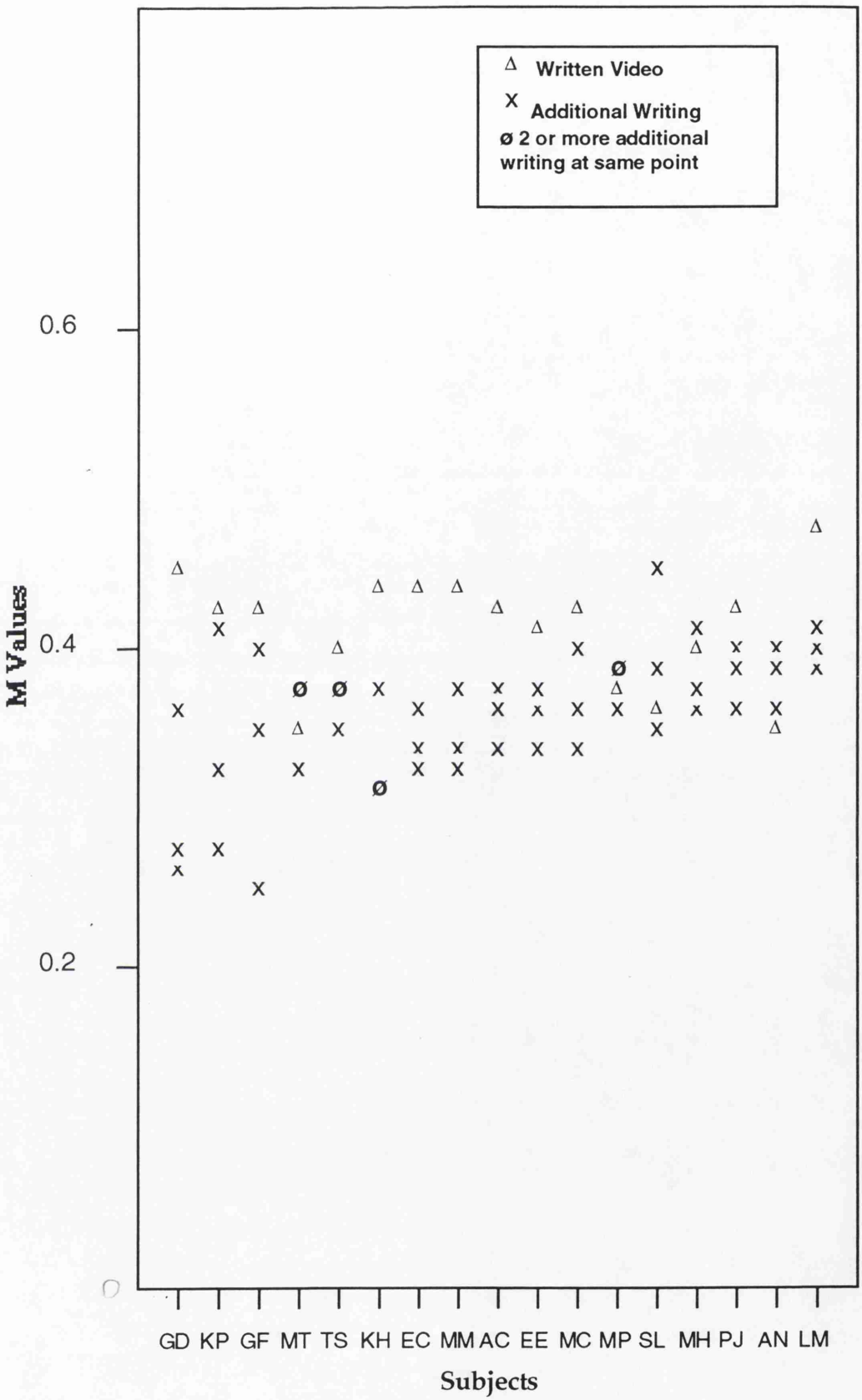


Figure 2.9(a) -23lw ratios arranged in ascending order by subject for sentence condition

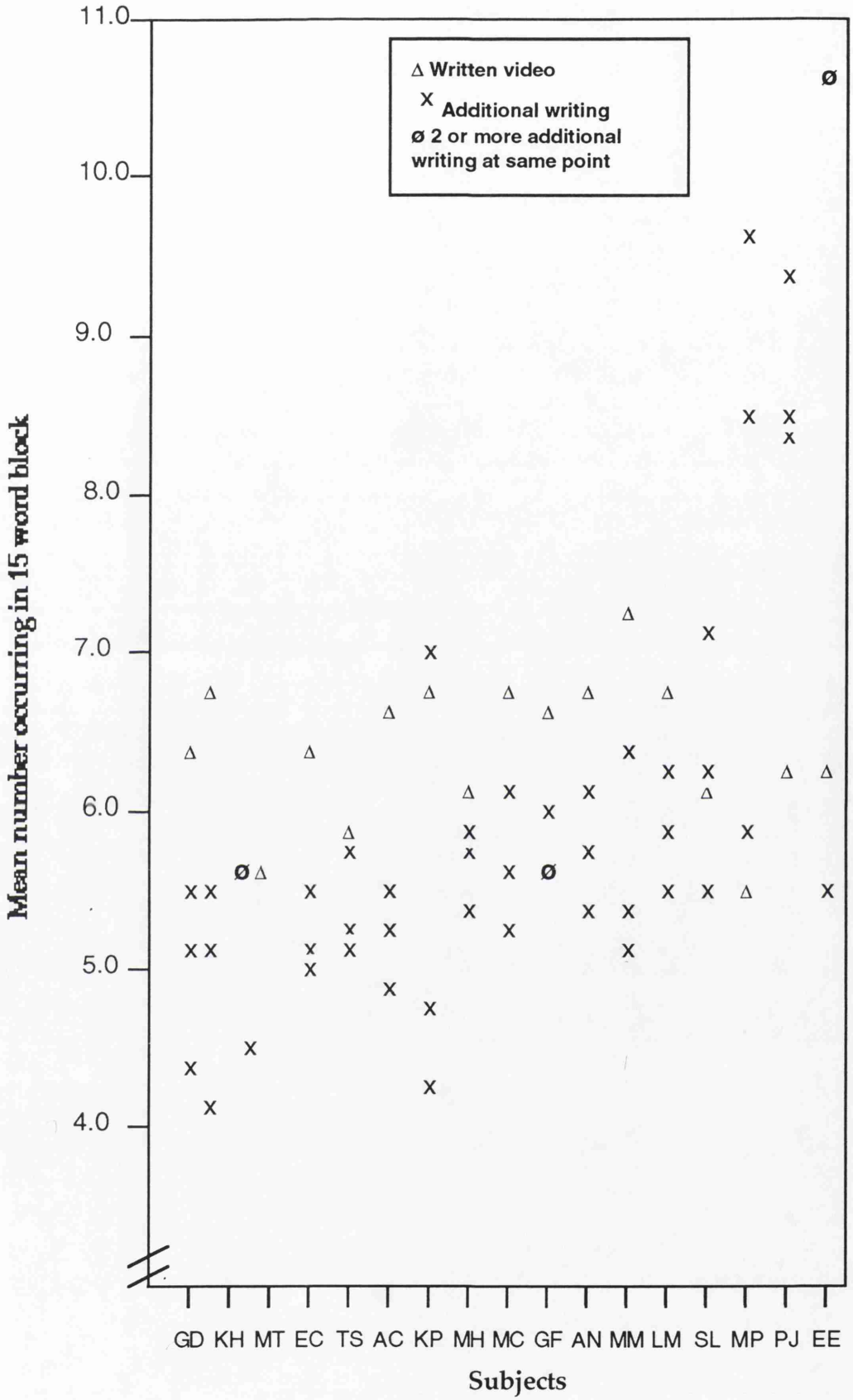
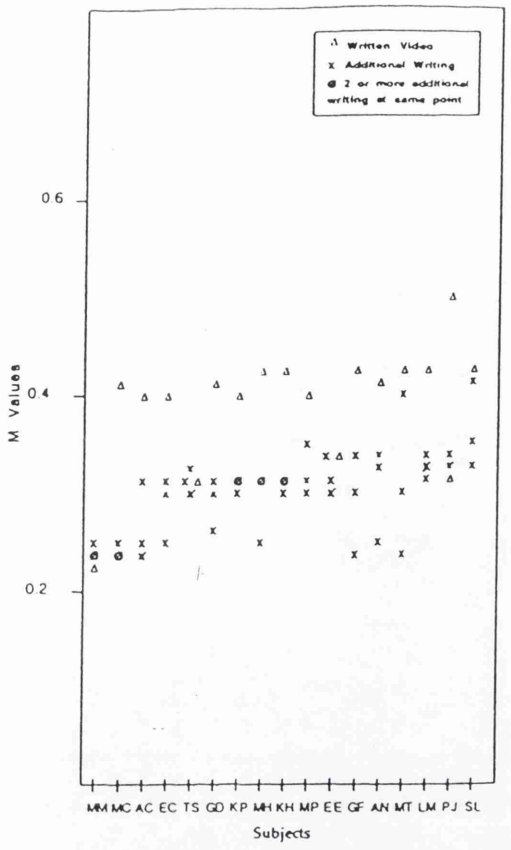
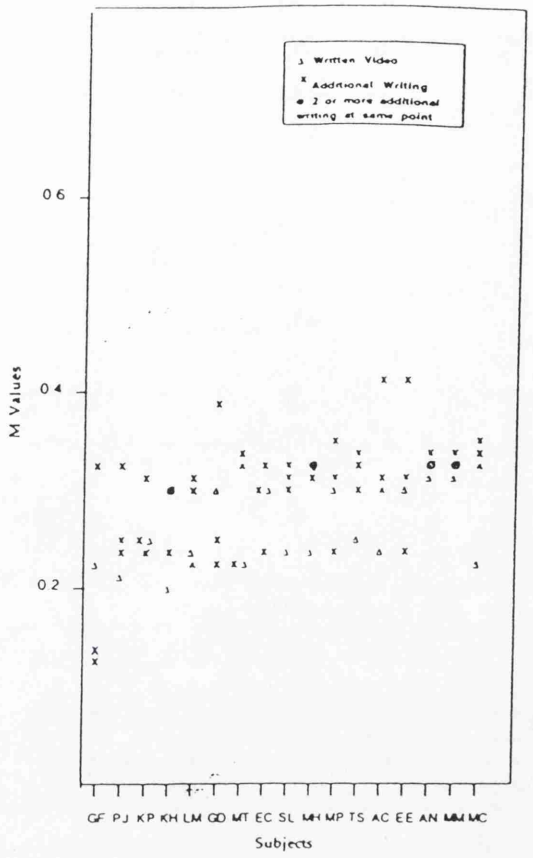
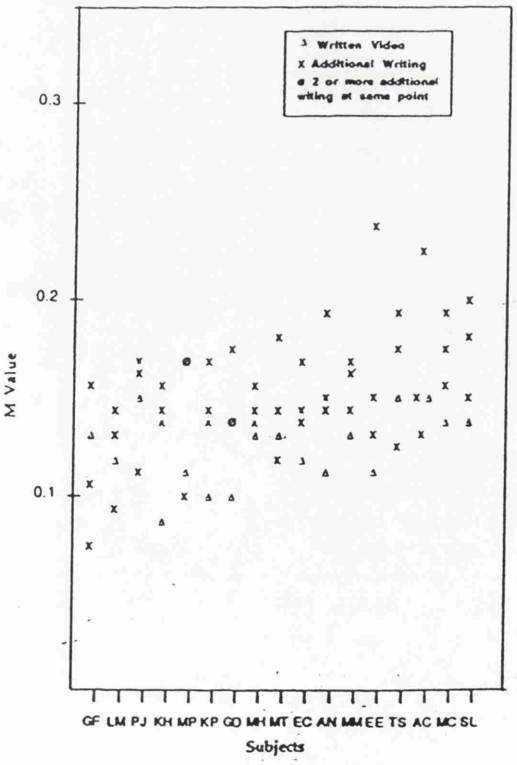


Figure 2.9(b): 23lw ratios arranged in ascending order by subject for block condition





Comparison of M Values (by Sentence) of Spoken and Written Conditions for lvw but not 23lw



Comparison of M Values (by Sentence) in Spoken and Written Conditions for nouns

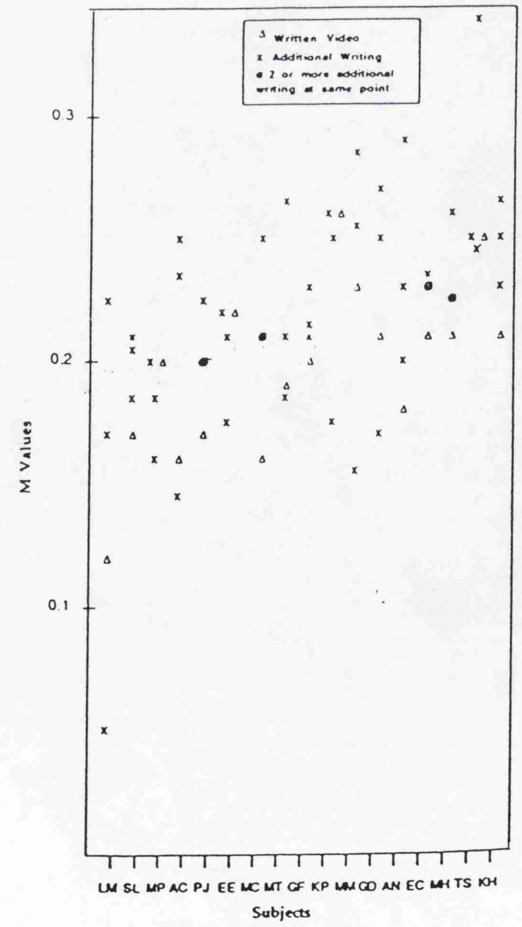
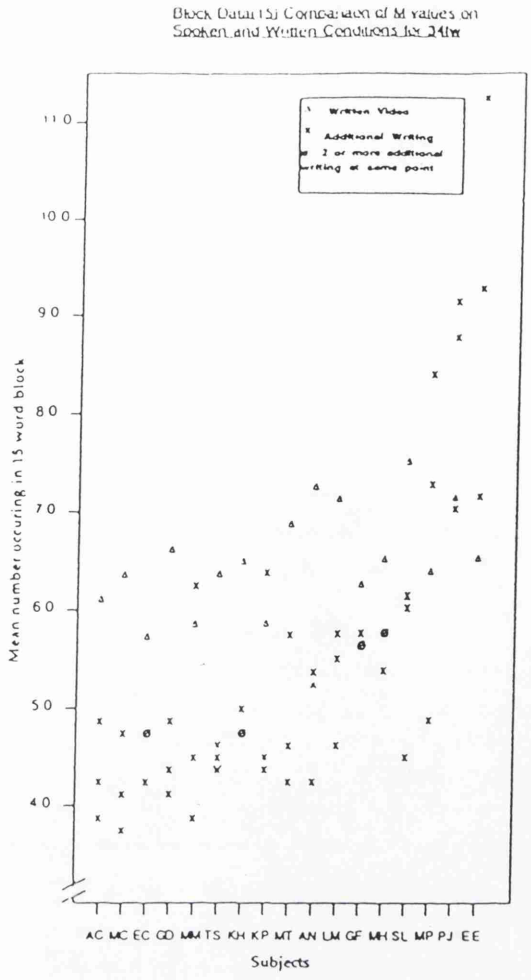
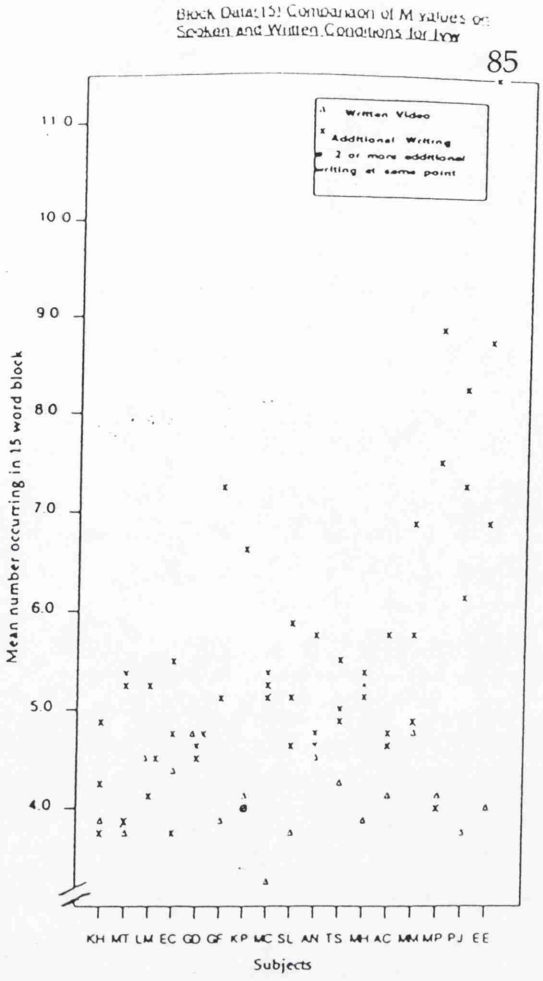
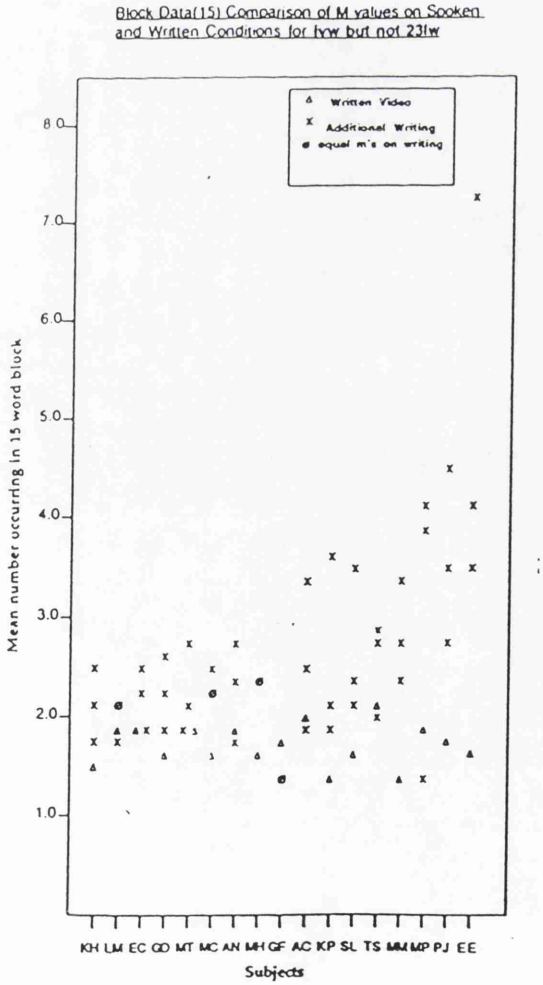


Figure 2.10(a) :Ratios arranged in ascending order by subject column for sentence data







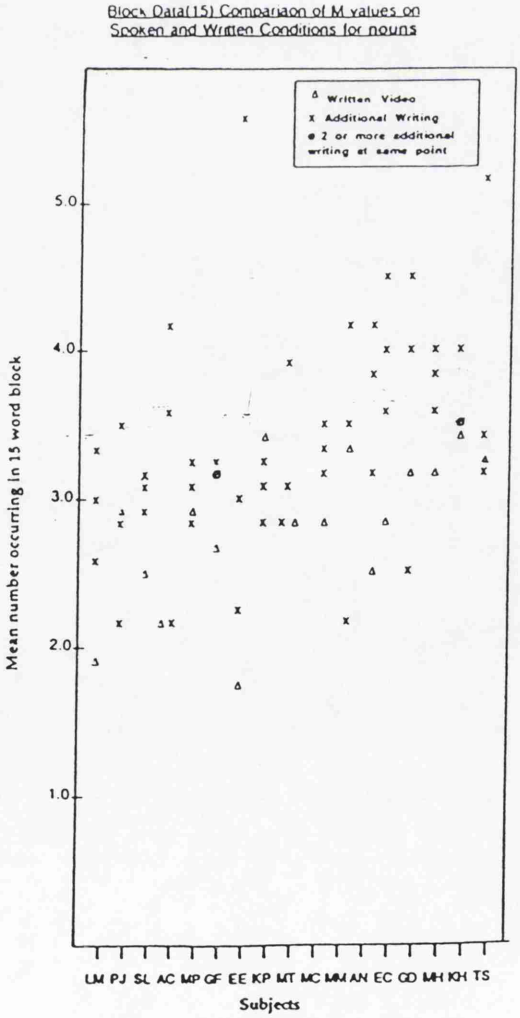


Figure 2.10(b) :Ratios arranged in ascending order by subject column for block data

The remainder of the graphs for the sentence condition are shown in Figure 2.10(a) and in Figure 2.10(b) for the block condition.

5. Results

The table below shows the results of the one way between subjects ANOVA on each of the habits.

	By sentence	By block
23lw	$F(3,48) = 9.389, p < 0.01$	$F(3,48) = 0.454, p > 0.5$
34lw	$F(3,48) = 33.951\ p < 0.01$	$F(3,48) = 4.058,\ p < 0.05$
ivw	$F(3,48) = 4.801, p < 0.01$	$F(3,48) = 3.932, p < 0.05$
ivwx	$F(3,48) = 6.968, p < 0.01$	$F(3,48) = 5.818, p < 0.01$
nouns	$F(3,48) = 3.302, p < 0.05$	$F(3,48) = 1.001, p > 0.1$

Table 2.4: Results of one-way within subjects ANOVAS on 5 habits

Significant differences were found within subjects on the ratio measure for all habits apart from 23lw and nouns in the block condition. Therefore, the ratio cannot be said to be constant within an individual.

C. Chapter Conclusion

Morton's claims are based on the idea that there are habits which may be more or less constant in individuals over a large extent of their lives. These habits, he claims, are revealed by the using the cusum technique. It is possible to test the idea of constant habit by the examination of simple ratios. This was done in the present study: the results show that variability within individuals over samples of writing is as great as is variability between individuals. Thus one of Morton's basic ideas is simply incorrect. One interpretation of Morton's claims is that the habit which is constant for a given person may be one of many, and that this

can only be known by carrying out a detailed examination of the writing of that individual. Such an argument still does not save the position of cusum as a tool for forensic linguistics. Constancy within an individual only makes sense if it goes along with constancies in another set of individuals, and if such a constancy is to have any discriminatory capacity, the values for individuals in the other sets must be different from the value for the individual in question. To make a claim about the constancy of a given habit for a given individual, several samples are needed for the individual, and several samples are also necessary for the other set. There is absolutely no point in basing claims about which habits are most characteristically constant for an individual on a single cusum. This study found significant differences in the ratios of habit to sentence length, for the four texts examined for each individual.

In conclusion, it would seem that the cusum technique is based on assumptions which are at best of limited reliability. Unless the constancy of a habit within an individual is established over a number of samples, then the crux of the whole method is destroyed.

Chapter 3: Cusum Measures with Single and Multiple Authorship

A. Introduction

Having tested the basic assumption behind the cusum, we now examine the effect of more than one author on the cusum curve. Figure 3.1 shows a cusum plot which extracts from the writing of two individuals are placed in succession, and the cusum of the whole is then calculated. This is an example that Morton gives (Morton & Michaelson, 1990) to illustrate the discrepancy which occurs in the cusum plot when writing of one individual is inserted into writing of another.

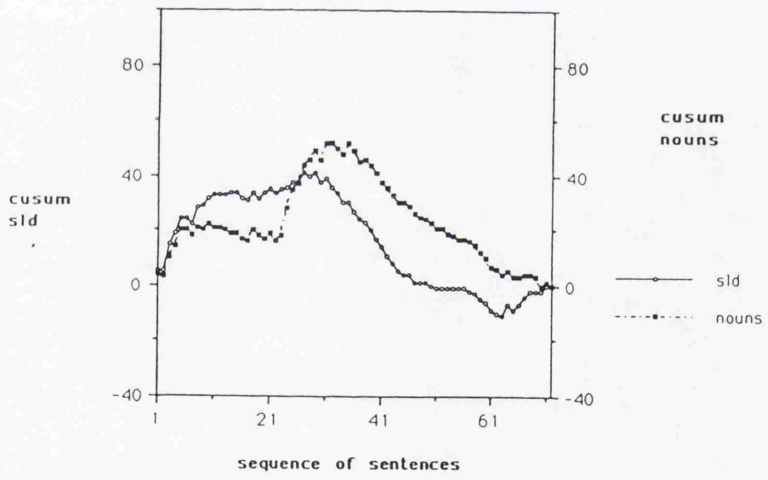


Figure 3.1: Cusum of sentence length and nouns for 22 sentences of Thomas Hardy in 50 of Walter Scott. From Morton & Michaelson, 1990)

Note the considerable discrepancy between the two curves. It is precisely this type of discrepancy which is at the heart of Morton's claims about the utility of the method in the forensic context, and his

papers contain many illustrations, asserting that the point at which the discrepancy emerges can be used to estimate where a foreign insert has been made. Figure 3.2 shows another example from Morton and Michaelson (1990) with an apparently large discrepancy at the mid point where the insertion has been made.

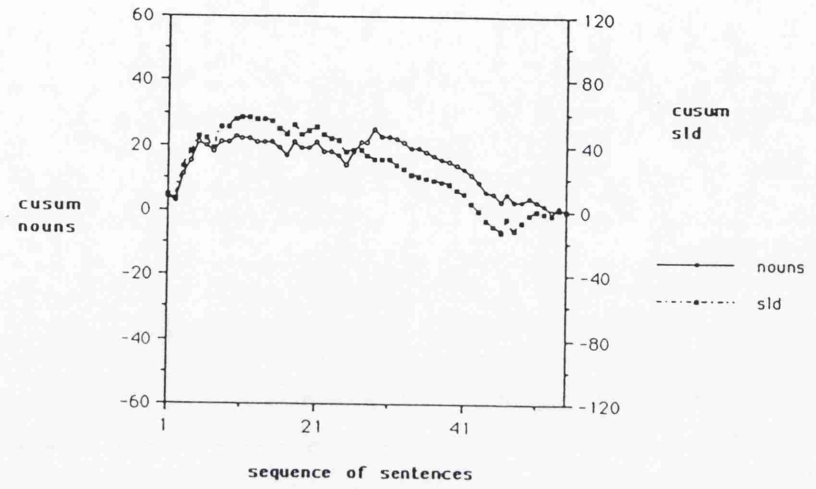


Figure 3.2: Cusum of sentence length and nouns for 5 sentences of Thomas Hardy into sample from Walter Scott at the mid-point

Morton and Michaelson (1991) argue that the good fit to the cusum is achieved by a constant ratio, in other words, the neat fit of the two cusums can only be expected in the presence of a fixed habit (ratio). So, if a piece of writing is tampered with by someone who has a different habit, this will distort the close relationship between the two curves. This alone must contribute to the potential appeal of the curves to a jury. For the scientific investigator, such complex curves can be a source of difficulty, however, particularly when an attempt is made to produce a mathematical index of the goodness of fit of the two curves.

Do cusum plots combining different discourses from the same individual show the good fit which is obtained from a single sample? Quite clearly, if the only mechanism underlying the agreement of cusum plots for habit and sentence length is the ratio (i.e. a linear relation with intercept zero) then this is equivalent to asking whether the ratio remains constant within an individual. If a piece of text with ratio M is inserted into one with ratio N, this should produce a disturbance in the cusum fit.

However, cusums generally preserve sentence order. This would be useful for the identification of where intrusions lie, but it might also be the case that there is additional information contained in a cusum which is supplementary to the global ratio. Therefore, an objective index of the fit of 2 cusum functions is required to determine the relationship between ratio differences and disturbance in the cusum fit.

1. Problems with Interpretation of Separation of two curves

The question being asked when interpreting the difference between the two components of the cusum graph is really "are these two lines the same or are they different?". Exactly what constitutes different is totally subjective, be it single or multiple author cusums (since we are initially unaware whether they are single or multiple author cusums) and the eye can be easily drawn to the separation in the wrong direction. Hardcastle (1993) points out that the different scales used on the y axis by Morton severely hampers this interpretation producing a uniformity between lines when the scale is condensed. If the scales are expanded, however, dissimilarity of lines can be reported when they would not have, had the scales been condensed.

Hardcastle also points out that a cluster of sentences which is shorter than average followed by one which is longer than average will produce a divergence which could be erroneously construed as a section of writing which has been tampered with. Similarly, because different texts from the same individual have differing ratios of sentence length to habit, joining two pieces of text together from the same person can give rise to a discrepancy which would imply multiple authorship.

The change in sentence length will also determine to a great extent the separation of the graph due to the extent to which an increase or a decrease in sentence length is sustained over a number of sentences. If there is a long run of increasing sentence-lengths (or of decreasing sentence lengths) then there will be a sustained opportunity for any divergence between two cusums to become visible, and to produce a steady increase in an area deviation measure.

A systematic investigation of how the change in sentence length affects the cusum chart and the insertion procedure was carried out. This demonstration is described next.

2. Demonstration 1

The objective here is to show how the divergence of sentence length and habit cusum curves is a function of sentence length. Such a demonstration would mean that the deviation measures are a function of more than just differences in habit in the sense of ratios.

Four "hypothetical texts" were invented representing four different authors, in a similar way to the linear data in chapter 2. These hypothetical texts were given four different sentence length to habit ratios which were 0.4, 0.7, 0.8, and 0.9. (For example, for the text with a ratio of 0.4, a sentence length would be 5 and the corresponding habit would be 2, since 2 is 4/10ths of 5.)

Initially the four ratio "texts" were grouped into two conditions, the CLOSE ratio condition (0.7+0.8), where the combined ratios were close together, and the FAR condition (0.4 +0.9), where the combined ratios were far apart. These represented two insertion conditions for the texts. The assumption was, that it should be possible to arrange the sentence lengths from each of these conditions, so that the resulting cusum curves could have a small separation or a large separation. The size of the separation would depend only on whether there was a large or a small change in successive sentence lengths. Sentence lengths in these conditions were thus arranged firstly so that there was a relatively small change in sentence length in successive sentences over the whole text, and secondly so that successive sentence had a large change in sentence length. Subsequently, there were 4 conditions:

1. Close condition, small change in sentence length
2. Close condition, large change in sentence length
3. Far condition, small change in sentence length
4. Far condition, large change in sentence length.

3. Results of the simulation

It was found as expected that using the same data (e.g. a combination of ratios 0.7 and 0.8 - close condition) it was possible to get cusum charts

where the area between the two curves were totally different, dependent only on the order in which the sentences were arranged. The two different separations for the close condition, where the ratios which were combined are similar are shown in Figures 3.3 (a) and (b) and for the far conditions, where the ratios combined are far apart, in 3.3(c) and (d).

