Interdisciplinary Approaches to Forensic Linguistic Analysis and Medieval Manuscript Studies: Developing a Working Framework for Research

by

Angela Corinne Loewenhagen Schrader

A Dissertation Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

Approved April 2021 by the Graduate Supervisory Committee:

Elly van Gelderen, Chair
Robert Bjork
Markus Cruse

ARIZONA STATE UNIVERSITY

May 2021
ABSTRACT

The field of forensic linguistics has burgeoned in the past several decades within a current-day framework of language use, ranging from dialectal analysis to legal language analysis in court cases, to trademark and authorship disputes, and more. When it comes to utilizing forensic linguistics techniques within a historical framework, however, there is still a great deal of research and work to be done. There is a gap in historical research that needs to be filled, to create a more cohesive whole when examining the past for understanding. Pioneers in historical authorship analysis are now using forensic linguistic methods more frequently in their manuscript analyses and research, and the results of those studies indicate that some linguistic variables can be statistically measured with a relative degree of accuracy for historical documents.

What is needed now is a forensic analysis which also comprehensively accounts for the challenges regarding various cultures’ definitions of ‘author’ and ‘authorship’ and translation methods in different time periods. For medieval manuscripts, these analyses must also consider the manuscript culture inherent in that time period. In this dissertation, I discuss the rift apparent in the framework of understanding where forensic linguistic analysis and manuscript analysis are not fully meeting in the middle. I address the need for a general methodology that allows academics in both disciplines to work together in finding variables for forensic testing which include the needs of the manuscript culture behind it, so that future research can more fully enrich the understanding of medieval history as a whole.

During that discussion, I analyze several completed authorship analyses surrounding the tenth century Old English gloss of the Lindisfarne Gospels, examining
the methods utilized by each researcher in accomplishing their chosen research goal. Then, I focus on developing a generalized methodology which can provide a framework for handling unique, individual analyses of medieval manuscripts which have questionable authorship attribution. This framework will help to create a more solid foundation for providing more accurate and effective data for historical authorship cases.
DEDICATION

I would like to dedicate this dissertation to my husband, Robert Schrader, who has patiently stood with me every step of the way through this long, difficult path that is my graduate education. He has supported me and pushed me, and he has always encouraged me when I most needed it – I am immensely grateful for everything you do for me.

I would also like to thank my family and friends, who have all been extremely supportive of me throughout my graduate school experience.
ACKNOWLEDGMENTS

I would like to thank my committee members, Elly van Gelderen, Robert Bjork, and Markus Cruse for their comments and contributions to my dissertation as I have worked my way through this project. I would also like to thank Elly van Gelderen for all of the emotional and professional guidance I have received from her over the last several years as I progressed through first my Master’s program and then my PhD program to this point; she kept me grounded mentally and did a wonderful job of guiding me through all of this as my chair and advisor – I am beyond grateful.
# TABLE OF CONTENTS

## CHAPTER

1. **INTRODUCTION** .................................................................................................................. 1

2. **LITERARY REVIEW: FORENSIC LINGUISTICS** ......................................................... 6
   2.1. Introduction ....................................................................................................................... 6
   2.2. What Is Forensic Linguistics?............................................................................................ 7
   2.3. Authorship Attribution in Forensic Linguistics................................................................. 9
   2.4. Stylistics Tests Utilized for Determining Authorship ..................................................... 13
   2.5. Methods and Testing for Authorship Analysis ............................................................... 18
       2.5.1. Statistical Tests ......................................................................................................... 20
       2.5.2. Choosing Data Sets .................................................................................................. 23
   2.6. Authorship Debates over Medieval Manuscript Texts ................................................. 27
   2.7. Relevance of Forensic Linguistics to Medieval Manuscript Studies ....................... 30
   2.8. Conclusion ...................................................................................................................... 31

3. **LITERARY REVIEW: MANUSCRIPT ANALYSIS** ....................................................... 33
   3.1. Introduction ...................................................................................................................... 33
   3.2. Definition ........................................................................................................................ 33
   3.3. Elements of Manuscript Production ............................................................................... 35
   3.4. Ordinatio and Communicare ........................................................................................... 38
   3.5. Early Medieval Culture and Manuscripts ........................................................................ 41
   3.6. Production Changes over Time ....................................................................................... 43
   3.7. Manuscript Creation and Transcription Errors ................................................................ 46
CHAPTER

5 CASE STUDY: THE OLD ENGLISH GLOSS TO THE LINDISFARNE GOSPELS ................................................................. 80

5.1. Introduction .................................................................................................................................................. 80

5.2. Literature Review ....................................................................................................................................... 81

5.2.1. History of the Lindisfarne Gospels ........................................................................................................ 81

5.2.2. Provenance of the Lindisfarne Gospels ................................................................................................. 83

5.2.3. Authorship Debates Surrounding the Lindisfarne Gospels ................................................................. 84

5.3. Data Collection Methods .......................................................................................................................... 86

5.4. Data Analysis ............................................................................................................................................... 87

5.4.1. Marcelle Cole’s (2017) Study ................................................................................................................. 87

5.4.2. Johanna Wood’s (2019) Study ................................................................................................................ 90

5.4.3. Tadashi Kotake’s (2018) Study ............................................................................................................... 92

5.4.4. Marcelle Cole’s (2018) Study ................................................................................................................ 95

5.5. Conclusion ................................................................................................................................................... 99

6 DEVELOPING A METHODOLOGY ................................................................................................................... 101

6.1. Introduction ................................................................................................................................................. 101

6.2. Current Forensic Linguistic Methodology Theory ..................................................................................... 102

6.2.1. Three Models of Analysis .................................................................................................................... 102


6.2.3. General Notes on Current Forensic Linguistic Methods ................................................................. 107

6.3. Medieval Considerations for Methodology Strategies ............................................................................... 109
# CHAPTER

6.3.1. Authorship Considerations ................................................................. 109

6.3.2. Manuscript Analysis Considerations .............................................. 111

6.4. Analysis of LiG Case Study Strategies ........................................... 112

6.5. Forming a New Methodology ......................................................... 117

6.6. Conclusion ......................................................................................... 122

7 LIMITATIONS AND CONCLUSIONS .......................................................... 123

7.1. Limitations ....................................................................................... 123

7.2. Conclusions ....................................................................................... 125

REFERENCES ............................................................................................ 128
CHAPTER 1
INTRODUCTION

The field of forensic linguistics has been burgeoning in the past several decades within a current-day framework of use with regards to language use, ranging from dialectal analysis to legal language analysis in court cases, to trademark and authorship disputes, and more (Coulthard & Johnson 2007). When it comes to utilizing forensic linguistics techniques within a historical framework, however, this fledgling field is more like an undeveloped realm of potential knowledge - there is still a great deal of research and work to be done. Pioneers in historical authorship analysis using forensic linguistic methods - such as Mosteller & Wallace (1964 and 1984), Jackson (1991; McMenamin 2002), and Yule (1944), among others - have shown that some linguistic variables can be statistically measured with a relative degree of accuracy for historical documents.

Chapter 2 will focus mainly on the field of forensic linguistics and some of the many ways in which it can benefit researchers working with medieval manuscripts.

However, there are certain challenges that inevitably arise when one works with historical manuscripts, which were handwritten in production, long before the advent of the printing press in the fifteenth century. For example, provenance (where a manuscript ‘resides’ or possibly originated) for a manuscript was not set in stone; both manuscripts and the highly-trained scribes and artists who created them were in high demand in the Middle Ages, and they were extremely mobile (von Daum Tholl 1995). Therefore, a researcher cannot understand the cultural context of a medieval manuscript (and thus ascertain linguistic certainty about authorship from a dialectal standpoint) until they also know what the cultural background and history of a manuscript is. Another major
problem in working with medieval manuscripts is that most of them have not survived to
the present day, which can make finding contemporary manuscripts for comparison in
writing styles, handwriting, etc. a real problem; in stark contrast to today’s vast array of
written and online resources, a medievalist and historical linguist must work with a very
small portion of what written resources would have been available in the Middle Ages.
There are also a whole host of issues which arise when one is examining handwritten
documents in a culture where editorial standards had not yet been set (which did not
occur until after the printing press was invented in the fifteenth century), which will be
discussed briefly in Chapter 3.

It is necessary, also, to consider the implications of cultural elements and attitudes
toward the idea or definition of the terms *author* and *authorship*, as well as the tendency
of humans to make either *corrections* or *errors* when creating a copy of an original
manuscript, or even a copy of a copy, or a copy of a copy of a copy, and so on.
Researchers such as Minnis (2010), Partridge and Kwakkel (2012), and Foucault (1969)
have spent a great deal of time analyzing and philosophizing about what it means to be an
*author* or a *scribe* when creating a manuscript, and bringing attention to the fact that
those terms had very different meanings and connotations in the Middle Ages, when
compared to how we view them today. This issue becomes even more complicated when
we take into account the other choices which need to be made whenever one creates a
translation of a text into another language, as we see in the tenth century *Lindisfarne
Gospels*, which includes an eighth century Latin text with a tenth century Old English
gloss (Nagucka 1997). This will be discussed further in Chapter 4, with a detailed
analysis of what it means to be an author or to attribute authorship to a medieval document.

There is a gap in historical research which needs to be filled for the sake of creating a more cohesive whole when one looks at the past for understanding. One of the reasons for this is that the members of our academic community all too often tend to over-specialize within their own career fields, and interdisciplinary practices can become a secondary thought process.

In our own era education has become divided into ‘subject areas’ such as history, English language, English literature and Latin, to the extent that many highly educated people are ignorant of areas outside their own specialisation and unaware of relevant interdisciplinary material in and around the period and region they are studying. Recent scholarly approaches to the medieval period, however, are emphasising the multilingualism, cross-culturalism and integration of disciplines which was taken for granted in the Middle Ages. This requires greater versatility on the part of students than heretofore. [Owen-Crocker 2009: 4]

In order to improve our understanding of history and all of the physical and cultural elements within it, we must continue to search for ways to increase and improve interdisciplinary research into the Middle Ages.

We are beginning to make some progress along these lines. There is some good research beginning to emerge from forensic linguists who wish to use their analytical skills not just for the present-day court cases we are seeing more and more frequently, but to help us anthropologically understand our past more fully through manuscript research and analysis. However, what I have not seen so far in any great detail, is a forensic analysis which also comprehensively takes into account the challenges discussed above regarding various cultures’ sense of authorship values and translation methods in
different time periods. There is a rift – a hole in the fabric of this framework of understanding where forensic linguistic analysis and manuscript analysis (both physical analysis and historical analysis) are not fully meeting in the middle. The academic world needs a good framework of methodology to work with which will allow both of those disciplines to work together, to find variables for forensic testing which will take into account the needs of the manuscript and the culture behind it, so that future research will more fully enrich our understanding of our history as a whole.

My dissertation focuses on developing a generalized methodology which can provide a framework for handling unique, individual analyses of medieval manuscripts that have questionable authorship attribution. During the course of my research into this subject, I analyze several completed authorship analyses surrounding the tenth century Old English gloss of the *Lindisfarne Gospels*, examining the methods utilized by each researcher for accomplishing their chosen research question. My discussion of these case studies is included in Chapter 5.

I then bring these various analyses together to begin the creation of my contribution to academia of new information and synthesis. In Chapter 6, I use the information from my forensic linguistic analyses and background research, along with my understanding of the physical and cultural demands involved in conducting manuscript research, to generate a beginning framework from which I can begin to form a set of methods, or ‘best practices’, in proceeding with future forensic analysis of historical manuscripts. It is my hope that this framework will help to create a more solid foundation for future manuscript studies, providing much more accurate and effective
data for historical authorship cases, and potentially other types of manuscript studies as well.
CHAPTER 2
LITERARY REVIEW: FORENSIC LINGUISTICS

2.1. Introduction.

In this chapter, I will briefly introduce the subfield of forensic linguistics and the aspects of this field which pertain to the attribution of authorship. To accomplish this, I will begin with a quick definition of the term *forensic linguistics* and its current usage in various court cases. I will then narrow down my discussion to one facet of forensic linguistics – authorship analysis and determination via stylistics. I will discuss some of the myriad ways in which authorship can be tested and analyzed, as well as a few methods for determining which of those tests would be appropriate for a certain type of document or situation.

The second goal of this chapter is to adequately discuss how and why this normally modern-day linguistics field may be utilized to provide a more scientific viewpoint on authorship debates over medieval texts. To achieve this, I will briefly discuss some of the authorship cases that have already been researched on medieval texts, and provide information on the types of testing and analysis that have been performed on them thus far. I will also explore the benefits of widening the scope of this subfield to include authorship studies that are not contemporary with modern publishing techniques, namely, that of medieval texts in the form of hand-created manuscripts.

The sources discussed in this chapter will cover some of the scientific research methods utilized in forensic linguistics as they are currently employed to generate the necessary data for analyzing modern documents for authorship markers. While I will briefly discuss the implications of using this type of testing for the sake of analyzing
medieval documents, the majority of that particular discussion will continue to mature in
Chapter 6, as part of my comprehensive discussion and analysis of all parts of this
dissertation.

2.2. What is forensic linguistics?

The field of forensic linguistics was officially coined when Jan Svartik (1968) published
The Evans Statements: A Case for Forensic Linguistics, which addressed linguistic
methods for testing and analyzing the authorship of several court documents integral to
the Evans murder case adjudicated in the United Kingdom in 1950 (Perkins & Grant
2013: 174; Coulthard & Johnson 2007). Timothy Evans was convicted of the murders of
his wife and baby; consequently, he was sentenced to death in 1950. Svartvik later
analyzed Evans’ documents of confession using forensic linguistic testing, and he
determined that the confession had several instances of drastically different grammatical
structures, especially regarding some incriminating parts of the confession. This analysis
demonstrated that the authorship of the confession was definitely in question. Ultimately,
Evans was exonerated posthumously in 1966 (Perkins & Grant 2013: 174). This pardon
may not have occurred if this forensic analysis had not been conducted and proven that
Evans’ confession statement had been altered.

So, what is forensic linguistics, exactly? According to the Encyclopedia of
Forensic Sciences, Second Edition:

Forensic linguistics is a branch of applied linguistics relating to the law and legal
processes. The discipline subsumes a wide variety of research and casework and it
is suggested by Coulthard, Grant, & Kredens [2010] that forensic linguistics is the
application of linguistics to three main areas; ‘written legal texts, spoken legal
practices, and the provision of evidence for criminal and civil investigations and
courtroom disputes’ (p. 529). [Perkins & Grant 2013: 174]
Since the context of court cases and linguistic analyses pertinent to those cases was the
initial research venue for forensic linguistics, it makes sense that the encyclopedic
definition of the term forensic linguistics would be assigned to only the three “main”
areas of application listed above. Indeed, most of the published cases discussing forensic
linguistic practices and methods have been within the purview of modern court cases and
investigations.

Coulthard & Johnson (2007) discuss many different uses for forensic linguistics
in modern day court cases. There are cases addressing morphological meaning and
phonetic similarities in trademark law, such as the McDonald’s Corporation v. Quality
Inns International, Inc. case (Coulthard & Johnson 2007: 121-3; Shuy 2002: 95-109);
cases involving the complexity of the syntax in a legal letter (Coulthard & Johnson 2007:
123-4; Levi 1993); and cases which are built around lexical or pragmatic meaning of
writings, both in business legal writing and in disputed personal communications.
Forensic linguistics is also used to analyze police reports and disputed statements of
confession (as in the Evans case above), and it can be employed to help a court determine
if a witness was coerced or led into a particular line of questioning rather than being
interviewed by police in a proper manner. It can also be utilized to identify fraudulent
documents and to analyze dialect variations and look for clues in an investigation.

For the purposes of this study, however, we will examine a more specific use for
forensic linguistics – that of authorship attribution. Coulthard & Johnson describe the use
of this subfield of forensic linguistics as a situation where:
The linguist approaches the problem of questioned authorship from the theoretical position that every native speaker has their own distinct and individual version of the language they speak and write, their own *idiolect*, and the assumption that this *idiolect* will manifest itself through distinctive and idiosyncratic choices in speech and writing. [Coulthard & Johnson 2007: 161; Coulthard 2004: 431-2; see Halliday et al. 1964: 75; Abercrombie 1969]

A linguist treats each piece of writing or speech, then, as having “linguistic fingerprints” or “impressions” as a way of identifying a particular author or speaker as being the creator of that utterance (Coulthard & Johnson 2007: 161).

### 2.3. Authorship attribution in forensic linguistics.

Authorship attribution - more formally referred to as *comparative authorship analysis* in forensic linguistics - is “where a linguist compares an anonymous text to the known writings of an individual to determine whether or not they are one and the same author” (Perkins & Grant, 2013: 175). Coulthard (1994; Olsson 2004: 15) alternatively referred to his own early methods of analyzing authorship markers forensically as *forensic discourse analysis*. According to McMenamin (2002; Coulthard & Johnson 2007: 171), there are two questions which may be asked in an authorship debate: “looking for *consistency*”, which addresses if a text or texts have just one author, or if they have several authors; and “looking for *resemblance*”, which addresses if a text or texts with unknown authorship attribution may possibly have been written by an author or authors who are part of a suspect group (a closed set scenario). McMenamin (2002) contends that there are two ways in which these two questions may be answered:

- The work is qualitative when features of writing are identified and then described as being characteristic of an author. The work is quantitative when certain indicators are identified and then measured in some way e.g. their relative
frequency of occurrence in a given set of writings. [McMenamin 2002: 76; Coulthard & Johnson 2007: 171]

As an example of an authorship attribution case, the Helander authorship case in Strängnäs, Sweden addressed the question of authorship regarding several anonymous and slanderous letters which were written in order to swing an election vote for a new bishop there in 1952 (Särndal 1967: 255). The letters served the purpose of emotionally influencing the clergy to favor Helander over the other candidates, and Helander was subsequently elected to the position of Bishop of Strängnäs. However, there was sufficient suspicion surrounding these letters that an investigation ensued, and a whole host of written material was later produced about the case. Särndal (1967) points us to three main bodies of reference on the case details, which I will quickly outline: S. Helander (1957), Runquist (1958), and Tenow (1963).

Essentially, Johannisson and Wellander (investigators into this case) employed linguistic statistical testing - in the form of very rare or abnormally used words and phrases - to compare the anonymous letters with some of Bishop Helander’s known writing samples. They concluded that there were enough similarities between the anonymous letters and Helander’s known writing that Helander must have been the author of the letters; Helander was subsequently removed from his newly-appointed office of Bishop of Strängnäs. It is worth noting that there were some problems with the research methods employed by Johannisson and Wellander, which might have provided more detailed and accurate information on this case if their methods had been more thorough. First, S. Helander (1957) demonstrated that many of the words and phrases which were used for the comparison of the known and anonymous writings are only
‘abnormal’ because they are context-specific to church writing, and there were many other clergymen who also used the same wordings in their own writing; so Helander’s known writing was potentially less identifiable with the anonymous letters after all on that basis. The other main problem stems from the fact that only Bishop Helander’s known writing was compared to the anonymous letters – there were no other suspects considered in the case, and therefore a proper methodical comparison could not be made between multiple suspects in the known and anonymous writings.

I use the term methods loosely in this discussion above, since Coulthard & Johnson (2007) have indicated that a complete methodology for the handling of forensic linguistic cases has not been developed as of yet. In the earlier days of the growing field of forensic linguistics in the twentieth century, there was a great deal of variability in how and when testing was utilized to determine authorship, and the methods varied significantly from case to case. Even in present-day court cases this continues to be the case. However, at least as it pertains to the forensic linguistic subfield of authorship attribution, McMenamin (2002:119-25) does put forward a quite comprehensive, flexible methodology for the identification of authorship in current forensic linguistic cases, which I will discuss in much greater detail in Chapter 6.

Perkins & Grant (2013) go on to note that:

Comparative authorship analysis of long texts is increasingly dependent heavily on multivariate computational techniques, which can be shown to be reliable but offer little explanation as to the outcome. This validity deficit means that such techniques tend not to be depended upon by the forensic analyst and in any case such techniques often require more text than is available in forensic casework.

[Perkins & Grant 2013: 175]
As will be discussed in more detail later, the lack of longer-length texts – evidenced by either incomplete or missing text – in documents is all too common when examining medieval manuscripts, making this an important factor to remember when developing a proper methodology for the use of forensic linguistics in manuscript studies.

In his article, “Quantifying evidence in forensic authorship analysis”, Grant (2007) discusses the importance of creating a transparent process for handling data in authorship cases:

… quantification requires considerable methodological clarity if it is to be applied to authorship analysis. Statistically rigorous quantification is insufficient without methodological clarity. … identification of differences between groups of texts, or classification by those differences, is not sufficient for rigorous authorship analysis. What is further required is an explanation of any differences found and argument that those differences represent authorship variation rather than some other factor. [Grant 2007: 9]

Forensic linguistics as an authorship profiler works well for court cases, as Shuy (2003) indicates:

It should be pointed out that linguistic profiling has been most effectively used to narrow down a suspect list rather than to positively identify a suspect. This is not to say that such positive identification is impossible but, rather, that the potential for variability in language use is great and the texts offered for comparison are sometimes dissimilar in genre, register, and size. [Shuy 2003: 686]

If this “potential for variability” that Shuy mentions applies to modern-day linguistic profiling – where we have so many less disadvantages to contend with, with regards to determining who the author is of a particular document – then we must ask ourselves the question: can we get more precise than this with the profiling of ancient manuscripts for authorship markers? Is that even possible? These are some of the pressing issues to be
discussed in further detail in Chapter 7, where I will analyze the limitations inherent in applying these modern techniques to medieval problems.

