STYLOMETRIC ANALYSIS
OF THE SHERLOCK HOLMES CANON

John Allen, 2018
(Last Revision 14 May 2018)

ABSTRACT

I describe a database designed specifically to compare the function word frequencies of the Sherlock Holmes adventures with the function word frequencies of Arthur Conan Doyle's non-Holmes fictional works. The results are startling.

Based on the function word frequency comparisons, Arthur co-authored *A Study in Scarlet* and wrote two of the short stories. The remaining Holmes adventures were written by either of two ghostwriters.

One of the ghostwriters, almost certainly Arthur's first wife, Louise, co-authored *A Study in Scarlet* and wrote each of the other early adventures.

The second of the ghostwriters, almost certainly Arthur's second wife, Jean, wrote all but two of the later adventures.

Arthur wrote most of the non-Holmes fictional works that have long been attributed to him, though Louise and Jean each wrote some of them.

I provide two independent measures of the method's accuracy, the two methods agreeing that the author prediction accuracy is equal to or greater than 97%.

I detail my methodology and provide a sample calculation, thus to allow others to replicate my effort and test my conclusions.

TABLE OF CONTENTS

LIST OF APPENDICES

CHAPTER 1
Background

In *The Boscombe Valley Mystery*, Sherlock Holmes cautioned Dr. Watson that "there is nothing more deceptive than an obvious fact." The obvious fact to be considered herein is that the Sherlock Holmes adventures flowed from the mind and pen of Arthur Conan Doyle.

The methodology employed for this study, author discrimination based on function word frequencies, is derived from the work of Frederick Mosteller and David Wallace, founders of the modern era of stylometric analysis. I patterned my work after their analysis as described in their 1963 seminal paper "Inference in an Authorship Problem."

The analysis detailed herein reveals that 1) Arthur wrote most of the non-Holmes fiction attributed to him, 2) a ghostwriter was responsible for the early Holmes adventures, and 3) a second ghostwriter was responsible for the later Sherlock Holmes adventures.

1.1 Suspicions

Early in the twentieth century, Bertram Fletcher Robinson, a journalist and author, claimed that he co-authored *The Hound of the Baskervilles*. More specifically, he claimed that he "wrote most of the first instalment for the *Strand*." Later, Robinson wrote his own detective stories and bylined them with "Joint Author with Sir Arthur Conan Doyle in his Best Sherlock Holmes Story *The Hound of the Baskervilles*." Robinson seems to be the first person to explicitly claim that Arthur was not the sole author of the Sherlock Holmes adventures.

In 1972, Sherlockian scholar D. Martin Dakin, in his *Sherlock Holmes Commentary*, made obvious his suspicions about the authorship of the last series of Holmes adventures: "This, the last series of Holmesian adventures demands a general consideration [...] since serious questions both of authorship and authenticity arise." [Dakin 249] Dakin argued that nine of the twelve stories may have been written by someone other than whomever wrote the early stories.

In 1981, Martin Gardner, skeptic and writer of popular works regarding math and science, became the first person to unambiguously pronounce to a mass reading audience that Arthur Conan Doyle neither created Sherlock Holmes nor wrote the adventures. Gardner made his case in his essay "The Irrelevance of Conan Doyle," which he included in his book *Science: Good, Bad, and Bogus*. Focusing on Arthur's belief in the paranormal, Gardner argued that Arthur was simply too gullible to have created any character so rational and skeptical as Sherlock Holmes. [Gardner 113-122]

In 1994, Sherlockian scholar W. W. Robson, in his introduction to *The Case-Book of Sherlock Holmes* (an anthology containing the last twelve adventures), declared that "it is hard to believe that any careful reader of the *Case-Book* would be prepared to testify that they all came from the pen of Conan Doyle." Robson thereby joined the small chorus of Sherlock Holmes enthusiasts suggesting that Arthur put his name to at least some adventures that he did not write. Robson even suggested a means for uncovering the truth. "We may wonder what would be the result of stylometric analysis

of the Holmes stories, the kind of treatment that has been given to the Platonic Dialogues, the plays of Shakespeare, or the Epistles of St. Paul." [Robson xvii-xviii]

1.2 My Involvement

Recently, I wrote my own stylometric analysis program to investigate the Holmes authorship issue. Based on historical, biological, and literary analyses, I had already concluded that Arthur's first wife, Louise Conan Doyle née Hawkins, was much more likely to have been Holmes's creator and the author of his early adventures. Recognizing the disparity in quality between the early and late adventures, I willingly conceded the late stories to Arthur.

I wrote my stylometric analysis program as a means of checking my conclusions. I was admittedly pleased and surprised by the early results of that analysis. Even now, after many modifications and enhancements to the program, the results remain fundamentally unchanged.

My earlier conclusion (and my self-admitted bias) that Louise wrote the early Holmes adventures is well supported by my stylometric analysis.

My earlier suspicion (and my self-admitted expectation) that Arthur wrote the later adventures is not, on the other hand, supported by my stylometric analysis. Instead, the analysis reveals a third distinct writing style, one found almost exclusively in the later Holmes adventures. For reasons to be described elsewhere in this paper, I now attribute the later Holmes adventures to Arthur's second wife, Jean Conan Doyle née Leckie.

Most unsettling, my stylometric analysis did not provide definitive results regarding *A Study in Scarlet*, the very first Sherlock Holmes adventure. Presuming the author of that first adventure is the actual creator of Sherlock Holmes, my stylometric analysis neither confirms nor refutes my early conclusion that it was Louise, rather than Arthur, who created Sherlock Holmes. I therefore use a chapter of this book to describe the mystery surrounding the first Holmes adventure and to present a more nuanced conclusion regarding who might have been Holmes's creator.

In 2017, I published *Shadow Woman: The Real Creator of Sherlock Holmes*. In that book, I use a combination of historical, biographical, and literary analyses to make my case the Louise created Sherlock Holmes and wrote the early adventures. As one portion of a single chapter in that book, I provided an overview of my stylometric analysis to buttress my more conventional arguments. [Allen 120-127] In the work before you, I describe my stylometric method and results in considerably more detail. I provide sufficient information herein that anyone with reasonable programming skills can check, recreate, and challenge my work.

CHAPTER 2
Structure of the Current Study

2.1 Mosteller and Wallace

I suggest that modern stylometric analysis began with the work of statisticians Frederick Mosteller and David Wallace. In their 1963 paper "Inference in an Authorship Problem," they described how they used discriminator word frequencies to determine authorship of the disputed *Federalist* papers. A year later, they expanded their description, considerably, in their book *Inference and Disputed Authorship: The Federalist*. Unless otherwise noted, references to Mosteller and Wallace are to their paper rather than their book.

Alexander Hamilton and James Madison each claimed credit for twelve of the papers. From the introduction to Mosteller and Wallace's paper:

> The *Federalist* papers were published anonymously in 1787-1788 by Alexander Hamilton, John Jay, and James Madison to persuade the citizens of New York to ratify the Constitution. Of the 77 essays, 900 to 3500 words in length, that appeared in the newspapers, it is generally agreed that Jay wrote five: Nos. 2, 3, 4, 5, and 65, leaving no further problem about Jay's share. Hamilton is identified as the author of 43 papers, Madison of 14. The authorship of 12 papers (Nos. 49-58, 62, and 63) is in dispute between Hamilton and Madison; finally, there are three joint papers, Nos. 18, 19, and 20, where the issue is the extent of each man's contribution.

Mosteller and Wallace then described various scholarly efforts to establish the authorship based on non-statistical methods, taking special note of the work of historian Douglass Adair, before explaining the root cause of the authorship dispute.

> No doubt the only reason the dispute exists is that Hamilton and Madison did not hurry to enter their claims. By the time those lists attributed to Hamilton were published, he was dead, and thereafter Madison waited a decade to make his own claim. Probably Hamilton and Madison were slow to claim their share in *The Federalist* because the arguments each gave there sometimes opposed political positions each adopted later.

Also in their introduction, Mosteller and Wallace discussed the challenge of discriminating between the writing styles of the two candidate authors.

> The writings of Hamilton and Madison are difficult to tell apart because both authors were masters of the popular *Spectator* style of writing—complicated and oratorical. To illustrate, in 1941 Frederick Williams and Frederick Mosteller counted sentence lengths for the undisputed papers and got means of 34.55 and 34.59 words respectively for Hamilton and Madison, and average standard deviations for papers of 19.2 and 20.3. These results show that for some measures the authors are practically twins.

In their conclusions, Mosteller and Wallace resolved the authorship dispute almost entirely in favor of Madison.

Our data independently supplement that of the historian. On the basis of our data alone, Madison is extremely likely, in the sense of degree of belief, to have written the disputed *Federalists*, with the possible exception of No. 55, and there our evidence is weak, suitable deflated odds are 80 to 1 for Madison. No. 56, next weakest, is a strong 800 to 1 for Madison. The data are strong for all the rest, including the two papers the historians feel weakest about, Nos. 62 and 63.

In both the paper and their book, Mosteller and Wallace provided sufficiently detailed explanation of their work that even someone with modest statistical and programming skills, such as myself, could adapt their methodology to an entirely different authorship dispute. Without reservation, I used their work as a model for my own.

2.2 The Text Collections

Mosteller and Wallace based their work on three collections of texts. One collection consisted of Hamilton's undisputed writings, another of Madison's undisputed writings, and the third of the twelve disputed papers.

As did Mosteller and Wallace, I also relied on three collections of text. For simplicity and clarity, and in light of the results, I now refer to my collections as Arthur, Louise, and Jean.

2.2.1 The Arthur Collection

The Arthur collection consists of 110 non-Sherlockian short stories, each of them fictional, each bearing Arthur's name as the author. The word counts range from 4,022 to 17,917 words. The first publication dates range from 1879 to 1930, a span of more than half a century.

Due to concern about temporal variation of an author's function word usage rates, I divided the Arthur collection into three sub-collections which I refer to as early-Arthur, intermediate-Arthur, and late-Arthur.

The early-Arthur collection, detailed in Appendix I, includes the 42 short stories from the Arthur collection that were first published before the end of 1893. The early-Arthur collection is intended to best correspond with the early Holmes adventures, those adventures consisting of *A Study in Scarlet*, *The Sign of Four*, and the first 24 short stories.

The intermediate-Arthur collection, detailed in Appendix II, includes the 48 short stories from the Arthur collection that were first published between the beginning of 1894 and 4 July 1906, when Arthur's first wife, Louise, died. The intermediate-Arthur collection is intended to best correspond with the intermediate Holmes adventures, those adventures consisting of *The Hound of the Baskervilles* and the short stories eventually collected in the anthology *The Return of Sherlock Holmes*.

The late-Arthur collection, detailed in Appendix III, includes the 20 short stories from the Arthur collection that were published after 4 July 1906. The late-Arthur collection is intended to best

correspond with the late Holmes adventures, those adventures consisting of *The Valley of Fear* and the Holmes short stories published after Louise's death.

Because the hypothesis being investigated is that Arthur put his name to stories that he did not write, I could not presume that each of these 110 short stories in the Arthur collection were, in fact, the work of Arthur Conan Doyle. I did presume, however, that the function word frequencies of each Arthur sub-collection, taken as a whole, are representative of Arthur's usage during the corresponding time period.

The 110 works of fiction in the early, intermediate and late-Arthur collections are listed in Appendices I through III, respectively.

2.2.2 The Louise Collection

The Louise collection consists of the first 24 Sherlock Holmes short stories, beginning with *A Scandal in Bohemia* in 1891 and ending with *The Final Problem* in 1893. The stories have word counts ranging from 6,499 to 12,718 words. I did not presume that all of these 24 short stories were written by the same person. I did, however, presume that the function word frequencies of the collection, taken as a whole, are representative of a single author, whether or not that author was Arthur.

The basis for the Louise collection is historical. Arthur Conan Doyle married Louise Hawkins in August of 1885. Eight months later, as if a month premature, Sherlock Holmes was born in the manuscript for *A Study in Scarlet*. Louise mentioned the story in a letter to Arthur's sister: "Arthur has written another little book, a little novel about 200 pages long, called 'A Study in Scarlet.' It went off last night." [Lellenberg 247]

Study was first published in the 1887 issue of *Beeton's Christmas Annual*. That first Sherlock Holmes adventure was followed by a second Holmes novel, *The Sign of Four*, first published in the February 1890 issue of *Lippincott's*. Those two novels were followed by the 24 Sherlock Holmes short stories that I chose to use as the Louise collection. In the last of those short stories, Holmes flees to Switzerland, pursued by Professor Moriarty, engaging him finally in mortal combat on a trail overlooking the Reichenbach Falls. Both Holmes and Moriarty plunge into the falls, and the Sherlock Holmes adventures presumably come to an end.

I suspect it is not coincidence that Holmes traveled to Switzerland to meet his fate at the same time that Louise traveled there to meet hers. According to Arthur's autobiography, Louise had recently been diagnosed with advanced tuberculosis. [Conan Doyle 120] Others report that she was given but three months to live. Having recently been in Switzerland, indeed peering into the very Reichenbach Falls where Holmes was to die, Louise returned to Switzerland, hoping that the cold, clean, thin air of the Swiss Alps might save her. Despite his suggestion to the contrary, Arthur did not travel with his supposedly terminally-ill wife. Instead, he remained in England to conduct a lecture tour. To his lecture agent, Gerald Christy, he wrote:

My wife has fallen ill, and has had to go to Davos. Of course, I shall let no private matter—however urgent—interfere with my engagements.

He signed his unfaithful letter: "Yours faithfully, A Conan Doyle." [Christie's 55]

It is tantalizing that Holmes was born eight months after Louise married Arthur and then died in Switzerland just as Louise traveled there to face her own mortality. Indeed, it was this realization that prompted me to envision Louise as a viable author of the early Holmes adventures, presuming of course that Arthur did not write them himself.