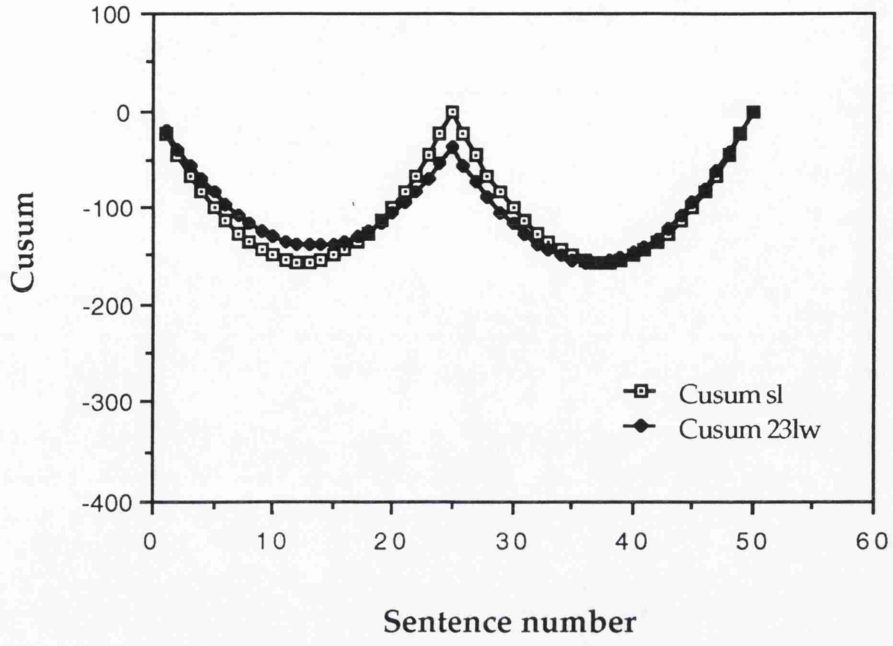


Figure 3.3 (a): Close ratio with large change in sentence length

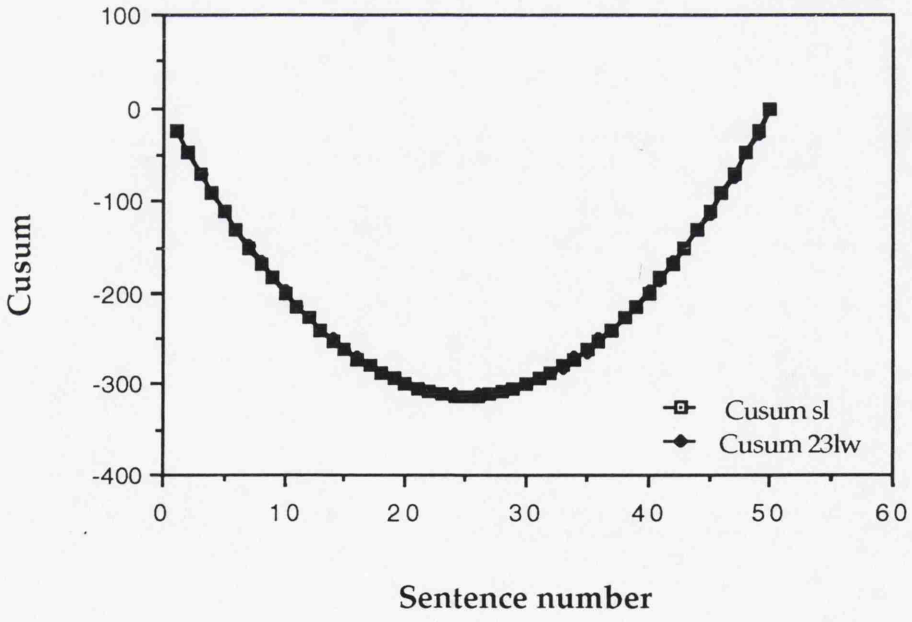


Figure 3.3(b): Close ratio with small change in sentence length

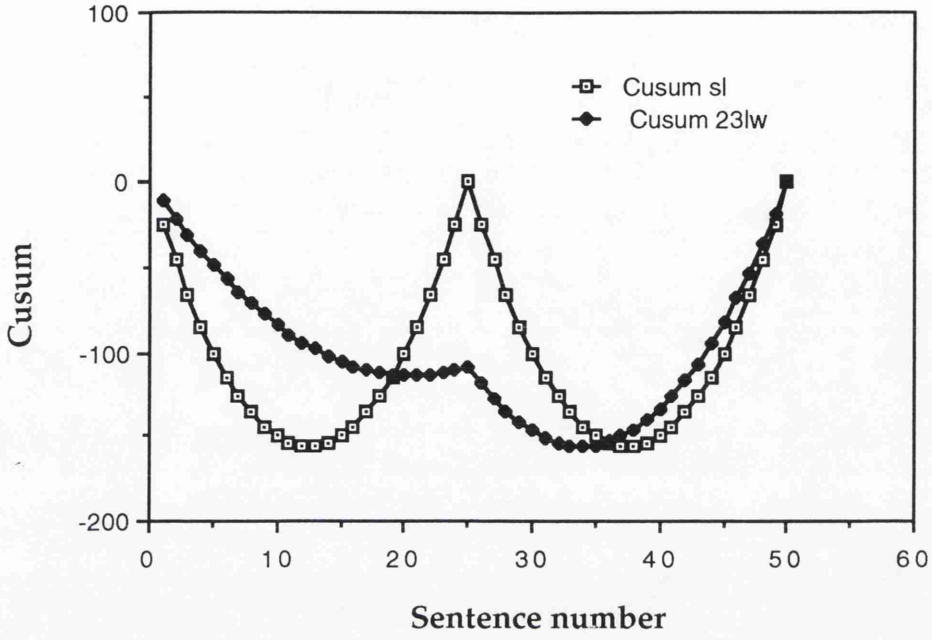


Figure 3.3(c): Far ratio with large change in sentence length

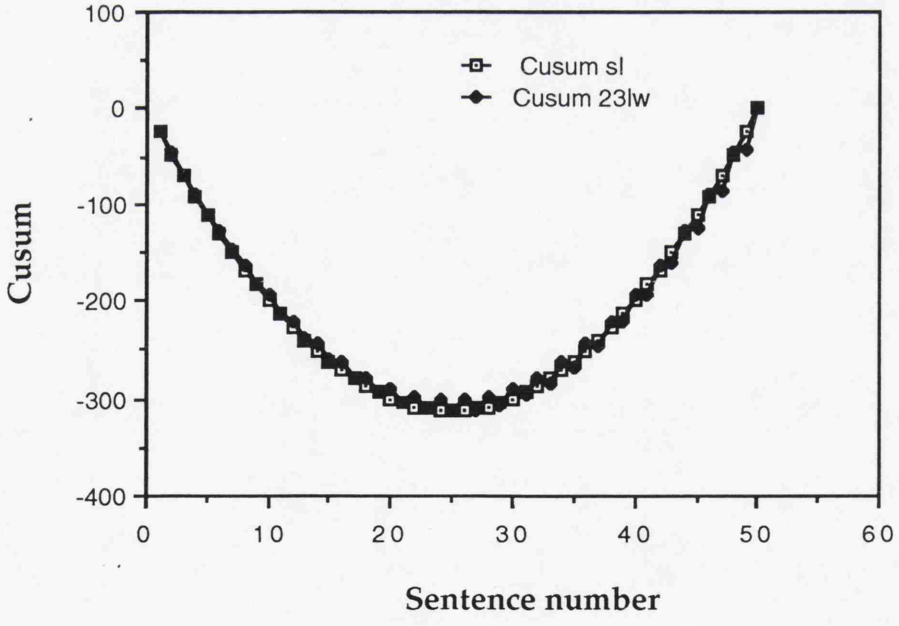


Figure 3.3 (d): Far ratio with small change in sentence length

These values in all 4 conditions are shown in Table 3.1 .

ratio	small change in sl	large change in sl
0.7/0.8	49.41	579.01
0.4/0.9	227.76	1833.48

Table 3.1: Total Deviation between cusum lines in 4 conditions

In both conditions, the total deviation with a large change in sentence length is greater than that with a small change in sentence length.

4. Discussion

These findings show that if there are only small changes in sentence length throughout the chosen text then there will be a close fit between the two curves, independent of whether the text has come from two sources with large or small ratio differences. Therefore, a large difference between the cusum curves is not a reliable indicator of text from a different individual, if sl to habit ratio determines the fit of the cusum.

There are other problems of a similar nature. Hardcastle (1993) points out that since the cusum chart does not represent the deviations from consistency directly, but only cumulative deviations, then the interline separations are more likely to be greatest near the middle of the chart than near to either end. Also, he suggests that the longer the text is, the greater the chance there is for the separation of the lines. In addition, Hardcastle argues that the scaling method which Morton uses is inaccurate since it is a function of the average sentence length and habit

word rate only within the portion of the text represented by the data points of the cusum chart between the highest peak and the lowest trough. If the difference between these is very small, then it may be unrepresentative of the data as a whole, and depending on whether the peak precedes the trough or vice-versa, the proportion of short and long sentences will differ. If the cusum lines diverge due to a change in ratio of the habits, the resulting separation will be carried forward to the rest of the cusum chart. Morton's solution to this is to use transparency overlay to correct this which inevitably results in many ad hoc interpretations.

Therefore, Hardcastle suggests that instead of plotting the cusum values, the values for the vertical separation between the two habits should be plotted. This would mean that if the proportion of habit words in the sample were equal those in the text as a whole, then there would be no widening or narrowing of the gap between the two curves. In effect, the slope of the sentence length to habit ratio should be plotted. Hardcastle terms the proportion of habit words in the sentence the OBSERVED habit words and the proportion of habit words in the text as a whole the EXPECTED habit words. The separation change is therefore the Observed habit words minus Expected habit words. Histograms of these values can be plotted and these show that there is considerable variation in habit rate occurrence from one sentence to the next and using this different scaling method can improve the cusum fit in some areas and make it worse in others. This method of scaling will be used in addition to Morton's method, in the next chapter. He also suggests that the problem of a deviation in one part of a text affecting subsequent parts could be averted by using a system of moving averages of the separation changes.

Calculation of Spearman's Rho between the cusum values for the two habits has also been used as a measure of cusum fit (Canter, 1992). The advantage of a correlational method such as this is that no scaling factor is required, and a direct test of whether the correlation is lower for the cusum values from two individuals rather than from one may be carried out. Canter uses a cut-off level of a correlation of 0.9 above which texts are assumed to be from one individual and below which, from two. On a comparison of single and multiple authored texts, for which the 23lw cusums had been calculated, Canter found that 48% of texts from a single author were inconsistent with the hypothesis of single authorship at a cut off point of a rho value of 0.9. Sixty five percent of texts derived from 2 authors were above the 0.9 criterion implying single authorship.

5. Investigation into the Insertion Procedure

5.1 Rationale

If, as Morton and Michaelson (1990) suggest, the basis of the cusum is ratio, and if the fitting procedures are good enough to capture the ratio, then as stated previously, it should follow that if a sample of writing based on slope M has added to it a sample based on slope N, the disruption to a common cusum graph should depend upon the difference between M and N. The larger the difference, the bigger the disruption should be. Furthermore, it should not matter whether M and N come from the same or from different authors.

So, we would expect the following:

1. **Same author, similar slope**: combined cusum shows good fit
2. **Different authors, similar slope**: combined cusum shows good fit

3. **Same author, different slopes**: combined cusum shows disruption
4. **Different authors, different slopes**: combined cusum shows disruption

Therefore disruption should only occur in conditions 3 and 4 where there are **different** slopes. Demonstrating this depends upon being able to put a number to the disruption (or goodness of fit) of the sentence-length and habit cusums. This experiment tests this idea with the data just described. Three conditions were used for the cusum calculation:

a) sentence length arranged in natural-order for the texts and the cusum calculated for the whole

b) sentence lengths arranged in random-order for the texts and the cusum calculated for the whole

c) sentence length arranged in ascending-order for the texts and cusum calculated for the whole

The order was arranged for the first text and the second separately, then put one after the other and the cusum of the whole calculated.

The reason for these conditions was as follows. Initially one might suppose a monotonic relation to hold between some simple measure of goodness of fit (like the area between two cusum curves) and the difference between M and N. However, this is not necessarily the case. Getting a divergence between two cusum curves has been shown to be dependent on the differences between successive sentence lengths. If however, ratio constancy is assumed to underlie the good fit of two cusum curves (Morton and Michaelson 1991) then the actual order of sentences used to produce a cusum should not have any material effect on goodness of fit. It should not matter whether the sentences are

arranged in rank order of length, in random order, or in their naturally occurring order, except for the caveat on shape discussed above. In fact, by arranging the sentences of a single discourse in ascending order of sentence length, a much smoother cusum can be obtained which contains a maximum change in shape. Figure 3.4 (a) shows this using artificial data having a single ratio arranged in ascending order of sentence length.

If two discourse samples are first arranged in rank order of sentence length separately , then put one after the other, and then the cusum of the whole calculated, a shape with a steady, large change can be obtained. Figure 3.4 (b) shows this with artificial data of two similar ratio values and 3.4 (c) with different ratio values. The data used in this simulation can be seen in Appendix 3. It can be shown that the area measure of deviation from this sort of plot *is* a monotonic function of the differences in the habit to sentence length ratios of the constituent discourse samples. Such ascending-ordered cusums are used in the investigation to be reported.

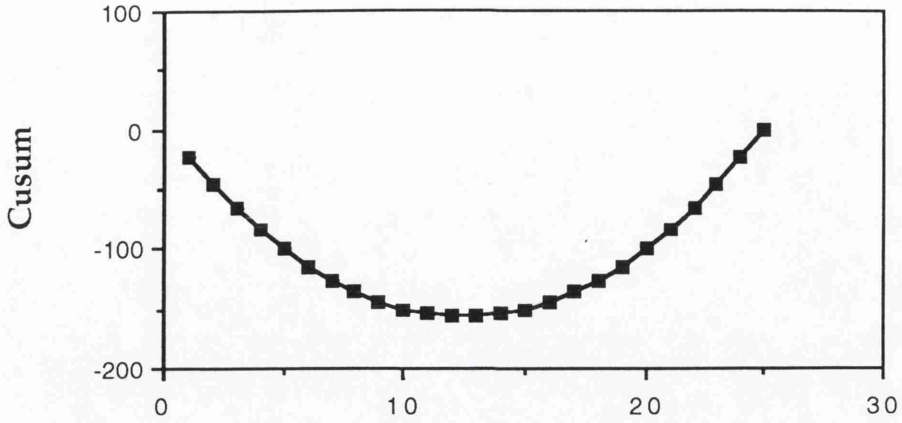


Figure 3.4(a): Artificial data for single ratio

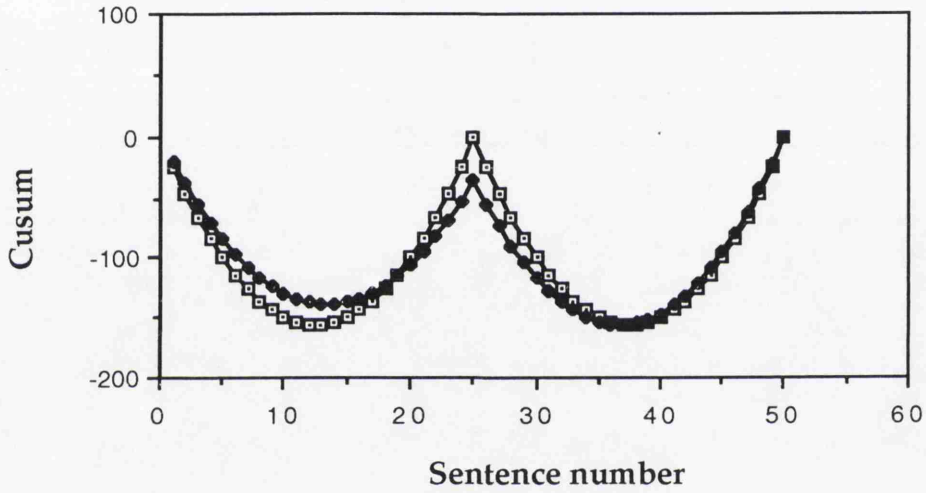


Figure 3.4(b): Artificial data with two similar ratios

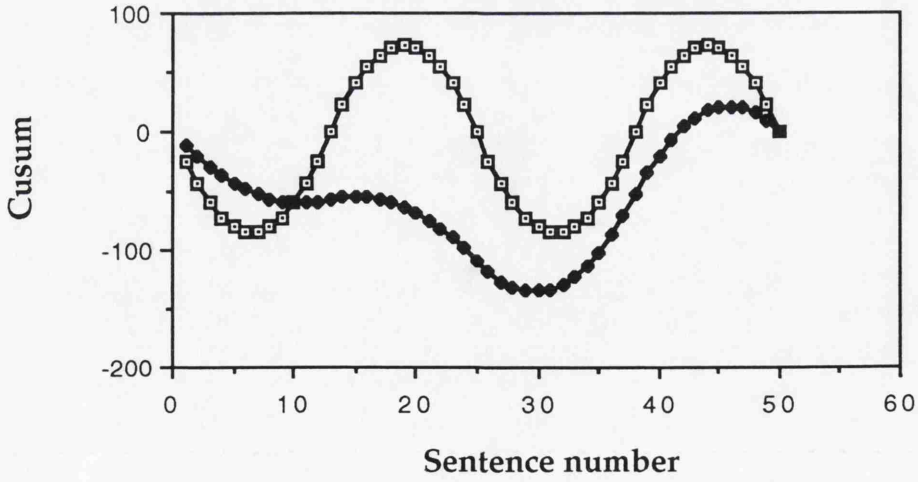


Figure 3.4(c): Artificial data with two dissimilar ratios

It could however be argued that some natural order information is of importance in the cusum. For instance, with 23lw and sentence length there might be slight but systematic deviations from linearity for given individuals which have only a marginal effect on a straight line fit, yet are somehow detectable in a cusum plot. This general possibility of real sequential effects having any influence is investigated by comparing the results of ascending order cusums, with those for natural-order and random-order. The random-order manipulation clearly eliminates any possibility of natural order effects.

B. Study 1: Test of Cusum fit in relation to ratio differences

1. Materials

The materials were 40 essays written under exam conditions by Glasgow University undergraduates. There were 5 essays from each of 8 subjects. All calculations for this study were using sentence length and the habit 23lw.

2. Design

Using the 40 essays, cusum plots were produced for the four combinations described previously , that is: two were produced by taking the 8 instances of most similar slopes for the **same** subjects, and 8 most similar for the **different** subjects (the close manipulation). The maximum slope difference was 0.014 (mean diff. = 0.007) for the **different**, and 0.001 (mean diff. = 0.0005) for the **same.** The other two were produced by taking the 8 most different slopes for **same** (mean diff. = 0.083) and the 8 most different slopes for **different** (mean

difference = 0.098) - the different manipulation. The result is thus 32 pairs of texts, 8 in each cell of the design. Three different cusum plots were produced for all 32 cases by making one plot of each pair run together in succession. The first plots used the natural order of the sentences. The second plots used a randomisation of sentence length in each pair, so as to destroy any order information. The final plots used ascending-ordered sentences in terms of ascending length (for the first pair and the second pair independently, as discussed).

3. Results

The separation of the two curves in each of the 40 (texts) x 3 (orders) was calculated as before from the absolute sum of the difference in value at each point regardless of sign, for each cusum plot. The mean values in each condition are shown in Table 3.2 and plotted in Figure 3.5.

Order	1Author/Close	1 Author/Far	2 Author/Close	2 Author/Far
Natural	461.59	691.72	537.03	723.02
Random	397.81	678.53	406.37	615.78
Ordered	439.49	1180.95	364.43	1400.36

Table 3.2 :Mean Cusum deviation in Three Order Conditions

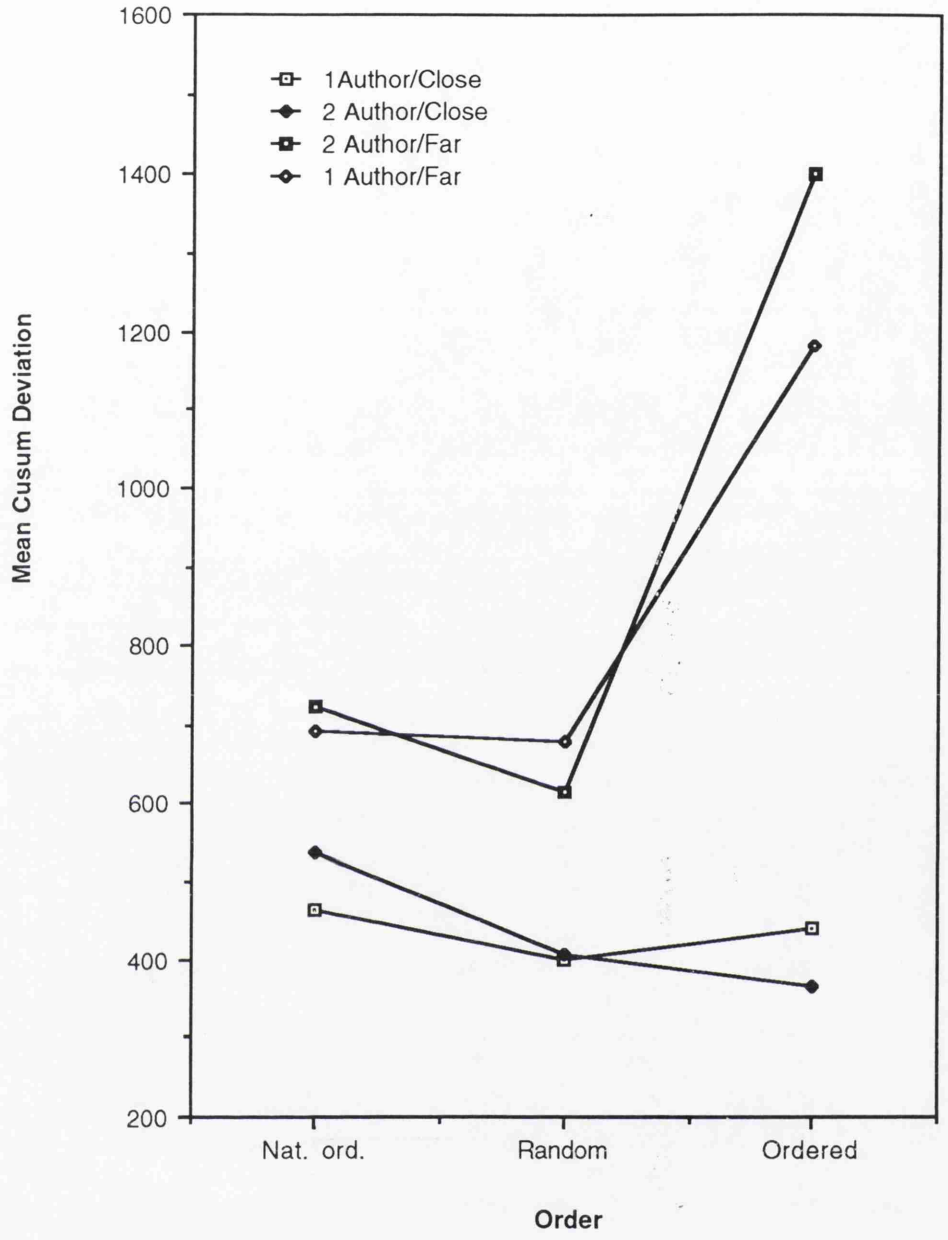


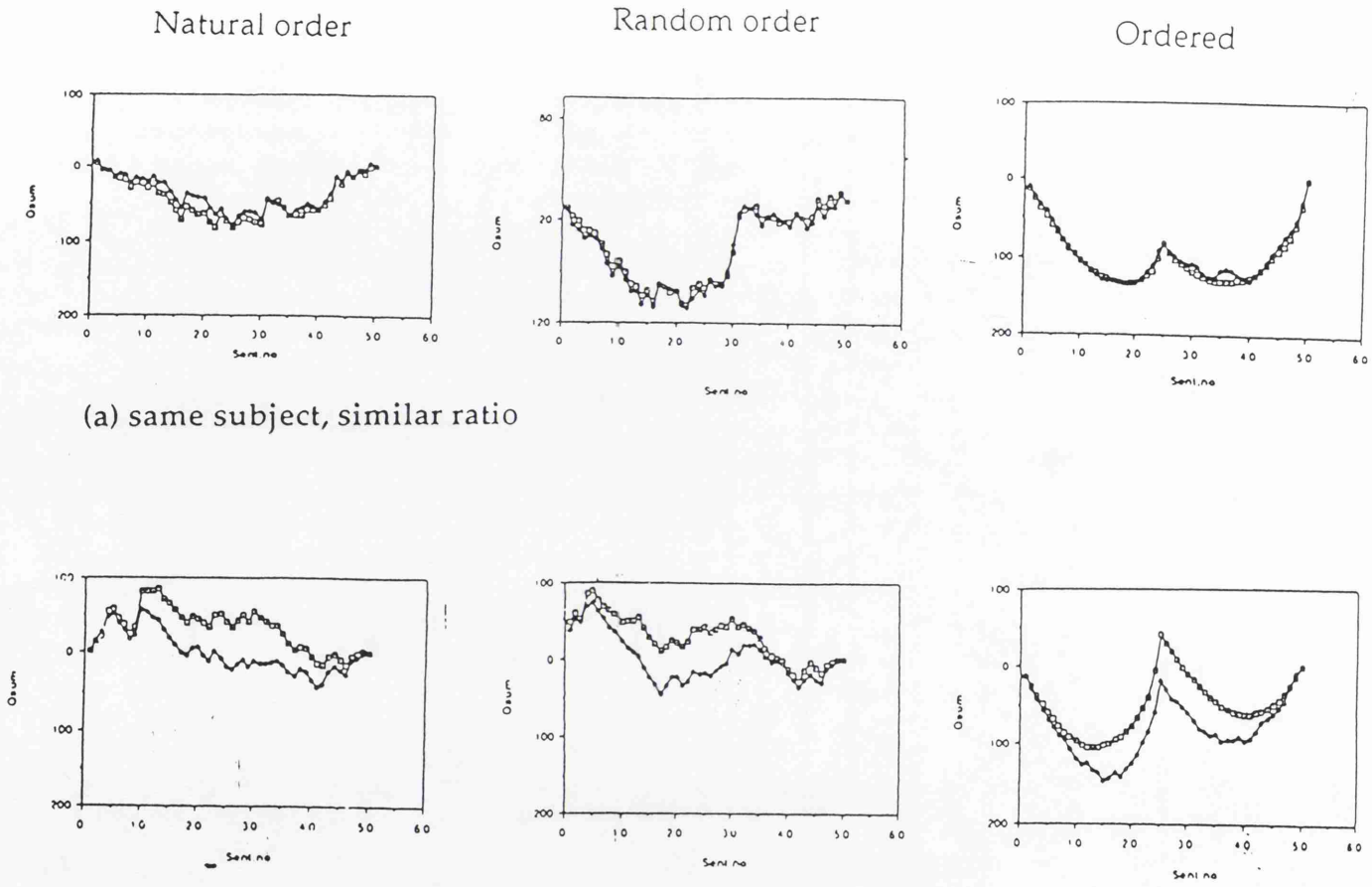
Figure 3.5: Mean Area Deviation of cusum curves in 3 order conditions

The results show that regardless of plot-type, ratio difference is the only factor which reliably influences deviation measure, apart from the order manipulation. This is particularly clear in the ordered data, for reasons discussed earlier. An analysis of variance on the deviation

measures supports this interpretation: while ratio difference (similar/ different) produces a reliable difference in deviation, with $F(1, 28) = 18.23$; $p < 0.01$, there is no reliable effect of subject (same/different), with $F(1, 28) = 0.194$; $p > 0.5$. As anticipated, the result is strongest for the ordered data, where a consistent direction of change maximises the chances of finding something on an area measure; this is shown by the interaction of measure with similar/different slope, with $F(2, 56) = 21.53$; $p < 0.01$.

Figure 3.6 (a) shows the results for the first two conditions, same subject, similar ratio and same subject different ratios while Figure 3.6(b) shows the same for the last two conditions , different subject similar ratio and different subjects, different ratios. Both are shown in all three order conditions: natural, random and ordered data. The patterns conform to the overall trends.[4]

[4] In addition to the measures made above, an attempt was made to carry out a more qualitative test by eye, by myself and a second judge, paying attention to the shapes of the deviations obtained. No obvious differences between same subject and different subject plots could be readily discerned.



(a) same subject, similar ratio

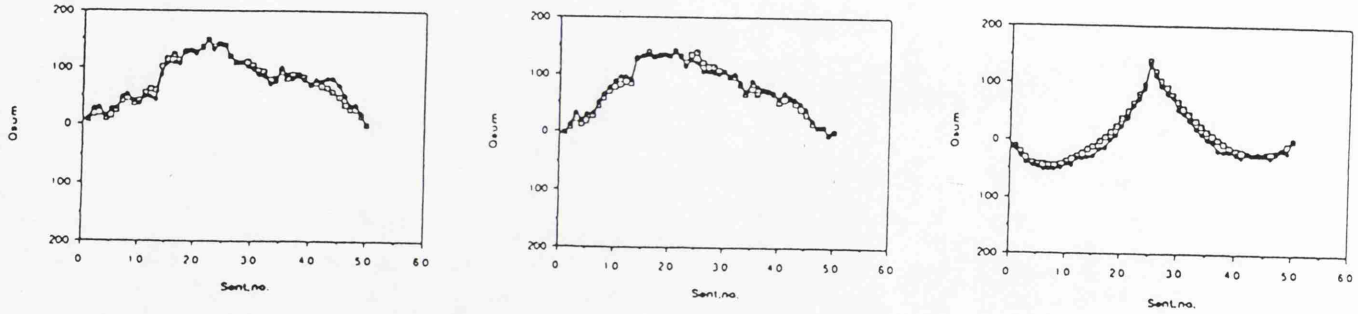
(b) same subject, different ratios

Figure 3.6(a): Examples of cusum plots in first 2 author and subject conditions and 3 order conditions

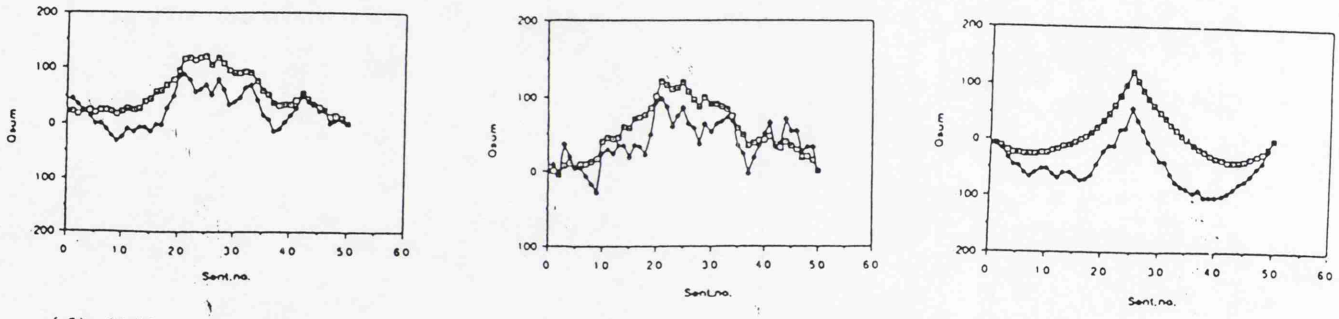
Natural order

Random order

Ordered



(c) different subjects, similar ratio



(d) different subjects, different ratios

Figure 3.6(b): Examples of cusum plots in second 2 author and subject conditions and 3 order conditions

4. Discussion

In order to test whether combinations of discourses from the same individual produce poorly fitting cusums depending on underlying ratio differences, it was necessary to devise a metric for goodness of fit. It was demonstrated that measures of deviation produce values which depend on the way in which sentence length changes as the discourse in question unfolds. Nevertheless, using three types of cusum plot, it was shown that the variation in fit was due to slope differences, and not to individuals. Taken together, the results obtained here suggest that a well-fitting cusum does not necessarily mean that the samples come from the same individual, and that a poorly fitting cusum does not necessarily mean that the samples come from different individuals.

These results hold for all three ways of producing cusum plots (natural order, random order, ascending sentence lengths) thus supporting Morton's assumption that natural order is not essential for a good fit to be obtained, and the ratio between sentence length and number of 23lw is at the root of the cusum. In the absence of evidence to the contrary, it can be concluded that preserving the natural order of sentences does not really contribute to the goodness of fit of cusums, although it is clearly necessary to preserve this ordering if there is to be any chance of detecting where an insertion has been made.

Although it is desirable to have a metric for the match of two cusum lines, these simulations illustrate how an obvious measure like area or total deviation is dependent upon the shape of the cusum itself. It is therefore very difficult to produce a simple metric.

C. Study 2: Investigation of Insertion Procedure using all 5 Habits

Having looked at the insertion procedure initially using 23lw and found that in all three order conditions, the variation in fit was due to ratio differences, and not to individuals, it was decided that all five habits (23lw, 34lw, ivw, ivwx and nouns) used in the previous chapter for variation within individual, would be tested. This time only natural order was used, although this increases the noise in the data. This was because the primary aim was to produce data applicable to insertions at unknown points.

1. Materials

The materials were the 4 written samples of text from 15 subjects; one text from each person was produced after watching the video and the other three contributed by the subjects as samples of their writing. This made a total of 60 texts.

2. Design

Cusums were calculated for excerpts of 15 sentences from each text. On the basis of the cusum curves, the total deviation measure was calculated. The Spearman's Rho value and ratio of sentence length to habit (M value) were also calculated from the original sentence length and habit data. The instance of each habit per 15 word block was also noted for a total of 15 blocks. This was equivalent to the 15 sentences used except there was no need to take into account the problems of punctuation.

On the basis of the M value cusum insertions were carried out. The two excerpts which showed the closest M values were combined together and the cusum and related measures were calculated. Then the two excerpts with M values that were furthest apart were combined in the same way. This was done both between and within subject. The result was a set of 60 combination texts, 15 in each cell of the design. Table 3.3 shows the design of the experiment.

Condition	Within subject	Between subject
Close	Combination for each subject of closest M's of 4 texts = 15 combinations	Combination of 15 closest M's of 60 texts = 15 combinations
Far	Combination for each subject of furthest M's 4 texts = 15 combinations	Combination of 15 closest M's of 60 texts = 15 combinations

Table 3.3: Experimental Design

Ten sentences of one text were inserted in another 15 sentence text after the 5th sentence making a text which was 25 sentences long. This 25 sentence text could therefore be made from 2 texts from the same author (**same** condition) or from 2 texts from 2 different authors (**different** condition). If the m values of the combined texts were close (range of difference = 0-0.17, mean = 0.021) they were put in the **close** condition, whereas if they were far apart (range = 0.003-0.097 mean = 0.044) then they were termed in the **far** condition.

The Total Deviation and Rho Values (Canter 1992) of the cusum of this resulting 25 sentence text were taken as a measure of goodness of fit.

3. Results

A 2(close/far) x 2(same/different author) x 5(habit measures) analysis of variance was carried out on the deviation scores, and showed that the deviations in the close condition were less than the deviations in the far condition ($F = 5.318$, $df = 1, 276$; $p < 0.05$). There was no difference for same/different author ($F < 1$; $df = 276$). The interaction of these two just failed to reach significance ($F = 3.78$; $df = 1, 276$) and there is little evidence to suggest that same/different author affects the results in any reliable way. Although there is a main effect of habit type ($F = 14.75$; $df = 4, 276$; $p < 0.01$), there is no interaction of this with any other variable. This means that all the habits contribute to the overall pattern and do not differ from one another. The overall pattern of results show that in the case of texts made out of two others, the deviation of the curves from one another is a function of underlying habit ratio, not of whether the text was derived from the same individual or two different ones.