2.4. Stylistics tests utilized for determining authorship.

While the official field of forensic linguistics is only a few decades old at this point, some of the measurements utilized to analyze texts for determining authorship have been around for much longer than that. It is becoming increasingly evident over the past century or so that literary analysis alone cannot produce a definite level of certainty as to the authorship of a document, without hard scientific evidence and analysis to back it up. More and more analyses of documents are including statistical and forensic evidence to back up claims of authorship, especially when legal matters are involved; this has prompted more testing of variables and forming of mathematical equations in the last several decades, to determine what factors are the most stable ones for providing data to use in determining authorship.

Morton (1978) provides a good explanation of some of the rationale behind this move toward scientific methods of authorship analysis:

…the writer’s choice of words is not an effective discriminator when it is taken in total as his vocabulary. His choice and rate of using frequent words is much more effective, especially when there are large amounts of text to sample and few contenders for the authorship of it. But by far the most effective discriminator between one writer and another is the placing of words. The placing can be looked at as the position within a sentence, the most effective method for inflected languages, or in immediate context, which is the method most effective in uninflected languages. Proportional pairs of words come somewhere between the frequent words and the placing of them in effectiveness; they have the great
advantages of being independent of punctuation and simple both to record and analyse. [Morton 1978: 107]

Since humans are well-known to be proverbial ‘creatures of habit’, it stands to reason that scientific analysis of written texts habitual patterns and frequencies of language use (both grammatical and lexical) by a writer should yield some good data to be used in assessing the authorship of those texts. Morton (1996) narrows down the parameters for the setting of parameters for these variable patterns:

…habits which occur frequently are unlikely to prove profitable indicators of authorship and habits which are too frequent also lack power to discriminate. Habits which make up a substantial proportion of the words in sentences, but are not too frequent, seem the most promising to investigate. [Morton 1996: 278]

Coulthard & Johnson (2007) discuss several of the documented tests performed for authorship purposes, which is now referred to more commonly as the forensic linguistics subfield of stylometrics. Historically, stylometrics came into being in 1851, when a biblical scholar asked S. E. de Morgan to find a way to determine if a set of letters was correctly attributed to St. Paul. Instead of addressing the authorship issue from a strictly literary standpoint, de Morgan opted to scientifically analyze the documents and look for linguistic features and patterns which could potentially be assigned to an individual’s writing style. His research led him to posit that authorship might be determined via average word length used by an author (de Morgan 1882; Coulthard & Johnson 2007: 164); this method would be utilized later by Mendenhall (1887) when he examined Shakespeare’s plays to assess their potential authorship by Shakespeare, Bacon, or Marlowe (Coulthard & Johnson 2007: 164).
Later, Yule (1938) would perform statistical tests on texts to determine what kinds of measurements would yield scientifically accurate results as to authorship. He found that measurements such as average sentence length and lexical richness (such as the number of different words used in a text, and type/token ratios) can be effective in sussing out the correct author from a group of writers (Coulthard & Johnson 2007: 164).

Yule (1944) continued his research into the statistical approaches which could be used to measure lexical richness in a text, and he developed the equation which came to be known as Yule’s Characteristic $K$ – a formula which provides an index of relatively stable results when used with texts of varying sizes (Yule 1944; Herdan 1960: 84-5; Tweedie & Baayen 1998; Mikros 2009: 64). It is shown as:

$$K = \sum \frac{n_i^2}{N^2} - \frac{1}{N},$$

where $N$ is the number of letters in the sample text and $n_i$ is the frequency of occurrence of the $i$-th letter. The variance for this formula is computed by:

$$\sigma^2_K = \frac{0.0168N + 0.1182(N-1)}{N(N-1)}$$

[Yule 1944; Herdan 1960: 84-5]

Honoré (1979) added the measurement of hapax legomena frequency (words that occur only once in a text) to the list of lexical richness items. He deduced an equation which would statistically assess the proportion of hapax legomena to other words in a text:

$$100 \times \log \frac{N}{1-V_1/V} - \text{where } N \text{ is the total length of the text in running words (tokens), } V_1 \text{ is the total of hapaxes and } V \text{ is the total vocabulary in terms of types}$$

[Coulthard & Johnson 2007: 165; Honoré 1979]

Honoré’s formula turned out to be faulty, however; it served to consolidate the measurement of both lexical and grammatical items. In the English language, a very few grammatical words tend to make up a very large portion of all texts; so, as the text size
grows longer, the proportion of grammatical words which are hapax legomena grows smaller. Winter & Woolls (1996) fixed this issue by altering the variable $V_1$ in the Honoré’s equation to read $L(\text{exical})V_1$, effectively restricting the variable to lexical items only (Coulthard & Johnson 2007: 165; Winter & Woolls 1996).

Clemit & Woolls (2001) further added *hapax dislegomena* (words which only occur *twice* within a text) to the list of measurements which can be used to determine authorship of a text (Coulthard & Johnson 2007: 166). Other measurements of lexical richness include lexical density, which is “the ratio of functional to content words frequencies in the text, also known as Functional Density” (Miranda & Calle 2007; Mikros 2009: 64), and relative entropy, which is “the quotient between the entropy of the text and its maximum entropy multiplied by 100” (Oakes 1998: 62; Mikros 2009: 64).

Later, Morton & Michaelson (1990) reported on the Cusum technique, which became the first test of authorship that was completely statistical in nature to be utilized as evidence for courtroom deliberations. Essentially, this measurement test assesses habitual writing items of authors:

Morton (1991) used the sentence, rather than the text, as his basic unit of measurement and then calculated the frequency of occurrence within each sentence, of variables like number of nouns, words beginning with a vowel, words consisting of three of four letters, or, most commonly, words consisting of two or three letters (2/3 lw). Each of these measurements was matched against sentence length as calculated in terms of orthographic words. A novelty of the method was that the results were not expressed, as one might have expected, in terms of the percentage of words of a given category per sentence. Rather, a calculation was made separately, both for the sentence and for the variable under consideration, of the CUMulative SUM of the deviation of both from the average for the whole
For this reason the method itself was labelled CUSUM. Graphs were then made from the resulting scores and one superimposed on top of the other. [Morton and Michaelson 1990; Coulthard & Johnson 2007: 167]

Function words are also posited as good markers for authorship attribution studies. Indeed, Koppel et al. (2003) determined through their research that function words were the most promising items to be used as a “universal, corpus-independent feature set for authorship attribution” (Mikros 2009: 66); they utilized the idea of ‘stability’ as the main determiner for their research, noting that function words are relatively unstable in that they can be synonymously switched out for other words as a choice of vocabulary. This would allow one to examine a document for a pattern of choice of words used as a measure of authorship. Mikros (2009) disagrees with Koppel et al, however:

Despite the wide acceptance of function words as authorship indicators, there are studies which demonstrate some serious drawbacks in using them in authorship attribution research. The oldest and most influential criticism was posed by Damerau (1975) who summarized the main deficiencies of the function words in the following arguments:

- The notion of function words is insufficiently precise to permit determination of what words to study.
- Not all function words are context insensitive.
- The distribution analysis of the function words revealed that many of them are not randomly distributed. [Mikros 2009: 66; Damerau 1975]

Instead, Mikros (2009: 65-7) introduces a measurement which narrows down function words to what he calls “Author-Specific Words (ASW)”, which are chosen from a text to provide the researcher with a list of words which both “characterize the vocabulary of each author and at the same time discriminate it mostly from the vocabulary of the other
authors” in an analysis grouping. The method by which ASW are produced will be discussed in more detail in the next section.

The sheer number of measurements which have been researched as potential markers for authorship attribution is vast. Indeed, Gerald McMenamin (2002) dedicated an entire book to stylistic measurements and the analysis of them as useful (or not useful) test subjects; there, he bibliographically lists 300 style markers that he has utilized in dozens of court cases surrounding authorship questions. Those markers were classified under the following broad categories: text format, numbers and symbols, abbreviations, punctuation, capitalization, spelling, word formation, syntax discourse, errors and correction, and high frequency words and phrases (McMenamin 2002, 216-31; Coulthard & Johnson 2007). Other researchers have also provided valuable feedback on the list of stylometric items, covering a spectrum which includes analysis of: spacing, capitalization, and end-marking (Philip 1940); frequency of characters in a text segment (Mikros 2009); top 100 most frequently used words (Stubbs 2002; Larner 2014); distribution of words by word classes, such as verb-adjective ratios (Buseman 1926; Särndal 1967); punctuation style (Thorndike 1948; Särndal 1967); and type-token ratios, or the ratio of one word’s occurrence to the total number of words in a text (Herdan 1960).

2.5. Methods and testing for authorship analysis.

It is evident from the brief synopsis of the last section that there are a great many ways to measure the writing in a document for the purposes of determining authorship, but how, exactly, are those measurements derived? What are the various methodologies employed to produce the data for these measurements? The answer is, once again, that there are
many ways to approach these questions. Coulthard (2004) describes the habits of human language use:

…whereas in principle any speaker/writer can use any word at any time, speakers in fact tend to make typical and individuating co-selections of preferred words. This implies that it should be possible to devise a method of linguistic fingerprinting – in other words, that the linguistic ‘impressions’ created by a given speaker/writer should be usable, just like a signature, to identify them. So far, however, practice is a long way behind theory and no one has even begun to speculate about how much and what kind of data would be needed to uniquely characterize an idiolect, nor how the data, once collected, would be analysed and stored; indeed work on the very much simpler task of identifying the linguistic characteristics or ‘fingerprints’ of whole genres is still in its infancy (Biber 1988, 1995; Stubbs 1996) [Coulthard 2004: 432]

The above quote indicates that each human speaker/writer will tend to make individual choices about which vocabulary and grammar rules to utilize when either speaking or writing, and that these choices will (more often than not) be habitual and characteristic to the individual. While the analysis of these choices and habits is still in its infancy, there are many statistical tests currently being tried out and refined, which can be useful in providing data for authorship determination. In this section, I will provide a few examples of various tests which have been utilized to provide good statistical data for use in authorship analysis through stylistic measurements. I will also briefly cover some good tips for choosing data sets; I will continue with a much more detailed discussion of methodologies in Chapter 6 of this dissertation.
2.5.1. Statistical tests.

Testing of stylistic measurements in determining authorship can range from simple formulas, such as the calculation of type-token frequencies – which involves a simple ratio equation – to some pretty complicated formulas which are best calculated using a computer software program, such as the uses of Bayes’ theorem, the Neyman-Pearson Lemma, or Chi-Square. Each of these tests serves a specific purpose in analyzing particular variables. Morton (1978) best describes the use of statistical testing in literary analyses:

Statistical method is based on the comparison of samples. A sample is a small group of observations taken from a large group, the population, and the statistics of the sample are useful in so far as they describe the statistics of the population. In literary studies, a sample may be a complete text, or it may be some part of a text; the population may be the whole corpus of an author’s writings or all the works he wrote in a particular literary form, …, or it may be the remainder of a text from which the sample has been extracted. [Morton 1978: 75]

Each authorship analysis, then, will necessarily be a unique case study in comparison with other case studies; this accounts for the myriad of ways in which to test literary data for variations.

Bayes’ theorem, for example:

…follows the principle of modifying one’s prior log odds by adding data information in the form of the weight of evidence (this term is used by Good 1950) obtained from the k selected discriminators, thereby arriving at posterior log odds which provide a measure of the ratio of the degrees of belief one ultimately holds for author A₀ as against author A₁. [Good 1950; Jeffreys 1961; Särndal 1967: 254]
Put into layman’s terms, Bayes’ theorem utilizes hypothesis testing to determine the probability that one writer is more likely than another writer to be the author of a text, within a certain margin of error. It relies on a “simple dichotomy” that either Author $A_0$ or Author $A_1$ was responsible for the creation of a text – there are no other possibilities with this distribution (Särndal 1967: 254). Särndal (1967: 259) stresses that it is a good practice to keep in mind when utilizing Bayes’ theorem for data testing, to use “discriminating words” which are going to be useful in subsidizing the evidence for authorship; the definition of *discriminating* in this case is that the words chosen for study are both highly distinctive and have a significant frequency of use in the disputed text.

However, when there is going to be difficulty in controlling the margin of error while using hypothesis testing - such as when the analysis expert goes into an authorship study with a certain belief that one author is more likely than another to be the author of a text, which changes the threshold at which the expert is willing to accept one test result over another – then the Neyman-Pearson lemma might be a better choice of test over Bayes’ theorem (Särndal 1967: 259). This form of hypothesis testing is predicated on the notion that it is beneficial to reduce the probability of error for one hypothesis, without necessarily raising the probability of error for the other hypothesis. (Särndal 1967: 260).

A third method of hypothesis testing is referred to by Särndal (1967: 263-7) as *diffuse alternative hypothesis*, which is utilized whenever an alternate author cannot be identified. This means the two hypotheses will not be a comparison between two potential authors, but an either/or situation: either Author $A_0$ is the writer of the disputed text, or he/she is not the writer of the text. In this situation, a Chi-square distribution analysis may be used (see also Lehmann 1959: 307). According to Butler (1985: 112-3),
“the chi-square test enables us to compare the frequencies we actually observe with those
that we should expect on the basis of some theoretical model, or of a hypothesis about the
distribution of the characteristics concerned.” Chi-square statistical testing is particularly
popular among the social sciences for the simple reason that it requires only nominal,
qualitative data to perform the testing, rather than requiring quantitative data; this allows
for the use of demographic data, characteristic data, etc. in analysis of a text (Faherty
2008: 141).

As was briefly mentioned in section 2.4 above, Mikros (2009) analyzed what he
called “Author Specific Words” (ASW), which when compiled would provide “the
frequency of the 80 most distinctive content words (20 per author) normalized in 1,000
words sample” (Mikros 2009: 65). In his discussion on how to properly compile good
ASW from a text, Mikros (2006, 2007) outlined a seven-step process:

1. Selection of the training corpus.
2. Formation of homogeneous subcorpora regarding the author of the included
texts.
3. Creation of frequency wordlists (FWL) for each of the subcorpora (for
example Author A FWL, Author B FWL, Author C FWL, and Author D
FWL).
4. Comparison of each FWL with the unified FWL of the remaining authors, i.e.,
comparison of Author A FWL with the FWL which has been created joining
Author B, Author C, and Author D FWLs.
5. Extraction of the $k$ most author-characteristic words that exhibit maximum
discriminating power between his vocabulary and the vocabulary of the other
authors. The extraction is performed using Log-Likelihood measure.
6. Repetition of the procedure (stages 4 & 5) by deploying the remaining
combinations of the available FWL comparisons.
7. Extraction of \( n \) words (in the previous example \( 4 \times k \)) which can be used as Author-Specific lexical variables in an authorship attribution training set.

[Mikros 2009: 67]

This methodology produces vocabulary lists for each author in a closed set which can be used to differentiate the writing of one author from another author in that grouping.

Mikros (2009: 67-8) stresses that the use of ASW only works with a closed set scenario, where the actual author of a text is definitely present in the group of authors whose writing is being analyzed; it will not work in situations where the author is a potentially unknown element or the group of suspected authors is not fully representative of the possible authors for a text.

2.5.2. Choosing data sets.

There is a significant portion of the forensic linguistics community who would rank lexis over other measurements as being paramount to the determination of authorship. Indeed, Larner (2014: 8) discusses this very thing, indicating that Gibbons (2003: 303) “claims that particular words may contribute to identifying an author” and Winter (1996: 165) “claims that vocabulary should not only be the starting point when differentiating authors, but ‘should be the standard approach to any linguistic analysis which compares texts’.” This section will cover a very few of the possible methods for determining what those vocabulary data sets should look like in an authorship study.

Mosteller & Wallace (1964: 249-251) developed a model of procedures for determining the data sets they would work with while analyzing the authorship of The Federalist papers, as paraphrased below:
1. For each suspect author, gather together a good amount of their known writings, to act as a baseline for comparison.

2. Choose what rules you wish to use in handling numbers, formatted items, quotations, etc., and be prepared to consistently apply those rules to all writings, of both known and unknown authorship.

3. For statistical ease of use and to gain more accurate data results in variation analyses, choose a suitable length of text (at least 1,000 words each) and divide each author’s writing into blocks of that length; be consistent in doing this with each text to be analyzed.

4. Before beginning your analysis, put together a list of words that you wish to analyze frequencies for, and stick to this list consistently for all texts.

5. For each word in your list of to-be-studied words, count and determine the frequency of occurrence of that word for every 1,000 words in the authors’ known writings; compare those frequencies with the other authors’ frequencies of that same word-use in their writings, and record the patterns of vocabulary use that appears.

6. For each word, compare the frequency of its use between and author’s known writings. If the variation of use is large, then the word should be discarded from the list as unsuitable; for one reason or another, it will not yield accurate results about the generalized usage of that word by that author.

7. Be mindful of the natural groupings an author’s writings may fall into, such as different time periods or different genres; the word usage a single writer may use in one genre or time period, for example, can vary greatly from one genre or time
period to another, and these vocabulary variations of some words may not be useful in an authorship analysis as a result – discard words which fall into this category.

These steps as outlined by Mosteller & Wallace (1964) can help with determining which words will be useful in an authorship analysis. Indeed, the forensic linguistic research conducted over the past several decades has served to provide several lists of modern English words which can be considered to be useful for testing in authorship cases. However, when working with ancient languages, such as Old English, Old Norse, Medieval Welsh, etc., those word lists have not yet been accumulated. More work will need to be done to accomplish this for future studies.

A common question in deciding on data set sizes is: what is the correct size a text sample needs to be in order to be effective? Morton (1978) addresses this with his discussion of the principles to employ in determining a bare minimum necessary sample size, indicating several factors:

1. It is necessary to take a sample with at least five occurrences of whatever event (words, phrases, grammatical element, etc.) is being observed. The sample size will differ for each experiment based on whatever sample size will provide at least those five examples of the event (Morton 1978: 84).

2. A minimum sample size should ensure that periodic effects are unimportant. Using serial correlation testing to indicate the presence of such effects enables the experimenter to adjust the sample size to account for them (Morton 1978: 84). Morton (1978: 86-7) discusses two different serial correlation tests that can be used for this purpose: the runs test and the differential distribution test.
3. Cusum charts can be helpful in assessing correct sample sizes, but they are rather cumbersome to put together. They can indicate where serial correlations exist and whether or not these correlations will be a problem for authorship analysis purposes. These charts are utilized to compile and show the tests of authorship performed throughout a text, and they were devised to indicate when there are changes in the mean of data measurements (Morton 1978: 87-9).

Now, with regards to early texts as opposed to more modern, printed ones, Farringdon (1996) has this to say about sample sizes:

The greatest contrast imaginable is the attribution of a complete book and of a single disputed sentence. … Theoretically, every single sentence should be analysed to ascertain that there have been no insertions, plagiarism, or half-conscious ‘borrowing’ in a book which appears under a single author’s name. For pre-Gutenberg texts, every sentence of what first appeared as a written manuscript must be analysed; but for modern texts, careful analysis of a number of samples from different parts of the text will give a high probability of authorship and is usually reliable. [Farringdon 1996: 183-4]

This distinction in sample sizes between handwritten texts and printed ones may be predicated on a number of different factors, including differences in use of punctuation, text formats, and spelling conventions (or lack thereof of any of the above). I would also posit that the differences in attitude and belief regarding the idea of authorship (as I will discuss later, in Chapter 4) in the medieval and early Renaissance time periods versus more modern writing practices would also play a role in determining necessary sample sizes for text analysis.

From the discussion above, it is clear that determining variables which can be viably analyzed in medieval manuscripts may turn out to be much more difficult that it would be
for modern documents, and it is definitely necessary to have a good working methodology which can help steer the researcher in the right direction for the selection of those variables for data sets. At least as it pertains to the use of forensic linguistics for authorship attribution, McMenamin (2002:119-25) does put forward a quite comprehensive, flexible methodology for the identification of authorship in current forensic linguistic cases, which I will discuss in much greater detail in Chapter 6. It is to be hoped that some of that framework can be utilized in developing a flexible methodology for working with medieval manuscripts as well.

2.6. Authorship debates over medieval manuscript texts.

During the course of my analysis of forensic linguistics in general and authorship attribution in particular, I have determined that there are other uses for the field of forensic linguistics, in providing supplementary analysis of an interdisciplinary nature. One of those possibilities lies within the field of medieval manuscript studies, in scientifically assessing the varying authorship debates of the Middle Ages. To date, most of these authorship debates have been adjudicated by the academic community via literary analysis alone, or via literary analysis with only the barest of linguistic references. One such example is that of Scattergood’s (2006) brief analysis of the authorship of the fourteenth century debate poem, The Boke of Cupide, in which he gave a detailed historical and literary account of the likelihood that Sir John Clanvowe was the author of the text, and only the briefest linguistic evidence of the dialectal elements. There have been precious few analyses performed from a more scientific linguistic perspective, but they are increasing in recent years – there is some hope that this line of
inquiry will continue to grow as the benefits of statistical, scientific testing become more apparent to the larger medieval academic community at large.

Debates over medieval authorship of manuscripts range in genre from liturgical texts, such as with the Old English gloss to the *Lindisfarne Gospels*; to apocryphal texts, such as that of several works attributed to Chaucer, which were later reascribed to other authors after closer review (Erdman & Fogel 1966); to legal canons, such as the Laws of Edward and Guthrum (written in the early eleventh century and revised as to its authorship several times thereafter) (Whitelock 1941; Erdman & Fogel 1966).

