

Chapter 1

Introduction



What value is “criticism” which consists merely of opinions without reference to facts? (Joseph McBride 1971, 32).

When experiment is pushed into new domains, we must be prepared for new facts, of an entirely different character from those of our former experience (P.W. Bridgman 1927, 2).

Every 10 years, *Sight & Sound* magazine carries out a survey to compile a list of critically acclaimed films. For the 2012 poll, the magazine’s editor explained that “we approached more than 1,000 critics, programmers, academics, distributors, writers and other cinephiles, and received (in time for the deadline) precisely 846 top-ten lists that between them mention a total of 2,045 different films” (James 2021). From 1962 to 2002, *Citizen Kane* (1941) topped *Sight & Sound*’s list, although in 2012 it came second to Hitchcock’s *Vertigo* (1958) and fell to third place in 2022. Despite the apparent frivolity in compiling a top ten, a list presents a straightforward way to arrange data (such as film titles) in a descending order of rank (determined by the number of mentions), and 846 responses constitute a sufficiently large data set to represent meaningful results. These polls confirm *Citizen Kane*’s status as one of the most critically acclaimed films in cinema history. The making of *Citizen Kane* has even been the subject of two feature films: *RKO 281* (Benjamin Ross, 1999) and *Mank* (David Fincher, 2020), while Orson Welles—the producer, director, and cowriter of *Citizen Kane* (as well as its star)—was and still is a celebrated public figure.

However, Welles’s cowriter credit on *Citizen Kane* has generated a long-running public dispute, between eminent critics and industry insiders such as Pauline Kael and actor John Houseman on one side, who argue that it is the sole work of Herman J. Mankiewicz, and equally eminent critics and industry insiders such as Andrew Sarris and Peter Bogdanovich on the other, who argue that Welles collaborated fully in the writing process. Mankiewicz’s biographer, Richard Meryman, sums up the dispute: “the authorship of *Citizen Kane* has become one of film history’s major controversies. And the question of who did what and how much opens up an extraordinary subdrama of jostling egos” (Meryman 1978, 237). The arguments for and against Welles’s coauthorship credit are merely asserted, driven by opinions, impressions, loyalties among friends, and feuds between enemies. Houseman maintained a 40-year feud with Welles, taking every opportunity to (quite literally)

discredit Welles's coauthorship of the *Citizen Kane* screenplay. Bogdanovich noted that, in his memoir *Run-Through*, Houseman (1972, 468) took his revenge on Welles via a strategic rhetorical move—Houseman goes out of his way to praise Welles as a director before proclaiming that Welles did not write any of *Citizen Kane*, a move that makes Houseman's account sound sincere, fair, and balanced (Bogdanovich 1972, 101). Houseman's revenge is matched by Kael's hostility—not only toward Welles but also toward Andrew Sarris, whose auteur theory canonized directors such as Welles: “The auteur critics are so enthralled with their narcissistic male fantasies,” Kael wrote, “that they seem unable to relinquish their schoolboy notions of human experience” (Kael 1963, 26). To counter Sarris's veneration, Kael glibly remarked that Welles “was to become perhaps the greatest loser in Hollywood history” (Kael et al. 1971, 46). Here, Kael combines uncertainty (*perhaps*) with hyperbole (*the greatest loser*). But by what criteria could Welles be defined as a loser? (Incompetence? Lack of success?) Kael uses the term (and boosts it with a superlative adjective) merely as an emotionally laden derogatory slur. But Sarris's and Bogdanovich's attempts to canonize Welles are equally unreserved. Bogdanovich begins his defense of Welles by quoting in passing the quintessential European film auteur Jean-Luc Godard, who said that “All of us will always owe [Welles] everything” (Godard, quoted in Bogdanovich 1972, 99). With *all of us*, *always*, and *everything* packed into one concise sentence, Bogdanovich could not have found a more hyperbolic statement to counter Kael's rhetoric. Bogdanovich then set out to discredit both Houseman and Kael, claiming, for example, that Houseman's career was undistinguished and that his only claim to fame was his brief association with Welles. Houseman, Bogdanovich argues, only gets himself noticed by continually repeating that Welles is a fraud: “For many years now, Houseman has been actively promoting the picture of Welles as a credit thief” (Bogdanovich 1972, 190). I untangle these feuds in Chap. 2, and in the following chapters, I bypass such hearsay and rumor and instead tackle the public dispute from a new perspective, one that subjects the language of the *Citizen Kane* screenplay to close scrutiny via statistical analysis.

