Barry Salt:

Statistical Style Analysis

The basic idea behind my methods of statistical style analysis is that the form of films noticeably differ from one to another, and that the variables used to study this should be based on the concepts that film-makers actually use.

Average Shot Length

The simplest to get is the Average Shot Length (ASL). This is the length of the film in seconds divided by the number of shots in it. I only quote one decimal place, except for films with an ASL under 2 seconds, for which I quote two decimal places. For older films I only take the length and count the shots from after the last title card in the front titles, which is usually the director's credit. For modern films that have a pre-credit sequence, I add in the length and number of shots in that as well. I do not include the length and shots in a credit sequence that has titles superimposed on action. In silent films, and indeed sound films, I count intertitles as shots.

For silent films, I measure the ASL at what seems to me to be the correct speed of projection, on the basis of the naturalness of the movement of the actors. That is, at the speed the film would have been shot at when it was made. This is admittedly a tricky point, and there is something to be said for quoting a value as though it was being projected at sound speed, because this value is more definite. But on the other hand, this produces the misleading impression in people who do not understand the situation, that the cutting is faster, in comparison to that of sound movies, than it really was when the film was made. And for movies that were cranked at 16 frames per second, like *Battleship Potemkin*, that gives an ASL 33% less than it is at the correct speed.

I count the number of reverse angle cuts (RA), which are defined as changes of camera direction of more than 90 degrees, and I express it as a percentage of the total number of shots in the film.

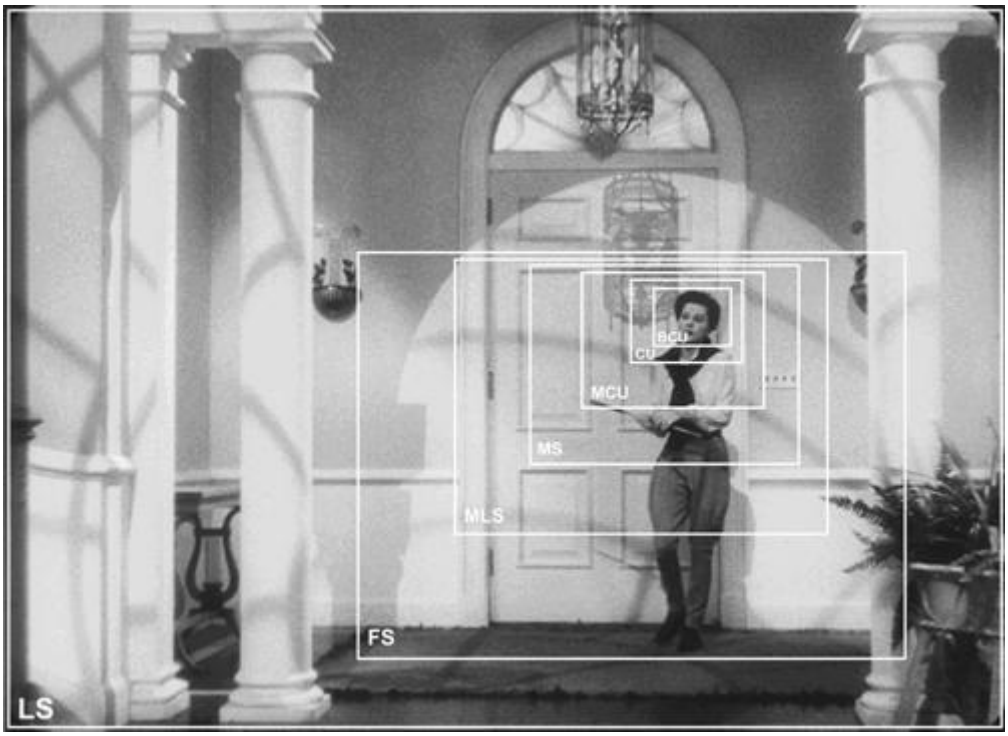
Another variable I collect is the number of cuts to or from a Point of View (POV) shot in the two films. POV shots are those taken from the position of one of the participants in the scene, or at any rate from a position on their line of sight. these are expressed as a percentage of the total number of shot transitions in the film, which equals the number of shots in the film.

Insert shots (INS) are shots of things, and do not include an actor's face. I count a very distant shot of a street scene, in which no principal actors are included, as an insert. They are also expressed as a percentage of the total number of shots.

Shot Scale

Although there is a small amount of disagreement about precisely what shot scale corresponds to each descriptive term, it is sufficient for the purposes of analysis to define carefully what one means by each category, and then stick to it. I am using categories of Scale of Shot like those commonly used in the film industry from the nineteen-forties, as follows: Big Close Up (BCU) shows head only, Close Up (CU) shows head and shoulders, Medium Close Up (MCU) includes body from the waist up, Medium Shot (MS) includes from just below the hip to above the head of upright actors, Medium Long Shot (MLS) shows the body from the knee upwards, Long Shot (LS) shows at least the full height of the body, and Very Long Shot (VLS) shows the actor small in the frame. In recent decades in film and television the vaguer term

‘Wide Shot’ has come to replace the various kinds of Long Shot described above, but I am keeping to the more finely graded terminology used when I first became involved with film-making nearly fifty years ago. This shows up characteristic distinctions between many films from earlier periods of cinema, and even today, certain important distinctions would vanish if one used only a three level scale.