The *Lindisfarne Gospels* authorship debate is one where a few case studies have been performed which may be classified as having utilized forensic linguistics. For example, Cole’s (2017) study on the subject and adjacency effects which provide the basis for the Northern Subject Rule (NSR) addresses linguistic variation and the diachronic timeline of the NSR, but it also provides valuable information regarding the authorship of the Old English gloss and whether Aldred was the only glossator to work on that manuscript or not. Kotake’s (2018) study on the source materials which were potentially used for the glossing of the *Rushworth Gospels* addresses the question of whether or not the glossator Owun copied his gloss directly from the *Lindisfarne Gospels*, or if he utilized other source materials for his work. I will elaborate more on these in Chapter 5’s case study.

While there are a few case study researchers who actively employ (either consciously or subconsciously) forensic linguistic testing in any manner on their data taken from medieval manuscripts, there are many, many more researchers who would greatly benefit from integrating such testing into their medieval case studies. For
example, Bald (1966) conducted an analysis of the manuscript of *The Booke of Sir Thomas More* (MS Harley 7368 in the British Library); his goal was to elaborate on the matter of whether or not Shakespeare had a hand in authoring this piece of writing. In addition to analyzing the manuscript’s qualities themselves, as well as the dating of the play, Bald also analyzed the evidence of handwriting and spelling within the manuscript, effectively bringing elements of forensic linguistics/stylistics into the examination. As can be said for any manuscript analysis, it is of course impossible to determine with 100% accuracy the authorship of such an old piece of writing; however, Bald does surmise that the analysis conducted strengthened the argument that Shakespeare was indeed the author of this document, while asserting that

> With the increase of knowledge, …, not only of the handwriting, printing, and spelling of the Elizabethans but of their habits of thought as well we gain a fuller understanding of both Shakespeare and his fellow dramatists. [Bald 1966: 171]

On the other hand, an authorship case which could benefit from a forensic linguistic analysis would be Sisam’s (1962) study on Lambard’s 1568 *Archaionomia*. Sisam analyzed this text and determined

> That Lambard’s texts of *I Athelstan* and *Ordinance* have a single source appears to be certain, for they are distinguished from the other texts of the *Archaionomia* by an extraordinary neglect of the rules of Old English accidence and syntax, and by a diffused caninity of expression. … as Lambard prints other parts of the manuscript with tolerable accuracy, it is not obvious why he or his printer should have produced from it the remarkable jargon of *I Athelstan* and the *Ordinance*. [Sisam 1962: 232-3]

In Sisam’s case study, his research may have benefited greatly from some statistical analysis of some of the instances where that ‘extraordinary neglect of the rules of Old
English accidence and syntax’ occurred, in comparison with a few other extant manuscripts in the *Archaionomia*, or even the exemplar for creation of the *Archaionomia*. However, here we run into a problem all too common when working with manuscripts: how do we prove which manuscript was utilized as an exemplar in the creation of Lambard’s version, when such a manuscript appears to no longer exist in physical form?

2.7. Relevance of forensic linguistics to medieval authorship studies.

As is typically the case with regards to medieval studies analyses utilizing syntax queries, the cases discussed above never mentioned the words *forensic linguistics* in conjunction with their studies. In fact, these types of studies seem to have been classified as being related to diachronic syntax or historical linguistics only. While they do indeed fall into one or both of these categories, they also would fall into the category of forensic linguistics. While forensic linguistics as an official field of linguistics came to be in response to a need for scientific inquiry regarding language use in all its forms, for the sake of mainly settling court cases and investigations, it is a flexible enough field to now be used for other purposes.

As an example, Hoy (1966) utilized some stylistic/forensic testing in his discussion of the authorship debates surrounding the seventeenth century Beaumont and Fletcher canon. In his essay, Hoy analyzed such linguistic items in the canon as forms of pronouns used, third-person verb forms used, and utilization of contractions, which served to aid in confirming some of the previously asserted attributions of authorship to these works. This case study, while also a matter of historical linguistics, qualifies as a forensic linguistic study as well, at least as far as the subfield of stylistics is concerned. The study focused on linguistic items in the seventeenth century Beaumont and Fletcher
canon and the findings were more robust with information for having had stylistic testing as a basis for its data analysis.

In reference to Malcolm Coulthard’s work on the 2005 Jenny Nicholl murder case in the United Kingdom’s court system (where Coulthard’s analysis of text messages from Jenny Nicholl’s cell phone both before and after her disappearance led to the conviction of her boyfriend for her murder), Perkins & Grant (2013) claim that “Coulthard’s approach focussing on within author consistencies of style and between-author distinctiveness has been recently shown to be amenable to statistical analysis to support the decision making of the forensic analyst” (Perkins & Grant 2013: 176). If it is possible to utilize statistical analysis to support the decisions made by forensic linguists when used in comparison with their other data taken – such as handwriting analysis, document analysis, syntactic structure analysis, etc. – for the purposes of determining authorship in modern cases, why can’t these techniques translate into a format that can be reliably used for medieval manuscripts as well?

2.8. Conclusion.

In this chapter, I have endeavored to cover a vast amount of information with regards to an entire field of linguistics, in a very short space. I began with a general definition of forensic linguistics as a whole, and then focused more fully on the subfield which addresses authorship attribution - stylistics. I examined some of the myriad ways in which stylistics can be used to test for authorship, elaborating on just a few of the many tests and measurements possible to choose in researching an authorship issue, as well as a few ways in which to determine how a data corpus should be chosen. I then provided some examples of historical documents which have either been studied using what could
be called forensic testing or which could benefit from having such testing applied to them for analysis.

Suffice it to say, the field of forensic linguistics is an extremely flexible field, with room to expand into new areas of research than what it is currently being used for. While this field of study began its official life as a means to settle court cases of varying types from a linguistic standpoint, I would posit that it can also be used within the interdisciplinary realm of medieval studies and manuscript studies, especially for authorship cases, but potentially for other linguistic studies as well.
3.1. Introduction.

In this chapter, I will discuss the various aspects of manuscript analysis and how it is important to medieval studies in general, from both a cultural and linguistic outlook. This very brief look at what is a vast field of study will help to set the stage for my later discussion of forensic linguistic methodology development for use with manuscript analysis, and the limitations that such a methodology will come up against when used with manuscripts. I will begin with a description of what constitutes manuscript analysis, and then provide some cultural background of the medieval time period and the conditions under which such manuscripts were produced. Then I will quickly outline both the codicological and palaeographical elements of manuscripts, and provide some information on how the production of manuscripts changed as the early medieval period led into the later medieval and early Renaissance period (and subsequently the invention of the printing press). Finally, I will discuss some of the types of errors which frequently occurred in both creating manuscripts and in copying them by hand.

3.2. Definition.

Manuscript analysis constitutes a wide array of information and subfields as it pertains to the study of ancient and medieval manuscripts. Diringer (1976) provides a clear, concise definition of what a medieval manuscript is:

The hand-written parchment or vellum codex is ‘the book’ par excellence of the Middle Ages. Such codices are generally termed ‘manuscripts’ of MSS., from the Latin term codices (or libri) manuscripti ‘codices (or books) written by hand’.

From the earliest codices extant, belonging to the third or fourth century C.E., to
the invention of the printing press in Europe, c. 1450, and the spread of printing in the second half of that century, giving birth to the ‘modern book,’ ‘the book’ of the Middle Ages covers a period of well over eleven hundred years. [Diringer 1976: 27]

While a great many of these medieval texts have, sadly, not survived this past millennium to be studied today, and many of those manuscripts which have survived and have been preserved are only partial bits and pieces of what they once were, we still have an enormous range of documents available to us for research into the anthropological aspects of medieval culture. Brown describes the purpose of the medieval book as a way “to introduce readers to the history, culture and art of the Anglo-Saxons by means of a survey of their surviving manuscripts” (Brown 2007: 7). Indeed, this description could be extended to cover much of what was written down on the Continent as well as the Insular islands during the medieval era.

Brown (2007: 7) goes on to describe various examples of surviving medieval texts, including ecclesiastical works such as bibles, gospel books, psalters, and liturgical texts (i.e., the Lindisfarne Gospels, Vespasian Psalter, and the Old English Hexateuch); the incitements of preachers (i.e., Ælfric and ‘the Wolf’); legal documents such as wills, charters, tax records, estate records, and codes of law; and literary epics such as that of Beowulf, the Exeter Book, and the Vercelli Codex. Also preserved are historical texts such as Nennius’s copy of the Historia Brittonum, Bede’s works, and the compilations comprising the Anglo-Saxon Chronicle. Finally, we have learning texts available to us which were utilized in monasteries and school environments (these locations were often one and the same during this time period), such as Latin and Old English glossaries, Pliny’s writings, and Aldhelm’s Riddles, among others.
Manuscript analysis itself can cover so many different areas and the term intersects with a wide array of various fields of study, ranging from physical production techniques, to determination of creation and/or provenance of a manuscript via various means, to language study, to scribal practices, to text-sharing practices. For example, Gullick (1995) analyzes several eleventh century manuscripts and the historical records associated with them to determine how fast scribes were expected to write. Owen-Crocker & Cesario (2009) outline many specific elements involved with the handling of Anglo-Saxon manuscripts, such as the size and shape of medieval books, the mise-en-page (layout) of a book, etc. Diringer (1976) discusses some of the environmental factors surrounding bookmaking in monasteries during the Middle Ages, providing valuable information about how scriptoriums were setup and how they were used in creating texts. Historical linguists can intersect their research with the study of scripts used and a text’s grammatical and other structural elements, in order to learn more about the language structures used in medieval writing. Palaeographers utilize the analysis of manuscripts to not only analyze a medieval text’s script and textual structure within the work, but to understand a transcription more fully by comparing the writing with the illuminations of the work and analyzing the text’s commentaries and emendations. Art historians study the illuminations in medieval texts, many of which are tactfully placed in manuscripts to tell a story of their own or to provide visual backup to what is written. The list of interdisciplinary uses of manuscript analysis goes on and on.

3.3. Elements of manuscript production.

In the field of manuscript analysis, medieval manuscripts are examined on a couple of different levels: a codicological level and a palaeographical one. It is entirely possible to
focus and specialize on either of these subfields of manuscript analysis; they each cover an enormous amount of ground in terms of research into manuscripts.

Codicology, or “the examination of the physical characteristics of the manuscript book or codex” (Rumble 2009: 29), involves detailed analysis of bindings and binding fragments, flyleaves, indexes, tables of contents, leaves, and any marks made of previous ownership for each manuscript. This form of analysis also covers the production process involving the planning and construction of the codex’s text-block, which involves the foliation and pagination of the text, collation of folio pages, numbering of quires, and research into any quire-signatures and catchwords found, as well as booklet formation.

Palaeography, on the other hand, is the study and “the classification and description of former styles of handwriting” (Rumble 2009: 29). This subfield covers the analysis of the various writing scripts used – depending on the time period, location of the manuscript production, and training of the scribe or scribes working on the manuscript, the shape of the letters conformed to whatever style was taught at that time. For example, which Anglo-Saxon script was utilized in creating a manuscript depended on whether the time period fell before or after the Benedictine Reform occurred in the second half of the tenth century. Prior to the reform, the most common Anglo-Saxon scripts used were the Uncial, Half-Uncial, and Insular miniscule. After the reform the scripts used were shifted to be more in line with the rest of Europe in its use of Caroline miniscule in Latin texts, although this was altered slightly by the early eleventh century and became what we know as Anglo-Saxon miniscule. (Rumble 2009)

On a more detailed level, palaeography also involves the analysis of specific aspects of a script used in writing out a manuscript, as well as the orthography,
punctuation, rubrication, and headings used. Some of those specific aspects of script type include the use of ascenders and descenders in letter formation, individual letter forms themselves, minims, ligatures (combined letters), and common abbreviations. The decorations analyzed include not only the picture drawings in a manuscript, but also decorated initials and diminuendo, which were often used to delineate the beginning of new sections in a text (Rumble 2009). Corrections and alterations perpetuated by scribes or copyists other than the original writer of a text after the completion of a manuscript are also intensely scrutinized; these types of items can lend valuable insight into the possible changes of, say, spelling conventions over time or potential changes in the cultural viewpoints of the next generation of scribes.

Scragg (2009) provides a few other items to keep in mind when we examine a manuscript, in addition to the ones mentioned above. These include book varieties and their intended purposes; scribal care in production as well as the types of errors they make; the fact that a particular text may have multiple surviving copies; the outside ornamentation of bindings; marginalia; and the manuscript’s origin(s) (i.e., where it was written) and its provenance(s) (i.e., where it was kept). The origin can be more than one place if, for example, the manuscript was written in its original language in one location and then later glossed at another location; in this case, what is considered the origin in an analysis would depend very much on which language in the manuscript was being examined (the original, or source, language or the gloss, or target, language). In realizing that the provenance could also change over time for a manuscript, it is important to remember that these texts were potentially highly mobile during their peak periods of use. I will talk more about that below.
The subfield of palaeography is also where we find an important interdisciplinary crossover with linguistics. Here, both linguists and palaeographers might wish to examine the language a manuscript was originally written in, as well as any subsequent glosses in a second or even third language were later added to it. Script changes are also examined in this context, as both a cultural and linguistic statement. Language analysis of a manuscript can cover many different avenues, including syntax used, dialectal variations in comparison with other texts, translation choices in glossing the original text, and identification of originating or correcting scribes and/or place of production via added colophons and commentaries added to the manuscript. It is this particular crossover I am concerned with for the purpose of this project, for this is where the analysis of attributed authorship of a manuscript would also necessarily fall.

3.4. Ordinatio and communicare.

In describing the elements (both codicological and palaeographical) of producing a medieval manuscript text, I would be remiss if I did not address the extremely important concepts of ordinatio (to order) and communicare (to communicate) as part of the process. Ordinatio refers to the actual layout of a text, including the pre-writing planning that went into deciding how a text would be separated into chapters and sections, and even into smaller elements such as paragraphs and sentences; it was in the ordinatio phase that the presentation of initials, paragraph markings, punctuation, etc. was also decided on (Snijders 2015). Also included is the script to be used, quality of the parchment or vellum to be used, and the planning of what illuminations would be created as part of this phase of planning, and stresses that “[t]he application of the ordinatio was seen as the most essential phase in writing, as a clear layout ensured that the scriptum
could transmit its message to the audience” (Snijders 2015: 37-8). Communicare is the transmission, or communication, of the text’s message to that audience.

Both Snijders (2015) and Scragg (2009) are very careful to point out that the notion of ordinatio in the Middle Ages, and indeed the entire idea of what an actual book comprised, is most decidedly not what we would envision them to be in today’s world of publishing. Many of these particular aspects I will address later, during my discussion of the concept of authorship, but I will put a few beginning thoughts as they pertain to ordinatio and communicare here as well.

One such thought regarding the idea of what a book is comes from Scragg (2009):

The medieval reader has a very different idea of a book, and therefore coming to surviving manuscripts with modern preconceptions has its dangers. Very many surviving vernacular manuscripts resemble miscellanies or commonplace books, and rarely have any sort of title. While it is always important to look at the manuscript context of any surviving text, contiguous items may not necessarily have any connection in terms of subject, at least to our eyes. … On the other hand, some manuscripts are very carefully organised around a definable theme. … Even when a manuscript appears to have an internal consistency, however, there can be no assurance that it was planned prior to the process of writing in the form in which it survives today. [Scragg 2009: 61]

This conceptualization of the medieval book indicates two things. The first is that the ordinatio and final format of a medieval text may not be established in an order which is readily obvious to our current day mode of thinking; subject matter which had a direct purpose at the time the manuscript was produced may not make sense to us now, in terms of how the materials were ordered or combined with other subject matter. Second, a medieval text as it survives today may be in a different form now than it was when it was
constructed originally; this happened quite frequently, when books were taken apart for
the sake of being copied or transcribed during the production of the next generation of all
or part of that manuscript’s materials. For example, if someone requested that a copy of
an existing manuscript be produced (i.e., he or she wanted a few selections copied from
the *Exeter Book*, as well as full copy of the *Vercelli Book*), then the resulting manuscript
may be both ordered differently than the original two texts were, and may also be
partially incomplete in comparison; but the new manuscript would be exactly what was
ordered when the medieval book collector made the request.

This attitude of painstakingly planning out the *ordinatio* and *communicare* of a
medieval text, producing said text, and then later on potentially taking apart and
rearranging that text seemingly at will – although it is clear that there was a reason for it
at the time, whether we can see what that purpose was or not – is very different from how
we view books today. It is also not properly appreciated by many modern-day editors
whose job it is to transform those bodies of medieval text into a uniform printed or digital
copy for modern readers:

The importance of a scriptum’s *ordinatio* is not always fully realised by present-
day scholars. For example, Fred Robinson has noted that even though the chapter
divisions in most Old English works were consciously placed in such a way as to
emphasise its most important sections, modern editors still consistently replace
those divisions according to their own modern insights, thereby discarding
valuable information regarding the conceptualisation and use of these works
during the Middle Ages. [Robinson 2005: 366-67; Snijders 2015: 38]
Kleinhenz elaborates on this point by emphasizing that modern editors need to integrate
their own knowledge of the text and various other factors, and challenge themselves to
“reconstruct” or “recreate” the original intent of the original work as faithfully as possible (Kleinhenz 1976: 21).

3.5. Early medieval culture and manuscripts.

It is virtually impossible to begin to understand the thinking and planning process that underlies the *ordinatio*, creation, and writing of any medieval manuscript without having at least a little background information on the cultural context surrounding its creation and subsequent use.

The literacy culture in Anglo-Saxon England was mainly contained within the ecclesiastic institutions, creating a minority of educated elites for which manuscripts were produced, most of whom were directly affiliated with the Christian church. That being said, however, “the breadth of the education enjoyed by that elite was remarkable, and these privileged few regularly crossed what today would be seen as academic boundaries” (Owen-Crocker 2009: 3-4). Many literate scholars were well-versed in different script formats and could read and write in more than one language (most commonly English and Latin in Anglo-Saxon England). Memorization, while part of the ecclesiastical curriculum, was not the only method of learning; these scholars were researching and notating results, glossing texts, and creating new textual materials. Indeed, within the texts created, scholars utilized what we merely see as different colored inks and pretty decorations as markers for readers, helping them to work their way through the texts; and illuminations were also utilized to either reinforce the textual information provided, or to tell an entire story on their own (Owen-Crocker 2009: 4).

Doyle (1990) provides some insight regarding manuscripts of the early medieval time period as being frequently ascribable to certain production centers based on
“distinctive habits of preparation, writing, punctuation, or decoration, associated with domestic training, team-work, and patterns” (Doyle 1990: 3). So, early medieval production centers had their own styles and manners of working on their manuscripts, and their habits were different enough from one location to the next as to make it possible to distinguish provenance of a document to a certain extent.

However, one cannot always assume provenance of a particular manuscript based on a production center’s distinct habits alone. It is also necessary to keep in mind that both competent scribes and well-made manuscripts were considered to be commodities in demand in the Middle Ages, and they were both highly mobile.

Distinctions between Insular and Continental book production were often blurred by the movement of text exemplars and models for illumination and by the far-reaching journeys of a few trained monastic scribe-artists, whose influence was not limited by geographic borders. Owing to this activity, the study of codicology, parchment preparation, script, and the contents of individual books can help to clarify the history of an early Continental writing center like that of Salzburg in the later eighth century. [von Daum Tholl 1995: 17]

It was not unusual at all for monasteries and other book production centers to lend out their texts to one another to be studied and copied (Lesne 1938: 428-34; Snijders 2015: 45), and they were well-versed in what one another’s current collections were, for the sake of making those trades back and forth. There is also some proof that some codices were also created and given as gifts in good faith, in order to create or maintain community relations within the monastic network (Mauss 2000; Bosterveld 2007: 104; Snijders 2015: 45).

Early medieval manuscripts were not the only mobile things within the book production community circuit; well-trained scribes were in high demand across this
community and were frequently either loaned out to other institutions for their services in creating manuscripts, or they travelled on their own to other locations in order to make a copy of a much-coveted text or to study it up close in greater detail.

This network influenced the layout, composition, and instability of scripta and manuscripts. The way a scriptum was laid out was closely connected the technical know-how and models that were available within the network. The reason to produce a new manuscript was frequently founded in considerations of rivalry or collaboration with a neighbouring community; changes in scripta were often made because the expectations and preferences of one monastic community would be slightly different from those of their neighbours. [Snijders 2015: 241]

Obviously, this networking between monasteries was a significant communal effort, indicating how very important the creation and dissemination of medieval texts really was.

3.6. Production changes over time.

The later medieval period saw some distinct changes in the book production environment, in both England and on the rest of the Continent. Derolez (2003) addresses the changes which occurred with regards to the production of manuscript books between the twelfth and fifteenth centuries. Due to the rise of more urban populations and the subsequent increase in administration, commerce, and trade, more of the urban upper-class people were learning to read and write, in addition to the already educated clergy. However, unlike the clergy, these laypeople were learning to read and write in the vernacular language available, rather than in Latin; this necessitated a new demand and supply of books written in the vernacular, both liturgical books and books for other purposes.
By the thirteenth and fourteenth centuries, monastic orders as book production centers (aside from a few remaining large centers) were declining, while universities and their surrounding municipalities were the new rising stars of generalized manuscript book production for the masses. Universities began to utilize a system of book copying (for textbooks and otherwise) which involved creating an original text or approved, corrected copy of an original, called an exemplar. The exemplar was then rented out to students in pieces, or peciae, and the students would copy it and then return it to the school and rent out the next portion of the text for copying. This peciae method facilitated swift mass production of a source text for many students (Destrez 1935; Boyle 1984; Bataillon et al. 1988; Derolez 2003).