In this study, I set out to answer two questions:

- (1) What distinguishes the writing of Mankiewicz from Welles?
- (2) What did each author contribute to the writing of the *Citizen Kane* screenplay?

This study has no vested interest in privileging one author over the other; instead, it is driven by curiosity and the unknown and is guided by a discovery procedure that aims to find the answers to these two research questions by employing data collection and statistical analysis. I aim to persuade film study scholars that it is only via statistical methods that we can resolve the long-running dispute over the authorship of the *Citizen Kane* screenplay.

Like all statistical analyses, the analysis guiding this study is an inquiry into unknown parameters: the distinctive features of Mankiewicz's writing, the distinctive features of Welles's writing, and each author's contribution to the *Citizen Kane* screenplay. These unknowns are reduced through the collection, measurement, classification, and statistical analysis of relevant data. This study puts to one side ad hoc

judgments and instead provides internal (textual) evidence of authorship, which supplements traditional methods of authorship attribution such as external historical evidence (discussed in Chap. 2). Moreover, this study examines the internal authorship of the *Citizen Kane* screenplay with unusual and special types of linguistic evidence not previously considered—new facts of an entirely different character (to paraphrase Bridgman). This evidence becomes accessible via statistical theories and methods that quantify those specific linguistic features that separate Mankiewicz from Welles, features that are then identified in the *Citizen Kane* screenplay. The simplest statistical techniques from the recent development of *new statistics* are employed to discover and quantify each author's contribution to *Citizen Kane*.¹ Quantifying entails translating a text's linguistic properties into numerical values, which can then be measured via frequency counts and transformed using other statistical operations (calculating the mean, standard deviation, percentages, ratios, confidence intervals, and effect sizes). My motivation behind this study involves replacing impressionistic accounts of Mankiewicz's and Welles's contributions to the *Citizen Kane* screenplay with a rigorous statistical analysis.

Attributing authorship to an anonymous or disputed text by quantifying stylistic features requires statistical inferences, for authorship is not an empirical property of a text. In other words, authorship cannot be measured directly and explicitly, for it is an indirect and implicit property that emerges from direct properties such as word frequencies. In logical terms, attributing authorship is not deductively entailed from a text's empirical properties but is nondemonstrative, an attribution supported but not demonstrated or proven by textual properties. A radical skeptic such as David Hume would deny the veracity of making a nondemonstrative or ampliative inference, for an inference of authorship extends beyond and is (always) underdetermined by the available data and cannot, therefore, be logically validated or justified. This type of skepticism raises metatheoretical issues, such as: Is there any type of data that can count as evidence in a statistical analysis of authorship? How does a statistical analysis validate the data it generates? And how do statistics manage indirect evidence and uncertainty? Statistics relies on a set of inductive reasoning processes that are explicitly defined and shared, which makes them secure, reliable, and robust. Statistics contains and controls uncertainty and rejects Hume's skepticism by turning a problem into an empirically testable research question by collecting, measuring, and quantifying an enormous amount of data and by identifying consistency (that is, patterns and trends) in that data. Statistical testing then draws inferences from that data, which provides evidence for the empirical research question. Within a statistical framework, linguistic features become independent variables that predict authorship. Controversially, authorship is reduced to a constant quantity, a *discriminant numerical value* comprising a linear combination of variables that maximizes the difference between authors. The likely success of such a study increases when the linguistic data are sufficiently comprehensive, when the distinctive linguistic variables are constant (or, at least, when their variation can be

¹ See, for example, Geoff Cumming (2014) and Cumming and Calin-Jageman (2017).

measured), when the data are drawn from only one genre of writing (the screenplay) and two authors (Mankiewicz and Welles), and when the study employs multiple statistical methods to corroborate and cross-validate the results (which reduces variability in the data and the bias of a single method).

