Handwriting Classification in Forensic Science

Michael Ansell

Recent methods of classification of features in handwriting for use in the forensic examination of documents are described, including the computerized system of classification. Developments in the statistical analysis of the way people construct numbers and lay out their writing are reviewed. The potential usefulness of these systems in quantifying the current document examiners' scale of probability for attributing questioned writings to particular authors is examined.

1.0 *Introduction*

In the past so-called experts have reached erroneous conclusions about handwriting classification often because they have studied overall similarities in style of writing rather than the detailed construction of letterforms. Fortunately, the current scientifically trained document examiner arrives at a more statistically justifiable conclusion. However, the science of document examination has advanced mainly as a result of developments in the chemistry and physics of documents, which can be applied to any document whether handwritten or not. The technique of electrostatic detection, for example, is used to recover hitherto invisible indented impressions of writing, or the technique of infra-red luminescence is used to recover obliterated writing. Only recently has significant progress been made in putting on a scientific basis the evaluation of authenticity in handwritten documents in terms of the handwriting per se.

Approaches to forensic handwriting comparison in general may be divided into the Osborn (1929) approach and the European approach. The European approach can be considered to be graphological. It requires an analysis of the psychological and medical condition of the writer of one document as compared with the writer of another. This leads to propositions such as: The writer of document A is a young woman in good health and that of

Visible Language, XIII 3 (1979), pp. 239-251.
Author's address: Metropolitan Police Forensic Science Laboratory, 109 Lambeth Road, London SE1 7LP, England.
0022-2224/79/0700-0239$02.00/0

document B is an old man, feeble, and suffering from brain damage. Therefore document A could not have been written by the writer of document B.

The Osborn (see also Harrison, 1966) approach is that practised by document examiners in the United Kingdom. It involves analysis of the physical result on paper caused by the writing process, without excursion into possible psychological or medical causes of unusual strokes or characters (except possibly for disguised writing). This paper is generally concerned with the latter approach. A direct comparison is made by the examination of writing on the questioned document and of writing on one or more samples of known authorship (control documents). This comparison is based on sets of features present in the handwriting that the examiner believes to be relevant to differences *between* individuals. At the same time the features chosen should not be likely to display variation from occasion to occasion *within* the writing of any one individual (this could lead to the inappropriate conclusion that two samples of writing by the same person — but taken at different times — were not of common authorship). To be practical the features chosen must also be easily identified and measured.

In the next two sections I consider systems for the classification of handwriting features; following that a section is devoted to problems involving statistics for the results of such classification. Such statistics are relevant to the efficient presentation of handwriting evidence in court.

2.0 *Classification Systems*

Whether as a result of differences in education, culture, and/or personal preference, people develop individual styles of writing. However, a general statement about overall style is usually not enough to discriminate between the writing of different people. Account must usually be taken of the details of construction of individual letters. Various sets of features or classification systems have been proposed for this purpose (for a review see Baxter, 1973). In this section we consider two illustrative studies.

2.1 *Ansell and Pritchard.* In collaboration with H. Pritchard at the Metropolitan Police Forensic Science Laboratory (London) I have used eighteen parameters to classify samples of normal block capital writing of 134 writers. Only three pairs of writing samples

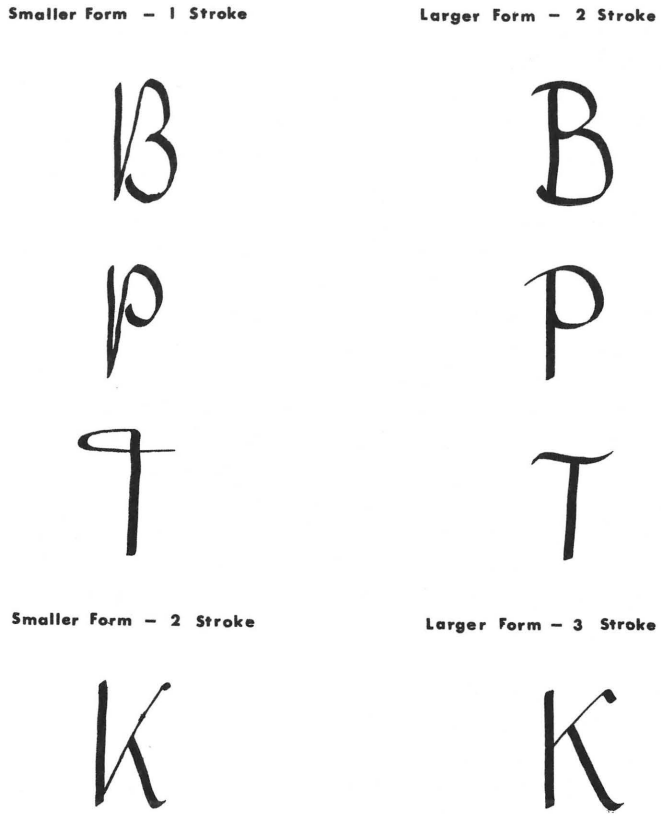

Figure 1. Comparison of the number of strokes required for regular (left) and very large handwriting styles.