For comparative purposes, so as to give an obvious relative measure of the likelihood of a director choosing a particular scale of shot in a film, I have taken the actual number of shots of each scale in a film, and then normalized the number to correspond to the number that would have occurred if the film was made up of 500 shots. If you want percentages, you can divide the figures by 5.

Camera Movement

I handle the objective treatment of camera movement by tabulating the numbers of shots with the different kinds of camera movement in each film, and again normalizing to the number that would be expected if the film in question contained 500 shots. The categories I use are Panning, Tilting, Panning and Tilting simultaneously, Tracking both without and with panning movements, movement involving the use of a camera crane, and Zooming. I have divided this last category into Zooming straight in or out, and Zooming with panning and/or tilting in my treatment of television shows. Only panning or tilting movements of more than 10 degrees are counted, as small movements to keep the actors well framed as they change their position slightly are made automatically by camera operators, and in general need no special thought about their relation to the director's ideas of staging. The same applies to small movements of a foot or so in the position of the rolling camera pedestal or dolly during the shot, and also of the height of the camera. I do not distinguish the different methods of supporting the camera, so that hand-held tracking and Steadicam tracking go in together with the traditional tracking with the camera on a dolly, or rolling pedestal in the case of TV.

Analysing the complete film is obviously much preferable, but you will not be too far out in general with a forty minute sample. Contrary to received ideas, the overall cutting rate for the second half most films is either the same speed, or not much faster than that of their first halves, though there are exceptions.

Practical Details

Ideally, the analysis should be done by recording the complete characteristics of each shot (scale, movement, length, etc.) in succession down the length of the film. This permits the most complete analysis of all the possible interrelationships between the variables. But although I initially tried this thirty odd years ago, I found that it took about three times longer than the method I have since used, so I have only used it for a few special cases since. My present method collects each quantity sequentially over the length of the film.

Up until recent years, I have always worked with prints of the films I was analysing, and indeed almost exclusively with 35 mm. prints, and I worked with them on Steenbecks and other flat-bed editing machines, which is obviously ideal for complete accuracy. But now I have turned to working from DVDs and VHS tapes. I feed these into a non-linear editing system (NLE), in fact Adobe Premiere on an ordinary PC (though a cheaper NLE would do just as well), and while they were being digitized in real time, I record the camera moves from the window in the digitizing programme screen. For the experienced analyst, this is just possible to do in real time, even for the fastest cut films. Then I go more slowly through the film in the NLE programme, recording the Scale of Shot, which usually requires some stopping and starting and going back, particularly for the films with very short ASLs. I also record the Inserts on this pass. Two more passes are necessary to get the numbers of reverse angles and POV shots. If I have a VHS tape, and one can always make one from a DVD, I usually do these last things on a VHS recorder with a jog-shuttle control, as I can usually manage recording these last two quantities at high speed for most films. Alternatively, it is possible to do the complete analytical process entirely on a VHS recorder with a jog-shuttle control, as I have done when analysing television programmes in the past. For my analytical procedure the standard control system for DVD players is very awkward to use when trying to work directly with the DVD disc.

There are important cautions to be made about the analytical process when working from tape recordings or DVDs of films. The first of these relates purely to the use of recordings made for the PAL television system. These are initially created from film prints that were shot at 24 frames per second when the original film were made, but are always transferred to the consumer medium at 25 frames per second. This means that their running time when played on PAL system devices is shortened by 4% of the original running time. This means that a correction factor has to be applied to the ASL has to be applied by multiplying it by a factor of 25/24. I have applied this correction in the above results. No correction is necessary for NTSC recordings. More important is the question of Scale of Shot determination from video and DVD copies of films. For old Academy screen ratio films, both 16 mm. copies and, even more so, video copies are cropped in all around the frame on transfer to a greater extent than the screen masking when they are shown in the cinema, or on a Steenbeck. The effect of this on the Scale of Shot is fairly slight, as it shifts a very small proportion of the CUs into the BCU category, and an even smaller proportion of the more distant Shot Scales into the next closer category. Since all American feature films made since 1954 are intended to be masked to widescreen on projection, or are shot in one of the anamorphic ‘Scope systems, or in a wide film system, the difficulty does not exist in

quite this form for wide screen films. The problem is that films made since then which are shot “flat”, i.e. with spherical lenses on the camera, may have the full Academy image, which was invariably recorded on the negative for American films, transferred to video, and not masked in to the widescreen proportions that were intended to be seen in the cinema. Despite the fact that DVD transfers are virtually always given the correct masking, and there is an increasing trend to releasing VHS copies properly masked in to wide screen, this problem has received a new boost from the shooting of many films in Super 35. In this process, the camera exposes what is called the “full” aperture in the gate of the camera, which is equivalent to the old silent period aperture. This image on the original film is masked in to widescreen or even to ‘Scope proportions by optical printing when making the release prints of the film.
Where possible, I used DVD copies when analysing the sample, and I also check with the VHS copy of the same film where possible. This check showed that in the case of *Deep Blue Sea*, which was shot in Super 35, the VHS copy had been taken from the full frame, and the DVD copy, like the cinema prints, was taken from the middle of the original frame in ‘Scope proportions. This meant that for any shot much more could be seen of the scene vertically in the VHS frame than could be seen in the DVD copy. That is, if I had analysed the VHS copy, I would have found that the film was shot from much further back than it really was, with respect to the intended cinema release framing. Another difficulty that can occur with films shots in Panavision, or other ‘Scope systems, (as opposed to merely being filmed with a Panavision camera with ordinary spherical lenses), is that full frame VHS copies can be made by “scanning and panning” the 1:2.35 ‘Scope frame. A pan made across the ‘Scope frame during the video transfer will show almost the true height of the frame, so creating no more of a problem than a video copy of an old Academy ratio film, but a scanning *cut* from one end of the ‘Scope frame to the other, which sometimes happens, introduces an apparent extra cut into the film which wasn’t there before. If there are a substantial number of these, this will affect the ASL slightly. Fortunately, the expert eye can detect most of them, but even I find I have missed some scanning cuts on re-examining a film. But to repeat, as long as you stick to DVD copies most of these difficulties can be avoided.