In this study, I define authorship in a literal way—in terms of a writer’s organization and the wording of a written text, which in turn produces that author’s style. Such a study assumes that writing is an orderly process that can be quantified and analyzed in precise terms. More specifically, an author’s style emerges from their recurring habits, a systematic series of linguistic choices that make their writing distinctive. (As we shall see, these choices are not necessarily made consciously, which is why we prefer the term *habit*.) Studying these linguistic habits from a statistical perspective takes place on several levels. For Alvar Ellegård (whose statistical authorship method is central to this study), the term *style* is synonymous with “features or combinations of features in an author’s way of writing” (Ellegård 1962, 9). Ellegård defines style as a distinctive combination of features that emerge from an author’s systematic set of linguistic habits. Similarly, Fiona Tweedie and her colleagues define style as “a set of measurable patterns which may be unique to an author,” adding that:

Almost every conceivable measure has been considered, ranging from sentence lengths to the number of nouns, articles or pronouns occurring in the text. The vocabulary of the author has also undergone scrutiny, with counts being taken of words that occur only once in the text, to the most common words that act as fillers. Between these two extremes are function words, certain non-contextual words occurring in the text. They can be used as “markers” for different authors (Tweedie et al. 1996, 1).

And in his definition of style, N. E. Enkvist compares frequencies:

Style is concerned with frequencies of linguistic items in a given context, and thus with contextual probabilities. To measure the style of a passage, the frequencies of its linguistic items of different levels must be compared with the corresponding features in another text or corpus which is regarded as a norm and which has a definite relationship with this passage (Enkvist 1964, 29).

In simple terms, statistical analysis identifies an author’s distinctive style by measuring and then quantifying a vast array of linguistic features via frequency counts. An author’s quantified stylistic features become meaningful only when compared to the quantified style of other authors. Style is therefore defined comparatively, as the quantifiable deviation of one author’s style from the style of other authors. Deviation is measured and quantified in terms of the frequency counts of letters, words, sentence lengths, the number of nouns, articles, pronouns, etc. Such a precise and measured analysis of style can assist us in identifying the author of a disputed text. In this study, a statistically based analysis of the authorship of the *Citizen Kane* screenplay identifies the discriminant variables, the most relevant or significant linguistic features that distinguish the style of Mankiewicz from Welles, for it is from these discriminant variables that inferences are generated to identify each author’s contribution to the screenplay. However, statistics cannot banish uncertainty but manages and reduces it to a measurable and knowable quantity.

Because this study uses statistics to measure and quantify writing habits, it forms part of the discipline of stylometry (the quantification of style)—more specifically, to a branch of stylometry devoted to authorship attribution, which in the past has examined in unprecedented detail the disputed authorship of classical texts (Plato, Aristotle), the New Testament, as well as Shakespeare’s plays, plus *De Imitatione Christi*, *The Letters of Junius*, *The Federalist* papers,² and numerous other disputed works, some of which have made international headlines—such as Don Foster’s successful unmasking of Joe Klein as the author of the anonymous novel *Primary Colors* and his mistaken attribution of the poem “A Funeral Elegy” to Shakespeare (Foster 2001). Like these previous studies in stylometry (outlined in Chap. 4), in this study, I employ descriptive statistical methods to quantify linguistic features to infer authorship. In the last 40 years, stylometric authorship attribution has been bolstered with the advent of computing and, more recently, with software tools, an integral part of which involves representing the results in tables and graphs. It is via these statistical procedures, software tools, and visual representation of information that we can discover new data relevant to determining the authorship of the *Citizen Kane* screenplay.

This type of research is evidence based. It avoids speculation, overgeneralization, and impressionistic judgments and is interdisciplinary, forming part of the Digital Humanities, which challenges the traditional ways of thinking embedded in the humanities. Digital Humanities research employs statistical solutions to seemingly intractable humanities problems. The present study introduces simple statistical methods to an arts and humanities readership and then demonstrates the value of those methods by carrying out a systematic study of the long-running coauthorship problem that has puzzled the film industry and film critics for decades.