The 24 short stories in the Louise collection are listed in Appendix IV.

2.2.3 The Jean Collection

The Jean collection consists of the last 19 Sherlock Holmes short stories, beginning with *The Adventure of Wisteria Lodge* in 1908 and ending with *The Adventure of Shoscombe Old Place* in 1927. These stories have word counts ranging from 4,496 to 11,458. I did not presume that all of these 19 short stories were written by the same person. I did, however, presume that the function word frequencies of the collection, taken as a whole, are representative of a single author, not necessarily the same author, or authors, as that of the Arthur and Louise collections.

The basis of the Jean collection is also historical. As it turns out, neither Holmes nor Louise died in Switzerland. Rather than plummet into the falls, Holmes had hidden himself on a ledge so that the world would believe he had perished along with Moriarty, thus to deceive Moriarty's henchmen, thus allowing Holmes to deal with them one at a time. We learn of his miraculous survival in *The Adventure of the Empty House*, published ten years after the tragic story told in *The Final Problem*.

Louise also survived, 13 years longer than expected, for reasons unexplained. For several years, Arthur joined Louise intermittently in Switzerland, then the two of them wintered once in Cairo before moving back to England. There, Arthur began (or perhaps continued) his (supposedly platonic) affair with Jean Leckie, she being more than 15 years younger than Arthur and more than 17 years younger than the woman she would supplant.

During the time that Arthur was married to Louise but in love with Jean, Sherlock Holmes appeared in one novel, *The Hound of the Baskervilles*, and 13 short stories. These intermediate stories could have, from a temporal standpoint, been just as easily written by Arthur, Louise, or Jean. They intermediate short stories are therefore not included in either the Louise or Jean collections.

Louise finally died of her tuberculosis on the 4 July 1906. Sherlock Holmes thereafter appeared in one more novel, *The Valley of Fear*, and 19 more short stories. Jean is therefore the most likely author of the late Holmes adventures, presuming, of course, that Arthur did not write them himself.

The 19 short stories in the Jean collection are listed in Appendix V.

2.2.4 Works Outside of the Collections

The Louise and Jean collections exclude the 4 Sherlock Holmes novels because, by design, the collections consist only of short stories ranging from 4,000 to 20,000 words in length.

The Louise and Jean collections also exclude the 13 short stories collected in the intermediate anthology *The Return of Sherlock Holmes*, since I had no historical basis for concluding that those short stories might have been written a single author. Arthur was married to Louise but dating Jean during the period when the stories appeared.

The 4 Holmes novels and 13 Holmes short stories that were evaluated as part of this study but excluded from both the Louise and Jean collections are listed in Appendix VI.

Taken together, the 24 Holmes adventures in the Louise collection, the 19 Holmes adventures in the Jean collection, the 13 uncollected Holmes short stories, and the 4 uncollected Holmes novels constitute all 60 stories in what is widely (and almost religiously) known as the Sherlock Holmes Canon. Each of the 60 Holmes adventures was assumed, for the purpose of this study, to be a disputed text. Only those Holmes adventures in the Louise collection, however, contributed to the function word frequencies presumed to be representative of the primary author of the Louise collection. Similarly, only those Holmes adventures in the Jean collection contributed to the function word frequencies presumed to be representative of the primary author of the Jean collection.

In addition to the 60 Holmes adventures and the 110 short stories in the Arthur collection, 18 non-Holmes novels and 37 non-Holmes short stories long attributed to Arthur Conan Doyle were also evaluated as part of this study. The 18 non-Holmes novels were excluded from the Arthur collection because their word counts exceeded 20,000. The 37 non-Holmes short stories were excluded from the Arthur collection because their word counts were less than 4,000. Those 55 non-Holmes works that were evaluated in this study but were excluded from the Arthur collection are listed in Appendix VII.

In summary, 225 stories in this study, all attributed to Arthur Conan Doyle, were evaluated as disputed texts; 110 were the short stories in the Arthur collection; 24 were the short stories in the Louise collection; 19 were the short stories in the Jean collection; 17 were Holmes adventures excluded from both the Louise and Jean collections; and 55 were non-Holmes novels and short stories excluded from the Arthur collection.

2.3 The Function Words

Mosteller and Wallace spent a fair portion of their paper describing how they settled on the discriminator words that they used for frequency comparisons among their three text collections. They eventually settled on 20: *although, as, at, by, commonly, consequently, considerable, destruction, enough, innovation, language, of, on, there, upon, vigor, voice, while, whilst,* and *would*.

The list includes both content words and function words. To explain the difference between the two categories of words, Mosteller and Wallace referred to the work of Miller, Newman, and Friedman as described in their 1958 paper "Length-Frequency Statistics for Written English."

> We will call these two classes the "function words" and the "content words." Function words include those which are traditionally called articles, prepositions, pronouns, numbers, conjunctions, and auxiliary verbs, plus certain irregular forms. The function words have rather specific syntactic functions which must, by and large, be known individually to the speaker of English. The content words include those which are traditionally called nouns, verbs, and adjectives, plus most of the adverbs. It is relatively easy to add new content words to a language, but the set of function words is much more resistant to innovations.

To better understand the important difference between function words and content words, consider, for example, a classic typing exercise: *The quick brown fox jumped over the lazy dog.* The sentence works as a typing exercise because it contains at least one instance of each letter in the alphabet. It is of passing interest here because it consists primarily of content words. Removing its three function words makes the sentence only slightly ambiguous: *Quick brown fox jumped lazy dog*.

One of Shakespeare's most famous lines, on the other hand, is a grand mass of function words: *To be or not to be, that is the question*. Removing the function words leaves but a one-word sentence with zero content: *Question*.

In *The Secret Life of Pronouns* (2013), linguist James W. Pennebaker estimates that the average English speaker has a vocabulary of 100,000 unique words. The overwhelming majority of those are content words: fewer than 500 are function words. The lowly function words, nonetheless, constitute more than half of the words actually uttered or typed. Everyone uses them in distinctive fashion, or so the stylometric argument goes. Authors supposedly leave them behind, unconsciously, like fingerprints on a knife.

In their paper, Miller, Newman, and Friedman provide their list of 363 function words. Mosteller and Wallace found that 12 of them provided the best discrimination between Hamilton and Madison: *although*, *as*, *at*, *by*, *enough*, *of*, *on*, *there*, *upon*, *while, whilst*, and *would*. They found that one function word in particular, *upon*, was particularly useful in discriminating between the two authors.

In the known writings of Hamilton, *upon* appears at the rate of 32 times per 10,000 words. In the known writings of Madison, the word appears barely at all, only twice per 10,000 words. The rarity of that word in the disputed documents therefore weighed heavily in Madison's favor.

Usually, an author's fondness for (or avoidance of) any single function word is substantially less noticeable than Hamilton's fondness for (or Madison's avoidance of) *upon*. Mosteller and Wallace were nonetheless quite taken with the utility of the function words in general, though they did offer a cautionary note in their paper.

We were surprised that in the end, it was the utterly mundane high-frequency function words that did the best job. Though we love them for their lack of contextuality, their final strength was as unexpected as it was welcome. [...]

The function words of a language appear to be a fertile source of discriminators, and luckily the high frequency words are the strongest. [...]

Pronouns and auxiliary verbs appear to be dangerously contextual; other function words are not entirely safe and should not be taken for granted, but should be investigated for contextuality.

For this study, I compiled the most frequently used function words from the Arthur collection, eliminated the personal pronouns that provide gender context (such as *he*, *she*, and *herself*), eliminated the auxiliary verbs that provide temporal context (such as *is* and *was*), then arbitrarily limited my discriminator word list to the 100 most frequently used function words remaining.

I did not test any of my function words beforehand to determine which of them might have been the best discriminators. Unlike Mosteller and Wallace, I used no content words in my list of discriminator words. My word list can therefore be described with equal accuracy as my function word list or my discriminator word list.

The 100 discriminator words used in this study are listed in Appendix VIII.

2.4 A Scoring Overview

Given a disputed text, a discriminator word, and two reference text collections, the next step in the stylometric analysis process is to determine, via a single scoring equation, whether the discriminator word rate in the disputed text is closer to the discriminator word frequency of the first or the second reference text collection.

Mosteller and Wallace devised their scoring equation (to be detailed in the next chapter) such that a "large" numerical result indicated Hamilton was the more likely author, while a "small" numerical result indicated Madison was the more likely author. Of their scoring algorithm, they wrote in their paper:

The weights are chosen so that when Hamilton is the author and the rates apply to him, the value [...] tends to be large (say), and when Madison is the author, the value small.

Again using the discriminator word *upon* as an example, Mosteller and Wallace's scoring equation returned a "small" value for the frequency of *upon* in every disputed text when weighed against the *upon* frequencies from each of the Hamilton and Madison reference text collections. For every disputed text, the infrequent use of the word *upon* provided a piece of evidence that Madison was in fact the author of the disputed text.

As statisticians, however, Mosteller and Wallace would never rely on a single word, even if that word was seemingly so revealing as *upon*. Instead they applied their scoring equation to each of their 20

discriminator words for each of the 12 disputed texts. For each disputed text, they summed their 20 discriminator word scores to arrive at a single disputed text score. If the text score met their rigorous definition of "large," then they attributed the disputed text to Hamilton. Otherwise, they attributed the disputed text to Madison.

For this study, I modified Mosteller and Wallace's word score equation such that a positive score pointed to one author while a negative score pointed to the alternative author. (I leave the details for the next chapter.) As did Mosteller and Wallace, I sum my discriminator word scores to obtain a disputed text score.

Rather than use 20 function and content words as discriminator words, the current study used 100 function words as discriminator words. Rather than evaluate 12 disputed texts, the current study evaluated 225. Rather than discriminate between two possible authors, the current study discriminated among three.

To discriminate among three authors, the current study relied on three separate sets of calculations. First, the disputed text was tested against the Louise collection and the period-appropriate Arthur sub-collection, then against the Jean collection and the period-appropriate Arthur sub-collection, and finally against the Louise and Jean collections. Only then could the calculated author be ascertained. The final result is presented in the format of the three examples below.

Sir Nigel; Arthur (AL=A257 JL=L50 AJ=A256)
The Yellow Face; **Louise** (AL=L319 JL=L246 AJ=J107)
The Second Stain; **Jean** (AL=L69 JL=J317 AJ=J321)

A discussion of the format reveals how the analysis discriminates among the three potential authors, presumably Arthur, Louise, and Jean.

Each result is presented on a single and separate line. The title of the fictional work appears first. Long titles are sometimes abbreviated. For works reprinted under different titles, I use the title from the earliest publication.

Next is the name of the calculated author. When the calculated author is listed as Arthur, the indication is that the story was most likely written by the person most responsible for writing the Arthur collection of stories, presumably Arthur himself. When the calculated author is listed as Louise, the indication is that the story was most likely written by the person most responsible for writing the Louise collection of stories. When the calculated author is listed as Jean, the indication is that the story was most likely written by the person most responsible for writing the Jean collection of stories.

When the calculated author is different than Arthur, the name of the calculated author is emboldened.

Each of the 225 disputed texts was analyzed three times to discriminate between two different pairs of possible authors. The first analysis indicated whether the story was more likely written by Arthur

or Louise, and the resulting score is on the right side of the "AL=" equation. The second analysis indicated whether the story was more likely written by Louise or Jean, and the resulting score is on the right side of the "JL=" equation. The third analysis indicated whether the story was more likely written by Arthur or Jean, and the resulting score is on the right side of the "AJ=" equation.

Rather than show the scores as either positive or negative numbers, I display each of them as a letter followed by the absolute value of the text score, with the letter indicating the more likely author of the head-to-head comparison.

There is no theoretical upper limit to the absolute value of the score. The highest score was 966, though few scores are over 500. Though it is difficult to equate a specific score with a specific confidence level, scores larger than 100 seem worthy of considerable confidence while scores less than 25 should be treated with considerable caution.

Since any one author is always considered during two, and only two, of the three possible head-to-head comparisons, the calculated author is the person deemed more likely in both of her head-to-head comparisons.

With respect to the non-Holmes novel *Sir Nigel*, Arthur is calculated to more likely be the author than either Louise or Jean, as shown by the "AL=" and "AJ=" equations. The Jean-versus-Louise comparison is thereby rendered inconsequential with respect to determining who among the three may have actually written the story. Of the entire set of 225 works of fiction attributed to Arthur, the analysis indicates that he wrote only 140 of them.

With respect to Sherlock Holmes mystery *The Adventure of the Yellow Face*, Louise is calculated to more likely be the author than either Arthur or Jean, as shown by the "AL=" and "JL=" equations. For the purpose of determining the person most likely to be the author, the Arthur-versus-Jean comparison is thereby rendered inconsequential. Of the 225 works of fiction attributed to Arthur, the analysis indicates that the primary author of the Louise collection wrote 40 of them.

With respect to the Sherlock Holmes mystery *The Adventure of the Second Stain*, Jean is calculated to more likely be the author than either Arthur or Louise, as shown by the "JL=" and "AL=" equations. The Arthur-versus-Louise comparison is thereby rendered inconsequential, at least with respect to determining the most likely author of the story. Of the 225 works of fiction attributed to Arthur, the analysis indicates that the primary author of the Jean collection wrote 44 of them.

The author of one non-Holmes work of fiction was calculated to be Ambiguous, meaning the program could not discriminate among Arthur, Louise, and Jean.

CHAPTER 3
Scoring Details

This chapter provides a mathematical discussion that is likely to be of interest only to those readers intent on understanding the details of the word scoring equations, perhaps with the intention of reproducing the work or writing their own program. The more casual reader should skip directly to Chapter 4.

