

The Case of Sherlock Holmes and Linguistic Analysis

Introduction

Sherlock Holmes, in all of his media incarnations, complicates the idea of the forensic scientist: he preserves and contaminates crime scenes, adheres to and violates contemporary industry ethics, and—most famously—eschews standard investigative policies in favour of abductive reasoning (mislabelled widely as ‘deduction’). The vagaries of Holmes’s engagement with scientific and police procedures are mostly the result of narrative necessity; outlandish crimes often require outlandish solutions, and a detective known for his novel and rebellious methods is undoubtedly more compelling in fiction than a detective of average abilities bogged down in the red-tape and paperwork of reality. In most of his forms, Holmes works grudgingly with the police (who are frequently portrayed as incompetents) to enhance the efficiency of their scientific investigative procedures (which are frequently portrayed as antiquated and ineffective).

It is in the original Sir Arthur Conan Doyle canon of short stories and novels, however, that Sherlock Holmes is perhaps the most indebted to the very structures he seems to undermine or improve. The solutions to some of these cases hinge upon innovation that are presumably of Holmes’s own design, or by knowledge possessed only by him. Holmes himself brags in *A Study in Scarlet*:

I have written a monograph upon the subject. I flatter myself that I can distinguish at a glance the ash of any known brand, either of cigar or of tobacco. It is just in such details that the skilled detective [i.e. Holmes himself] differs from the Gregson and Lestrade type [i.e. the average police detective].¹

Indeed, his analysis of cigar ash provides vital breakthroughs (or is at least referenced) in many of the Conan Doyle texts and in later adaptations.² In reality, however, much of the scientific technique and knowledge attributed to Holmes in Conan Doyle’s texts are actually just various early or even established forms of forensic science which were utilised by

contemporaneous police departments in Britain.³ These are generally passed off to the reader as the brain-children of Holmes alone, as he eschews the more stringent methodologies, procedures, and ethical practices that would be required of police officers and other investigators.

There is, however, one element of forensic science that *was* truly innovative on the part of Conan Doyle in the *Sherlock Holmes* canon: representation of what we would now call the field of forensic linguistics. This article argues that—despite his loose understanding and acknowledgement of established forensic science at the time of his writing—Conan Doyle at the *fin de siècle* somewhat anticipated the development of forensic linguistics, or at least some practices in and elements of forensic linguistics, roughly eighty years before the field was identified. Although Conan Doyle's understanding of the incipient field is nowhere near perfectly aligned with its actual practices and developments, his anticipation of and engagement with it is one of the few (if not the sole) scientific innovations that could reasonably be ascribed to Sherlock Holmes. This also serves to demonstrate that at the *fin de siècle*, in a society interested in advancements in justice and forensic investigations, there was a credible societal belief that linguistic analysis can be useful in forensic contexts.

Much as its name suggests, forensic linguistics is the application of linguistic knowledge to a forensic context; this can range from analysing the language in which a law is written, to studying what detainees understand when they are read their legal rights, to examining the language of threatening writing to attempt to determine its authorship. Forensic linguistics, as will be explained more fully below, is a relatively new and broad discipline with an ongoing evolution. It had previously been argued that forensic linguistics should be considered a form of applied linguistics rather than a distinct field, although the increasing prevalence of application in the legal and civil spheres means that it has gained acceptance as its own discipline.⁴ As will be explored, the popularity of Conan Doyle's

portrayals in the 1890s and 1900s represent underlying precepts of forensic linguistics, which have fed into the shaping of the field in the latter half of the twentieth century.

This article takes an interdisciplinary approach to the *Sherlock Holmes* canon to interrogate Conan Doyle's engagement with and occasional rejection of the scientific process in his development and representation of forensic linguistics. Further, this particular dialogue between fiction and the scientific process is characteristic of Sherlock Holmes as a character, whose own intellectual processes are augmented by obfuscation, theatricality, and individuality. Much as with his near-superhuman protagonist, Conan Doyle imbues forensic linguistics with preternatural abilities and a heightened narrative potential far beyond the realms of possibility—a practice that continues in art and media to this day and has very real consequences for actual forensic linguistic analysis.

This article will analyse five short stories from the *Sherlock Holmes* canon as case studies: “A Scandal in Bohemia” (1891), “The Man with the Twisted Lip” (1891), “The Boscombe Valley Mystery” (1891), “The Adventure of the Reigate Squire” (1893), and “The Adventure of the Dancing Men” (1903). Early elements of forensic linguistics appear in a great many more *Sherlock Holmes* stories, but these five stories in particular illustrate Conan Doyle's innovation surrounding language and the detective process. The first four stories were written in the early 1890s—only about five years after Sherlock Holmes made his first appearance in literature in *A Study in Scarlet* (1887)—and they demonstrate Conan Doyle's interest in the relationship between language and criminal investigation. The last story, “The Adventure of the Dancing Men”, was written more than a decade later and speaks to the expansion and crystallisation of Conan Doyle's development of this branch of linguistics.

In particular, these short stories were selected not only because of the depth of their engagement with forensic linguistics (in that the understanding of language is a major feature

of the narrative, or is integral to unravelling the mystery) but also because they present a range of adjacent fields, sub-specialisms, issues, or offshoots of the field itself, illustrating that Conan Doyle also anticipated the sheer breadth of the discipline as we know it today. In particular, “A Scandal in Bohemia” deals with sociolinguistic profiling and Native Language Influence Detection (NLID), “The Man with the Twisted Lip” and “The Adventure of the Reigate Squire” deal with handwriting analysis (which is a distinct, separate field from forensic linguistics, but has relevance as an adjacent field), “The Boscombe Valley Mystery” deals with lexical priming, while “The Adventure of the Dancing Men” deals with semiotics and the linguistic significance of context.⁵ Each of these will be addressed in turn below as the stories are analysed individually.