Chapter 2, *The Trials of Coauthorship*, investigates the dispute between Herman J. Mankiewicz and Orson Welles, focusing on Welles’s claim to the status of coauthor of the *Citizen Kane* screenplay, and attempts by his adversaries to deny him this status, which (they argue) he appropriated unfairly and deceptively. I frame my discussion of this apparent case of modern-day pseudepigrapha (false ascription of authorship to a written text) via two institutions: the current guidelines from the Writers Guild of America (WGA) and the current copyright law formulated by the United States legislature. For both institutions, authorship is premised on self-contained individualism—on writing as a solitary and individual act of creation that produces an original text that (in Aristotle’s formulation) reflects the writer’s character, thereby conferring on the writer the right to claim the ownership of that text. The chapter also examines the assumptions behind the concept of coauthorship, which complicates authorship attribution by challenging the commonplace idea that a written text has a single origin (the mind of one author). Coauthorship involves sharing different writing tasks and roles, such as outlining, planning, drafting, editing, and revising. Chapter 2 ends by discussing Robert Carringer’s authoritative

²Standard overviews of stylometric authorship attribution include Susan Hockey (1980, Chapter 6); David I. Holmes (1994); Patrick Juola (2008); Efstathios Stamatatos (2009); and Michael P. Oakes (2014).

study of the (co)authorship of seven different versions of the *Citizen Kane* screenplay (Carringer 1978).

Chapter 3, Screenplays: Words on a Page, presents the control set of screenplays—screenplays of known authorship, which will be analyzed (in Chap. 5) to determine an effective set of linguistic variables that distinguish the writing of Mankiewicz from Welles. The control set consists of two screenplays known to be written by Mankiewicz (*Made in Heaven* (1943–45) and *A Woman's Secret* (1949)) and two known to be written by Welles (*The Other Side of the Wind* (1970) and *The Big Brass Ring* (1982)). Each screenplay sample is 20,000 words long—the first 20,000 words of each screenplay, minus the character cues and other standard formatting marks, as explained in the chapter. Each sample is therefore 40,000 words, which constitutes a control corpus that serves to identify stylistic features that distinguish Mankiewicz from Welles. To assist in this effort, data from other screenplays are presented in passing, including *The Magnificent Ambersons* (Welles, 1941), *Touch of Evil* (Welles, 1957), and *Man of the World* (Mankiewicz, 1931). For comparative purposes, data from two other screenplays not written by Mankiewicz or Welles are included: *His Girl Friday* (1940, written by Charles Lederer) and *All the President's Men* (1974, written by William Goldman). Mankiewicz's sample consists of an adapted screenplay—*A Woman's Secret*. It will be compared to its source material (the novel *Mortgage on Life* by Vicki Baum) using a software program called WCopyfind,³ which evaluates two documents by matching the overlapping words and phrases. This will establish how similar the screenplay is to the adapted source material.

The first half of Chap. 4, The Statistical Analysis of Style: Aims and Methods, begins by outlining three fundamental distinctions central to statistics: descriptive and inferential statistics, sample and population, and statistical tests and effect sizes. I then present stylometry's basic premises to arts and humanities scholars: namely, that relevant linguistic features should ideally be quantifiable, high rate, context-free (not dependent on the subject matter), multiple, subconscious, distinctive, and stable (consistent and regular). In the second half of the chapter, I present an overview of stylometric methods that have previously been successful in discriminating between authors and the range of data they used—including punctuation, unigrams, contractions, vocabulary analysis (word frequency profile, the type/token ratio, distinctiveness ratio), collocational analysis, and frequency distribution of sentence length.

The authorship attribution process comprises a training phase, followed by a testing phase. Chapter 5, Distinguishing Mankiewicz from Welles: Training Phase Results, presents the training phase, where the stylometric methods introduced in Chap. 4 are applied to the control group of screenplays to establish which methods are successful in identifying and quantifying the stylistic features that distinguish Mankiewicz from Welles. Several statistical concepts and methods proved useful: relative frequencies, the distinctiveness ratio, sample means, sample standard