By the fourteenth and fifteenth centuries, the monastic houses had taken on a much more uniform manner of writing and creating manuscripts; Doyle (1990: 3) indicates that this might mean that scribes and other religious members were by that time receiving a good deal of their education in the same learning centers as that of the growing secular and lay literate scholars. This also means that it became much more difficult to attribute a text to a certain book production center merely by examining the techniques and script used in its creation. Commercial book production in the lay community was more traditional than that of the universities and their peciae system, but they also became more standard in terms of the physical aspects of book-making, including materials sourced, layouts of written materials, and techniques of decorating and structuring texts (Derolez 2003).

The second half of the fifteenth century saw the advent of the printing press. Much of the more modern scholarship and research performed on texts produced with
this new invention in the later Middle Ages/early Renaissance period have focused a great deal on the idea that the printing press and all of the book production changes that necessarily came with it caused a mass standardization of both the format and content of books (Hellinga 1983; Scattergood 2006). It was suddenly possible to mass produce virtually identical books for larger populations than ever before; unique, one-of-a-kind, handwritten manuscripts gave way to uniformly printed pages where script forms, layouts, font sizes, and printing materials were very similar to one another. Putting books into print served to eventually bring about the standardization of the vernacular languages as well, as it became easier to create uniform texts and provide them to a much broader network of people if printers could edit out dialectal differences and more archaic features, as well as stabilizing grammatical structures and punctuation rules.

While this world-changing invention did make it possible to provide reading materials to anyone who could afford to purchase them for reading and learning, rather than catering to an elite or upper-class group of scholars, there were also some serious drawbacks that came with that. Hellinga (1983) addresses this by inferring that a general attitude began to arise in the general population, where “with the invention of printing, the manuscript form in which the culture of the past, the classics, patristics, and anything else written in the past millennium, had been preserved, had ceased to have any importance once it had appeared in print” (Hellinga 1983: 3). Also, that “stabilizing” effect caused by the standardization of texts and the subsequent linguistic form of the vernacular which would come to be used in them, “significantly contributed to the marginalization of archaic and dialectal features of the language” (Scattergood 2006: 25).
3.7. Manuscript creation and transcription errors.

Medieval manuscripts, being handwritten and hand-created in every way, were by nature of their very existence, unique to one another. Even a copy of a manuscript could not be a perfect replica of its predecessor.

Before the appearance of print technology, each occurrence of a written work was unique, the product of a specific, historically conditioned intersection between one or more authors, one or more scribes, and the material conditions of a particular manuscript’s creation, which might include the state of the exemplar; the skill of the scribe or scribes in terms of calligraphy, illumination, and graphic design; and the proposed or *de facto* literary context, that is, which works were chosen or simply available to be copied and bound together to form a manuscript book. [Lionarons 2004: 67]

I will not go into the methods of constructing the physical aspects of manuscripts here; that would take entire volumes of discussion and it is not really within the purview of this particular study. For an excellent, detailed description on how a medieval manuscript is physically created, refer to Timothy Graham and Raymond Clemens’ (2007) book, *Introduction to Manuscript Studies*, for a thorough explanation of that subject.

I will, however, quickly touch on some of the practices employed by scribes when selecting layouts and performing the writing of their texts. According to Snijders (2015), within certain environmental constraints, scribes generally had significant influence on the manuscript *ordinatio* process, through the expression of subdivisions of text, the punctuation they used, ink colors and where they were utilized in the text, and the physical dimensions of the text on the page. In this way, a scribe could create *communicare* almost as effectively with the look of the layout and what they put in it as they did with the words they wrote in the text. It should be noted, though, that this
freedom to be able to select the layouts and mediums of expression as befit the text and the scribe(s) involved with the book production was available prior to the Benedictine Reform of the English church in the late tenth century; after the reform was complete, monasteries began to enact a more rigorous rule system regarding the look and layout of manuscript texts being created in England. I suspect that this reform is one of the reasons for the eventual decline of the monastic orders in the twelfth and thirteenth centuries as the pseudo-official large places of learning and book-making in Anglo-Saxon England, but that is a discussion for another time.

One aspect of scribal influence which was perhaps unconscious during the course of the making of a manuscript was that of human error. As I mentioned above, unlike with modern printing, where production of many identical, faithful copies of an original work may be made, manuscripts were necessarily handwritten, and each one was unique. Human error was not only likely, but virtually guaranteed, and the more generations of copies made of an original work, the more errors and deviations from that original text were likely to show up in extant texts. These errors might be caught by the scribe(s) working on the original text in its initial creation stages or after it, and emendations to the text may be made to correct them; or later copies of the original may reflect errors, emendations, corrections, changes to the text, etc., depending on the scribe currently working with the textual materials and making decisions regarding how it was to be treated.

The use of *apopeciae* (a copy of a copy) of a text has an interesting role to play in the study of medieval manuscripts:
The place of apopeciae in establishing a critical text of a scholastic author of the 13th or 14th century is an important one. Apopeciae on their own may not establish exactly what an author wrote – that will depend very much on the fidelity of the stationers’ peciae to the epipeciae from which they were copied and on the fidelity of the epipeciae exemplar to the author’s autograph or apograph – but they will allow one to establish the text of the author as it was in common circulation and was taught or commented on and used in university circles. [Boyle 1986: 74]

Now, normally it might seem that human error and the correction of it in medieval book-making might be a drawback, because a text could never be “perfect” for a reader. However, it is the very nature of human error in the use and writing of language which can be a factor of identification for the studious historical linguist who is looking to compare one text with another in search of clues as to its dating, provenance, location of creation, or even its authorship. There are, of course, other factors which may be examined (syntax, grammar use, dialect, etc.), but errors in a document provide a great deal of information as well.

3.8. Language use and glosses.

Anglo-Saxon manuscripts were, for the most part, written in Latin or in Old English, and many were glossed either in their entirety or in part, to aid the reader in translating the text into or out of the vernacular as needed. According to Wieland (2009), it is necessary to understand both the Latin and the Old English manuscripts produced in Anglo-Saxon England, in order to better comprehend how Anglo-Saxons saw their culture and the place of books and educational information in it. Both languages were important to that society, and while early medieval Latin manuscripts tended to be largely ecclesiastical or historical in nature, the religious and the mundane were frequently able to overlap one
another. Old English manuscripts tended to be counterpart to much of the material covered in Latin manuscripts, and when the Old English manuscripts were direct translations of Latin texts, they usually tried to retell the Latin versions from an Anglo-Saxon perspective (Wieland 2009: 157).

It is also, however, important to keep in mind that, ironically, on a more personal and individual level of knowledge, just because a manuscript was created or owned by someone in Anglo-Saxon England, this only proves that the work was potentially known there; it does not prove that the manuscript was utilized in any kind of educational or reading format:

A manuscript may reflect the personal interests of one reader, or the collective interest of a monastery or cathedral. A pristine manuscript with no annotations allows practically no determination of how it was used. Moreover, many manuscripts were destroyed either during the Anglo-Saxon period or in the centuries afterwards. Their destruction and their disappearance will never allow a full reconstruction of the holdings of Anglo-Saxon libraries and of the texts available to the Anglo-Saxons. Any look at the Anglo-Saxon past will at best be ‘through a glass darkly’. [Wieland 2009: 157]

It is the extra materials added to a manuscript, in addition to the script and the text material itself, which can give us the most information about how a manuscript was actually used while it was owned or held in a collection. This shows up in the notations in the margins and between the lines, in the glosses, through the commentaries and colophons, and in the emendations and corrections to a text over time. They are definitely not to be discounted in terms of their usefulness when determining authorship, provenance, or a number of other things about the manuscript and its creators.
This brings us to the addition to an original text of a gloss or glosses. “Any entry of script the purpose of which is to help explicate the original text of the manuscript may be called a gloss” (Graham 2009: 159). For Anglo-Saxon manuscripts, this could either be manifest as part of the *ordinatio* of a text, where the originating scribe(s) always meant to create at least a partially bilingual text and included the gloss as part of the planning process; or the gloss was added much later on, by other scribes who wished to make the text easier to read for its current generation of scholars. It is this second class of glosses which can help shed some light on how later generations of historians and other scholars received and processed the information from the medieval period that had survived long enough to reach them for analysis (Graham 2009: 159). Linguistically, it is especially interesting to note that glosses were applied to different manuscript languages, depending on the time period. Anglo-Saxons chose to gloss Latin texts with Old English, to provide vernacular help in reading Latin, while readers from later time periods would gloss Old English manuscripts, either in Latin or in a later version of English, to aid in comprehending Old English (Graham 2009: 159).

### 3.9. Conclusion.

I have not used the term *author* in conjunction with medieval manuscripts as of yet, preferring to use the terms *scribe* or *originator* for the time being; the reason for this will become clear in my later discussion regarding the meaning and sense of authorship in the Middle Ages versus today’s definition of the term. Suffice it to say for now, that the idea of the *author* in the Middle Ages was in many ways extremely different from the concept as we understand it in today’s society, and this makes a distinct difference in how we treat the question of authorship with regards to written materials from the Middle Ages.
In this chapter, I have endeavored to cover an entire field of study in very brief detail, as a building block to the discussion I will create in later chapters. This information is meant to provide a bit of a foundation toward that discussion, and it in no way covers all elements of the discipline of manuscript analysis. By quickly glossing over some of the myriad ways in which manuscript analysis may be utilized, and providing a few of the elements necessary for the production of medieval books, as well as the creation of the *ordinatio* or planning phase of book making, I hope to have made it clear that there are a number of potential limitations which may possibly be applied when a linguistic analysis is performed on a medieval text. This should also become more evident in later discussions when the added components of cultural processes and changes over time to those processes is taken into consideration.

I also touched upon the choices made in medieval bookmaking with regards to which language was used and why glosses were added either during the initial production process or in later generations. It is also very important to keep in mind that each and every handwritten medieval text will likely endure human errors in production, either in the original text or in later commentaries, colophons, and emendations made to that text, and even in the glosses added to a text. These errors, as well as the various text types and their language structures in which they occur, are important linguistically when one is performing any kind of analysis as to the authorship, provenance, dating, and use of a medieval text. Grammatical structure and vocabulary choices in writing are also extremely important to the analysis of any text, as was also discussed in detail in my chapter on forensic linguistic practices.
CHAPTER 4

LITERARY REVIEW: AUTHORSHIP

4.1. Introduction.

In this chapter, I will delve into the enormously gray area that is the concept of authorship, from both a modern and a medieval perspective. I separate these two time periods because they are extremely different from one another in terms of perception regarding this subject. It is well documented that people who lived in the Middle Ages absolutely did not view the notion of ownership of the materials contained within a written work in the same manner that our society views it today; and there is a great deal of scholarship and analysis on that point already available within the academic world, from many different viewpoints. I will endeavor to touch on several of those viewpoints as they pertain to my discussion regarding the interdisciplinarity of manuscript analysis and linguistics.

First, I will provide some background research on the “definition”, per se, of the term authorship, from both a modern-day perspective and a medieval one. Then I will begin to focus on fleshing out the idea of medieval authorship from a number of different angles. I will discuss a few markers of medieval authorship, and the use of medieval “anonymity” in medieval writing. From there, I will discuss how the oral traditions preceding the age of literacy helped shape and inform many of the written works of the Middle Ages, and make some comparison remarks about a few of the various roles a medieval writer can be seen to assume. I will also address the studies done on how modern-day editors treat with medieval texts in converting them into more current reading mediums. Finally, a quick discussion about how this chapter’s material coincides
with research into manuscript analysis and forensic linguistics will tie the information here together with the rest of this project.

4.2. Definition of modern authorship.

Much of what needs to be covered in this discussion with regards to the idea of modern authorship can be neatly summarized within the philosophical writings of Barthes (1988: 142-3; Barthes 1968), who describes the *modern author* as “a product of our society insofar as, emerging from the Middle Ages with English empiricism, French rationalism and the personal faith of the Reformation, it discovered the prestige of the individual, of, …, the ‘human person’.” In this sense, the author and his/her writing are related to one another as a parent is to their child; the parent-author is the progenitor, the creator of the writing, and “he exists before it, thinks, suffers, lives for it” (Barthes 1988: 145; Barthes 1968). This modern attitude toward the vision of what an author is, is also beginning to war with modern society’s notion of the modern *scriptor*:

…the modern scriptor is born simultaneously with the text, is in no way equipped with a being preceding or exceeding the writing, is not the subject with the book as predicate; there is no other time than that of the enunciation and every text is eternally written *here and now*. [Barthes 1988: 145; Barthes 1968]

Here, the scriptor and the writing are much more loosely related to one another, and the scriptor is *not* seen as being a creator of the work; indeed, the scriptor does not become a *scriptor* until the text is laid down on the page – the information in the text exists in the universe already, and the scriptor merely relates it, rather than creating it.

Now, it can be argued that the *author*, being that progenitor of a work, is directly responsible for that writing and therefore can take full credit for its creation and what it conveys (although not necessarily for the responses of its readers). The modern *scriptor*,...
on the other hand, is assumed to be created at the same time as the written work, which presupposes that the scriptor takes no responsibility for what is written, as they both come into being at the same time; does this also suggest that the scriptor in this case also cannot take credit for the written work? If this is indeed the situation, the how is authorship ascribed? There are many other concerns that may be raised about this particular line of thought with regards to responsibility and credit due, but those are, perhaps, better left for another venue of discussion.

Another role within the realm of modern authorship is that of the narrator, which – along with a story’s characters – is notated as a ‘paper being’, and we are cautioned that “the (material) author of a narrative is in no way to be confused with the narrator of that narrative. … who speaks (in the narrative) is not who writes (in real life) and who writes is not who is” (Barthes 1988: 111-12; Barthes 1966; Lacan 1977). This role of narrator, incidentally, has not actually changed all that much since the Middle Ages; Barthes (1988: 111, n2; Barthes 1966) acknowledges this in his admission that many narratives throughout history are registered as anonymous works, such as bardic epic tales, oral narratives, recitations, and mythological folktales.

Finally, it is necessary to at least mention the relatively recent creation of the droit d’auteur, or copyright, in conjunction with the modern viewpoint on authorship. Again, Barthes (1988: 160-1; Barthes 1971) addresses this invention, which came into being during the time period of the French Revolution. Since the author had come to be considered the parent figure for his/her work, he/she was also the owner of it; “literary science therefore teaches respect for the manuscript and the author’s declared intentions, while society asserts the legality of the relation of author to work” (Barthes 1988: 160-1;
Barthes 1971). Prior to the French Revolution, there was no legal precedence toward ownership of one’s writing, and what we would today consider to be plagiarism happened quite frequently in one guise or another, as a normal happenstance regarding writing.

4.3. Definition of medieval authorship.

When discussing the definition of authorship, it is easy to misconstrue this term as being a fixed idea in time, when in reality, it has changed significantly over time. This misconception can result in the misattribution of one society’s viewpoint and understanding of the term from one time period onto another, which will result in inaccurate perception of the other society’s structure around the notion of authorship as it stands in that other time period. Minnis (2010) and Dronke (1974) elaborate on this point:

In recent years, in discussions of late-medieval literature, it has become fashionable to employ a number of critical terms which derive their meaning from modern, not medieval, literary theory.¹ … There are many major aspects of medieval texts which cannot be discussed adequately in the terminology and framework of those sources of medieval rhetoric and poetic which have to date enjoyed full scholarly attention. [Minnis 2010, 1; ¹ Dronke 1974, 13]

Rather than maintaining our current viewpoint regarding authorship (i.e., an idea of ‘ownership’ of what is written, and the legal ability to do so), Minnis (1984) and Lionarons (2004) point out that the medieval perspective on authorship did not focus on the innovative uniqueness of a text; instead, the medieval writer sought to create a perception of ‘authority’ over the written material’s subject (Minnis 1984; Lionarons 2004: 68-9).

Continuing along this line, Lionarons (2004: 69) describes the attainment of auctoritas, or the combination of ‘authority’ and ‘authorship’, where a medieval writer
created texts that were essentially true accounts, thus giving the writer’s work the distinction of being ‘authoritative’.

Attributions of authorship might be omitted in manuscripts if they were deemed unnecessary to establish the auctoritas of a work, while in other manuscripts the attribution of works to known auctores, even if such attributions were false by modern standards, served to provide legitimizing credentials to works by unknown of unfamiliar writers. [Lionarons 2004: 69]

An example of auctoritas as an attainable status in authorship attribution can be found in Minnis’s (1984) work regarding the utilization of Latin in Gower’s Confessio Amantis, which was also reanalyzed in 1989 by Pearsall; Minnis asserts that Gower’s use of Latin was a method for showing the scholastic community that he was a serious, committed writer and poet who had the authority and right to be considered reputable (Minnis 1984: 275; Pearsall 1989: 13-4). So, it may be said that, depending on the text type being written and the more specific time period and society within which it was created during the Middle Ages, auctoritas was achieved by the writer when he succeeded in creating a feel of respectability and accuracy in a work, thereby communicating to the learned community that the writer knew what he was talking about and therefore had the authority to write about it as well.

In other forms of medieval writing, however, we see a different approach to the idea of authorship, primarily because of the fluid, still-changing nature of an oral tradition as it is transmitted into a written form. This can be seen in relation to Old Norse sagas, both in verse and in prose forms. Here, there is the distinct possibility that what is written by one ‘author’ in a saga, or part of a saga, may not be what ends up as the final iteration of that work; in this situation, this ‘author’ may have built on previous material
(both oral and/or written) and other ‘authors’ may build on what this ‘author’ wrote – the
structure of words and story may still change (be it imperceptibly or significantly) before
the final copy of that manuscript for that saga is completed (O’Donoghue 2005: 13-4).
One method of detecting if this is the case for a saga is to examine verb tenses in the
writing, to determine if the ‘author’ is writing the narrative as if it was historical rather
than fictional (O’Donoghue 2005: 51).

A modern concept regarding authorship that is frequently overlaid by literary
critics on the framework that is medieval writing is the concept of intertextuality. A
medieval author ordinarily did not create a unique piece of writing when he or she wrote;
usually the manuscript text was being created according to an already-existing written
manuscript, and the author was merely trying to be as faithful as he or she could to the
original while working on the new version or copy (Mäkinen 2004: 152). “Medieval
intertextuality is insensitive to authorship, but not unconsciously so: the dialogue between
texts (textual or structural parallels) is intended, and even demanded of the new
reiterations of the existing texts and their traditions” (Bergner 1995: 48; Mäkinen 2004:
152). This does not, however, mean that the medieval author did not make changes;
indeed, the original material may have been altered significantly between its first
authorial writing and its final one, by numerous ‘authors’ throughout the Middle Ages.
This can happen when a medieval author created a translation of a text, or a corrected
copy, or a commentary on a piece, or a compilation of previously separate works (cf.
Plett 1991; Mäkinen 2004: 152). Medieval manuscripts have, in just about any iteration,
been altered via some form of editing practices (Mäkinen 2004: 152).
Finally, it is well worth noting that our modern conception of plagiarism (as well as the negative connotations associated with it) is definitely not what the societies of the Middle Ages registered it to be. Scattergood (1999), while analyzing Riddle 47, illustrates this point:

The words of others become one’s own by digestion and rumination so that they can inform and enrich one’s own writing. If reading remains inert and unchanged it is not one’s own. If it remains unaltered in the memory it still belongs to someone else, and to use it unaltered is what in the Middle Ages would have constituted their version of plagiarism, that is to say, literary theft. Either way, as a destroyer of books or as a misuser of them, the theiving worm represents a serious threat to literary culture. [Scattergood 1999: 127]

Medieval readers, then, are responsible for taking in the information on the written page, processing it, and then reusing it in their own writing; it is expected to change in the retelling, becoming essentially one’s own reflection on the material learned. While this is similar to how scholastic researchers today will carefully quote and paraphrase other scholars in their own written works, it is not the same as the medieval method of writing. Today, we would carefully document where the quote or paraphrased information came from and give full credit for it to the legal owner of the work – the original author.

During the Middle Ages, on the other hand, it was not necessary at all to attribute the ‘borrowed’ information to anyone, much less the original author. This lack of attribution to give credit for a copied written text (either a piece or a whole) to an author can severely complicate an authorship study on that text, unless there are other records which may help corroborate the authorship claim.

Circling back to the reminder that the notion of authorship has changed over time, this is also true within the Middle Ages itself. As more and more oral traditional material
was transmitted and fixed into written manuscript form, and more diverse groups of the medieval population increasingly learned to read and write, that concept of plagiarism with regards to ascribing proper ownership to written work began to grow and change into something resembling what we understand as that term today. This appears to have really begun to take hold by at least the fourteenth century, when we see the commercial scribal habit of taking a personal copy of an author’s exemplar for himself, perhaps in payment for his scribal services in making copies for the author, or in order to sell copies of it for his own profit. While this practice was seen as a relatively acceptable act per se, it did begin to elicit claims of ‘piracy’ during that time period (such as those complaints lodged by such authors as Boccaccio, Pecock, and Deguileville); it also had the added effect of putting a great deal of textual variation of a manuscript into public circulation which may in many cases be unauthorized by the actual author himself (Edwards & Pearsall 1989: 259).

4.4. Markers of authorship.

Here we come to discuss the question, what are the different markers by which we may discover authorship of a document? Most of this, of course, was discussed in my chapter about the basics of forensic linguistic testing, with the description of stylistics and what that study entails in terms of deciding authorship. So, in this section I will discuss some of the more philosophical and less linguistic aspects of authorship which may serve as markers to the researcher.