Warren Buckland:

[From: *Studying Contemporary American Film: A Guide to Movie Analysis*, by Thomas Elsaesser and Warren Buckland, pp. 101-16; bar charts omitted]

3.4. Statistical Style Analysis: Theory

The statistical style analysis of motion pictures is primarily a systematic version of *mise en scène* criticism – or, more accurately, *mise en shot* criticism. We have already seen that Eisenstein invented the term *mise en shot* to focus attention on the way shots are staged – that is, the way the parameters of the shot translate the actions and events into film. The advantage of statistical style analysis over *mise en scène/shot* criticism is that it offers a more detached, systematic, and explicit mode of analysis. Statistical style analysis characterizes style in a numerical, systematic manner – that is, it analyzes style by measuring and quantifying it. At its simplest, the process of measuring involves counting elements, or variables, that reflect a film's style, and then performing statistical tests on those variables.

More specifically, there are three standard aims of statistical style analysis: (1) to offer a quantitative analysis of style, usually for the purpose of recognizing patterns, a task now made feasible with the use of computer technology. In language texts, the quantitative analysis of style and pattern recognition is usually conducted in the numerical analysis of the following variables: word length, or syllables per word, sentence length, the distribution of parts of speech (the different percentage of nouns, pronouns, verbs, adjectives, and so on in a text), calculating the ratio of parts of speech (for example, the ratio of verbs to adjectives), or by analyzing word order, syntax, rhythm, or metre; (2) for the purposes of authorship attribution, in cases of disputed authorship of anonymous or pseudonymous texts (see Foster 2001); and (3) for purposes of identifying the chronology of works, when the sequence of composition is unknown or disputed (e.g., Plato, Shakespeare's plays).

The first aim, the quantitative analysis of style, involves descriptive statistics, and the second and third (authorship attribution and chronology) involve both descriptive and inferential statistics. As its name implies, descriptive statistics simply describes a text as it is, by measuring and quantifying it in terms of its numerical characteristics. The result is a detailed, internal, molecular description of a text's (or group of texts') formal variables. Inferential statistics then employs this formal description to make predictions. That is, it uses this data as an index, primarily an index of an author's style, or to put the author's work into chronological order on the basis of measured changes in style of their work over time. Whereas descriptive statistics produces data with complete certainty, inferential statistics is based on assumptions the statistician makes on the basis of the descriptive data. The assumptions the inferential statistician makes only have degrees of probability rather than certainty.

3.4.1. The quantitative analysis of style

One of the few film scholars to apply statistical style analysis to film is Barry Salt. In his essay 'Statistical Style Analysis of Motion Pictures' (Salt 1974), and later in his book *Film Style and Technology* (Salt 1992), Salt describes the individual style of directors by systematically collecting data on the formal parameters of their films.

Salt then represents the quantity and frequency of these formal parameters in bar graphs, percentages, and Average Shot Lengths (there will be more on these methods in section 3.5). When he compares and contrasts the form of the films of different directors, he moves into the realm of stylistic analysis. Style in this sense designates a set of measurable patterns that significantly deviate from contextual norms. As just one example, Barry Salt calculated that the average shot length of a film in the 1940s is around 9-10 seconds. A 1940s film with an average shot length of 30 seconds therefore significantly deviates from the norm, and is therefore a significant indicator of style.

3.4.2. Authorship attribution .

Authorship attribution is a long-standing, traditional subject in New Testament scholarship, study of the Classics, literary scholarship as well as in the legal context (for inferring whether the defendant wrote his or her confession, or whether it was 'co-authored' with the police, for example). Statistical style analysis has contributed its computerised statistical methods to these areas with controversial results.

One of the principles behind authorship attribution of written texts is that the stylometrist should not focus on a few unusual stylistic traits of a text, but on the frequency of common words an author uses – particularly minor or function words, whose use are independent on the subject matter or context. These include words such as prepositions (*of, to, in*) as well as synonymous function words such as *kind* vs *sort*, or *on* vs *upon*. One author may be prone to use *on* instead of *upon*, or *kind* rather than *sort*. (Stylometric analysts usually look for dozens of synonymous pairs in an author's work.)