3.1 Scoring Details of the Mosteller and Wallace Study

Though they used different nomenclature and terminology than used herein, Mosteller and Wallace calculated the numerator of their Word Score as (fH - fM) x fD, where fH represents the frequency of the discriminator word in the Hamilton text collection, fM represents the frequency of the discriminator word in the Madison collection, fD is the frequency of the discriminator word in the disputed text, and x is used as the multiplication symbol.

For the disputed text, there is no substantial dispute with respect to the proper frequency to be used in the Word Score calculation; fD is simply the number of occurrences of the discriminator word within the text divided by the total number of words within the text.

Since the reference text collections consist of multiple texts, and since each text within the collection can (and probably does) have a different frequency within the text, there are multiple ways in which the discriminator word frequency for the entire text collection might be determined. Mosteller and Wallace used either of two possible methods: "[...] we use median rates in the screening set for each author unless one or both of the medians is zero, and then the average rate for each author is used."

The median value is frequently described as the middle value. In the collection of numbers {1, 3, 3, 6, 7, 7, 8) the median value is 6 while the average value is 5. Statisticians frequently prefer the median value because it is not so easily skewed by an outlier value. Consider now the similar collection of numbers {1, 3, 3, 6, 7, 7, 85}. The 85 is an outlier, probably a data error rather than a representative value within the collection. In this exaggerated example, the median is unchanged; it is still the middle value, 6. The average value, however, balloons from 5 to 16. The median value is more likely representative.

Consider now the collection of numbers {0, 0, 0, 0, 3, 5, 6}. The median value is 0 and the average value is 2. In such a case, Mosteller and Wallace would eschew the median value and use instead the average value from both of the reference text collections.

To gain a sense of how Mosteller and Wallace's Word Score numerator worked, consider their discriminator word *upon*. Recall that Hamilton used that word 32 times per thousand words while Madison used it only twice per thousand words. Assuming a disputed text in which *upon* was used at a rate of 20 times per thousand words, Mosteller and Wallace's Word Score numerator would look something like (32 - 2) x 20, which equals 600, which is a "large" score that Mosteller and Wallace would presumably deem indicative of Hamilton. For comparison, consider a disputed text in

which *upon* never appears. The Word Score numerator would look something like (32 - 2) x 0, which equals zero, which is presumably a "small" score indicative of Madison.

For several good reasons, the Word Score consists of both a numerator and a denominator. Proper selection of the denominator can make the Word Score non-dimensional, which is convenient, and less responsive to an author's natural variation in word frequency patterns from one text to the next. For their Word Score denominator, Mosteller and Wallace used a denominator equal to rH^2 + rM^2, where rH is equal to the range of the discriminator word frequencies found within the Hamilton text collection, and rM is equal to the range of the discriminator word frequencies found within the Madison collection.

If, for example, the frequency of *upon* within the texts in the Hamilton collection ranged from 26 to 41 occurrences per thousand words, then rH^2 would be (41-26) x (41-26) = 225. (The units would be in occurrences per thousand words, squared, identical to the units of the numerator if the variables therein had been expressed in occurrences per thousand words. Assuming consistent units in the numerator and denominator, the Word Score will always be non-dimensional.

If we simply assume that Madison always used the word *upon* at the rate of 2 occurrences per thousand words, then the rM^2 would be equal to zero, the Word Score denominator would be 225, and the Word Score would be reduced accordingly. If our hypothetical Hamilton, however, had been more consistent with his usage of the word *upon*, then the frequencies in his collection might have ranged from 27 to 37 occurrences per thousand words, his rH^2 would be equal to 100 (rather than 225), and the absolute value of the Word Score would have been more than twice as large. It thereby becomes clear that discriminator words used at consistent rates throughout a reference text collection will tend to have greater influence in determining the calculated author.

In their paper, Mosteller and Wallace explained why they used the square of the frequency ranges rather than square of the standard deviations (also known as the variances).

> To replace the variances we used squares of ranges (range is largest minus smallest rate). Our reasons for using ranges were to cut down on total amount of calculation and to defend against contextual words. Recall that ranges are somewhat more affected by outliers than standard deviations, and the more contextual the word, the more wildly its rates vary.

I now present, without further discussion, the scoring algorithm successfully employed by Frederick Mosteller and David Wallace. Only the nomenclature is mine.

> Word Score = (fH - fM) x fD / (rH^2 + rM^2)
> Disputed Text Score = Sum(Word Scores)
> If Disputed Word Score "large", then author was Hamilton
> If Disputed Word Score not "large", then author was Madison

3.2 Scoring Details of the Current Study

My biggest concern with the Mosteller and Wallace algorithm is its reliance on "large" or "small" values in determining authorship. In their paper, they offer little insight into how one might categorize scores as either large or small.

> The weights are chosen so that when Hamilton is the author and the rates apply to him, the value [...] tends to be large (say), and when Madison is the author, the value small.

This lack of specificity introduces a tantalizing pathway for confirmation bias. If the stylometrist preferred one author over another, it would be easy to rationalize making "large" just a bit larger, knowing all the while that such a decision would push the outcome in the preferred direction.

Also deviating from Mosteller and Wallace, I chose to use neither the median value nor the average value for determining the discriminator word frequency of a reference text collection. For ease of computation, and for intuitive comfort, I used the weighted average of the word frequencies across the collection. That weighted average value was easily calculated by dividing all the occurrences of the discriminator word within the entire collection by the total number of words within the entire collection. Total occurrences divided by total words equals the weighted average frequency of any word within any collection of texts.

I manage the concern over possible outliers in a collection by means other than equating median values with text collection frequency values. In the current study, every story within a text collection is treated also as a disputed text. If a story within a reference collection is a significant outlier with respect to a particular significant discriminator word, then the outlier story will tend to stand out from others in the reference collection when that outlier story is evaluated as a disputed text.

Using a comparison between the Arthur and Louise collections as an example, the numerator of the Word Scoring equation for the current study is $(fD - fL)^2 - (fD - fA)^2$, where fD is the frequency of the discriminator word in the disputed text, fL is the frequency of the discriminator word in the Louise collection of texts, and fA is the frequency of the discriminator word in the Arthur collection texts. If the frequency in the disputed text is closer to the frequency in the Arthur collection than to the frequency in the Louise collection, then the result will be positive. If, on the other hand, the frequency in the disputed text is closer to the frequency in the Louise collection than the frequency in the Arthur collection, then the result will be negative.

To sense how the numerator works, assume the following frequency set: {fD, fA, fL} = {6, 9, 5}. In that case, the frequency of the disputed text would be closer to the frequency of the Louise collection, and the numerator would be correspondingly negative: $(6 - 5)^2 - (6 - 9)^2 = -8$.

Now, for contrast, assume a slightly different frequency set: {fD, fA, fL} = {6, 9, 11}. The frequency of the disputed text would be closer to the frequency of the Arthur collection, and the result would be correspondingly positive: $(6 - 11)^2 - (6 - 9)^2 = 16$.

For comparisons between Arthur and Jean, the numerator is unsurprisingly similar: $(fD - fJ)^2 - (fD - fA)^2$.

For comparisons between Louise and Jean, the numerator is $(fD - fL)^2 - (fD - fJ)^2$.

The numerators are structured such that a positive result always points to the author of the Arthur collection as the likely author of the disputed text. The numerators are structured such that a negative result always points to the author of the Louise collection as the likely author of the disputed text.

In part because of computing power issues, Mosteller and Wallace used frequency ranges (squared) in their denominator. Having substantially more computing power at my fingertips, I used standard deviations (squared) in the denominators. Rather than $rH^2 + rM^2$ of the Mosteller and Wallace study, the current study uses $vL + vA$ (for the Arthur versus Louise comparison, as an example) where vL is equal to the population variance of the frequencies within the Louise collection, and vA is equal to the population variance of the frequencies within the Arthur collection. By definition, the population variances are equal to the square of the population standard deviations.

The denominator for the current study performs the same functions as did the denominator for the Mosteller and Wallace study. It causes the resulting Word Score to be non-dimensional, and it rewards those collections that have reasonably consistent discriminator word frequencies across their texts.

To allow the results to be reported without a decimal point, all Word Scores for this study have been multiplied by ten. The resulting Disputed Text Scores will therefore range from several hundred negative to several hundred positive.

This current formulation removes a source of potential bias by eliminating the subjective decision of what constitutes a "large" or a "small" number. In the current formulation, the author of collection A is indicated whenever the word frequency from the disputed text is closer to the frequency in collection A than to the frequency in collection B, and vice versa.

Finally, the discriminator word scoring equation used in this study (for an Arthur-to-Louise comparison) is:

$$\text{Word Score} = 10 \times [(fD - fL)^2 - (fD - fA)^2] / (vL + vA)$$

A sample Disputed Text Score calculation is provided in Appendix IX, for *A Scandal in Bohemia*, discriminating between Arthur and Louise as possible authors.

CHAPTER 4
Accuracy

The accuracy of the current study has been assessed in three different fashions: re-evaluation of *The Federalist* papers; frequency of ambiguous results; and frequency of chronologically troublesome results. The result of the three assessments, taken both individually and as a whole, is that the stylometric analysis described herein accurately discriminates among Arthur, Louise, and Jean as possible authors of the 225 works of fiction long attributed to Arthur Conan Doyle alone.

4.1 Re-Evaluation of the Federalist Papers

To test the discriminator word list and the scoring equations used by the current stylometric analysis software, I adapted the current method to evaluate *The Federalist* papers.

Instead of using Mosteller and Wallace's 20 function and content words as discriminator words, I used my list of 100 function words as discriminator words.

Instead of using the Arthur, Louise, and/or Jean reference collections, I used a Hamilton reference collection comprised entirely of his 52 undisputed *Federalist* papers and a Madison reference collection comprised entirely of his 14 undisputed *Federalist* papers.

Instead of evaluating 225 works of fiction allegedly written by Arthur Conan Doyle, each of which I treated as a disputed text, I evaluated 81 *Federalist* papers published anonymously under the pen name Publius, plus 9 other papers (allegedly by Hamilton) published anonymously under the pen name Pacificus, plus 5 other papers (generally attributed to Madison) published anonymously under the pen name Helvidius. I treated each of the 95 papers as a disputed text, whether or not they were included as part of the Hamilton or Madison reference text collection.

The *Federalist* papers have already been described, but the Pacificus and Helvidius papers need an introduction. From the "Summing Up" section of Mosteller and Wallace's paper:

> The main study is robust against changes in type of essay. The method performed satisfactorily on both *Federalist* material and on essays exterior to *The Federalist*. Such robustness is encouraging in considering the method for other studies and in assessing the generality of the present inference. The disputed papers could be as contextually disparate as the Pacificus or Helvidius papers for all we know.

On 22 April 1793, President George Washington issued a *Proclamation of Neutrality* in which he declared the United States to be neutral with respect to the military conflict between Great Britain and France. The official proclamation prompted considerable debate regarding its wisdom and its constitutionality. Alexander Hamilton, then the Secretary of the Treasury, supported both neutrality and the proclamation. Thomas Jefferson, then the Secretary of State, supported neutrality but opposed proclaiming it in public; he eventually resigned over the issue.

Hamilton began making his case to the public via a series of anonymous essays published under the pen name *Pacificus*. Jefferson, desperate to counter these essays, pleaded with a reluctant James

Madison to once again take up his pen, this time arguing against Hamilton rather than in parallel with him.

> Nobody answers him, & his doctrine will therefore be taken for confessed. For god's sake, my dear Sir, take up your pen, select the most striking heresies, and cut him to pieces in the face of the public. There is nobody else who can & will enter the lists with him.

Madison finally responded after the fifth of the *Pacificus* papers, turning out five papers under the pen name *Helvidius*. Hamilton responded with two *Americanus* papers, signing them also as *Pacificus*.

For this sub-study, I considered each of the 14 *Pacificus / Helvidius* papers to be a disputed text. Coupled with the 81 *Federalist* papers, this brings to 95 the total number of essays evaluated as part of this sub-study to assess the accuracy of current method's scoring equation and discriminator word list.

I present the complete results in Appendix X and summarize them here. For 93 of the 95 texts, the stylometric method presented herein calculated the correct author, at least the person generally deemed to be the author by historians and/or Mosteller and Wallace. The current method calculated the correct author for each of *The Federalist* papers. For *Helvidius I* and *Helvidius IV*, the method calculated Alexander Hamilton, rather than James Madison, to be the more likely author.

It is extremely unlikely that Hamilton wrote either of those two papers, but the program was forced to choose between Madison or Hamilton. A different interpretation of the incorrect results is that the papers were written by an unknown third party whose writing style was more similar to that of Hamilton than Madison.

Whoever the actual author of *Helvidius I* and *IV*, the calculations of current method for those two essays fly in the face of the generally accepted historical evidence. In the absence of other evidence to the contrary, the two calculations are deemed erroneous. Therefore, based solely on the 95 Hamilton and Madison papers included as part of this sub-study, the estimated accuracy of the stylometric method presented herein is 93 correct out of 95 possible, or 98%.

4.2 Ambiguous Results

Returning to the main study of discriminating among Arthur, Louise, and Jean, there are a least two categories of detectable errors: ambiguous results and chronologically troublesome results. I'll discuss the ambiguous results here and the chronologically troublesome category in the next section.

For each of the 225 Arthur-attributed stories evaluated as a part of this study, there were 8 possible outcomes.

> Arthur (AL=A JL=L AJ=A)
> Arthur (AL=A JL=J AJ=A)
> **Louise** (AL=L JL=L AJ=A)
> **Louise** (AL=L JL=L AJ=J)
> **Jean** (AL=A JL=J AJ=J)
> **Jean** (AL=L JL=J AJ=J)
> **Ambiguous** (AL=A JL=L AJ=J)
> **Ambiguous** (AL=L JL=J AJ=A)

There are 2 possible results in which the author calculation would be ambiguous on its face. Two hypothetical examples are shown below.

