

Authorship Attribution in Forensic Linguistics

(Date of completion: 2015. 01.07.)

Table of Contents

Abstract.....	2
1. Introduction.....	3
2. Authorship Attribution.....	3
2.1. Non-forensic authorship attribution studies	3
2.2. Authorship attribution in forensic linguistics	5
2.2.1. Challenges	5
2.2.2. Idiolect and the notion of linguistic fingerprints.....	6
3. The methodologies of forensic authorship attribution.....	9
3.1. Qualitative analysis.....	9
3.2. Quantitative analysis	11
4. Legal background	14
5. Conclusion	19
References.....	20

Abstract

Forensic linguistics, the application of linguistics to legal issues and the law is still a relatively new field of applied linguistics. Forensic authorship attribution is one of its most important sub-fields, and the main focus of this paper. Although literary and forensic authorship analysis shares some methodology, the two has to deal with completely different challenges and so nowadays there is not much interaction between these fields. Authorship analysis is connected to the notion of idiolect, whether it exists, and if it does, whether and to what extent it can be used to determine the author of a text of questioned authorship. Depending on whether one subscribes to the weak or strong theory of idiolect, there are two different approaches to authorship attribution: the stylistic approach building on the weak theory and therefore preferring qualitative analysis, and the cognitivist approach that builds on the strong theory and uses quantitative methods. After the Daubert criteria was established in 1993 in the United States, quantitative analysis became preferred, and this defined the evolution of the field in a fundamental way. Today it is generally believed that for a method to be accepted as scientific, it has to be quantified, and even qualitative analysis accepts it as part of the process.

1. Introduction

This library research paper aims to provide a brief overview of one of the major sub-fields of forensic linguistics: forensic authorship attribution. Each section will examine an important area: history, general ideas and challenges, methodologies, and criticism, in order to determine how viable forensic authorship attribution is as a branch of forensic science, and how it might continue evolving in the near future.

The term forensic linguistics was coined by a Swedish professor, Jan Svartvik, in his 1968 paper *The Evans Statements: A case for forensic linguistics* that not only named the field, but also marked its birth. Olsson (2008) defined forensic linguistics as “the interface between language, crime and law, where law includes law enforcement, judicial matters, legislation, disputes or proceedings in law, and even disputes which only potentially involve some infraction of the law or some necessity to seek a legal remedy” (p. 3).

Forensic linguistics draws from a wide variety of linguistic areas such as “phonetics and phonology, lexis, syntax, semantics and pragmatics, discourse and text analysis, computational and corpus linguistics” (Coulthard, 1998a, p. 118), “language and memory studies, Conversation Analysis, Discourse Analysis, theory of grammar, Cognitive Linguistics, Speech Act Theory” (Olsson, 2008, pp. 3-4) and even cognitive neuroscience (Grant, 2010).

2. Authorship Attribution

2.1. Non-forensic authorship attribution studies

The study of literary authorship precedes its forensic counterpart by several hundreds years. Although nowadays the two rarely intersect, forensic authorship attribution borrowed heavily from the methodologies developed for literary analysis, and thus giving a short summary of the major studies of this field is essential for a better understanding of the current state of forensic methodology.

The oldest known mentions of cases of questioned authorship come from ancient Greek literature (Olsson, 2008); however, it was not until the early 18th century that the study of disputed authorship finally started to attract academic interest. In 1711, H.B. Witter, a German pastor analyzed the *Pentateuch* and arrived at the conclusion that the different names of divinity in the text might be an indication that the books might have been written by different authors (as cited in Olsson, 2008). Other Biblical texts that had been analyzed for

authorship include the *Letter to the Hebrews* and the Pauline Epistles. In 1851, mathematician Augustus De Morgan proposed that measuring sentence length might be a way of distinguishing authors in order to help settle the debate concerning these writings. This was the first time a statistical method was suggested for authorship analysis (as cited in Holmes, 1994).

Aside from Biblical texts, another controversy that received a significant amount of attention was the question of the real authorship of plays generally attributed to William Shakespeare. The first one to suggest that Sir Francis Bacon might be the real author was James Wilmont in 1785 (as cited in Olsson, 2008); however, as he did not wish to take part in the heated argument that was inevitable to happen, he decided against publishing his theory. It was Janus Corton Cowell who first presented the theory to the public, in 1805 (as cited in Olsson, 2008). Thomas Corwin Mendenhall was the first who attempted to further the Shakespeare/Bacon debate using a statistical method, by analyzing the frequency distributions of word-length in Shakespeare's plays and in the known works of Bacon (as cited in Holmes, 1994).

Building on De Morgan's method, Yule (1944) used mean sentence-length as a marker in his analysis of *The Imitation of Christ*; however, his results showed that this was not a reliable method of identifying authorship. An important issue according to Holmes (1994), especially in the case of older texts, is that since sentence length depends on punctuation, either the original punctuation of the author needs to remain intact, or the entire text has to be edited by one person; otherwise, the results of the analysis will not be accurate.