emerged as being indistinguishable in respect of all the parameters. These parameters may be summarised:

A Angular or Broad?
A, B, D, etc. Do they start with a downstroke or not?
G Does it have a tail?
K Is it made with 2 or 3 strokes, are they radial?
M Is the centre short or long?
W Is the centre short or long?
U Does it have a tail?

We then used this system to consider the effects of writing on a vertical or horizontal surface. Cases occur in which an offence

such as obscene writing on a wall is committed where, understandably, the offending "document" cannot be conveniently submitted to the laboratory. A further problem arises that control samples based on the same instrument and surface are rarely submitted by the investigating officer! It is within our normal field of experience to consider and assess the significance of resemblances between such questioned documents and control samples of normal horizontal writing on paper. However, where there are differences, are they due to different authorship or are they due to variations that arise when one and the same person writes on different surfaces and in different planes?

An experiment was run with fifteen people who were asked to provide two samples of the phrase BUS PARKING PROHIBITED in block capital letters. One sample in large writing with black felt-tip pen was written on a card placed at a convenient height on the wall. A ballpoint pen was used to provide the second sample written normal size on a card placed on a desk top. We found that only 17% of letter pairs differed in construction between the two samples. The difference was almost always the use of a letter in large writing which increased the number of strokes required (Figure 1).

It may, therefore, be concluded that with appropriate choice of features it is possible to devise handwriting classification schemes that are sensitive to differences between individuals. Moreover they can give consistent results within individuals even where the writing situation is significantly changed from one sample to the next.

2.2 *Hensel, Khan, and Dizon.* The study discussed above involved block capital handwriting which is often contrasted with cursive (or joined-up) writing and numerals. Disconnected or script handwriting is not encountered quite as often in adults as in children and may be classified in a similar manner to cursive writing or individual letters with no need to consider the method of joining successive letters.

Hensel, Khan, and Dizon (1973) considered the problem of classification of non-roman script. If the document examiner whose native language is English has to examine writings in Cyrillic script (such as Russian or Serbo-Croat) where some letters are similar to roman but others are quite different, a typed or block capital transliteration is essential. This is because in the examina-

tion of unfamiliar cursive words it is difficult to establish where one letter ends and another begins, or how many letters there are in each word. As an example, in English the word minimum can be written with fifteen joined identical curved segments. If the word were not recognisable in the context of a sentence, the individual letters — not to mention strokes — could not be identified. Scripts such as Arabic (far removed from roman writing) present much greater problems to a document examiner whose native language is English. Although the identification of individual letters is very difficult, the fact that the writing reads right to left presents no additional difficulty.

To compare two or more Arabic writings to establish common authorship a document examiner without any knowledge of the language or the script needs to be sure that he is comparing like with like. He will also want to know the variations that are likely to occur between different people's writings and which letter forms are common or rare.

When this has been done, he can then apply the scientific principles of observation and deduction with which he is familiar, whatever the language. Hensel, Khan, and Dizon used the method of comparison of an "intelligent illiterate" who uses the criminalistics techniques commonly employed with toolmarks, firearms, and footmarks. They also encountered complications introduced by the illiteracy of native writers. In such cases a scribe or professional letter writer will, for a fee, prepare applications, write petitions, and fill in forms which are then "signed" by a thumbprint. The authors examined one fraud case involving 104 documents bearing 147 signatures of illiterate labourers. All these documents had been written by a single clerk!

3.0 *Computers and Handwriting Classification.*

The computer offers the means of storing a large amount of information about handwriting features. This information may then be rapidly scanned for statistical purposes such as those discussed in the next section or for direct comparisons of individuals' writing (after classification).

A system in use at the Zurich Kantonspolizei Laboratory (Angst and Erismann, 1972) employs a large number of features. A one-paragraph sample of writing is copied by the individual concerned and later classified by a document examiner. Classification takes at least an hour, but once the results are entered into the computer,

comparisons with a large data bank may be made very rapidly to select other entries with varying degrees of correspondence. Each sample is classified according to the broad class of writing skill, line quality, slope, size, width, angularity, type and degree of connection, position of accents, and detailed construction of various cursive and block capital letters. Unlike systems where the classifier has to indicate which type of any particular letter or digit is present (e.g., Ansell and Strach, 1975), the Zurich system depends upon a series of "yes" or "not present" answers; e.g., "Is A angular?" if so, check item 53; "Is A broad?" if so, check item 54. This works well in practice for the Swiss-German letters encountered.