> *Hypothetical Story 1*; **Ambiguous** (AL=A7 JL=L6 AJ=J5)
> *Hypothetical Story 2*; **Ambiguous** (AL=L120 JL=J110 AJ=A120)

In the first hypothetical, Arthur is more likely than Louise, Louise is more likely than Jean, and Jean is more likely than Arthur. All three cannot be true at the same time, and the author is identified as Ambiguous and placed in bold to show that the calculated author was other than Arthur.

In the second hypothetical, Louise is more likely than Arthur, Jean is more likely than Louise, and Arthur is more likely than Jean. Again, all three cannot be true at the same time, and the author is identified as Ambiguous in bold font.

In only one instance was the program unable to discriminate an author among Arthur, Louise, and Jean. The story was *The Field Bazaar*.

> *The Field Bazaar*; **Ambiguous** (AL=A44 JL=L35 AJ=J10)

The Field Bazaar is a very short story. At only 1,052 words, it is shorter than every other story in the study other than *How Watson Learned the Trick* at 503 words, that very short story being scored unambiguously for Jean.

Seemingly more significant than the story length of *The Field Bazaar* seems to be the low value of the three scores. The function word usage doesn't seem particularly close to that of Arthur, Louise, or Jean. In any case, the calculated result for the story is deemed a detectable error.

4.3 Chronologically Troublesome Results

In his 150,000 word autobiography, Arthur claimed that he first met Louise, whom he chose to never identify by name, early in 1885 [Conan Doyle 70-71], apparently in March, soon before their marriage in August. However, as explained in *Shadow Woman*, Arthur was a fabulist. With respect to their first meeting, Arthur's autobiographical claim is not only suspicious on its face, given Victorian norms, it is contradicted by his much earlier semi-autobiographical novel *The Stark Munro Letters*. Therein, Stark Munro, obviously Arthur, meets Winnie LaForce, obviously Louise, during a train ride in March of 1882.

The date of their first meeting is significant with respect to assessing the accuracy of the results presented herein. Louise clearly could not have written any work published under Arthur's name prior to the date of their first meeting. Any result to the contrary would be an obvious error. Of the 225 stories considered in this study, 24 were published prior to August of 1885, when Louise married Arthur, and 4 were published prior to March of 1882, when Arthur may have first met Louise. The analysis indicates that none of them were written by Louise.

Similarly, any result identifying Louise as the author of a story published after her death would be suspicious, but not conclusively in error, since publication could have been withheld. Of the 60 stories in the study that were published after Louise's death in July of 1906, the analysis indicates that one of them was written by Louise.

   How it Happened (1913); **Louise** (AL=L31 JL=L25 AJ=J46)

Once again, the story is short, only 1,409 words, the scores are low, and the story was almost certainly not written by Louise. The result is another detectable error.

With respect to Jean, any result identifying her as the author of a story published after her death would be suspicious. Jean, however, outlived both Louise and Arthur, thereby rendering any such study result impossible.

Finally, any result identifying Jean as the author of a story published under Arthur's name prior to their first meeting would be an obvious error. On 16 March 1904, Arthur wrote a letter to his mother in which he noted that he and Jean had "been together seven years from yesterday now, and our love has grown with the years." [Lellenberg 522] Some of Arthur's chronologists therefore place the date of their first meeting as 15 March 1897, one day after Jean turned twenty-three years old. Such dating depends both on Arthur's honesty (a dangerous crutch) and his intended meaning of "been together," as compared to "met." Jean was, at the time, living in Kidbrooke, a London suburb, and Arthur was living in Hindhead, some fifty miles distant. Other evidence suggests that the two met earlier, and no evidence precludes their meeting anytime after Arthur and Louise moved to London in 1891, Jean then being eighteen years old and living less than ten miles distant.

If Arthur and Jean did first meet in March of 1897, then the study involves 113 stories published under Arthur's name prior to their first encounter. The analysis identifies Jean as the author of two of them.

> *Question of Diplomacy* (Jun 92); **Jean** (AL=L43 JL=J67 AJ=J118)
> *The Parasite* (Nov 94); **Jean** (AL=A74 JL=J143 AJ=J41)

For the purpose of accuracy estimation, these two calculated authors are deemed erroneous, though they may both in fact be correct.

4.4 An Estimate of the Method's Accuracy

We can estimate the error rate of the author predictions by comparing the actual number of detectable errors to the expected number of detectable errors. Recall that for each of the 225 Arthur-attributed stories evaluated as a part of this study, there were eight possible outcomes.

> Arthur (AL=A JL=L AJ=A)
> Arthur (AL=A JL=J AJ=A)
> **Louise** (AL=L JL=L AJ=A)
> **Louise** (AL=L JL=L AJ=J)
> **Jean** (AL=A JL=J AJ=J)
> **Jean** (AL=L JL=J AJ=J)
> **Ambiguous** (AL=A JL=L AJ=J)
> **Ambiguous** (AL=L JL=J AJ=A)

Assuming randomness, those 22 stories written before Arthur (allegedly) met either Louise or Jean had a 75% chance of a detectable error: 2 ambiguous possibilities plus 4 chronologically troublesome possibilities out of 8 total possibilities. The expected number of errors would be 22 x 0.75 = 16.5. There no chronologically troublesome errors in the group of 22 stories.

Assuming randomness, those 91 stories written after Arthur (allegedly) first met Louise but before Arthur (allegedly) first met Jean had a 50% chance of a detectable error: 2 ambiguous possibilities plus 2 chronologically troublesome possibilities out of 8 total possibilities. The expected number of errors would be 0.50 x 91 = 45.5. There were two detectable errors in this group of 52 stories: *A Question of Diplomacy* and *The Parasite*.

Assuming randomness, each of the 52 stories written while Arthur was married to Louise but seeing Jean had a 25% chance of a detectable error: 2 ambiguous possibilities out of 8 total possibilities. The expected number of errors would be 0.25 x 52 = 13. There was one ambiguous result in this group of 91 stories: *The Field Bazaar*.

Assuming randomness, those 60 stories written after Louise's death also had a 50% chance of a detectable error: 2 ambiguous possibilities plus 2 chronologically troublesome possibilities out of 8 total possibilities. The expected number of errors would be 0.50 x 60 = 30. There was one chronologically troublesome result in this group of 91 stories: *How it Happened*.

By comparing the actual number of detectable errors against the expected number of detectable errors, we can estimate the inaccuracy of this study as 4 detectable errors out of 105 expected, or approximately 2.5%. The estimated accuracy of the stylometric analysis method described herein is therefore greater than 97%, based on the frequency of detectable errors. This number compares well with the 98% accuracy rate predicted based on the sub-study of the Hamilton and Madison papers.

Based on the stylometric analysis alone, we cannot be certain whether Arthur, Louise, or Jean wrote any particular Sherlock Holmes adventures. However, assuming an accuracy rate of 97%, we can easily calculate that there is a 43% chance that Louise (or someone who used function words in a fashion similar to the Louise collection) wrote each and every one of the 28 Holmes adventures credited to her by the current stylometric study.

Similarly, we can easily calculate that there is a 41% chance that Jean (or someone who used function words in a fashion similar to the Jean collection) wrote each and every one of the 29 Holmes adventures credited to her by the current stylometric study.

Conversely, assuming the error rate of this study is 3%, we can easily calculate that there is zero chance that Arthur wrote each and every one of the Holmes adventures.

CONCLUSIONS

Adopting the methodology of Mosteller and Wallace, I have developed a stylometric analysis tool to test the authorship of 225 fictional works long attributed to Arthur Conan Doyle. Among the works evaluated were the 60 Sherlock Holmes adventures comprising what some describe as the Sherlock Holmes Canon.

The accuracy of the new methodology is calculated to be 98%, based on its application to the *Federalist* papers. The accuracy is calculated to be 97%, based on the infrequency of ambiguous or chronologically troublesome results from its application to the Conan Doyle works of fiction.

Based on the stylometric analysis, Louise is calculated to have co-authored the first Holmes adventure, *A Study in Scarlet*, wrote the second Holmes novel*, The Sign of Four*, wrote the third Holmes novel, *The Hound of the Baskervilles*, wrote the 24 earliest Holmes short stories, and wrote two of the intermediate short stories. She also wrote ten of the non-Holmes short stories usually attributed to Arthur.

Based on the stylometric analysis, Jean wrote all but five of the Holmes adventures published after *The Hound of the Baskervilles*. She also wrote fifteen of the non-Holmes short stories usually attributed to Arthur.

Based on the stylometric analysis, Arthur wrote three of the Holmes adventures and most of the non-Holmes fiction generally attributed to him.

The disputed text scores for all 165 non-Holmes works of fiction evaluated as part of the study are presented in Appendix XI.

The disputed text scores for all 60 Holmes adventures evaluated as part of this study are presented in Appendix XII.

REFERENCES

Allen, John: *Shadow Woman: The Real Creator of Sherlock Holmes*, Allen & Allen Semiotics, 2017

Barquin, Alexis: *The Arthur Conan Doyle Encyclopedia* website. This is a spectacular, readily available source of information regarding Arthur and his writings. It includes a digital copy of the text for each of the stories evaluated as part of this study.
available online at www.arthur-conan-doyle.com

Christie's: *The Conan Doyle Collection: Wednesday 19 May 2004*. This is a 179 page catalog prepared for an auction of Conan Doyle papers.

Conan Doyle, Arthur: *Memories and Adventures*, Antony Rowe Limited, 3rd edition, 1988, available online at
http://gutenberg.net.au/ebooks14/1400681h.html

Dakin, D. Martin: *A Sherlock Holmes Commentary*, Drake Publishers Inc, New York, 1972

Gardner, Martin: *Science: Good, Bad, and Bogus*, Avon Books, 1981

Lellenberg, Jon; Stashower, Daniel; and Foley, Charles: editors of *Arthur Conan Doyle: A Life in Letters*, Penguin Press, 2007

Miller, G. A.; Newman, E. B.; and Friedman, E. A.: "Length-Frequency Statistics for Written English" found in *Information and Control*, December 1958, available online at
http://www.sciencedirect.com/science/article/pii/S0019995858902298

Mosteller, Frederick and Wallace, David L.: "Inference in an Authorship Problem," *Journal of the American Statistical Association*, June 1963, pp 275-309, available online at
http://www.stat.cmu.edu/~vlachos/courses/724/final/mosteller.pdf

Mosteller, Frederick and Wallace, David L.: *Inference and Disputed Authorship: The Federalist*, Addison-Wesley Publishing Company, 1964

Robson, W. W.: Introduction and explanatory notes for *The Case-Book of Sherlock Holmes*, Oxford University Press, 1994

APPENDIX I
The Early-Arthur Collection

1. *The Mystery of Sasassa Valley*; Sep 1879; 4,684
2. *A Night Among the Nihilists*; Apr 1881; 4,618
3. *That Little Square Box; Dec* 1881; 7,433
4. *Gully of Bluemansdyke*; Dec 1881; 8,694
5. *Bones, the April Fool of Harvey's Sluice*; Apr 1882; 11,164
6. *Our Derby Sweepstakes*; May 1882; 8,918
7. *That Veteran*; Sep 1882; 5,150
8. *My Friend the Murderer*; Dec 1882; 7,356
9. *Captain of the Polestar*; Jan 1883; 10,080
10. *Gentlemanly Joe*; Mar 1883; 4,449
11. *A Winning Shot*; Jul 1883; 12,131
12. *Selecting a Ghost*; Dec 1883; 6,839
13. *The Silver Hatchet*; Dec 1883; 5,859
14. *An Exciting Christmas Eve*; Dec 1883; 6,339
15. *J. Habakuk's Jephson's Statement*; Jan 1884; 13,159
16. *The Heiress of Glenmahowley*; Jan 1884; 6,581
17. *The Blood-Stone Tragedy*; Feb 1884; 4,032
18. *John Barrington Cowles*; Apr 1884; 9,526
19. *The Tragedians*; Aug 1884; 9,882
20. *Crabbe's Practice*; Dec 1884; 4,830
21. *The Man from Archangel*; Jan 1885; 9,624
22. *The Lonely Hampshire Cottage*; May 1885; 4,269
23. *The Great Keinplatz Experiment*; Jul 1885; 6,270
24. *Elias B. Hopkins*; Dec 1885; 5,752
25. *The Fate of the Evangeline*; Dec 1885; 7,915
26. *Touch and Go: A Midshipman's Story*; Apr 1886; 4,022
27. *Cyprian Overbeck Wells*; Dec 1886; 7,124
28. *Uncle Jeremy's Household*; Jan 1887; 15,662
29. *The Stone of Boxman's Drift*; Dec 1887; 6,387
30. *John Huxford's Hiatus*; Jun 1888; 8,827
31. *The Ring of Thoth*; Jan 1890; 7,134
32. *A Physiologist's Wife*; Sep 1890; 7,026
33. *The Surgeon of Gaster Fells*; Dec 1890; 8,122
34. *A Pastoral Horror*; Dec 1890; 7,170
35. *Our Midnight Visitor*; Feb 1891; 9,338
36. *A Straggler of '15*; Mar 1891; 5,386
37. *The Colonel's Choice*; Jul 1891; 4,315
38. *A False Start*; Dec 1891; 5,042
39. *Out of the Running*; Feb 1892; 5,472
40. *A Question of Diplomacy*; Jun 1892; 5,023