With the emergence of Darwinism in the second half of the nineteenth century, authorship attribution started to receive even more attention, mainly to research Biblical texts, which led to a quick development of the field (Olsson, 2008). Holmes (1994) listed several methods created during this period, such as the statistical analysis of distribution of parts of speech, function word frequencies, or methods that were designed to measure the richness of the author's vocabulary, such as the type-token ratio (the number of lexical units in a text divided by the number of units in the same text), distribution of hapax legomena (words that occur exactly once in a text) and hapax dislegomena (words that occur exactly twice in a text), word frequencies, and multivariate statistical approaches (such as factor analysis, discriminant analysis, and cluster analysis).

Olsson (2008) suggested that although many of these methods achieved some degree of success at identifying the author, researchers failed to find one method that could consistently and reliably distinguish between authors, and this, at least partially can be

attributed to the fact that most of these methods were developed by mathematicians and statisticians who did not study linguistics, and so lacked the knowledge to identify some of the inherent flaws in their theories.

2.2. *Authorship attribution in forensic linguistics*

2.2.1. Challenges

Authorship analysis is one of the major areas where a forensic linguist may be able to provide help during the criminal procedure. However, forensic authorship attribution is often very different from the analysis of literary texts, and it comes with its own share of challenges and limitations. Firstly, cases can be drastically different from each other. In one case the analyst might have to figure out who authored a collection of incriminating letters, in another analyze a will that may or may not have been forged by a third party, and then next time investigate the claim of potential plagiarism. Maybe they will have to deal with what Chaski (2007) calls the “keyboard dilemma”. When textual evidence is found on a confiscated computer, it is common for the defense to claim that it was not written by the defendant. As it is usually relatively easy for a third party to gain access to a computer, unless the text itself gets analyzed for authorship, this claim can be used effectively to weaken its reliability as evidence. These are just a few examples of the kind of cases a forensic linguist might face.

Secondly, the linguist has to keep in mind what is at stake. In forensic sciences, there is no room for guesswork. Linguists acting as expert witnesses need to present an opinion to the court, and they must make it clear that linguistic analysis of forensic evidence alone is never enough to decide the outcome of a case, only to influence it (Coulthard, 1998b), and so forensic authorship analysis is best used for confirmatory purposes. When the analysis cannot provide a clear answer, this must be explained to the court as unambiguously as possible so the report is not misunderstood by them as either verification or rejection of the questioned authorship case.

Thirdly, while literary authorship cases usually deal with a number of longer texts, forensic linguists often have access only to a very limited number of texts of either known or questioned authorship they can work with during a case. Therefore, the usual statistical methods that only work when analyzing long writings have no use here. Having to work with only a handful of short texts, often no longer than 100-150 words each, is business as usual for the forensic linguist (Olsson, 2008).

Finally, as it will become clear from the next section, the terms *authorship attribution* and *authorship identification* are somewhat problematic. Olsson (2008) pointed out that because of the short length of forensic texts, and because the existence of the notion of idiolect, and by extension that of the linguistic fingerprint, is highly disputable, there is no way to attribute a text to one specific author with absolute certainty. He also suggested that *authorship comparison* would be a more apt name, since forensic linguists can only say that out of a certain pool of potential authors, one is more likely to be the author of the questioned texts than the others, assuming the author is definitely part of the original pool. In cases where there is only a single suspect writer, only the likelihood of them being the real author can be determined, with no identification or attribution being done.

These last two points are summarized well by Richard Weld Bailey, who named three universal conditions that necessarily have to be met for authorship analysis to be possible:

- (i) the number of putative authors should constitute a well-defined set;
- (ii) the lengths of the writings should be sufficient to reflect the linguistic habits of the author of the disputed text and also those of each of the candidates;
- (iii) the texts used for comparison should be commensurate with the disputed writing. (as cited in Holmes, 1994, pp. 87-88)

2.2.2. Idiolect and the notion of linguistic fingerprints

It is commonly believed that every individual is to a certain extent unique and special. Our personality is shaped by different influences, which means that no two human views the world exactly the same way. This uniqueness is reflected in the way we use and perceive language, and thus, every person speaks a dialect unique to them. This personal dialect is called idiolect. McMenemy (2002) defined idiolect as “the individual’s unique combination of linguistic knowledge, cognitive associations, and extra-linguistic influences” (Chapter 3, Section 4, Subsection 3, para. 1). The concept of idiolect gave birth to the idea that by analyzing the speaker’s way of using language, we might be able to construct what is called a *linguistic fingerprint*, an analysis of the speaker’s idiolect that could be used to identify them based on linguistic evidence they produce. There are, however, several problems with this theory. First, the name itself is misleading, as it implies a reliable and relatively errorless scientific method akin to fingerprint or DNA analysis (Coulthard, 1998a).

Even if we accept the existence of idiolects, it would be a mistake to equate it with something like DNA that is not only guaranteed to be unique to the person under normal

circumstances, but each of us was born with it (McMenamin, 2002). Language on the other hand is something we acquire, and keep on acquiring throughout our life, and as the learning process never completely stops, an individual's idiolect is similarly never fixed; it is constantly changing and evolving (Olsson, 2008). Holmes (1994) pointed out that not even the style of established writers necessarily remains the same. Early in their careers, many authors like to experiment with different genres and styles; however, it is often the case that the level of innovation gradually drops with time, and they start preferring certain words and patterns.