4.0 *Statistics and the Presentation of Handwriting Comparison Information*

Following comparison of questioned and control documents, a document examiner will usually be able to express his degree of certainty in his conclusions at one of four levels:

1 Definitely written by the same person.
2 High probability that they were written by the same person.
3 Could have been written by the same person.
4 No evidence that they were written by the same person.

These expressions have implications that are appreciated by document examiners, but may not be fully comprehended in the courtroom. Indeed, members of the legal profession would probably be very happy to see the introduction of a sixteen-point system such as used in fingerprint examination. In this section we consider the difficulties in the reliability of documentation of any particular piece of forensic handwriting comparison. The underlying interest in the following studies is to what extent they tell us about questions such as: What is the likelihood that two different people will write the same way? or How unusual in the population is a particular distinctive handwriting feature or set of features?

4.1 *Livingstone.* Livingstone (1959, 1963) described a system of "pen printing" classification which included cursive writing. In the later work he describes some statistics drawn from 200 samples stored in his collection of edge-punched cards. As well as layout he treated features or groups of letters such as g, y, and z.

Table I. Classification of letter features. Livingstone (1963)

	System A		*System B*	
	Complete tail	*Abbreviated tail*	*One or more angles*	*Without angle*
Letters g, y, z	82%	18%	30.5%	64%

He considered in particular their tails, and divided these into complete or abbreviated (that is, without lower loop) on one system (A) and incomplete plain loop without angle or with one or more angles in another system (B) (Table I).

Some care must be exercised in applying a system developed in one country to nationals of another. For example, Livingstone considers the features of dating layout in the United States; e.g., 23 November 1962 is the dating style used only 4% of the time in the USA but is in frequent use in the United Kingdom.

4.2 *Harvey and Mitchell.* In the case of the murder of Nicola Brazier, Harvey and Mitchell (1973) were faced with a protracted examination and elimination of suspect writings from a large number of people. (The author has on occasion been faced with a similar problem involving 4,000 writers.) Their questioned document was a cheque for five pounds, four shillings, and ten pence dated 7th September 1970. They chose six characteristics from the cheque and scrutinised each of 1,046 samples for these features only:

Layout

1 Presence of double dashes in the sum of money £5 = 4= 10 (38 out of 1046)
2 Position of the "th" in the "7th" of the date. High (366) Low (166) Absent (514)
3 Presence of indentation of the word "five" following the printed word "pay" (883)

Letter Design

4 Short staffed g (4)
5 The "ce" in pence with the e larger than the c (15).
6 The "x" in Essex having bottom left to right stroke longer (36)

Table II. Frequency of joint occurrence of double dash and indentation. Harvey and Mitchell (1973)

	Double dashes	*Not double dashes*	*Total*
No indentation	11 [5.92]	152 [157.07]	163
Other	27 [32.07]	856 [850.92]	883
Total	38	1008	1046

They then considered the hypothesis that the features Double dashes/No indentation were independent. This required determination of the frequencies of joint occurrence of double dash and indentation (Table II). If two features are independent, row and column totals may be used to give expected frequencies of the joint occurrences (in brackets).

The chi square statistic may then be used to determine the statistical reliability of departures in the data from the theoretical predictions based on the assumption of independence,

$$\chi^2 = \sum_{I=1}^{K} \frac{(O_I - E_I)^2}{E_I}$$

O being the observed result, E the expected result, and K the number of cells in the table. The value $\chi^2 = 5.345$ obtained indicates a reliable degree of dependence. The rejection of independence in this case may be contrasted with application of the test to "low th" versus "no indentation." There $\chi^2 = 0.00018$, indicating no signs of dependence. This kind of information is useful to the document examiner who doesn't want to spend extra time evaluating another possible feature, if that feature provides information depending strongly on a feature already assessed. On the other hand, if two features that in the population are highly dependent diverge in both questioned and control documents, this information is evidence for common authorship.