Forensic Linguistics: A History

Before an analysis of Conan Doyle’s short stories can be undertaken, we must first provide a potted history of forensic linguistics in order to better clarify what Conan Doyle may have known given the state of the field at the time of his writing, how he deviates from that contemporaneous knowledge, and what a modern application of forensic linguistics can illustrate. Although much has been made of Conan Doyle’s early training as a medical doctor and his integration of this scientific knowledge into his works of fiction, there is no known formal connection between Conan Doyle and the linguistics field.⁶ It must be assumed that any predictions he makes about the development of forensic linguistics through his writing is accidental—a by-product of his professional interests in language, crime, and the scientific process.

The formal history of forensic linguistics is a recent one, and was certainly not recognised during Conan Doyle’s lifetime: it first gained an established, clear identity roughly seventy-to-eighty years after the publication of the stories examined in this article,

and nearly forty years after Conan Doyle's death. One can trace elements of what is now understood to be the field of linguistics back millennia: we see examples of it with shibboleth in the Biblical story of Jephthah and the Ephraimites, with Plato's dialogue *Cratylus*, or with the works of Al-Kindi, the ninth-century Arab philosopher, to name just a few.⁷ But linguistics and the study of language, as with the above examples and for much of its history, was largely subsumed into other fields, including philosophy, ancient and classical philology, and the developments of logic, rhetoric, and grammar.⁸ By the 1880s, however, when Conan Doyle began writing his *Sherlock Holmes* stories, the field of applied linguistics was at least partially established and operating, if in its infancy.⁹ The field continued to develop and solidify during the forty years that Conan Doyle wrote the series and it would go on to gain traction in the early twentieth century with the works of Ferdinand de Saussure, considered one of the formal founders of the discipline.¹⁰ The field continued to evolve, with two main approaches developing; prescriptive and descriptive linguistics. The division between them is almost as old as the study of language itself, and can be traced to the Priscian and modistae approaches to grammar from around the twelfth century.¹¹ This division in approaches was largely bred by differences in application, with the Priscian being focused predominantly on the learning of Latin, and the modistae being more focused on the understanding of how grammars are used. In northern Europe the modistae developed into speculative grammarians where the "term *speculative* is based on Latin *speculum* "mirror; image" because speculative grammars sought to mirror reality".¹² Holmes' analysis seems to be more in-keeping with the descriptivist approach which is inherent to forensic linguistics.

Forensic linguistics is a sub-field of interdisciplinary field of applied linguistics. The term '*Applied Linguistics*' emerged in the USA in the 1940s when linguistic analysis was used to solve practical language teaching problems.¹³ The field evolved to include using linguistic knowledge and theory to tackle a wide range of real-world problems.¹⁴ Forensic

linguistics could be considered to be a sub-area of applied linguistics, focusing loosely on language law and criminal contexts, though it is now largely considered its own field.¹⁵ It was with the publication of Swedish author and scholar Jan Svartvik's work *The Evans Statements: A Case for Forensic Linguistics* (1968), that forensic linguistics made its first strides towards becoming a distinct field.¹⁶ As will be explored in the literature analyses below, many of these fractures and applications—and, indeed, the necessity of such fractures and applications—were predicted by Conan Doyle well in advance of the maturation of the linguistic field.

“A Scandal in Bohemia”, Sociolinguistic Profiling, and Native Language Influence Detection

“A Scandal in Bohemia”, perhaps one of the more famous *Sherlock Holmes* cases, is the only one in which Holmes is outsmarted and unable to solve his case successfully. Holmes is approached by the King of Bohemia to retrieve a photograph (implied to be sexual in nature) of the king with his former lover, Irene Adler. Adler, who is in possession of the photograph, intends to blackmail the king with it if his scheduled diplomatic marriage to a Scandinavian princess goes forward. Despite Holmes's best attempts, the photograph remains in the hands of Adler, with the blackmail deferred for the moment but not ruled out as a possibility.

It is in this *Sherlock Holmes* short story that forensic linguistics is perhaps most clearly utilised through the methodology of sociolinguistic profiling. Sociolinguistic profiling, which is arguable the most commonly represented area of forensic linguistics in fiction, is a technique used when investigators are presented with an anonymous text, but no key list of suspects. In other words, investigators have no material written by known suspects to compare with the anonymous text in order to identify more definitive authorship. Instead, the investigator identifies linguistic features in that anonymous text in order to develop (as

the name suggests) a profile of the potential social characteristics of the author.¹⁷ Based on the linguistic conditions of the anonymous text, the linguistic analyst can assess the likelihood of the writer belonging to a certain gender, age group, class, nationality, ethnicity, or determine their job, other languages spoken, and more; in short, an individual's spoken and written language is influenced by social factors and their background.¹⁸ When developing a sociolinguistic profile, the investigator must essentially reverse engineer this process to predict the likely social influences based on the linguistic features in the anonymous writing samples provided.

Sociolinguistic profiling appears clearly and early—but only briefly—in “A Scandal in Bohemia”: Holmes receives an anonymous letter requesting help and, in his usual grandstanding fashion, is able to predict the exact identity of the author (the King of Bohemia, who further attempts to obfuscate his identity by arriving disguised in a mask) largely through the author's writing style. The note reads:

There will call upon you to-night, at a quarter to eight o'clock [...] a gentleman who desires to consult you upon a matter of the very deepest moment. Your recent services to one of the royal houses of Europe have shown that you are one who may safely be trusted with matters which are of an importance which can hardly be exaggerated. This account of you we have from all quarters received. Be in your chamber then at that hour, and do not take it amiss if your visitor wear a mask.¹⁹

It's important to note that Holmes and Watson take other non-linguistic conditions of the letter into account in the development of their profile (the writer's class is guessed by the luxury of the paper; the writer's country of origin is guessed by the watermark monogram of the paper company). However, the third and most definitive characteristic is linguistic analysis through-and-through—Holmes uses what we now would call Native Language Influence Detection (NLID), which is an element of sociolinguistic profiling and can loosely be defined as the detection of an author's native language from the way they write in a second language. Holmes says:

And the man who wrote the note is a German. Do you not the peculiar construction of the sentence—‘This account of you we have from all quarters received.’ A Frenchman or Russian could not have written that. It is the German who is so uncourteous to his verbs. It only remains, therefore, to discover what is wanted by this German who writes upon Bohemian paper and prefers wearing a mask to showing his face.²⁰

Through this quotation, Conan Doyle anticipates NLID by several decades and illustrates that there is, and perhaps always has been, an innate social understanding that one can identify a person’s native language (in this case, German) from the way they use a second language (in this case, English). Sherlock Holmes performs a rudimentary comparative analysis, basing his conclusions on the positioning of one verb part. He identifies that in German the verb is split, that the past participle becomes the final item in the sentence; the splitting of the verb is a well-known difference between English and German, for which one only needs a very rudimentary knowledge of German to understand.²¹ Leaving aside the now archaic lexical choices, it would be more fluent to say in English, ‘We have received this account of you from all quarters’.²²

The analysis that Holmes performs is not a key component in solving the case, as it serves neither to convict nor even to identify a suspect, nor to drive the momentum of the case forward in any way. The king, though masked, introduces himself to Holmes as a Bohemian nobleman, instantly confirming (and rendering irrelevant) Holmes’s NLID and part of his sociolinguistic profile. Further, it is not the king’s identity, nor even that of the guilty party, that needs identifying: Irene Adler is known to be the culprit from the very beginning of the case and it is only the recovery of the photograph that is the goal, not her arrest and conviction. Further still, the sociolinguistic profiling represented in this short story not only had no bearing on Holmes’s case, but the case itself remained unresolved. The significance of this brief and inconsequential sociolinguistic profiling is to illustrate both the intellectual prowess of Conan Doyle’s protagonist as well as to indicate another potential tool available in police detection. The certainty with which Holmes states his sociolinguistic

profile (and, indeed, the certainty with which Holmes states all of his abductive observations) would be problematic in a modern forensic situation and certainly would not meet the standards required for evidence in most jurisdictions, as we will see in the analysis of further *Sherlock Holmes* short stories below. However, the observations that Holmes makes are perhaps intended by Conan Doyle to be more investigatory rather than evidentiary, which alters the level of certainty required. In investigative situations, observations with lower levels of certainty can be useful in suggesting avenues of investigation, even if those observations could not be upheld in a court of law (this is seen in both psychological and linguistic profiling).²³ Although one would hesitate to measure Conan Doyle's knowledge of both forensic science and law against modern or even contemporaneous standards, such as they were, one of Holmes's greatest assets as a private detective is that he is not beholden to the same ethical measures and standardised procedures as the normal police force. As previously discussed, his ability to rebel against established and rigorous structures, and to use unconventional scientific processes in order to solve problems of varying magnitudes, makes Holmes successful as both a private detective in the reality of his own world and as a compelling protagonist in the reality of Conan Doyle's.

“The Man with the Twisted Lip”, “The Adventure of the Reigate Squire”, and Handwriting Analysis

Handwriting analysis is the site of Conan Doyle's interaction with forensic science in two of his short stories, “The Man with the Twisted Lip” and “The Adventure of the Reigate Squire”, although Conan Doyle relies considerably less on handwriting as a science in the former than he does in the latter. Instead, he uses handwriting analysis in “Twisted Lip” as the locus of confidence that the public tends to place on forensic evidence. Despite the distinct difference between the fields of handwriting analysis and forensic linguistics, there are two key similarities; the data and how issue of expressing certainty levels.²⁴ There is a

long-recognised problem in forensic sciences about how forensic analysis is represented in fiction as unambiguous, infallible, and quickly processed. In a modern context, one of the more notable examples is in TV programmes which centre around crime, such as the *CSI* franchise, where complex and nuanced analysis is reduced to a computer running a few lines of code before the word ‘MATCH’ appears on the screen. In many of these examples, a case is solved on the strength of this single piece of unimpeachable evidence. While evidence from forensic linguistics certainly can play a major part in solving crimes, both contemporarily and historically, it is rarely the only evidence upon which one could make a conviction. Additionally, the analysis of that evidence is also considerably more nuanced than most popular accounts lead the general public to believe. The general standard of good forensic linguistic practice would be to assign *probability*, rather than the certainty, of the authorship of a text in most cases, and to acknowledge that it does not necessarily follow that the authorship of a text guarantees that the author committed a crime.²⁵ This issue of nuance is especially prevalent in terms of handwriting analysis in Conan Doyle’s story “The Man with the Twisted Lip”.

The narrative follows the disappearance of Neville St. Clair, a wealthy businessman, whose wife is convinced that she spotted him in the window of an opium den. Upon police investigation, only a beggar, Hugh Boone, is revealed to be in the room. Evidence is found near Boone that seems to incriminate him for Mr St. Clair’s disappearance and Boone is sent to jail. The mystery stagnates until Mrs St. Clair receives a letter from her husband, in his handwriting and including his wedding ring, telling her that he is safe but giving no details as to his whereabouts. It is revealed that Mr St. Clair and Hugh Boone are the same person—he was never a respectable businessman, but rather a professional beggar who did well enough to pose as a gentleman, start a family, and lead a double life during the working hours.

It is the letter and subsequent handwriting analysis which are most central to unravelling the mystery. The majority of early forensic analysis of communication was focused on handwriting as a determining factor for gathering information about unknown authors, so it is unsurprising that Doyle focuses on graphological elements over more contemporary understandings of linguistic features. Although handwriting analysis—in which it was thought that personality traits, age, and other indicators of an author's background and status could be determined—is now controversial and has been largely discredited in contemporary forensic fields, it was very prevalent in the Victorian period and, indeed, well into the twentieth century.²⁶ It must be noted, however, that neither “Twisted Lip” nor this article touch upon handwriting analysis as a formal discipline.²⁷ Rather, it is Conan Doyle's willingness to inject doubt and nuance into a forensic science realm that is important in his developing engagement with the field.