This individual variation that every speaker exhibits is what is called *within author variation* or alternatively *intra-author variation*, and depends on several factors, such as mode, vocabulary, time, personal circumstances, and social changes (Olsson, 2012). Mode refers to the way the text was produced. The mode does not simply entail whether a text was written or spoken, but also how different samples from the same author can show significant variation. A handwritten text can be very different from one written on a computer, or as Grant (2013) warns, even SMS messages written by the same author can show significant differences depending on whether they were written on a mobile phone with a keypad, or on a smartphone with a touchscreen and a proper QWERTY keyboard. Text type greatly affects what vocabulary the writer will utilize. For example when composing a formal letter, the writer is more likely to use academic or business vocabulary than when writing to friends. Similarly, fiction and non-fiction books can be strikingly different when comparing their vocabularies. Time, as explained above, means that the more time passes between the writing of two texts, the more different they will be. A change in personal circumstances, such as health or marital status can also greatly alter personal style. Meanwhile, social changes cause certain words and phrases to become preferred for a time within a specific community, until they eventually fall out of favor (Olsson, 2012). There is one more factor, sociometric parameters, the speaker addressing their partner differently based on their gender, social status, or depending on the context of the conversation, which can greatly affect language use (Olsson, 2008).

Exactly how unique is an idiolect? Olsson (2008) believes that a few hundred years ago we might have been able to distinguish people purely based on their language use; however, this is no longer the case. With universal education and global media, languages are now becoming more and more homogenized. Nowadays it is nearly impossible to tell apart native English speakers of different ethnicity and nationality by simply examining their writings, unless the topic they are writing about somehow hints at their origin. Furthermore,

the idiolect reflects the sociolect of the groups the speaker associates themselves with. The evolution of sociolects is a group phenomena; as members of a group interact, they unavoidably influence each other's language use and form the sociolect together. Therefore, as the sociolect evolves, the idiolect of each member changes along with it (McMenamin, 2002; Olsson, 2008).

It is important to consider what style means in the linguistic context to decide if it can be a reliable indicator of authorship. Olsson (2008) suggested that it is either a set of markers consciously picked by the speaker, or it is a set of unconscious habits that a linguist or anyone who knows what to look for can nonetheless measure. If it is the former, then the author can attempt to disguise their style, and it would also be possible for another person to successfully imitate them. However, if it is the latter, are markers universal, or does each author have their own set with some markers that are unique to them? If every author has their own set of markers, then there would be no way to decide what to look for before analyzing the text at first, making it necessary for the forensic linguist to approach each case on an individual basis.

Even if every writer has an idiolect, it will not necessarily be reflected in every single text they produce (Grant, 2010a). In order to reproduce something like a linguistic fingerprint, a large amount of texts would be needed from the same author, which is a luxury forensic linguists very rarely have (Coulthard, 1998a).

In his 1990 paper, Morton proposed that by applying the CUSUM (cumulative sum) analysis technique, it becomes possible to tell authors apart with absolute certainty, confirming the existence of linguistic fingerprints (as cited in Olsson, 2008). He believed that by analyzing the distribution of two and three letter words, and also vowel-initial words in sentences, the speaker's language use can be visualized, and anomalies, such as additional parts inserted later by another writer, would become easily noticeable. He claimed that his technique was a reliable solution, and worked in all cases, including both long and short text analysis. Since the CUSUM method seemingly offered a straightforward, relatively simple magic solution for authorship identification, it gained immense popularity within no time. However, Morton was not a linguist, and he failed to realize that his method was based on false assumptions; for example, he failed to distinguish between written and spoken language or take intra-author variation into account. As Grant (2010b) pointed out, even though in reality the technique did not work at all, it looked convincingly scientific enough for people to believe that it really was scientific. Nevertheless, linguists eventually started to question the validity of this method, and what became known as the "CUSUM controversy" turned into the biggest controversy in the history of forensic linguistics, and received significant media

attention (Holmes & Tweedie, 1995). As it was eventually proven not to work, the CUSUM method was completely discarded by the forensic linguist community (Olsson, 2008). There had been some attempts, albeit mostly by statisticians, to improve the method, such as weighted CUSUM analysis and weighted variance estimation methods (Bissell, 1995). However, these methods still failed to consistently and reliably indicate differences between authors in multiple-authored texts (Canter & Chester, 1997).

As of now, no one had been able to construct a linguistic fingerprint, or even prove that such a thing is actually possible (Coulthard, 1998a). However, forensic linguists generally work with a limited pool of potential authors, and so in practice analyzing texts for consistency, and looking for traits that can help distinguish the writers from each other usually gets the job done adequately (Grant, 2010a; Olsson, 2008).

Depending on how one interprets idiolect, there are two main approaches to forensic authorship analysis: the stylistic method, generally referred to as qualitative authorship analysis, and the cognitivist method, also called quantitative authorship analysis (Grant, 2010a). The cognitivist approach is built on a strong theory of idiolect, which believes that linguistic competence is the most important determining factor of the speaker's language use. On the other hand the stylistic approach, which was built on a weak theory of idiolect, denies this concept. Instead they believe that a person's idiolect is in constant change, and can show drastic differences at different points in the person's life. While at first this might not seem like a very important distinction, this theory of idiolect fundamentally defines how qualitative and quantitative analysts approach the task of authorship analysis.