Figures 3.7 (a) - (d) show these trends for 23lw, 34lw, ivw and ivwx where the total deviation for the close condition is less than for the far condition, independent of number of authors.

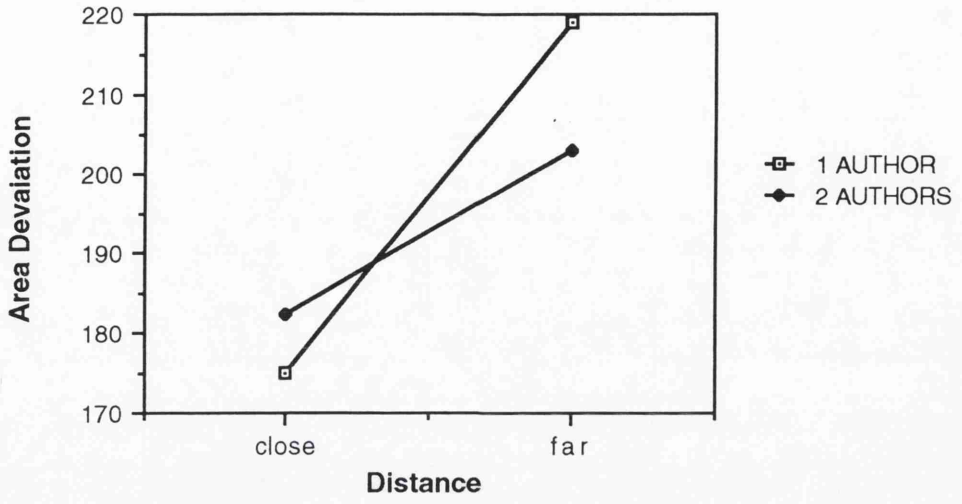


Figure 3.7(a) - 23lw

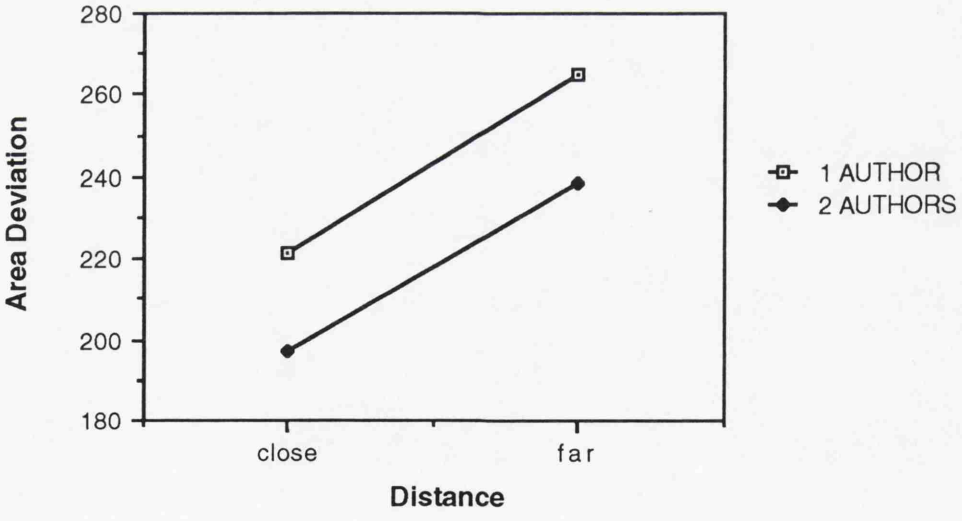


Figure 3.7(b) - 34lw

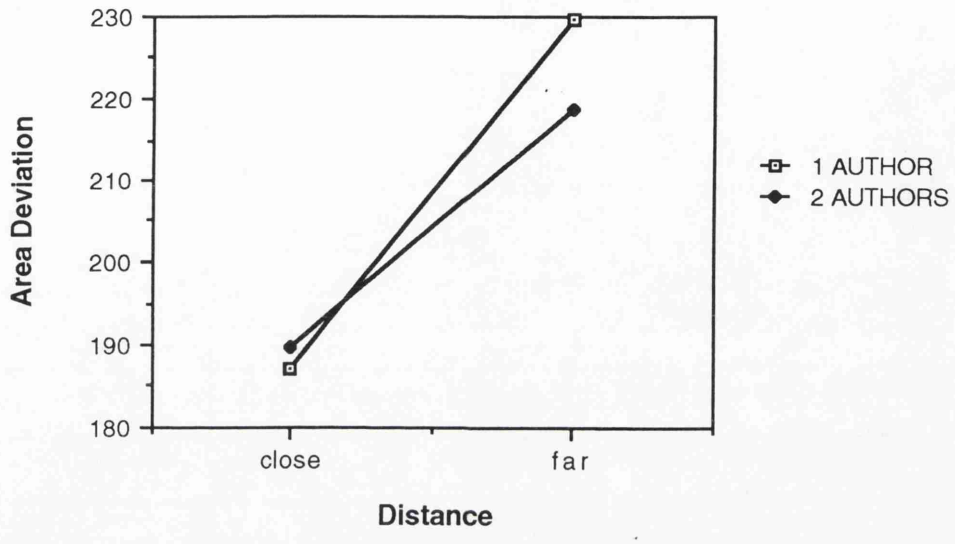


Figure 3.7(c) - ivw

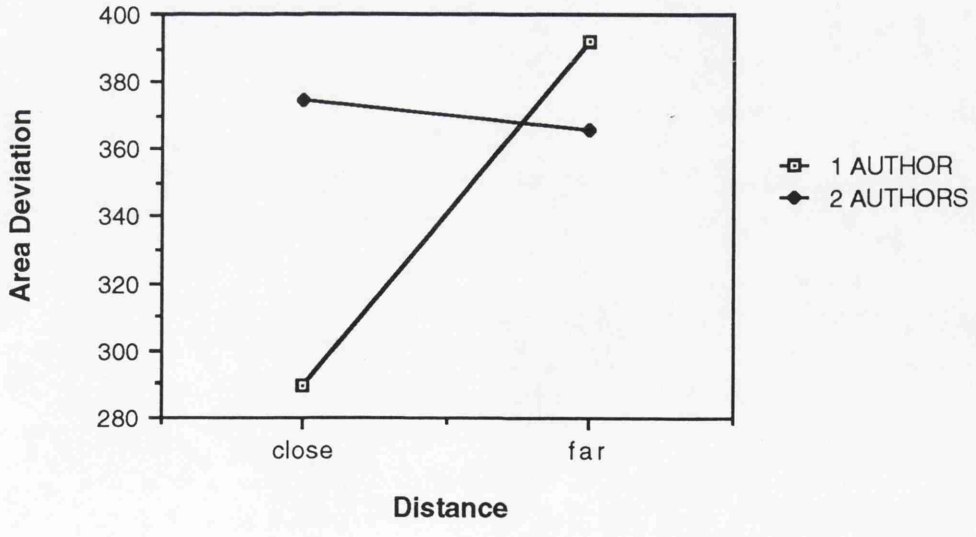


Figure 3.7(d) - ivwx

4. Discussion

The argument made previously was essentially that the ratios obtained using the techniques described above predict some of the behaviour of the cusum curve in just the way Morton would predict it. If writing from a person with a radically different habit (ratio) from another is substituted into the writing of the other person, great separation of the curves is likely to occur. On the other hand, if there is little difference in habit (ratio) between two people, then intersubstitution should have little effect. Using 23lw ratios, in the first experiment it was shown that this was the case, and also that if the writing of the same individual is substituted into another piece of writing showing a big difference in ratio, then that too shows a great deviation. Given that within person variability is as great as between person variability, this completely destroys the credibility of Morton's technique.

In the second analyses, the same procedure was used to examine the behaviour of all five habits. It was shown that for all habits tested, the results from the first study held. The fit of the cusum curve is not dependent on whether one or two authors are combined. It is the difference in ratio between the two texts combined which determines the good fit or otherwise of the cusum curve.

CHAPTER 4 : Two further Studies using CUSUM

A. Introduction

Two further studies are reported here. In the first, the contention that the habit (ratio) of an individual does not differ for speech and writing, is tested. In the second, a very large scale study of cusum deviation within and between individuals, is reported.

1. Comparison of Speech and Writing

According to Morton and Michaelson (1990) the cusum technique can be employed equally well to both speech and writing. The example he uses to illustrate this is one in which he combines a series of excerpts taken from the novels of Iris Murdoch and combines them with an interview which she gave on television. Morton apparently examined several habits and found that there is no change in habit between the samples for use of nouns. From this one example he maintains that the cusum method holds over both the spoken and written forms and indeed he upholds this conviction when queried.[5] This, as Morton acknowledges "confronts one of the most formidable obstacles to the acceptance of a scientific system by writers or critics. It is the profound psychological conviction that their intention to write, or speak differently to different people in different circumstances, must change their style." (Morton 1990, p 26).

---

[5] e.g. at first conference of Society for Forensic Linguistics, University of Birmingham, November, 1991.

Although both the spoken and written word can be said to be analogous on some level, in that they both constitute **language** and may be used to say the same things, there are ways in which a distinction between the two must be made. Firstly, there are different functions for which each is appropriate. The context in which writing is appropriate is not always the same as where speech would be appropriate and vice versa. Indeed this would only mean that the two forms were duplicating the function of the other. Secondly, many prosodic and paralinguistic features that are present in speech, are not present in writing and in speech there are none of the 'signals' that are present in writing. Speech is usually spontaneous whereas writing is not. Writing is usually planned, whereas speaking often, is not.

These observations add up to the fact that even on the surface it is very unlikely that the sum total of constituents within writing and speaking are comparable by a method which plays on many of the features which are reflected in the speech writing distinction.

Two issues must be confronted if the cusum technique is to be used with speech. First, a practical one: the transcription of speech. The second one is theoretical: there is good a priori grounds to suppose that the 'habit structure' of speech and writing will differ.

2. Problems Transcribing Speech

The problems which arise on the transcription of spoken text are clear to most researchers in the area. Apart from the editing out of sounds and non-words produced by the speaker, it is not always easy to decide whether pauses in speech should mark sentence boundaries or whether

the whole text makes only a few loosely constructed sentences. The comparative absence of any sentence markings in speech as compared with the abundance present in writing, would raise questions as to whether it is possible to compare the two, and also as to the validity of the sentence in the analysis of spoken language. In addition, speech is often ungrammatical (e.g. see Anderson et. al., 1991, on Map Task corpus).

Halliday (1990) argues that the basic unit of syntax is the clause. He says that the clause can occur singly or in a clause complex and that clauses and clause complexes are both essential for the study of spoken and written language. Other researchers also recognise that there is a problem with the use of sentences in the analysis of spoken language (e.g. Quirk et al(1985)). Linell (1988) reasserts the lack of clear cut sentences in spoken language and also points out that speech consists of phrases and clauses loosely related to each other which combine into structures which are less clear and hierarchical than the structures which are normally dealt with in grammars.

Conversely, there are those who defend the reality of observable sentences in spoken language. Chafe and Danielewicz (1987) assert that both clauses and sentences have their place in the analysis of spoken language. They stress the importance of what they call "prototypical intonation units" consisting of a single intonation contour which may often be a phrase or simply a fragment of syntax. They also say that although the sentence is a controversial unit of analysis in the spoken genre, speakers do appear to produce sentence-final intonation when they think that they have come to the end of some coherent sequence.

However, there may not always be agreement in what intonation signals the end of a coherent sequence. Miller (in press) quotes an experiment by Wackernagel-Jolles (1971) which found that when students were asked to indicate sentence boundaries in unpunctuated narrative transcripts, agreement as to the position of sentence endings ranged from 13 out of 20 in one text to 6 out of 29 in another. The current controversy about what text units can be recognised in spoken language for transcription still rages.

3. The Issue of Lexical Density

Morton claims that constancy of habit holds over different modalities so that the writing of a specific individual will have the same properties as that of his or her spoken material. The lexical density hypothesis (e.g. Halliday, 1990) would suggest otherwise. This well-substantiated hypothesis shows that lexical density, that is, frequency of content words as distinct from closed class function words such as prepositions and connectives, is higher in the written modality than in the spoken. In terms of the habits Morton uses, this means that there should be a higher ratio of nouns and correspondingly, a lower ratio of function words in writing then in speech. This runs counter to Morton's assertion that the ratios should be equal in the two modalities.

B. Study 1 - Comparison of Spoken and Written Material

The issue investigated in this experiment, is whether or not the ratios of habits are consistent over spoken and written language, as Morton holds, or whether, in line with the lexical density hypothesis, they would be different. It is expected that the habit indicators which reflect the

proportion of function words would differ in speech and in writing. Since 2 and 3 letter words, 3 and 4 letter words and initial vowel words generally tend to be function words, the frequency of these should be higher in spoken language, even within the same individual. On the other hand, content words such as nouns, should have a lower frequency.

1. Subjects

The same 20 subjects were used as in the previous experiment. Ten were students from the University of Glasgow and 10 were staff. (All spoken and written texts were available from these subjects. The number dropped to 17 after this data had been collected.) Half of each subgroup carried out the two recall tasks in a given order.

2. Material and Method

The design of the experiment was that used in chapter 3. To recap briefly, the subjects were required to watch a video which consisted of a number of episodes which revolved round a series of people carrying out various activities and hobbies. There was no dialogue in the video and it was set to music. It lasted about 7 minutes from start to finish. After the video half the subjects were required to describe verbally what had happened and half were required to describe this in writing. The next day they returned, and those who had spoken on the first occasion now produced a written version, while those who had spoken produced a written version. All of the spoken monologues were recorded and later transcribed.

The object of this experiment was the comparison of the spoken and written video material. Because of the problems with transcription, cusums were calculated on two different bases. The first defined sentence boundaries purely on the judgement of the transcriber of the tape. The second group of calculations was carried out on successive blocks of 15 words regardless of what grammatical fragments they contained, so that the issue of sentence boundary did not arise.

3. Basic Calculations

Pauses (such as er, uhm, etc.) were edited out of the speech transcripts and full stops were inserted at what appeared to be sentence boundaries. Again, the purpose written program COUNT (see Appendix 2) was applied to both the spoken and written materials and the following habits were calculated.

(a) proportion of 2 and 3 letter words in each sentence of each sample (23lw)

(b) the proportion of 3 and 4 letter words in each sentence of each sample (34lw)

(c) the proportion of words in each sentence that started with a vowel in each sample (ivw)

(d) the proportion of ivw which were not 23lw in each sentence, in each sample (ivwx).

In addition a hand count was made of

(e) the proportion of nouns (excluding pronouns but including classifiers) in each sentence in each sample.

This data along with the sentence length data, was used to calculate the cusum curves and the cusum measures as described in the general chapter on cusum.

4. Fixed Sample Calculations

Since it could be argued that sentences are arbitrary in speech the ratios were also calculated based on a fixed sample, which disregarded sentence boundaries. In the present experiment, the fixed sample was 15 words so the number of occurrences of each habit per successive samples of 15 words was calculated for both the speech and writing samples. This allowed the comparison to be made without concern for defining sentence boundaries in speech.

5. Results and Discussion

A direct comparison was made between the ratios produced for the spoken and written materials. These are shown in Figure 4.1.

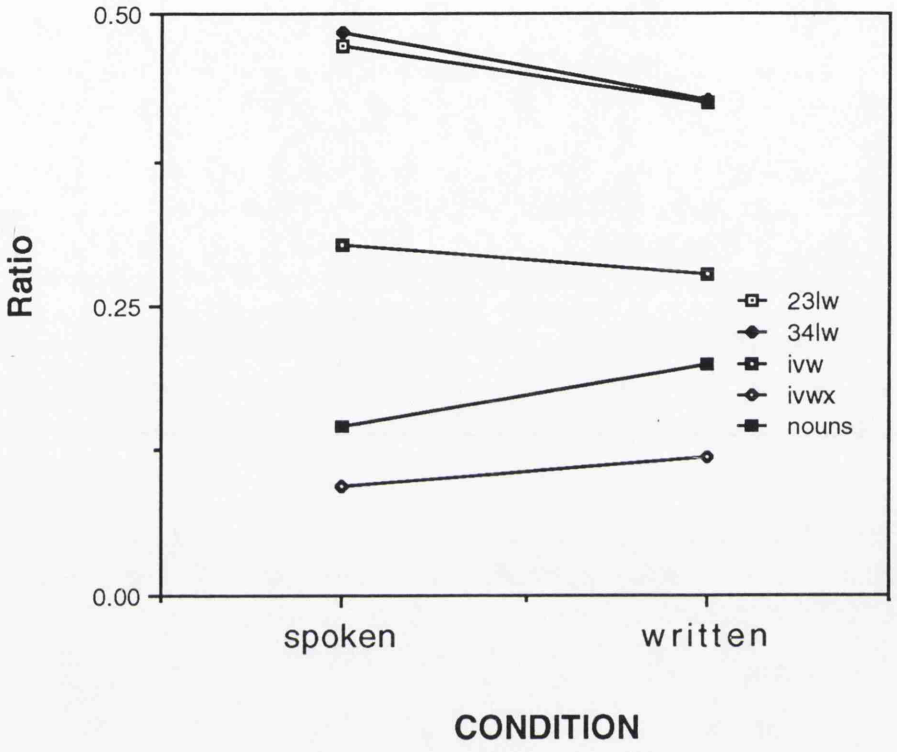


Figure 4.1: Comparison of Ratio on spoken and written samples in sentence condition

Contrary to Morton's claim, several of the results show statistically reliable differences. The ratio of 23lw to total number of words is higher in the spoken than in the written versions ($t = 4.38$; $df = 19$, $p < 0.01$); so is the ratio of 34lw ($t = 2.13$; $df = 19$, $p < 0.05$); the ratio of ivw is lower also ($t = 3.29$; $df = 19$, $p < 0.01$). These results are consistent with the lexical density hypothesis rather than with Morton. Also consistent with this is the fact that the ratio for nouns is higher in the written version ($t = 6.39$; $df = 19$, $p < 0.0001$).

Figure 4.2 shows a comparison of the average number of each habit word per 15 word block, in the two conditions, and shows an almost identical pattern.

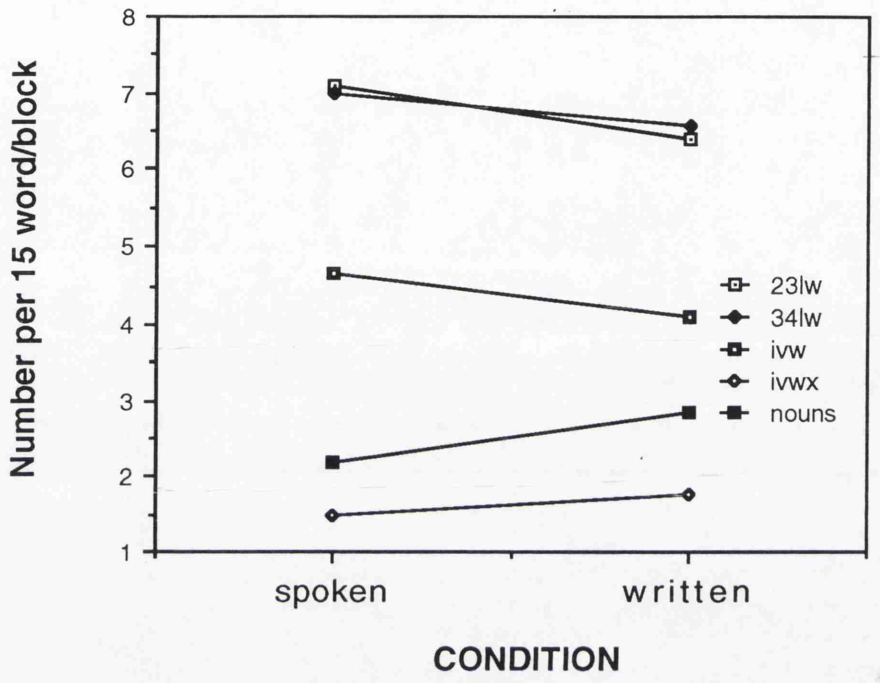


Figure 4.2: Comparison of spoken and written samples in block condition

The results in Figure 4.2 are consistent with lexical density differences and not with Morton's contention. In the spoken condition, there are statistically reliable results showing higher ratios for: 23lw ($t = 4.23$; $df = 19$, $p < 0.01$), 34lw ($t = 2.34$; $df = 19$, $p < 0.05$) and ivw ($t = 3.31$; $df = 19$, $p < 0.01$). The ratio of nouns is again lower in the spoken condition ($t = 4.88$; $df = 19$, $p < 0.01$). Therefore, whether using sentence boundaries, or a fixed sample calculation support is for the lexical density hypothesis and not for Morton's claim that ratios will be constant across modalities.

C. Study 2: A Large Sample Comparison of Same and Different Author Cusums

Other investigators into Cusum (e.g. Hardcastle, 1993; Canter, 1992) have focused directly on the shortcomings in interpreting cusum charts per se, mainly because it is this method which is presently being used in the courts as expert witness testimony. A criticism could be made of the present investigations up to this point, for the fact that they have not focused on the cusum technique itself. Rather, they have concentrated on underlying assumptions about habit. This is reasonable because the investigations have shown that the deviation between the cusum curve for sentence length and that of the habit is very dependent upon the order of sentences of different lengths and the linear correlation.

The present study therefore provides an investigation into the claims of the technique using the cusum procedure itself. Although cusum comparisons have been made, the sample sizes used have never been large enough that the results could be conclusively put down to the inadequacies of the cusum technique itself rather than noise in the data.

The present study is designed to test whether there are global differences using two of the goodness of fit measures, Total Deviation and Rho, for samples from same or different authors.

1. Rationale

The aim of the experiment is to produce a large amount of cusum deviation data for both same-subject and mixed subjects text.

2. Materials

The materials used were a subset of those used in the previous experiment. They consisted of the written material which had been obtained from the 17 subjects participating in the video experiment. Each subject had 4 written materials, one of which had been produced as a result of watching of the video, while the other three had been provided as additional samples of his or her work. It was decided that the spoken material should be left out for this experiment due to the large difference between the speech and writing samples which were shown in the previous study.

3. Design

The experiment was designed to test whether differences were evident in the cusum fit measures used between same author and different author cusums over a large number of combinations. In order to do this, every written text was combined with every other written text independently of whether the individual texts were from the same or different authors. The cusums of these combinations could be calculated along with Total Deviation and Rho measures and a comparison made between the cusum combinations resulting from the two types of combination.

A program was specially designed for this purpose, a more sophisticated version of the COUNT program. The COUNT program originally counted the number of letters in each word and listed the number of words of each length for each sentence. It also determined the number of words which started with a vowel and therefore by default the number of 2 and 3 letter words that did not start with a vowel.

The additional program (called CUSUM) therefore used the output of the COUNT program which had already been applied to the texts. Initially it was ensured that equal numbers of sentences were being compared when the two texts were combined since this could inevitably affect the total deviation measure. It was decided that the combination samples should be 25 sentences long in total since this was an adequate sample size according to Morton (1992) on which to apply the cusum technique. In order to ensure this, 10 sentences from the COUNT output of one text was inserted into the middle of 15 sentences of the COUNT output of another text. This was done individually by the program for the four habits being tested (23lw, 34lw, ivw and ivwx).

CUSUM therefore separated out the middle 10 and the middle 15 sentences of each text and inserted the 10 after sentence 7 of the middle 15 sentences of the text giving a COUNT output of 25 sentences. This was done for every text and the combination was made with every other text including itself and for each of the four habits. All habits were compared with sentence length for the cusum calculation. If any text happened to be shorter than the minimum length of 15 sentences then it was omitted from the

calculation. In this way, two very large sets of cusums would be calculated for both same and different author combinations (same = 264, different = 4224).

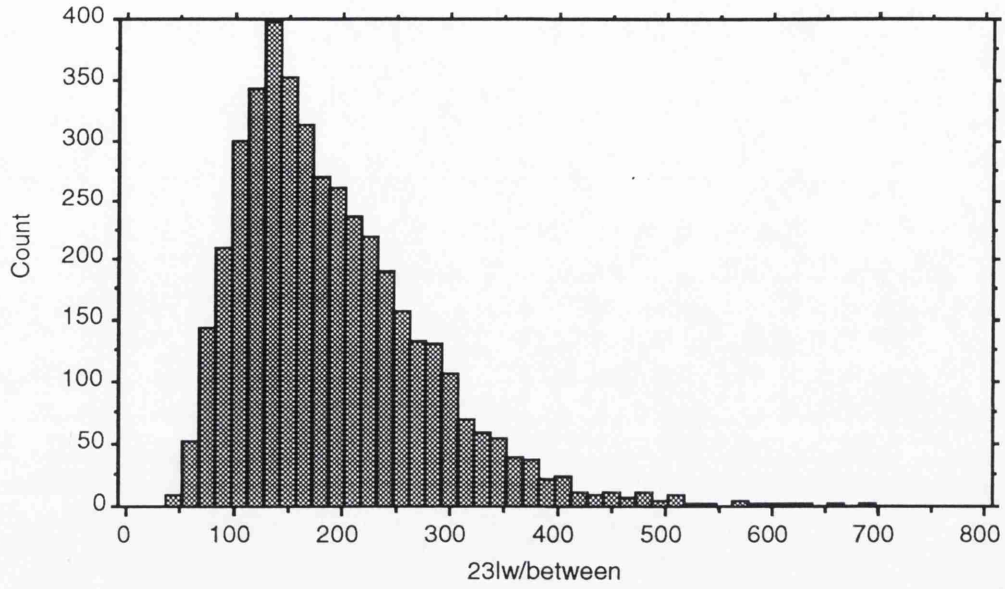
After the texts had been combined, CUSUM calculated the cusum on each of the combination texts. From this, the Total Deviation measure and a measurement of Rho was also calculated the results combinations separated in to those consisting of the same and different authors. An example output from CUSUM can be seen in Appendix 4.

Thus end product from the CUSUM program was a large number of Total Deviation and Rho scores, both for the same author group and the different author group. These were subsequently arranged in to a frequency distribution; one for each habit (23lw, 34lw, ivw, ivwx) and for each measure (area deviation or rho). There were thus 8 frequency distributions in total, four for Area Deviation and four for Rho.

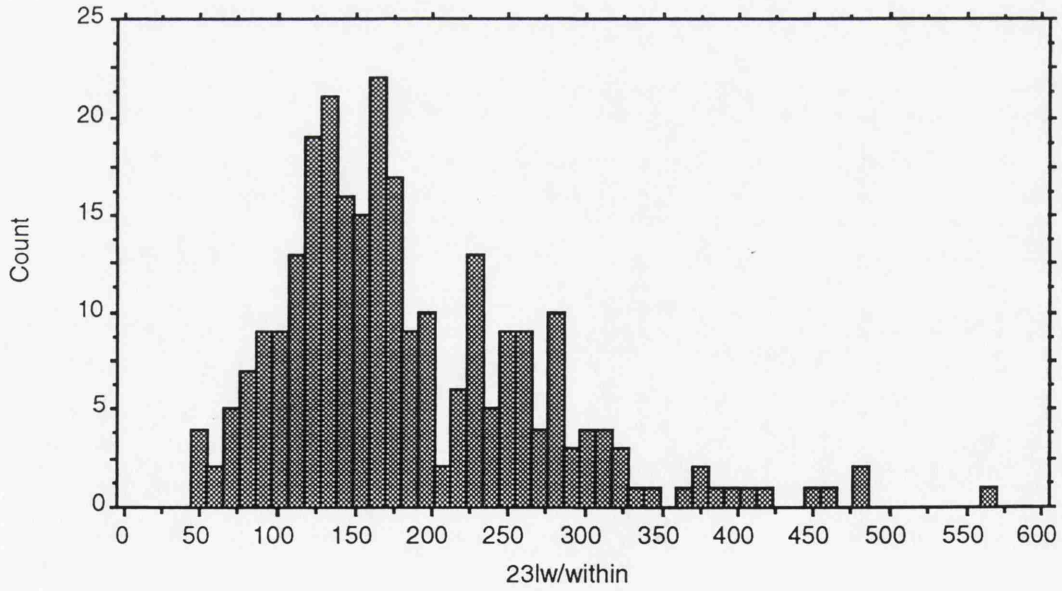
If the cusum technique distinguishes between same and different author combinations, the frequency distributions on goodness of fit scores should be statistically different. If on the other hand, the cusum technique does not differentiate between same and different author combinations, no significant differences will be found between the two distributions in any condition.

4. Results

The frequency distribution graphs for 23lw, both between and within subjects for Total Deviation using Morton's scaling method can be seen in Figure 4.3(a) and Hardcastle's scaling method in Figure 4.3(b), and the same for Rho in Figure 4.4.

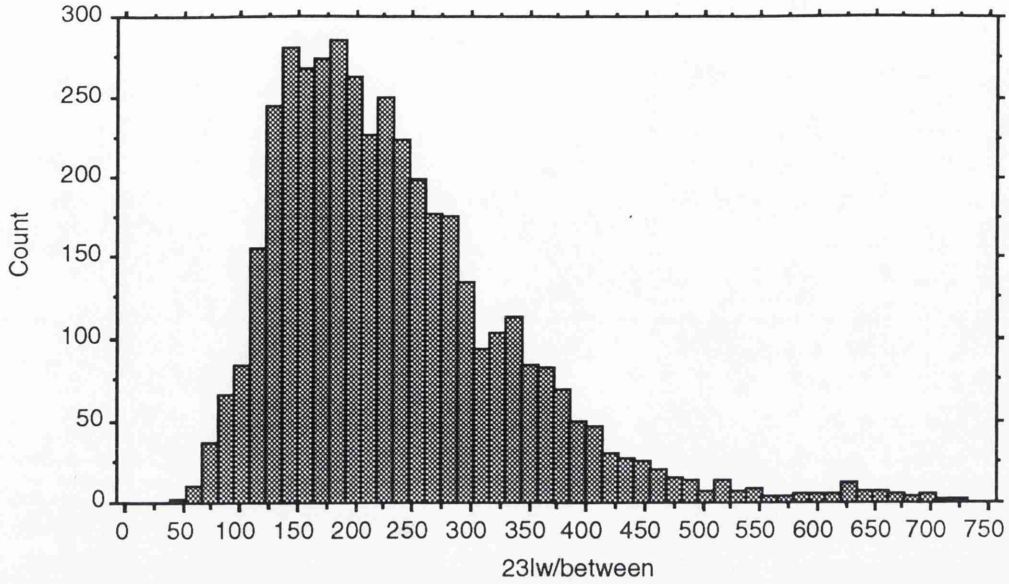


Frequency distribution for 23lw (different subject insertion)

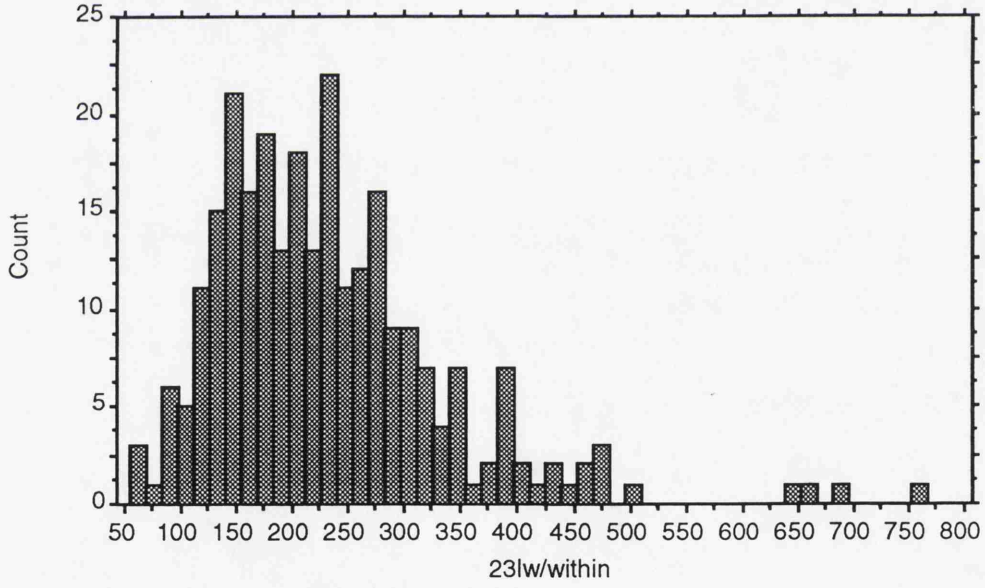


Frequency distribution for 23lw(same subject insertion)

Figure 4.3 (a): Total Deviation (Morton's scaling method)

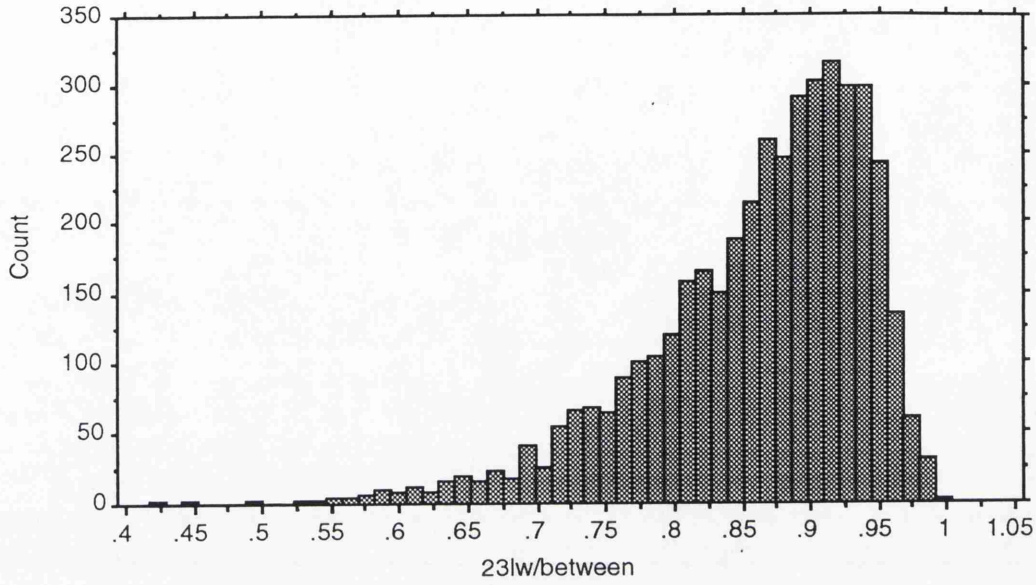


Frequency Distribution for 23lw (different subject insertion)

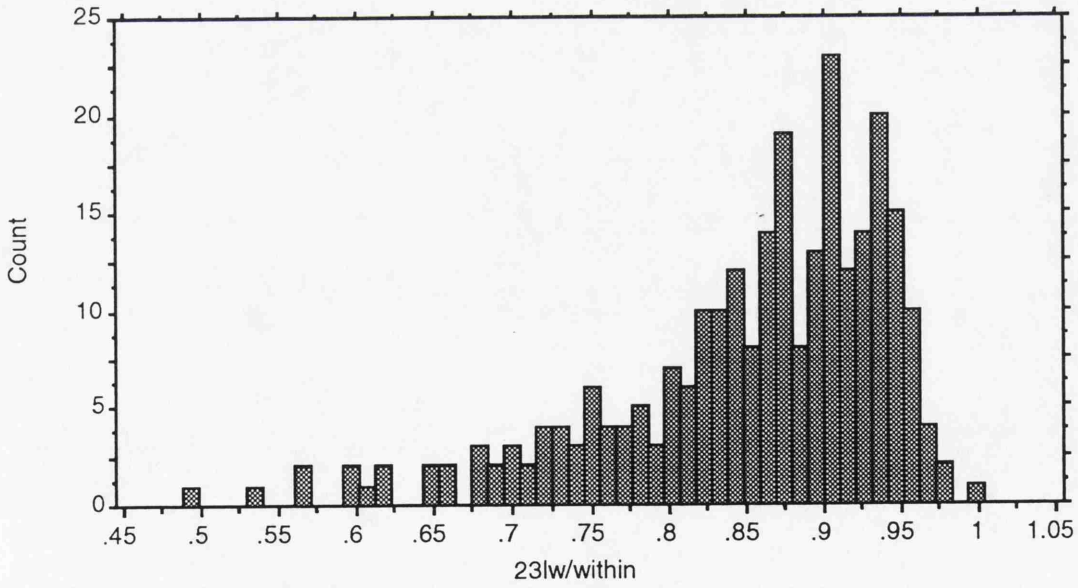


Frequency Distribution for 23lw (same subject insertion)

Figure 4.3 (b): Total Deviation (Hardcastle's scaling method)



Frequency distribution for 23lw (different subject insertion)



Frequency distribution for 23lw (same subject insertion)

Figure 4.4: Rho

Figure 4.5(a) show a comparison of the between and within Frequency distributions for 34lw, ivw and ivwx using the Total deviation measure with Morton's scaling method. Figure 4.5(b) shows the same with Hardcastle's scaling method and Figure 4.5(c) shows the same for Rho.