³<https://plagiarism.bloomfieldmedia.com/software/wcopyfind/>

deviation, confidence intervals, and effect size. In traditional statistical terms, *relative frequency* quantifies the frequency of variables within the same sample.⁴ In contrast, the *distinctiveness ratio* compares the relative frequency of the same variable in two different samples. Distinctive variables are identified by dividing a variable's relative frequency in one sample into the same variable's relative frequency in the other sample.⁵ The higher the ratio, the more distinctive the variable. The *sample mean* estimates the population mean, the *confidence interval* represents variation around that estimated mean, and the *effect size* measures the scale or magnitude of the difference between two samples. In the following study, and with frequent reference to Ellegård's theory of authorship attribution, these statistical methods are employed to construct models or *statistical profiles* of Mankiewicz's and Welles's writing styles: relative frequency identifies variables (linguistic features) that have a high or low frequency in Mankiewicz's and Welles's writings; the distinctiveness ratio compares Mankiewicz's and Welles's relative frequencies in order to identify the most distinctive linguistic features that distinguish the two authors; confidence interval represents the range or variation that linguistic features can take around the mean of an author's statistical profile, with the upper and lower limits corresponding to the boundaries of that profile; and effect size measures the size of the separation between the two authors.

From this training phase, I draw up a list of distinctive linguistic features that distinguish Mankiewicz from Welles. An initial list of 77 distinctive linguistic features and the final list of 44 features are presented in the Appendix to Chap. 5. These distinctive features are divided into two groups—the plus group represents 22 variables distinctive of Mankiewicz (in relation to Welles), and the minus group represents 22 variables distinctive of Welles (in relation to Mankiewicz).

In Chap. 6, Comparing Mankiewicz and Welles to the *Citizen Kane* Screenplay (1), I employ Mankiewicz's and Welles's statistical profiles to assign authorship in a precise way to the *Citizen Kane* screenplay. Whereas in Chap. 5 I distinguish Mankiewicz from Welles using 44 distinctive linguistic features, in Chap. 6 I employ the same features to establish the *similarities* between each author and the *Citizen Kane* screenplay. It is in this chapter that I discover the quantity of writing that Mankiewicz and Welles contributed to *Citizen Kane*. I analyze the screenplay three times: firstly, as a single document; secondly, by segmenting it into 4000-word samples; and thirdly, by dividing it into its 13 major scenes. On each occasion, I compare the *Citizen Kane* screenplay to Mankiewicz's and Welles's statistical profiles. In other words, the 22 distinctive linguistic features of Mankiewicz's statistical profile and the 22 distinctive features of Welles's profile are counted in the *Citizen Kane* screenplay and converted into relative frequencies; these *Citizen Kane* relative frequencies are then divided into Mankiewicz's and Welles's relative frequencies to calculate their distinctiveness ratio. If the ratio is small, this signifies similar

⁴Relative frequency of x = the observed frequency of x in the sample divided by the total number of words in that sample.

⁵The distinctiveness ratio = relative frequency of x in sample₁ / relative frequency of x in sample₂.

authorship; if the ratio is large, it suggests different authorship, where similar/different authorship is defined in terms of ratios and the boundaries of the confidence interval. I identify the exact sections in the screenplay where Welles's stylistic signature dominates and where Mankiewicz's signature dominates. I use these results to revise and update some of the conclusions in Carringer's authoritative study by promoting a hypothesis that he did not consider: whether Welles wrote part of the *Citizen Kane* screenplay *before* Mankiewicz began writing the first draft.

In Chap. 7, Comparing Mankiewicz and Welles to the *Citizen Kane* Screenplay (2), I examine sentence length, clusters, and type/token ratios and use the software program Linguistic Inquiry and Word Count (LIWC) to compare Mankiewicz and Welles to *Citizen Kane*. LIWC measures and quantifies 92 linguistic features of texts, including grammatical categories such as pronouns, verbs, and function words, together with punctuation, informal expressions, and words expressing positive or negative sentiment. The chapter also compares two screenplays not written by either author (the two mentioned above—*His Girl Friday* and *All the President's Men*) to their statistical profiles to ensure the statistical tests do not falsely attribute these screenplays to Mankiewicz or Welles, and I analyze other texts known to be written by Mankiewicz and Welles to see how well they match each author's profiles.

The conclusion considers how nondemonstrative inferences of authorship can be evaluated and, more generally, reflects on the role and limits of statistics in solving problems in Arts and Humanities research.