At first it may seem odd to distinguish writing style by analyzing an author's consistent use of frequent function words, which he or she is not conscious of using. But as A.Q. Morton argues, these words offer the stylometrist a common point of comparison between authors: 'A test of authorship is some habit which is shared by all writers and is used by each at a personal rate, enabling his work to be distinguished from the works of other writers' (Morton, in Farringdon 1996: 274). So it is the quantity, or personal rate, of common words that is important, rather than their absence or presence in an author's writing. Furthermore, we can argue that a stylometric analysis is analogous to fingerprinting or to DNA testing. Humans share an enormous amount of DNA with other animals. It is only the minute details that distinguish humans from animals. Furthermore, human beings can be distinguished from each other on the basis of DNA testing or, more conventionally, on the basis of other small details – particularly fingerprints. One of the most common metaphors of stylometric authorship attribution is that it is fingerprinting authors. Anthony Kenny writes: 'What would a stylistic fingerprint be? It would be a feature of an author's style – a combination perhaps of very humble features such as the frequency of *such as* – no less unique to him [or her] than a bodily fingerprint is. Being a trivial and humble feature of style would be no objection to its use for identification purposes: the whorls and loops at the ends of our fingers are not valuable or striking parts of our bodily appearance' (Kenny 1982: 12-13).

A writer's style can therefore be measured in terms of a constant use of language

features, or a combination of features. Just one example, on Raymond Chandler:

Chandler 's style, like that of any author, consists of the *conjunction* of its constituent elements … . Much of the action and color in Chandler's stories is conveyed by dialogue, which comprises, on average, 44% of all the words in a story; for every thousand words of text, there are, on average, approximately 30 verbal exchanges, which last approximately 15 words apiece. For every thousand words of text, Chandler's stories also contain approximately one argot word, three similes, one vulgarity, no obscenities at all, and 38 coordinating injunctions. (Sigelman and Jacoby 1996: 19)

This information identifies Chandler's style – at least from a quantitative perspective, and can be used as the norm by which to attribute an anonymous story to Chandler.

If we think of the descriptive possibilities of stylometric authorship studies for film analysis, we note that, as with *mise en scène* criticism, statistics can be used to make auteur criticism more rigorous – that is, detached, systematic, and explicit. The auteur critic should then focus on the frequency of the common stylistic parameters a director uses – whose use are independent on the subject matter or context – rather than on a few unusual stylistic traits of a film. In other words, it is possible to use the descriptive dimension of authorship attribution to identify the series of invariant stylistic traits in a director's work (again, the traits linked to the parameters of the shot, in the first instance). It is imperative to think of a director's unique style in terms of the combination of all the parameters related to the shot (what statisticians call multivariate analysis).

The inferential dimension of authorship attribution has a more limited application to film, but some films such as *Poltergeist* have disputed authorship (was it directed by Tobe Hooper or Steven Spielberg?). By systematically analyzing the parameters of the shots in *Poltergeist*, and then comparing the results to samples from Hooper's and Spielberg's other films, it may be possible to identify the film's authorship (defined in terms of *mise en shot*, that is, the parameters of the shot). Of course, because we move from descriptive to inferential statistics, then the result can never be certain, but only predicted with a degree of probability. Only the descriptive aspect of the analysis remains beyond doubt.

On a cautionary note, the variables chosen to determine a director's style need to be valid (Salt has covered this problem by collecting data on the variables under a director's control). Secondly, the results need to be statistically significant, rather than due to chance occurrence. Many statistical tests are in fact tests for significance.

3.4.3. Chronology

The third area of statistical style analysis is chronology. Here again the statistics used can be either descriptive or inferential. A description quantifies and measures the changes in a body of work, usually of a single author. The point here is that an author's work changes in a predictable manner. An inferential study uses these descriptions of change to place an author's work into chronological order where that chronology is unknown or disputed. By identifying a pattern of change, and by measuring and quantifying that change, the author's work can then be put in

chronological order. An assumption behind inferential chronological studies is that an author's work is rectilinear, in other words, there is a linear progression in the change in an author's style. Furthermore, the idea of change needs to be reconciled with the idea of the author's style remaining constant in author attribution studies.

In film, chronology studies can be used descriptively to identify a change in style across a director's work. The most obvious example is charting the change of any shot parameter across a director's career, such as average shot length, distribution of shot scales, use of camera movement, and so on.

3.5. Statistical Style Analysis: Method

In his *Film Quarterly* essay 'Statistical Style Analysis of Motion Pictures' (Salt 1974), Barry Salt aimed to identify the individual style of a director by systematically collecting data on the formal parameters of films, particularly those formal parameters that are most directly under the director's control, including:

- duration of the shot (including the calculation of average shot length, or ASL)
- shot scale
- camera movement
- angle of shot
- strength of the cut (measured in terms of the spatio-temporal displacement from one shot to the next).

Salt collected data from these parameters by laboriously going through the film shot by shot. For most of his analyses, he in fact collected data on all the shots that appear in the first 30 minutes of each film, because this is a representative sample from the film. We shall employ (and test the viability of) this practice in our statistical style analysis of *The English Patient* in section 3.6. Salt is also interested in combining the results of each parameter. For example, he argues that it would be useful to combine 'duration of the shot' with 'shot scale' for each film (or indeed, a director's entire output), in order to determine 'the relative total times spent in each type of shot' (Salt 1974: 15), 'giving an indication of the director's preference for the use of that type of shot' (Salt 1974: 15). So, a director may use close-ups for a total of 20 minutes during a film, long shots for 30 minutes, and so on.

After analyzing a sample of films from four directors, Salt finds that both shot scale and ASL are significant and defining characteristics of a director's style. (Calculating the ASL involves dividing the duration of the film by the number of shots.) However, the distribution of shot scale is similar for the four directors he analyses.