3. The methodologies of forensic authorship attribution

3.1. Qualitative analysis

Usually speakers of a language have more than one way of wording their intended message, this is called linguistic variation. Which one out of these the speaker settles on largely depends on their idiolect. Thus the main goal of qualitative analysis is to identify all meaningful variation within the text, and to find the author by explaining how and why the variation happened (McMenamin, 2002). Qualitative analysis takes a case study approach to authorship analysis, as it is highly dependent on intuition, and unlike quantitative analysis, it has no automatic methods that could be applied to all cases.

To measure linguistic variation, the most important step is to identify the linguistic variables, the units that store information regarding the writing style of the author, or the community the author is part of (McMenamin, 2002). Labov (1966) suggested that for analysis of variation, the most useful variables are those that are “high in frequency, have a certain immunity from conscious suppression, are integral units of larger structures, and may be easily quantified on a linear scale” (p. 32). To this list of criteria McMenamin also added that variables should be immune from conscious suggestion, and should be widely distributed throughout a particular population or individual. As Labov explained, in actual texts, variables are not encountered (as the concept of variables is an abstraction), only their variants. Because of the qualitative analysis has to adapt to the challenges of each case, the methodologies used can show significant differences between cases.

The method preferred by Coulthard (1998a) and also used by Olsson (2008) is fairly straightforward. The forensic linguist examines the questioned text or texts, and marks the features they believe are in certain ways characteristic to the author. For each selected feature, the linguist tries to find alternative ways of expressing the same meaning. After this, a corpus is used to check which of these features are regular (unmarked), and which ones are unusual (marked) forms. The same process is repeated with every known text as well, if there are any, after which the preferred stylistic features in the questioned and known texts are compared. The more marked features appear in a text, the more likely it is that the potential author will be identified, provided the real author is part of the pool of suspected authors. There are no shortcuts, and it is important that the linguist does not overlook any important features; therefore, qualitative analysis is always time-consuming, and it has to be handled with meticulous care.

Forensic stylistics is the application of stylistics, the linguistic study of style, to a forensic context, in this case to forensic authorship attribution of written textual evidence (McMenamin, 2002). To investigate authorship, the forensic stylist analyzes style-markers that appear in the texts. There are two kinds of style-markers: variation within the norm, or deviation from the norm (McMenamin, 2010). The models used by McMenamin, one of the leading forensic linguists of the field, for stylistic analysis had been originally developed by Robert Stanley Wachal in 1966 (as cited in McMenamin, 2002). These three models are: resemblance, consistency, and population. The resemblance model is the most common out of the three, it is used when there is only one or a couple of potential authors, and the only task is to either identify the person who is most likely to be the author, or exclude them based on the results of the analysis. The consistency model is used when it is unclear whether two texts had

been written by the same author. And finally, the population model is used when there are more than a few suspected authors, and there is no other way (no nonlinguistic evidence is known of that could help) to exclude them. In that case the resemblance model is used repeatedly to analyze, exclude or confirm each author one after the other until only one remains.

There are cases when it becomes impossible not to take some graphological features of the writing itself into consideration as well. This is when authorship profiling happens (Olsson, 2008), when instead of the usual comparison or confirmation analysis, the analyst tries to reconstruct the profile of the writer based on everything that can be learned from the way they use language. Although this sometimes means handwriting and some basic psychological analysis is also done and taken into consideration, authorship profiling is nonetheless neither.

In summary, qualitative analysis does not have a method that could be used for all cases. This is because according to the weak theory of idiolect, by creating a list of pre-selected features to look for, the forensic linguist might miss important clues that could potentially significantly alter the results of the analysis. Qualitative analysis is therefore time-consuming, and is prone to human mistakes.

3.2. Quantitative analysis

Over the past thirty years, computers became more and more widely available, which led to the quick development of the field of computational sciences. This eventually gave birth to a plethora of new interdisciplinary research areas such as computational psychology, computational biology, and also computational linguistics (Olsson, 2008). The main objective of computational linguistics is to create computer software that can be used for specific linguistic tasks, for example analyzing texts for the purpose of authorship attribution (Chaski, 2010). As quantitative analysis subscribes to the strong theory of idiolect, they believe that since all features necessary to identify the author can be found in all speakers of a language, computerized authorship attribution can be just as, if not more effective than the same work carried out by human analysts.

Stylometry is the statistical analysis of language use, focusing on language features that are countable and computable, such as word length, sentence length, phrase length, vocabulary frequency, and distribution of words of certain length in the text. Although originally it was not created for forensic authorship studies (most methods mentioned in

section 2.1 were also stylistic), some of its methods have been proven to work in the forensic context, or had been modified to better suit the needs of the discipline by computational forensic linguists such as Chaski (2005; 2010).

