3. Stylistics
3.1 The study of style
3.1.1 History
The study of style has its roots in the ancient study of rhetoric. One of the seven liberal arts, rhetoric was formally codified in the fifth century B.C. in Sicily, although it had been practiced for hundreds of years before this time (Lanham, 1991). Classical rhetoric consists of five parts, the third of which is *lexis* (style). The first two parts of rhetoric, *heuresis* (invention) and *taxis* (arrangement), are concerned with finding and organizing the topic of discourse; the last two, *mneme* (memory) and *hypokrisis* (delivery), are concerned with the oral presentation of the topic, since originally, rhetoric was associated with oratory. Rhetoric was eventually applied to written discourse, particularly after the invention of the printing press (Corbett, 1971). Due to the subsequent shift in emphasis of rhetorical study to text, and the association of invention and arrangement with logic, the study of rhetoric now focuses on style (Lanham, 1991).

As I said in section 2.4.10, style involves a writer's linguistic choices. Although style has been much discussed, however, none of the major rhetoricians actually attempted to define it, which illustrates the vagueness of the term since its inception. Rather, discourse on the classification of style (generally into one of three categories: low, middle, or high), and debates about the functional versus the decorative nature of style, Asianism (highly ornamental style) versus Atticism (brief, witty style), written versus spoken style, and economy versus copia of words were popular. Discussions about how to achieve the translation of thoughts into words centred around choice of words, the composition of words into phrases and clauses, and the use of figures of speech (Lanham, 1991; Corbett, 1971).

3.1.2 Models of style
Despite the words spilled on the topic, after 2000 years of study, style remains an imprecise, multiply defined, and not well understood entity. The theory and practice underlying the study of style has been adversely affected by the variety of different emphases given to the conception of style, due to the resulting number of often-contradictory theories[7] (Quirk, 1969). These various theories have arisen in part because of the overlapping interests of different disciplines in the study of style. Of these theories, two are prominent and influential in the current literature.[8] The first of these has an underlying evaluative model of style. Value judgments are made about writing style according to prescriptive stylistic guidelines. The emphasis is on teaching and imposing norms of style, which tends to discourage the development of an individual style. This understanding of style originated in Greek times (Milic, 1967), and no doubt owes its popularity to many pedagogues over the centuries who, faced with the difficult task of imparting knowledge about style to their students, have opted for the simplest, and most measurable, approach. The evaluative model is espoused in style and rhetoric manuals designed to help students learn to apply stylistic norms in their writing. Since such manuals form their basis, the evaluative model also underlies commercial style checkers. The other, contradictory, theory views style as an individuating factor. Style is considered to be the distinctive expression of a writer, a group, a society, or a combination thereof (Cluett, 1990). Although this model has its roots with Greek rhetoricians (Milic, 1967), it was strongly influenced and popularized during the Romantic period due to the Romantic preoccupation with individuality (Milic, 1991). The aim of this model is to identify or understand writers or groups of writers by studying their writing style (Milic, 1967). Work done in *stylistics* (see below) embodies this second model of style.

[7] For a discussion of various conceptions of style from ancient to modern times, see Enkvist (1964) or Milic (1967).

[8] These two models seem to be the most useful ones for developing software to help collaborative writers merge their styles, since they both embody the belief that one can discover linguistic features of style, unlike, for example, the Crocean view that style and content cannot be separated, or the Senecan view that morality is closely linked with writing style (Milic, 1967).

There are two reasons that I have chosen to investigate the latter rather than the former model of style to find a way in which collaborative writers might be helped to accomplish the task of merging their various writing styles. First, imposing a single normative style on writers would be inappropriate for many genres (and sub-genres) of writing, since there are different stylistic norms associated with different types of writing. In addition, texts are written for different purposes, which will also influence the stylistic choices their authors make. The alternative is to use the evaluative model to develop a variety of normative styles designed to reflect the possible types of writing that the collaborators could be doing (e.g., technical writing, academic writing, etc.). However, such a task is overwhelming, given the variety of possible styles even within one such category (compare, for example, academic writing in computer science with that in literary studies). The difficulty with this approach, as Lanham (1974) points out, is that, "successful prose styles vary as widely as the earth" (p. 17). Second, even if the development of a variety of norms turns out to be an easier task than it appears, I am interested in revealing stylistic differences to help collaborative writers produce more stylistically consistent writing, rather than imposing yet another style on their writing.