In a statistical style analysis of Max Ophuls' films (Salt 1992, Chapter 22), Salt uses a standard stylometric tests to analyze the distribution of stylistic parameters in each film. Firstly, the histograms, or bar charts, representing the number of each shot type in each film (the number of close-ups, long shots, etc.). Secondly, he takes equal lengths of film, calculates the expected number of shots and shot types in each section, and then counts the actual number of shots and shot types in that section, to determine if they conform to the average (the mean) or deviate from it. There are several ways to select the equal section intervals:

1. Salt recommends intervals of one minute (i.e. 100ft intervals on 35mm film);
2. If calculating shot types one can define the intervals in terms of no. of shots (e.g. 50) and calculate the expected no. of shot types, and the actual no. of shot types;
3. Take the ASL of the whole film, and then analyze it scene by scene (each scene is defined in terms of spatio-temporal unity and in terms of events). Work out the expected no. of shots and shot types for each scene, and count the actual no. of shots. If the ASL is 10 seconds, and the scene lasts 2 minutes, the expected number of shots for that scene is 12.

In his analysis of *Letter From an Unknown Woman*, Salt notes the following:

For instance, in scene 1 five shots would be expected if the cutting were even throughout every part of the film, but in fact there are only three shots. Contrariwise, in scene no. 5, while only seven shots would be expected, there are actually fourteen. (Salt 1992: 309)

This type of analysis can also be applied to the expected no. of shot types in each scene and the actual no. of shot types. Salt's analysis of Ophuls' film *Caught* shows how this information can be useful in analyzing a film's style:

Caught is the first Max Ophuls film in which there is a very definite reduction in the amount of variation in Scale of Shot and cutting rate from scene to scene, and this becomes very apparent if a breakdown into 100ft sections is made on a 35mm. print. After the point in the film at which Leonora has married Smith-Ohlrig and been left alone in his mansion, we have for the next half hour of screen time very little departure from the average Scale of Shot distribution, and the cutting rate is also very steady for lengths of several minutes at a time, despite the occurrence of scenes of quite varied dramatic nature. It is only in the last 12 minutes of the film, when the most dramatic twitches of the plot take place, that there are any strong deviations from the norms. (Salt 1992: 310)

Salt is able to determine, not only how the shot lengths and scales are distributed across the whole film, but also how this film compares to Ophuls' other films ('Caught is the first Max Ophuls film in which there is a very definite reduction in the amount of variation in Scale of Shot and cutting rate from scene to scene'). Salt develops this historical analysis by considering Ophuls' later films, and notes that Ophuls pairs down variation in shot scale even more (relying more and more on the medium long shot), and using longer and longer takes, often combined with extensive camera movements.

For example, in *La Ronde*, with the scene between the Young man and The Chambermaid we get, after the first 11 shots, long strings of up to 10 shots each with the same camera distance in every shot. Most of these are also in the Medium or medium Long Shot scale, and the film continues in the same manner after this scene. At one point there is a string of 15 consecutive close ups, which is the sort of thing that just did not happen in other people's films in the same period, as a little checking

will show. (Salt 1992: 311)

In summary, statistical style analysis is a very precise and accurate tool for determining both the stability and the change in style that takes place across a filmmaker's career. Statistical style analysis focuses the research on how films are put together, rather than how they are perceived or comprehended.

Barry Salt carried out his statistical analysis by hand, which limited the types of tests he could perform on the data he collected. With the exponential growth in computer technology and software over the last decade, statistical style analysis can now be carried out using computer technology and powerful software programs. In the following analysis of *The English Patient*, data was still collected by hand, but it was then entered into the software program SPSS for Windows (Statistical Package for Social Scientists). SPSS is a spreadsheet program, with rows and columns. In film analysis, each row (which is automatically numbered) represents a shot, and each column represents a parameter of that shot. The parameters recorded include: shot scale, shot length, camera movement, direction of moving camera, and camera angle. Once the data has been entered, it can be represented both numerically and visually, and then numerous statistical tests performed on it.

The following analysis of *The English Patient* will consist of both the visual and numerical representation of data (particularly bar graphs, and frequency and percent tables). Then a few simple statistical tests will be applied: measure of the mean or average shot length; measure of the standard deviation of shot length; and the skewness of the values for shot length and shot scale. (The results will also be compared to a similar analysis of *Jurassic Park*.) The mean is a measure of central tendency, of the average value of a range of values. Standard deviation is the reverse of measuring the mean, for it is a measure of dispersion, or distribution-spread of values, around the mean; if the value of the standard deviation is large, this means that the values are widely distributed. Skewness measures the degree of non-symmetrical distribution of values around the mean. If the values are perfectly distributed, then the skewness value will be zero. If more of the values are clustered to the left of the mean (that is, if their value is less than the mean), then the distribution is positively skewed. If the values are clustered to the right of the mean, the distribution is negatively skewed.

These tests properly apply only to ratio data (where zero is an absolute value – zero weight, zero time, etc.). Only shot length is, strictly speaking, ratio data. In the shot scale, numbers have been assigned to the categories, which means they constitute a nominal scale (e.g., Very Long Shot is 7, but there is not reason why it couldn't be 1). However, by using the nominal scale consistently (1 = big close up, 2 = close up, 3 = medium close up, etc.) the norm, standard deviation, and skewness do at least have some heuristic value.