Upon reading the note sent ostensibly by her missing husband, Mrs St. Clair identifies it as being written in her husband's handwriting. Mrs St. Clair, no doubt clinging to the new evidence out of worry and placing expectations on it that it could not reasonably fulfil, draws a very different conclusion to Holmes's analysis of the situation:

‘And you have no doubt that it is your husband's hand, madam?’

‘None. Neville wrote those words.’

‘And they were posted to-day at Gravesend. Well, Mrs. St. Clair, the clouds lighten, though I should not venture to say that the danger is over.’

‘But he must be alive, Mr. Holmes.’

‘Unless this is a clever forgery to put us on the wrong scent. The ring, after all, proves nothing. It may have been taken from him.’

‘No, no; it is, it is his very own writing!’

‘Very well. It may, however, have been written on Monday and only posted to-day.’

‘That is possible.’

‘If so, much may have happened between.’

‘Oh, you must not discourage me, Mr Holmes’.²⁸

Mrs St. Clair concludes that her husband must still be alive, and on the surface this seems reasonable, as the letter is postmarked from earlier that day. Holmes, however, quickly highlights the fallacy in placing too much weight on the conclusions of analysis. He correctly surmises that identifying Mr St. Clair as the definite author of the note—which Holmes is reticent to do fully, as we can see through his acknowledgement that it may be “a clever forgery to put us on the wrong scent”—does not connect in any way to when the note was written, nor if Mr St. Clair was the one to post it. Despite Conan Doyle’s recognition of the false confidence the general public has in analytical findings, it should be noted that he is first and foremost an author of fiction. As such, Conan Doyle very frequently minimises the uncertainty of conclusions reached by his detective.

“The Adventure of the Reigate Squire”, written two years after “Twisted Lip”, gives much more credence to handwriting analysis. The story focuses on the murder of a coachman, who is found clutching a partial note in his hand. Holmes—a self-declared expert in handwriting analysis, among many other things—quickly realises the note was written by two different authors who alternated words. He creates a handwriting profile of the note writers (one old, one young, and related to each other), by which he is able to focus his investigation on two suspects: the father and son of the household that employed the coachman. By tricking the father into producing a handwriting sample, Holmes confirms his suspicions and the wrongdoers are brought to justice. This links to the question in forensic linguistics (and particularly authorship analysis) of co-authorship. Holmes tricking the father into producing a handwriting sample could also have a parallel in collecting naturally occurring language from a suspect, in order to compare it to the original questioned document to look for comparative features (though in FL this would be linguistic features, rather than graphological ones).

Although Conan Doyle could not have predicted the ultimate rejection of handwriting analysis as a viable scientific field, he does manage to connect Holmes's interrogation of handwriting in "Reigate Squire" to legitimate forensic practices—in this instance, document analysis—to determine that two authors contributed to the authorship of a single text. Although document analysis does not aim to predict personality traits nor give a full author profile (like handwriting analysis claims to do or sociolinguistic profiling does), issues of handwriting are still pertinent to document analysis. For instance, handwriting pressure can give indications about authorship, while the slant of letters can allow an analyst to determine the chances of an author being left- or right-handed.

No doubt aware of the characteristics of handwriting analysis—some of which would stand the test of time better than others—Conan Doyle writes:

You may not be aware that the deduction of a man's age from his writing is one which has been brought to considerable accuracy by experts. In normal cases one can place a man in his true decade with tolerable confidence. I say normal cases, because ill-health and physical weakness reproduce the signs of old age, even when the invalid is a youth. In this case, looking at the bold, strong hand of the one, and the rather broken-backed appearance of the other, which still retains its legibility although the t's have begun to lose their crossing, we can say that the one was a young man and the other was advanced in years without being positively decrepit.²⁹

Although Holmes's certainty about his handwriting profile in "Reigate Squire" runs counter to his brief commentary on false confidence in "Twisted Lip", Conan Doyle nevertheless anticipates a form of analysis that is useful in forensic situations; he has merely chosen the wrong features of interest. In the above extract, Holmes attempts to answer questions that are now far more the province of sociolinguistic profiling; indeed, considerable work *has* been done on predicting a person's age from their writing, but linguistic elements are the focus, rather than handwriting.³⁰

In particular, Conan Doyle touches on the concept of an *ecolect* when his protagonist determines from the handwriting on the note that the two authors are related to one another.

An *ecolect* is used to refer to the language of a small, closed group, such as family members who live together, as the father and son murderers in the story do. *Ecolects* are related to the principles of *sociolect*, in which groups of people use similar language and linguistic features: the idea that an individual has unique ways of using language, which might be informed by social factors, as well as their own history and personal preferences. Although Holmes's determination that the authors of the note share a familial bond is something that *does* have a foundation in forensic science, he erroneously attributes the discovery of this bond to handwriting (which forms no part of *ecolect* analysis today) instead of to linguistics (which is the sole determining characteristics of *ecolects*).

Conan Doyle also touches on a particular methodology in forensic science called *elicitation*, although, again, we consider this from the context of forensic linguistics, rather than handwriting analysis. *Elicitation* in this context is the gathering of new data (for example a handwriting sample, or language data, which has been written specifically for this purpose) from a particular group of people, which is then used for comparative purposes.³¹ Conan Doyle writes, "I managed, by a device which had perhaps some little merit of ingenuity, to get old Cunningham to write the word 'twelve,' so that I might compare it with the 'twelve' upon the paper [in the dead coachman's hand]".³² *Elicitation* has been used in a few forensic cases over the years, although its validity in such cases is now contested and largely avoided, with 'naturally occurring' data being preferred in most contexts as more linguistically valid.³³

Eliciting comparison data is considered problematic in a forensic case because it relies upon the target person or group being unaware of why a language sample is required, or how it will be used – and in particular what linguistic feature might be analysed. If the targeted person or group suspected what the investigator was looking for, they might consciously or unconsciously alter their language. Conan Doyle seems aware of the *observer's paradox*, which indicates that an individual will change their behaviours (and in

this case their language) when they are being observed. He builds this into Holmes's investigative methodology by having Holmes misdirect the suspect by getting the suspect to write the word 'twelve' in circumstances that are seemingly innocuous and disconnected from the murder investigation. Although the ruse works and eventually leads to a confession of murder, it would likely *take* a confession of murder for such a case to be solved and prosecuted in today's forensic environment in which such shaky methodology would never stand up to the rigours of a murder trial.