Ishihara (2011) tested the efficiency of Likelihood Ratio (LR)-based evidence evaluation for the authorship analysis of SMS messages. The LR test can show how likely an assertion is to occur if it is assumed to be true, relative to the probability of it happening provided it is not true. In forensic linguistics, this method was first used for speaker recognition and voice comparison, but Ishihara found that its application to authorship attribution also shows promising results; however, presently its feasibility is still limited. The most important issue is that the accuracy of this method is heavily dependent on the available pool of sample texts. When the pool is large enough (has at least 2200 words), the results are generally accurate. However with a small sample size (less than 1000 words), the results are very poor. This means that for this method to be viable, the linguist needs access to a sizeable corpus of SMS messages, which as of now does not exist. On the other hand, LRs have a huge benefit that should not be ignored; they can be combined with the results of LR tests done for other fields, for example the aforementioned voice analysis, or blood-stain analysis. By doing this, an overall LR probability, combining the LRs of almost all fields of the evidence investigation process, could be presented to the court.

De Vel, Anderson, Corney, and Mohay (2001) were interested in developing a method for email analysis. While not an unusual issue in forensic linguistics, one of the challenges of email texts is that they generally tend to be short, which means some of the usual computational methods, such as measuring vocabulary richness, are not feasible. Also, because emails usually have a drastically different composition style than regular texts, only other emails can be used to construct the suspected author's profile. They concluded that using a machine learning algorithm might be a good potential solution to deal with this problem. They picked the relatively new, but in performance outstanding, and even among machine learning algorithms "unfortunately complex" Support Vector Machine (SVM) algorithm (Joula, 2006, p. 285). The SVM is first provided with a number of example documents, this is where it learns what classifiers it should be looking for. In the experiment, a total of 170 style-markers, and 21 structural attributes were extracted from the training documents. With this knowledge, the SVM was then used to analyze a new batch of "questioned" documents. They found that the SVM was very effective in discriminating between authors, although the style-markers were significantly more useful in the analysis

than the structural attributes. This method was found to be effective even when the emails involved multiple topic categories.

Chaski (2005) was well aware that those features that are easily computed, like sentence-length or word-length, are usually not much help in differentiating between authors. Other features, such as phrasal types, are on the other hand much more useful and grounded in linguistic science; however, they are also much harder to compute. Furthermore, there is also the problem of “the curse of dimensionality”. The texts that need to be analyzed might be extremely short, so the usual number of variables used for the analysis might need to be reduced, and have to be selected especially carefully; otherwise, the feature set might end up having more dimensions than cases. (It might be worth noting here that one of the main benefits of the previously mentioned SVM is that because of the way it works, it completely avoids this problem (Joula, 2006).) Chaski was the first to suggest analyzing the syntactic layer for markedness instead of surface layers such as vocabulary, spelling or document layout (Olsson, 2008). For this purpose, she created a computer software that can use a syntax-based analysis method. This makes her method stand out, as usually stylometric analysis is very rarely grounded in linguistic theory to this extent.

Automated Linguistic Identification and Assessment System_Syntactic Author Identification (ALIAS_SynAID) is the authorship attribution module of Chaski’s much larger ALIAS software (http://www.aliastechnology.com/ALIAS_Technology/ALIAS_Answering_Forensic_Questions.html). According to Chaski (2010), the system was based on several fields of linguistic research, such as “sociolinguistic data gathering, psycholinguist experimental design, phonotactics, syntax, semantics, pragmatics, discourse analysis, corpus linguistics, computational linguistics, and statistics” (“Author Identification”, para. 1). Syntactic tagging can be done manually by anyone with the proper linguistic knowledge; however, using ALIAS_SynAID’s automated tagging can significantly speed up the process and also improve consistency (Chaski, 2010). Once all of the words in the text are tagged, ALIAS uses Head Driven Phrase Structure Grammar as a model to categorize headwords as either marked or unmarked (Olsson, 2008). The next step is analyzing syntactically-classified punctuation, which means instead of counting each type of punctuation (colons, commas, etc.), ALIAS simply checks what kind of boundary they mark, and puts them into categories such as EOC (end-of-clause) or EOP (end-of-phrase) (Chaski, 2005). The final step is counting average word-length. After each step, ALIAS exports the collected data to a spreadsheet, and after all questioned and known texts had been analyzed this way, the numerical output is statistically

analyzed to determine how well the texts can be differentiated from each other. This can be useful to test two sets of known texts to see if they can be differentiated from each other before adding in the questioned texts as well (Chaski, 2010). The method ALIAS uses obtained 95% accuracy (Chaski, 2005).

In summary, quantitative analysis uses numerous alternative statistical methods that are nowadays almost always computerized. This is possible because the strong theory of idiolect allows the creation of a list of pre-selected features that can be universally applied to all cases. The generally fast speed of these methods allows the forensic linguist to potentially use more than one, and compare the results with each other.