41. *Lot No. 249*; Sep 1892; 13,148
42. *The Green Flag*; Jun 1893; 5,794

APPENDIX II
The Intermediate-Arthur Collection

1. *The Doctors of Hoyland*; Apr 1894; 4,414
2. *The Lord of Chateau Noir*; Jul 1894; 4,600
3. *The Parasite*; Oct 1894; 17,917
4. *A Medical Document*; Oct 1894; 4,416
5. *The Curse of Eve;* Oct 1894; 4,338
6. *How the Brigadier Won his Medal*; Dec 1894; 10,035
7. *The Recollections of Captain Wilke*; Jan 1895; 5,660
8. *How the Brigadier Held the King*; Apr 1895; 10,111
9. *How the King Held the Brigadier*; May 1895; 9,453
10. *How the Brigadier Slew the Brothers of Ajaccio*; Jun 1895; 8,765
11. *How the Brigadier Came to the Castle of Gloom*; Jul 1895; 9,056
12. *How the Brigadier Took the Field*; Aug 1895; 9,485
13. *How the Brigadier was Tempted by the Devil*; Sep 1895; 9,321
14. *How the Brigadier Played for a Kingdom*; Dec 1895; 8,751
15. *The Three Correspondents*; Oct 1896; 13,276
16. *How the Governor of Saint Kitts Came Home*; Jan 1897; 4,764
17. *The Dealings of Captain Sharkey*; Mar 1897; 4,860
18. *The Voyage of Copley Banks*; May 1897; 4,320
19. *The Striped Chest*; Jul 1887; 5,395
20. *Fiend of the Cooperage*; Oct 1897; 4,980
21. *The New Catacomb*; Jan 1898; 5,439
22. *The Story of the The Beetle Hunter*; Jun 1898; 6,181
23. *The Story of the Man with the Watches*; Jul 1898; 7,034
24. *The King of the Foxes*; Jul 1898; 5,519
25. *The Story of the Lost Special*; Aug 1898; 7,000
26. *The Story of the Sealed Room*; Sep 1898; 5,036
27. *The Story of the Black Doctor*; Oct 1898; 7,243
28. *The Story of the Club-Footed Grocer*; Nov 1898; 7,463
29. *The Story of the Brazilian Cat*; Dec 1898; 8,099
30. *A Shadow Before*; Dec 1898; 4,682
31. *The Story of the Japanned Box*; Jan 1899; 4,918
32. *The Story of the Jew's Breast-Plate*; Feb 1899; 7,505
33. *The Story of B.24*; Mar 1899; 6,274
34. *The Story of the Latin Tutor*; Apr 1899; 12,360
35. *The Story of the Brown Hand*; May 1899; 6,142
36. *The Croxley Master*; Oct 1899; 15,950
37. *The Crime of the Brigadier*; Jan 1900; 5,637
38. *The Debut of Bimbashi Joyce*; Jan 1900; 7,119
39. *Playing with Fire*; Mar 1900; 4,841
40. *The Coming of the Huns*; Sep 1901; 4,138

41. *How the Brigadier Lost His Ear*; Sep 1902; 8,574
42. *The Leather Funnel*; Nov 1902; 4,672
43. *How the Brigadier Saved the Army*; Nov 1902; 8,776
44. *How the Brigadier Rode to Minsk*; Dec 1902; 7,796
45. *How the Brigadier Bore Himself at Waterloo*; Jan 1903; 15,365
46. *How the Brigadier Triumphed in England*; Mar 1903; 6,974
47. *How the Brigadier Captured Saragossa*; Apr 1903; 8,402
48. *The Last Adventure of the Brigadier*; May 1903; 5,651

APPENDIX III
The Late-Arthur Collection

1. *The Pot of Caviare*; Mar 1908; 5,069
2. *The Lord of Falconbridge*; Aug 1909; 9,019
3. *The Homecoming*; Dec 1909; 4,930
4. *The Terror of Blue John Gap*; Aug 1910; 7,172
5. *The Lift*; Sep 1910; 4,263
6. *The Blighting of Sharkey*; Apr 1911; 4,778
7. *Giant Maximin*; Jul 1911; 5,736
8. *One Crowded Hour*; Aug 1911; 5,001
9. *The Fall of Lord Barrymore*; Dec 1912; 4,669
10. *Borrowed Scenes*; Sep 1913; 4,169
11. *The Horror of the Heights*; Nov 1913; 6,673
12. *Danger!*; Jul 1914; 11,791
13. *The Prisoner's Defense*; Jan 1916; 4,973
14. *Three of Them*; Apr 1918; 9,378
15: *The Bully of Brocas Court*; Nov 1921; 4,975
16: *When the World Screamed*; Apr 1928; 11,311
17. *The Disintegration Machine*; Jan 1929; 4,835
18. *The Death Voyage*; Sep 1929; 9,752
19. *The End of Devil Hawker*; Aug 1930; 9,676
20. *The Last Resource*; Aug 1930; 4,282

APPENDIX IV
The Louise Collection

1. *A Scandal in Bohemia*; Jul 1891; 8,606
2. *The Red-Headed League*; Aug 1891; 9,209
3. *A Case of Identity*; Sep 1891; 7,022
4. *The Boscombe Valley Mystery*; Oct 1991; 9,694
5. *The Five Orange Pips*; Nov 1891; 7,372
6. *The Man with the Twisted Lip*; Dec 1891; 9,298
7. *The Adventure of the Blue Carbuncle*; Jan 1892; 7,881
8. *The Adventure of the Speckled Band*; Feb 1892; 9,891
9. *The Adventure of the Engineer's Thumb*; Mar 1892; 8,349
10. *The Adventure of the Noble Bachelor*; Mar 1892; 8,169
11. *The Adventure of the Beryl Coronet*; May 1892; 9,746
12. *The Adventure of the Copper Beeches*; Jun 1882; 10,004
13. *The Adventure of Silver Blaze*; Dec 1892; 9,664
14. *The Adventure of The Cardboard Box*; Jan 1893; 8,738
15. *The Adventure of the Yellow Face*; Feb 1893; 7,540
16. *The Adventure of the Stockbroker's Clerk*; Mar 1893; 6,824
17. *The Adventure of the Gloria Scott*; Apr 1893; 7,913
18. *The Adventure of the Musgrave Ritual*; May 1893; 7,626
19. *The Adventure of the Reigate Squire*; Jun 1893; 7,222
20. *The Adventure of the Crooked Man*; Jul 1893; 7,163
21. *The Adventure of the Resident Patient*; Aug 1893; 6,499
22. *The Adventure of the Greek Interpreter*; Sep 1893; 7,057
23. *The Adventure of the Naval Treaty*; Oct 1893; 12,178
24, *The Adventure of the Final Problem*; Dec 1893; 7,205

APPENDIX V
The Jean Collection

1. *The Adventure of Wisteria Lodge*; Aug 1908; 11,458
2. *The Adventure of the Bruce-Partington Plans*; Dec 1908; 10,764
3. *The Adventure of the Devil's Foot*; Dec 1910; 10,054
4. *The Adventure of the Red Circle*; Mar 1911; 7,631
5. *The Disappearance of Lady Frances Carfax*; Dec 1911; 7,740
6. *The Adventure of the Dying Detective*; Nov 1913; 5,822
7. *His Last Bow*; Sep 1917; 6,124
8. *The Adventure of the Mazarin Stone*; Nov 1921; 5,707
9. *The Problem of Thor Bridge*; Feb 1922; 9,662
10. *The Adventure of the Creeping Man*; Mar 1923; 7,717
11. *The Adventure of the Sussex Vampire*; Jan 1924; 5,995
12. *The Adventure of the Three Garridebs*; Oct 1924; 6,239
13. *The Adventure of the Illustrious Client*; Nov 1924; 9,833
14. *The Adventure of the Three Gables*; Sep 1926; 6,091
15. *The Adventure of the Blanched Soldier*; Oct 1926; 7,760
16. *The Adventure of the Lion's Mane*; Nov 1926; 7,237
17. *The Adventure of the Retired Colourman*; Dec 1926; 5,562
18. *The Adventure of the Veiled Lodger*; Jan 1927; 4,496
19. *The Adventure of Shoscombe Old Place*; Mar 1927; 6,298

APPENDIX VI
Holmes Adventures Excluded from the Collections

1. *A Study in Scarlet*; Nov 1887; 47,291
2. *The Sign of Four;* Feb 1890; 34,706
3. *The Hound of the Baskervilles*; Aug 1901; 59,570
4. *The Adventure of the Empty House*; Sep 1903; 8,773
5. *The Adventure Norwood Builder*; Nov 1903; 9,319
6. *The Adventure of the Dancing Men*; Dec 1903; 9,697
7. *The Adventure of the Solitary Cyclist*; Dec 1903; 7,909
8. *The Adventure of Priory School*; Jan 1904; 11,516
9. *The Adventure of Black Peter;* Feb 1904; 8,194
10. *The Adventure of Charles Augustus Milverton*; Mar 1904; 6,775
11. *The Adventure of Six Napoleons*; Apr 1904; 8,394
12. *The Adventure of the Three Students*; Jun 1904; 6,509
13. *The Adventure of the Golden Pince-Nez*; Jul 1904, 8,995
14. *The Adventure of the Missing Three-Quarter*; Aug 1904; 8,089
15. *The Adventure of the Abbey Grange*; Sep 1904; 9,215
16. *The Adventure of the Second Statin*; Dec 1904; 9,738
17. *The Valley of Fear*; Sep 1914; 57,999

APPENDIX VII
Non-Holmes Works Excluded from the Collections

1. *The Cabman's Story*; May 1884; 3,243
2. *The Mystery of Cloomber*; Dec 1888; 49,183
3. *Micah Clarke*; Jan 1889; 175,733
4. *The Firm of Girdlestone*; Oct 1889; 138,120
5. *The White Company*; Jan 1891; 152,255
6. *The Voice of Science*; Mar 1891; 2,558
7. *A Sordid Affair*; Nov 1891; 3,475
8. *The Doings of Raffles Haw*; Dec 1891; 38,321
9. *Beyond the City*; Dec 1891; 39,786
10. *De Profundis*; Mar 1892; 3,994
11. *A Regimental Scandal*; May 1892; 3,139
12. *Between Two Fires*; Nov 1892; 2,145
13. *Jeland's Voyage*; Nov 1892; 2,776
14. *The Los Amigos Fiasco*; Dec 1892; 2,813
15. *The Great Shadow*; Dec 1892; 41,137
16. *The Refugees*; Jan 1893; 123,801
17. *The Slapping Sal*; Jul 1893; 3,269
18. *The Case of Lady Sannox*; Nov 1893; 3,785
19. *Sweethearts*; Jun 1894; 2,421
20. *The Third Generation*; Nov 1894; 3,837
21. *The Surgeon Talks*; Nov 1894; 2,830
22. *His First Operation*; Nov 1894; 2,205
23. *Behind the Times*; Nov 1894; 1,688
24. *A Foreign Office Romance*; Nov 1894; 3,892
25. *The Stark Munro Letters*; Sep 1895; 78,124
26. *Rodney Stone*; Jan 1896; 91,886
27. *The Field Bazaar*; Nov 1896; 1,052
28. *Uncle Bernac*; Dec 1896; 57,969
29. *The Tragedy of the Korosko*; Jun 1897; 47,312
30. *The Confession;* Jan 1898; 2,338
31. *The Retirement of Signor Lambert*; Dec 1898; 3,140
32. *A Duet, with an Occasional Chorus*; Apr 1899; 67,985
33. *An Impression of the Regency*; Aug 1900; 3,050
34. *The First Cargo*; Nov 1901; 2,982
35. *Sir Nigel*; Dec 1905; 131,556
36. *The Silver Mirror*; Aug 1908; 3,981
37. *The Marriage of the Brigadier*; Sep 1910; 3,863
38. *The Last Galley*; Nov 1910; 2,984
39. *The Last of the Legions*; Nov 1910; 2,422

40. *The Red Star*; Jan 1911; 3,560
41. *The Contest*; Feb 1911; 2,869
42. *Through the Veil*; Apr 1911; 2,000
43. *An Iconoclast*; May 1911; 2,595
44. *The Lost World*; Apr 1912; 76,542
45. *The Poison Belt*; Mar 1913; 29,435
46. *How it Happened*; Sep 1913; 1,409
47. *A Point of View*; Dec 1918; 1,694
48. *The Nightmare Room*; Dec 1921; 2,577
49. *The Centurion*; Nov 1922; 1,997
50. *A Point of Contact*; Nov 1922; 2,510
51. *Billy Bones*; Dec 1922; 3,318
52. *How Watson Learned the Trick*; 1924
53. *Land of Mist*; Jul 1925; 74,096
54. *The Maracot Deep*; Oct 1927; 41,393
55. *The Parish Magazine*; Nov 1930; 3,043

APPENDIX VIII
The Function Words

The 100 function words used in this study are presented below in order of their frequency as found within the 110 stories of the Arthur collection. The frequency is expressed as occurrences per million words. The list begins with the word *the* at 59,652 occurrences per million words, and ends with the word *else* at 206 occurrences per million words.

As an example, there are 47,823 occurrences of the word *the* within the 110 stories of the Arthur collection. The total word count for that collection is 801,705. The frequency of the word *the* in the Arthur collection is therefore calculated as 47,823 / 801,705 x 1,000,000 = 59,652 occurrences per million words.