"The Adventure of the Dancing Men", Semiotics, and the Significance of Context

Conan Doyle's "The Adventure of the Dancing Men" (1903) deals with the study of semiotics, or symbols and what they signify. The field of semiotics often runs parallel to, but is distinct from, linguistics in that it investigates communication but deals with specifically non-linguistic signs and methods of communication).³⁴

The "Dancing Men" case begins when a concerned husband starts to find drawings of stick figure men in various poses (giving the appearance of dancing) around his property. These drawings terrify his wife and allude to a dangerous episode in her past, of which she refuses to speak. The husband's lack of faith in the police leads him to approach Holmes to solve the case. The drawing is revealed to be a monoalphabetic substitution cypher disguised as an innocuous children's drawing increasing the chance that it is overlooked by the casual observer, with only the writer and intended reader(s) aware of the code: "the object of those who invented the system has apparently been to conceal that these characters convey a message, and to give the idea that they are the mere random sketches of children".³⁵ Holmes himself says, "At first sight it would appear to be some childish prank. It consists of absurd little figures dancing across the paper upon which they are drawn. Why should you attribute any importance to so grotesque an object?"³⁶ However, the communicative importance of the

drawings is demonstrated to the husband through the reaction they produced in his wife, who clearly understood the cipher and was hence likely the intended reader of the message.

Holmes's investigation reveals that the drawings were left by the wife's former fiancé, who had been a member of her father's gang, the organisation from where this particular cipher originated. The narrative ends with the former fiancé killing the husband, seriously wounding the wife, and being arrested after Holmes's timely cracking of the cipher.

Conan Doyle again anticipates much later linguistic research in his application of a monoalphabetic substitution cypher to a gang context. Gangs have long been linked to specific vernaculars, often termed 'argot', that serve as a sort of code in order to help obfuscate plans and identities from observers outside that gang; this intentionally limits the audience to people from a closed sociocultural group and links language to identity.³⁷ Conan Doyle's use of this cipher not only expands further on linguistic and semiotic tools in a forensic context, but also helps to build his characterisation through the idea of language as identity. The wife's early classification as an innocent victim in this case is complicated through her former in-group status with her father, ex-fiancé, and the gang, and her withholding of knowledge from her new social in-group: her husband and the police. The wife, herself, explains to the husband during their courtship:

I have had some very disagreeable associations in my life [...] If you take me, Hilton, you will take a woman who has nothing that she need be personally ashamed of; but you will have to be content with my word for it, and to allow me to be silent as to all that passed up to the time when I became yours.³⁸

Although terrorised and threatened by a criminal, the wife continues to protect linguistic knowledge from outside members; this linguistic protection places her not only in a liminal space in terms of her social status (she is both a respectable English wife and an American gang affiliate) but also in a liminal space in terms of actual legality (her silence abets a

wanted criminal and leads directly to a murder and assault, even though she is also the expressed intended victim of the crime she is abetting).

With significant overlap between the wife's guilt and innocence, and between the child-like and sobering nature of the drawings, Holmes is brought in to bring clarity to the situation. Here the process of linguistic analysis is better explored—or at least given more space—than in other Holmes narratives. Holmes says:

I am fairly familiar with all forms of secret writings, and am the author of a trifling monograph upon the subject, in which I analyse one hundred and sixty separate ciphers; but I confess that this is entirely new to me. The object of those who invented the system has apparently been to conceal that these characters convey a message, and to give the idea that they are the mere random sketches of children. Having once recognised, however, that the symbols stood for letters, and having applied the rules which guide us in all forms of secret writings, the solution was easy enough. The first message submitted to me was so short that it was impossible for me to do more than to say with some confidence that the symbol



stood for E. As you are aware, E is the most common letter in the English alphabet, and it predominates to so marked an extent that even in a short sentence one would expect to find it most often. Out of fifteen symbols in the first message four were the same, so it was reasonable to set this down as E. It is true that in some cases the figure was bearing a flag and in some cases not, but it was probable from the way in which the flags were distributed that they were used to break the sentence up into words. I accepted this as a hypothesis, and noted the E was represented by



But now came the real difficulty of the enquiry. The order of the English letters after E is by no means well marked, and any preponderance which may be shown in an average of a printed sheet may be reversed in a single short sentence. Speaking roughly, T, A, I, N, S, H, R, D, and L are the numerical order in which letters occur; but T, A, O, and I are very nearly abreast of each other and it would be an endless task to try each combination until a meaning was arrived at. I, therefore, waited for fresh material. In my second interview with Mr. Hilton Cubitt he was able to give me two other short sentences and one message, which appeared – since there was no flag – to be a single word.³⁹

Holmes then goes on at length to walk the reader through his subsequent decryption of the messages to Mrs. Cubitt, revealing that the notes said, “ELSIE COME”, “AM HERE ABE SLANEY”, and, “ELSIE PREPARE TO MEET THY GOD”.⁴⁰