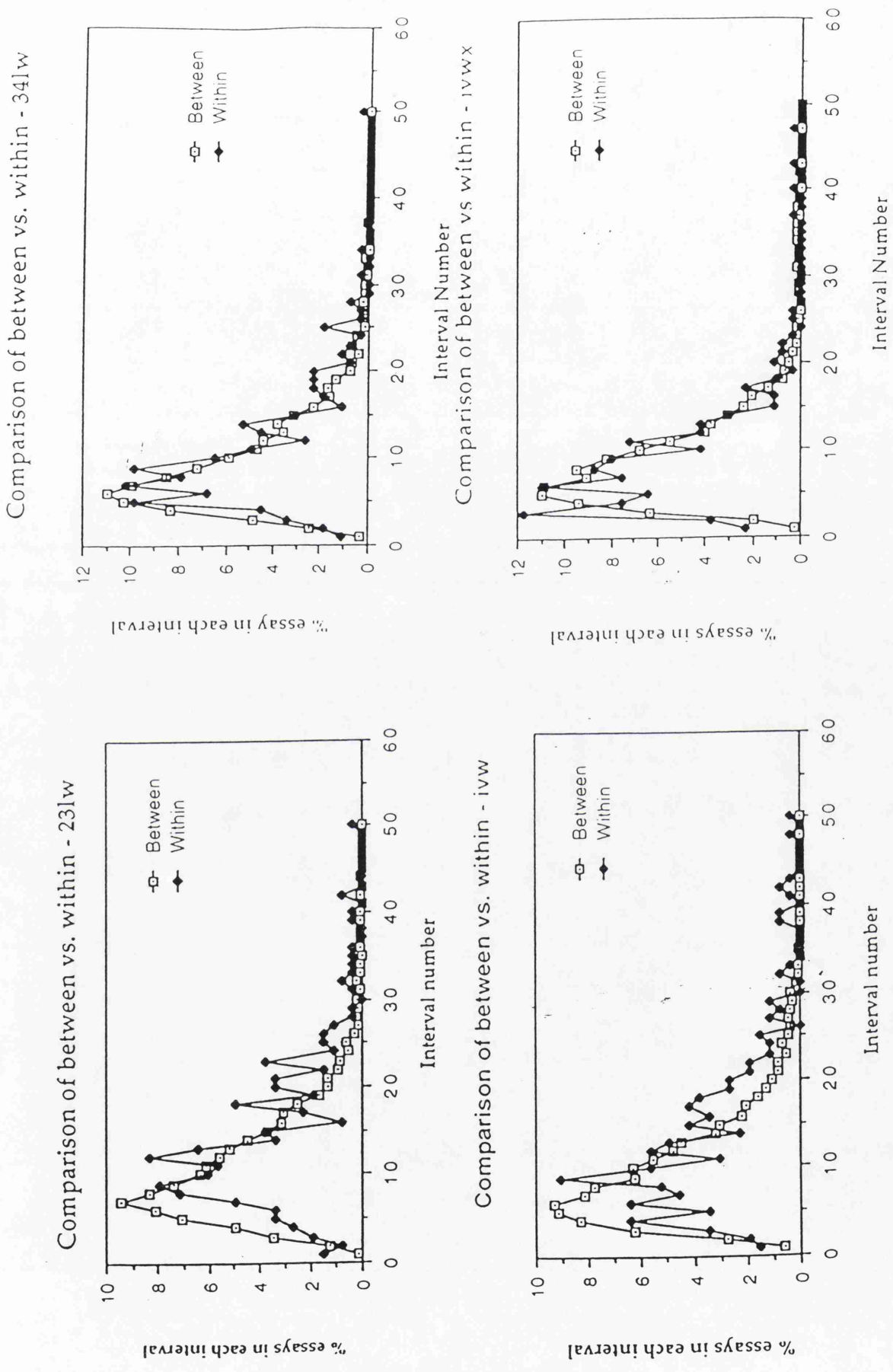


Figure 4.5(a): Frequency distributions using Total Deviation Scores for 34lw, ivw and ivwx (Morton's sclaing method).

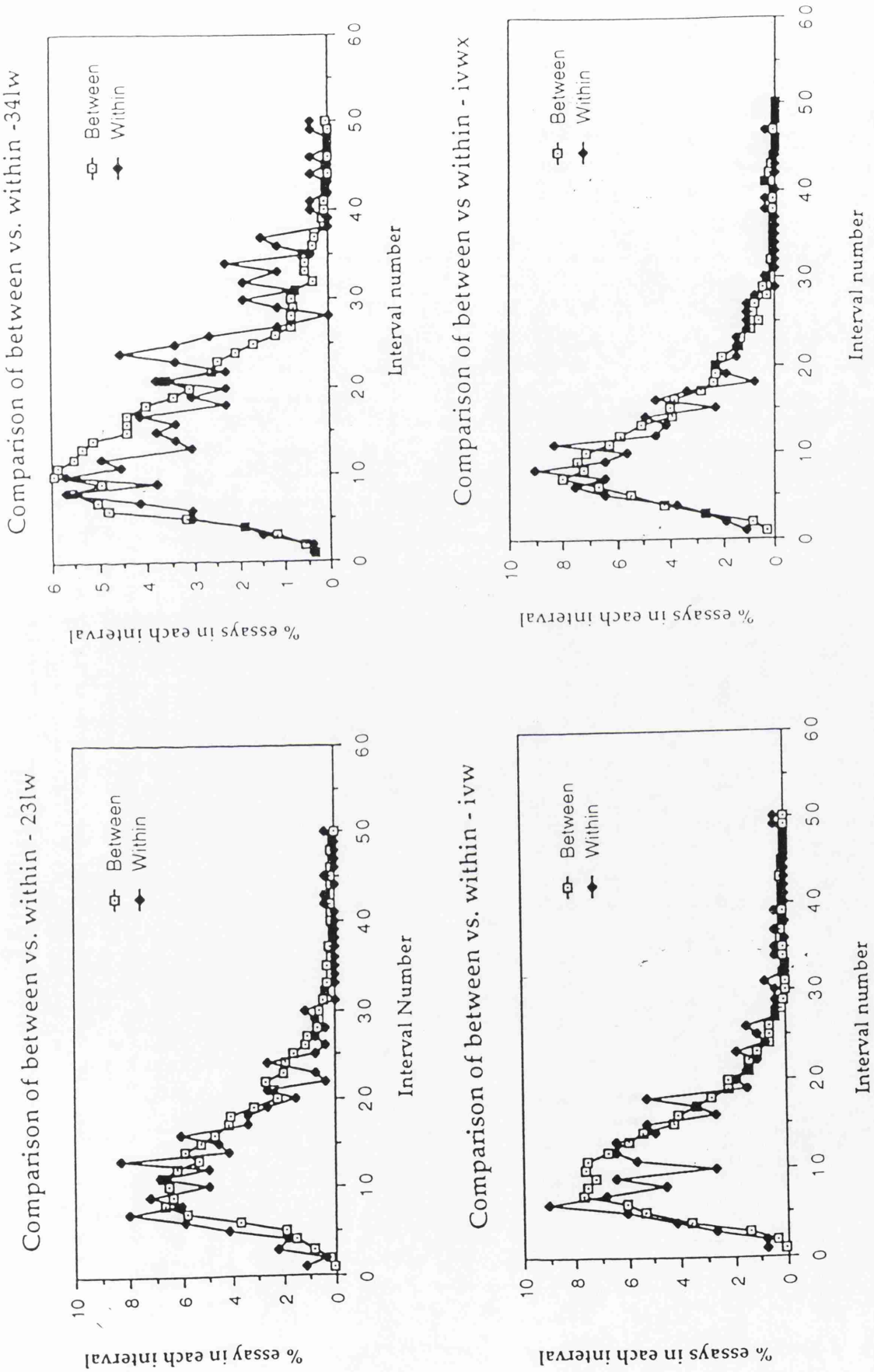


Figure 4.5(b): Frequency distributions using Total Deviation Scores for 341w, ivw and ivwx (Hardcastle's scaling method).

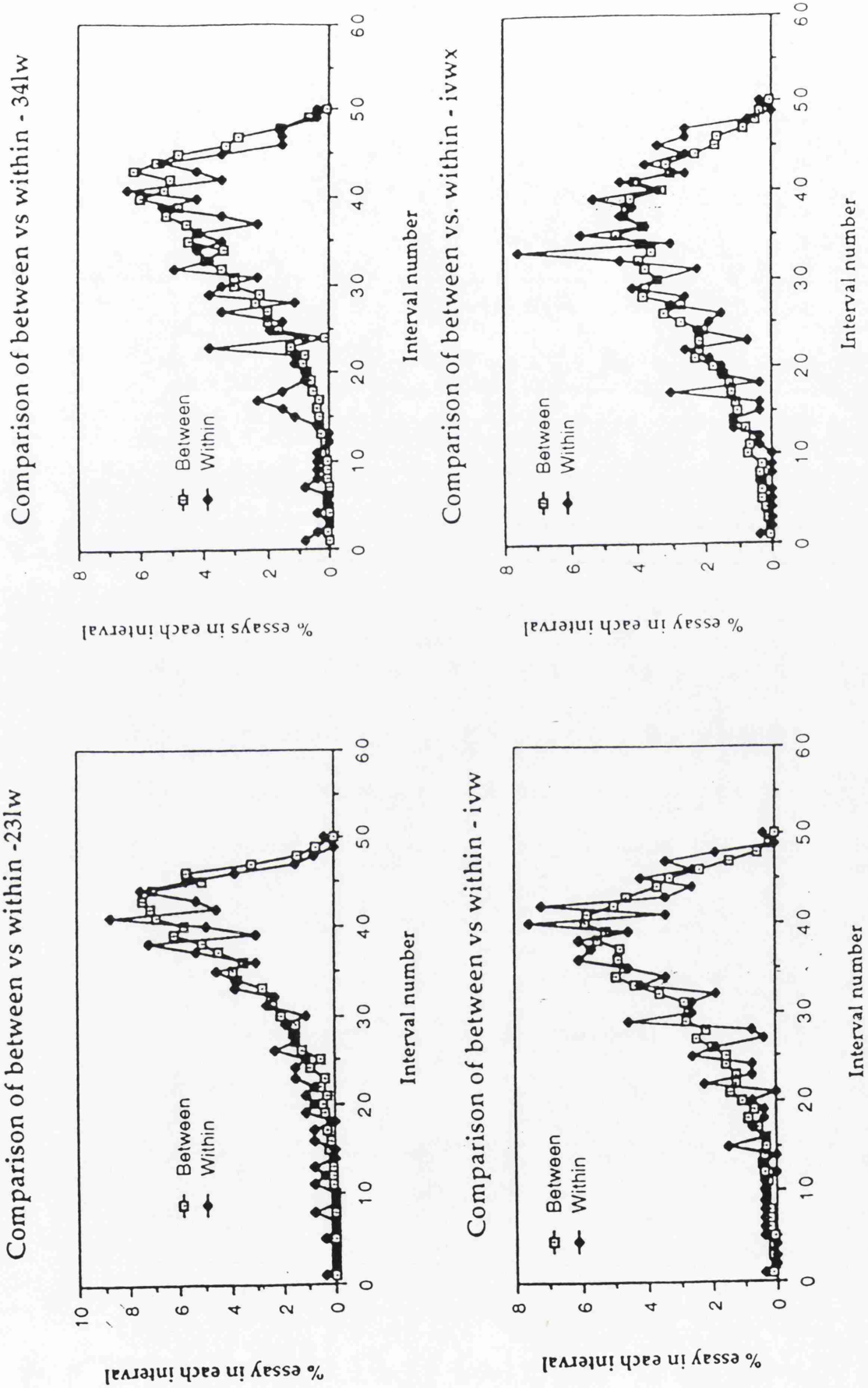


Figure 4.5(c): Frequency distributions using rho scores for 34lw, ivw and ivwx (Rho)

Chi-squared tests

A comparison was made of the percentage of essays falling into each of the 50 intervals, for each of the conditions. The comparison was thus between within and between group conditions for each habit, for total deviation and for rho. Both Morton's and Hardcastle's scaling methods were examined for the Total Deviation condition, since Rho values were unaffected by the method of scaling. The method of comparison was the Chi-squared test between observed and expected frequencies. The model used specifies the **same** subject condition as the expected frequency since this was the baseline score of only one author. The between subject condition was tested against this. Hence the **same** subject conditions was the **expected** frequency and the **different** subject condition the **observed** frequency. A good fit between observed and expected frequencies, or within and between subject conditions would be indicated by a small chi-squared value.

The Chi-squared results are shown in Table 4.1.

Area Deviation: Morton's (1991) scaling method

Habit	Chi-squared	p <
23lw	41.406	0.771
34lw	15.95	0.999
ivw	5.407	0.927
ivwx	16.917	0.999

Area Deviation : Hardcastle's(1993) Scaling Method

Habit	Chi-squared	p <
23lw	28.815	0.990
34lw	17.473	0.999
ivw	21.487	0.999
ivwx	10.757	1.000

Rho

Habit	Chi-squared	p <
23lw	18.667	0.999
34lw	23.275	0.999
ivw	29.331	0.988
ivwx	19.145	0.999

Table 4.1 : Results of Chi-squared for 3 conditions and four habits

All of the chi-squared values show a very good fit, thus there is no significant different between the observed or different subject condition and the expected or same subject condition, for any of the habits.

As an additional test of similarity, the mean and mode values of each of the frequency distributions were compared. Table 4.2 shows the mean and mode values for the frequency distributions of each habit in the three conditions. The skewness figures for the same and

different author plots were also noted for each condition as a measure of the position of the distributions around the mean. These tables can be seen in Table 4.3.

Area Deviation : Morton's scaling method

	Different author		Same author	
Habit	**Mean**	**Mode Interval**	**Mean**	**Mode Interval**
23lw	192.3	128.6 - 143.7	185.6	160.2 - 170.7
34lw	237.5	156.3 - 178.5	221.4	152.2 - 169.2
ivw	203.5	133.4 - 150.6	197.7	140.8 - 152.4
ivwx	767.1	416.4 - 475.7	720.7	299.3 - 363.4

Area Deviation : Hardcastle's scaling method

	Different author		Same author	
Habit	**Mean**	**Mode Interval**	**Mean**	**Mode Interval**
23lw	233.46	177.02 - 190.88	232.87	226.21 - 240.44
34lw	286.51	204.51 - 220.40	278.19	158.40 - 170-52
ivw	268.19	166.13 - 184.92	269.85	151.00 - 167.68
ivwx	426.57	273.25 - 301.19	429.961	308.78 - 337.09

Rho

	Different author		Same author	
Habit	**Mean**	**Mode Interval**	**Mean**	**Mode Interval**
23lw	0.862	0.909 - 0.921	0.854	0.901 - 0.911
34lw	0.814	0.890 - 0.902	0.806	0.882 - 0.892
ivw	0.825	0.874 - 0.884	0.812	0.861 - 0.872
ivwx	0.663	0.709 - 0.724	0.647	0.636 - 0.653

Table 4.2: Mean and Mode Intervals for 3 conditions and 4 habits showing same and different author combinations

Area Deviation: Morton's scaling method

habit	Different author	Same author
23lw	1.436	1.233
34lw	1.126	1.695
ivw	1.359	1.441
ivwx	1.410	1.772

Area Deviation: Hardcastles scaling method

habit	Different author	Same author
23lw	1.326	1.581
34lw	0.798	0.741
ivw	1.621	1.521
ivwx	1.410	1.471

Rho

habit	Different author	Same author
23lw	-1.120	-1.262
34lw	-0.881	-0.864
ivw	-1.015	-1.202
ivwx	-0.56	-0.619

Table 4.3: Skew scores for 3 conditions and 4 habits showing same and different author combinations

By inspection, there is no significant difference between the mode intervals or the skew scores in the same and different author conditions.

5. Conclusions

This investigation was based on the assumption that if there is any merit in the cusum technique, then there should be a discernible and reliable difference between the distributions of deviations in cusum curves for the same and different author conditions. The results showed that there were no discernible differences whatsoever.

There was a high goodness of fit of the distributions of the deviation scores, using either measure, or either scaling method, between same and different author conditions.

By statistical methods alone, therefore, it would seem unlikely that the distributions come from different populations. The goodness of fit of the cusum is likely to be as poor or as good whether a text has been formed by the same or by different authors. Thus the results of the previous chapters are confirmed, but the large sample of texts used in this study makes this conclusion undeniable; the fit of the cusum is an unreliable method of determining whether a text has come from one or more than one author.

D. Conclusion to Part 2

1. Recent Studies on the Cusum technique

Over the last few of years there has been a series of investigations from various quarters into the validity of the cusum technique. These investigations have attempted to address the claims of the cusum technique which briefly, are as follows:

1. The claim that the "habit" that is chosen for a particular individual will be consistent over time, that is, over an entire life-span.

2. The claim that circumstances or moods will have no effect on the chosen habit nor will the genre of the sample to be studied - whether it is taken from speech or from writing.

3. The claim that the technique as a whole is able to distinguish whether a text has been written by one or more than author, and also if there is additional material by the questionable author available, who the author of the text is likely to be.

4. The claim by Morton that the cusums render it impossible to disguise the authorship of a piece of text and that the technique is 100% accurate in determining whether a text is of single or multiple authorship.

As expected these claims have encouraged a lot of research, and, as with any other proposed scientific technique they have been

systematically tested. Many questions were raised, originating from issues of statistics, subjectivity, punctuation, the interpretation of the cusum charts, and indeed the long-standing debate of speech/writing differences.

The studies quoted below have all been published or are about to be published in the past 2 years, and have started from a neutral point of view attempting to test the claims that Morton has made.

Perhaps the most significant in-road to the question of the Cusum's validity, and especially to its validity in the court room situation, has been made by Canter et al. at the University of Surrey. In addition to collaborating with the law departments in order to provide a critical assessment of the cusum technique. As previously discussed, Canter developed the use of the Spearman's Rank Order Correlation coefficient (Rho) as a measure of validity of the Cusum technique.

Initially, he found that when an independent judge was used to asses a wide variety of texts for single or multiple authorship, 65% of these texts were judged to be of more than one author.

Canter then calculated the Rho values between the 2 and 3 letter word cusums and the sentence length cusums on a large number of single author texts on the assumption that a high Rho value (over 0.90) would imply single authorship. He found that in 48% of cases, the rho value did not reach the criterion and the text would mistakenly have been judged as being from more than one author.

When hybrid texts of more than one author were engineered, only 35% of these were found to be within the criterion of multiple authorship, implying that the other 65% would mistakenly have been judged as being from a single author using this criterion.

Canter concluded therefore, that the Cusum technique was less likely to show multiple authorship for more than one author than for single authorship and that it was "just as likely to identify two authors when there is one as when there are two". (Canter, 1992, p17).

Hardcastle (1992) also questions the interpretation of the cusum chart, and proposed that it would be more sensible to plot the vertical separation between the two habits so that if the proportion of habit words in the sample were equal to those in the text as a whole, then there would be no widening or narrowing of the gap between the two curves. Hardcastle also suggests that the scaling procedure which should be used should involve multiplication by the slope of sentence length to habit ratio.

Hilton and Holmes (1993) also present an alternative method to the cusum technique which is based on the use of weighted cumulative sums of sentence length and establish the difference between two texts by conducting quasi-t-tests on the cumulative sum weight for each text. A weighted cusum according to a specific formula is plotted on the vertical axis and the cumulative sentence length divided by the average sentence length is plotted on the horizontal axis. In this way a formal, statistical testing procedure can be carried out on the differences between two texts. Hilton and

Holmes' results indicated that using this technique cumulative sums "are not consistently reliable indicators of authorship." (Hilton & Holmes, 1993, p2)

Jamieson and Aitkin (1993) used three different methods on two groups of text to try and assess the validity of the cusum technique. The first group of texts was samples of 50 sentences from three books by the Bronte sisters and the second was from three books by Jane Austen. Using visual comparison methods, Canter's Spearman's Rho and the t-statistics used by Holmes, they found no consistency in the determination of authorship.

2. Shortcomings of these studies

The majority of these studies have failed to take into account the fact that the fundamental assumption of the cusum technique has no validity. They concentrate on the faults associated with the application of the technique itself, since there do seem to be high correlations in natural data between sentence length and the habits which Morton uses for his investigation. In addition, it is the technique as it stands which is presently being used in the courts, therefore, it is the technique as it stands which should be subject to testing.

However, it has been shown here that although these high correlations do exist, and they are extremely constant within subjects, they do not show any individual differences *across* subject. This means that before the cusum process has been applied, there are problems with the nature of the data to be used. If there are no

differences in the use of these habits across individuals, then the question of whether a piece of text is written by one or more authors is immediately confounded.

The majority of these studies have also failed to use a large enough sample size. There is always the possibility that results can be put down to noise in the data rather than any inadequacy in the cusum technique itself.

3. Comments by Morton

Morton and Farringdon have tried to address the criticisms raised against the Cusum technique. In "Identifying Utterance" (Morton & Farringdon, 1992) it is acknowledged by the authors that "the analysis has its limitations" (p6). The authors go on to quote situations in which the text is heterogeneous and therefore the cusum technique can be considered less reliable. The implication is, however, that if these situations are avoided, the cusum technique can be considered very dependable. The authors also address three objections which they say are always raised against the analysis. These, briefly, are as follows. Firstly, the assertion that "people speak and write differently on different occasions". Secondly, the assertion that individuals are incapable of deliberately changing their style, and thirdly, the problem that punctuation which is not original will not reflect the true sentence length to habit ratio for that author.

All three problems are addressed by Morton and Farringdon, by means of visual illustration. A cusum chart of part of a spoken

interview and a personal letter associated with the case of the Birmingham Six is used to illustrate the closeness of habit in the two modalities. The second illustration is from "The Booker Book" where the author (Brett, 1989) tries to imitate the style of four other Booker prize winners. Again the cusum chart illustrates the closeness of habit showing that whoever Brett is imitating, his own habit comes through.

Finally, the problem of punctuation is addressed and it is acknowledged that non- original punctuation renders the results "at least speculative" (p7). However, it is pointed out by Morton and Farringdon that the occasion when punctuation causes a problem in the analysis, is in fact, very rare, since it would only matter if the occurrence of the habit in question changed from one part of a sentence to another. It is only the proportion of habit words per sentence which matter so that only "wildly eccentric systems of punctuation would invalidate comparisons of this kind" (p7). Again this is illustrated using a visual example where punctuation has been tampered with.

Morton and Farringdon are therefore aware of some of the criticisms of the cusum technique. However, their response has been to address these criticisms using the very method which is being condemned. The illustrations used rely on fundamental assertion that there are habits which are constant within an individual, but which show variation across individuals. The latter has been shown to be without good foundation. The scaling methods of the charts and their interpretation have also been severely disputed, and yet this is an area which is ignored by the authors in an effort to

explain other objections. Neither has there been any attempt to explain, either linguistically, or psychologically, the production of a constant habit for a particular individual, nor has the proportion of cases that the technique has been unsuccessful been reported.

It is evident then, that there are still major drawbacks with the cusum technique which have not been addressed by its exponents.

4. Conclusions

There have been a number of critical studies of the cusum technique all of which have failed to find any results which substantiate the claims made by Morton. All have advocated that the courts should consider more fully the implications of using a technique such as this, without first subjecting it to a great deal of rigorous testing.

PART 3: ANALYSIS OF INDIVIDUAL CONSTANCIES IN SURFACE VARIABLES

CHAPTER 5 : Individual Differences in Stylecheck indices

A. Introduction

Part 2 of this thesis has focused on the Cusum technique as a measure to differentiate one writer from another. However, this technique is a very specific and somewhat convoluted measure of writing style. There have been other simpler measures proposed as indices of writing style, which take account of surface variables in the text such as word type or sentence beginnings. Programs which examine the incidences of surface variables of this type are termed stylecheckers. Cherry et al. (1980) looked at the stylechecker STYLE™ in this light and provided a full description of its terms and usage. On this basis, STYLE was used for the present study as a means of detecting and counting various potentially interesting stylistic parameters within a text.

In relation to the primary question - whether there are constancies in the writing of individuals - the comparison of these variables between and within subject provides an obvious initial basis for investigating the concept of individual constancies in writing style.

1. The stylechecker "STYLE"

The stylechecker STYLE™ was supported on the UNIX system and produces a series of counts of a total of 41 variables which fall into five categories:

1) readability indices,
2) sentence information,
3) sentence type,

4) word usage and

5) sentence beginnings.

1) The readability indices are the standard measures used to estimate the reading skills needed by the reader to understand a document. It has been shown that high scores are measures of stylistic difficulty. (Cherry & Vesterman, 1980) The formulae reported are the Kincaid Formula, Automated Readability Index, Coleman-Liau Formula and Flesch Reading Ease Score. These indices were omitted from the analysis as it was felt that they represented an amalgamation of a number of vague concepts and therefore would be of little help in the quest to pin down variables which showed constancies within an individual.

2) Sentence information concerns sentences in the text as a whole. It gives information on how many sentences and words a text consists of, their average length, how many questions, imperatives, and non-function words there are and the average length of non-function words and finally whether the sentence is classed as short or long. (Short = less than 21 words, long = more than 36 words). Several of these counts were omitted from the analysis due to the fact that they were extremely context dependent.

3) Sentence types fall into four groups; simple, complex, compound and compound-complex. A simple sentence contains one verb and no dependent clause; a complex sentence contains one independent clause and one dependent clause, each with one verb; a compound sentence contains more than one verb and no dependent clause; while a compound-complex-sentence contains either several dependent clauses or one dependent clause plus a compound verb in either the dependent or

independent clause (Cherry and Vesterman, 1980). The program detects these classes of sentences and derives the percentage of each in the text.

4) Word usage. This consists of counts of verb types (to be, auxiliary, infinitives) as a percentage of total verbs and passives as a percentage of non infinitive verbs. It also counts the word types as a percentage of the total words (pronouns, adverbs, conjunctions, nouns, adjectives, prepositions and nominalisations). Nominalisations are defined as verbs which are changed to nouns by the addition of a suffix but these were omitted for analysis from this section because even on prior inspection they showed almost no variation across subject. The majority of subjects used either no nominalisations at all or if they were used, they were in very low numbers.

5) Sentence beginnings. This lists the numbers of different ways in which a subject or sentence is opened. Quoted subject openers were nouns, pronouns, adjectives, articles and total subject openers. A problem arose however since only total subject openers was quoted as a percentage. It proved very difficult to identify specific subject openers and hence calculate percentages for the other openers. Because of this, and since all other variables used were quoted as a percentage only total subject openers were used. Sentence openers (prepositions, adverbs, verbs, subordinating conjunctions, conjunctions, expletives) are counted under this heading also however expletives as sentence openers were also excluded since like nominalisations, they almost never occurred therefore any analysis using them would produce a floor effect.

A example of the output of the stylechecker can be seen in Appendix 5.

2. Materials and Method

STYLE ™ was used to "trawl" on 125 pieces of text taken from 25 individuals (a total of 5 each). This originally produced 43 measures for each text. As mentioned above, number of the variables measured by the stylechecker were left out for the reason that either they were too context dependent (e.g. number of words/sentences), or because they measured very global aspects of the text within which any individual differences would be very hard to detect. (e.g. readability indexes). The remaining variables, numbering 23, were investigated for the 125 texts. The variables used can be seen in Appendix 6.

The problem to be tackled was whether or not subjects could be classified into groups using the measures produced by the stylechecker. Several methods will be used to address this problem, including Cluster Analysis, Correlation, and Factor Analysis. Initially, Cluster analysis will show whether essays can be clustered by subject according to their incidences of the variables measures by the stylechecker. Correlation will then be used for an initial investigation of the relationships existing between the variables, and Factor Analysis will investigate this more thoroughly, identify any underlying themes at work within the variables.

The problems to be investigated are therefore:
a) whether the stylecheck variables allow the essays to be separated out into 25 distinct subject clusters,
b) what correlations exist between the stylecheck variables themselves, and,

c) whether the stylecheck variables can be reduced to a smaller number of themes which will identify a "style space" into which normal individuals may fall.

B. Study 1 - Cluster analysis

This analysis was carried out in order to determine whether the 125 essays could be separated into 25 groups by subject on their use of the stylecheck variables.

Overview of Cluster Analysis

Cluster analysis can encompass several different classification methods, but in general it is a means of organising data into meaningful structures which are dependent on the distances or similarities between groups of objects in multi-dimensional space.

Hierarchical tree method

This method begins with the classification of each object into a single class. A linking process is then applied where the threshold for the decision as to whether two of these objects are the same or different is relaxed in such a way that larger and larger clusters of increasingly dissimilar objects are formed. The linking process continues until all of the "objects" are joined and a hierarchical tree results.

Distance Measures

The tree joining method uses measures of similarity or distance in order to join the objects. The distances used can be based on a single dimension or on multiple dimensions. Euclidean distances are a measure of the actual geometric distances between objects in two or three dimensional

space and are calculated when using the tree joining method of cluster analysis (Jackson, 1983).

Linkage rule

When each object represents a single cluster the measure of the Euclidean distances is straightforward. However, when the objects begin to be clustered into groups a specific linkage rule must be applied. Two clusters can be linked using the "nearest neighbours" method where linkage occurs if any two objects in the two clusters are closer together than the distances within the clusters. This is called single linkage and produces a chain of clusters each linked by a single object. Alternatively complete linkage links objects in the clusters which are furthest away from each other. A full description of the linkage rule can be found in Hartigan (1975).