Therefore, I will focus on work done in *stylistics*, an area of language study in which the dominant understanding of style is as an individuating factor.

3.2 Stylistics

Although how we express ourselves is restricted by many aspects of language, language use does involve a significant amount of choice. Writers (and speakers) can select from among a variety of words and other linguistic features the ones needed to get their message across to their audience. Linguistic choices are made on the basis of both personal characteristics of the writer (e.g., dialect), and contextual constraints (e.g., medium). The study of linguistic choices is called *stylistics*. In its most general usage, stylistics is: "The study of any situationally distinctive use of language, and of the choices made by individuals and social groups in their use of language" (Crystal, 1992, p. 371). The aim of stylistics is to identify stylistically significant features of language, or *style markers*, and the functions they fulfill (Crystal & Davy, 1969). Stylistics has its origins in linguistics, but is used widely in literary studies.

3.2.1 Stylostatistics

One area of stylistic investigation that has received increasing attention in recent years is *stylostatistics*, also referred to as *stylometry* or *statistical stylistics*. Stylostatistics is concerned with describing a writer's style quantitatively. Rather than relying on a scholar's subjective, and often vague, responses, the goal of stylometry is to find information about style from countable features of text (Potter, 1991). The foundation of the statistical theory of stylistics is that style is a probabilistic concept, and therefore stylistic tendencies can be revealed within the variability of actual writing samples (Dolezel, 1969).

Most of the work done in stylostatistics involves research on the statistical structure of literary texts (Crystal, 1991), although biblical studies (e.g., Radday & Shore, 1985), as well as more general linguistic investigations have also been undertaken (e.g., Biber, 1988). Computers are increasingly used to compile and analyze the information in such studies, which has led to the adoption of the appellation *computational stylistics*. In fact, one of the reasons that stylostatistics has gained in popularity over the last few decades is that computers have taken much of the drudgery out of compiling information from text, and have made statistical analysis less time-consuming and less demanding. Since automated word-tagging became possible in the mid-1980's, it has become even easier to conduct certain types of stylometric investigations (Potter, 1991).

3.2.2 History

The collection of quantitative data from literary works has an old tradition. Alexandrian scholars, who standardized the Homeric text, compiled lists of singly-occurring and unique words from the text. Biblical scholars counted words and verses of each book in the Masoretic text of the Bible to determine its middle word and middle letter (Milic, 1967). The modern idea of statistical stylistics has been around for about 150 years. The first known reference to this idea comes from Augustus de Morgan, who was interested in the authenticity of disputed texts ascribed to St. Paul. He wrote in 1851, “I should expect to find that one man writing on two different subjects agrees more nearly with himself than two different men writing on the same subject. Some of these days spurious writings will be detected by this test.” (cited in Bailey, 1969, p. 217). De Morgan suggested that word length might prove to be a distinguishing trait of a writer (Bailey, 1969). The first person to actually test this hypothesis was a geophysicist named Mendenhall (Kenny, 1982). Since that time, researchers in stylometry have studied diverse linguistic features in their search for style markers.[9]

Although there has been over a century of research in statistical stylistics, and the last thirty years have seen a marked increase in such research, along with the technological advances that have facilitated data collection and analysis, there has not been a great deal of progress made in statistical stylistics. There are many problems that are associated with stylometric research that have impeded its development.

3.2.3 Problems in stylostatistical research

The problems affecting stylostatistical occur at all stages of research, from theory to evaluation. I will briefly summarize them to provide a cautionary note, before summarizing research relevant to my investigation.

Theory The first of the problems associated with statistical stylistics is the paucity of literary theory underlying stylistic investigations (Potter, 1991). However, since my use of stylometric techniques is not literary, I will simply refer the interested reader to van Peer (1989) and Potter (1991).

Amount of research The second problem in stylostatistics is the small cohort of researchers, and the resulting lack of research in this area (Potter, 1991). There are a number of biases in literary studies that have discouraged or impeded the pursuit of this type of research by academics. First, Bailey (1969) claims that the explicitness of the assumptions that are necessary to any statistical investigation are not highly valued in literary disciplines, and therefore, tend to be looked upon with suspicion. Second, literary studies encourage the examination of details in text, whereas statistical analysis is better suited to the investigation of broader tendencies (Bailey, 1969). Third, the empirical rather than theoretical emphasis in quantitative studies diverges from modern literary theories (Potter, 1991). Finally, the understanding of statistics is not promoted among students in the humanities. This deficiency has limited both access to such research, and the number of academics who are capable of conducting statistical studies (Bailey, 1969; Potter, 1991).