Holmes’s assertion that he is an expert on the field is perhaps wishful thinking or ignorance on the part of Conan Doyle, as this assertion does not match completely with Holmes’s own observations. An expert in cryptology—even in 1903—would suspect a simple letter replacement cypher early on, if not immediately, when presented with a series of figures that seemed to have the communicative purpose and properties of a note; letter replacement is one of the oldest cyphers and tend to be relatively simple to decode.⁴¹ What Holmes also engages in during this passage is something called ‘letter frequency analysis’, which is not strictly linguistic in the sense that we understand the field today, but it *does* rely on linguistic principles and has several linguistic applications— as well as being, as seen in “The Dancing Men”, a key component in the field of cryptography. The frequencies of letters cited by Holmes are based on long established patterns and remained similar to those in Robert Edward Lewand’s *Cryptological Mathematics*.⁴² This is worth noting because, as Lewand argues, the letter frequency rankings vary depending on the dataset you take into account indicating that Holmes (via Conan Doyle) chose a reliable source to base his analysis on.⁴³ So, for example, analysis of the *Concise Oxford Dictionary* in 2012 determined that, unlike Holmes’s letter frequency rankings, the letters of the alphabet occur in the following order: E, A, R, I, O, T, N, S, L, C, U, D, P, M, H, G, B, F, Y, W, K, V, X, Z, J, Q. However, the problem with this study—which Lewand argues and which Conan Doyle seems to have understood—is that the dictionary is not representative of standard English usage. The frequency of letters in a list of every word in the English language is not the same as understanding the frequency of letters when factoring in the frequency of *word*-usage, as well. In particular, a dictionary would only list each pronoun once, when in reality the sheer

amount of pronoun usage in spoken and written English would bump those letters used to a higher placement in the frequency order. As this type of cypher was in extremely common usage in the nineteenth century and, indeed, long before, it is not unusual that Conan Doyle would have had such a detailed understanding of this concept of both letter- and word-usage.⁴⁴

Despite Holmes's initial and unlikely confusion about the nature of the cypher, he not only provides a surprisingly accurate and astute letter frequency analysis, but also illustrates the need for sufficient data in linguistic analysis. When initially contacted by Mr Cubitt, Holmes is given a short letter of fifteen characters. Given Holmes's frequent intellectual grandstanding—often through the outlandish parsing of very small clues—it would have been easy from a narrative perspective (and in keeping with Holmes's established character traits) for Conan Doyle to have Holmes crack the cypher from a single fifteen-character message. Whether through Conan Doyle's legitimate engagement with linguistics or through his desire to build narrative tension by stringing out the mystery of the code for longer, Conan Doyle writes in to Holmes's cryptography speech an implicit or explicit understanding of having sufficient data for analysis, which remains relevant in both cryptology and in forensic linguistic analysis. As Holmes correctly states, it would be very difficult for even the most accomplished codebreaker to determine the letter substitutions without sufficient length of text, because the content would have too great an impact on the overall word frequencies. Conan Doyle's use of the note "ELSIE COME" is particularly astute in this context, as there remains a higher rate of 'e' usage, but all the other letters are only used once. A higher volume of text, as Holmes eventually receives over the course of the story, stands a greater chance of normalising the effect of content-exclusive terms.

"The Boscombe Valley Mystery" and Audience Design

One major component of language analysis is the context in which words are spoken; this particularly includes something called audience design, which is when a person's linguistic style changes in response to the audience to whom they are speaking. In Conan Doyle's "The Boscombe Valley Mystery", audience design is centred on a single word: the Australian cry 'cooee', which is a particular call used to attract attention. In "Boscombe Valley", an Australian expatriate, Charles McCarthy, is murdered and his estranged son, James, is the main suspect. Witnesses claim to have seen Charles walk into the woods, followed by an armed James. James confirms the story, but claims he had gone into the woods to hunt and had heard his father call 'cooee', indicating that his father had gone to meet a third party. James states that he then partially overheard his father's murder and last words: 'a rat'. Using James's testimony as a lead, Holmes investigates the woods and finds the footprints of a third party. Holmes abduces that Charles met with another Australian ("cooee" being distinct to Australia) and that his last words, 'a rat', were all that James was able to overhear of his father saying "Ballarat", an Australian city. Holmes tracks down the suspect and extracts a confession of murder, freeing James.

Much of Holmes's case is built on the linguistic significance of 'cooee'; James, as much of an Australian expatriate as his father, states that the word 'was a usual signal between my father and myself', but Holmes finds evidence that Charles had no idea that James was hunting nearby and therefore couldn't possibly be calling to him. Holmes says,

Well, obviously it could not have been meant for the son. The son, as far as he knew, was in Bristol. It was mere chance that he was within earshot. The 'Cooee!' was meant to attract the attention of whoever it was that he had the appointment with. But 'Cooee' is a distinctly Australian cry, and one which is used between Australians. There is a strong presumption that the person whom McCarthy expected to meet him at Boscombe Pool was someone who had been in Australia.⁴⁵

Here again, Holmes seems extremely sure of his appraisal of the situation; there is none of Conan Doyle's commentary on false confidence as seen in "Twisted Lip", as here 'cooee'

and the reference to ‘a rat’ (the end of Ballarat) are easily-solved and perhaps narratively conspicuous clues that lead Holmes directly to another Australian suspect. Indeed, this lack of false confidence even brings about the further modern linguistics issue of lexical priming, in which the audience will usually hear something that relates to what they expect to hear. This is a particularly fraught issue in modern forensic linguistic and policing circles in which there is a great risk that a listener will mishear something which they believe to indicate guilt, based on the expectations or previous knowledge they have of the speaker or the situation in which they overhear the speaker. Both James—and Holmes, based on James’s second-hand testimony—are satisfied that Charles actually called ‘cooe’, based on his status as an Australian expat. While James supposes that ‘a rat’ was meaningless and merely the ravings of his father’s near-death delirium, Holmes far more tenuously (yet correctly) assumes that Charles must have meant ‘Ballarat’, again because Holmes is lexically primed in his knowledge of Charles as an Australian, but does not question that priming. Once again, a modern reader can see that Conan Doyle is more concerned with the mechanics of the plot rather than the realities of linguistics, which serve as occasionally useful set dressing for Holmes’s cases.