Other stylistic issues that can be raised (but won't be for this exercise) is to enter the number of scenes in the SPSS program, and then calculate the average number of shots per scene, and therefore calculate the expected number of shots per scene, and the actual number. Other useful data can be collected on: positional reference (for example, what position do close ups typically take in a film? – the first, second, third shot?) or contextual reference (do close ups usually follow long shots?). Percentiles

are also a useful tool. They measure the number of variables at regular intervals of a text. For example, at every five percent, count the number of variables (e.g., close ups) in the film. This will reveal if the variables are evenly distributed throughout the film, or concentrated in a particular part of it. One of the most interesting tests, however, is to determine the correlation between variables. For example, what is the correlation between shot length and shot scale? We would expect some correlation, because close ups usually appear on screen only for a short time, whereas a very long shot usually has a long duration on screen. But we can determine if there is a correlation between any of the variables – camera movement and shot length, or camera movement and shot scale, for example.

3.6 Statistical Style Analysis: *The English Patient*

Data was recorded from the following five parameters of the shot over the first 30 minutes of *The English Patient*: shot length, shot scale, camera movement, camera direction, and camera angle. For comparative purposes, the same data were recorded from the first 30 minutes of *Jurassic Park*. Barry Salt has already argued that 30 minutes is a representative sample to analyze. To test this hypothesis, we shall compare the results of the statistical style analysis of the first 30 minutes of *Jurassic Park* with the statistical style analysis of the whole film.

The statistical tests applied in this section to the collected data are the simplest ones available on SPSS: calculating the frequency of variables (that is, counting them), representing those frequencies as percentages, calculating the mean, the standard deviation, and the skewness of the results.

The first 30 minutes of *The English Patient* (up to the moment where Caravaggio introduces himself to Hana, and they go into the kitchen of the monastery) consists of 356 shots. In terms of shot length, the main values are to be found in Table 1.

The first column indicates shot length values (1 second, 2 seconds, and so on); the second column the number of times this shot length appears in the first 30 minutes of *The English Patient* (1 second shots appear 41 times, 2 second shots 84 times); and the third column indicates the percentage of shots with each value (1 second shots constitute 11.5 % of all the shots in the sample, while 2 second shots represent 23.6% of all the shots in the sample).

Table 1 only represents shots of length 1 to 10 seconds. There are additional values, up to 129 seconds (the opening credit sequence shot), but the frequency of shot lengths above 10 seconds is usually very small – one or two examples. Shots of length 1 to 10 seconds constitute 92% of all the shots in the sample.

Table 2 shows that the mean (the average) value of shot length of this sample is 5.1. In other words, the average shot length (ASL) of the film is 5 seconds (there is, on average, a cut every five seconds). The standard deviation of shot length is 8, indicating a wide dispersion of values around the mean, while the skewness of values is 10.97, indicating a very strong postive skewedness of values, favouring those values below the mean. What this means, in effect, is that there are a large number of shots in the range 1-4 seconds. All of this information can also be represented visually

(Figure 1).

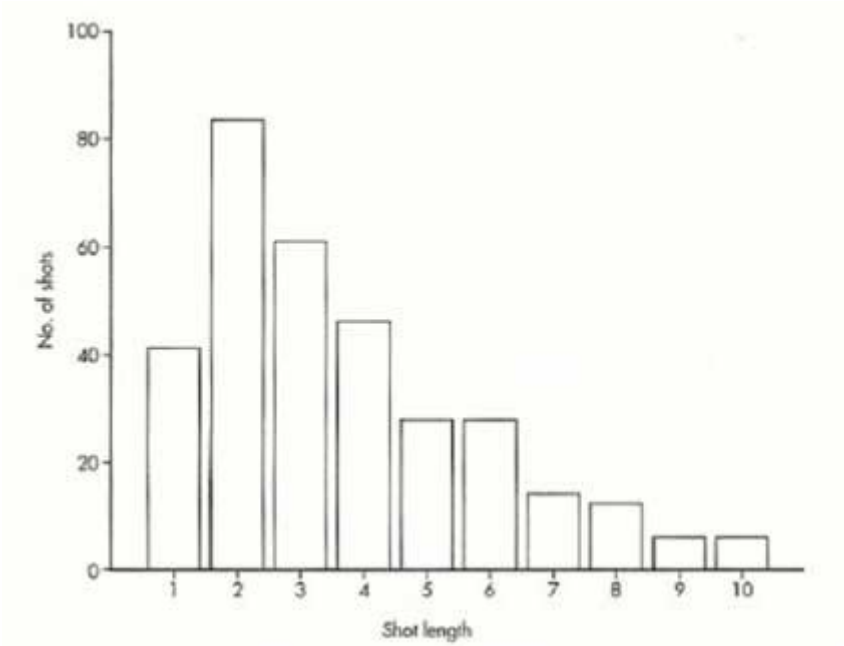


Figure 1. Shot length for the first 30 minutes of *The English Patient*

The value of this information may not be readily apparent. One of the best ways to make sense of it is to conduct a comparative analysis. The first 30 minutes of *Jurassic Park* (up to the end of the scene where Grant, Sattler, Malcolm, and Gennaro see a dinosaur egg hatch in the lab) consists of 252 shots, in comparison to *The English Patient*'s 356, a difference of 104 shots. This indicates that *The English Patient* has 40% more shots than *Jurassic Park*, a surprising result considering that *The English Patient* is a highbrow mega-movie imitating Art cinema aesthetics, while *Jurassic Park* is a blockbuster full of fast action.