Two-way joining method

It is possible to link cases, variables, or in the case of two-way joining, both cases and variables. The two way joining method is useful when it is hypothesised that both the cases and the variables will contribute simultaneously to any meaningful patterns which may emerge from the cluster (Hartigan, 1975).

K-Means clustering

This is the clustering method which should be applied when a hypothesis has already been formed about the number of clusters expected among the objects. In effect it is the reverse procedurally of an analysis of variance. An ANOVA would evaluate the between group variability against the within group variability, as opposed to the k-means clustering method which is trying to minimise variability within clusters and maximise

variability between clusters by moving the objects in and out of specific clusters in order to get the most significant ANOVA results.

1. The present data

Cluster analysis, and in particular the K-Means clustering method is tailored for the analysis of the present data. The hypothesis is that if similar style variable values hold for each subject for each of their five essays, then the 125 essays will be easily grouped into clusters according to subject. If k=25 (for the 25 subjects), then each of the 25 clusters will consist of five essays each from a single subject. Before the cluster analysis was performed, the raw data scores were standardised. (This means that each value was expressed in terms of its difference from the mean, divided by the standard deviation).

2. Results

The resulting clusters did not come close to representing a division by subject. The clusters of essays, shown by subject, are listed in Table 5.1, overleaf.

Cluster	Subjects contained in cluster
1	S1, S11, S15, S18, S19
2	S2, S5, S5, S5, S13, S14, S14, S15, S20, S22, S25
3	S6, S7, S8, S10, S24, S25
4	S1, S19, S24
5	S2, S2, S2, S4, S4, S6, S7, S9, S12
6	S14, S20, S21
7	S1, S2, S4, S4, S6, S9, S10, S12, S14, S19, S21, S24
8	S14
9	S5, S8, S8, S8, S10, S15, S16, S23
10	S7, S12, S17, S22
11	S9, S11, S12, S12, S13, S15, S18, S19, S23, S24
12	S3, S8, S9, S10, S10, S17, S17
13	S3, S3, S13 S13, S13, S22, S22, S23, S25
14	S6, S6, S7, S9, S11, S19
15	S1, S4, S15, S20, S25, S25
16	S16, S16
17	S1, S3, S11, S18, S24
18	S18, S20, S21, S23
19	S23
20	S7, S20, S21
21	S17
22	S5, S11, S18
23	S13, S17
24	S16, S16, S22
25	S21

Table 5.1: Clusters of essays produced by cluster analysis, shown by subject

Figure 5.1 below shows the number of essays from any one subject which occur together in any one cluster. The maximum number of essays from the same subject is five and the minimum is one as shown on the x-axis. Eighty-nine of the 125 essays occur on their own within a cluster. There are 12 occurrences of two essays from the same subject together, and only four times are there more than 2 essays by the same subject occurring together.

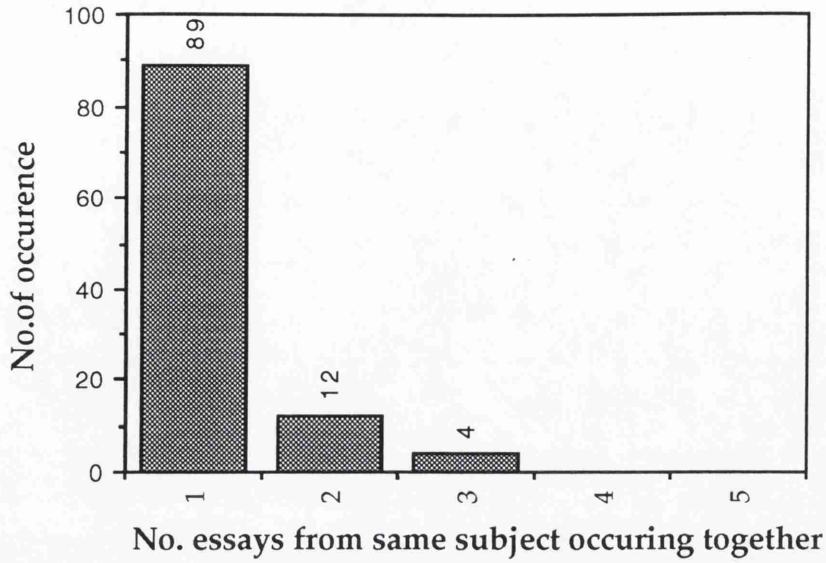


Figure 5.1 : Number of essays from any one subject occurring together

In order to determine the criteria that formed each cluster, the means for each variable on each cluster were noted. This allowed each cluster to be named by particular variables.

Table 5.2 shows the variables which have the highest loadings for each cluster and their mean values. Loadings were taken to be "high" if they exceeded the value of +1 but if no variables reached this value, the most extreme values were taken.

CLUSTER	No. CASES	HIGHEST MEANS	MEAN VALUES
1	5	V10 - Infinitives	1.92
		V22 - S.O.sub-con	1.36
2	11	V12- Prepositions	0.81
		V13 - Conjunctions	0.81
		V17 - Pronouns	-0.84
3	6	V5 - Complex	-1.17
		V7 - Compound	1.34
		V18 - S.O. total	1.22
		V20 - S.O. Adverb	-1.09
4	3	V1 - NFW	-1.55
		V6 - Compound complex	2.92
		V14 - Adverbs	1.90
		V15 - Nouns	-1.53
5	9	V18 - S.O. total	1.03
		V14 - Adverbs	-1.02
6	3	V5 - Complex	-2.01
		V21 - S.O. Verb	1.44
7	12	V7 - Compound	-0.64
		V18 - S.O. Total.	0.67
		V20 - S.O. Adverb	-0.62
8	1	V9 - Auxiliary Verbs	10.62
9	8	V5 - Complex	1.15
		V17 - Pronouns	1.06
		V20 - S.O. Adverb	1.14
10	4	V1 - N.F.W.	1.08
		V21 - S. O. Verb	1.75
11	10	V13 - Conjunctions	1.10
		V16 - Adjectives	-0.94
12	7	V3 - Long sent	-1.03
		V18 - S.O. total	-1.20
		V19 - S.O. prep	1.82
13	9	V5 - Complex sent	1.34
		V20 - S.O. Adverbs	1.09
14	6	V12 - Prepositions	-1.14
		V17 - Pronouns	1.11
		V22 - S. O. sub-con	1.88
15	6	V5 - Complex	1.12
		V8 - Verb to be	1.60
		V12 - Prepositions	1.14
		V14 - Adverbs	1.12
16	2	V2 - Short sent	1.87
		V8 - Verb to be	-1.87
		V10 - Infinitives	2.52
		V11 - Passives	-1.88
		V23 - S.O. Sub-con	2.88
17	5	V6 - compound complex	1.59
		V7 - Compound	-1.10
		V22 -S.O. Sub-con	-1.14
18	4	V1 - N.F.W.	2.35
		V4- Simple sent	1.80
		V15 - Nouns	2.16
19	1	V1 - N.F.W.	4.04
		V2 - Short sent	-3.40
		V4 - Simple sent	3.20
		V15 - Nouns	4.06
20	3	V4 - imple sent	1.12
		V7 - Compound	-1.37
		V11 - Passives	2.47
		V12 - Prepositions	1.64
		V18 - S.O.total	1.27
21	1	V7 - Compound	2.92
		V18 - S.O. total	-3.26
		V19 - S.O.Preps	3.22
		V20 - S.O. Adverbs	2.81
22	3	V7 - Compound	2.24
		V11 - Passives	2.24
23	2	V13 - Conjunctions	1.88
		V16 - Adjectives	-2.32
24	3	V23 - S.O. conjunctions	4.87
25	1	V21 - S.O. verbs	5.86

Table 5.2 : Highest Loadings on each of the 25 clusters

3. Discussion

Eleven of the clusters (2,5,7,9,11,12,13,14,15,16 and 24) did contain more than one essay from the one subject but in the majority of these cases, with the exception of clusters 16 and 24, the clusters were large and contained between 6 and 11 essays. Because of this, the probability of having two essays from the same subject in a large cluster is inevitably increased. In only four of the clusters which contain more than one essay from the same subject, is this number over 2. Clusters 2, 5, 9, and 13 contain three essays from the same subject but this is the maximum number from the same subject contained in any one cluster. Unless the clusters are clearly partitioned by subject, then it is likely that the pattern results are not due to a consistency in the use of these variables within subject. It can be concluded that the essays are not being clustered by an individual use of the style variables within subject.

The means for each variable were also examined. From this it was possible to determine on what criteria the clusters were being formed. The clusters were formed with patterns of variables with positive or negative weightings on the means for each variable. The highest means for each cluster (the most extreme, or exceeding + 1) were assumed to be the ones on which the cluster was being formed. Some of the clusters were formed on the basis of one variable alone (e.g. cluster 8 - auxiliary verbs, cluster 25 - verbs as subject openers) suggesting that they did not "fit" any where else. The majority of clusters were formed by a combination of variables. The essays in the clusters are thus related through high incidences of these combined variables.

Finally, a cursory examination of the topic of each essay was made. It was thought that the clustering may be related to specific topics of essay rather than anything at the more microscopic level of the style variables. This appeared perhaps to be the case only in cluster number 3 where 4 out of the 6 essays in the cluster were on schizophrenia and the remaining two were on psychological classification methods. Even this case contained more than one topic, so clearly other factors were at work.

In conclusion then, there was no distinct clustering by subject on these style variables.

C. Study 2- Intra-variable Correlation

The patterns of variation between the stylecheck variables themselves were the next line of investigation. If relationships between these variables can be identified for a large sample, then this may help in the identification of what could be termed a "style space" into which normal individuals may fall. Any deviation in these relationships for a given individual may be regarded as a characteristic of writing style.

1. Procedure and results

The scores for every text on 23 of the variables were intercorrelated using a Spearman's Rank order correlation to establish whether any trends were present. Spearman's Rho was used since some of the variables were not normally distributed but occurred either in low or in fairly high proportions. Table 5.3 below shows the significant correlations at the 0.05 and 0.01 levels of significance.

Non Function Words	Nouns	0.557**
	Adjectives	0.762**
	Pronouns	-0.544**
Short Sentences	Long Sentences	0.571**
Simple Sentences	Complex Sentences	-0.692**
	Compound Complex	-0.502*
Complex Sentences	Nouns	-0.439*
Verb "to be"	Passives	0.630**
Passives	Pronouns	0.480*
Prepositions	Nouns	0.432*
Adverbs	Nouns	-0.577**
Nouns	Pronouns	-0.588**
Adjectives	Pronouns	-0.532**
Subject Opener Tot.	S. O. Prepositions	-0.628**
	S.O. Adverbs	-0.606**
	S.O. Sub-Conjunctions	-0.423*

Table 5.3 : Significant Correlations between Stylecheck Variables (*sig at 0.05, **sig at 0.01)

2. Discussion

There are several interesting observations which can be made from the correlation study. The correlations which are reasonable in relation to the normal rules of grammar will be dealt with first followed by those which are more interesting in relation to writing style.

Non function words correlate positively with nouns (0.557) and adjectives (0.762). Since adjectives and nouns are in effect non function or content

words they produce a positive correlation. Pronouns produce a negative correlation (-0.544) since these are function words.

Simple sentences are highly inversely correlated with compound complex and complex sentences (-0.502 and -0.692 respectively). Since a simple sentence contains only one verb and no dependent clause, and both complex and compound complex sentences contain more than one clause, the correlation existing will automatically be an inverse one.

The verb "to be" is highly correlated with passives (0.630). The passive form of a verb will increase as the proportion of the verb increases.

Prepositions are correlated with nouns (0.432) since these two word types tend to be found together, with the preposition usually preceding the noun.

Nouns are negatively correlated with pronouns (-0.588). This is reasonable since if a text contains more naming words it will contain fewer pronouns and vice-versa, provided noun-phrase rate remains roughly constant.

Total subject openers are negatively correlated with three types of sentence openers; prepositions, adverbs and subordinating-conjunctions. This implies that as the number of subject openers which are nouns, pronouns, adjectives or articles increases, the number of adverbial, prepositional and subordinating conjunctional sentence openers will decrease.

The correlations up until now have been helpful in establishing the "normal" relationships between variables in a text but have been uninteresting in

that they are deducible from the rules of grammar. The correlations to follow are potentially more interesting in that they indicate parameters of writing style that are not deducible in this way.

The significant correlation between short and long sentences (0.571) does at first seem counter-intuitive. What this means is that in general, if an individual uses a high percentage of long sentences with his/her text, then he/she will also use a high percentage of short sentences. The definitions as used by the stylechecker of short and long sentences, are not strictly complementary, in other words, if a sentence is not short, it is not therefore automatically long. There is a gap between the definitions of long and short sentences, into which a sentence can fall. In this way, it is possible to have a high percentage of both short and long sentences. Individuals therefore use extremes of both short and long sentences.

Complex sentences correlate negatively with nouns (-0.439). It seems therefore that the more nouns there are in a sentence, the less complex it becomes. The number of nouns therefore varies inversely with the complexity of the sentence.

Adverbs are inversely correlated with nouns (-0.577). This would imply that when an individual is frequently describing the way in which a verb is "done" (the definition of an adverb) then the number of nouns they use is decreased. Similarly, if many nouns are in use then verb descriptions are decreased.

Adjectives are inversely correlated with pronouns (-0.532) suggesting that since adjectives are usually used to describe nouns, when a pronoun is

used instead of a noun, there is less likely to be an adjective to go along with it.

3. Comment

It must be stressed that many of these observations are directly implicated from word types. The value of this study lies in the fact that the movement of the variables in relation to each other is now evident. It is possible that an individual or a group of individuals could use these variables in ways which are different from the normal patterns demonstrated. That particular use could then be regarded as a function of the individuals' writing style. Factor analysis will help to establish any more complex relationships between the variables.

D. Study 3- Factor Analysis

Why Factor Analysis?

Two points have, thus far been established. Cluster analysis showed no sign of grouping of essays by subject and within the essays there are several expected movements of the variables with respect to each other. The logical progression of the investigation is to establish whether the 23 style variables can be reduced to a set of concepts which describe in a more concise way the variations in the "style space". It is expected that by using factor analysis, the 23 variables can be reduced to at least some sensible factors which will remain constant within an individual.

Factor analysis was chosen as the method of analysis because it allows the data to be reduced to a smaller number of variables. It is ideally used

when dealing with a data set which has no dependent or independent variables and for which common factors are required. If it can be found that for a set of say 30 variables, 6 common factors will explain a large percentage of the common variance then it is possible to use these 6 scores instead of the original 30 without losing any of the essential information.

Factor Analysis versus Principal Components Analysis

There are two basic types of factor analysis differing in the way that they explain the variance. **Factor analysis** allows for specific and joint factors. These therefore take into account unique, as well as common variance and allow for factor rotation. **Principal components analysis,** on the other hand, is less sophisticated and calculates a factor which will explain the maximum variance in all the variables. The next factor is then calculated so that it explains the maximum amount of the remaining variance. The second factor in a principal components analysis cannot therefore be correlated with the first one. In other words, they will be *orthogonal.* This process is continued until the variance of all the variables has been explained, i.e. until the number of factors is equal to the number of variables, thus meaning that the data set is not any smaller than it originally was. Factor analysis is therefore most useful for data reduction whereas principal components is more useful for detecting structure within the data. A fuller description of the distinctions between factor analysis and principal components analysis can be found in Jackson (1983).

1. The present data

Since data reduction was the aim of the study, Factor Analysis, as opposed to Principal Components analysis was the method used. The raw data scores on the 23 variables (see appendix 6) were noted for a total of 125 essays. These consisted of 5 essays from each of 25 subjects. As explained above, the subjects were all undergraduate students at Glasgow University and the essays had all been written under exam conditions before they were stylechecked to produce the scores on the 23 variables. The resulting data matrix of raw scores was standardised and then factor analysed to produce a factor matrix.

The factor matrix

The original factor matrix shows the first stage at which the variables used will have a common meaning. Each variable received a loading on each factor which is its correlation with the factor. There were 8 factors produced originally from the raw data. The factor matrix is available in appendix 7.

Halting the extraction process: Cattell's Scree Test

This test is a method of determining the optimum number of factors extracted before the intrusion of non-common variance becomes significant (Cattell 1966). A graph is plotted of eigen values against factor number, and the point at which the graph levels out is the point at which to stop extracting factors. The graph can be seen in Figure 5.2 below.

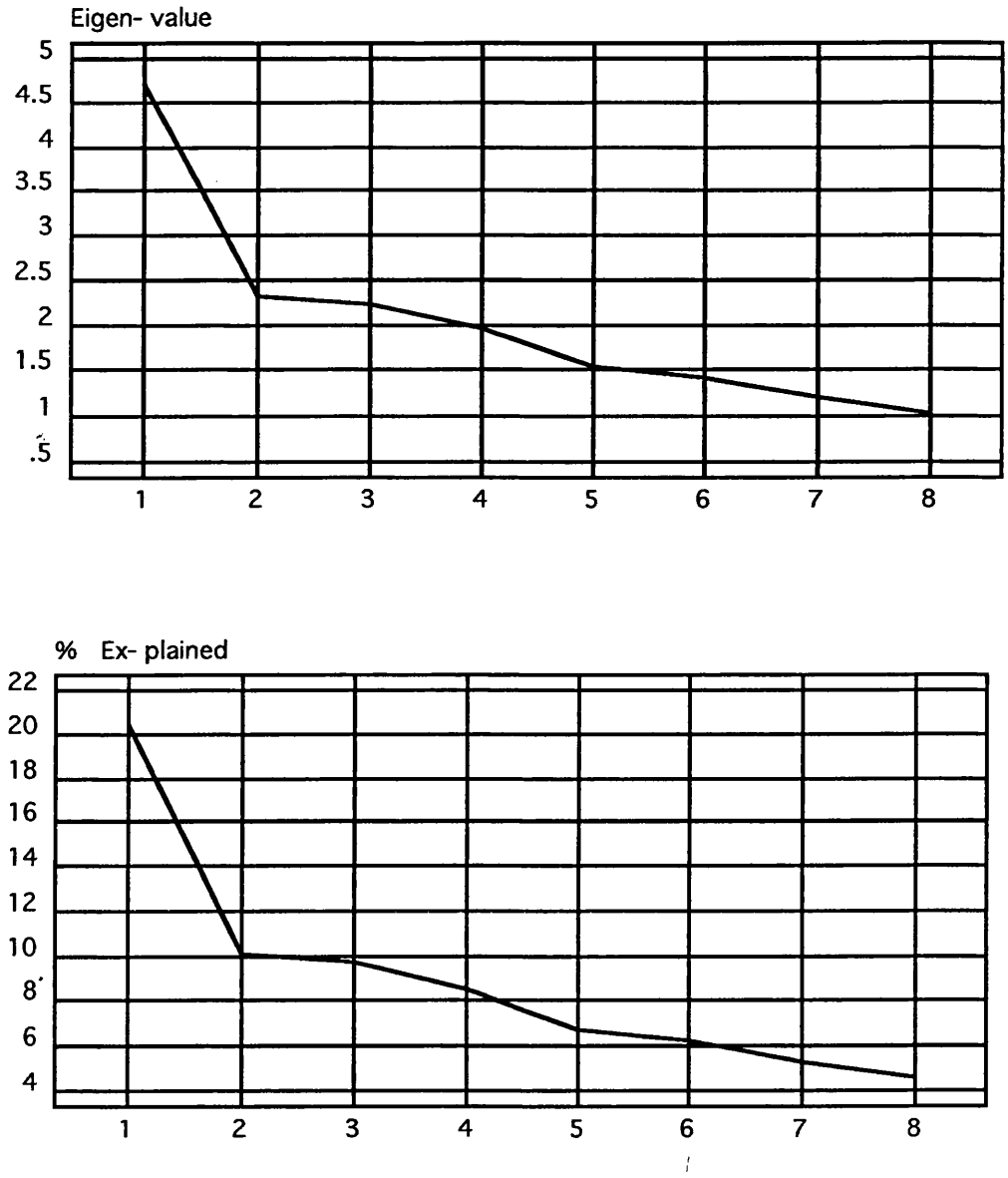


Figure 5.2 : Plot of Eigen Values and Percentage of Variance explained

The cut off point in the plotted scree test could occur after 2 which explained 30.39% of the variance, or after 5 factors, which accounted for 55.53% of the variance. It was decided that since there was little point in categorising individuals into two groups on the two factors, that five factors would be retained. The 5 factor were then rotated using Varimax rotation.

The rotated factor matrix

Rotation of the factor matrix allows the simplifying of factors so that each of the variables will load highly on only one of the factors. In the original analysis, the first factor was calculated to maximise the amount of variance that it could explain. This means, however that it may have been distorted slightly to accommodate some of the variance of variables which are not an integral part of the factor. Rotation of the factor solution helps to correct any distortion. A rotation was carried out as this maximises the tendency of each variable to load highly on any one factor. This therefore makes the variables more extreme in their loadings, thus aiding in the interpretation of the results. The rotated factor matrix appears in Appendix 8.

Unique and Common Variance

A common factor is one which has two or more items with significant loadings and the variance associated with the factor is common variance. A factor with only one significant loading is a unique factor and the unique variance is the part of the total variance of the test resulting from only the unique properties possessed by the test. If the first variable loads highly on factor one, then it is important in determining the factor score for factor one. Usually, the first factor to be extracted correlates fairly highly with all the variables, but it is not advantageous to have the first factor having many significant variables and the others to have none.

2. Results: the rotated factors

The rotation of the matrix allowed the interpretation of the factors. Obviously, the more variables to have high loadings on any one factor, the more clearly the factor may be defined. Factors which have both negative

and positive loadings are called bipolar factors and the variables they account for are contrasting. These contrasts contribute to the eventual labelling of the factor. Which variables grouped together under each factor were of interest so the underlying style feature may be discovered. All factor loadings exceeding a arbitrary positive criterion level of .4 were noted, as well as all the negative loadings in excess of -.4

The following sections give an account of:

(a) The amount of variance explained by each factor.
(b) Which variables are loaded onto which particular factors.
(c) The strength of each of the loadings (in brackets).

Factor 1

Factor 1 was bipolar and accounted for 20.39% of the variance and had significant loadings on complex sentences, prepositions, nouns, pronouns and conjunctions.

complex sentences (-0.546)

prepositions (0.681)

nouns (0.758)

pronouns (-0.698)

conjunctions (0.543)

Factor 2

Factor 2 accounted for 10.09% of the variance and had significant positive loadings on the verb "to be" , passives and conjunctions.

verb "to be" (0.812)

passives (0.790)

conjunctions (-0.430)

Factor 3

Factor 3 accounted for 9.74% of the variance and a had significant positive loading on short sentences, long sentences, subject openers conjunctions and subject openers verbs.

short sentences (-0.679)

long sentences (-0.706)

subject openers conjunctions (-0.724)

subject openers verbs (-0.450)

Factor 4

Factor 4 was bipolar and accounted for 8.55% of the variance. It had a significant positive loading on subject openers total, subject openers prepositions and subject openers adverbs.

subject openers total (-0.873)

subject openers prepositions (0.752)

subject openers adverbs (0.619)

Factor 5

Factor 5 was bipolar and accounted for 6.75% of the variance. It had a significant loading on non function words, simple sentences, compound complex sentences and adjectives.

non function words (0.675)

simple sentences (0.626)

compound complex sentences (-0.806)

adjectives (0.487)

3. Discussion of Results from Factor Analysis

The purpose of a factor analysis is to decrease the number of variables. In this case it was hoped that clear dimensions might emerge which would enable the separation of essays by subject in a further analysis. The results

of the factor analysis showed that the variables can be categorised into five factors. These factors can be explained and named to some extent.

Factor 1, which loaded negatively on complex sentences, and pronouns, and positively on nouns, prepositions and conjunctions could be named the **Sentence Complexity factor**. This factor identifies the word types which are associated with a sentence which is more complicated than a simple sentence. As a sentence becomes more complex, that is, when it involves a greater number of clauses, there is a difference in the proportion of different word types. Pronouns vary in the same direction as complex sentences, whereas nouns, prepositions and conjunctions vary in the opposite direction: they increase as the sentence becomes less complex. This first factor hence reflects the word groups which are associated with the complexity of a sentence.

Factor 2 showed positive loadings on the verb "to be" and passives and a negative loading on conjunctions. It could be named the **Passivity factor**. Logically, the passive form of a verb will increase as the proportion of that verb increases. This is accompanied by a decrease in the proportion of conjunctions indicating that conjunctions have some bearing on this relationship.

Factor 3 consists of two variables relating to length of sentence (short and long sentences) as well as conjunctions and verbs as subject openers. It can therefore be thought of as a **Modulation factor.** It has already been shown in the correlation section that both short and long sentences vary in the same direction. This result is replicated in the factor analysis confirming that there is a definite parameter in the use of long and short sentences, but here it is also apparent that how the sentence is opened has

an effect on this relationship. Opening a sentence with a verb or a conjunction has a direct influence on the modulation of sentence length.

Factor 4 can be called the **Opener factor**. The total number of subject openers counted by the stylechecker varies inversely with prepositional sentence openers and adverbial sentence openers . This is again confirming a result produced by the correlation study where total subject openers were found to vary inversely with conjunctional, prepositional and adverbial sentence openers.

The final factor, factor 5, is the most difficult to interpret. It accounts for only 6.75% of the variance and therefore is the weakest factor. It is again related in some way to sentence complexity since there is an inverse relationship between simple sentences and compound complex sentences. The number of non-function words and adjectives vary in the same direction as simple sentences which implies that compound complex sentences, which have a greater number of dependent clauses, tend to have fewer of these types of words. This factor con be thought of as a **weak complexity factor.**

4. Conclusion

The 23 original variables were reduced to 5 factors which seemed to describe some basic concepts at work in the so-called normal "style space". Following this, it was considered whether it would be advantageous to re-apply these 5 factors in another cluster analysis to determine whether essays grouped by subject using these parameters instead of the 23 variables. According to the rules of probability, subjects were unlikely to group any more reliably on a smaller number of variables. On

consideration, then, it was decided that the factor analysis had served to identify the main concepts within the normal style space and to group the essays on these 5 concepts alone, especially since the latter factors were weak, would not be useful in the quest to identify individual differences.

E. Chapter Conclusions

Three methods of investigation were applied to the variables produced by the stylechecker STYLE. Initially, and perhaps most importantly, cluster analysis was used to test whether the 125 essays could be separated out into 25 subject groups. If the style variables had shown a constant usage within subject, then this type of analysis would have come close to producing 25 equal groups of essays. However, no evidence for subject constancy was found as the essays were grouped not on subjects' individual use of variables, but on variations in the incidence of the variables themselves.

The next method of investigation was an analysis of the relationships between the style variables themselves. Spearman's rho, on the 125 essays produced patterns of correlation which were sensible of writing of this sort, and hence could be useful in the detection of any deviation from the norm or the identification of a normal "style space".

Finally, Factor analysis was applied to the 23 variables. These were reduced to five factors which could be explained and accounted for 55.53% of the variance.

In conclusion then, there did not seem to be any identifiable individual constancies using the style variables from the program STYLE.

The question remained however as to whether there are any enduring features at all within individuals, which bear some relation to these variables. The next chapter addresses this question from the viewpoint of personality.

CHAPTER 6: Style Variables and Personality

A. Introduction

Some of the ideas explored in this thesis and elsewhere are that each individual has their own distinctive writing style, and that this writing style varies not only with circumstances and the purpose of writing but also with the person writing. The previous chapters have tried to identify this variation in the surface characteristics of text produced by different writers. While it is important to know if each individual has their own unique writing style as measured by such surface characteristics, an explanatory account of writing style requires more than just a mapping between a person and their personality characteristics. What is it about a particular person that leads them to write in the way that they do? Personality influences the way in which we express ourselves and therefore must also influence our writing style. The purpose of this chapter is to discuss to what extent personality is reflected in writing.

Before carrying out this exploration, the concept of personality assessment will be discussed .

What is Personality?

At present, there is no single agreed-upon definition of personality, but generally it can be said to represent those characteristics of an individual that account for consistent patterns of behaviour or action. The scientific study of personality attempts to explain similarities and differences between individuals (Pervin, 1984).

There are many theories of personality and each tends to look at a different behaviour, or to look at the same behaviour in a different way. Some of these will be reviewed briefly in order to explain the way in which personality was assessed in the experiment described in the next section.

Theories of Personality

In an effort to distinguish the basic dimensions of personality theorists have identified a conceptual unit which they term a trait. A trait is a broad disposition by an individual to respond to a situation in a certain way. There are two main trait theorists, Hans J. Eysenck and Raymond B. Cattell. The work of both these theorists will be reviewed briefly since both are relevant to this study.

Hans J. Eysenck

Eysenck used the statistical measure of factor analysis to develop a classification of certain traits within an individual. He applied a large number of test items to a large number of individuals to produce a series of variables for the factor analysis. If variables move together, then it can be inferred that they have some kind of common feature behind them, or that they belong to the same unit of personality functioning. Eysenck proposed that the position to be adopted as regards personality was "..... of some type of typological approach, the delineation of certain important dimensions of personality along which individuals can be ranged" (Eysenck, 1982, p3). Using factor analysis, Eysenck defined the basic dimensions of personality which he called **types**. In the course of his research he defined three dimensions of personality which he labelled **Introversion-Extroversion, Neuroticism,** and **Psychoticism**.

Several predictions can be made using this theoretical system. According to Eysenck, an individual high on the extroversion dimension is sociable, has many friends, craves excitement, acts on the spur of the moment and is impulsive. In contrast, an introverted individual tends to be quiet, introspective , reserved, reflective and distrustful of impulsive decisions and will prefer a well ordered life to one filled with risk.

Eysenck gave major attention to the extroversion-introversion dimension and discussed fully the biological-genetic basis for individual differences and the relationship of these individual differences to a wide variety of areas.

Raymond B. Cattell and the 16 P.F..

Cattell also used factor analysis and multivariate analysis as tools for the identification of the basic dimensions of personality. He also was concerned with traits which he defined as "....what a person will do when faced with a defined situation." (Cattell 1979, p 14). He was particularly interested in what he termed source traits which represent associations among behaviours that vary together. He argued that, with a few exceptions, the same source traits can be identified within individuals using 3 types of data: life record data (L-data), questionnaire data (Q-data) and objective test data (OT-data). He developed the 16 Personality Factor Inventory (16 P.F.) to measure personality using questionnaire data.

The 16 P.F. was constructed using the dimensions found using the ratings on the L-data as a source of hypothesis for test items (Cattell

1959b).Thousands of questionnaire items were written and administered to a normal population, and tests were run to see which of these items moved together. Factor Analysis was carried out and the result was the 16 P.F..

The 16 P.F. contains 187 questions including (a) and (b) below:

(a) Often I get angry with people too quickly

a. yes,

b. in between,

c. no

or

(b) I spend much of my spare time talking with friends about social events enjoyed in the past.

a. yes,

b. in between,

c. no.

These are weighted and added up to give a score for each of the 16 factors.

Cattell did not label these factors but they have subsequently been labelled in order to avoid any misinterpretation. The main factors are reserved-outgoing, relaxed-tense, submissive-dominant, concrete thinking-abstract thinking, practical-imaginative, and conservative-liberal. These are thought to cover a wide range of personality functioning and in addition seem to be related to Eysenck's type dimensions (Pervin 1984).

If personality is reflected in writing style we would expect at least some correlations to emerge between the main factors and the style variables. We would expect more outgoing individuals to use longer, more complex sentences. We might also expect them to write with different types of words, perhaps reflecting the way in which they might speak. This might involve the use of more conjunctions to create longer sentences and more descriptive words such as adverbs and adjectives. Individuals high on the venturesome scale might have trends in this direction also. Practical or controlled individuals might use simpler sentences and write in a more structured, constrained way while casual or expedient individuals might have a less grammatical approach to writing.

1. Rationale of the experiment

This experiment was designed to measure the relationship between personality as measured by the 16 P.F., and writing style. It was expected that there would be at least a few significant correlations between the personality factors and the scores on the writing variables since writing is after all a mode of expression and the way in which one expresses oneself is clearly affected by personality.

Cattell's 16 P.F. (Form A, 1986) was used as the measure of personality because it was easy to administer and gave a good scope of dimensions for correlation. The scores produced by the stylechecker "STYLE" were used as the other dependent variable.

B. Intercorrelations between STYLE variables and Factors on the 16 P.F..

1. Subjects

The subjects were a group of 29 undergraduates and members of staff in the Department of Psychology.

2. Design

Each of the subjects had written either two or three essays which were made available for the experiment. The essays written by the undergraduates were written under exam conditions while those written by members of staff were scientific papers. It was recognised that there may be differences in these pieces of writing due to the different circumstances in which they were written, however the assumption was made that if personality effects were going to emerge at all over the whole set of essays then they would by quite enduring and independent of essay type or context.

The essays were put through the stylecheck program STYLE which was used in chapter 5, (see appendix 5) thus providing a set of scores for each essay. Scores were calculated for each subject by averaging these scores for all that subject's essays. The subjects were also required to complete the 16 P.F. questionnaire and the raw scores were used so that no information was lost due to standardisation.