Pons-Sanz (2013: 265) corroborates what I have been saying thus far about the difficulty of attributing authorship to most Old English texts; attribution records are extremely rare, stylistics tests can be untrustworthy because of authority issues, and
copyists frequently (and with no compunctions whatsoever, since the notion of plagiarism was not really an issue of importance) took huge chunks of text from original source materials (see Pons-Sanz 2007a: ch. 1; Pons-Sanz 2007b). However, this does not mean that attribution is virtually impossible for Old English manuscripts; it is just much more complicated than attribution might be for more modern-day texts, and there are more factors and limitations which must be taken into consideration. For example, cross-referencing stylistic markers with other source texts which have better attribution records and performing more varied analyses on a text can potentially solidify evidence toward or away from a potential author for a medieval text; or examining both linguistic and more literary similarities to an author’s known style of writing can help strengthen an argument for attribution.

An item which can potentially be a good marker for authorship testing is an addition to a traditional story which matches with a known author’s other writing. As Liestol (1930) points out about Snorri Sturluson’s writing in the Edda:

The splendid speech that Snorri puts into the mouth of Thorgny Lagmann seems to be a pure invention; and the same may be said of many other speeches in Snorri. Undoubtedly speeches and conversations often make more impression on us, and give more substance to the picture we form in our minds of bygone days, than mere descriptions of the events that occurred.” [Liestol 1930: 244; 1 speech from Beckman 1918: 278-286]

In this case, Snorri’s added speeches and how the style in which they were written could be compared with other works of his, to determine if they were, indeed, his own writings, versus the rest of the historical account of events which he would have been potentially relating from an oral account into a written one. This could also be useful if a researcher
wished to simply determine which speeches in a work were in the author’s own words, as compared to being a retelling of a speech made by someone else.

Another item which has merit as a possible authorship marker within a medieval text is the use of specific sigils and characters to denote various things. For example, Voigts (1989) examines the use of such sigils and characters as abbreviations and signifiers in scientific and medical texts for the calculation of measurements. She prefaces her work with the admonishment that “[i]t would be presumptuous to assume that medieval authors, compilatores, and scribes were unaware that the sigils they employed were extra-linguistic, or at least bilingual” (Voigts 1989: 102). The use of symbols to shorten or represent words, phrases, and even whole ideas is not new to writing, and it would not be out of the ordinary for a scribe or author to utilize such symbols for their own methods of writing. Indeed, I believe it is possible to potentially analyze the use by a writer of specific symbols (when they are used frequently enough or in patterns of writing) as markers of authorship.

4.5. Medieval “anonymity” in writing.

We cannot assess the medieval viewpoint on authorship without addressing the very common occurrence in that time period of texts whose ownership of the content of their work was deliberately left anonymous, or without an official author ascribed. In addition to the types of works discussed above by Barthes (1988; 1966) which were given this status of anonymous work, various poetry types were also frequently classified this way, because many poems had been passed down orally for generations before being committed to a written form. Therefore, it would make good common sense that the
person committing those poems to the page might not see themselves (and society would not either) as an *author* of the text.

Quinn (2016: 47) discusses two different groupings of anonymous poets regarding medieval poetry in manuscripts. The first is the group where there *is* a name to be ascribed as the author of the poetry, but for whom there is no longer a surviving record of it as proof. The second group is the one where the poetry never had an author ascribed to it, because it is the written result of those several generations of oral transmission (i.e., many people had a hand in shaping it into what it was when the transmission to written form occurred). Old Norse Eddic poetry, for example, is often anonymous, for good reason: because, according to what we know of the medieval Norse society to date, there was no controlled manner of (or official institution or group handling the rules for) the conversion of traditional poetry from oral to written form, nor were there any set rules for those conversions. This means that those oral traditional stories and poems (both anonymous and author-ascribed) were likely still being performed and changed and updated, even as the first initial manuscripts were coming out into the world to be read (Quinn 2016). Essentially, “possibly from the late twelfth century onwards, … performances based on oral tradition would have continued independently of, or in tandem with, the written transmission of verse” (Quinn 2016: 48). In the case of the traditional stories, the performances of what would be known to medieval Norse society as ‘old stories’ could not be seen as belonging to one author, and therefore the written texts of them would also be considered anonymous.

However, with that being said, if an author *is* ascribed to a medieval text written from an oral tradition for some reason or other, and the oral performance of the story or
poem was still also in play at the time of the manuscript creation, there does not necessarily need to be any conflict about the ascription of an author to it:

[W]e must not regard these indications of an ‘author’ as evidence of any disagreement between the written sagas and the oral narratives upon which they rest. Whether a work is in writing or merely handed down by word of mouth, the traditional and the new personal elements always blend in some degree. But the personal element will generally appear in inverse proportion to the strength and fixity of the tradition. [Liestol 1930: 33]

In these cases, then, while many medieval texts drawn from oral traditions are considered anonymous because, essentially, those stories belong to the society in which they grew and drew strength in their telling and retelling, the traditional texts which are ascribed will certainly show something of the author’s (as storyteller in written form) style in them, but the traditional story will always remain strong and clear.

Anonymity in the attribution of medieval texts did not only reside with oral traditions and their transmission into written form in manuscripts. Within the religious traditions, also, we often see anonymous attributions to written materials. The tenets of the Church often encouraged its members to practice anonymity and discouraged ‘self-advertisement’ in taking credit for their work; also, while in general the religious houses did keep good records of their collections of books and other written materials, once they were dissolved there was always the chance that those meticulous records might disappear or simply not survive for us to trace back (Doyle 1989: 109). So, for many religious texts, the anonymous nature of their authorship could either be due to an aversion to self-aggrandizement, or the records of their authorship simply no longer exist.
4.6. Oral traditions vs. written works.

In the above section, I dealt with oral tradition and its transition to the written word in the Middle Ages, as it pertains to the anonymity of authorship. Now I will elaborate a little more on the transfer of oral tradition to a written format in more generalized terms. The most distinct difference in the delivery of an oral telling versus the delivery of a written one is best summed up by Orton (2005):

The writer writes; but unlike the speaker – and all users of language were, of course, speakers in the oral phase of English history, before literacy arrived – the writer cannot see, or indeed identify with certainty, the reader or readers who will receive his or her utterance. The writer, furthermore, must recognize that he or she will not be present when his or her written utterance is read, unlike a speaker, who will always be able to at least see the people he or she addresses. [Orton 2005: 206]

Basically, in the oral tradition, the skald or bard is right there in front of his intended audience, gauging responses to the story and embellishing or cutting back as needed in order to keep their attention and tell the story in the best way possible for that specific group of people. In the written manuscript, on the other hand, the author is removed from his audience, unable to gauge the audience’s reactions and react to them; the words on the page are fixed and must be enough to engage and keep the audience going.

The Middle Ages were an interesting time, then, for the transmission of information as it transitioned from an entirely oral tradition (in many societies) toward a more literate society where a portion of the population was gleaning their knowledge from various manuscripts, beginning with small groups and then growing over time to more significant numbers of people. As I noted above, there was certainly overlapping oral performance and written material for these traditional utterances, and remnants of
that oral tradition remain in many medieval manuscripts. For example, Orton (2005) studied the colophons of several tenth century Anglo-Saxon manuscripts, analyzing why they occasionally utilize first person pronouns as the colophons refer to the manuscript they are written for, such as “Wulfwi me wrat” – “Wulfwi wrote me” (Orton 2005: 195). Orton determined that some of the colophons were manuscript-deictic (basically attached to the specific manuscript at hand and non-transferable to future copies of that manuscript in terms of ownership of the text there), and some were work-deictic (attached to the work itself – these attributions could be transferred to other copies of the manuscript for future authorship claims). The use of the first-person pronouns in all of these cases to refer to the manuscript and its ownership was, Orton (2005) surmised, employed in order to preserve some of that ancestral oral tradition, a holdover from when the storyteller would be right in front of his audience, speaking directly to them.

Another example of overlap between oral and written materials lies with Chaucer’s works. Erzgräber’s (1985) analysis of Chaucer’s House of Fame involves discussion over the difficulties of creating dual-purpose written material; he describes the fourteenth century as a time when poets such as Chaucer had to consider how to create a manuscript which could be recited by authors or entertainers, and also be accessible for silent reading. This meant that Chaucer (and the other poets of his caliber during that time period) would have needed to be very well-educated indeed, having both the ability to perform his own works for an audience, and being trained in the corpus of textual materials available in the fourteenth century in order to utilize them in producing the written word. Whether the language used in these manuscripts was the Latin language (for religious materials) or the local vernacular language, there is no denying that a great
deal of learning would be involved in order to become an author during the later Middle Ages.

Regarding which language to use in writing a manuscript during the Middle Ages, that choice was apparently predicated by the use to which the manuscript would be put, as can be seen with Caedmon’s *Hymn*. O’Brien O’Keefe (1994: 249) indicates that “[t]he differing level and nature of extralinguistic cues in Latin and Old English imply that Caedmon’s *Hymn* was read with different expectations, conventions, and techniques than those for the Latin verses with which it traveled.” She goes on to explain that Caedmon’s *Hymn*, an oral poem that was transferred into a written format, the process was not a one-time transfer of mediums; indeed, within the text there was “a considerable admixture of oral processes” (O’Brien O’Keefe 1994: 250). This suggests that, although this is only one example of such a process, there was room for a great deal of fluidity in the transference of oral material into a written form, and it took some time before a final written form was reached.

4.7. Roles as an “Author” in the Middle Ages.

This chapter thus far has been designed to demonstrate how fluid the term *author* actually was for the medieval writer. Most people believe today that the author has a couple of roles to play in putting together his or her work for an audience, and it would be an easy thing to overlay that idea onto how medieval society viewed the concept of authorship; this would be a grave mistake, however. In reality, the *auctor* or author of the Middle Ages could take on a number of different roles in his or her capacity of writing and creating a manuscript text, including that of translator/glossator, copyist or scribe, compiler or *compilatio*, authority, moralist, narrator, and even spectator. I will use this
section to discuss examples of these different roles in medieval writing as an *auctor*, to
demonstrate how very different they all could be in function.

Minnis (2010: 73-117) dedicates an entire chapter to the discussion of authorial
roles in the ‘literal sense’, applying Aristotle’s four major causes to the notion of
thirteenth century medieval writing and the place of the *auctor* or author within it. He
summarizes the four causes as: the efficient cause, where the *auctor* motivates the
creation and formation of the text; the material cause, which includes the materials the
*auctor* uses while writing the text; the formal cause, which covers everything about the
*auctor’s* literary style and structure as he writes; and the final cause, which addresses
why the *auctor* wrote the text in the first place, his objective and end goal in writing
(Minnis 2010: 5). Minnis (2010) also elaborated on the changes in the general viewpoint
of where the *auctor* received his inspiration to write, which occurred between the twelfth
and thirteenth centuries. The twelfth century, for example, saw a great deal of liturgical
material produced; the commentaries on these texts indicate that the *auctor* was a conduit
for God’s word, and that God was the one who sent down what was to be written. By the
thirteenth century, however, it is evident that the *auctor* was beginning to be seen as a
human being whose own ideas were what was being written into the Scriptures; the Bible
at this time became more of a literary source than an absolutely divine one (Minnis 2010:
5).

4.7.1. *Auctor* as Translator/glossator.

In the early Middle Ages, we see more instances of glosses on manuscript texts, where
the original text required an either full or partial translation directly applied to the text for
reference and reading purposes; this could be seen in several Latin texts where an Old
English gloss was applied to it, such as with the *Lindisfarne Gospels* or the *MacRegol Gospels* (also known as the *Rushworth Gospels*). Later in the Middle Ages, as Middle English replaced Old English in use, we begin to see more outright translations of texts, where the text is not glossed, but a new manuscript is created in either the new vernacular language (such as Middle English) or in Latin. In fact, many of the Middle English manuscripts surviving today were originally written in Latin or French, according to Jones (1990: 21), who states that modern-day editors who work with those Middle English texts also work frequently with the original texts from which they derive, when they are available to be studied. This can cause complications in the preparation of a Middle English text for print publication, for “rarely will source or translation be in the hand of author or translator, but will be found in a number of manuscripts in both source and target languages” (Jones 1990: 21).

This brings attention to the difficulty inherent in translating a piece of writing, which was apparent in the translations of the later Middle Ages. Hanna (1989) addresses this in his analysis of *compilatio* as it was applied to the *Wife of Bath* as a text. In that discussion, it becomes clear that the original Latin version of the story is different from its later iteration in a Middle English manuscript:

In the first place, English Latin, like the Wife’s Jerome, had been cut free from its authors and their context. It could no longer necessarily rely upon the support of the Latin tradition. Texts might get used in ways which were foreign to their controlled use in clerical culture. … Secondly, such texts are substantially resituated. They enter canons of different confirmations, new contexts which again shift meaning and use.” [Hanna 1989: 10]
As so often occurs in any language translation situation, cultural influences behind linguistic choices and the use of grammar and syntax itself can change the look, presentation, and feel of a text quite easily, effectively changing it from one story to a different one; and if you also add in the fact that the translation is being made sometimes one or several generations after the original text was written, there will be changes based on cultural change over time which will show through in the translation as well. Taking this into consideration, in many ways, the writer of translations in the Middle Ages does indeed qualify as auctor as translator.

4.7.2. Auctor as Copyist.

Depending on the manner in which a medieval copyist was working on a text, they could be construed as an auctor at times. Vinaver (1976) does caution against looking at copyists as being exactly like other auctors in terms of much of what they do, however:

… we naturally assume that a scribe has certain characteristics which distinguish him from any author; otherwise we should not hold him responsible for altering the author’s work. What are those characteristics? The people who copy manuscripts do not belong to any distinct human type. They may have the same sympathies and idiosyncrasies, the same tastes and reactions as the authors themselves. But in one respect they are unquestionably different: the mechanism of their work is quite unlike that of original writing. [Vinaver 1976: 141]

Here, Vinaver refers to the fact that the author maintains his or her focus on the one written page that is being created; the copyist, on the other hand, must shift his attention back and forth between the text he is copying and the new page he is writing out. The mechanism of production, then, is the major difference in what these two roles entail (Vinaver 1976).
With that being said, however, copyists did sometimes have other roles which placed them into the wheelhouse of auctor along with the other roles discussed here. Kwakkel (2012: 57) identifies Parkes (1991a; 1991b), Huot (1987), and Westphal (1993; 1999) as having studied the activities of many copyists of the later Middle Ages. They determined that not only did late medieval copyists perform their main duties of simply copying texts and exemplars, but there were definite instances where some copyists also performed the task of compiler, effectively placing scribes in those instances into the role of auctor as compiler and copyist.

4.7.3. Auctor as Compiler or compilatio.

In his production of the Retraction at the end of the Canterbury Tales, Chaucer sets himself up as a ‘compiler’ of the text, rather than an ‘author’ (Partridge 2012). Partridge stresses that this is not actually a rejection of the authorship role in this case, because of Chaucer’s literary training within French society in the later Middle Ages; in that cultural setting, “the verb ‘compiler’ could be used to refer not only the act of assembling material from existing writings, but also to a writer’s ordering of his own works and directing their disposition on the page” (Partridge 2012: 133). From this perspective then, Chaucer’s self-assignation to the role of compiler was an admission to an what was viewed in French culture as an essential aspect of authorship. Minnis (1984; 1988) and Hanna (1989) also assign this role of compiler (or compilatio) to Chaucer in his work on the Canterbury Tales.

4.7.4. Auctor as Authority.

Much of what needed to be said about the role of auctor as an authority on a text’s subject matter was discussed above in my analysis of what medieval authorship was. A
good example of this attainment of a status of ‘authority’ as it pertains to authorship is that of Robert Henryson’s role in writing his *Morall Fabillis*, where he reconstructs Aesop’s fables in the vernacular Middle Scots. Higgins (2012) works through this analysis, studying how Henryson began his translation of the fables by acknowledging Aesop as an authority of the stories (since Aesop was the originator of them); this attribution measure automatically gives the fables weight and value in the eyes of the fifteenth century reader. Henryson slowly appropriates the material as he progresses with his storytelling and brings the content into the fifteenth century in terms of its relevance to that time period, rather than telling the tales from an older perspective. By doing this, he sets himself up slowly as a more accessible authority figure to his medieval readers, reaching more of an authority status than that of Aesop himself by the time he finishes the writing. In this way, Henryson becomes less of a narrator of Aesop’s tales and more of an *auctor* as ‘authority’ of the fables for his medieval audience (Higgins 2012: 198-231).

4.7.5. *Auctor* as Moralist.

The later Middle Ages was a time of instability and great change, as cities grew larger and the merchant class became more powerful within their various trades; there was a significant shift at this time of power from the previously all-powerful government administrators and aristocracy to the sharing of political controls between those government entities and the merchant class – increased trade changed everything. Booton (2010: 193) addresses one of the odd fashionable results of that change in class structure – the advent of “courtesy or conduct manuals” which circulated among the aristocracy. In these manuals were instructions for noblemen in how to educate themselves on living a
“good Christian life”, how to be an effective statesman, and how to behave in society; noblewomen had their manuals as well, which instructed them on leading “a charitable and obedient life, with proper Christian decorum” (Booton 2010: 193). This would uphold their elevated place in society over that of the merchant class. In her analysis, Booton assigns this role of auctor as moralist to such medieval authors as Jean Gerson and Philippe de Mézières for manuals addressing noblemen on their conduct, and Christine de Pizan and Geoffrey de La Tour Landry for manuals addressed to noblewomen.

4.7.6. Auctor as Narrator.

I return here to Liestol’s (1930) assessment of Snorri Sturluson’s work as a storyteller in his writing of Old Norse historical sagas during the early thirteenth century, such as the Heimskringla. Here, Liestol is careful to emphasize that Snorri was relating historical accounts, not creating or reciting fictional stories in his writing. As such, he could embellish a bit by adding dialogue and speeches to his accounts, but “he could not very well alter, still less play fast and loose with what people knew to be true” (Liestol 1930: 217). In fact, Snorri acknowledged this need for accuracy in the Prologue to the Heimskringla, where he wrote:

‘Engi myndi þat þora, at segja sjálfum honum þau verk hans, er allir þeir, er heyrði, vissi, at hégómi væri ok skrök, ok svá sjálfir hann’. No one would have dared to recite to the man himself exploits which he and all the hearers knew to be false and feigned. [Liestol 1930: 217; Prologue to the Heimskringla – translation is Liestol’s]

Here, Snorri took on the mantle of auctor as narrator in his writing, retelling a story with all of its facts and only a little of the embellishment for added flavor for his intended
audience. The text itself was not necessarily his to own, as it was a well-known set of oral accounts that he endeavored to set down in writing.

4.7.7. Auctor as Spectator.

Snorri’s exploits as an author at the same time extended even further from the active idea of ownership of a text to the role of auctor as spectator. Liestol best describes this phenomenon with regards to Old Norse texts:

Sturlunga arranges the material artistically, so that we can see an incident vividly with the mind’s eye; the author is master of the old saga technique by which we are kept in suspense, because we are unaware of the real significance of events until it becomes known to the actors themselves. The saga writer is a spectator, not an interpreter; the plot of the story develops as it does in real life. [Liestol 1930: 73]

Even in his accounts of historical information, Snorri maintains that storytelling quality of writing, relating the events in the order in which they occurred and letting the historical story itself unfold with the telling; he is just as much a spectator in this sense as his audience.

4.8. Modern editors and treatment of medieval authors’ texts.

Present-day readers of a medieval text must keep in mind an extremely important tenet when they are choosing a printed version of that text to read: that “medieval text” is very likely no longer just that, medieval. Modern-day editors, who are tasked with converting the textual matter of medieval manuscripts into either a printed format or (recently) a digital one, find themselves in a position of great power with regards to how that material is transmitted to the printed or digital page.
According to Scragg (2009: 83-4), “the common practice among editors is to try to establish the ‘author’s’ text, a dangerous undertaking in the medieval period where anonymity is usual.” The danger inherent in this establishment of an author’s text is clear here. For one thing, it is not always clear or obvious that the medieval texts that have survived long enough to be converted into a print or digital form are actually ‘original’ in the first place. “The researcher should not take it for granted that a published author has worked with the original text; the number of identical errors in different books demonstrates that this is not so” (Owen-Crocker & Cesario 2009: 9). Indeed, most manuscripts remaining to us today are copies of originals or even copies of copies – and it is very safe to assume that the people who copied those texts have altered them from the originals in various ways, through the addition of emendations, commentaries, ‘corrections’ to what was written earlier, and even updates to the texts. So, what a modern editor is likely dealing with when they have a manuscript text in front of them for conversion is not an original of the author’s work at all; and even if it does happen to be classified as what we believe is an original, it may still be impossible to verify this as it may be a work with anonymous authorship attribution tied to it.

This conundrum places a new role upon modern editors as they navigate how to accomplish this conversion task.

Since we lack autograph manuscripts or authorially-approved editions, our interpretation of … practically every author from these early periods, is the immediate and direct consequence of the way in which the editor has presented their works. We put our faith in his ability to give us, if not the exact words and thoughts of an author, at least the closest possible approximation of the author’s
original intention in his work; if he is at fault, then so too are we in our evaluation of the text. [Kleinhenz 1976: 19-20]

It is incumbent, therefore, for editors to be aware of their role almost as authors themselves when they take on the work of transmitting these texts into new mediums; the creative process they must go through to accomplish this is in many ways very similar to that of the original author of the text, as they “reconstruct” or “recreate” the original text (Kleinhenz 1976: 21).