We can make many other comparisons. *Jurassic Park*'s values for shot length can be found in Tables 3 and 4. The shot lengths in the range 1 to 10 seconds only constitute 80% of all the shots in the sample, suggesting that Spielberg's film has a wider variety of shot lengths. This is reflected in a skewness value of 2.68 (the mean value is 7 seconds and standard deviation is 6.69). Whereas the skew value of *The English Patient* is 10.97, in *Jurassic Park* it is only 2.68. This shows that the shot length values are more evenly distributed around the mean of 7. There is still a bias towards lower values (lower than the mean), but the bias is far smaller than in *The English Patient*. This information can also be represent visually, which makes the point more clearly (see Figure 2).

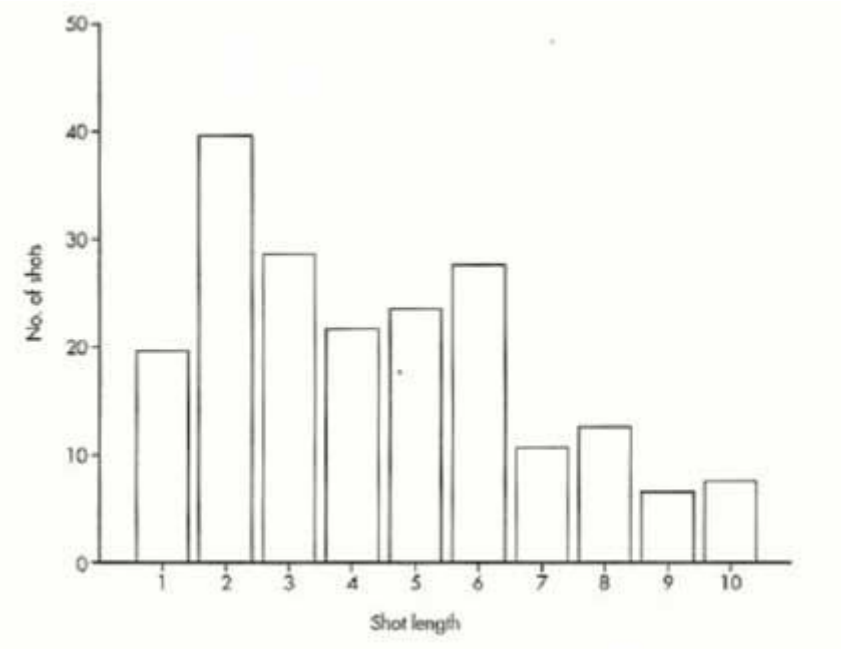


Figure 2. Shot length for the first 30 minutes of *Jurassic Park*.

We can explore this difference in shot length values further. In *The English Patient*, 52% of the shots fall in the range 1 to 3 seconds. In *Jurassic Park*, only 35% of the shots fall within this range. We have to include the values up to 5 seconds before *Jurassic Park* reaches the same percentage (in fact shots falling in the range 1 to 5 seconds constitute 54% of the film's total). However, by looking at the bar graphs, we can detect a similar pattern: a low value for 1 second, rising steeply for 2 seconds, and then falling gradually for the values 3 and 4 seconds. Furthermore, no shot length above 4 seconds in *The English Patient* and no shot length above 6 seconds in *Jurassic Park* constitute more than 10% of the total values. Whether these results only represent patterns common to *The English Patient* and *Jurassic Park*, are common in filmmaking, or are an anomaly will require further research.

With the above tests we are simply scratching the surface of what can be achieved with statistical style analysis. It is also possible to apply the same tests to the results obtained from the other four parameters of the shot. But because this would make the chapter even longer than it already is, we shall instead consider camera movement and shot scale. With the data collected on camera movement, we can test John Seale's claim that he avoids moving the camera unless absolutely necessary. The first 30 minutes of *The English Patient* contains the following values for camera movement:

	Frequency	Percent
still camera	302	84.8
pan	31	8.7
track	22	6.2
crane	1	0.3

The still camera is by far the most common value (85% of all shots), with only 15% of the shots containing camera movement. This seems to confirm John Seale's claim

that he likes to keep the camera still.

In comparison, *Jurassic Park* contains the following values for camera movement:

	Frequency	Percent
still camera	187	74.2
pan	33	13.1
track	26	10.3
crane	4	1.6
pan and track	2	0.8

These results may surprise some readers, especially the high percentage of still shots in an action blockbuster. But the percentages are significantly different to *The English Patient*, since *Jurassic Park* has 11% more moving shots than *The English Patient*.

Finally, in terms of shot scale, the distribution in both films confirms to what statisticians call a 'normal distribution', with high values in the middle (the mean) and progressively lower values on either side (see Figure 3). The result of these normal distributions is that the standard deviation and skewness values are low. Both directors favour medium close ups (28% in *Jurassic Park*, and 33% in *The English Patient*) and medium shots (21% in *Jurassic Park*, and 20% in *The English Patient*), although *Jurassic Park* only contains half as many close ups as *The English Patient* (9% in *Jurassic Park*, 18% in *The English Patient*). *Jurassic Park* compensates with almost three times as many long shots as *The English Patient*.