Conclusion

In many ways, the utilisation of linguistic analysis in crime fiction is ultimately beneficial for the field of forensic linguistics, no matter how poorly understood or misrepresented it is. Its mere inclusion raises the awareness of the field, but its poor use often can serve to obfuscate the true practices of linguistic analysts; one might worry that representation of forensic linguistics in fiction might lead to more educated perpetrators who can better hide their crimes (the television show *CSI* has been credited as an influence for a higher percentage of perpetrators of premeditated crimes wearing gloves to hide their fingerprints, for example.⁴⁶ However, in terms of forensic linguistics it is likely that perpetrators of

crimes would only alter surface-level language features, which are usually not of much interest to forensic linguistic analysis.⁴⁷

Although Conan Doyle's engagement with linguistic analysis was, itself, mostly interested in only surface-level language features, he contributed to a significant dialogue between the arts and the forensic sciences at the fin de siècle. The sheer popularity of the *Sherlock Holmes* canon and the multiplicity and variety of Conan Doyle's interrogation of forensic linguistic fields has led to linguistics' further crystallisation in the minds of the public as a viable and necessary method of investigation. Indeed, forensic linguistics is the perfect intersection of literature and detection, and it is arguably through Conan Doyle's appreciation of both, that the field was better able to solidify as its own discipline and become a hallmark of crime fiction to the present day.

¹ Sir Arthur Conan Doyle, *A Study in Scarlet* (London: Ward Lock & Co, 1887), 33-34.

² Of the *Sherlock Holmes* canon by Conan Doyle, see *A Study in Scarlet* (1887), *The Hound of the Baskervilles* (1902), and "The Boscombe Valley Mystery" (1891). Various adaptations which include this knowledge as part of Holmes's oeuvre includes: *Elementary* (Series 1, Episode 15: "A Giant Gun, Filled with Drugs", 2013); *Sherlock* (Series 2, Episode 1: "A Scandal in Belgravia", 2012 and Series 3, Episode 2: "The Sign of Three", 2014); the video games *Sherlock Holmes: The Case of the Silver Earring* (Frogwares, 2004), *Sherlock Holmes: The Awakened* (Frogwares: 2007), *The Testament of Sherlock Holmes* (Frogwares: 2012) and *Sherlock Holmes: Crimes and Punishments* (Frogwares: 2014), among many others.

³ Dean Wilson and Mark Finnane, "From Sleuths to Technicians? Changing Images of the Detective in Victoria" in *Police Detectives in History, 1750-1950*, ed. Clive Emsley and Haia Shpayer-Makov (Aldershot: Ashgate, 2006), 148; Carlo Ginzburg, "Morelli, Freud and Sherlock Holmes: Clues and Scientific Method", *History Workshop* 9 (Spring 1980), 8; Marino C. Alvarez, "Sherlock Holmes: Blood Identification and the Writing Machine", *Criminology and Police Science* 61:3 (1970), 3 (footnote 8).

⁴ Alison Johnson and Malcolm Coulthard. "The Routledge Handbook of Forensic Linguistics" in *The Routledge Handbook of Forensic Linguistics*, eds. Malcolm Coulthard and Alison Johnson (London: Routledge, 2010), 1.

⁵ Hannes Kniffka, *Working in Language and Law* (Houndmills: Palgrave Macmillan, 2007), 62.

⁶ Lawrence Frank, "The Hound of the Baskervilles, the Man on the Tor, and a Metaphor for the Mind", *Nineteenth-Century Literature* 54:3 (Dec. 1999), 339-40; James Reed, "A Medical Perspective on the Adventures of Sherlock Holmes", *J Med Ethics: Medical Humanities* 27 (2001), 76-81.

⁷ Salikoko S. Mufwene, "The Origins and the Evolution of Language" in *The Oxford Handbook of the History of Linguistics*, ed. Keith Allan (Oxford: OUP, 2013), 15-18; Judges 12.1-6.

⁸ Keith Allan, *The Western Classical Tradition in Linguistics* (Sheffield: Equinox Publishing, 2007), 11.

⁹ Andrew Linn, "The Birth of Applied Linguistics: The Anglo-Scandinavian School as Discourse Community", *Historiographia Linguistica* 35, 342-84.

¹⁰ Norman Fairclough, *Language and Power* (Abingdon: Routledge, 1989), 7.

¹¹ Allan, *Western Classical Tradition*, 163.

¹² *Ibid*, 164.

¹³ Janie Rees-Miller, "Applied Linguistics" in *The Handbook of Linguistics*, eds. Mark Aronoff and Janie Rees-Miller (Oxford: Blackwell Publishing, 2001, Kindle Edition), 9,376.

¹⁴ Li Wei, "Introducing Applied Linguistic" in *Applied Linguistics*, ed. Li Wei (Oxford: Wiley-Blackwell, 2014), 1-25.