Style Variables

The list of variables used for this study was reduced from the total number produced by the stylechecker. Two of the readability measures (Kincaid and Auto indexes) were dropped because they were designed to give a more reliable reading on technical documents such as training manuals and therefore were not suitable for the exam scripts being used in this study. Average number of questions was also dropped because of incomplete data. It was decided that average sentence length and average word length were too dependant on context and on time factors, both of which were likely to play a large part in the composition of exam scripts. Percentages were used rather than absolute frequencies which make comparisons between essays difficult. This left a total of 24 style variables to be correlated with the 16 personality factors.

3. Results

A non-parametric Spearman's Rank Order Correlation was used to correlate the average variable score with the 16 raw scores on the 16 P.F.. The non-parametric correlation was used since not all the variables were normally distributed and the test was two-tailed since no specific predictions were made a priori. The significant correlations are detailed in Table 6.1. The bold type indicates high correlation with the style variable.

Style Variable	Personality Factor		R
Flesch reading index	Practical-**Imaginative**		0.495
Complex sentences	Cool-**Warm**		0.492
Complex sentences	Sober-**Enthusiastic**		0.414
Verb "to be"	Shy-**Venturesome**		0.584
Auxiliary Verbs	**Expedient**-Conscientious	-ve	0.507
Conjunctions	Casual v- **Controlled**		0.470
Pronouns	Cool-**Warm**		0.507
Nominalisations	**Cool**-Warm	-ve	0.475
Nominalisations	Expedient-**Conscientious**		0.377
Prepositions as S.O.	Conservative/**Expmenting**		0.380
Adverbs as S.O.	**Conservative**/Expmenting	-ve	0.409
Verb as S.O.	**Practical**-Imaginative	-ve	0.397

Table 6.1: Significant correlations between style variables and personality factors at the 0.05 level of significance.

Chance expectation would suggest that for a correlation matrix of 16x24 items, 19 of these correlations could be significant. This is not a problem for the present dataset, since the correlations reported here were also amongst a set obtained in a previous pilot study. This study correlated the style scores and the 16 P.F. scores for 14 undergraduates and produced 20 significant correlations, including the ones reported here. The probability of the same correlations proving significant for both studies is very low, in fact, there is a 1/368 chance that any one of these correlations will appear significant in both studies.

Figure 6.1 on the following two pages shows plots of these significant relationships.

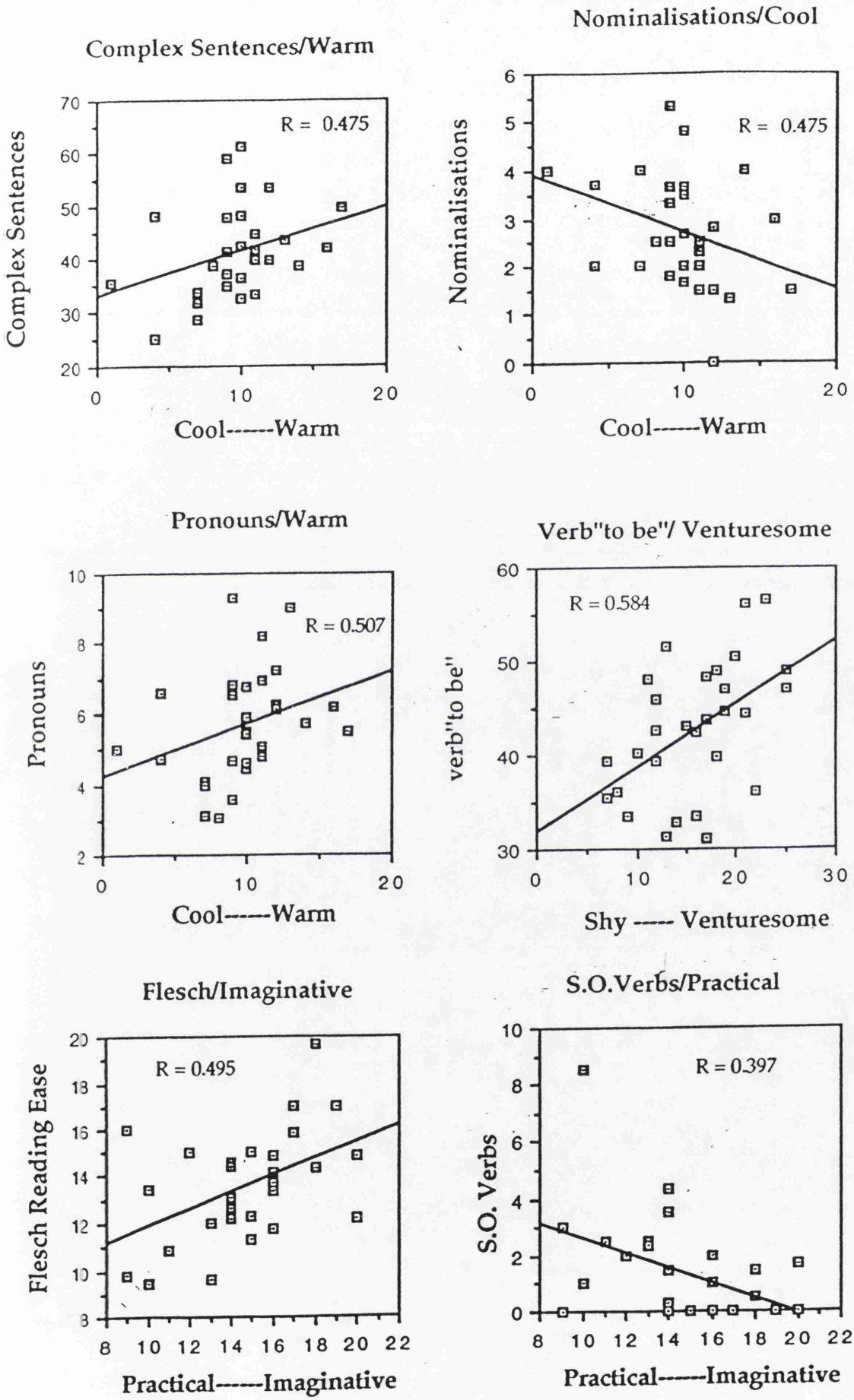


Figure 6.1: Plots of significant correlations between style variables and personality factors.

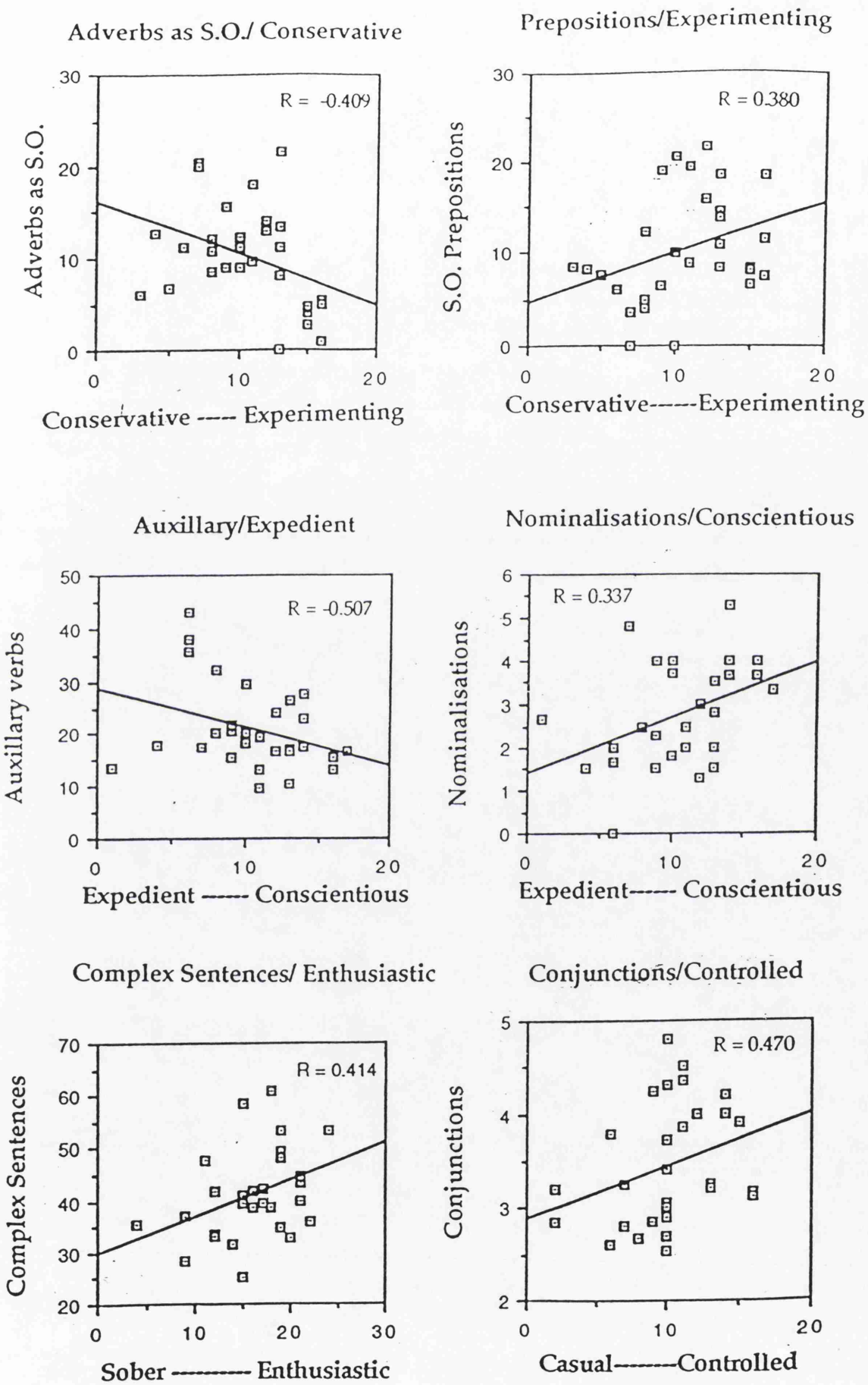


Figure 6.1: Plots of significant correlations between style variables and personality factors cont.

4. Discussion of Results

Twelve significant correlations were found. These covered 7 of the personality factors and 10 of the STYLE variables. Warm/outgoing correlates highly positively with the percentage of pronouns and negatively with the percentage of nominalisations suggesting a relationship between these two style variables. Warm/outgoing also correlates with complex sentences. The practicality factor correlated with both the Flesch reading ease scores and with verbs as subject openers. The percentage of Prepositions as subject openers and percentage of Adverbs as subject openers are correlated with the conservative-experimenting personality dimension. The expedience-conscientious dimension correlates both with percentage of auxiliary verbs and with percentage of nominalisations. Additional correlations are found between enthusiastic and complex sentences, shyness and the verb "to be", and controlled with conjunctions.

It is possible to speculate on the reasons for these results. Many of the correlations seem to be associated with factors in the 16 P.F. which are associated with Eysenck' introversion-extroversion dimension. The factors which relate to this dimension are cool-warm, shy-venturesome and sober-enthusiastic. Five of the 12 correlations involve these factors.

The correlation of **complex sentences** with the **warm** end of the sociability scale is easy to explain. It would seem that more sociable, outgoing or extroverted people tend to express themselves using more involved expressions, perhaps similar to the way that they would speak. Complex sentences are also correlated with **enthusiastic** which is also associated with the extroversion dimension, thus a similar

explanation would be appropriate.**Pronouns** also correlate with the **warm** end of the sociability scale. This could be due to more outgoing people feeling it less necessary to be constrained by the *naming process* and hence substitute pronouns for nouns more often. Pronouns could also be considered to be a less formal mode of expression and warm outgoing people are also considered to be less formal.

Nominalisations are correlated with the **cool** end of the cool-warm scale. The definition of a nominalisation is a verb which is changed into a noun by adding one of the suffixes "ment, "ance", "ence" or "ion". When a writer transforms a nominalised sentence to a non-nominalised sentences, she/he increases the effectiveness of the sentence in several ways. The noun will become an active verb and usually one complicated clause becomes two shorter clauses. The following is a good example:

1. Their inclusion of this provision is admission of the importance of the system. (nominalisations)
2. When they included this provision, they admitted the importance of the system. (no nominalisations)

((Cherry & Vesterman, 1980)

People who use nominalisations could be thought of as rigid in their way of writing and very rule bound. They may tend to restrict their sentences by nominalising their verbs. They are perhaps more introverted then, which would explain the correlation with **cool.** There is also a correlation of nominalisations with **conscientious.** The above explanation goes some way to explaining this also since individuals who are conscientious will probably also follow grammatical rules "to the book" .

The **verb "to be"** is correlated with the **venturesome** end of the shy-venturesome scale. Again this scale is associated with introversion-extroversion. Perhaps individuals who use this verb in large amounts are always themselves "doing" some activity. The verb "to be" can be used with many other verbs, implying an active individual and hence the correlation with venturesome.

Verbs as sentence openers correlated significantly with **practical.** A sentence opened by a verb is very direct and to the point, for example:

"Go and open the window"

Individuals who use this mode of expression are perhaps very practical and to the point. They do not "beat about the bush" when expressing themselves.

There is a significant correlation between **conjunctions** and **controlled.** A conjunction is defined as a word which joins sentences, clauses, phrases or words. It can be a) a co-ordinating conjunction which joins things of equal rank, or b) a subordinating conjunction which joins a dependent clause to a principal clause of a higher rank. An example of each is shown below.

a) Jack **and** Jill went up the hill
b) This is the house **that** Jack built

Controlled can also be defined as " following self image, socially precise and compulsive". The significant correlation of conjunctions

with controlled could perhaps be explained by the fact that individuals with this personality trait are very socially precise and careful. The use of conjunctions gives the impression of exactness and not careless writing. "This is the house that Jack built" could be expressed without a conjunction as

" This is Jack's house" which is not so defined or correct.

Auxiliary verbs correlate significantly with **expedient.** An example of a sentence using an auxiliary verb is:

"I have thought about this all week"

or "

I do like you"

where *have* and *do* are the auxiliaries.

Auxiliaries are therefore used to indicate more than past or present tense. They indicate the tense when it is not indicted by inflection. This correlation seems surprising in some ways since initially one might think that the use of an auxiliary verb indicates a more conscientious or rule-bound mode of expression. However, a possible explanation lies in the fact that, individuals who score highly on the expedient end of the scale, are also self-indulgent. The use of the auxiliary verb changes the emphasis to nearer the beginning of the sentence, so that in the above examples, the focus is more on the "I" . This may be indicative of a more self-centred, indulgent individual.

Finally, the use of different sentence openers seems to indicate some features of personality. Opening a sentence with an **adverb** is correlated with **conservative** and opening with a **preposition** is correlated with **experimenting**. As these are polar opposites, this

would imply that adverbs and prepositions are used by opposite personality types (though this is not directly shown by the correlation). An example of a sentence beginning with a preposition is:

"Under the table the mouse crouched quivering".

An example of sentence beginning with an adverb is:

"Quickly, she closed the curtains".

Opening a sentence with an preposition is probably therefore more unusual than opening it with and adverb. The prepositional opening gives the impression of a more poetic, flowery style and could be considered more experimenting than using an adverb in the same position.

As the above discussion shows, it is at least possible to speculate about the correlations between the style variables and the personality factors. However, it should be remembered that the source of the correlations is not known, that is, it is not known whether personality is a function of writing or whether the way we write is affected by personality. This is characteristic of the typical problem with personality where a circularity of explanation occurs in which a trait is used to explain a behaviour which is in turn the basis for the concept of the trait in the first place.

It seems more likely that the way we write is affected by our personality and hence it is in this direction that the correlations are explained. The important point however is that it has been shown that quantitative measures of personality (regardless of one's view of their validity) correlate with quantitative measures of writing style, using computerised statistical procedures.

C. Chapter Conclusions

This chapter has provided an initial investigation of personality in relation to writing style on the prime assumption that if personality influences the way in which we express ourselves, it must also influence our writing style. The aim of this chapter was therefore to discuss to what extent personality is reflected in writing. The present study was designed to determine whether there were any significant correlations between the personality traits identified in Cattell's 16 P.F. test and the style variables identified by STYLE.

It was expected that if personality is reflected in writing style at least some correlations would emerge between the main factors and the style variables. This was indeed the case and 12 significant correlations were found involving 10 style variables and 7 of the personality factors. These correlations were explainable to some extent in that many of the significant style variables correlated with factors which were associated with Eysenck's introversion-extroversion dimension. This allowed an explanation to be based on the characteristic of extroverts and their typical interactions with others whether by speech or by writing.

This study has made a number of issues evident. Firstly, there do seem to be reliable relationships existing between personality characteristics and writing style. The fact that there are variables within an individual's writing which are related to these constructs is one which has never been tested constructively in this way. Intuitively, one feels that personality *must* be evident in writing, and indeed it is often said

that one can imagine what a person is like by reading their writing. But it has never been clear what exactly gives this impression.

The question is still open as to whether extreme numbers of the style variables which are associated with a personality characteristic would give a strong impression of that personality. It may be possible to manipulate numbers of style variables within a piece of writing and test this. Would a large number of complex sentences or pronouns in a piece of text give the impression that the author is outgoing? This could have far reaching consequences, for instance, in job applications, if one was able to give the impression of being a particular type of person by using more or less of a certain style variable. It may be, however, that the question is more complicated, and some readers are good judges of personality while others are not. The implication then, is that it is the personality of the reader which is affecting his perception of the personality of the writer.

In conclusion then, there are reliable associations between the writing style variables and the personality constructs produced by the 16 P.F.. Since these personality characteristics are not theoretical constructs, but real predisposition's to behave in a certain way, the existence of relationships between them and measures of writing style should not perhaps, be surprising.

PART 4: SUBJECTIVE JUDGEMENTS OF WRITING STYLE

CHAPTER 7: Subjective judgement of style features

A. Introduction

In this chapter the stylistic features of 60 of the essays were examined in more detail. In the previous section, automated counts had been made of surface variables by the stylechecker. These had shown some differences across subject but not enough to separate individuals. It was thought that there may be more subjective features in the essays which might show individual differences to a greater extent than the statistical variables examined in the previous section.

1. Materials

The materials were 60 essays written by undergraduate students at the University of Glasgow, under examination conditions. There were 12 authors each with 5 essays.

2. Procedure

Initially each essay was read through carefully. Any idiosyncratic features which seemed to occur frequently in that essay were noted down. Having noted these in one essay, these features were sought in the following essay as well as any additional features which the author of that essay might use. This procedure was performed for all sixty essay so that a pool of features was formed, and ultimately each essay was checked for them.

The features which were noted tended to relate to surface variables of the text rather than ones which related to the actual semantics or syntax of the essay. These features included several punctuation marks, the use of signs instead of words, or numbers instead of words, the use of questions, abbreviations and the use of the first person. The full list of features examined can be seen below.

1. Brackets
2. Inverted commas around words
3. Colons
4. Semi-colons
5. Grammatical Hyphens
6. Slash
7. Abbreviations such as U.S.A. or E.E.G.
8. Use of Questions
9. Use of 1st person
10. Double word abbreviations (D.W.A.) such as "don't" or "can't"
11. Plus sign for "and"
12. Numbers instead of words
13. Use of emphasising script such as capitals or underlining
14. Hyphenated words such as "well-liked"

Three undergraduate judges were then enlisted to check the features in the sixty essays. They were asked a) to note whether each of the features was present or absent in each of the essays and b) if present, to count the incidence of each of the features in each essay. The number of words in each essay was then counted since the length of the essay would inevitably affect the incidence of the features. The **proportion** of the features in the checklist could then be calculated for each essay by

dividing the incidence of the habit by the number of words in the essay. Therefore each essay had a score for the proportion of each of the 14 features in the checklist.

3. Rationale

This investigation was therefore to determine whether a) the subjective style variables grouped in any way into sensible configurations which might define a certain "style space" and b) whether subjects could be separated uniquely using these variables.

Three different clustering methods were used in order to answer these questions. These were multidimensional scaling, factor analysis and cluster analysis.

The initial analysis was a non-metric multidimensional scaling technique which used the lowest level data (whether a feature was present or absent from the essay in question), in order to determine which of the features seemed to group together and on what basic dimensions.

B. Study 1: Multidimensional Scaling (MDS)

This method uses a procedure which moves the variables around in space until a configuration which best approximates the relationship between all the variables is reached and the goodness of fit is maximised.

Non-metric multidimensional scaling was deemed an ideal procedure to provide spatial evidence for the relationships between these subjective style variables.

Overview of Multidimensional Scaling. (MDS)

In general, the goal of MDS is to detect any meaningful underlying dimensions between objects or variables. It is similar to Factor Analysis in this respect, but differs in that it imposes fewer restrictions in terms of linear relationships and normally distributed data.

Goodness of fit : Stress

Usually MDS measures goodness of fit by means of the *stress* measure. The raw stress value is *Phi* and this basically measures the sum of the squared deviations of observed distances from the expected or reproduced distances (reference). Thus, the smaller the stress value, the better the fit of the reproduced distance matrix to the observed distance matrix. The goodness of fit can be evaluated by means of a Shepard diagram which plots the observed distances against the reproduced distances. If the reproduced distances are identical to the observed distances, all the points on this plot will fall on a straight line.

Number of Dimensions

Like Factor Analysis, MDS requires that the number of dimensions on which to produce the distance matrix, is specified. In general, if more dimensions are used, the fit of the matrix will be better but harder to interpret (Jackson, 1983). In fact, if there were as many dimensions as there were variables, then the observed distance

matrix could be reproduced perfectly. However, like factor analysis, the aim is to reduce the number of variables and explain the distance matrix in terms of a small number of dimensions.

The scree test first proposed by Cattell (1966) can again be used to decide on the number of dimensions. If the stress value is plotted against the different number of dimensions, the point at which the plot levels off is usually taken as the best number of dimensions. The interpretability of the dimensions is also a factor to be taken into account.

Interpreting the Dimensions

The final step of the analysis is to interpret the produced distance matrix in terms of the dimensions. The actual orientation of the axes from the MDS analysis are arbitrary and can be rotated in any direction. A two dimensional scatterplot is usually crucial to interpretation.

Applications of MDS

MDS is a useful technique in that it can be applied to a number of situations and data. It has been very popular in psychological research on person perception (see, for example, Rosenberg, 1977), in marketing research (Churchill, 1987), in order to detect the nature of dimensions underlying the perceptions of different brands or products (see Greene & Carmone , 1970, for examples) and also recently in the field of psychological profiling (e.g. Canter, 1992).

1. The present data

The initial data consisted of a presence/absence score for every variable on each of the 40 essays. Every variable was then correlated with every other over the 40 essays using a non-parametric (gamma) correlation. Gamma correlation was used as this takes into account the large number of tied ranks which exist within the data.

The result was a 14 * 14 similarity matrix which was put into the Multidimensional scaling module of CSS STATISTICA and the analysis run using one two and three dimensions. The similarity matrix can be seen in Appendix 9.

Number of dimensions to retain

The stress values were considered against the number of dimensions in the form of a scree test (Cattell, 1966) as an aid to determining the number of dimensions to retain. This is shown in Figure 7.1.

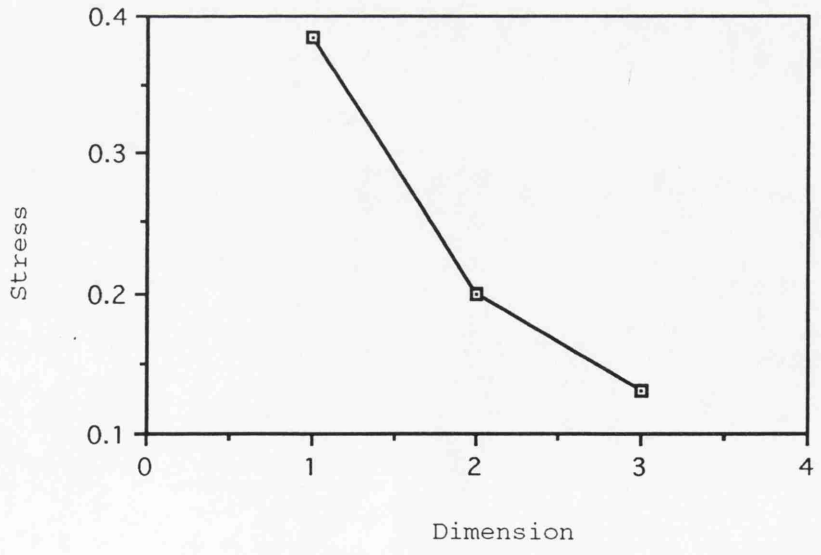


Figure 7.1: Plot of Scree values for each of Three Dimensions

Two dimensions were retained since the scree graph appears to level off slightly at this point, and also on consideration, two dimensions proved to be more interpretable than three.

2. Results

The two dimensional solution is shown in a scatterplot in Figure 7.2.

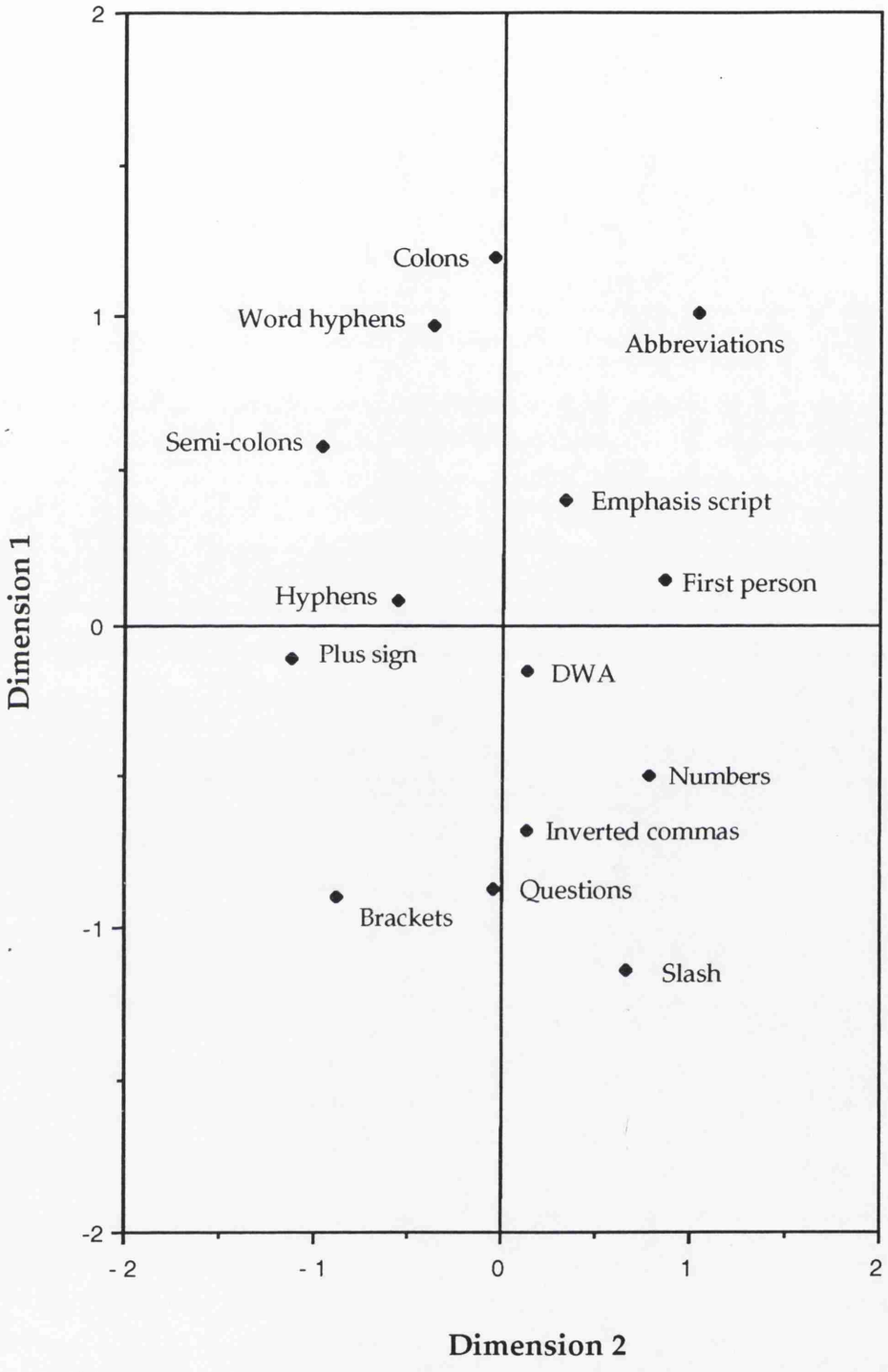


Figure 7.2 : Scatterplot of 2 Dimensional solution

3. Discussion of MDS

The two dimensional solution can be interpreted as follows.

Dimension one seems to reflect an *informality* dimension in which reflects many of the variables which might be used in speech, such as use of the first person and questions. They also tend to use inverted commas in their writing and possibly double word abbreviations such as "can't" or "don't" . At the other end of this dimension individuals will use more grammatical variables such as colons and semi-colons, and hyphenated words. This interpretation of this dimension would imply that there are individuals who tend to write in an informal manner, perhaps more as they would speak, while others write more formally, possibly as they have been taught, an are aware of when to use, for instance, a colon or a semi-colon.

The second dimension would seem to reflect a *scientific /literary* dimension. At one end of this dimension people use concise number oriented modes of expression, such as plus signs or numbers in place of words, abbreviations such as "USA" or "UV light" and emphasis of script using capitals or underlining. At the other end, they use a more literary style which involves the use of colons and semi-colons before explanations and hyphenated words.

Considering these dimensions, Figure 7.3 would perhaps illustrate the solution more clearly.

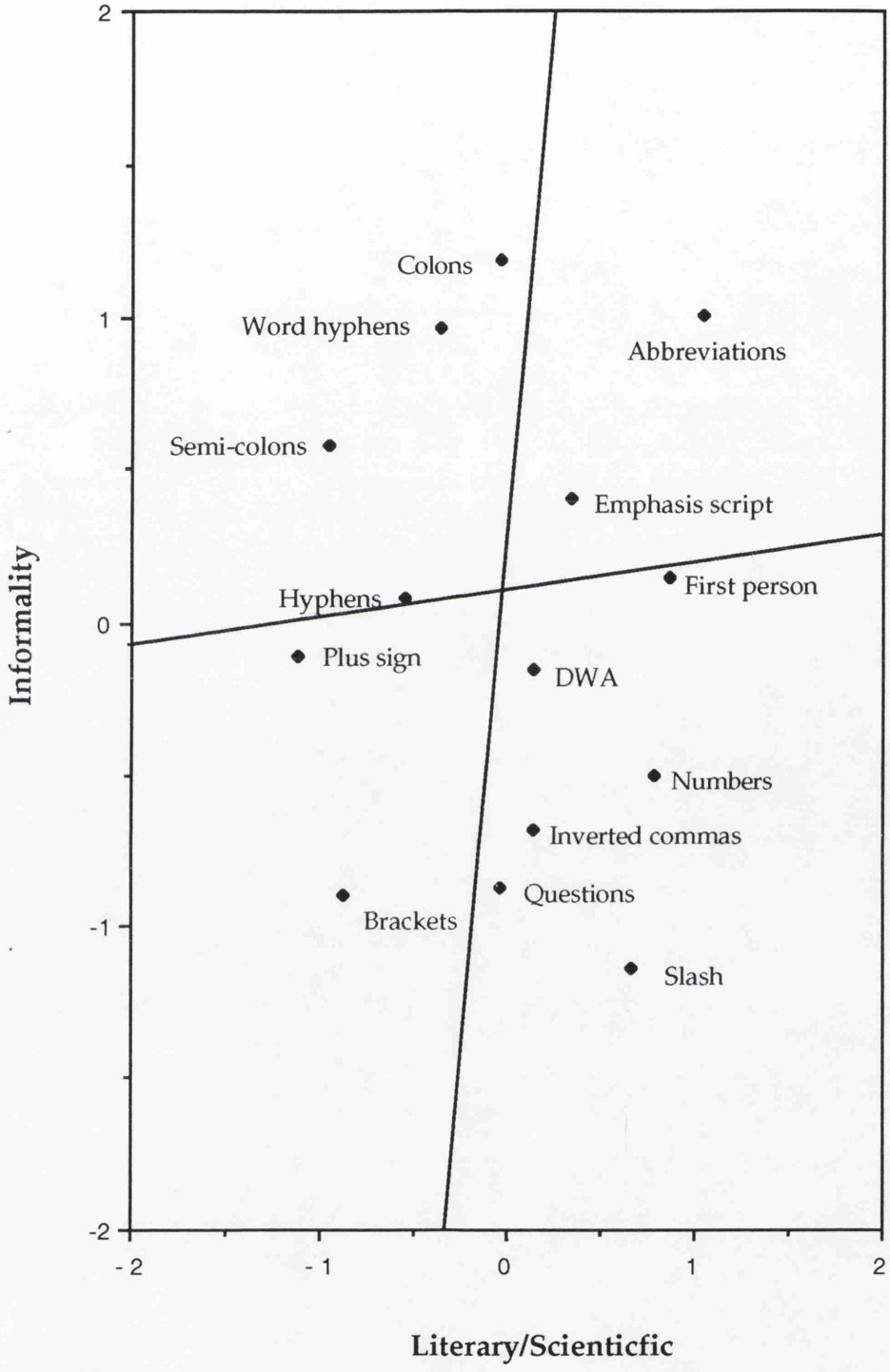


Figure 7.3: Arbitrary axes on 2 Dimensional Solution

4. Conclusion to MDS

It seems that the 14 variables can be grouped into two interpretable dimensions which are reflected in an individual's use of writing.