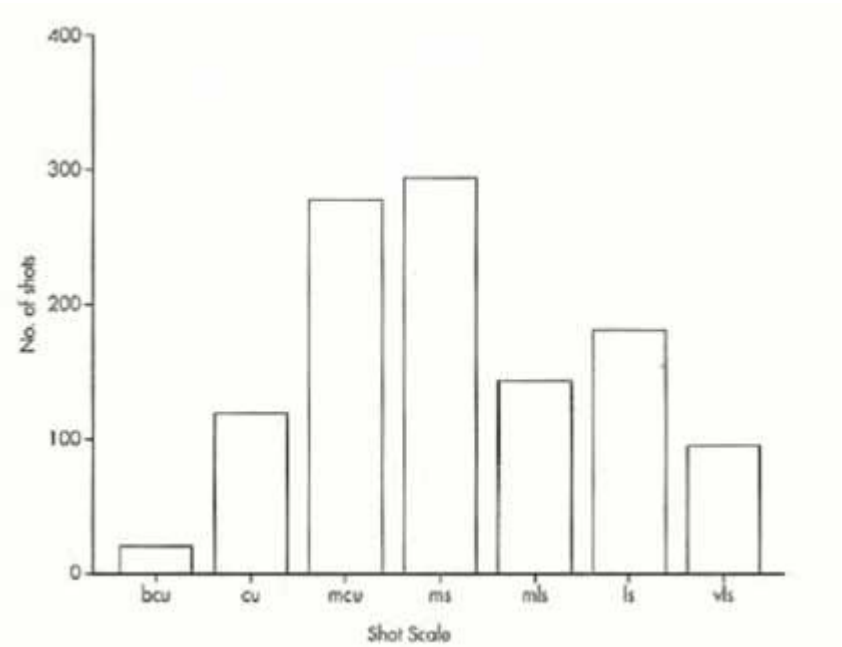


Figure 3. Distribution of shot scale in *Jurrasic Park* (whole film)
(bcu = big close-up; cu = close-up; mcu = medium close-up; ms = medium shot; mls = medium long shot; ls = long shot; vls = very long shot)

In summary, *The English Patient* contains a short range of shot lengths averaging out

at 5 seconds, heavily biased towards shots of 1-3 seconds, with a very high percentage of still shots. *Jurassic Park* has a much wider distribution of shot lengths, which average out at 7 seconds, with a bias (but not as much as in *The English Patient*) towards shots below this value, with a slightly more percentage of camera movement. For the record, 71% of shots in *The English Patient* are at eye level, compared to 81% in *Jurassic Park*. Furthermore 7% of shots in *The English Patient* are from a low angle, compared to 11.5% in *Jurassic Park*. This similarity is surprising, for Spielberg is well known for using low camera angles. The values for shot scale are more 'stable' in both films, and conform to the normal distribution of values.

One final task needs to be carried out to check the viability of the above results – the representative nature of the first 30 minutes of a film. Here we shall simply note major similarities and differences between a statistical style analysis of the first 30 minutes of *Jurassic Park*, and an analysis of the whole film. (When two figures are quoted, the first one always refers to the 30 minute sample and the second to the whole film.) Firstly, shot length. The mean for the first 30 minutes is 7 seconds (252 shots divided by 1800 seconds), whereas for the whole film it is 6 seconds (1145 shots divided by 6870 seconds), suggesting that the cutting rate increases as the film progresses. This increase in cutting is not surprising for an action film with its usual climatic ending, but what is surprising is that the increase is small. Standard deviation remains stable between the two samples, whereas skewness increases from 2.68 to 3.58, suggesting a increase in bias towards shots of shorter length in the whole film. And indeed, when we look at the percentage of 1 second shots, we note that, in the 30 minute sample, they constitute 8% of shots, whereas in the whole film, they constitute 14.5%. The other low values of shot length also increase slightly in the whole film. Whereas, as reported above, 54% of shots in the 30 minute sample fall between 1 and 5 seconds, in the whole film 54% of shots fall between 1 to 4 seconds. Put another way, shots between 1 and 5 seconds in the whole film constitute 63% of shots (as opposed to 54% in the 30 minute sample). Shot scale remains almost identical in both samples, as does camera movement (surprisingly, the number of still shots only falls 1% to 73% in the whole film, despite the increase in action). Significantly, the percentage of low camera angles almost doubles when we take into consideration the whole film – from 11.5% to 21%.

The information that the SPSS software has yielded is simply the raw material for writing about the style of *The English Patient*, and for comparing its style to the style of other films. The above analysis only presents a small sample of data and even fewer tests on the stylistic patterns to be found in the film. The primary difference between this analysis and more conventional *mise en scène* analysis is that statistical style analysis is more systematic and rigorous, and is more narrowly focused, for it exclusively analyzes shot parameters. When reading the results of a statistical style analysis, we need to keep in mind that both the computer and statistics are merely tools, means to an end to analyzing data on style, a way of quantifying style and making the recognition of underlying patterns easier.

References

Farringdon, Jill (1996), *Analysing for Authorship: A Guide to the Cusum Technique* (Cardiff: University of Wales Press).

Foster, Don (2001), *Author Unknown: On the Trail of Anonymous* (London: Macmillan).

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

Salt, Barry (1974), 'The Statistical Style Analysis of Motion Pictures' *Film Quarterly*, 28, 1: 13-22.

____ (1992), *Film Style and Technology: History and Analysis* (London: Starword).

Sigelman, Lee, and William Jacoby (1996), 'The Not-So-Simple Art of Imitation: Pastiche, Literary Style, and Raymond Chandler', *Computers and the Humanities* 30, 1: 11-28.