It is worth noting, also, that editors come up against an academic barrier of sorts when they work with these texts for the sake of putting them in the hands of scholars for research purposes. Kleinhenz (2009: 20-1) draws attention to the fact that, in today’s academic society, the role of the editor is considered to be below that of the role of the academic critic in terms of importance and power. This is a problem in that – when utilizing a printed or digital text - the academic critic is technically usually at least two times removed from the experience of reading the original text (more times, if the manuscript was a copy of the original); the editor, on the other hand, is only at least one time removed from it. Therefore, since the critic is dependent on the editor for an accurate and faithful transmission of the manuscript, the editor should be considered (and hoped) to have more direct experience with the medieval text, and perhaps have more power over how it is transmitted and, ultimately, interpreted. The editor in this case would need to have considerable knowledge of the time period the manuscript hales from, in order to successfully produce a viable printed or digital copy of that manuscript.
4.9. Relevance to manuscript analysis.

The study of authorship and its relation to medieval culture and thought is very closely related to the analysis of manuscripts, and while both avenues of research yield valuable insights on how medieval society functioned regarding books and learning, utilizing them both together can provide a more complete picture of the subject. Owen-Crocker’s (2009) research on manuscripts shows that many surviving medieval works are probably not original texts. The poetic codex *Exeter Book*, for example, in its current surviving form was written in one hand (i.e., there was one author or scribe who worked on the writing of it), but the book itself contains materials from potentially different locations and time periods (Owen-Crocker 2009: 5). We are discovering that this was a more common practice than was previously thought; it does make sense, though, considering the immense amount of time and effort that had to go in to the production of any book in the Middle Ages – of course these materials would be used and reused as often as possible over the course of potentially several generations. It was not infrequent, indeed, that “additional texts were sometimes copied into spare space in the margins of existing texts” (Owen-Crocker 2009: 5).

Kwakkel (2012: 57) also provides some insight from a similar perspective, noting that scribes (as compilers of works) would carefully plan through their *ordinatio* process for the adaptation of the structure of their compilations of works; the original compilation would constitute a cohesive, well-planned whole, but future copies of that original could then be competently copied in pieces if need be. This means that scribes were in the habit of planning for the eventuality that a text might be pulled apart (the original or a copy of it) so that parts of it could serve a new, separate purpose in the future. The logic inherent
behind these compilations in the Middle Ages was often an internal process to the production center in which it was created, and the structural planning was also often affected not only by the needs of the written word to be properly placed, but also by the supply and demand of physical materials for the project (Gillespie 1989).

Occasionally, …, a compiler explains his intentions as he marshals his material, indicating the rationale of the collection. … Such thematic cohesion is by no means the rule, however, even in well produced books. … The important criterion is not the integrity of the text, but its utility in context.” [Gillespie 1989, p.325-6]

Gillespie (1989: 324) also adds to this the idea that an author’s work with a text was potentially not complete once the text was written into the manuscript, and, conversely, that completed text might also be altered with no further input whatsoever on the author’s part, with no indication that this editorial process occurred. Finally, Gillespie (1989: 324) also stresses that religious manuscripts were frequently subjected to a range of uses for which they were not originally intended; this can further muddy the waters of understanding of medieval context and intent.

So, we may ask: what does all of this have to do with the concept of medieval authorship? The rest of this chapter has demonstrated how fluid the medieval definition of authorship is – so many different types of roles can be attributed to the position, depending upon the context of the situation in which the auctor finds himself or herself. The fact that there are so many facets to manuscript analysis which pertain to how the texts that authors have their hand in, in whatever capacity, shows that in order for us to truly understand both manuscript analysis from a palaeological standpoint and authorship from a literary viewpoint, it is necessary to take all of those physical and philosophical points of detail into consideration as we attempt to determine what the authorship status
is for a particular manuscript. Obviously, every manuscript will be a different and separate case, based on factors of provenance, materials used, scribal records available for the production center, known aspects of the ordinatio of the book, and a whole host of other factors, along with the philosophical analysis on which version of auctor we are looking at for that manuscript, in an attempt to understand literary context as well.

4.10. Conclusion.

We must think of authorship in the Middle Ages in a much different vein than we do of authorship in today’s society. Today, we are focused on whether the words and contexts themselves ‘belong’ in a sense of ownership to an attributed author. Regarding a medieval manuscript, however, we must focus on whether the document was created by the author, or whether the gloss was created by the attested glossator, etc. The actual words in a manuscript could have been copied from another long-lost original manuscript or even an exemplar created from an original, or they could have come from an oral tradition. It is also important, at the same time, to keep in mind which role the author is playing for a particular manuscript when he or she was helping with its creation; this will make a difference in how we view the document itself in our attempt to attribute any kind of authorship claim to it.

Our goal in utilizing forensic linguistics on these manuscripts, then, must be to employ more of a focus on manuscript attribution, rather than on word and context attribution to any particular author. The medieval ‘author’ (or auctor) becomes, in many cases, a transferor of previously recorded knowledge. What we test for, then, will still include the words, syntax, and grammatical elements of the text, in comparison with other potentially known works of the author. In addition to that, we will also need to
incorporate other information about the document itself, which may help corroborate that textual evidence.
CHAPTER 5

CASE STUDY: THE OLD ENGLISH GLOSS TO THE LINDISFARNE GOSPELS

5.1. Introduction.

In order to properly examine the sorts of methods which have been utilized to conduct forensic analyses on medieval manuscripts, it will be necessary to scrutinize several studies performed on one manuscript and then compare and contrast them. The tenth century Old English gloss for the eighth century *Lindisfarne Gospels* (*LiG*) provides an excellent example of an ongoing medieval authorship study, one in which the manuscript is still extant and which has had quite a few research papers written about it over the past several decades.

In this chapter, I will provide a brief discussion of the history of this document, as well as some of the literary and linguistic theories regarding the generation of its Old English gloss. I will then quickly outline the methods by which I will be analyzing various research studies on the gloss. Then, I will discuss the methodologies utilized by several different researchers in their pursuit of better understanding as to the forensic linguistic aspects of the gloss’s creation. During the course of that discussion, I will also elucidate on how these methodologies perhaps could have been improved upon through the use of a more generalized, structured approach to the data collection and analysis phases of the conducted research. The limitations inherent in each study will also be analyzed in the context of manuscript studies and the conceptualization of medieval authorship.
5.2 Literature Review.

5.2.1. History of the *Lindisfarne Gospels*.

The *Lindisfarne Gospels* manuscript (Cotton Nero D.iv) is housed at the British Library (Richards 1994: 209), and it began its existence as a Latin gospel book, covering the Christian gospels of Matthew, Mark, Luke, and John. The actual date of its creation is somewhat debated; for example, Backhouse (1981: 14; 1995: 5) claims it was created at the end of the seventh century, but Brown (2007: 10) contends that the year of creation for this manuscript was probably closer to 715-720. It was created ‘for God and for St Cuthbert’, according to a colophon to the manuscript which was written in 970 AD by “Aldred, the son of Alfred and of a worthy woman’ (alfredi natus alredus uocor: bonae mulieris/i. tilw’, filius eximius loquor)” (Boyd 1975: 2); this Aldred was the provost of Chester-le-Street at the time this colophon was written. Via the colophon, Aldred named in total four people as the creators of the document: Eadfrith (Bishop of Lindisfarne from 698-721) as the writer of the original Latin manuscript; Aethelwold (Bishop of Lindisfarne from 721-740) as the binder of the manuscript; Billfrith the Anchorite as the one who embellished the book with all of its jeweled ornaments; and Aldred himself as the translator who wrote the tenth century interlinear Old English gloss for the manuscript (Backhouse 1981: 12; Brown 1989).

Even with Aldred’s colophon in place to assert the original Latin manuscript’s authorship, as well as the authorship of the later Old English gloss, there is still some ambiguity as to the sense of its wording. Backhouse (1981) puts this into perspective: Some scholars have argued that Eadfrith and Ethelwald did not themselves make the manuscript but merely commissioned someone else to do so. However,
Symeon of Durham, writing in the early twelfth century but with the full weight of local tradition behind him, says quite categorically that Eadfrith wrote the book with his own hand. [Backhouse 1981: 13]

Verey (1989: 149-50) supports this assertion, indicating that “the *Lindisfarne Gospels* … were uncompromising in the transmission of the Italian tradition imported in the archetype of the Italo-Northumbrian text”. He indicates also that the same style is reflected in various other manuscripts produced at Wearmouth/Jarrow during the early eighth century, as well as at Lindisfarne (when Eadfrith was the Bishop there and would have been writing the *LiG* out in tribute to Lindisfarne’s patron St Cuthbert) and at Durham.

According to Verey (1989: 144), “the Gospels were by far the best known and copied book in the early middle ages, and this familiarity was not always wholly conducive to the precise transcription of an exemplar; there was a widespread process of local editing or ‘improving’.” As Bruce-Mitford (1989: 175; Kendrick et al. 1956-60: 89-106, 246-50) discusses, the scriptorium at Lindisfarne was well-known and well-respected indeed in the eighth century, as the *Echternach Gospels* (Paris Bibl. Nat. MS lat. 9389, c. 690) and the *Durham Gospels* (Cathedral Library MS A.II.17, c. late seventh century) were also made there, and it is commonly believed that they were worked on by the same scribe/artist who worked on the *LiG*. The assertion that the same scribe worked on all three of these gospel books does not preclude the notion that they were actually created in the same scriptorium, however. Both Verey (Brown & Verey 1972) and Bruce-Mitford (1989: 185-6) note that even though the *LiG* and the *Durham Gospels* shared the same corrector, this does not necessarily mean that they were both written in the same scriptorium, since it was common practice for both manuscripts and correctors/scribes to
travel between scriptoriums, the former to serve as exemplars for new manuscripts or to be borrowed for other purposes, and the latter to better serve their communities as a whole as their skills as scribes were in high demand during the middle ages.

5.2.2. Provenance of the *Lindisfarne Gospels*.

Brown (1989: 152-3) asks the question: can “any manuscripts be securely placed at Lindisfarne?” She asserts that Aldred’s colophon is probably the most compelling evidence that the *LiG* was indeed produced there, citing the fact that Bishops Eadfrith and Aethelwold were stationed at Lindisfarne during a period of time which matches the dating of the manuscript, and the “stylistic and textual affinities (of the Gospels) all point to Lindisfarne as a likely place of production” (Brown 1989: 152-3); Brown also contends that the tenth century provenance of Chester-le-Street in Durham of the Old English gloss to the Gospels is supported by the features of the manuscript.

So, thus far I have discussed the original Latin manuscript that is the *LiG* and its relatively firm provenance of origin at the Lindisfarne scriptorium. However, as I mentioned previously, provenance of a medieval manuscript can change over time. For example, the *LiG* was created at Lindisfarne, but over the next several centuries its provenance changed several times. The monks of St Cuthbert were forced to leave Lindisfarne behind in 875, while Northumbria was falling to the Danes and Viking raiding, taking various relics (and the gospels) with them. In 882, the St Cuthbert monks settled in at Chester-le-Street and remained there until 995, when more Viking raiding forced them to move again, this time to Ripon, and then to Durham. During their time in Chester-le-Street, Aldred is attested to have added his word-for-word Old English translation gloss to the original Latin of the *LiG*, as well as his famous colophon; Graham
(2009: 163-4) calls the gloss “not so much a fluent vernacular translation of the gospel text as a vocabulary key to the Latin.” Aldred also glossed the *Durham Ritual* (Durham, Cathedral Library MS A.IV.19) during his time at Chester-le-Street, a manuscript which was a “service book containing collects and scriptural readings for the Hours of the Diving Office throughout the liturgical year” (Graham 2009: 165).

After the *LiG* reached Durham, no more mention is made of its whereabouts until the Early Modern period, when speculation about its possibly being seized by Henry VIII’s commissioners occurred in 1537, during the breaking up of the church system in England; the rumor is that the manuscript was taken to London at that time because of its great worth (Backhouse 1981: 87-8). The manuscript changed hands several times over the next few decades, and in the early seventeenth century it was entered into the collection of Sir Robert Cotton, and in 1703 was donated along with the rest of Sir Cotton’s collection to the British Museum branch that was later to become the British Library (Backhouse 1981: 88).

So, in this case, the *LiG* has technically had two separate points of provenance: one for the original Latin manuscript in Lindisfarne, and one for the Old English gloss in Chester-le-Street. It has also had at least seven separate points of provenance over the centuries: Lindisfarne, Chester-le-Street, Durham, possibly London (and the activities it was utilized for there), Sir Robert Cotton’s estate, the British Museum, and the British Library.

**5.2.3. Authorship debates surrounding *Lindisfarne Gospels*.

Now that a bit of history and the provenance points of the *LiG* have been covered, I can address the authorship debates which surround the Old English gloss of this manuscript.

84
One of the extant debates surrounds another manuscript in conjunction with the LiG. Graham (2009: 163) indicates that “of the twenty-seven books of the New Testament, it was only the four gospels that received a complete gloss, and only in two surviving manuscripts, each of which was glossed long after its original production.” These two surviving manuscripts were the Lindisfarne Gospels and the Macregol Gospels/Rushworth Gospels (Oxford, Bodleian Library, MS Auct. D.2.19).

The Latin manuscript of the Macregol Gospels/Rushworth Gospels (hereafter Ru1 and Ru2, as explained below) originated in Ireland in the late eighth/early ninth century. It was then glossed in the second half of the tenth century by two scribes. Farmon glossed all of Matthew, the beginning of Mark, and a little bit of John – his parts of the gloss are referred to as Ru1. Owun glossed the rest of Mark and John, and all of Luke – his glosses are commonly referred to as Ru2. Graham (2009: 165) notes that both scribes are mentioned in the colophon at the end of Matthew.

The debate that surrounds the LiG and Ru1/Ru2 is whether or not Owun copied his Old English gloss in Ru2 from Aldred’s gloss of the LiG. The long-standing belief is that Owun did indeed copy his version in Ru2 from Aldred’s LiG gloss (see as far back as Skeat 1871-1887: IV, vii). However, more recent studies performed by researchers like Kotake (2018) indicate that not only did Owun not copy from Aldred’s gloss, but there is a possibility that both glosses stemmed from an older manuscript of the gospels which probably did not survive to be researched today, which would corroborate Skeat’s (1871-1887: III, ix; IV, vii) assertion that there was one original exemplar for the Anglo-Saxon gospel glosses which was used for them all, but not Skeat’s assertion that Owun used the LiG when writing the Ru2 gloss.
The second main authorship debate that is currently ongoing concerns only the *LiG* itself. The question here is whether or not Aldred was the only scribe who worked on creating the Old English gloss to the *LiG*, or if the significant variation shown in the *LiG* is due to source material influence instead. This particular discussion has been analyzed from many different angles, including that of palaeography and philology (see Ross et al. 1960; Blakeley 1949-1950; Skeat 1871-1887) and syntax (see van Bergen 2008; Brunner 1947-1948). Most studies have determined until recently that any change of scribal assignment in the *LiG* must have happened in that hypothetical older exemplar manuscript and filtered down as it was copied for the *LiG* gloss, not that the change may have happened with the *LiG* itself (Cole 2018: 169). Cole’s (2018) research into the use of linguistic variation in the *LiG* as a tool for the attestation of authorship is a groundbreaking step in the right direction for using forensic linguistic methods to analyze medieval authorship attribution.

5.3. Data Collection Methods.

For the sake of this case study, I will analyze several already-completed research articles regarding the *LiG* Old English gloss and its authorship attribution issues. Depending on the project at hand, each researcher has employed a different set of methods while performing their analysis on this manuscript. I will briefly outline each article’s purpose and methodology and then discuss the benefits and limitations inherent in each study. The goal will be to determine what factors in these studies would be beneficial to use for a more generalized methodology to utilize with other potential future studies, and to refine factors which might not be as ideal for more general use with medieval manuscripts.
5.4. Data Analysis.

5.4.1. Marcelle Cole’s (2017) study.

Cole’s (2017) study on the subject and adjacency effects which provide the basis for the Northern Subject Rule (NSR) addresses “variation between the present-tense markers -ð and -s in Late Old Northumbrian and discusses evidence which indicates that the subject and adjacency effects at the crux of the NSR were already operative in Old Northumbrian with different morphological material” (Cole 2017: 131). Regarding authorship attribution, the results of this study indicate that Aldred did utilize other source materials when he wrote the Old English gloss to the LiG, and that the LiG gloss does not demonstrate a consistent dialect of Northern origin, nor does it show an idiolect (Cole 2017: 136).

The methodology utilized in this study is laid out as follows:

1. **Determine/decide “whether the text of the gloss can be treated as a single, homogenous linguistic piece”** (Cole 2017: 135).

2. **Assess “the distributions of linguistic variants in Lindisfarne as a diagnostic for establishing demarcations in the script”** (Brunner 1947-1948; Cole 2017: 135).

3. **Analyze the data.**
   
   a. Create a corpus of the variant data, taking “every instance of a plural and third-singular present form with an -s or -ð ending … from all four Gospels, including the forms found in the prefaces” (Cole 2017: 136).
b. “Following Brunner’s (1947/8) methodology, the data were then divided into 64 (arbitrarily determined) equal sections to determine the general distribution of -s and -ð across the text” (Cole 2017: 136).

c. Rule out testing which is not viable and provide reasoning for it. In this case, Cole decided that using logistic regression analysis on each of the 64 sections was not going to work (Cole 2017: 136).

d. Determine which testing is viable and use it. Cole (2017: 137) chose to conduct the statistical analysis of the data with Rbrul (Johnson 2009a, 2009b) – this program is derived from R, an open-source software program used for statistical analysis (R Development Core Team 2009).

4. Discuss statistical data.
   a. Grammatical context effects
   b. Subject, type, person, and number effects
   c. Adjacency effects

5. Present and discuss findings.

6. Conclude the study.

Two inherent complications to this study are readily apparent, and are directly tied to some of the issues I discussed in Chapters 3 and 4, regarding the notion of medieval ‘authorship’ and the analysis of medieval manuscripts. First, Cole (2017: 134-5) points out that interlinear glosses to Latin texts are the only written records available for the analysis of the Late Old Northumbrian dialect; this means that the Latin in the original manuscript can potentially affect how the Old English translation comes out grammatically. Cole (2017: 138; Benskin 2011: 170) addresses this possible limitation,
where the verbal morphology of Latin can force the use of certain spellings and grammatical structures in an Old English gloss, in order to produce a more precise translation of the Latin, rather than providing a more grammatically correct Old English translation; the findings therefore have “the potential influence of a Latin priming effect” (Cole 2017: 138).

The second limitation inherent in this study is directly tied to the practice of medieval manuscript production itself. Since the production of these texts is almost always a collective effort requiring the expertise and work of more than one person, it becomes “very difficult to talk about homogeneous dialectal features” (Cole 2017: 135). Cole (2017: 137) addresses this merely by adding the caveat that the data being analyzed for this study cannot verify a homogeneous dialect.

A couple of benefits do come out of this study, however. First, Cole’s results present important linguistic information with regards to subject and adjacency in the earliest history of the English language; this work confirms that the grammatical NSR was in effect much earlier than had previously been assumed, and that it was much more clear-cut as well (Cole 2017: 146, 149). This predicates the second benefit of this study, which directly impacts on my work on this dissertation: that it is indeed possible to utilize forensic linguistic data testing to answer grammatical change questions in the field of comparative historical linguistics. This second benefit does come with a caveat here, though, that Cole’s study did not verify the possibility of using forensic linguistics to determine authorship, as the data chosen for testing did not allow for the proper determination of an individual’s specific idiolect.
5.4.2. Johanna Wood’s (2019) study.

Wood’s (2019) study on the subjunctive mood in the LiG Old English gloss was designed to cover three separate goals: to examine the Latin third person singular imperfect subjunctive in the LiG and how it is glossed in Old English; to then assess how much the gloss translation is affected by the Latin; and to add this information about the subjunctive to the authorship discussion of the LiG gloss (whether Aldred was the only glossator or not) by looking at variances in that use (Wood 2019: 165). Regarding the authorship question, Wood determines that the multiple glosses may be a good factor for more analysis:

[C]onvincing evidence to support multiple authors or different varieties of English would be clustering of any of the significant phenomena. … More interesting is the distribution of the doublets: 7 of the 8 examples cluster in Sections 4 and 5 [Wood 2019: 188]

The methodology used in this study was not expressly given, but the structure of the article was such that I was able to piece together the general idea of what the methodological steps were as I analyzed it. What I was able to glean from the study as to the methodology used is as follows:

1. **Determine which variable(s) will be analyzed for the sake of the study, and why.** In this case, Wood (2019: 165-166) chose to focus on the subjunctive for three reasons: “there is no ambiguity in form between indicative and subjunctive, … which removes the need for speculation whether the subjunctive is intended or not”; “auxiliary be and copula be are frequent and both may be
considered functional”; and the subjunctive “gives a manageable amount of data for microanalysis”.

2. **Determine the size of the corpus samples/groupings to be used in the study.**

Here, Wood (2019: 167) utilized the feature demarcations in the text that Brunner (1947-1948: 52) had already established were there, which do not correspond with the literary demarcations of the gospels themselves, producing 4 commonly-supported divisions in the text, from the beginning of the text to MtGl \((LiG)\) 26.16; MtGl \((LiG)\) 26.17 to MkGl \((LiG)\) 5.40; MkGl \((LiG)\) 5.41 to LkGl \((LiG)\) 2.9; and LkGl \((LiG)\) 2.10 to the end of the text. Wood then also incorporates Cole’s (2016: 181) additional division, splitting the last section into LkGl \((LiG)\) 2.10 to JnGl \((LiG)\) 3.13 and JnGl \((LiG)\) 3.14 to the end of the text, giving her five clear-cut sections of text to analyze.

3. **Determine which variables will not be analyzed and why.** In this case, Wood (2019: 170) decided not to pursue the spelling variations of the past tense *be*, explaining that their variances are more orthographic rather than syntactic in nature and would not be relevant in the current study.