4.3 *Ansel and Strach.* Ansell and Strach (1975) concentrated on the apparently simple task of classifying the methods of writing numbers (Figure 2). The manner of writing 0 and 8 by 993 people is summarised in Table III. Consider classes 2, 3, and 4 in Table

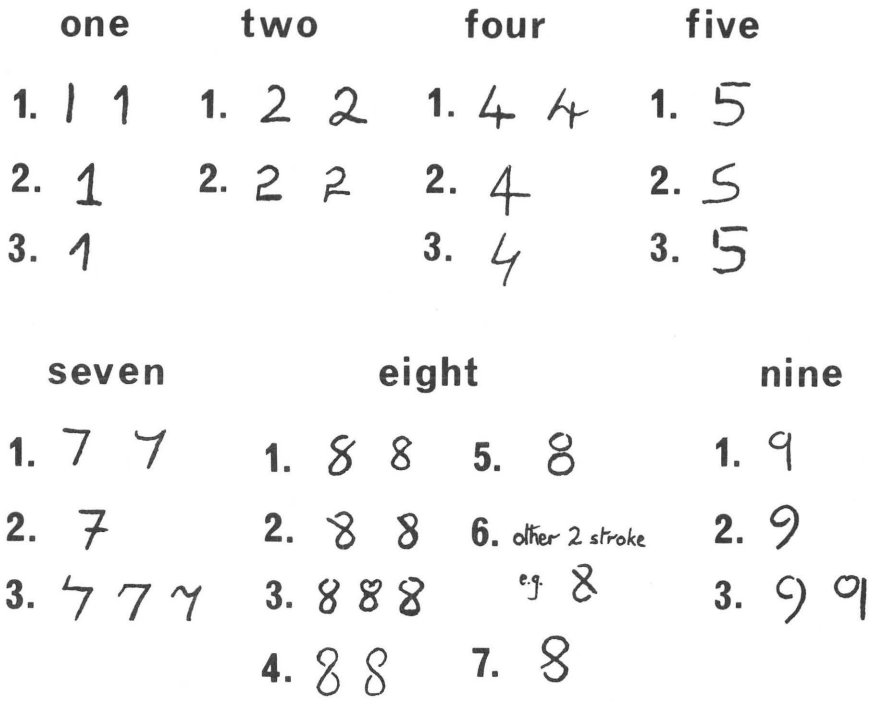

Figure 2. Classification of methods used to write the numbers 1, 2, 4, 5, 7, 8, and 9. Ansell and Strach (1975)

III where 0 and 8 are classified according to the position of any opening, join, or discontinuity.

The number of people writing a particular class of number 0 in combination with a particular class of number 8 is given in Table IV. For example the entry in the bottom right-hand corner of the table means that 43 people joined both the 0 and 8 at the top. The (35) indicates the frequency expected on the basis of Table III assuming independence of class assignment. For these data $\chi^2 = 30.01$, which indicates a reliable departure from independence in manner of construction of these two digits.

Further research on these lines has shown that classification of the numeral 0 is unsound because many people write several or even all classes of this number. Taking into account the fact that people wrote several classes of each number and adjusting the classification to be more discriminating, we can see the results of the modified classification of 8 in Table V. We considered 90 peo-

Table III. Classification of the manner of writing the numbers 0 and 8 according to position of any opening, join, or discontinuity. Ansell and Strach (1975)

Number 0			*Number 8*		
Class 1	Join ambiguous	34%	Class 1	Join ambiguous	18%
Class 2	North West	36%	Class 2	North East	51%
Class 3	North East	7%	Class 3	North West	14%
Class 4	North	26%	Class 4	North	14%
Class 5	Two stroke forms	0.8%	Class 5	Middle	1%
Class 6	Other forms.	0.9%	Class 6	Separate circles	0.6%
Class 7			Class 7	Other two stroke forms	0.5%
Class 9			Class 9	Others	1.2%

ple who wrote five examples of this number (but of unknown handedness), 140 people who wrote two examples of this number (also of unknown handedness), 833 people who were right-handed and wrote one example of this number, and 101 left-handed people writing one example of this number. Unfortunately this system is still limited in that about 14% of writers would have their numbers (1, 2, 4, 5, 7, 8, and 9) assigned to the highest frequency class for these digits.

Table IV. Classification comparison of people writing the numbers 0 and 8. Ansell and Strach (1975) (See Table III)

Classes of number 0	*Classes of number 8*			
	1,5-9	2	3	4
1,5-9	78(66)	157(166)	51(48)	42(48)
2	65(67)	154(170)	64(49)	53(49)
3	15(14)	45(35)	4(10)	5(10)
4	37(48)	137(121)	22(35)	43(35)

Table V. Modified classification of handwriting number 8 according to position of any opening, join, or discontinuity. Ansell and Strach (1975)