A second way of grouping the variables is through use of factor analytical techniques. The next study then is a factor analysis using the numerical data on the same variables, on the assumption that the same sort of factors or dimensions would emerge as from the MDS.

C. Study 2 : Factor Analysis

The binomial data in the previous study indicated that there were various themes at work within these variables. Factor analysis was chosen as a second method to investigate which of the variables seem to group together, however this time, the numerical data was used.

The rationale was similar to that in Chapter 5, where the intention was to reduce the 23 stylecheck variables to a lesser number of concepts which could be used to describe the data in a more concise way.

1. The present data

The proportion per essay of each of the 14 variables as agreed by the three judges was noted for each of the 60 essays. The resulting data matrix of raw scores was standardised by converting them to z values (mean of 0.00, standard deviation of 1.00) and then factor analysed to produce a factor matrix. This factor matrix is shown in Appendix 10.

The original factor matrix shows the first stage at which the variables used will have a common meaning. Each variable received a loading on each factor which is its correlation with the factor. The unrotated solution produced a series of 6 factors. These explain 66.6% of the variance. A table of eigen values and percent of variance explained can be seen in Table 7.1 below.

Factor	Eigen Value	% Variance explained	Cumulative %
1	2.15	15.36	15.36
2	1.70	12.14	27.50
3	1.62	11.62	39.12
4	1.44	10.25	49.38
5	1.36	9.74	59.12
6	1.05	7.48	66.60

Table 7.1: Table of Eigen Values for 6r original factors and % Variance explained

It was decided that 5 factors should be retained on the following basis. The Scree test would indicated that only 2 factors should be retained as the graph begins to level out after factor 2. The other criterion for deciding how many factors should be retained is the Kaiser criterion which proposes that all factors with an eigen value of greater than 1 should be retained. In this case this would mean retaining all 6 factors. The factor solution which makes the most sense with regard to interpretation is the retention of 5 factors and it was felt that this was justified since the graph levels off until this point and then drops again. The graph of Variance explained can be seen below in Figure 7.4.

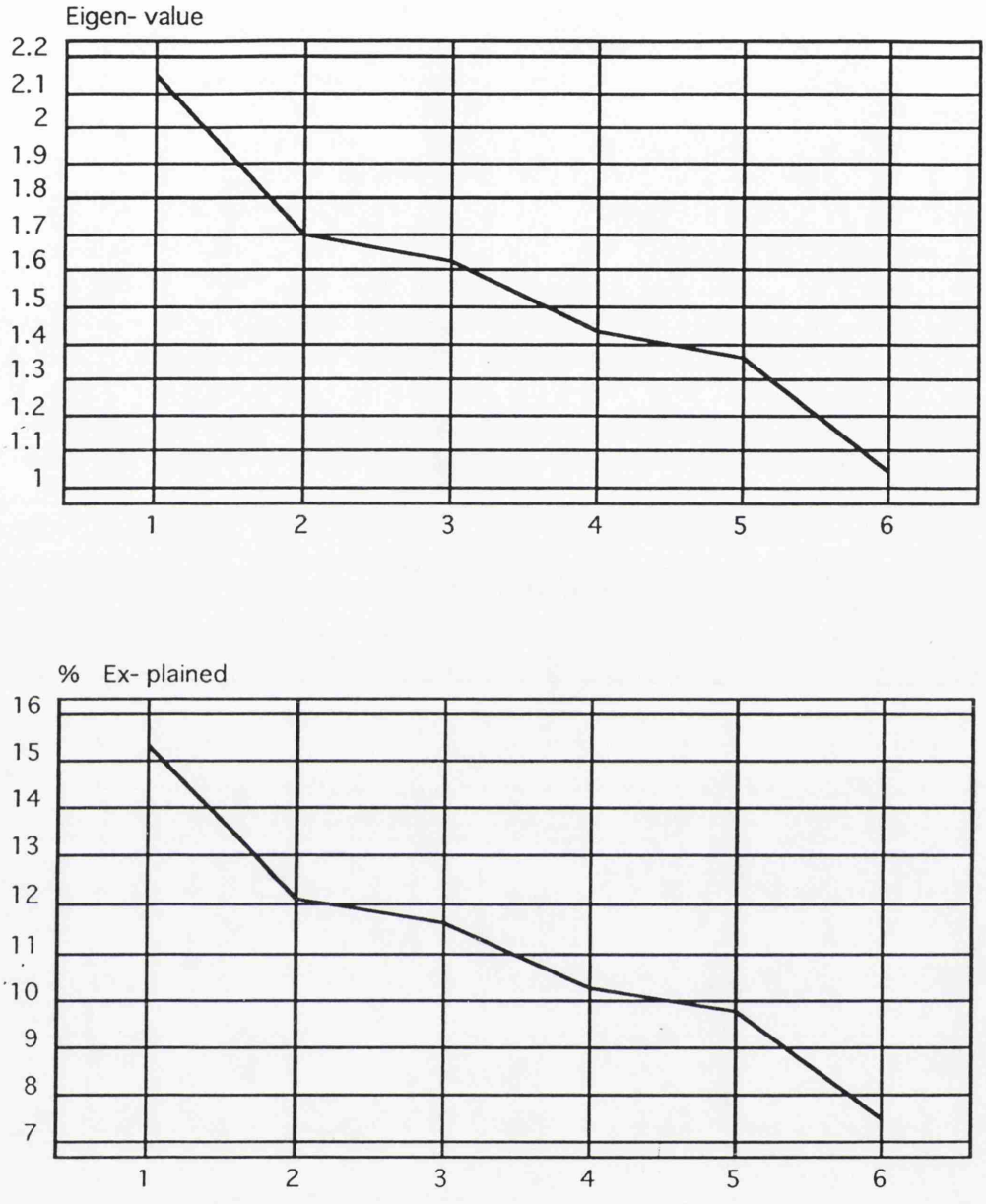


Figure 7.4: Plot of Eigen values and Variance explained

The 5 factors were then rotated using Varimax Rotation. The rotated factor matrix is in Appendix 11. The results are shown below.

2. Results

All factor loadings exceeding an arbitrary positive criterion level of 4% were noted, as well as all the negative loadings in excess of -4%. Under each factor is an account of:

a) The amount of variance explained by each factor.
b) Which variables are loaded onto which particular factors.
c) The strength of each of the loadings (in brackets).

Factor 1

Factor 1 accounted for 14.52% of the variance and had significant loadings on Double word abbreviations, First person, Numbers and Slash.

D.W.A. (-0.610)

First person (-0.424)

Numbers (-0.843)

Slash (-0.810)

Factor 2

Factor 2 was bipolar and accounted for 12.09% of the variance and had significant loadings on Emphasis script, Plus sign and Grammatical Hyphens.

Emphasis script (-0.636)

Plus sign ((0.507)

Grammatical Hyphen (-0.748)

Factor 3

Factor 3 accounted for 11.67% of the variance and had significant loadings on Colons and Semi-colons and DWA.

Colons (0.754)

Semi-colons (0.813)

DWA (0.449)

Factor 4

Factor 4 was bipolar and accounted for 10.7% of the variance and had significant loadings on Abbreviations, Inverted Commas, and Questions.

Abbreviation (0.796)

Inverted Commas (0.796)

Questions (-0.40)

Factor 5

Factor 5 was bipolar and explained 10.09% of the variance and had significant loadings on brackets, hyphenated words & first person.

Brackets (-0.639)

Hyphenated words (-0.786)

First Person (0.413)

3. Discussion of Results

The results of the factor analysis showed that the variables can be categorised into five factors. These can be interpreted as follows.

Factor 1 could be thought of as the **informality factor**. It loads significantly on double word abbreviations (e.g. can't, don't) and on the use of the 1st person. Both these variables are frequently used when the writer feels he is actually talking to the reader. The other variables involved in this factor are the use of the slash and numbers instead of words. A slash is often used when the writer wants to give an alternative suggestion for a particular word and again this could thought of as something which would be more likely to occur in speech, though in speech this is done easily, without the need for a punctuation mark. Numbers are usually used for two reasons. Firstly, in scientific contexts and secondly, when it does not matter if the text is

not literary, in other words when the writing is in note form. The second reason could also be connected to informality.

Factor 2 can be said to represent an **emphasis** factor. The different scripts used for emphasis such as underlining and capitals plus hyphens used in the middle of sentences can be thought of as mechanisms which emphasis the point that the writer is making. This is more obvious in the case of script type, but hyphens usually indicate that there is more to be said on a subject and that the writer does not want to add this information using a conventional punctuation mark. He/she will usually restate the point he wishes to after the hyphen, for instance, from essay 48:

" He found that if the sender gave direct instructions which were easily and quickly carried out then the receivers felt dissatisfied and disgruntled as a result - that they did not feel as though they were participating in a team effort."

The other variable which loads highly on this factor, though it has the lowest of the three loadings, is the plus sign. It loads in the opposite direction from the other two variables in this factor, indicating that when the incidence of these is high, the incidence of the plus sign is low, and vice versa. Use of a plus sign could perhaps be thought of a less emphatic way of adding information than using the word "plus" or " and" which would explain the bipolar relationship in this factor.

Factor 3 is clearly the **colon factor** since there are high loadings on both colons and semi-colons. Individuals who use a lot of colons in their writing also tend to use many semi-colons and vice-versa. The

two types of colons can therefore be easily reduced to one concept. This factor also loads weakly on DWA, but this is probably incidental in this factor since it already loads highly in Factor 1.

Factor 4 has high loadings on Abbreviations, Inverted Commas and questions. It can be thought of as a **speech mimic** factor. The variables which load highly on this factor are devices used to make writing more like speech. Inverted commas are a means of emphasising particular words in speech. They are a device used in writing which allow a word or phrase to be emphasised or indeed to indicate the beginning of direct speech. The question mark indicates a change in intonation which in a way copies the intonation used in speech. Abbreviations are usually used when the writer is aware that he will have to repeat a lengthy word or phrase throughout the text and wishes to do this more succinctly. Alternatively, there may be standard abbreviations which are used for specific concepts which the writer will adopt. For whatever reason, abbreviations make the text less cumbersome allowing the writer to talk about or emphasise a point more easily as he would be able to do in speech.

Factor 5 is harder to explain. It is bipolar and accounts for 10.09% of the variance, therefore it is the weakest factor. It loads on Brackets, Hyphenated words and use of the First person. It can be argued that if an individual hyphenates words where they are not usually hyphenated, this could reflect a stylistic habit. However, if a word is actually spelt with a hyphen, it can be considered part of the grammar of the English language. If so, then hyphenated words cannot really be considered a feature of style. Hyphenated words and brackets load in the opposite direction, implying that a high incidence of one would be

found with a low incidence of the other. This might indicate the two ends of a dimension where the individual either uses very correct style of writing, signified by correctly hyphenated words, and at the other end he uses a style which is more akin to note taking, or less correct English, and this is signified by devices such as the use of brackets. Use of the first person only loads marginally in the direction of the hyphenated words and thus is associated with the dimension in this direction. However the marginal loading and the fact that it also loads on Factor one, would imply that it is only incidental to this final factor. We will call this factor the **weak informality factor.**

4. Conclusions on Factor analysis

The five factors produced can be seen to reflect to some extent the results of the MDS in the previous study. Again there is a factor which has been named the informality factor, which groups together variables such as the use of the 1st person and double word abbreviations. This does re-affirm the assumption that people use these informal variables together or else they would use more grammatical or literary variables such as colons or semi-colons. Other factors also emerged which reduced the variables to a smaller number, such as the emphasis factor which involved the use of different script types and hyphens. A factor which seemed to represent some of the devices used in speech was also identified as well as the two types of colon being reduced to one factor.

D. Study 3- Cluster Analysis

The judges had also noted the number of times each of the check-list features appeared in each of the 60 essays enabling the proportion of each feature in each essay to be calculated. These proportions were again standardised as before and a cluster analysis could be carried out in the same way as for the stylecheck data in chapter 5. A k-means (where k=12 for the 12 subjects) clustering algorithm was applied to the 60 essays and the number of essays by the same subject in each cluster was noted. The results are shown in Table 7.2 below.

1. Results

CLUSTER	SUBJECTS IN CLUSTER
1	S12
2	S1, S1, S4, S7, S8, S11, S12
3	S9
4	S4, S7, S7, S12
5	S1, S2, S3, S3, S3, S3, S3, S4, S5, S5, S5, S5, S5, S6, S7, S8, S8, S10, S10, S10, S11, S11, S11, S12
6	S6, S6, S6, S6, S9, S9, S9, S10, S10
7	S9
8	S1, S1, S8, S8, S11
9	S2, S2, S4, S12
10	S7
11	S4
12	S2, S2

Table 7.2 : Details of subjects appearing in each cluster

The number of times an essay by the same subject occurs in the same cluster can be seen in Figure 7.5 .

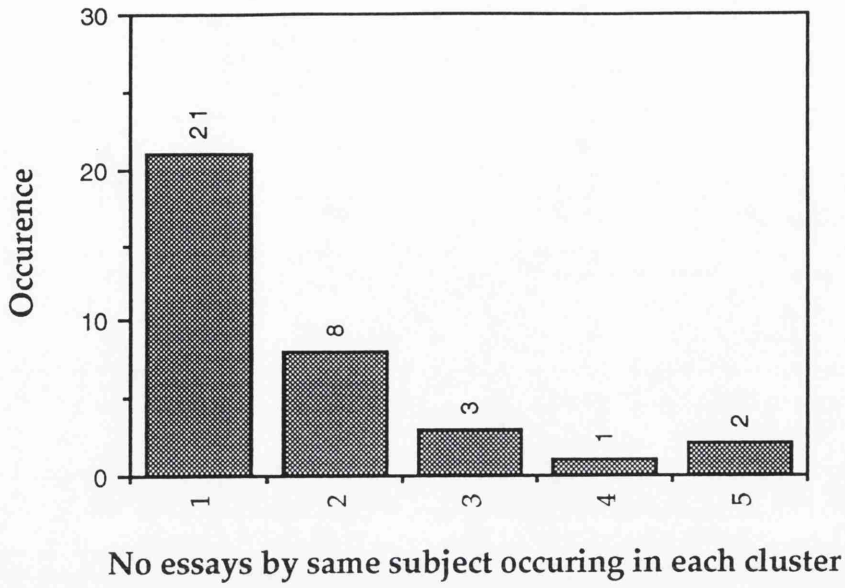


Figure 7.5: No essays by the same subject occurring in the same cluster

It can be seen that on two occasions all five essays by the same subject occur in the same cluster. Four essays by the same subject occur in the same cluster on one occasions , three on three occasions, two on eight occasions and one on the majority of occasions, numbering 21.

When the means for each variable are examined, it is evident what features each cluster is loading on. The highest loading means for each cluster are shown in Table 7.3 below. In most clusters there was a clear highest loading on one of the variables. In cases where there was not (for instance, in cluster 5) the highest means are shown.

CLUSTER	VARIABLES	MEAN VALUES
1	Questions	7.62
2	First Person	1.52
	Semi-colons	1.08
3	D.W.A.	4.47
	Slash	3.16
4	Brackets	3.01
5	Abbreviations	-0.353
	Grammatical Hyphens	-0.447
	First person	-0.337
	Hyphenated words	-0.336
	Plus sign	-0.332
	D.W.A.	-0.324
6	Plus sign	2.03
7	Numbers	7.62
8	Hyphenated words	2.81
9	Grammatical hyphens	2.15
	Abbreviations	1.47
10	Colons	6.16
11	Inverted commas	6.53
	Abbreviations	4.44
12	Emphasis script	4.63

Table 7.3: Variables loading on each Cluster

Only clusters 2, 3, 5, 9 and 11 did not have a clearly identifiable highest mean on one of the variables and cluster 5 showed very diffuse loadings on all of the variables.

2. Discussion of results

The number of essays by the same individual occurring in the same cluster was greater than the number occurring when the 23 style variables were used, but still the results were extremely poor. All five essays by the same subject did occur twice in the same cluster, but this cluster had 24 essays in it altogether. As pointed out previously, this will inevitably increase the likelihood of this occurrence. In the same cluster (cluster 5) there was also two occurrences of 5 essays from the same subject. On inspection of the means for each cluster it could be

seen that there were no loadings on any of the factors which were extreme. All the loadings on this cluster were around the same figure of -0.3 and all were negative. This would imply that this is a very general factor into which many essays fall. All these 24 essays therefore used comparable amount of these variables.

Four essays from the same subject also occurred in cluster 6. This cluster loaded highly on the use of plus signs.

Since subject matter is likely to play a part in the clustering of essays also, the topic of each essay in each cluster was noted. No cluster contained essays on a particular subject and only very broadly could it be said that the first cluster contained essays which were psychological and cluster 9 contained essays which were scientific or biological. This fits in with the dimensions produced by the MDS and the factors produced by the Factor analysis in that Abbreviations seemed to be associated with the scientific dimension and thus is the variable which loads highly on cluster 9.

On initial inspection then, it would appear therefore that the cluster analysis of the subjective variables was more successful than of the stylecheck variables, since these are more clusters which contain 4 or 5 essays from the one individual. However, the analysis still only very weakly separates essays by subject, especially since a large number of the essays fall into the one cluster, cluster 5.

It is clear then that there are many essays in the sample which have very high or very low incidences of these subjective variables, so that although the analysis is including essays by the same subject in the

same cluster, it is including essays by a number of other subjects as well. This means that many of the variables used will not be useful in separating out individuals, that is, for identifying individual differences between subjects, since they are constant across the essays of more than one individual.

Again, some clusters are being formed on the basis of one essay alone (e.g. cluster 7 -on use of numbers) implying that these essays do not fit in any of the other clusters.

E. Chapter Conclusions

This chapter has focused on subjective features of writing style, some of which can only be picked up by visual inspection.

The initial analysis examined only if these features were present or absent in the texts being considered, and determined whether there were any underlying dimensions within the features using Multidimensional Scaling techniques. Two were identified which seemed to exemplify an informality dimension and a scientific / literary dimension.

A Factor analysis on the proportion of these features in an essay, produced results which confirmed those from the MDS. Five interpretable factors were produced one of which could be identified as the informality factor in the previous study. These results indicated that the features used could successfully be reduced to a smaller number of concepts.

The final analysis examined whether essays could be grouped or clustered by subject. There was evidence of several essays by the same subject occurring in the same cluster and hence having high loadings on the same features. However, these clusters tended to be very large, thus increasing the likelihood of this occurrence, and only one large cluster contained the 5 essays from a single subject. Clustering the essays by subjects was therefore no more successful than the surface style features in chapter 5.

When examining subjective variables such as these, there is a large element of the analysis which is affected by the genre of the materials in question. Perhaps even more than when using the stylecheck variables, the subject matter of the essay will determine for instance whether formal or scientific variables are used. Both these dimensions were picked up using the MDS and the Factor analytical techniques, which would imply that this is a fundamental characteristic of the data. The essays were written on a variety of topics covering art, science and social science. These topics must therefore determine to some extent the dimensions identified. However, the resulting clusters were not identified strongly on topic, so the influence the style variables is greater than that of genre.

In conclusion then, there was little or no evidence to suggest that subjects could successfully be separated using these variables.

CHAPTER 8 : Discussion

In general, this thesis has examined the evidence for individual constancies in writing style. It has covered a variety of measures, including individual differences in the use of surface variables in writing, and the relationships existing between these surface variables, and, for example, personality. However, the majority of the thesis has perhaps been dominated by an analysis of the claims made by Morton and Michaelson (1990) about the Cumulative Sum Technique. This focus on Cusum was mainly due to the extent of the claims made by Morton, and the consequences of these claims for a study of writing style, should they turn out to be reliable. However, a detailed examination of these claims showed that the majority were void, and as a method for detecting multiple authorship, it has been shown to be vacuous.

One way or another, studies of writing style have failed to produce a reliable measure for the identification of one particular author as opposed to another. This implies that a quest which was initially thought to be simple has many more problems associated with it than was originally suspected. The search for a measure to uniquely determine authorship began in the late 19th century with the study of sentence length. It is continuing now in the late 20th century with methods such as the cusum technique. Arguably, we are still no further forward in formulating a foolproof method of authorship identification for forensic or literary purposes. Indeed, it is possible to document the successes and failures of this objective, since the outset of these studies. Figure 8.1 overleaf shows some of these studies.

Date	Exponent	Method proposed to detect authorship	Outcome for authorship studies
1859	De Morgan	Average word length would vary from to person	Average word length did not vary sufficiently between individuals and was found to be too dependent on context
1887	Mendenthall	Frequency distributions of word lengths	Only distinguished groups of authors who used a certain length of word most often rather than separating any author uniquely.
1939 1957	Yule Wake	Sentence length distributions	Some degree of success in that this was an objective measure though still not enough variability from sample to sample.
1962	Ellegard	Comparison of characteristic words and expressions or "marker words"	Again, some degree of success, however, method was was regarded as very subjective and not suitable for all types of texts.
1964	Mosteller & Wallace	Frequent words which were used by most authors but at different rates	Successful in that words studied were independent of context and conclusions could be drawn on certain texts which were regarded to be "beyond reasonable doubt". However, still not regarded as a foolproof method of authorship identification.
1970's	e.g. Morton	Collocational methods	Position of words is still affected by context and raised many problems about type of texts to use and comparisons to be made.
1970-date	e.g. Stratil & Oakley	More sophisticated quantitative analyses such as clustering techniques	Produced varied results sometimes showing similarities and differences between authors. However, tended to find these differences for single words and therefore methods could not be regarded as coming close to a test for authorship.
1990's	Morton & Michaelson	Cumulative Sum Technique	Only method claimed to have 100% succes rate in authorship identification but now shown to be unreliable and be as likely to identify two authors as one and one as two.

All of the studies shown in Figure 8.1 have been reviewed in Chapter 1 of this thesis. If one was to pass judgement on the most successful group of studies over the years in terms of authorship determination, the honour would possibly go to those which have concentrated on distinctive characteristics or marker words of authors. Although these studies were regarded as the least scientific and most subjective, they perhaps did not receive adequate acclaim. It is interesting to note, however, that procedures using marker words are now being applied again by forensic linguists.[6]

Among the most recent methods is the cusum technique. The criticisms which can be levelled at this procedure are many and varied, since the fundamental ratio assumption has been shown to be void. The ratio between sentence length and habit for any individual is not constant over different texts from that individual, nor is there any reason to assume that a separation in a cusum chart implies multiple authorship. The separation of the cusum curve has been shown to be dependent on the order of sentence lengths and the correlation between sentence length and the chosen habit. In short, the very method which looked most promising for authorship identification is debatably further from the mark than any of the other studies.

Despite the failure of these methods to absolutely determine authorship, there is still a fundamental belief that there *are* differences in writing style, and that these differences are

---

[6] For example, at the First Conference of Forensic Linguistics in Birmingham, November, 1991, several papers on the use of marker words in criminal cases were presented by delegates.

discernible merely by reading a piece of text. The following five examples, which were used in this thesis, would generally be considered as displaying a range of writing styles.

Example 1:

> The problem with a lot of medical "research" that is done these days is that it is sponsored by drugs companies, and tends to get "de-funded" if the results are "unreliable" !! "My God", they say, "a doctor-yuppie with morals!!". Yes, I know!! I try to be a money-grubbing, superficial, elitist, narrow-minded git, but being married to a scientist (and a social scientist at that) makes it nigh-on impossible to maintain the facade !!

Example 2

> In addition to these letters I have also replied promptly to every letter that you have sent me regarding information about my address. My reason for contacting you three times and for replying promptly to your letters was that I did not want to be issued with a community charge demand for a large amount of money half way through the year. Despite all my attempts this has failed as I have now received an initial community charge bill for £161.25, plus seven monthly payments of £32.25!

Example 3

> Before I discuss the framework of investigation I would like to state why PREPOSITIONS? and why SPACE? Firstly Prepositions _ well as I have said they provide a complete language domain for investigation that is apparently simple and narrow yet is surprisingly complex. In other words simple analyses of meaning result in many "exceptions". You could say that there are two approaches to semantic analysis of prepositions - the MINIMAL SPECIFICATION approach and the FULL SPECIFICATION APPROACH.

Example 4

> Treatment of AN is of variable efficacy in that, like alcoholism and drug addiction, it is only when the patient truly wishes to recover that the treatment will be successful. If the sufferer's weight has fallen below 75% of her ideal weight-for-height, then she requires treatment. Such treatment consists of a behaviour modification programme involving rewards (e.g. T.V. privileges), punishment (i.e. prevention from leaving the hospital), and informational feedback about weight status.

Example 5

> The use of a definite NP is marked since most subject referents in on-line discourse are already known, understood, and easily retrievable from the context (see footnote 2) and are, as a result, only minimally coded in the narrative discourse (i.e. through a ø-anaphor, or through verb agreement). The use of SV word

> order is marked since it occurs with highest frequency in situations where thematic continuity is disrupted. Again, the maintenance of thematic continuity can be assumed to be the norm since more propositions maintain rather than disrupt this continuity.

These five examples illustrate a number of points. There are some very obvious difference between the examples which could be classed in terms of formality. It would be possible to order the texts in terms of formality, but it is likely that this order would vary somewhat across individuals. This exemplifies the subjectivity of the perception of writing style. Some of the features identified previously as associated with informality, are visible in these examples. In particular, Example 3 shows the use of different script types for emphasis. Example 1 shows ample use of hyphenated words and quotation marks. Although Examples 1 and 2 are both from letters, they are different in terms of formality. Example 2 is very formal and more on a par with Example 4 or 5, which are not from letters. Examples 3, 4 and 5 are from examination texts but Example 3 seems more informal than 4 and 5, despite being written for a similar purpose and in a similar context.

The scientific/literary dimension identified by the Multidimensional scaling in chapter 7, is also apparent, especially in Examples 4 and 5. There is an increased use of abbreviations, numbers and brackets in both.

All five examples perhaps give a different impression of their writers though in fact, Example 1 and Example 4 are written by the same author.

The analysis of writing style in this thesis focused on trying to establish what features within the texts made them appear to be different, and whether or not these features remained constant within an individual. The assumption was that if features were characteristically constant, then it might be possible to use them to distinguish between authors. The obvious starting point of the investigation was to examine some universal style measures such as word type or sentence beginning. However, no separation of individuals was found using these measures. The next step was to turn to some subjective measures of writing style, or features which "stood out" for a particular individual, especially in light of the comparative success of the early "marker word" studies. The hope was that these would be more constant within an individual and hence be more likely to provide a separation between authors. Again, however, no discrimination was found using these features.

The study of the style variables and personality was undertaken for two reasons. Firstly, for the more basic reason that it is generally believed that the personality of an individual is evident through his or her writing. Any correlations found between style variables and personality would confirm this assumption. The second reason for a study of personality, was that personality is the main field in which there are well-established individual differences. Personality studies mirrored the objectives of the thesis, in that they establish how individuals are the same, and how they are different. The present research aimed to do this too, by looking at the specific area of writing. The outcome was that several correlations were found between personality and style which may or may not, have some biological basis according to present research.

In addition to correlations between the style features and the personality factors, the style features were found to be related to each other in various ways. Some of these relationships were sensible in relation to sentence structure (e.g. compound sentences increased with the use of conjunctions), but others seemed to indicate more complex associations between the variables. Although the relationships existing between the style variables are of no benefit to the forensic approach to authorship, they do indicate that these sorts of features can be reduced to a smaller number which move together. This fact raised the possibility that there is a particular "style space" into which normal individuals might fall. This style space might be relevant to the study of writing style in that individuals which use different incidences of the variables defining the style space, may not have only an unusual writing style, but may also have what is considered a "bad " writing style. The definition of a style space using these and other variables, could then be useful for the formulation of guidelines for writing style, perhaps in a style or grammar checker.

The possibility of a normal style space was examined a little further by means of factor analysis on the universal style variables and multidimensional scaling on the subjective variables. However, it is not yet resolved what the notion of a style space means for the study of writing style, except perhaps that a deviation from this style space could be considered an unusual style.

There are other possibilities for the study of writing style. There is a long-standing notion that people might write in the same way that they speak. This is supported by the fact that it is often helpful to

read a piece of writing aloud in order to determine whether or not it is acceptable or grammatical.

Punctuation is also an area which shows an element of individual difference. This again has been said to relate to the way in which an individual speaks (e.g. Chafe, 1988) and is just one of the ways in which to indicate different rhythmic and prosodic structures within writing. Prosodic features such as these have been under examined in relation to the study of writing style, and could definitely be considered good candidates for individual difference since they do appear likely to add up to some sort of style signature.

Indeed, pilot work on individual differences in punctuation was carried out near the onset of the present research. Subjects were asked to add commas to a short passage from which all commas had been removed. Other punctuation, such as full-stops and colons, were left in the passage. A large degree of variability was found in the punctuation of the passage, however, some structure was detected in the position of commas. When the comma positions were "mapped out" diagramatically, it was found that there were connections between the comma positions: if a comma was put in one position, then it was very likely to be put in another position also. Although the study was not replicated over several passages from one person, it is possible that these relationships may be constant across texts for an individual. It is also possible, however, that these relationships may only divide individuals into a small number of groups, but that they may contribute to the "style space" for that text. Of course, there is the additional possibility that individuals are only punctuating in the way in which they were

taught, which limits the usefulness of punctuation in a study of writing style.

Conclusion

The present research finds little evidence to suggest that individuals can be identified uniquely by any of the measures of writing style which have been examined. However, the results do point to the possibility of a certain degree of categorisation of the studies on writing style. This categorisation would be according to whether or not a unique writing style is advocated and whether or not a comparison of known and unknown authorship is required or just an examination of individuals statistically using a large number of variables.

Four different positions can be taken then, and can be evaluated in terms of the ability of their associated methods to distinguish individuals through their writing. There is a section of studies which would conform to the **weak forensic** position, which states that it is possible to compare a text whose authorship is suspected but not known, with text from 2 or 3 likely individuals, by means of visual inspection. Judgements can be made using a variety of means, but techniques using marker words or expressions are obvious candidates in this position. This position is open-ended and atheoretical, since texts can be compared on anything which seems to be a possible discriminator, however idiosyncratic.. Studies such as Miron (1981) have used this approach. Miron (1981) advocated the use of a variety of indices in order to determine the authorship of any type of communication document. Their method " rests upon the assumption

that the language of a given communication potentially represents a unique configuration of attributes of the source of that communication; that the language of the communication is the signature of its source" (Miron, 1981, p406).

The analysis uses what they termed signature words (characteristic words),whose "occurrences are so distinctive that they narrow considerably their possible origins" (Miron, 1981, p437). Choice of vocabulary was one of many other methods examined.

There is therefore some evidence to support this position as the studies using marker words are relatively successful when compared with other more statistically based methods.

The second position which can be identified is the **strong forensic** position which would support that individuals are unique in their writing style and can be identified as such. The prime contender in this position is the Cusum technique which claimed to be able to identify with 100% accuracy whether or nor one or more individuals had written a certain piece of text. The present research has found no evidence for the validity of this position.

The third position which can be taken is one which advocates **strong individual style.** If this were true, we would have expected to be able to distinguish individuals uniquely using incidences of particular variables in their writing. The application of statistical methods such as cluster analysis would have allowed the separation of individuals on a large number of surface variables. Therefore again this research

has found no evidence that such a strong claim can be made for strong individual style.

The weaker claim in relation to such a position would be a **weak individual style** position, which was found to have some validity in the study of writing style. This entails the notion that there are certain constancies in people's writing, but that these constancies are not very distinctive, in other words, individuals cannot be separated completely using measures of writing style, but may perhaps be clustered in groups.

Four positions can therefore be identified. The assumptions of both the strong forensic position and the strong individual style position have been found to be too extreme, and no evidence to support either of them has been found by any of the methods applied in this thesis. There is some substance however, to both the weak positions. Both showed some element of success in this thesis and elsewhere, in the comparison of marker words or expressions and in the ability to separated individuals into groups, if not uniquely. It is to these two positions, the weak forensic position, and the weak individual style position, that any further research should be directed.