4. **Determine what rules/guidelines will be used in analyzing the corpus data.**

Wood (2019: 170-1) lays out the definitions of what constitutes indicative mood and what constitutes subjunctive mood based on previous researchers’ findings on the subject, to aid in situations where the explanation may not be clear, as to why a certain Latin subjunctive form may not be rendered as such into Old English.
5. **Test the data.** In each of the five delineated sections of text, Wood (2019: 172) counted each instance of the Latin subjunctive verb form *esset* and its Old English gloss, and determined if the mood of the glossed term had a defined explanation. The data was then tabulated for later analysis.

6. **Analyze the data.**

   a. Wood (2019: 173-80) analyzes the variations found in assigned gloss word for each instance of Latin *esset* found and provides explanations for anomalous items.

   b. Wood (2019: 183-6) then analyzes the eight instances where there were multiple glosses provided for the Latin *esset* (such as *wæs ɫ were* or *were ɫ uæs*)

7. **Conclude the study.**

   Obviously, while this study focused on syntactic elements of the *LiG*, it was based generally in a manner of explanation and analysis which relied more on grammatical explanation than statistical variation research. This provides the benefits of word-by-word or phrase-by-phrase analysis of the corpus items chosen for study. It does not, however, provide much statistical data which could help provide a more well-rounded discussion about the “clustering” of variations that Wood indicates would support the idea that there was more than one glossator for the *LiG*.

5.4.3. **Tadashi Kotake’s (2018) study.**

Kotake’s (2018) study on Owun’s potential source materials for the glossing of *Ru2* continues to address the question of whether or not Owun copied his gloss directly from the *LiG*, or if he utilized other source materials for his work on *Ru2*. After comparing
various aspects of the *LiG* and *Ru2* with several other extant Vulgate manuscripts, taking into consideration both the Latin wording as well as the Old English translations, Kotake comes to the conclusion that Owun may have used the *LiG* as one of his source materials for the *Ru2* glossing project, but that he did *not* copy the *LiG* gloss.

The methodology used by Kotake for this study was not expressly given, nor were the testing criteria used in gathering the corpus data for analysis. This is, unfortunately, more common than one might wish when looking through medieval research studies regarding manuscripts; it is my hope that by providing a more stable framework of methodology to medieval researchers which can be utilized in various types of manuscript case studies, we may see more scientific approaches begin to appear in this field of study. What I could determine from this study with regards to methodology is as follows:

1. **Create a corpus containing words/phrases (in Latin and/or in Old English) which differ between the *LiG* and *Ru2*, including corrections, omissions, and word choice variations.** Kotake (2018: 381) indicates that a good place to start in creating that corpus is to use Bibire & Ross’s (1981) data list, which includes various items from *Ru2* which clearly did not come from the *LiG*.

2. **Separate the corpus into two separate groups to be analyzed, based on the type of data present** (Kotake 2018: 381):
   a. Type 1: Corrections made by Owun to the *Ru2* Latin text, which do not match the *LiG* manuscript
   b. Type 2: Translation choices made by Owun in the Old English gloss to *Ru2*, which do not match the *LiG* manuscript. This group was more
complicated to analyze, so this data was further split into three separate example groupings (Kotake 2018: 386):

i. Word insertions, excluding ‘ו’, ‘and’ and prepositions

ii. Word order variations

iii. Wording variations

3. **Compare both groups of corpus data with 30 other Vulgate manuscripts available for analysis, looking for comparison datapoints where the non-`LiG` glossing choices made by Owun may also appear in these other potential source materials.**

4. **Analyze and compile the resulting comparison data into cohesive tables and examples for discussion purposes.**

   a. Type 1 group: Even though Owun worked on parts of Mark and John, and all of Luke, _all_ of the corrections Owun made to the Latin text which also did not show up in the _LiG_ were from Luke; there were other corrections made in Mark and John as well, but these were also reflected in the _LiG_.

   b. Type 2 group: These types of variations between the _LiG_ and _Ru2_ are more spread out, and were found in all three of the gospels that Owun glossed.

5. **Discuss limitations to the study.**

6. **Conclude the study.**

   This study, while not strictly adhering to a stated research structure, still followed certain methods of analysis, especially that of organizing a specific corpus for use in comparison with other extant texts. It is important to note that other texts which were not
of Owun’s or Aldred’s writing were included in that group; as was discussed previously in Chapter 2, this is a vital step in providing good baseline forensic material for authorship studies. The obvious drawback to this when working with medieval manuscripts, of course, is that these comparison texts may not always have survived to be analyzed.

Kotake (2018: 384) addresses the other limitations inherent in this study, which causes the complications in analyzing the Type 2 data sets above, citing Liuzza’s (1994-2000: II, 26) work on the gospel glosses: “accidental omissions, deliberate alterations, and additions for clarity undoubtedly arise spontaneously in the course of the long labour of translation”. Kotake (2018: 384) goes on to stress that interlinear glosses may provide less of a translation issue such as this, or at least less instances of it, but there is still a margin for such choices to be made by the glossator in his work; these translation choices and alterations can also happen in conjunction with a slightly different original Latin version being read, or the glossator’s preference for a certain turn of phrase when making a translation. This can, incidentally, also often be the case with regards to a glossator’s preference for utilizing punctuation and conjunctions in their work as well, depending on their viewing of the context of a text or out of a desire to clarify what they are translating.


Cole’s (2018) study addresses the authorship question surrounding the LiG Old English gloss and whether or not Aldred was the sole glossator who worked on it. To accomplish this, she focuses on the linguistic variations in the LiG gloss. Previous viewpoints (see Ross et al. 1960) assume that any changes in glossatorial hands were perhaps predicated in whatever exemplar or source material was utilized to generate the LiG gloss, rather
than during the production of the gloss itself; this has allowed researchers to consider Aldred to be the only glossator on the *LiG*, despite the significant amounts of linguistic variation which are present in the *LiG*. Cole’s (2018) findings corroborate the researchers such as Brunner (1947-1948), Blakeley (1949-1950), and van Bergen (2008), who have demonstrated that Aldred almost certainly did utilize some sort of source material(s) for his glossing of the *LiG*, materials which are now lost to us.

The structure of this article is very straightforward, and it was very easy to determine the methodology Cole used here:

1. **Examine the non-linguistic evidence.**
   a. Aldred’s colophon, where Aldred takes credit for glossing the *LiG*. Cole (2018: 171-2) briefly discusses how unusual it is for glosses to have colophons attached to them, so Aldred’s motivation (as has been discussed by many other researchers as well) is circumspect here; Brown (2003: 99; Cole 2018: 172) suggests that Aldred may have written the colophon as a sign that this gloss was his *opus dei*, intended to secure him a permanent, significant place in the Cuthbertian community at Chester-le-Street.
   b. Palaeographical evidence. Cole (2018: 172-5) outlines the significant number of variations in palaeographical elements evident in the *LiG* gloss, which indicate that either Aldred played a supervisory role for at least part of the glossing process, as Skeat (1871-1887: III, x) suggests, or Aldred wrote the gloss entirely on his own, but utilized source materials of the vernacular translations of the Latin gospels for his portions of his work on the *LiG*, as Ross et al. (1960: 11, 22) suggests.
2. **Examine the linguistic evidence.** Cole (2018: 175-8) quickly outlines the evidence for linguistic variation demarcations put forth by Brunner (1947-1948), which showcase certain groupings of clear variants in certain parts of the text. Both Brunner (1947-1948: 52) and Blakeley (1949-1950) agree based on their findings that “either the gloss was written by one scribe at different stages of his life or – more probably – different scribes from the same scriptorium wrote the gloss” (Cole 2018: 176; Blakeley 1949-1950: 27). Other linguistics studies are also put forward on this subject. Cole (2018: 177) suggests that there has not been enough study on this as of yet, to make a firm decision one way or the other, regarding the demarcations and linguistic variations; hence the reason for this current study.


4. **Decide whether to assess the text as a whole or assess each individual gospel in the text.** Cole (2018: 181) chose to use Brunner’s (1947-1948) method of separating the text “into 64 (arbitrarily determined) equal sections to determine the general distribution of -s and -ð across the text” for the plural and third-person singular present tense verb form with either an -s or a -ð ending in the text. She then uses Blakeley’s (1949-1950) text divisions into four sections, and adds one more based on her own observations of variates, for a total of five main sections of text for comparison purposes.
5. **Assign coding to type-tokens “according to the subject type with which the verb form occurred”** (Cole 2018: 181-182). There were seven separate subject types associated with the verb tokens taken as the corpus for this study: personal pronouns, indefinite pronouns, demonstrative pronouns, full noun phrases, relative pronouns, null subjects co-occurring with a present indicative form, and null plural imperatives.

6. **Calculate rates of appearance of the -s ending by subject type, for each of the five demarcated sections of text and tabulate the results.**

7. **Perform logistic regression analysis** (collapsing sections 2-5 together to test for data samples that would be big enough to get decent results).

8. **Perform chi-square testing on the two subject types most notably different in behavior and tabulate the results.** In this case, it was personal pronouns and full noun phrases which were tested by Cole (2018: 183-4), demonstrating that Section 1 of the data did not demonstrate a subject-effective variation significant enough to note, but Sections 2 through 5 did show significant variation between the two subject types, with the most variation showing in Sections 2, 4, and 5.

9. **Analyze the results.** Cole (2018: 184-5) discusses the results of the chi-square testing, which indicates that the gospel of Matthew (up to MtGl (LiG) 26.16) is distinct from the other gospels in language use, as is all of the gospel of John.

10. **Interpret the results from a linguistic and a cultural viewpoint.** Cole (2018: 185) states that “[a]brupt changes would suggest a change of scribes, while a gradual transition away from one variant to another would be more in line with a change in the same scribe’s practice (van Bergen 2008: 291”).
11. **Conclude the study.** (Cole (2018: 187-8) closes out this study by stating that the distribution testing performed here support the theory that Aldred utilized source materials for at least part of his glossing of the *LiG*, and in some places he copied the source materials (variant forms and all), rather than relying on his own translation of the Latin text in the *LiG*.

The benefits provided by this study are quite clear: logical statistical testing served to provide excellent linguistic information regarding the distribution of specific subject types in the *LiG* gloss, which can now be used to enrich the current canon of information regarding this authorship debate. Cole (2018: 186) also takes care to include and discuss vital cultural information about the manuscript and the scriptorium and scribal culture surrounding this debate, stressing that the statistical information cannot be used in a vacuum – it must be taken into consideration alongside the non-linguistic evidence present as well. She also makes sure to note that we cannot assume that what we understand linguistically about the language and culture of the Middle Ages at Lindisfarne is how the linguistic community actually worked at that time, based on the analysis of one manuscript.

5.5. **Conclusion.**

This chapter has served to present a taste of how various methods of conducting manuscript analyses on a manuscript such as the *LiG* can attain the goal of the specific study at hand, but in some cases could potentially benefit from further statistical analysis offered up through the use of forensic linguistics. I began the chapter with a brief overview of the history and background of the *LiG* and discussed the authorship debates surrounding the document. I then quickly outlined my plan for analyzing four case
studies on one or the other of the authorship debates and determining what each case
study’s methodology was for conducting the gathering of data, testing of the data (if any),
and analyzing the data. I then examined each case study in depth; each one displayed a
different set of methods and each had its separate limitations and benefits of use. I will be
analyzing these results as part of Chapter 6, where I will be determining in much more
depth how to form the working methodology which drives this dissertation.
CHAPTER 6
DEVELOPING A METHODOLOGY

6.1. Introduction.

This chapter will be a culminating discussion of the materials from Chapters 1 through 5, an effort to bring together the thoughts from them all into one cohesive whole. During that discussion, I intend to add some background information about the formation of forensic linguistic methodology and how it is used in modern cases to form a logical research construction. From there, I will incorporate the difficulties inherent in assessing medieval manuscripts and the authorship culture evident in the Middle Ages (in western Europe, anyway). The end goal will be to present the beginnings of a working methodology which may be utilized in developing good forensic linguistic research into the authorship of medieval manuscripts.

To this end, I will begin with a section on current forensic linguistic methods for running stylistic tests on modern documents. Next, I will re-address some of the many factors of medieval culture and thought that we need to keep in mind when generating a forensic linguistic methodology for use on medieval manuscripts. The following section will be a quick analysis of the Lindisfarne Gospels case study presented in Chapter 5, where I will discuss the research studies I presented there and how the findings determined by the studies’ authors may have benefited more strongly by the use of these modern methodologies, albeit with the necessary modifications to allow for medieval circumstances. I will then put forward the general methodology I have formed which will allow manuscript researchers to apply modern forensic techniques to their case studies successfully.
Every project has its limitations, of course, which must be discussed in an effort to set boundaries for the findings determined. This dissertation is absolutely no exception. I will include the most serious limitations of this work in the conclusions chapter and put forward some decent solutions for their mitigation or at least incorporation into future and continuing research into the subject at hand. The work of bringing an interdisciplinary approach into any particular field is not something that can be done in a single study; it will be an ongoing effort for quite some time to come.

6.2. *Current forensic linguistic methodology theory.*

As with any field of study which involves analysis of data, there is no one ‘perfect’ way of accomplishing a research task. So it is with the current theories surrounding the ‘right’ way to perform a forensic linguistics study. In this section, I will quickly run through a few of the current models of thought employed by linguists who work with authorship cases, as well as a couple of outlines of methodological process which are suggested for modern authorship cases.

6.2.1. *Three models of analysis.*

According to Wachal (1966: 4) and McMenamin (1993: 173; 2002: 117-8), there are three models of thought which can be used for an unbiased assessment of authorship: the population model, the consistency model, and the resemblance model. The population model (Wachal 1966: 4) can be used in situations where there is no one or two likely candidate or candidates to put forward as the author of a piece of writing; in this case, evidence is examined to narrow down the number of possible candidates to a grouping that is less than the entire population. McMenamin (2002) gives an example of an authorship case where:
…a high-level State of California official (and many other people) received a large set of writings exposing, in lurid detail, his personal relationships with office employees. The official knew from document examination evidence that the writer was likely to be one of 22 people who had access to a particular copy machine; thus, the population of possible writers was 22. [McMenamin 2002: 118]

The consistency model (Wachal 1966:4), on the other hand, assesses the opposite extreme situation, where two or more items are examined to see if they were written by the same author or not. McMenamin (2002: 118) puts forward the example of having multiple letters in question of authorship; this model is used to determine whether those letters (no matter what the signatures claim or what the suspects claim) were written by one person or by many.

Finally, the resemblance model of thought (Wachal 1966: 6) is the most commonly used of these three models. It is used in situations where nonlinguistic proof can help the investigator narrow the number of suspects to either one or a very small number of people as authors of a piece of writing; after this group is determined, linguistic testing is used to analyze the questioned writing to either identify or to rule out each suspect in the short-list. Again, McMenamin (2002) provides an apt example of this:

…a large U.S. telephone company received threatening letters containing information that only two people in the company had: the recipient and the suspected writer. For the analysis, both were considered possible authors. [McMenamin 2002: 118]

These models come directly back around to McMenamin’s (2002: 171) questions – discussed briefly in section 2.3 – regarding consistency and resemblance; the population
model is an addition to those queries. This brings us back also to the methodology that McMenamin (2002) outlines regarding how to structure modern authorship analyses.


McMenamin (2002: 119-125) describes a detailed, seven-step process by which modern authorship analysis can be accomplished. This procedure is designed with the idea in mind that the results would need to be properly documented and backed up with data in order to be presented in court case situations. Oddly enough, the detailed manner in which it is designed will lend itself well to my later discourse on developing a way to use it in a medieval manuscript analysis context, despite the modern tone it is casfed in.

The procedure is structured as follows (I have split the fourth step as described by McMenamin (2002) into two separate steps 4 and 5 as they both require detailed work to put in and really should be on their own in the process):

1. **Organize the case materials.** This includes getting a detailed and thorough description from the client which entails the following: the case study request; what the client believes the authorship issue is; the documents to be analyzed; and a contract for services as an expert witness in court. This also includes making the initial working copies of documents, organizing the materials in chronological order for use, labelling things to keep, etc. (McMenamin 2002: 119).

2. **Determine what the authorship problem is, from your perspective.** This may or may not coincide with the authorship problem as the client sees it, so some discussion may be necessary with the client to get both of you on the same page. This also entails a determination of what the research questions will be – both for quantitative and for qualitative analysis - keeping the idea in mind that these can
(and probably will) change as the analysis proceeds (McMenamin 2002: 119-20). This step also involves the determination of who the candidates are for the suspect-author group whose writings need to be examined.

3. **Determine the method(s) for answering the research questions.** This involves examination of the questioned document(s) and comparing them with any known writings, noting any stylistic variations present and determining any markers which may require more detailed analysis later on. Utilizing concordancers (computer software which can scan a document and pull corpus material out as requested through commands entered by the user) can help immensely in forming these lists of variations and comparison materials (McMenamin 2002: 120).

4. **Gather the data.** This includes putting together information about style markers (format of the document(s), punctuation, spelling, syntax, lexical and semantic variations, etc.) and doing qualitative and quantitative testing. Qualitative testing can involve the determination of dialect and/or idiolect features. Quantitative testing may include statistical tests such as frequency distributions of various types of words, discriminant analysis, or examination of letter patterns (McMenamin 2002: 121-3).

5. **Analyze the data results and apply them to each of the author-candidates.**

   Consider the following two polar opposite hypotheses for each potential suspect:

   a. **Is the author-candidate not the author of the document in question?**

      To test this hypothetical question, one must compare the data from the questioned document to the data from the author-candidate’s known writings and look for dissimilarities of both the linguistic and the
statistical kind; if there are enough dissimilar results between the two, then this author-candidate might be excluded from the suspected authors list.

McMenamin (2002) defines significant dissimilarities as follows:

A linguistically significant dissimilarity is a repeated individualizing characteristic present in one set of writings but absent in the other. A statistically significant dissimilarity is a characteristic present in questioned and known writings at respective relative frequencies of occurrence that are significantly different. [McMenamin 2002: 124]

b. Is the author-candidate the author of the document in question?

To test this second hypothesis, one must examine any similar markers between the known writings of the author-candidate and the document in question, and determine if enough similarities exist to potentially attribute authorship to that person. McMenamin (2002) defines significant similarities as follows:

A linguistically significant similarity is a repeated individualizing characteristic shared by the respective sets of writings. A statistically significant similarity is a characteristic shared by the respective sets of writings that is repeated in the questioned and known writings at respective relative frequencies of occurrence that are not significantly different. [McMenamin 2002: 124]

6. Draw conclusions about the findings and results. This step, within the context of modern court case studies, entails making a decision as to the degree with which each suspect probably did or probably did not write the document in question. In the United States judicial system:
Conclusions regarding authorship are stated in terms of identification or exclusion on a five-, seven-, or nine-point continuum, such as the nine-point scale approved and used by the American Society for Testing and Materials (ASTM), or the similar scale … by the Scientific Working Group for Forensic Document Examination [SWGDOC 2000; McMenamin 2002: 124]

7. **Present your opinion of attributed authorship of the document in question.**

For court case reports, this opinion should be presented at the beginning of the report and then backed up throughout the report with the procedures and data developed in steps 1 through 6 above (McMenamin 2002: 125).

8. **Write a detailed report regarding steps 1 through 7 above.** This step will be a descriptive compilation of the above steps, presenting the problem and question(s) at hand regarding the questioned document; describing what materials were analyzed, who the suspects are, and which methods were used to test the materials for data; comparing the findings and determining which suspects can be ruled out and which ones are still in the running for authorship attribution; drawing conclusions from the data; and presenting an informed opinion to the court as to who was most likely to have authored the document (McMenamin 2002: 125).

6.2.3 General notes on current forensic linguistic methods.

There are a couple more general notes to remember with regards to current forensic linguistic methodologies regarding authorship studies, which I will quickly touch on here. The first is to keep in mind that an authorship study should be as impartial as possible. This would seem to be a straightforward, relatively obvious assumption, but in many cases that is just not so. It is all too easy to go into a research project of any kind with a
preconceived notion of what the outcome will be, and when it comes to cultural analysis - and especially in high-profile court cases situations, where information (both accurate and wildly false) seems to come from everywhere – preconceived outcomes are fairly common. Mosteller & Wallace (1964) ran into this very problem when they began their research into the *Federalist* Papers:

> Many have suggested that the meanings of words should provide additional discrimination. Unfortunately for this suggestion, the marker words are chosen in large part because one author uses the marker and the other rarely does. The point is that in choosing these categories, the attempt is to find a method of classification that does discriminate between the authors. [Mosteller & Wallace 1964: 253-4]

The point here is that it is important to make sure to choose testing methods which will help to equitably discern which writer’s style and manner of writing (be it through grammar usage, punctuation usage, spelling usage, etc.) most accurately matches an unknown piece of writing, rather than choosing the tests which are more likely to produce the resultant outcome which has been pre-assumed by the researcher to be true.

The other point I would like to place here for consideration regards statistical methods of testing for corpus data, which is dependent on good sampling of text materials. Morton (1978) stresses the extreme importance of understanding the corpus data one chooses samples from, and remembering that *how* the samples are divided and chosen is just as important as making sure what is sampled is relevant to the overall study. He admonishes:

> The ideal sampling method would be one which made a complete and continuous inspection of the text, an inspection which would isolate any sections of the text which differed by a statistically significant amount from the remainder of the text.
Such a method exists but it by no means relieves the scholar of responsibility for his critical judgements. The method is also subject to the usual restraints of all sampling systems. … For example, if a sequence of 1,000 sentences is being examined for some occurrence, then it must be borne in mind that this sequence could be divided into 100 sequences of 10 sentences, or 20 sequences of 50 sentences, and so it is only to be expected that statistically significant differences will appear by the operation of chance alone. [Morton 1978: 77]

Therefore, it is necessary to fully understand what the research is meant to accomplish and to carefully choose how the corpus material is to be sectioned out for statistical testing, so that the resulting data findings are meaningful in the manner intended for the sake of the study at hand.