1. the; 59,652
2. and; 28,161
3. of; 30,625
4. a; 24,853
5. to; 23,403
6. in;15,719
7. it; 13,965
8. that; 13,939
9. with; 8,824
10. as; 8,116
11. for; 7,344
12. at; 6,825
13. which; 6,554
14. but; 6,262
15. from; 4,675
16. not; 4,631
17. there; 4,614
18. upon; 4,600
21. this; 4,125
19. one; 4,048
20. on; 3,982
22. so; 3,667
23. all; 3,663
24. by; 3,496
25. an; 3,312
26. no; 3,293
27. would; 3,063
28. what; 3,012
29. could; 2,965
30. when; 2,935

31. out; 2,870
32. up; 2,864
33. if; 2,745
34. who; 2,527
35. then; 2,493
36. into; 2,454
37. some; 2,328
38. very; 2,264
39. down; 2,213
40. or; 2,19
41. more; 1,916
42. over; 1,897
43. now; 1,746
44. about; 1,663
45. two; 1,603
46. before; 1,569
47. like; 1,476
48. only; 1,375
49. never; 1,318
50. other; 1,296
51. any; 1,293
52. than; 1,280
53. such; 1,219
54. through; 1,144
55. might; 1,094
56. after; 1,065
57. here; 1,010
58. still; 1,009
59. off; 1,004
60. away; 979
61. last; 964
62. again; 948
63. where; 937
64. every; 921
65. first; 907
66. nor; 906
67. much; 866
68. too; 831
69. nothing; 813
70. yet; 801
71. three; 755
72. most; 741
73. those; 737

74. until; 727
75. however; 723
76. while; 723
77. many; 717
78. against; 711
79. why; 706
80. just; 684
81. under; 674
82. same; 656
83. something; 655
84. without; 631
85. though; 607
86. another; 596
87. between; 590
88. few; 575
89. enough; 544
90. each; 515
91. always; 493
92. across; 485
93. both; 457
94. also; 422
95. together; 419
96. hardly; 362
97. rather; 349
98. next; 339
99. indeed; 282
100. else; 206

APPENDIX IX
A Sample Calculation

Each record shows one of the 100 function words and its word score for an Arthur-to-Louise comparison for *A Scandal in Bohemia*.

word score = 10 x [(fX - fL)^2 - (fX - fA)^2)] / (vL + vA)

fX = frequency of the function word in *A Scandal in Bohemia*, expressed in occurrences per million words

fL = frequency of the function word in the Louise collection, in occurrences per million words

fA = frequency of the function word in the early-Arthur collection, in occurrences per million words

vL = population variance of the function word from the Louise collection, in occurrences per million words squared

vA = population variance the function word from the early-Arthur collection, in occurrences per million words squared

Negative word scores are suggestive of Louise; positive word scores are suggestive of Arthur.

The function word records are ordered by increasing word score, from function words most suggestive of Louise to function words most suggestive of Arthur, at least for *A Scandal in Bohemia*.

1. then; -38.94
(fX=4,067; fA=2,022; fL=3,272; vA=407,880; vL=503,467)

2. not; -26.54
(fX=6,972; fA=3,671; fL=5,397; vA=1,396,111; vL=1,774,984)

3. it; -23.11
(fX=17,894; fA=12,655; fL=15,650; vA=5,711,118; vL=3,985,475)

4. to; -15.62
(fX=29,050; fA=23,774; fL=25,768; vA=7,674,237; vL=3,248,591)

5. just; -12.47
(fX=1,511; fA=660; fL=1,172; vA=215,279; vL=272,481)

6. off; -10.35
(fX=349; fA=1,146; fL=819; vA=303,949; vL=97,162)

7. though; -9.27
(fX=232; fA=860; fL=601; vA=195,047; vL=83,167)

8. through; -8.90

(fX=465; fA=1,121; fL=834; vA=194,829; vL=135,285)

9. like; -8.41
(fX=697; fA=1,488; fL=1,008; vA=485,463; vL=142,495)

10. nothing; -7.97
(fX=1,278; fA=747; fL=1,043; vA=160,858; vL=123,400)

11. indeed; -6.66
(fX=1,278; fA=338; fL=417; vA=101,300; vL=112,328)

12. what; -6.52
(fX=3,718; fA=2,773; fL=3,570; vA=1,088,907; vL=248,804)

13. over; -6.18
(fX=1,046; fA=2,151; fL=1,882; vA=584,800; vL=261,259)

14. rather; -5.82
(fX=813; fA=448; fL=611; vA=77,411; vL=81,854)

15. many; -5.80
(fX=349; fA=757; fL=367; vA=259,191; vL=27,473)

16. after; -5.26
(fX=581; fA=1,288; fL=1,058; vA=306,904; vL=211,866)

17. who; -5.00
(fX=3,486; fA=2,299; fL=2,612; vA=945,828; vL=342,623)

18. the; -4.69
(fX=51,824; fA=60,101; fL=54,238; vA=85,749,731; vL=47,918,440)

19. on; -4.44
(fX=3,370; fA=4,534; fL=3,972; vA=1,806,344; vL=430,847)

20. three; -4.33
(fX=1,162; fA=589; fL=730; vA=209,946; vL=123,917)

21. most; -4.20
(fX=1,278; fA=747; fL=864; vA=142m642; vL=119,965)

22. if; -3.45
(fX=2,905; fA=2,222; fL=2,736; vA=813,022; vL=458,064)

23. else; -3.36
(fX=465; fA=229; fL=283; vA=24,989; vL=42,745)

24. as; -3.21

(fX=6,972; fA=8,556; fL=8,018; vA=2,984,253; vL=1,426,010)

25. upon; -3.13
(fX=2,905; fA=4,534; fL=4,012; vA=3,469,653; vL=1,100,601)

26. about; -3.00
(fX=1,046; fA=1,958; fL=1,728; vA=921,469; vL=299,413)

27. could; -2.96
(fX=2,905; fA=2,376; fL=2,840; vA=474,076; vL=454,048)

28. together; -2.49
(fX=232; fA=477; fL=387; vA=95,811; vL=47,129)

29. always; -2.48
(fX=581; fA=383; fL=501; vA=64,141; vL=68,180)

30. in; -2.24
(fX=18,127; fA=16,407; fL=16,926; vA=4,037,579; vL=2,740,809)

31. never; -2.00
(fX=697; fA=1,349; fL=1,192; vA=601,127; vL=300,421)

32. away; -1.93
(fX=697; fA=1,124; fL=993; vA=337,914; vL=151,673)

33. when; -1.91
(fX=3,602; fA=2,892; fL=3,183; vA=1,202,327; vL=517,809)

34. next; -1.71
(fX=116; fA=328; fL=273; vA=95,931; vL=23,294)

35. but; -1.53
(fX=6,158; fA=5,452; fL=6,132; vA=2,305,159; vL=958,261)

36. against; -1.47
(fX=581; fA=750; fL=536; vA=117,965; vL=63,466)

37. too; -1.46
(fX=813; fA=1,043; fL=705; vA=238,424; vL=43,196)

38. with; -1.27
(fX=7,901; fA=8,582; fL=7,934; vA=2,501,441; vL=1,142,471)

39. might; -1.18
(fX=1,162; fA=956; fL=1,197; vA=203,549; vL=143,686)

40. yet; -0.91

(fX=1,394; fA=725; fL=745; vA=189,800; vL=103,627)

41. again; -0.72
(fX=581; fA=895; fL=834; vA=223,079; vL=259,873)

42. for; -0.70
(fX=5,345; fA=6,988; fL=6,911; vA=1,874,598; vL=1,612,090)

43. hardly; -0.62
(fX=349; fA=496; fL=452; vA=107,372; vL=69,916)

44. few; -0.58
(fX=465; fA=731; fL=705; vA=131,302; vL=95,971)

45. at; -0.53
(fX=7,204; fA=6,727; fL=7,418; vA=1,436,402; vL=1,993,207)

46. some; -0.50
(fX=1,511; fA=2,541; fL=2,502; vA=809,759; vL=745,919)

47. also; -0.49
(fX=349; fA=242; fL=427; vA=61,982; vL=46,927)

48. without; -0.47
(fX=1,162; fA=670; fL=685; vA=246,508; vL=70,427)

49. now; -0.37
(fX=1,975; fA=1,745; fL=2,115; vA=440,091; vL=451,541)

50. from; -0.37
(fX=5,229; fA=4,598; fL=4,692; vA=2,088,179; vL=875,820)

51. an; -0.37
(fX=3,718; fA=3,346; fL=3,346; vA=949,398; vL=652,252)

52. something; -0.36
(fX=232; fA=683; fL=670; vA=209,407; vL=94,416)

53. such; -0.30
(fX=1,046; fA=1,191; fL=993; vA=440,136; vL=180,858)

54. by; -0.24
(fX=3,602; fA=3,835; fL=3,540; vA=1,410,574; vL=684,173)

55. before; -0.22
(fX=1,627; fA=1,530; fL=1,633; vA=260,199; vL=173,414)

56. each; -0.18

(fX=349; fA=428; fL=412; vA=83,209; vL=44,557)

57. enough; -0.02
(fX=581; fA=551; fL=606; vA=94,624; vL=74,733)

58. out; -0.01
(fX=3,137; fA=3,107; fL=3,133; vA=956,271; vL=356,850)

59. other; 0.00
(fX=1,394; fA=1,359; fL=1,430; vA=293,330; vL=350,469)

60. both; 0.02
(fX=581; fA=509; fL=506; vA=128,791; vL=45,099)

61. every; 0.10
(fX=813; fA=882; fL=730; vA=131,072; vL=102,337)

62. however; 0.10
(fX=581; fA=1,037; fL=1,043; vA=379,968; vL=167,335)

63. those; 0.13
(fX=465; fA=599; fL=611; vA=106,250; vL=139,537)

64. across; 0.16
(fX=465; fA=464; fL=417; vA=89,078; vL=54,646)

65. why; 0.17
(fX=697; fA=715; fL=780; vA=228,364; vL=162,041)

66. more; 0.19
(fX=1,859; fA=1,929; fL=1,713; vA=570,061; vL=315,808)

67. all; 0.19
(fX=3,370; fA=3,310; fL=3,560; vA=1,180,666; vL=557,334)

68. until; 0.21
(fX=581; fA=737; fL=760; vA=250,771; vL=111,253)

69. another; 0.22
(fX=465; fA=576; fL=338; vA=133,275; vL=34,915)

70. between; 0.37
(fX=465; fA=512; fL=566; vA=70,267; vL=149,265)

71. into; 0.46
(fX=3,486; fA=2,647; fL=2,617; vA=504,535; vL=619,645)

72. where; 0.50

(fX=1,046; fA=950; fL=1,226; vA=261,317; vL=203,071)

73. of; 0.54
(fX=27,423; fA=29,010; fL=25,317; vA=23,340,880; vL=12,026,111)

74. any; 0.66
(fX=465; fA=1,427; fL=1,455; vA=449,097; vL=391,860)

75. and; 0.67
(fX=29,747; fA=30,968; fL=28,112; vA=11,009,207; vL=6,718,057)

76. same; 0.70
(fX=349; fA=612; fL=640; vA=110,672; vL=116,800)

77. there; 0.87
(fX=3,718; fA=4,405; fL=4,528; vA=1,450,578; vL=664,454)

78. one; 0.89
(fX=3,137; fA=3,813; fL=3,898; vA=905,341; vL=462,866)

79. still; 0.90
(fX=1,046; fA=1,188; fL=809; vA=298,716; vL=97,689)

80. a; 1.02
(fX=26,028; fA=26,241; fL=24,517; vA=16,539,261; vL=5,474,254)

81. which; 1.09
(fX=6,391; fA=6,701; fL=7,333; vA=4,664,144; vL=2,613,745)

82. than; 1.26
(fX=1,278; fA=1,346; fL=1,524; vA=238,026; vL=204,644)

83. nor; 1.42
(fX=1,046; fA=960; fL=685; vA=646,454; vL=217,195)

84. much; 1.69
(fX=813; fA=969; fL=1,137; vA=307,799; vL=168,432)

85. while; 1.87
(fX=930; fA=960; fL=660; vA=299,783; vL=83,305)

86. would; 1.87
(fX=2,324; fA=2,579; fL=2,959; vA=907,320; vL=898,071)

87. two; 1.95
(fX=1,278; fA=1,433; fL=1,648; vA=366,895; vL=213,062)

88. this; 2.18

(fX=3,951; fA=3,977; fL=4,786; vA=1,939,062; vL=1,255,240)

89. down; 2.24
(fX=2,556; fA=2,586; fL=2,060; vA=793,730; vL=301,419)

90. last; 3.42
(fX=116; fA=934; fL=1,023; vA=249,642; vL=198,207)

91. very; 3.84
(fX=2,556; fA=2,225; fL=3,510; vA=842,007; vL=1,241,559)

92. up; 4.04
(fX=3,718; fA=3,352; fL=2,835; vA=1,387,945; vL=210,104)

93. or; 4.36
(fX=2,440; fA=2,335; fL=1,787; vA=548,256; vL=404,376)

94. here; 4.85
(fX=349; fA=1,101; fL=1,261; vA=285,838; vL=262,947)

95. so; 4.89
(fX=2,789; fA=3,339; fL=3,808; vA=683,219; vL=821,576)

96. first; 6.81
(fX=232; fA=805; fL=1,038; vA=233,276; vL=237,156)

97. under; 7.89
(fX=930; fA=650; fL=462; vA=111,857; vL=66,773)

98. no; 38.95
(fX=1,859; fA=2,995; fL=3,475; vA=821,911; vL=657,063)

99. only; 10.91
(fX=930; fA=989; fL=1,653; vA=218,243; vL=258,763)

100. that; 29.15
(fX=12,085; fA=112,362; fL=16,936; vA=4,529,631; vL=3,517,732)

The disputed text score is the sum of all the Arthur-versus-Louise word scores for *A Scandal in Bohemia* is -155.38. The result is presented as AL=L155, indicating that Louise is more likely the author of *A Scandal in Bohemia* than Arthur.

Though not detailed herein, the text score for the Jean-versus-Louise comparison for *A Scandal in Bohemia* is JL=235, indicating that Louise is more likely the author than Jean.

Since Louise is calculated to be both more likely than Arthur and more likely than Jean, Louise is calculated to be the author of *A Scandal in Bohemia*.

APPENDIX X
Results of the Current Method Applied to
The Hamilton and Madison Papers

For each of the 61 anonymous writings generally attributed to Alexander Hamilton, the stylometric analysis method presented herein calculated Hamilton to be the author of each of them.