-
- ¹⁵ Roger W. Shuy, “Forensic Linguistics” in *The Handbook of Linguistics*, ed. Mark Aronoff and Janie Rees-Miller (Oxford: Blackwell Publishing, 2001, Kindle Edition), 683.
- ¹⁶ Hannes Kniffka, *Working in the Language and the Law*, 28; Jan Svartvik, *The Evans Statements: A Case for Forensic Linguistics* (Gothenburg: University of Gothenburg Press, 1968).
- ¹⁷ Ria Perkins and Tim Grant, “Forensic Linguistics”, in *The Encyclopaedia of Forensic Sciences*, 2nd ed., ed. Jay A. Siegal et al. (London: Elsevier, 2013), 175.
- ¹⁸ Miriam Meyerhoff, *Introducing Sociolinguistics*, 2nd ed. (Abingdon: Routledge, 2011), 1; Janet Holmes, *An Introduction to Sociolinguistics*, 4th ed. (Abingdon: Routledge, 2013), 3.
- ¹⁹ Sir Arthur Conan Doyle, “A Scandal in Bohemia” (1891) in *The Best of Sherlock Holmes* (London: Collector’s Library, 2009), 12.
- ²⁰ *Ibid*, 13.
- ²¹ Michael Swan, *German in Learner English: A Teacher’s Guide to Interference and Other Problems*, eds. Michael Swan and Bernard Smith (Cambridge: CUP, 2001, Kindle edition), 1,130.
- ²² In some German editions of “A Scandal in Bohemia”, some creative translating is undertaken. The King’s shibboleth sentence, ‘This account of you we have from all quarters received’, is purposefully awkward in English because its grammar is translated directly from what the King would have said in his native German (‘Diesen Bericht von dir haben wir von allen Seiten erhalten’). However, when this is translated back into German for a German edition of the short story, it would read as grammatically correct and fluent to a native German speaker. Holmes, in the German version, would therefore have no reason to pick up on any foreign-language influence in the note. Some translators have therefore changed it to read ‘Dies haben wir von allen Seiten gehört über Sie’, or ‘This have we from all sides heard about you’. This translation renders the sentence as faithful as possible to Conan Doyle’s original wording while still including the necessary unusual construction to allow Holmes to draw an observation about the ‘impoliteness of verbs’ and to merit Holmes’s sociolinguistic profiling of the author. It is extremely unlikely that a native German speaker would write that phrase in that manner, and the translator is therefore asking the German reader to suspend their disbelief in order to get the necessary sense of disfluency and corresponding linguistic analysis.
- ²³ Tim Grant, “Approaching Questions in Forensic Authorship Analysis” in *Dimensions of Forensic Linguistics*, eds. J. Gibbons and M.T. Turell (Philadelphia: John Benjamins Publishing Company, 2008), 224; Ria Perkins, “Native Language Identification (NLID) for Forensic Authorship Analysis of Weblogs” in *New Threats and Countermeasures in Digital Crime and Cyber Terrorism*, ed. Maurice Dawson (Hershey, PA: IGI, 2015), 215.
- ²⁴ A.P.A. Broeders, “Some Observations on the Use of Probability Scales in Forensic Identification”, *Forensic Linguistics* 6:2 (1999), 228-41.
- ²⁵ In the example of, say, analysing a kidnapping note, a good forensic linguist report would report the *likelihood* that a particular suspect wrote the note or—in the absence of suspects and the need to develop a sociolinguistic profile, the *likelihood* of such suspect characteristics. Further, a good forensic linguist report would acknowledge that, even in the event that the authorship of a kidnapping note was definitely determined, that does not ‘prove’ that the author actually committed the kidnapping.
- ²⁶ Barry L. Beyerstein and Dale F. Beyerstein, eds. *The Write Stuff: Evaluations of Graphology, the Study of Handwriting Analysis* (Amherst, NY: Prometheus Books, 1992); Alfred O. Mendel, *Personality in Handwriting: A Handbook of American Graphology* (Oxford: Stephen Daye Press, 1947).
- ²⁷ The field of graphology can be dated to the 1622 publication of *Trattato Come Da Una Lettera Missiva Si Conoscano La Natura E Qualità Dello Scrittore* by the Italian philosopher Camillo Baldi (though as was common to the era, different spelling variations were often used). The book focused on how handwriting analysis can determine the character of the person who wrote it. In 1812 Swiss physiognomist Hocquet produced a more popular volume which was more widely received. Subsequently a circle of powerful French clergy formed to systematically examine the links between a person’s handwriting and their character. The French circle was at its peak around when Doyle was writing, so it is possible that he may have been aware of the publications and outputs from the circle.
- ²⁸ Sir Arthur Conan Doyle, “The Man with the Twisted Lip” (1891) in *The Best of Sherlock Holmes* (London: Collector’s Library, 2009), 144.
- ²⁹ Sir Arthur Conan Doyle, “The Adventure of the Reigate Squire” (1893) in *The Adventures of Sherlock Holmes* (Ware: Wordsworth Classics, 1996), 374.
- ³⁰ Natalie Schilling and Alexandria Marsters, “Unmasking Identity: Speaker Profiling for Forensic Linguistic Purposes”, *Annual Review of Applied Linguistics* 35 (March 2015), 195-214.
- ³¹ This contrasts to ‘collected’ or ‘naturally occurring’ data which was not produced specifically for the circumstances – for example a pre-existing letter.
- ³² Conan Doyle, “Reigate Squire”, 375.
- ³³ Tim Grant and Nicci Macleod, “Assuming Identities Online: Experimental Linguistics Applied to the Policing of Online Paedophile Activity”, *Applied Linguistics* 37:1 (February 2016), 50.

-
- ³⁴ Umberto Eco, *Einführung in die Semiotik* (München: UTB Wilhelm Fing Verlag, 1994); Umberto Eco, *Segno* (Milano: Istituto Editoriale Internazionale, 1973).
- ³⁵ Sir Arthur Conan Doyle, "The Adventure of the Dancing Men" (1903) in *The Best of Sherlock Holmes* (London: Collector's Library, 2009), 320.
- ³⁶ *Ibid*, 296.
- ³⁷ Peter L. Patrick and Samuel W. Buell, "Competing Creole Transcripts on Trial", *Essex Research Reports in Linguistics* 32 (2000), 103-32; Diego Gambetta, *Codes of the Underworld: How Criminals Communicate* (Princeton, NJ: Princeton University Press, 2009).
- ³⁸ Conan Doyle, "Dancing Men", 298.
- ³⁹ *Ibid*, 320-21.
- ⁴⁰ *Ibid*, 322-23.
- ⁴¹ Chuck Easttom, *System Forensics, Investigation, and Response*, 3rd ed. (Burlington, MA: Jones & Bartlett Learning, 2019), 117.
- ⁴² John R.G Hassard, "Cryptography in Politics", *The North American Review* 128:268 (1879), 315-25; Robert Edward Lewand, *Cryptological Mathematics* (Washington, D.C.: Mathematical Association of America, 2000).
- ⁴³ Lewand, *Cryptological Mathematics*, 37.
- ⁴⁴ David Kahn, *The Codebreakers* (New York: Scribner, 1967), 775-76.
- ⁴⁵ Sir Arthur Conan Doyle, "The Boscombe Valley Mystery" (1891), in *The Best of Sherlock Holmes* (London: Collector's Library, 2009), 111.
- ⁴⁶ Amanda Vicary and Yuliana Zaikman, "The CSI Effect: An Investigation into the Relationship between Watching Crime Shows and Forensic Knowledge", *North American Journal of Psychology* 19:1 (February 2017), 51-64.
- ⁴⁷ Shuy, "Forensic Linguistics", 683.