Appendices

Appendix 1: Example of Cusum Calculation for sentence length for a sample of 25 sentences

No of Sentence	No. words	Diff. from average	Cusum
1	49	2.12	**2.12**
2	15	-31.88	**-29.76**
3	84	37.12	**7.36**
4	127	80.12	**87.48**
5	7	-39.88	**47.60**
6	17	-29.88	**17.72**
7	34	-12.88	**4.84**
8	8	-38.88	**-34.04**
9	148	101.12	**67.08**
10	57	10.12	**77.20**
11	51	4.12	**81.32**
12	56	9.12	**90.44**
13	84	37.12	**127.56**
14	26	-20.88	**106.68**
15	30	-16.88	**89.80**
16	36	-10.88	**78.92**
17	179	132.12	**211.04**
18	17	-29.88	**181.16**
19	48	1.12	**182.28**
20	6	-40.88	**141.40**
21	11	-35.88	**105.52**
22	40	-6.88	**98.64**
23	19	-27.88	**70.76**
24	10	-36.88	**33.88**
25	13	-33.88	**0.00**

Average No. Words = 46.88

Appendix 2: Output of the COUNT PROGRAM

Sentence Condition

Sent		1	2	3	4	5	6	7	8	9	10	Total
1	V	0	2	2	1	2	0	0	0	0	0	7
1	C	0	2	3	1	4	1	0	0	0	0	12
1	W	0	4	5	2	6	1	0	0	0	0	19
2	V	1	2	0	0	0	0	1	0	0	0	4
2	C	0	3	1	1	3	3	1	0	0	1	13
2	W	1	5	1	1	3	3	2	0	0	1	17
3	V	1	1	2	0	0	0	0	0	0	0	4
3	C	0	1	2	2	4	4	0	2	0	0	15
3	W	1	2	4	2	4	4	0	2	0	0	19
4	V	0	1	0	1	0	0	0	0	0	0	2
4	C	0	0	1	0	1	0	1	0	0	0	3
4	W	0	1	1	1	1	0	1	0	0	0	5
5	V	0	0	1	0	0	0	0	0	0	0	1
5	C	0	0	0	2	1	0	0	2	0	0	5
5	W	0	0	1	2	1	0	0	2	0	0	6
6	V	1	1	1	0	0	0	0	0	0	0	3
6	C	0	2	0	1	1	4	3	0	0	1	12
6	W	1	3	1	1	1	4	3	0	0	1	15
7	V	1	0	1	0	0	0	0	0	0	0	2
7	C	0	0	0	0	1	1	3	0	0	0	5
7	W	1	0	1	0	1	1	3	0	0	0	7
8	V	0	4	1	1	0	1	0	0	0	0	7
8	C	0	2	2	4	3	2	3	0	1	0	17
8	W	0	6	3	5	3	3	3	0	1	0	24
9	V	0	2	0	1	0	0	1	0	0	0	4
9	C	0	2	1	4	1	1	4	0	0	0	13
9	W	0	4	1	5	1	1	5	0	0	0	17
10	V	1	0	0	0	0	0	0	0	0	0	1
10	C	0	1	0	1	0	1	1	0	0	0	4
10	W	1	1	0	1	0	1	1	0	0	0	5
11	V	0	1	0	0	0	0	0	0	0	0	1
11	C	0	1	1	1	1	1	1	0	0	1	7
11	W	0	2	1	1	1	1	1	0	0	1	8
12	V	1	1	0	2	0	0	0	0	0	0	4
12	C	0	2	3	2	1	0	1	1	0	0	10
12	W	1	3	3	4	1	0	1	1	0	0	14
13	V	1	0	1	0	0	0	0	0	0	0	2
13	C	0	2	0	2	2	1	0	3	0	0	10
13	W	1	2	1	2	2	1	0	3	0	0	12
14	V	0	0	0	1	0	0	0	0	0	0	1
14	C	0	2	1	0	1	0	0	1	0	0	5
14	W	0	2	1	1	1	0	0	1	0	0	6
All	**V**	**7**	**15**	**9**	**7**	**2**	**1**	**2**	**0**	**0**	**0**	**43**
All	**C**	**0**	**20**	**15**	**21**	**24**	**19**	**18**	**9**	**1**	**3**	**131**
All	**W**	**7**	**35**	**24**	**28**	**26**	**20**	**20**	**9**	**1**	**3**	**174**

KEY

V = Words beginning with a vowel

C = Words beginning with a consonant

W = Number of words with particular length occuring in sentence

1-10 - number of letters in word

Block Condition

Block		1	2	3	4	5	6	7	8	9	10	Total
1	V	0	1	0	1	2	0	0	0	0	0	4
1	C	0	2	3	1	3	1	0	0	0	0	11
1	W	0	3	3	2	5	1	0	0	0	0	15
2	V	1	3	2	0	0	0	0	0	0	0	6
2	C	0	2	1	0	3	2	0	0	0	1	9
2	W	1	5	3	0	3	2	0	0	0	1	15
3	V	1	0	1	0	0	0	1	0	0	0	3
3	C	0	2	1	2	1	4	1	1	0	0	12
3	W	1	2	2	2	1	4	2	1	0	0	15
4	V	0	2	1	1	0	0	0	0	0	0	4
4	C	0	0	2	1	5	1	1	1	0	0	11
4	W	0	2	3	2	5	1	1	1	0	0	15
5	V	1	1	1	0	0	0	0	0	0	0	3
5	C	0	2	0	2	2	2	1	2	0	1	12
5	W	1	3	1	2	2	2	1	2	0	1	15
6	V	1	0	2	0	0	0	0	0	0	0	3
6	C	0	0	0	1	2	4	5	0	0	0	12
6	W	1	0	2	1	2	4	5	0	0	0	15
7	V	0	2	1	1	0	0	0	0	0	0	4
7	C	0	2	1	4	1	1	2	0	0	0	11
7	W	0	4	2	5	1	1	2	0	0	0	15
8	V	0	3	0	1	0	1	0	0	0	0	5
8	C	0	1	1	3	2	0	2	0	1	0	10
8	W	0	4	1	4	2	1	2	0	1	0	15
9	V	1	1	0	0	0	0	1	0	0	0	3
9	C	0	2	1	3	0	2	4	0	0	0	12
9	W	.1	3	1	3	0	2	5	0	0	0	15
10	V	0	1	0	2	0	0	0	0	0	0	3
10	C	0	2	4	1	1	1	2	0	0	1	12
10	W	0	3	4	3	1	1	2	0	0	1	15
11	V	2	1	1	0	0	0	0	0	0	0	4
11	C	0	2	0	2	3	1	0	3	0	0	11
11	W	2	3	1	2	3	1	0	3	0	0	15
12	V	0	0	0	1	0	0	0	0	0	0	1
12	C	0	3	1	1	1	0	0	2	0	0	8
12	W	0	3	1	2	1	0	0	2	0	0	9
All	**V**	**7**	**15**	**9**	**7**	**2**	**1**	**2**	**0**	**0**	**0**	**43**
All	**C**	**0**	**20**	**15**	**21**	**24**	**19**	**18**	**9**	**1**	**3**	**131**
All	**W**	**7**	**35**	**24**	**28**	**26**	**20**	**20**	**9**	**1**	**3**	**174**

Appendix 3: Artificial data used in Manipulation of ratio and change in sentence length

Data for Figure 3.4(a) Cusum of Single ratio data (0.4)

sl	23lw	CUSUMsl	CUSUM23lw
2	0.8	-24	-24
4	1.6	-46	-46
6	2.4	-66	-66
8	3.2	-84	-84
10	4	-100	-100
12	4.8	-114	-114
14	5.6	-126	-126
16	6.4	-136	-136
18	7.2	-144	-144
20	8	-150	-150
22	8.8	-154	-154
24	9.6	-156	-156
26	10.4	-156	-156
28	11.2	-154	-154
30	12	-150	-150
32	12.8	-144	-144
34	13.6	-136	-136
36	14.4	-126	-126
38	15.2	-114	-114
40	16	-100	-100
42	16.8	-84	-84
44	17.6	-66	-66
46	18.4	-46	-46
48	19.2	-24	-24
50	20	0	0

Data for figure 3.4(b) - similar slopes (0.7/0.8) Rank order separately and cusum of the whole

sl	23lw	CUSUMsl	CUSUM23lw
2	1.4	-24	-19.93
4	2.8	-46	-38.31
6	4.2	-66	-55.16
8	5.6	-84	-70.46
10	7	-100	-84.22
12	8.4	-114	-96.44
14	9.8	-126	-107.12
16	11.2	-136	-116.26
18	12.6	-144	-123.85
20	14	-150	-129.91
22	15.4	-154	-134.42
24	16.8	-156	-137.39
26	18.2	-156	-138.72
28	19.6	-154	-138.72
30	21	-150	-137.06
32	22.4	-144	-133.87
34	23.8	-136	-129.14
36	25.2	-126	-122.86
38	26.6	-114	-115.05
40	28	-100	-105.69
42	29.4	-84	-94.79
44	30.8	-66	-82.35
46	32.2	-46	-68.37
48	33.6	-24	-52.84
50	35	0	-35.78
2	1.6	-24	-55.49
4	3.2	-46	-73.43
6	4.8	-66	-89.61
8	6.4	-84	-104.04
10	8	-100	-116.70
12	9.6	-114	-127.6
14	11.2	-126	-136.73
16	12.8	-136	-144.11
18	14.4	-144	-149.72
20	16	-150	-153.58
22	17.6	-154	-155.67
24	19.2	-156	-156
26	20.8	-156	-154.57
28	22.4	-154	-151.38
30	24	-150	-146.41
32	25.6	-144	-139.71
34	27.2	-136	-131.23
36	28.2	-126	-120.99
38	30.4	-114	-108.99
40	32	-100	-95.23
42	33.6	-84	-79.71
44	35.2	-66	-62.42
46	36.8	-46	-43.38
48	38.4	-24	-22.57
50	40	0	0

Data for figure 3.4(c) - different slopes (0.4/0.9) Rank order separately and cusum of the whole

sl	23lw	CUSUMsl	CUSUM23lw
2	0.8	-24	-10.75
6	2.4	-44	-20.43
10	4	-60	-29.05
14	5.6	-72	-36.60
18	7.2	-80	-43.07
22	8.8	-84	-48.48
26	10.4	-84	-52.82
30	12	-80	-56.10
34	13.6	-72	-58.30
38	15.2	-60	-59.43
42	16.8	-44	-59.50
46	18.4	-24	-58.50
50	20	0	-56.43
48	19.2	22	-54.59
44	17.6	40	-54.43
40	16	54	-55.03
36	14.4	64	-56.70
32	12.8	70	-59.43
28	11.2	72	-63.24
24	9.6	70	-68.12
20	8	64	-74.06
16	6.4	54	-81.07
12	4.8	40	-89.15
8	3.2	22	-98.30
4	1.6	0	-108.52
2	1.8	-24	-118.60
6	5.4	-44	-126.28
10	9	-60	-131.56
14	12.6	-72	-134.43
18	16.2	-80	-134.90
22	19.8	-84	-132.96
26	23.4	-84	-128.62
30	27	-80	-121.88
34	30.6	-72	-112.73
38	34.2	-60	-101.17
42	37.8	-44	-87.22
46	41.4	-24	-70.85
50	45	0	-52.09
48	43.2	22	-35.53
44	39.6	40	-19.37
40	36	54	-6.61
36	32.4	64	3.74
32	28.8	70	11.69
28	25.2	72	17.23
24	21.6	70	20.37
20	18	64	21.10
16	14.4	54	19.43
12	10.8	40	15.36
8	7.2	22	8.88
4	3.6	0	0

Appendix 4: Output of program CUSUM

B	C	D	E	F	G	H	I	J
17	6	-9.10	-4.40	-9.10	-8.45	-4.40	0.65	0.65
31	14	4.90	3.60	-4.20	-1.54	-0.80	2.66	2.66
60	27	33.90	16.60	29.70	30.35	15.80	0.65	0.65
26	10	-0.10	-0.40	29.60	29.58	15.40	-0.02	0.02
23	10	-3.10	-0.40	26.50	28.81	15.00	2.31	2.31
17	6	-9.10	-4.40	17.40	20.36	10.60	2.96	2.96
24	12	-2.10	1.60	15.30	23.43	12.20	8.13	8.13
23	6	-3.10	-4.40	12.20	14.98	7.80	2.78	2.78
21	7	-5.10	-3.40	7.10	8.45	4.40	1.35	1.35
19	6	-7.10	-4.40	-0.00	-0.00	-0.00	0.00	0.00

Bave 26.1 Cave 10.4
Fmax 29.7 Fmin -9.1 Frange 38.8
Hmax 15.8 Hmin -4.4 Hrange 20.2
F/H 1.92079 Jsum 21.5198 Rho 0.89626

KEY:

Bave = Average of column B
Cave = Average of column C
Fmax = Maximum value in column F
F min = Minimum value in column F
Frange = Range of column F
Hmax = Maximulm value in column H
Hmin = Minimum value in column H
Hrange = Range of column H
F/H = Frange/Hrange
Jsum = Sum of Deviations between Cusum curves

B = sentence length
C = 23lw
D = diff average sl
E = diff average 23lw
F = Cusum sl
G = scaled cusum 23lw
H = unscaled cusum 23lw
I = Deviation between F - G
J = Absolute deviation F - G

Appendix 5: Output of Stylechecker STYLE

readability grades:
(Kincaid) 7.8 (auto) 10.1 (Coleman-Liau) 7.7 (Flesch) 6.9 (80.9)
sentence info:
no. sent 9 no. wds 212
av sent leng 23.6 av word leng 4.20
no. questions 0 no. imperatives 0
no. nonfunc wds 123 58.0% av leng 5.31
short sent (<19) 33% (3) long sent (>34) 11% (1)
longest sent 36 wds at sent 2; shortest sent 13 wds at sent 8
sentence types:
simple 33% (3) complex 44% (4)
compound 11% (1) compound-complex 11% (1)
word usage:
verb types as % of total verbs
tobe 30% (6) aux 10% (2) inf 15% (3)
passives as % of non-inf verbs 12% (2)
types as % of total
prep 9.9% (21) conj 5.2% (11) adv 5.7% (12)
noun 28.8% (61) adj 15.6% (33) pron 8.0% (17)
nominalizations 0 % (0)
sentence beginnings:
subject opener: noun (1) pron (4) pos (1) adj (0) art (1) tot 78%
prep 0% (0) adv 11% (1)
verb 0% (0) sub_conj 11% (1) conj 0% (0)
expletives 0% (0)

Appendix 6 - Variables used from STYLE™

(all as percentages)

1. Non Function Words
2. Short Sentences
3. Long Sentences
4. Simple Sentences
5. Complex Sentences
6. Compound Complex Sentences
7. Compound Sentences
8. Verb to be
9. Auxiliary verbs
10. Infinitives
11. Passives
12. Prepositions
13. Conjunctions
14. Adverbs
15. Nouns
16. Adjectives
17. Pronouns
18. Total Subject Openers
19. Prepositional Sentence Openers
20. Adverbial Sentence Openers
21. Verb Sentence Openers
22. Subordinating Conjunction Sentence Openers
23. Conjunctional Sentence Openers

Appendix 7

Original Factor matrix from Factor Analysis of 23 stylecheck variables

variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
N.F.W.	.794298	-.082915	.017092	.093755	.247716	-.418830	-.030010	.081915
Short Sent	-.393080	.456746	-.323841	-.286174	.259995	.139130	.124587	.298957
Long Sent	-.356939	.457870	-.421401	-.255047	.241496	-.146590	-.050697	.207398
Simple Sent	.697473	-.219293	-.082421	-.092974	.301092	.326079	.006851	-.062948
Complexsent	-.586424	-.033699	.405793	.375053	.008339	-.083603	.029404	.104741
Comp.com	-.464849	.276968	-.253331	-.328342	-.459949	-.129490	-.143866	-.002526
Comp.sent	.249466	.111370	-.184051	.017441	-.207445	-.435338	.531050	.078262
Verb to be	-.097875	.765114	.372335	-.005801	-.005955	.176051	.070773	-.241103
Auxillliary	-.113149	.202773	-.037593	-.142624	-.244366	-.375040	-.053083	-.096398
Infinitives	-.222228	-.320993	-.268301	.181389	.008019	-.146221	-.201054	.694124
Passives	.138571	.768051	.185958	.180845	-.017879	.081329	-.153422	.037283
Prepositions	.249890	.304333	-.501738	.429604	-.139186	.262532	.094350	.049916
Conjunctions	.155883	-.080761	-.632753	-.273957	-.386758	-.170034	.055921	-.276528
Adverbs	-.498709	-.132061	.198849	-.201709	.260987	-.338263	.224974	-.120427
Nouns	.785881	-.096660	-.346152	.175003	-.108052	-.010327	-.158611	-.146768
Adjectives	.566033	.365981	.294293	.151715	.276068	-.312017	.088579	.144822
Pronouns	-.649791	-.409928	.153817	-.322780	.076751	.234930	.210848	-.077610
S.O. Total	.640327	.032160	.272735	-.581483	-.133375	.114551	.077107	.227100
S.O. Prep.	-.313087	-.010703	-.281386	.646619	-.099165	.311615	.186653	.003341
S.O. Adverb	-.387620	-.031202	-.201056	.425737	.320146	-.400484	.146426	-.283226
S.O. Verb	-.030531	.034285	-.200396	-.116507	.438095	-.101789	-.599003	-.231422
S.O. Sub-con	-.393668	-.044078	.179309	.157152	-.300142	-.171776	-.505310	.008380
S.O. Conj.	-.261102	.061956	-.503641	-.184171	.467570	.130756	.056097	-.081494
Variance	.203936	.100866	.097416	.085519	.067579	.062087	.052784	.045428

Appendix 8

Rotated Factor Matrix from factor Analysis of 23 Stylecheck Variables

variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
N.F.W.	.400175	.071571	.076911	-.285736	.674838
Sort Sent	-.094060	.178761	-.679214	.049851	-.335968
Long Sent	.002021	.143304	-.705886	.080451	-.335724
Simple Sent	.307746	-.129297	-.055538	-.367622	.625297
Compl. sent	-.544758	.156584	.249784	.499555	-.131165
Comp. Comp.	.021406	-.040544	-.155712	-.030595	-.805637
Compound	.373350	.020652	.046440	-.078719	-.060651
Verb to be	-.105280	.812482	-.049832	-.046424	-.240485
Auxilliary	.056185	.091014	-.005897	-.077045	-.343480
Infinitives	-.014902	-.388416	-.039137	.323521	-.009304
Passives	.203028	.790088	-.036840	.040354	-.090899
Prepositions	.680643	.125145	-.098116	.349496	-.005108
Conjunctions	.543150	-.430607	-.158651	-.186311	-.339877
Adverbs	-.603144	-.087370	-.180850	.044878	-.091410
Nouns	.758364	-.130621	.122859	-.157036	.396410
Adjectives	.197488	.570080	.047652	-.190778	.487355
Pronouns	-.697930	-.406147	-.074574	.010244	-.257445
S.O. total	.161799	.103363	.160375	-.872508	.132272
S.O. preps.	.150776	-.066041	.064822	.751966	-.092289
S. O. Adverbs	-.147651	-.048622	-.239434	.619365	.101419
S. O. Verbs	-.069570	-.036367	-.450271	.003015	.197109
S.O. Sub. Con	-.225920	-.006887	.280791	.251846	-.332406
S. O. Conj.	-.047154	-.184731	-.724079	.132467	.004521
Variance	.135135	.100568	.088972	.113399	.117243

Appendix 9: Similarity Matrix for Multidimensional Scaling of 14 Subjective Variables

variables.	BRACKETS	INVERTED	COLONS	SEMIC	HYPHENS	SLASH	ABBREV	QUEST	FIRSTP
BRACKETS	1.000000	.346535	-.400000	.250836	.142857	.300000	-.503356	-.058366	-.433526
INVERTED	.346535	1.000000	-.285714	.360000	.344538	.388889	-.186441	.770751	.573964
COLONS	-.400000	-.285714	1.000000	.409091	.307692	-.666667	.139535	.214286	.508108
SEMIC	.250836	.360000	.409091	1.000000	.674419	-.285714	-.230769	-.250836	-.111111
HYPHENS	.142857	.344538	.307692	.674419	1.000000	.285714	-.043478	.049180	.269841
SLASH	.300000	.388889	-.666667	-.285714	.285714	1.000000	-.078947	.254237	.333333
ABBREV	-.503356	-.186441	.139535	-.230769	-.043478	-.078947	1.000000	-.488372	.085714
QUEST	-.058366	.770751	.214286	-.250836	.049180	.254237	-.488372	1.000000	.573770
FIRSTP	-.433526	.573964	.508108	-.111111	.269841	.333333	.085714	.573770	1.000000
DWA	.107914	.428571	.230769	.120000	.052632	.096774	-.090909	.103448	.515152
PLUS	.142857	-.083095	-.207317	.290323	.465649	-.363636	-.147982	.226586	-.218274
NOS	.088757	.081967	.169492	-.355450	.217391	.405405	.130435	.428571	.142857
WORDHYP	.327511	.225532	.630435	.215190	.264368	-.250000	.142857	-.072464	.256198
EMPHASIS	-.072464	.330049	-.285714	.036437	.746725	-.071429	1.000000	.072464	.221374

	PLUS	NOS	WORDHYP	EMPHASIS	DWA
BRACKETS	.142857	.088757	.327511	-.072464	.107914
INYERTED	-.083095	.081967	.225532	.330049	.428571
COLONS	-.207317	.169492	.630435	-.285714	.203769
SEMIC	.290323	-.355450	.215190	.036437	.120000
HYPHENS	.465649	.217391	.264368	.746725	.052632
SLASH	-.363636	.405405	-.250000	-.071429	.096744
ABBREY	-.147982	.130435	.142857	1.000000	-0.090909
QUEST	.226586	.428571	-.072464	.072464	.103488
FIRSTP	-.218274	.142857	.256198	.221374	.515152
DWA	.180328	.310881	-.176471	.395349	1.00000
PLUS	1.000000	.342466	.051383	.386282	.180328
NOS	.342466	1.000000	-.111111	.328467	.310811
WORDHYP	.051383	-.111111	1.000000	.006211	-.176471
BOLD	.386282	.328467	.006211	1.000000	.395349

Appendix 10

Original Factor martix from Factor Analysis of 14 subjective style features

variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
EMPHASIS	-.177635	-.550122	.032660	-.324229	.091770	.447320
HYPHENWO	.098032	-.238267	-.297675	.491872	.542638	.019261
DWA	-.680969	.216647	.304561	.021605	-.182989	.165863
NUMBERS	-.720084	.227444	-.338090	.143529	-.136630	-.037384
PLUSSIGN	.217535	.361773	.116244	.157276	-.272131	.619544
FIRSTP	-.305826	.175683	-.293285	.328814	.203213	-.219426
QUESTION	-.221381	-.009099	.112696	-.399130	.279741	-.264749
ABBREVIA	-.043933	-.783688	-.010187	.387538	-.206207	.015953
SLASH	-.759809	.160577	-.289793	.004789	-.025967	.049762
GRAMHYPH	-.564756	-.449268	.309288	-.172317	.247980	.230739
SEMICOLO	-.040764	.218411	.717412	.245809	-.189430	-.105317
COLONS	-.161367	-.055913	.718812	.337600	.312097	-.236831
INVERTED	-.091027	-.319297	-.085921	.480896	-.577221	-.086973
BRACKETS	-.050086	-.292076	-.006921	-.436424	-.494002	-.429378
Variance	.153568	.121413	.116251	.102594	.097417	.074752

Appendix 11

Rotated factor matrix from factor Analysis of 14 subjective variables

variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
BOLD	.081923	-.635842	-.087511	.068255	-.159621
HYPHENWO	.089484	-.141244	-.128504	.167603	.786654
DWA	-.610424	-.066776	.449002	-.046528	-.237739
NUMBERS	-.843353	.007135	-.085525	.063420	.034538
PLUSSIGN	.052064	.507244	.130025	.022882	-.114314
FIRSTP	-.424072	.071916	-.062887	.015198	.413073
QUESTION	-.068861	-.354526	.040136	-.399404	-.088244
ABBREVIA	.127673	-.386225	.028996	.796897	.086904
SLASH	-.809685	-.154253	-.077543	-.046926	.013157
GRAMHYPH	-.209232	-.747840	.320192	-.014851	-.043031
SEMICOLO	.043676	.242628	.754294	.048330	-.168162
COLONS	.083841	-.167804	.812895	-.013178	.247133
INVERTED	-.147671	.120578	.027491	.796035	-.106263
BRACKETS	.032708	-.206904	-.195834	.177010	-.639380
Variance	.145239	.120965	.116700	.107443	.100895

References

Anderson, A. , Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S, Isard, S., Kowotko, J., McAllister, J., Miller,J., Sotillo, C., Thomson, H., Weinhart, R.,(1991). The HCRC Map Task Corpus, *Language and Speech,* 34, 4, 351-366.

Baker J.C., (1988). Pace: A Test of Authorship Based on the Rate at which New Words Enter an Authors Text, *Literary and Linguistic Computing,* 3(1), 36-39.

Bayne, R., & O'Neil, F., (1988). Handwriting and personality: A test of some expert graphologist's judgements. *Guidance and Assessment Review,* 4(4), 1-3.

Black, H.C., Nolan, R.J. & Nolan-Haley, J.M. (1990). *Black's Law Dictionary,* 6th edn., St. Paul, MN, West Publishing.

Brett, S. (1989). *The Booker book,* London: Sidwick & Jackson.

Burrows, J.F., (1987). Word Patterns and Story Shapes, The Statistical Analysis of Narrative Style, *Literary and Linguistic Computing,* 2(2), 61-70.

Burrows, J.F.,(1986). Modal Verbs and Moral Principles: An Aspect of Jane Austen's Style, *Literary and Linguistic Computing,* 1(1), 9-23.

Bushnell, I.W.R., (1993). A Comparison of thenValidity of Handwriting Analysis with that of the Cattell 16P.F.. Presented at B.P.S. London, 1992 and Open University Postgraduate Conference , 1993.

Canter, D., (1992). An Evaluation of the "Cusum" Stylistic Analysis of Confessions. *Expert Evidence*, 1, 93- 99.

Canter, D., (1994). *Criminal Shadows*. Harper Collins.

Cattell, R,B., (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.

Cattell, R.B., (1979). *Personality and learning theory*. New York: Springer.

Cherry, L.L. & Vesterman, W. (1980). Writing Tools- The Style and Diction Programs.

Churchill, G.A. (1987). *Marketing Research: Methodological Foundations.* New York: The Dryden Press.

Crumbaugh, J. C., & Stockholm, E. (1977). Valuation of graphoanalysis by global or holistic methods. *Perceptual and Motor Skills*, 44, 403-410.

Damerau, F.J., (1975). The Use of Function Word Frequencies as Indicators of Style. *Computers and the Humanities*, 9, 271-280.

De Morgan, A., Personal letter cited in Morton (1990).

Ellegard, A., (1963). *Who was Junius?* Almquist & Wiksell: Stockholm.

Expert Evidence (1992). Editorial Introduction: Some legal issues affecting novel forms of expert evidence. *Expert Evidence*, 1(3),79-83.

Eysenck, H.J. (1982). *Personality genetics and behaviour.* New York: Praeger.

Greene, P.E. & Carmone, F.J., (1970). *Multidimensional scaling and related techniques in marketing analyis.* Boston: Allyn & Bacon.

Halliday, M. A. K., (1990). *Spoken and written language.* Oxford University Press.

Hardcastle, R.A., (1993). Forensic Linguistics: an assesment of the CUSUM method for the determination of authorship. *Journal of the Forensic Science Society*, 33, 95-106.

Hartigan, J.A.,(1975). *Clustering Algorithms.* .: New York:John Wiley & Sons.

Hilton, M. L. & Homes, D.I. (1993). An Assessment of Cumulative Sum Charts for Authorship Attribution. *Literary and Linguistic Computing*, 8(2), .

Jackson, MacD:P.,(1991). George Wilkins and the First Two Acts of *Pericles*: New Evidence from Function Words. *Literary and Linguistic Computing* , 6(3), 155-163.

Jackson, B (1983) *Multivariate data analysis*:. Homewood, Illinois: Richard D. Irwin Inc..

Jamison & Aitken (1993). A Survey of Approaches in Authorship Studies. Report from University of Edinburgh, Department of Computer Science.

Johnson, P.F., (1973). The Use of Statistics in the Analysis of the Characteristics of Pauline Writing. *New Testament Studies,* 20, 92-100.

Kenny, A., (1982). *The Computation of Style*. . Oxford:Pergamon Press.

Lewis, P.A. (1965). Some Results on Tests for Poisson Processes. *Biometrika*, 52, 67-77.

Linell, P. (1988). The impact of literacy on the conception of language: the case of linguistics. In R. Saljo (ed.) *The written world.* Berlin: Springer, 41-58.

Loewenthal, K. (1982). handwriting as a guide to character. In D.M. Davey, & M. Harris (Eds.): *Judging people: a guide to orthodox and unorthodox method sof assesment.* Maidenhead: McGraw-Hill.

Mackay, I.. Unpublished Masters thesis, University of Glasgow.

McMenamin, G. (1993). Forensic Stylistics. Amsterdam: Elsevier 1993.

McPhail, I.D. (1987). Evidence. A Revised Verson of a Research Paper on the Law of Evidence in Scotland. The Law Society of Scotland, Edinburgh. MacDonald Printers Ltd.

Mendenthall T.C., (1887). The Characteristic Curves of Composition, *Science* , 9, 237-249.

Merriam, T., (1986). The Authorship of Sir Thomas More, *Literary and Linguistic Computing,* 1(2), 104-108.

Merriam, T., (1982). The Authorship of Sir Thomas More. *ALLC Bulletin,* 10, 1-7.

Michaels, G., Maze, M. , & Hodos, D., (1989). *Handwriting analyst.* Berkeley, Cal: Ciasa.

Michaelson, S. & Morton, A.Q, (1972). Last Words. A Test of Authorship for Greek Writers, *New Testament Studies,* 18, 192-208.

Michaelson, S., Morton,A.Q., (1977). Hamilton Smith, N., To Couple is the Custom. Unpublished report, University of Edinburgh.

Michaelson, S. & Morton, A.Q.,(1973). Positional Stylometry, *The Computer and Literary Studies,* 3, 69-83.

Miller, J. (in press). Does spoken language have sentences? Ed. Palmer, F. Festschrift for Sir John Lyons: Cambridge University Press.

Miron , M.S., (1981). The Resolution of Disputed Communication Origins. *Speech and Language: Advances in Basic Research and Practice,* 5, 505-466.

Morton, A.Q. (1978). *Literary Detection,* London: Bowker.

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), 1-8.

Morton, A.Q. & Farringdon, M.G. (1992). Identifying Utterance. *Expert Evidence,* 1, 84-92

Morton, A.Q. & Michaelson, S., (1990). *The Q-sum plot.* University of Edinburgh, Computing Science Department, Internal report no. CSR-3-90.

Morton, A.Q., (1991). *Proper Words in Proper Places,* University of Glasgow, Computing Science Department, Departmental Research report 1991/R18.

Mosteller, F. & Wallace, D.L., (1964). *Inference and Disputed Authorship: The Federalist.* Addison-Wesley Publishing Company.

Moxey, L.M. & Cookson S., (1987). *The Adaptive Stylechecker.* Alvey Project: Adaptive Dialogue working paper.

Pervin, L.A., (1984) *Personality: Theory and Research..* John Wiley & Sons: New York.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language.* London: Longman.

Rosenberg, S., (1977). New approaches to the analyis of personal constructs in person perception. In A. Landfield (Ed.), *Nebraska symposium on motivation* , 24. Lincoln, NE: University of Nebraska Press.

Sanford, A.J. & Moxey, L.M., (1989). Language Understanding and the Cognitive Ergonomics of Style. In (Williams, N. & Holt, P., eds.) *Computers and Writing; models and tools*. Intellect Books. Oxford: Blackwell Scientific Publications Ltd.

Smith, M.W.A.,(1982). The Authorship of Pericles: An Intitial Investigation. *The Bard*, 3, 143-176.

Smith, M.W.A., (1987). The Authorship of *Pericles*: New Evidence for Wilkins. *Literary and Linguistic Computing*, 2(4), 221-229.

Smith, M.W.A, (1987). Hapax Legomena in Prescribed Positions: An Investigation of Recent Proposals to Resolve Problems of Authorship. *Literary and Linguistic Computing*, 2(3), 145-153.

Smith, M.W.A., (1985). An Investigation of the Basis of Morton's Method for the Determination of Authorship. *Style*, 19, 3, 341-368.

Smith, M.W.A., (1988). The Authorship of Acts I and II of Pericles: New Approach Using First Words of Speeches. *Computers and the Humanities*, 22, 23-41.

Smith, M.W.A., (1989). Forensic styleometry: A theoretical basis for further developments of practical methods. *Journal of the Forensic Science Society;* 29(1), 15-33.

Stratil, M. & Oakley, R.J., (1988). A Disputed Authorship Study of Two Plays attributed to Tirso de Molina, *Literary and Linguistic Computing,* 2 (3), 153-160.

Taylor, G. & Wells, S., (1987). *William Shakespeare: A Textual Companion.* Oxford: Clarendon Press.

Totty, R.N., Hardcastle,R.A., & Pearson, J., (1987). Forensic Linguistics: The Determination of Authorship from Habits of Style. *Journal of the Forensic Science Society,* Series A, 27, 13-28.

Wackernagel-Jolles, B. (1971). *Untersuchungen zur gesprochenen Sprache: Beobachtungen zur Verknupfung spontanen Sprechens.* Goppingen: Alfred Kummerle.

Wake, W.C.,(1957). Sentence Length Distributions of Greek Authors,. *Journal of the Royal Statistical Society,* Series A., 120 (3), 341-346.

Williams, N. & Holt, P.,(Eds.)(1989). *Computers and Writing: models and tools.* Intellect BooksBlackwell Scientific Publications Ltd: Oxford.

Woodward, R.H. & Goldsmith, P.L., (1964).Cumulative Sum Techniques. *Mathematical and Statistical Techniques for Industry,* Monograph No.3. Edinburgh: Oliver and Boyd.

Yule, G. U. (1939). *Biometrika*, 30, 363-384.