The Federalist 1; Hamilton (H325)
The Federalist 6; Hamilton (H108)
The Federalist 7; Hamilton (H335)
The Federalist 8; Hamilton (H163)
The Federalist 9; Hamilton (H54)
The Federalist 11; Hamilton (H244)
The Federalist 12; Hamilton (H192)
The Federalist 13; Hamilton (H267)
The Federalist 15; Hamilton (H309)
The Federalist 16; Hamilton (H358)
The Federalist 17; Hamilton (H216)
The Federalist 21; Hamilton (H293)
The Federalist 22; Hamilton (H354)
The Federalist 23; Hamilton (H181)
The Federalist 24; Hamilton (H320)
The Federalist 25; Hamilton (H125)
The Federalist 26; Hamilton (H355)
The Federalist 27; Hamilton (H237)
The Federalist 28; Hamilton (H197)
The Federalist 29; Hamilton (H416)
The Federalist 30; Hamilton (H299)
The Federalist 31; Hamilton (H385)
The Federalist 32; Hamilton (H75)
The Federalist 33; Hamilton (H227)
The Federalist 34; Hamilton (H388)
The Federalist 35; Hamilton (H226)
The Federalist 36; Hamilton (H224)
The Federalist 59; Hamilton (H193)
The Federalist 60; Hamilton (H282)
The Federalist 61; Hamilton (H191)
The Federalist 65; Hamilton (H308)
The Federalist 66; Hamilton (H286)
The Federalist 67; Hamilton (H157)
The Federalist 68; Hamilton (H194)
The Federalist 69; Hamilton (H205)
The Federalist 70a; Hamilton (H214)

The Federalist 70b; Hamilton (H243)
The Federalist 71; Hamilton (H204)
The Federalist 72; Hamilton (H305)
The Federalist 73; Hamilton (H292)
The Federalist 74; Hamilton (H275)
The Federalist 75; Hamilton (H270)
The Federalist 76; Hamilton (H297)
The Federalist 77; Hamilton (H349)
The Federalist 78; Hamilton (H173)
The Federalist 79; Hamilton (H124)
The Federalist 80; Hamilton (H162)
The Federalist 81; Hamilton (H212)
The Federalist 82; Hamilton (H161)
The Federalist 83; Hamilton (H217)
The Federalist 84; Hamilton (H133)
The Federalist 85; Hamilton (H220)
Pacificus I; Hamilton (H216)
Pacificus II; Hamilton (H130)
Pacificus III; Hamilton (H161)
Pacificus IV; Hamilton (H300)
Pacificus V; Hamilton (H239)
Pacificus VI; Hamilton (H83)
Pacificus VII; Hamilton (H309)
Americanus I; Hamilton (H169)
Americanus II; Hamilton (H88)

For each of the following 19 anonymous writings generally attributed to James Madison, the stylometric analysis method presented herein calculated Madison to be the author of 17 of them.

The Federalist 10; Madison (M239)
The Federalist 14; Madison (M175)
The Federalist 37; Madison (M274)
The Federalist 38; Madison (M150)
The Federalist 39; Madison (M304)
The Federalist 40; Madison (M188)
The Federalist 41; Madison (M184)
The Federalist 42; Madison (M232)
The Federalist 43; Madison (M289)
The Federalist 44; Madison (M223)
The Federalist 45; Madison (M232)
The Federalist 46; Madison (M270)
The Federalist 47; Madison (M358)
The Federalist 48; Madison (M296)

Helvidius I; **Hamilton** (H16)
Helvidius II; Madison (M65)
Helvidius III; Madison (M40)
Helvidius IV; **Hamilton** (H161)
Helvidius V; Madison (M34)

For each of the 12 disputed *Federalist* papers that Mosteller and Wallace calculate to be the work of James Madison, the stylometric analysis method presented herein calculated Madison to be the author of each of them.

The Federalist 49; Madison (M158)
The Federalist 50; Madison (M213)
The Federalist 51; Madison (M326)
The Federalist 52; Madison (M216)
The Federalist 53; Madison (M255)
The Federalist 54; Madison (M193)
The Federalist 55; Madison (M96)
The Federalist 56; Madison (M305)
The Federalist 57; Madison (M225)
The Federalist 58; Madison (M180)
The Federalist 62; Madison (M223)
The Federalist 62; Madison (M180)

For each of the 3 *Federalist* papers generally considered to be a collaboration between Hamilton and Madison, and for which Mosteller and Wallace determine Madison to the more likely primary author, the stylometric analysis method presented herein calculated Madison to be the author of each of them.

The Federalist 18; Madison (M182)
The Federalist 19; Madison (M215)
The Federalist 20; Madison (M134)

In summary, for 93 out of 95 texts, the stylometric analysis method described herein calculated the author in a fashion consistent with the generally accepted author, and consistent with the work of Mosteller and Wallace. For two papers generally attributed to Madison, *Helvidius I* and *Helvidius II*, the method calculated Hamilton to more likely be the author, keeping in mind that Hamilton was the only alternative open to the stylometric method.

Based solely on the evaluation results for the 95 anonymous papers generally attributed to Hamilton and/or Madison, the accuracy of the stylometric analysis method described herein is 98%.

APPENDIX XI
Calculated Authorship of the Non-Holmes Stories Attributed to Arthur Conan Doyle

Event: Arthur sets up medical practice in Southsea
Note: Begin use of the early-Arthur collection

The Mystery of Sasassa Valley; Arthur (AL=A163 JL=L327 AJ=A366)
A Night Among the Nihilists; Arthur (AL=A157 JL=L41 AJ=A152)
Gully of Bluemansdyke; Arthur (AL=A463 JL=L324 AJ=A630)
That Little Square Box; Arthur (AL=A153 JL=L228 AJ=296)

Event: Arthur and Louise possibly meet for first time

Bones; Arthur (AL=A350 JL=L297 AJ=A515)
Our Derby Sweepstakes; Arthur (AL=A134 JL=L416 AJ=A424)
That Veteran; Arthur (AL=A338 JL=L333 AJ=A563)
My Friend the Murderer; Arthur (AL=A291 JL=L360 AJ=A538)
Captain of the Polestar; Arthur (AL=A210 JL=L289 AJ=A371)
Gentlemanly Joe; Arthur (AL=A273 JL=L304 AJ=A466)
A Winning Shot; Arthur (AL=A156 JL=L288 AJ=A340)
The Silver Hatchet; Arthur (AL=A265 JL=L214 AJ=A386)
Selecting a Ghost; Arthur (AL=A284 JL=L185 AJ=A392)
An Exciting Christmas Eve; Arthur (AL=A136 JL=L264 AJ=A334)
J. Habakuk's Jephson; Arthur (AL=A297 JL=L470 AJ=A597)
The Heiress of Glenmahowley; Arthur (AL=A162 JL=L163 AJ=A258)
The Blood-Stone Tragedy; Arthur (AL=A186 JL=L318 AJ=A336)
John Barrington Cowles; Arthur (AL=A100 JL=L187 AJ=A218)
The Cabman's Story; Arthur (AL=A186 JL=L290 AJ=A363)
The Tragedians; Arthur (AL=A134 JL=L157 AJ=A201)
Crabbe's Practice; Arthur (AL=A161 JL=L151 AJ=A224)
The Man from Archangel; Arthur (AL=A280 JL=L532 AJ=A620)
Lonely Hampshire Cottage; Arthur (AL=A165 JL=L319 AJ=A381)
Great Keinplatz Experiment; Arthur (AL=A82 JL=L238 AJ=A226)

Event: Arthur marries Louise

Elias B. Hopkins; Arthur (AL=A405 JL=L317 AJ=A592)
The Fate of the Evangeline; Arthur (AL=A57 JL=L423 AJ=A322)
Touch and Go; Arthur (AL=A200 JL=L679 AJ=A636)
Cyprian Overbeck Wells; Arthur (AL=A221 JL=L264 AJ=A397)
Uncle Jeremy's Household; Arthur (AL=A88 JL=L312 AJ=A286)
The Stone of Boxman's Drift; Arthur (AL=A136 JL=L275 AJ=A307)
John Huxford's Hiatus; Arthur (AL=A148 JL=L526 AJ=A470)
The Mystery of Cloomber; Arthur (AL=A80 JL=L172 AJ=A184)

Micah Clarke; Arthur (AL=A279 JL=L252 AJ=A424)
The Firm of Girdlestone; Arthur (AL=A133 JL=L257 AJ=A285)
The Ring of Thoth; Arthur (AL=A240 JL=L215 AJ=A372)
A Physiologist's Wife; Arthur (AL=A58 JL=L50 AJ=A60)
The Surgeon of Gaster Fells; Arthur (AL=A298 JL=L328 AJ=A477)
A Pastoral Horror; Arthur (AL=A141 JL=L268 AJ=A329)
The White Company; Arthur (AL=A264 JL=L223 AJ=A387)
Our Midnight Visitor; Arthur (AL=A264 JL=L192 AJ=A381)
A Straggler of '15; Arthur (AL=A280 JL=L34 AJ=A270)
The Voice of Science; Arthur (AL=A95 JL=L106 AJ=A138)
The Colonel's Choice; Arthur (AL=A60 JL=L63 AJ=A91)
A Sordid Affair; **Louise** (AL=L182 JL=L80 AJ=J123)
A False Start; **Louise** (AL=L13 JL=L82 AJ=A37)
Beyond the City: **Louise** (AL=L137 JL=L104 AJ=J73)
The Doings of Raffles Haw; **Louise** (AL=L48 JL=L78 AJ=J7)
Out of the Running; Arthur (AL=A28 JL=L215 AJ=A169)
De Profundis; Arthur (AL=A107 JL=L210 AJ=A227)
A Regimental Scandal; **Louise** (AL=L42 JL=L51 AJ=J47)
A Question of Diplomacy; **Jean** (AL=L43 JL=J67 AJ=J118)
Lot No. 249; Arthur (AL=A112 JL=L162 AJ=A218)
The Fate of Fanella; Arthur (AL=A72 JL=L228 AJ=A190)
Jetland's Voyage; Arthur (AL=A100 JL=L299 AJ=A299)
The Great Shadow; Arthur (AL=A12 JL=L409 AJ=A280)
The Los Amigos Fiasco; Arthur (AL=L107 JL=L143 AJ=J10)
The Refugees; Arthur (AL=A60 JL=L216 AJ=A200)
The Green Flag; Arthur (AL=A454 JL=L398 AJ=A705)

Louise, diagnosed as terminal, moves to Switzerland

The Slapping Sal; Arthur (AL=A642 JL=L507 AJ=A966)
The Case of Lady Sannox; Arthur (AL=108 JL=J102 AJ=A17)

Note: End use of early-Arthur collection
Note: Begin use of intermediate-Arthur collection

The Doctors of Hoyland; **Louise** (AL=L6 JL=L141 AJ=A117)
Sweethearts; Arthur (AL=A159 JL=L135 AJ=A293)
The Lord of Chateau Noir; Arthur (AL=A141 JL=L234 AJ=A294)
The Parasite; **Jean** (AL=A74 JL=J143 AJ=J41)
A Medical Document; Arthur (AL=A41 JL=L197 AJ=A184)
The Third Generation; Arthur (AL=A61 JL=L47 AJ=A102)
The Surgeon Talks; Arthur (AL=A54 JL=L305 AJ=A258)
His First Operation; Arthur (AL=A117 JL=L268 AJ=A319)
The Curse of Eve; Arthur (AL=A71 JL=L293 AJ=A286)

Behind the Times; **Louise** (AL=L28 JL=L311 AJ=A258)
A Foreign Office Romance; Arthur (AL=A59 JL=L192 AJ=A205)
Brigadier Won his Medal; Arthur (AL=A138 JL=L317 AJ=A381)
Recollections of Captain Wilke; **Louise** (AL=L261 JL=L302 AJ=A2)
Brigadier Held the King; Arthur (AL=A129 JL=L290 AJ=A347)
The King Held the Brigadier; Arthur (AL=A190 JL=L325 AJ=A427)
Brigadier Slew the Brothers; Arthur (AL=A146 JL=L217 AJ=A320)
Brigadier Came to the Castle; Arthur (AL=A97 JL=L181 AJ=A245)
Brigadier Took the Field; Arthur (AL=A88 JL=L96 AJ=A150)
The Stark Munro Letters; **Louise** (AL=L8 JL=L190 AJ=A142)
Brigadier was Tempted; Arthur (AL=A159 JL=L306 AJ=A349)
Brig Played for a Kingdom; Arthur (AL=A172 JL=L121 AJ=A278)
Rodney Stone; Arthur (AL=A159 JL=L148 AJ=A263)

Event: Louise survives and returns to England

The Three Correspondents; Arthur (AL=A125 JL=L136 AJ=A221)
The Field Bazaar; **Ambiguous** (AL=A44 JL=L35 AJ=J10)
Uncle Bernac; Arthur (AL=A112 JL=L41 AJ=A140)
Governor of Saint Kitts; Arthur (AL=A248 JL=L183 AJ=A341)