	90 people 5 examples Unknown handedness	*140 people 2 examples Unknown handedness*	*833 people 1 example Right-handed*	*101 people 1 example Left-handed*
1. North East	55.8%	63.87%	46.74%	58.4%
2. North West	5.1%	14.24%	12.29%	3.0%
3. North	36.4%	15.69%	36.88%	33.6%
4. Middle	—	2.18%	1.44%	2.0%
5. Separate Circles	1.1%	1.82%	0.48%	—
6. Other two stroke forms	1.1%	0.74%	0.84%	—
7. South West	0.21%	—	0.84%	2.0%
9. Other	0.25%	1.46%	0.48%	—

4.4 *Allan, Pearson, and Brown.* In a recent experiment Allan, Pearson, and Brown (1978) used 52 people each writing a short length of prose. Eight measurements were taken:

1 Number of lines.
2 Margin width.
3 Paragraph indentation.
4 Length of last ten spaces.
5 Length of last eleven words with spaces.
6 Length of first ten spaces.
7 Length of first eleven words with spaces.
8 Ratio of relative height of letters with ascenders.

It will be noted that none of these involve details of method of letter formation. They then used a measure

$$D \star \sqrt{\sum_{i=1}^{n} \frac{(x_i - y_i)^2}{m_i^2}}$$

to characterise the difference between control and questioned samples, where x_i and y_i are the measurements in the "i" the dimension, m is the mean value over subjects for that dimension, and n is the number of dimensions. A computer program was used

to compare measurements. The program calculates the "distance" between the measurements on the known and questioned samples in the dimensions required. Comparisons were made of a person's handwriting with other samples of that person's writing that had been disguised or had been written after an interval of one year. Only 5% of writings by other people were computed to be closer to the known person's writing than were his time lapsed or disguised samples.

It may be expected that work such as that of Allan, Pearson, and Brown and of Ansell and Strach will in future lead to a clear statement in the statistical confidence of handwriting evidence given in court. At present it is limited to suggesting a reply to questions such as: Is this a common letter T? Is this a common style of writing or layout? The answers would be in the form: No, it occurs to an extent of less than x% in samples submitted for examination.

In practice the work of Ansell and Strach is used in a negative sense if the defence claims resemblances between the writers of, say, a questioned 2 and 4 of a particular person as being important, but the resemblances can be shown to be statistically not significant.

Conclusion

In this short review a limited selection of the available work has been considered to give some idea of the range of forensic handwriting research. It ought, however, to be mentioned that in the field of forensic science it is normal to refer to the person carrying out handwriting examinations and comparisons not as a "handwriting expert" but as a "document examiner." Only rarely can comparison of handwriting be carried out without wider examination of the writing materials or instruments used, or perhaps needing to clarify or restore the writing before it can be compared.

So what of the future? More detailed statistics are needed as to letter construction, style, dimensions, and letter and word spacing; also more work on the dependence of construction of a particular letter upon the method of construction of other letters by the same author. I would like to see the document examiner being able to make statements in court of the kind: only x% of the population's writing bears the same number of resemblances to the control writing as does the suspect's.

Finally, although experienced document examiners often possess subjective (but hopefully reliable) ideas of the effects of disguise, illness, infirmity, wrong-handedness, duress, alcohol, or drugs on a person's handwriting, little objective statistical work has been done (although Allan, Pearson, and Brown make a start on the effects of disguise and of time lapse). A helpful approach could be for a psychologist to classify a subject's writing using one or more of the systems mentioned in this review (or a modified version of the system) and a statistician to assess the usefulness of the results.

References

Allan, A.R., Pearson, E.F., & Brown, C. A comparison of handwriting characteristics. 8th Meeting of International Association of Forensic Sciences, Wichita, 1978.

Angst, E., & Erismann, K. Auswertung von anonymen und pseudonymen handschriften mit electronischer datenverabeitung. *Kriminalistic*, 1972, *2*, 60.

Ansell, M., & Strach, S.J. The classification of handwritten numerals 7th Meeting of International Association of Forensic Sciences, Zurich, 1975.

Baxter, P.G. Classification and measurement in forensic handwriting comparison. *Medicine, Science and the Law*, 1973, *13*, 3, 166-184.

Harrison, W.R. *Suspect Documents*. London: Sweet and Maxwell, 1966.

Harvey, R., & Mitchell, R.M. The Nicola Brazier Murder. *Journal of the Forensic Science Society*, 1973, *13*, 157.

Hensel, E.B., Khan, I.A., & Dizon, J. Forensic examination of popular handwriting systems. *Journal of the Forensic Science Society*, 1973, *13*, 2, 143-152.

Livingstone, O.B. A handwriting and pen-printing classification. *Journal of Criminal Law, Criminology and Police Science*, 1959, *49*, 5, 487-506.

Osborn, A.O. *Questioned Documents*. London: Boyd, 1929.