4. Legal background

In the early days of forensic linguistics, linguists used to accept jobs without having much idea what the requirements of giving expert evidence were, treating it almost like a hobby instead of a profession (Coulthard, 1998a). As the field was not yet established, each linguist had to come up with their own methodology for authorship attribution, or other tasks related to the many sub-fields of forensic linguistics. There was not much quality control either, methods grounded in scientific research (Svartvik, 1968), coexisted with others that at best were unscientific and thus unsuited for forensic analysis (Holmes, 1994), and at worst were faux-scientific, offering easy solutions that were too good to be true, and appeared so convincing in the process that they managed to fool the entire community for a long time (Holmes & Tweedie, 1995). As the field started to grow, methodology too began to develop and improve at a very fast pace (Coulthard, 1998a). The age of peaceful coexistence quickly came to an end, with legal requirements for giving expert evidence changing in a way that ended up disqualifying some previously accepted methods, and making it harder for others to stay relevant. Stylometry had little difficulty meeting the new requirements, and so a rift was born between the followers of the qualitative, and the followers of the quantitative approach. What may be called a rivalry is in reality an often very bitter feud that has been going on for as of now more than two decades. Although the previously heated argument is seemingly starting to die down, it is unlikely to be fully resolved any time soon. As previously mentioned, Grant (2010) explained that the two sides view the concept of idiolect and individual language use in fundamentally different ways, so unless conclusive evidence appears to settle the debate, reconciliation might actually never happen.

For a long time, the definitive precedence for providing expert evidence used to be the *Frye* test (Olsson, 2008). The *Frye* test states that expert evidence is admissible in a court of law provided it is a method that is accepted by the scientific community of the field it belongs to (Olsson, 2008). Although passing this test was only mandatory for cases handled by the Federal Court, *Frye* had been adopted by most state courts, as well. The main criticism this criteria received was that controversy is often a natural part of the evolution of science (Tiersma & Solan, 2010), and that even methods that were accepted by their respective scientific communities, for example hypnosis, could still be applied in ways that gave completely false results (Olsson, 2008). To address some of these flaws, the *Federal Rules of Evidence* added *Rule 702* to clarify and replace the *Frye* test:

If scientific, technical, or other specified knowledge will assist a trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case. (as cited in Joula, 2006, p. 312)

However, the rule did not have the effect its creators had hoped for. It was not clear for every federal court whether it was supposed to replace *Frye* or not, and so some adopted it and some did not, leading to a great confusion that lasted for years (Tiersma & Solan, 2010). This eventually changed in 1993, with the *Daubert v. Merrell Dow Pharmaceuticals, Inc.* case (as cited in Tiersma & Solan, 2010, p. 29), where the Supreme Court clarified that *Rule 702* has indeed replaced *Frye*, and also added that for an evidence to be scientifically valid it does not necessarily have to be generally accepted by the scientific community, but instead the method itself has to be grounded in scientific techniques (Tiersma & Solan, 2010). *Daubert* specified four nonexclusive rules for defining a scientifically valid method:

1. whether the theory offered had been tested,
2. whether it has been subjected to peer review and publication,
3. the known rate of error; and
4. whether the theory is generally accepted in the scientific community. (as cited in Tiersma & Solan, 2010, p. 30)

The *Daubert* criteria is now used by all US federal courts, and also by some state courts (Tiersma & Solan, 2010). Outside of the United States different countries have their own sets of requirements, but nonetheless the significance of the *Daubert* cannot be denied, as

it ended up shaping the evolution of the field of forensic authorship analysis (Olsson, 2008). Furthermore, Coulthard (2010) noted that it is very likely that other countries, including the United Kingdom, will eventually follow the American example and adopt new rules that work similarly to the *Daubert* test. He believes that forensic linguists around the world should prepare for this eventuality and work on creating a methodology that is compatible with the standard.

The *Daubert* criteria received mixed reactions from the forensic linguist community. Those linguists who prefer quantitative methods, such as Chaski (2005), considered the new standards a welcome change, as quantitative methods can meet the requirements fairly easily. Some quantitative analysts nonetheless expressed some concerns. Joula (2006) admitted that on the one hand the main ideas behind *Daubert* are not things any professional should complain about, as every single one of the requirements set by the criteria, for example using reliable methods that are backed up by sufficient data, are all goals scholars should naturally strive for in the first place. On the other hand, some methods, for example principal component analysis (PCA) cannot easily provide error rates, and because of this weakness it would fail the *Daubert* test, even though it satisfies all other criteria. Since *Daubert* encourages quantitative methods, Rudman (1998) warned against using statistics just for the sake of making an analysis appear more scientific than it really is. Another interesting point raised by McMenamin (2002) is that the primary audience of the expert analyst is the lay jury, and for their sake it is important to make sure the analysis is presented in a way that they can cope with. Coulthard (2010) used the same reasoning when he argued against likelihood ratios, stating that they might be too complex for the jury to understand. Joula (2006) also expressed similar concerns regarding the quantitative methods that offer the highest accuracy (machine learning algorithms, neural networks), and suggested that future quantitative research should either try to find new methods that are easier to understand without having to sacrifice accuracy, or find a better way to explain complex methods such as the SVM just well enough to allow the jury to form their own opinion based on the analysis.

Practitioners of forensic stylistics and other qualitative methods however criticized several aspects of *Daubert*. Olsson (2008) for example pointed out that there can be cases where it is simply not possible to quantify the research data, referring to his theory that there is no such thing as a linguistic fingerprint; the only thing that could possibly guarantee quantifiable results for every instance when authorship is questioned.