Event: Arthur begins affair with Jean Leckie

Dealings of Captain Sharkey; Arthur (AL=A355 JL=L124 AJ=A415)
The Voyage of Copley Banks; Arthur (AL=A358 JL=L255 AJ=A509)
The Tragedy of the Korosko; Arthur (AL=A140 JL=L106 AJ=A210)
The Striped Chest; Arthur (AL=A96 JL=L247 AJ=A278)
Fiend of the Cooperage; Arthur (AL=A41 JL=J14 AJ=A37)
The New Catacomb; Arthur (AL=A66 JL=L80 AJ=A127)
The Confession; Arthur (AL=A459 JL=L83 AJ=A478)
The Beetle Hunter; Arthur (AL=A79 JL=L137 AJ=A169)
The Man with the Watches; Arthur (AL=A107 JL=L90 AJ=A176)
The King of the Foxes; Arthur (AL=A174 JL=L421 AJ=A470)
The Lost Special; **Louise** (AL=L32 JL=L163 AJ=A100)
The Sealed Room; **Jean** (AL=A3 JL=J44 AJ=J52)
The Black Doctor; Arthur (AL=A50 JL=L147 AJ=A156)
The Club-Footed Grocer; Arthur (AL=A123 JL=L169 AJ=A244)
A Shadow Before; Arthur (AL=A234 JL=L40 AJ=A249)
The Story of the Brazilian Cat; Arthur (AL=A151 JL=L112 AJ=A223)
The Retirement of Signor Lambert Arthur (AL=A93 JL=L4 AJ=A103)
The Japanned Box; Arthur (AL=A101 JL=J48 AJ=A88)
The Jew's Breast-Plate; Arthur (AL=A90 JL=L98 AJ=A222)
The Story of B.24; Arthur (AL=A120 JL=L149 AJ=A222)
A Duet, with an Occasional Chorus; **Jean** (AL=L44 JL=J76 AJ=J7)

The Story of the Latin Tutor; Arthur (AL=A82 JL=L59 AJ=A117)
The Story of the Brown Hand; Arthur (AL=A136 JL=L40 AJ=A150)
The Croxley Master; Arthur (AL=A191 JL=L73 AJ=A241)
The Crime of the Brigadier; Arthur (AL=A279 JL=L45 AJ=A297)
The Debut of Bimbashi Joyce; Arthur (AL=A69 JL=L115 AJ=A170)
Playing with Fire; Arthur (AL=A128 JL=J136 AJ=A39)
The Impression of the Legacy; Arthur (AL=A339 JL=L134 AJ=A439)
Coming of the Huns; Arthur (AL=A334 JL=L268 AJ=A497)
The First Cargo; Arthur (AL=A43 JL=L325 AJ=A286)
Brigadier Lost His Ear; Arthur (AL=A335 JL=J29 AJ=A293)
Brigadier Saved the Army; Arthur (AL=A219 JL=L45 AJ=A241)
The Leather Funnel; **Jean** (AL=A93 JL=J133 AJ=J14)
Brigadier Rode to Minsk; Arthur (AL=A190 JL=J32 AJ=A152)
Brigadier Bore Himself; Arthur (AL=A264 JL=L166 AJ=A361)
Brigadier Triumphed in England; Arthur (AL=A113 JL=J96 AJ=J23)
Brigadier Captured Saragossa; Arthur (AL=A140 JL=L77 AJ=A156)
Last Adventure of the Brigadier; Arthur (AL=A213 JL=J43 AJ=A142)
Sir Nigel; Arthur (AL=A257 JL=L50 AJ=A256)

Event: Louise dies
Note: End use of intermediate-Arthur collection
Note: Begin use of late-Arthur collection

The Pot of Caviare; **Jean** (AL=A222 JL=J317 AJ=J9)
The Silver Mirror; **Jean** (AL=L53 JL=J2 AJ=A85)
The Lord of Falconbridge; Arthur (AL=A169 JL=L78 AJ=A173)
The Homecoming; Arthur (AL=A310 JL=J32 AJ=A263)
The Terror of Blue John Gap; Arthur (AL=A84 JL=L150 AJ=A161)
The Marriage of the Brigadier; Arthur (AL=83 JL=J138 AJ=A3)
The Lift; Arthur (AL=A324 JL=J148 AJ=A176)
The Last Galley; Arthur (AL=A510 JL=L2 AJ=A472)
The Last of the Legions; Arthur (AL=A342 JL=J139 AJ=A248)
The Red Star; Arthur (AL=A128 JL=L159 AJ=A184)
The Contest; Arthur (AL=A212 JL=J43 AJ=A182)
The Blighting of Sharkey; Arthur (AL=A425 JL=L19 AJ=A404)
Through the Veil; Arthur (AL=A291 JL=J80 AJ=A224)
An Iconoclast; Arthur (AL=A212 JL=J70 AJ=A161)
Giant Maximin; Arthur (AL=A437 JL=L33 AJ=A447)
One Crowded Hour; Arthur (AL=A246 JL=L217 AJ=A374)
The Lost World; Arthur (AL=A79 JL=L84 AJ=A115)
The Fall of Lord Barrymore; Arthur (AL=A199 JL=J212 AJ=A74)
The Poison Belt; Arthur (AL=A45 JL=L41 AJ=A64)
How it Happened; **Louise** (AL=L31 JL=L25 AJ=J46)

Borrowed Scenes; Arthur (AL=A108 JL=L117 AJ=A158)
The Horror of the Heights; Arthur (AL=A30 JL=L172 AJ=A137)
Danger!; Arthur (AL=A166 JL=L268 AJ=A322)
The Prisoner's Defense; **Jean** (AL=L150 JL=J59 AJ=J163)
Three of Them; Arthur (AL=A188 JL=J98 AJ=A138)
A Point of View; **Jean** (AL=L24 JL=J78 AJ=J43)
The Bully of Brocas Court; Arthur (AL=A312 JL=L64 AJ=A297)
The Nightmare Room; **Jean** (AL=A102 JL=J263 AJ=J52)
The Centurion; **Jean** (AL=A52 JL=J189 AJ=J49)
A Point of Contact; Arthur (AL=A271 JL=J37 AJ=A220)
Billy Bones; Arthur (AL=A288 JL=L171 AJ=A370)
How Watson Learned the Trick; **Jean** (AL=L269 JL=J332 AJ=J548)
The Land of Mist; **Jean** (AL=A96 JL=J165 AJ=J19)
The Maracot Deep; Arthur (AL=A101 JL=L127 AJ=A164)
When The World Screamed; Arthur (AL=A155 JL=L54 AJ=A161)
The Disintegration Machine; **Jean** (AL=A88 JL=J204 AJ=J65)
The Death Voyage; Arthur (AL=A277 JL=J37 AJ=A255)

Event: Arthur Dies

The End of Devil Hawker; Arthur (AL=A169 JL=J68 AJ=A106)
The Last Resource; Arthur (AL=A295 JL=L56 AJ=A279)
The Parish Magazine; **Jean** (AL=L75 JL=J171 AJ=J193)

APPENDIX XII
Calculated Authorship of the Sherlock Holmes Adventures

Event: Louise marries Arthur
Note: Begin use of the early-Arthur collection

A Study in Scarlet; Arthur (AL=A93 JL=L208 AJ=A228)
A Study in Scarlet (Sherlock Holmes chapters only)
  Part 1, Chapter 1; Arthur (AL=A203 JL=L39 AJ=A193)
  Part 1, Chapter 2; Arthur (AL=A12 JL=L110 AJ=A123)
  Part 1, Chapter 3; Arthur (AL=A87 JL=A52 AJ=A94)
  Part 1, Chapter 4; **Louise** (AL=L29 LJ=L300 AJ=A187)
  Part 1, Chapter 5; **Louise** (AL=L32 LJ=L58 AJ=J9)
  Part 1, Chapter 6; Arthur (AL=A113 LJ=L107 AJ=A170)
  Part 1, Chapter 7; **Louise** (AL=L46 LJ=L107 AJ=A32)
  Part 2, Chapter 6; **Louise** (AL=L127 LJ=L315 AL=A82)
  Part 2, Chapter 7; **Louise** (AL=L236 LJ=L18 AL=A86)
A Study in Scarlet (Utah narrative only)
  Part 2, Chapter 1; Arthur (AL=A351 LJ=312 AJ=A498)
  Part 2, Chapter 2; Arthur (AL=A438 LJ=315 AJ=A640)
  Part 2, Chapter 3; **Jean** (AL=A13 LJ=J8 AJ=J10)
  Part 2, Chapter 4; Arthur (AL=A211 LJ=416 AL=A533)
  Part 2, Chapter 5; Arthur (AL=A287 LJ=L527 AL=A632)

The Sign of Four; **Louise** (AL=L52 JL=L44 AJ=A39)

Note: The next 12 stories are usually anthologized as
The Adventures of Sherlock Holmes

A Scandal in Bohemia; **Louise** (AL=L155 JL=L235 AJ=A7)
The Red-Headed League; **Louise** (AL=L191 JL=L206 AJ=J35)
A Case of Identity; **Louise** (AL=L344 JL=L155 AJ=J238)
Boscombe Valley Mystery; **Louise** (AL=L222 JL=L185 AJ=J99)
The Five Orange Pips; **Louise** (AL=L71 JL=L105 AJ=J11)
Man with the Twisted Lip; **Louise** (AL=L119 JL=L246 AJ=A69)
The Blue Carbuncle; **Louise** (AL=L248 JL=L126 AJ=J115)
The Speckled Band; **Louise** (AL=L155 JL=L187 AJ=J1)
The Engineer's Thumb; **Louise** (AL=L136 JL=L283 AJ=A56)
The Noble Bachelor; **Louise** (AL=L331 JL=L228 AJ=J176)
The Beryl Coronet; **Louise** (AL=L335 JL=L155 AJ=J211)
The Copper Beeches; **Louise** (AL=L245 JL=L163 AJ=J131)

Note: The next 12 stories are usually anthologized as

The Memoirs of Sherlock Holmes

Silver Blaze; **Louise** (AL=L66 JL=L189 AJ=A41)
The Cardboard Box; **Louise** (AL=L171 JL=L217 AJ=J18)
The Yellow Face; **Louise** (AL=L319 JL=L246 AJ=J107)
The Stockbroker's Clerk; **Louise** (AL=L240 JL=L214 AJ=J78)
The Gloria Scott; **Louise** (AL=L158 JL=L316 AJ=A89)
The Musgrave Ritual; **Louise** (AL=L145 JL=L250 AJ=A39)
The Reigate Squire; **Louise** (AL=L212 JL=L249 AJ=J1)
The Crooked Man; **Louise** (AL=L198 JL=L286 AJ=A39)
The Resident Patient; **Louise** (AL=L139 JL=L236 AJ=A31)
The Greek Interpreter; **Louise** (AL=L105 JL=L272 AJ=A91)
The Naval Treaty; **Louise** (AL=L212 JL=L187 AJ=J73)
The Final Problem; **Louise** (AL=L142 JL=L298 AJ=71)

Event: Louise's tuberculosis diagnosed as possibly terminal
Event: Louise moves to Switzerland
Note: End use of early-Arthur collection
Note: Begin use of intermediate-Arthur collection

The Hound of the Baskervilles; **Louise** (AL=L20 JL=L19 AJ=J11)

Note: The next 13 stories are usually anthologized as
The Return of Sherlock Holmes

The Empty House; Arthur (AL=A61 JL=L70 AJ=A91)
The Norwood Builder; **Jean** (AL=L55 JL=J109 AJ=J158)
The Dancing Men; **Louise** (AL=L95 JL=L22 AJ=J97)
The Solitary Cyclist; **Louise** (AL=L67 JL=L83 AJ=J9)
The Priory School; **Jean** (AL=L50 JL=J158 AJ=J190)
Black Peter; **Jean** (AL=A16 JL=J185 AJ=J132)
Charles Augustus Milverton; Arthur (AL=A7 JL=J9 AJ=A14)
The Six Napoleons; **Jean** (AL=L94 JL=J16 AJ=J111)
The Three Students; **Jean** (AL=L225 JL=J120 AJ=J337)
The Golden Pince-Nez; **Jean** (AL=L33 JL=J146 AJ=J152)
The Missing Three-Quarter; **Jean** (AL=L77 JL=J57 AJ=J140)
The Abbey Grange; **Jean** (AL=L13 JL=J53 AJ=J47)
The Second Stain; **Jean** (AL=L69 JL=J317 AJ=J321)

Event: Louise dies
Note: End use of the intermediate-Arthur collection
Note: Begin use of the late-Arthur collection
Note: The next 7 stories are usually anthologized as *His Last Bow*

Wisteria Lodge; **Jean** (AL=L77 JL=J116 AJ=J149)
The Bruce-Partington Plans; **Jean** (AL=L16 JL=J355 AJ=J275)
The Devil's Foot; **Jean** (AL=L57 JL=J78 AJ=J127)
The Red Circle; **Jean** (AL=A11 JL=J174 AJ=J100)
Lady Frances Carfax; **Jean** (AL=L47 JL=J267 AJ=J238)
The Dying Detective; **Jean** (AL=L24 JL=J309 AJ=J213)
His Last Bow; **Jean** (AL=A63 JL=J188 AJ=J88)

The Valley of Fear; **Jean** (AL=A3 JL=J119 AJ=J77)

Note: The next 13 stories are usually anthologized as
The Casebook of Sherlock Holmes

The Mazarin Stone; **Jean** (AL=L76 JL=J296 AJ=J259)
The Problem of Thor Bridge; **Jean** (AL=L108 JL=J259 AJ=J277)
The Creeping Man; **Jean** (AL=L18 JL=J253 AJ=J183)
The Sussex Vampire; **Jean** (AL=L6 JL=J193 AJ=J126)
The Three Garridebs; **Jean** (AL=L99 JL=J218 AJ=J236)
The Illustrious Client; **Jean** (AL=L70 JL=J238 AJ=J228)
The Three Gables; **Jean** (AL=L69 JL=J343 AJ=J292)
The Blanched Soldier; **Jean** (AL=L126 JL=J111 AJ=J185)
The Lion's Mane; **Jean** (AL=A5 JL=J167 AJ=J127)
The Retired Colourman; **Jean** (AL=L62 JL=J277 AJ=J236)
The Veiled Lodger; **Jean** (AL=A19 JL=J121 AJ=J75)
Shoscombe Old Place; **Jean** (AL=A24 JL=J173 AJ=J98)

Event: Arthur dies