6.3. Medieval considerations for methodology strategies.

As was discussed in Chapters 3 and 4 of this dissertation, there are many factors to take into consideration when performing any kind of an analysis on medieval manuscripts, authorship attribution or otherwise. These factors fall into various categories, both physical and cultural in nature, and most of them are alien to the way in which current-day documents are seen and handled. I will recap a very few of those considerations here, for the sake of this chapter’s discussion on new methodologies.

6.3.1. Authorship considerations.

First, the concepts of author and authorship were much more complicated in the Middle Ages, and very different from today’s definitions. Minnis’s (1984) assessment of the status of auctoritas, for example, provided insight into why most works were not signed by an author – it is possible that the author either did not feel it to be necessary to draw attention to the fact that the work even had an author (it was not the primary focus of the work itself) or the author did not have the sense of respectability or reputation in the
community to feel it possible to attach his name to the work. With this background ‘sense’ of authorship in mind, it is not surprising, therefore, to understand that our modern idea of *copyright* (nor, in that case, that of *plagiarism*, at least until the fourteenth century) did not exist in the Middle Ages – nor was it deemed a necessary thing to have as an institution.

Second, I also discussed the modern concept of intertextuality as a common overlay on medieval works when scholars examine those texts. I will reiterate: “Medieval intertextuality is insensitive to authorship, but not unconsciously so: the dialogue between texts (textual or structural parallels) is intended, and even demanded of the new reiterations of the existing texts and their traditions” (Bergner 1995: 48; Mäkinen 2004: 152). This does not, however, mean that the medieval author did not make changes; indeed, the original material may have been altered significantly between its first authorial writing and its final one, by numerous ‘authors’ throughout the Middle Ages. Thus, it is very important to keep in mind when one is studying the potential changes (or lack thereof) made from one manuscript copy to the next, and remember that those changes may or may not be conscious decisions made by the scribe working on a particular manuscript.

Finally, attribution of authorship to medieval texts is exponentially more difficult that it would be for more modern documents, for a myriad of reasons. For example, as I discussed in section 4.4, Pons-Sanz (2013: 265) corroborates what I have been saying thus far about the difficulty of attributing authorship to most Old English texts; attribution records are extremely rare, stylistics tests can be untrustworthy because of authority issues, and copyists frequently took large chunks of text from original source
materials (see Pons-Sanz 2007a: ch. 1; Pons-Sanz 2007b). I will reiterate that this does
not make attribution studies impossible for medieval texts, but it definitely raises the
difficulty level a few notches and involves more detailed analysis and cross-analysis of
any and all related materials available (and surviving) for the manuscript being studied.
Cross-referencing stylistic markers with other source texts which have better attribution
records and performing more varied analyses on a text can potentially solidify evidence
toward or away from a potential author for a medieval text. Examining both linguistic
and more literary similarities to an author’s known style of writing can also help
strengthen an argument for attribution.

6.3.2 Manuscript analysis considerations.

Modern document production, at least as far as printed materials go, is a relatively
uniform process, with certain fixed formatting rules, punctuation usage guidelines, and a
printing press method of physical page and binding creation. In sharp contrast to this
uniformity of book generation, medieval manuscripts are – one and all – unique,
handmade products which followed no hard and fast rules whatsoever when it came to
their creation, beyond the generalized process of the binding of the work itself (and even
then, those bindings could and were pulled apart to reassess and reorder manuscripts as
needed). The process of the manuscript creation must be taken into consideration as part
of any manuscript study for the sake of attribution of authorship, as well as provenance
and source materials (if and when they can be identified for the manuscript in question).

It is also extremely important to glean other important pieces of data from the
manuscript, such as the aspects of ordinatio (the actual layout of a text, which includes
such items as punctuation rules chosen for the text and text segmentation rules) and
communicare (the transmission, or communication, of the text’s message to the intended audience). I will reiterate what I discussed in section 3.4 with regards to Scrugg’s (2009) thoughts on the matter: that the ordinatio and final format of a medieval text may not be established in an order which is readily obvious to our current day mode of thinking; subject matter which had a direct purpose at the time the manuscript was produced may not make sense to us now, in terms of how the materials were ordered or combined with other subject matter. Also, a medieval text as it survives today may be in a different form now than it was when it was constructed originally; this happened quite frequently, when books were taken apart for the sake of being copied or transcribed during the production of the next generation of all or part of that manuscript’s materials.

The main point here is: when we consider how to generate a generalized methodology for handling medieval manuscripts on a scientific analysis basis, we absolutely must include a detailed discussion about the known history of the physical manuscript itself, as well as its known relationships to other contemporary manuscripts and the scriptoriums they may have been housed at for any length of time. Thus, a medieval manuscript authorship analysis will become much more well-rounded and full of information than we might ordinarily see with a more modern document study.

6.4. Analysis of LiG case study strategies.

As I discussed in the Chapter 5 case study on the methodologies used to analyze the Old English gloss for the Lindisfarne Gospels (LiG), the current research being done on this manuscript using any kind of forensic linguistic testing is quite varied with regards to the methodologies used to accomplish this task. In this section, I will quickly summarize
each of the four studies’ methodological processes, for use in the next section’s formation of a new method to follow.

Cole’s (2017) study on subject and adjacency effects on the LiG utilized the following method of analysis:

1. Determine/decide whether the text of the gloss can be treated as a single, homogenous linguistic piece.
2. Assess “the distributions of linguistic variants in Lindisfarne as a diagnostic for establishing demarcations in the script” (Brunner 1947-1948; Cole 2017: 135).
3. Analyze the data:
   a. Create a corpus of variant data.
   b. Divide the corpus into equal sections.
   c. Rule out non-viable tests.
   d. Determine the tests that are viable and use them.
4. Discuss the statistical data.
5. Present and discuss the findings.
6. Conclude the study.

Wood’s (2019) study on the subjunctive in the LiG Old English gloss followed this method of analysis:

1. Determine which variable(s) will be analyzed for the study, and explain why.
2. Determine the size of the corpus samples/groupings to be used.
3. Determine which variable(s) will not be used, and explain why.
4. Determine what rules/guidelines will be used in analyzing the corpus data.
5. Test the data.
6. Analyze the data findings.

7. Conclude the study.

Kotake’s (2018) study revisiting Owun’s glossing methods on the *Rushworth Gospels* (*Ru2*) adhered to the following analysis method:

1. Create a corpus containing words/phrases (in Latin and/or in Old English) which differ between *LiG* and *Ru2*, including corrections, omissions, and word choice variations.

2. Separate the corpus into two separate groups to be analyzed, based on the type of data present.
   a. Type I: Corrections Owun made to the *Ru2* Latin original manuscript that do not match the *LiG* Latin text
   b. Type II: Translation choices Owun made in the Old English gloss to *Ru2* that do not match the *LiG* Old English gloss

3. Compare both groups of corpus data with 30 other Vulgate manuscripts available for analysis, looking for comparison datapoints where the non-*LiG* glossing choices by Owun may also appear in these other potential source materials.

4. Analyze and compile the resulting comparison data into cohesive tables and examples and discuss the findings.

5. Discuss the limitations of the study.

6. Conclude the study.
Cole’s (2018) study on linguistic variation as a tool for determining authorship of the *LiG* gloss was structured with this methodological process:

1. Examine the non-linguistic evidence.
2. Examine the linguistic evidence.
3. Assess previously researched variables for quantitative statistical analysis.
4. Decide whether to assess the text as a whole or assess each individual Gospel separately in the text.
5. Assign coding to type-tokens “according to the subject type with which the verb form occurred” (Cole 2018: 181-2).
6. Calculate the rates of appearance of the -s ending by subject type, for each of the five demarcated sections of text and tabulate the results.
7. Perform logistic regression analysis.
8. Perform chi-square testing on the two subject types most notably different in behavior and tabulate the results.
9. Analyze the results.
10. Interpret the results from both a linguistic and a cultural viewpoint.
11. Conclude the study.

Each of these methodologies had good points, as well as areas where more could have been done using forensic linguistic methods.

Cole (2017) successfully utilized forensic linguistic testing in her study to analyze subject and adjacency effect on the *LiG* gloss, but she needed to spend more time focusing on the non-linguistic evidence we have available on the *LiG*; creating a
synthesis between the two would have strengthened the meaning of the results presented in the data.

Wood (2019) focused on which variables would and would not work for testing purposes in her study on the subjunctive in the LiG gloss, and she set good guidelines and sample groupings for her analysis; but she needed to do more statistical testing (such as a chi-square analysis or a regression analysis), which may have helped support her discussion about the “clustering” of variations suggesting that there was more than one glossator for the LiG gloss.

Kotake (2018) in his study revisiting Owun’s glossing methods in Ru2 took advantage of contemporary manuscripts available for analysis and compared not only the Ru2 and the LiG manuscript variations with one another, but also with thirty other Vulgate manuscripts; this effectively could give baseline data on which one of Ru2 or LiG may have used other source materials for the glosses. However, he did not specify which linguistic tests were used to gather the data he analyzed, and he did not spend any time discussing any of the historical data we have on either Ru2 or LiG, both factors which may have helped to support with more certainty his discussion of the data results.

Finally, Cole (2018) was the most complete methodological structure of the four case studies I examined. Her study on linguistic variation as a tool for determining authorship of the LiG gloss gave detailed information regarding the known history and debates surrounding the manuscript as well as the known linguistic evidence; assessed previous statistical data already gathered on the gloss; properly determined which variables to use and how to separate sections of the text out for testing; demonstrated statistical testing through chi-square and logical regression analysis; and scientifically
followed through on discussion and analysis of the results before presenting a conclusion. However, as complete as this study was from a forensic linguistic perspective, it only gave a cursory nod in the direction of medieval authorship definitions, and a little more discussion in that arena would have benefited the consideration of the data results that much more fully.

6.5. Forming a new methodology.

In developing a generalized methodology for using forensic linguistics techniques with medieval manuscript analysis for authorship studies, a melding of analyses will be required. The usual techniques used for forensic linguistic testing on modern documents will need to be adapted to work with and include various aspects of manuscript analysis, while also taking into consideration the cultural differences inherent in the Middle Ages with respect to the notion of author and authorship. While this may seem like an enormous task to undertake (and it really is), I believe we can begin to make some headway here on getting that path started.

After taking into consideration the above chapters and sections and their detailed discussions regarding forensic linguistics, manuscript analysis, authorship considerations, and methodological theory, I propose the following methodology as a general framework for generating a well-rounded medieval authorship attribution study:

1. **Organize all case materials.**
   a. This includes gathering available extant manuscripts in a format which will be easily prone to generating corpus data (many surviving manuscripts are now available to examine online, which can make this task much easier than it used to be up until recently).
b. Include at least one contemporary manuscript to compare with the manuscript in question, to provide comparative data; if that contemporary manuscript has already been attributed to a known author, all the better.

2. Define the authorship problem as you see it, along with the specific research questions you wish to answer. Be prepared to reformulate these questions as data starts coming in and is analyzed later. Also, decide on an applicable authorship model, as discussed in Section 6.2.1 (population, consistency, or resemblance).

3. For each research question:
   a. Determine which variables will best serve the purpose of the study at hand and explain why you chose those particular variables.
   b. Depending on the research question(s), the authorship model, and the variable(s) to test, determine which statistical testing method will work and explain why; also, briefly consider a few other test methods and why they will not work for the present study.

4. Examine non-linguistic cultural information relating to the manuscript(s) to be studied.
   a. Consider such factors as the scriptorium where the manuscript was created, known provenance(s) of the manuscript and any potential contact the manuscript or scribe(s) may have had with other contemporary manuscripts, source materials, etc.
b. Consider the community in which the manuscript resided, including connections between scriptoriums and other local provenances, where cultural connections may have affected how the manuscript was used and what its intent and audience was.

c. Consider and connecting texts which might indicate notes on the authorship of the manuscript(s) in question (i.e., colophons, scriptorium records and annals, etc.).

d. Consider the manuscript’s place in the authorship community, taking into consideration whether its purpose would lend to there being any records or indication as to who the author could possibly be, to provide information on who the suspected author or authors could be.

5. **Examine the linguistic evidence available for the manuscript in question.**

Include cursory information about the other manuscripts included in the study as well, to provide a comparative baseline for answering the research questions.

6. **Determine how you are going to divide each manuscript up into testable segments** (i.e., 1,000 words each, separated into 10 groups of 100 words, or 20 groups of 50 words, etc.). This segmentation process should be based on careful thought about what you are trying to accomplish as well as providing equality to the texts for accurate analysis and comparison.

7. **For each manuscript in the study:**
   a. For each segment, analyze the writing and make note of all variations found which utilize the variables chosen.
b. For each segment, apply the statistical test(s) chosen for the study and tabulate the results in a meaningful way for later discussion and analysis.

c. Make note of patterns found within the variations found for comparison with the other extant manuscripts you are examining.

8. Compare the data gathered for each segment of each manuscript (questioned and known), drawing inferences from emerging patterns of variations and deviations from the statistical norms. Identify which style markers provide the most information about those variations and determine whether those variations are deliberate or mistakes in the manuscripts. Look for recurring habits in the variations.

9. Present and discuss the findings. As with McMenamin’s (2002: 121-124) analysis of findings:

   a. Identify the style-markers which present variations and are pertinent to the current study’s research questions. This can include syntax markers, punctuation and/or spelling variations, manuscript format differences, variation in lexical or grammatical choices, etc.

   b. Describe all of the qualitative variations noted in the data gathering stage, including any dialectal or idiolectal features noted as well as other cultural or class-based characteristics of the questioned and the known manuscripts, and compare them with any cultural information you have on the questioned manuscript.
c. For any statistical testing performed on the manuscripts, describe the quantitative results of those measures and analyze how the results may be applied to your research question(s).

10. **Interpret the results.**

   a. Assess the findings of the results in a cultural context, comparing known historical information with the variations, comparisons, and statistical data gathered on the questioned manuscript, to determine if the cultural information is sufficient to aid in attributing authorship to anyone in the author-candidate grouping.

   b. For each author-candidate (as per McMenamin’s (2002: 124) methodology) assess the following two questions:

   i. What if this candidate is not the questioned manuscript’s author? (i.e., will the evidence presented allow you to exclude this candidate from the suspect group?)

   ii. What if this candidate is the questioned manuscript’s author? (i.e., is there enough evidence to warrant keeping this candidate in the suspect group?)

11. **State your opinion.** Once you have assessed the author-candidate list in this manner and narrowed down your options, assess the short-list members and make a final determination as to who you believe is the correct author(s) of the manuscript in question, or if an author can or cannot be accurately identified. Explain in detail why. Make sure the reasoning includes careful
analysis of the cultural, physical, and linguistic attributes of the manuscript
are all taken into consideration in this explanation.

6.6. Conclusion.

This chapter was designed to bring some elements from each of the previous chapters
together, to provide background support to the generalized methodology I have presented
in section 6.4. I briefly discussed some of the methodological theory used in modern
forensic linguistic case studies on authorship attribution. Then I included quick
summaries of Chapters 3 and 4, including some of the many issues in manuscript analysis
and authorship considerations which should be taken under advisement when conducting
any manuscript study. I then recapped the methods I noted in Chapter 5’s case studies on
the Lindisfarne Gospels in four separate research papers on that manuscript, analyzing
where the methodologies used could be improved. Finally, I presented my own
comprehensive methodology which may be utilized in generating a complete analysis of
the authorship of a questioned medieval manuscript.
CHAPTER 7
LIMITATIONS AND CONCLUSIONS

After navigating this long journey through three separate fields of research, we come at last to the end of this dissertation. In this chapter, I will summarize some of the many limitations inherent to the beginning of this new interdisciplinary research, discussing some of the pitfalls to such study and some of the solutions which may be applied to them. Then I will quickly summarize the study and briefly discuss my conclusions with regards to this dissertation. I will also direct the reader’s attention to potential future research which can be engaged in to further the work I have started here.

7.1. Limitations.

When a researcher works on a modern-day authorship analysis case, they may think of limitations such as the difficulty of determining who should be considered as a suspect-author of a piece of writing. Essentially, the researcher would analyze the questioned document for unusual variations from what would be considered a normal baseline for the dialect in which the writing was produced, determining what idiolect the document was written in. Then, they would compare each suspect’s own writing with that of the questioned writing and make a determination based on a number of forensic stylistic tests of which suspect(s) might be the most likely author-candidate.

By contrast, someone studying a medieval manuscript for authorship attribution may need to gather information on the document itself in terms of provenance, source materials, history of the document itself, and the social constructs surrounding the manuscript, before they can even come up with a list of potential candidates for authorship. The questioned manuscript must then be analyzed for variations via forensic
testing. The researcher’s job is not yet completed; they must also track down any extant examples of writing for those candidates (if any still exist) and then determine if those examples were copies made from even earlier source materials. If those source materials still exist (and in many – if not most - cases, they may not), the researcher must then determine how much of the variation found in the questioned manuscript actually came from the earlier materials. At that point, there may possibly be a determination made as to which author-candidates might be the author of the questioned text – or there may be a conclusion drawn that there is not enough extant information available to attribute authorship to any of the candidates at all. Throughout all of this analysis and pulling-together of information, the researcher must also remember to take into consideration the complicated social structure that surrounds the notions of authorship and author in medieval writing, and assess the data accordingly in making any attribution of authorship to a medieval manuscript.

The point here is that authorship attribution is complicated enough when applied to modern texts. When it is directed toward medieval manuscripts, however, the difficulty level increases significantly with regards to variables which much be considered in order to produce viable results. The amount of information which can be gathered for any medieval manuscript – or lack thereof in many cases – can also limit the types of forensic testing that can be applied to these documents. It is possible to test many of these texts in one way or another statistically, but it is very important for the researcher to choose the variables and tests carefully, and also to interpret the data results in conjunction with the other anthropological information gathered along with the document itself.
This brings me to what is perhaps the biggest limitation in these medieval manuscript studies – that of predetermined outcomes. It is a well-known fact that statistical data is only as good as the interpretation of its results, and a predetermined outcome is a dangerous factor to interpretation. It is all too easy for a researcher to become convinced of what the outcome of a research project should be, rather than allowing the data and the science to lead the way to a logical answer. Without an impartial analysis on a scientific study, the science becomes conjecture only. Therefore, the importance of keeping as equitable an outlook as possible on a study such as this cannot be too strongly encouraged, from both a scientific and an ethical standpoint.

7.2. Conclusions.

This dissertation began with an explanation of purpose – it has been a journey to discover a better methodology for applying forensic linguistic testing to determine authorship attribution for medieval manuscripts. I began with three full chapters which briefly covered three very full, rich fields of study in their own rights, each of which have many whole volumes of text dedicated to their discussion and practical analysis. The first covered the modern field of forensic linguistics, which strives to apply linguistic study to court cases of various kinds, including authorship attribution of modern-day texts. The second literary chapter covered the field of medieval manuscript analysis, describing the ways in which manuscripts were made, distributed, utilized, and viewed by the public. Finally, the third attended to the study of medieval authorship, analyzing the myriad of ways in which the terms authorship and author were applied to how texts were created as well as how they were received and handled in the Middle Ages.
These literary chapters served to set the stage for the next stage of analysis: that of a case study, where I analyzed the methodological processes utilized in four separate authorship studies regarding the Old English gloss to the *Lindisfarne Gospels*. I looked for anthropological analysis, forensic analysis and statistical testing, and searched for a melding of the two sets of data in their results. This investigation yielded varied results for each study, and each study demonstrated differing levels of use of statistical forensic testing as well as comparison with other sociological factors in discussing their results and findings.

This led to the results of my research for this entire dissertation: that of the production of a new methodology which would take into consideration modern forensic linguistic statistical testing techniques and cultural authorship analysis when conducting studies into medieval authorship attribution. I presented a comprehensive methodology which melded all three disciplines into its steps and discussed why they were all needed in order to generate a comprehensive, well-rounded, scientific analysis for authorship determination on medieval manuscripts. My next steps beyond this dissertation would necessarily entail the application of this new methodology to a case study where I may test it out and see how well it works in providing comprehensive results regarding a medieval authorship dispute.

Ultimately, the evidence presented by the *Lindisfarne Gospels* case study provides an excellent example of why my dissertation is so important to take into consideration for future studies into medieval authorship attribution – the varied approaches to the question of authorship determination indicate that a generalized methodology for handling such studies is needed and vital to the growth of the
interdisciplinary field growing here. I say ‘interdisciplinary’ because the study of medieval manuscripts for authorship purposes really needs to be undertaken with all three fields of forensic linguistics, manuscript analysis, and authorship studies in mind at all times.

Obviously, this also means that the researchers undertaking such studies must be knowledgeable enough in each of these fields to at least be able to apply the theories and concepts presented by each of them. To date, I have not found such a program officially available at any higher education institution, which could provide the kind of training necessary to prepare scholars for this path. To that end, then, I would also endeavor to develop a teaching curriculum – or even a certificate or degree program – which would accommodate such training for scholars who wished to take this path of research and learning.
REFERENCES


Damerou, F.J. 1975. The use of function words frequencies as indicators of style. 


*Durham Gospels*. Durham, Cathedral Library, MS A.II.17.


*Echternach Gospels*. Paris Bibliothèque Nationale, MS lat. 9389.