McMenamin (2002) added that for forensic stylists there are three possible ways to handle the *Daubert* situation: either give up trying to provide expert evidence using stylistic

methods entirely, and instead when needed offer only nonscientific testimony; try to convince the federal courts to broaden the definition of what constitutes as expert evidence; or use this opportunity to further research forensic stylistic methodology and find something that can satisfy the requirements set by *Daubert*.

As a result of the test, after 1993 forensic stylistics has come under attack by many other linguists. McMenamin and his method received some harsh criticism by Crystal (1995) who felt the methodology he described was too vague to be considered seriously, and questioned if forensic stylistics can be considered a real science. Further questions were raised concerning the viability of forensic stylistics as a forensic science after the *U.S. v. Van Wyk* case in 2000 (as cited in McMenamin, 2002, Chapter 9, Section 3, para. 5). In *Van Wyk*, the Court refused to allow the expert witness to testify after he failed to meet the standards (McMenamin, 2002). Tiersma and Solan (2010) commented on the case and decided that the Court correctly recognized that forensic stylistics at that time was very similar to handwriting analysis in that it used to be more of an art than an actual science. They added that without taking base rates into account, possibly with the help of relevant corpora, the stylistic analysis will lack any scientific validity. For the best results, it is also important to analyze both similarities and differences between the different texts. Finally, the analyst should warn the court about the limits of forensic stylistics, to avoid unintentionally misleading them with the testimony.

Critics of the stylistic method like to refer to this case when discussing its shortcomings. For example, Chaski (2010) mentioned it to point out that unlike forensic stylistics, her own computational method has a flawless track record at passing the *Daubert* test. However, McMenamin (2002) made a good case for dismissing *Van Wyk* when discussing the problems of forensic stylistics. First of all, the witness himself was not a linguist, but an FBI agent who had no degree in linguistics, or received any kind of professional training in linguistic stylistics. Moreover, the method he used for his analysis had already been rejected as nonscientific by the forensic linguist community. He also seemed to confuse forensic stylistics with another, unrelated field: textual analysis. It is apparent as he used both terms interchangeably in his report, further demonstrating his lack of knowledge in the field of forensic linguistics. Finally, the witness relied on background information for his analysis, which is something forensic linguists try to avoid if possible. Taking all of these factors into consideration, it is easy to see that if the *Van Wyk* case proved anything, it is that the witness was unqualified for the job, and not that forensic stylistics is in any way unsuitable as a science. The Court apparently shared the same sentiment, as although they did

not let the witness to finish his testimony, the analysis of markedness he had prepared was nonetheless accepted as evidence in the case.

Only a year later, in *U.S. v. Spring* (as cited in McMnamin, 2002, Chapter 9, Section 3, para. 8) the Court once again ruled that stylistic analysis of questioned authorship can be accepted as scientific evidence. Nevertheless, a few years earlier Coulthard (1998b) wrote that an analysis that focuses only on stylistic features can only ever show the opinions of the analyst, and never the facts; and despite the favorable court ruling, it seems McMnamin (2002; 2010) shared some of his concerns, as ever since *Daubert* and *Van Wyk*, he turned his attention to perfecting his method in ways that make it more compatible with the current criteria. He recognized that there was a need for at least some level of quantitative analysis, not only because it helps meeting the judicial requirements, but also because it makes the testing process significantly faster and easier. He kept his method relatively unchanged, the texts are analyzed manually, and the style-markers are still picked by hand; however, once this is done, he uses a variety of established computational and statistical methods, such as frequency distribution, standard Error of Difference, or the proportion test to quantify his results.

This revised method of forensic stylistics was generally well-received by fellow linguists, Chaski (2010) even praised it as “by far the best application of the forensic stylistic method” (“New Computational Tools for Author Identification”, para. 3), although added that McMnamin should consider applying a shorter list of preselected style-markers for better consistency. She argued that without establishing a standard protocol for checking features, even if the results are quantified, the method itself will still depend too much on intuition, which makes it susceptible to methodological subjectivity and confirmation bias. Tiersma and Solan (2010) concluded that qualitative methods, mainly those practiced by McMnamin and Coulthard are perfectly adequate for scientific use, especially if they are supported by statistical analysis. They also approved of the quantitative method of Chaski, as it has the potential to lead to the development of even more sophisticated techniques. They suggested that even though right now there is no way to resolve the idiolect debate that is the most fundamental difference between the two sides, combining the quantitative syntactic analysis and the qualitative research of idiosyncratic features into one method should nevertheless be considered.

5. Conclusion

In this paper, I gave a brief summary of the current state of the field of forensic authorship attribution in hope of determining its viability as a forensic science. It is clear that despite its short history, the field went through considerable evolution throughout the years, and it is unlikely that this evolution will slow down any time soon. Although the forensic linguist community has not yet managed to find a method that could satisfy all requirements and at the same time gain universal approval, there is much less chaos now than there used to be as little as ten years ago. If the current trends continue, forensic sciences in general will become even more computerized in the future, and computer software tools far more advanced and sophisticated than what exist now will be readily available to the analysts. Some linguists (Joula, 2006; Rudman, 1998) argued against the development of a general purpose authorship attribution system, as they believe the process of analysis should never become fully automated. However if ALIAS is any indication, the technology will arrive sooner rather than later. Whether this is a positive thing or something that should be a cause for concern remains to be seen. Nevertheless, the field is now well established, and has a large number of active practitioners and researchers who keep pushing the limits and improving forensic linguistics. Looking at what already has been achieved, and also at ongoing researches, it is hard not to be optimistic about the future of forensic authorship attribution.