Rather than rely on point estimates (such as the p values of null hypothesis significance testing) to determine the relevance and importance of the results, this study follows the new statistics by employing confidence intervals and magnitude or effect sizes, which are particularly appropriate for the nondemonstrative inferences generated in the Arts and Humanities. The difference between two data samples (e.g., Mankiewicz and Welles) is quantified in terms of degrees (a range of values located within an interval) and effect size (a ratio value, measured using *Cohen's d*),⁶ which replaces the single all-or-nothing p value. An estimate determines *to what extent* the data support an inference rather than stating that the data either support or does not support the inference.

By employing methods that have become central to the new statistics in his study of *The Letters of Junius* in 1962, Ellegård was ahead of his time. His methods are also straightforward to understand and sufficiently powerful to offer a precise solution to the authorship of the *Citizen Kane* screenplay.

⁶Confidence intervals and ratios are the same in the way they quantify differences. Whereas a ratio measures the difference between two samples numerically, a confidence interval represents that ratio visually. A ratio of one, for example (no difference between two samples), is located in the center of an interval. *Cohen's d* measures the difference between two samples in terms of standard deviation. A *Cohen's d* value of 0.5 quantifies the difference between two samples as 0.5 standard deviations, while a *Cohen's d* value of 0 signifies identity between the two samples (as does a ratio of 1). *Cohen's d* is defined more formally in Chap. 4.

References

- Bogdanovich, Peter. 1972. The Kane Mutiny. *Esquire* (October): 99–105; 180–190.
- Bridgman, P.W. 1927. *The Logic of Modern Physics*. New York: The Macmillan Co.
- Carringer, Robert L. 1978. The Scripts of *Citizen Kane*. *Critical Inquiry* 5 (2): 369–400.
- Cumming, Geoff. 2014. The New Statistics: Why and How. *Psychological Science*, 25 (1): 7–29.
- Cumming, Geoff, and Robert Calin-Jageman. 2017. *Introduction to the New Statistics: Estimation, Open Science, and Beyond*. New York: Routledge.
- Ellegård, Alvar. 1962. *A Statistical Method for Determining Authorship: The Junius Letters 1769-1772*. Gothenburg: Gothenburg Studies in English.
- Enkvist, N. E. 1964. On Defining Style. In N. E. Enkvist, J. Spencer & M. J. Gregory, *Linguistics and Style*, 1–56. Oxford: Oxford University Press.
- Foster, Don. 2001. *Authorship Unknown: On the Trail of Anonymous*. London: Macmillan.
- Hockey, Susan. 1980. *A Guide to Computer Applications in the Humanities*. London: Duckworth.
- Holmes, David I. 1994. Authorship Attribution. *Computers and the Humanities* 28 (2): 87–106.
- Houseman, John. 1972. *Run-Through 1902/1941*. New York: Touchstone.
- James, Nick. 2021. How We Made the Greatest Films of All Time Poll: <https://www.bfi.org.uk/sight-and-sound/polls/greatest-films-all-time/introduction>
- Juola, Patrick. 2008. Authorship Attribution. *Foundations and Trends in Information Retrieval* 1 (3): 233–334.
- Kael, Pauline. 1963. Circles and Squares. *Film Quarterly* 16 (3): 12–26.
- Kael, Pauline, Herman Mankiewicz, and Orson Welles. 1971. *The Citizen Kane Book*. Boston: Little, Brown.
- McBride, Joseph. 1971. Rough Sledding with Pauline Kael. *Film Heritage* 7 (1): 13–16; 32.
- Meryman, Richard. 1978. *Mank: The Wit, World, and Life of Herman Mankiewicz*. New York: William Morrow.
- Oakes, Michael P. 2014. *Literary Detective Work on the Computer*. Amsterdam: John Benjamins Publishing Company.
- Stamatatos, Efstathios. 2009. A Survey of Modern Authorship Attribution Methods. *Journal of the American Society for Information Science and Technology* 60 (3): 538–56.
- Tweedie, F. J., S. Singh, and D. I. Holmes. 1996. Neural Network Applications in Stylometry: The *Federalist Papers*. *Computers and the Humanities* 30 (1): 1–10.