References

- Bissel, A. F. (1995) Weighted cumulative sums for text analysis using word counts. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 158(3), 525-545.
- Canter, D., & Chester, J. (1997). Investigation into the claim of weighted cusum in authorship attribution studies. *International Journal of Speech Language and the Law*, 4(2), 252-261.
- Chaski, C. E. (n.d.). ALIAS modules: Software for forensic linguistic analysis. In *ALIAS Services*. Retrieved April 15, 2014, from http://www.aliastechnology.com/ALIAS_Technology/ALIAS_Answering_Forensic_Questions.html
- Chaski, C. E. (2005). Who's at the keyboard? Authorship attribution in digital evidence investigations. *International Journal of Digital Evidence*, 4(1), 1-13.
- Chaski, C. E. (2007). The keyboard dilemma and author identification. In S. Shinoi & P. Craiger (Eds.), *Advances in digital forensics III* (pp. 133-146). Orlando, FL: Springer.
- Chaski, C. E. (2010). Linguistics as a forensic science: The case of author identification. In S. J. Behrens & J. A. Parker (Eds.), *Language in the real world: An introduction to linguistics* (Chapter 10) [Kindle Keyboard version]. Retrieved from Amazon.com
- Coulthard, M. (1998a). Making texts speak: The work of the forensic linguist. *Studia Anglica Posnaniensia*, 33, 117-130.
- Coulthard, M. (1998b). Identifying the author. *Cahiers de Linguistique Francaise*, 20, 139-161.
- Coulthard, M. (2010). Experts and opinions: In my opinion. In M. Coulthard & A. Johnson (Eds.), *The Routledge handbook of forensic linguistics* (pp. 473-486). Abingdon, Oxon: Routledge.
- Crystal, D. (1995). [Review of the book *Forensic stylistics*, by G. R. McMenamin]. *Language*, 71(2), 381-385.
- de Vel, O., Anderson, A., Corney, M., & Mohay, G. (2001). Mining e-mail content for author identification forensics. *ACM SIGMOD Record*, 30(4), 55-64.
- Grant, T. (2010a). Text messaging forensics: Txt 4n6: Idiolect free authorship analysis? In M. Coulthard & A. Johnson (Eds.), *The Routledge handbook of forensic linguistics* (pp. 508-522). Abingdon, Oxon: Routledge.

- Grant, T. (2010b, August 25). Tim Grant on forensic linguistics [Video file]. Retrieved from <https://www.youtube.com/watch?v=HitHpC1PSVc>
- Grant, T. (2013). Txt 4n6: Method, consistency, and distinctiveness in the analysis of sms text messages. *Journal of Law and Policy*, 21(2), 467-494.
- Holmes, D. I. (1994). Authorship attribution. *Computers and Humanities*, 28, 87-106.
- Holmes, D. I., & Tweedie, F. J. (1995). Forensic stylometry: A review of the cusum controversy. *Extrait de la Revue Informatique et Statistique dans les Sciences humaines*, 31(1), 19-47.
- Ishihara, S. (2011). A forensic authorship classification in SMS messages: A likelihood ratio based approach using n-gram. *Proceedings of Australasian Language Technology Association Workshop (ALTA 2011)*. Australian National University, Canberra, Australia.
- Joula, P. (2006). Authorship attribution. *Foundations and Trends in Information Retrieval*, 1(3), 233-334.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- McMenamin, G. R. (2002). *Forensic linguistics: Advances in forensic stylistics* [Kindle Keyboard version]. Retrieved from Amazon.com
- McMenamin, G. R. (2010). Forensic stylistics: Theory and practice of forensic stylistics. In M. Coulthard & A. Johnson (Eds.), *The Routledge handbook of forensic linguistics* (pp. 487-507). Abingdon, Oxon: Routledge.
- Olsson, J. (2008). *Forensic linguistics* (2nd. ed.). London, England: Continuum International Publishing Group.
- Olsson, J. (2012). *Wordcrime: Solving crime through forensic linguistics*. London, England: Continuum International Publishing Group.
- Rudman, J. (1998). The state of authorship attribution studies: Some problems and solutions. *Computers and the Humanities*, 31, 351-365.
- Svartvik, J. (1968). *The Evans statements: A case for forensic linguistics*. Gothenburg, Sweden: University of Gothenburg Press.
- Tiersma, P. M., & Solan, L. M. (2010). *Speaking of crime: The language of criminal justice (Chicago series in law and society)*. Chicago, IL: The University of Chicago Press.
- Yule, G. U. (1944). *The statistical study of literary vocabulary*. Cambridge, England: Cambridge University Press.