Knowledge of related work Stylostatistical researchers often have little knowledge of research related to their own work. This problem is illustrated by the fact that investigations are often repeated due to ignorance, rather than to replicate, improve, or verify earlier work (Potter, 1991). The main cause of this ignorance appears to be the lack of communication among academics conducting stylometric research. There are several reasons for this lack of communication. First, stylostatistical studies are conducted by scholars in a variety of disciplines. Since many academic fields do not encourage cross-disciplinary work, some researchers remain unaware of related studies done in other disciplines. Second, work done in certain areas may not be easily accessible to

---

[9] See Milic (1967) Chapter II “The Problem of Style”, Bailey (1969), or Kenny (1982) Chapter 1 “The Statistical Study of Literary Style” for more detailed histories of modern stylostatistics.

academics in others. Most literary critics for example, even those who have taken courses in statistics, lack the necessary statistical background to understand some of the more complex work done by mathematicians and statisticians (Potter, 1991). Finally, the range of styles and variety of languages that have been analyzed is another barrier to accessibility. Without a native-like grasp of the language under investigation, readers might not be able to benefit from the study's insights (Dolezel & Bailey, 1969).

Quality The quality of some of the research is questionable, mainly because the lack of statistical background of most researchers in this area has had an adverse effect on the use of statistics in their studies.[10] First, for the most part, only simple statistics are used, which has limited the scope of the investigations. Second, sampling procedures are at times suspect, which could have resulted in skewed results. Finally, inflated claims are often made by researchers who do not fully understand how to interpret statistics (Smith, 1987).

Replication Deliberate replication or improvement of studies happens rarely (Potter, 1991). Replication is difficult, if not impossible, primarily because of the lack of standards for linguistic analysis. Researchers often invent their own codes, but do not define them explicitly enough for other researchers to use. The use of automatic part-of-speech tagging allows replicable results, but taggers vary in the level of linguistic detail,[11] as well as in how some tags are defined.

Another contributing factor to this problem is that the methodology in these studies is not always described clearly enough for researchers to accurately replicate it. Since there are no explicit standards, and adherence to the ones which have been gradually evolving is inconsistent, assumptions about methodology cannot be made (Potter, 1991).

Evaluation Finally, there has been little subsequent critical evaluation of stylostatistical work (Smith, 1987; Potter, 1991). This deficit can be partially attributed to some of the already-mentioned problems. Since there are few researchers in this area, and many of them lack the necessary breadth of knowledge, the number of people who can fully understand and critique these studies is limited. In addition, the lack of standards for linguistic analysis and the diversity of studies under the stylostatistical umbrella make the comparison of studies complicated even for researchers who have developed the needed expertise. Moreover, being faced with new codes and methods in study after study makes even understanding the research difficult and time-consuming, which discourages critiques of even single studies.

There is some exemplary work in stylostatistics (e.g., the classic Mosteller and Wallace investigation of the disputed Federalist papers, 1964), but few studies meet these standards. Therefore, caution must used in interpreting results of stylometric studies. The above-mentioned problems are serious and many changes must be made in the field before they are alleviated. However, traditional methods of stylistic investigation are also rife with problems, as they rely primarily on academics' subjective responses. Stylometry at least provides an additional, and more concrete, method of investigation, which is particularly important when intuitive judgements are in opposition to one another (Crystal & Davy, 1969). Furthermore, recent retrospective reviews (e.g., Milic, 1991; Potter, 1991), critical analyses of previous work (e.g., Smith, 1987), and articles that demonstrate greater attention to related studies when conducting new research (e.g., Irizarry, 1990; Laan, 1995) suggest a trend towards more critical awareness and self-evaluation among stylostatistical researchers.

---

[10] The present author is not immune from such criticism.

[11] For example, the Penn Treebank has 47 tags whereas the London-Lund Corpus of Spoken English has 197 (Marcus et al., 1993).

3.3 Stylostatistical research

Statistical stylistics has been used in three main areas of investigation: the identification of characteristics of authors' writing styles, the search for stylistic sets of markers associated with different genres; and the description of stylistic features of historical periods, including investigation of diachronic language change. There is a tension that exists among these three areas, each vying for more explanatory power (Cluett, 1976).

Since collaborative writers are writing at the same (historical) time, and are working on the same text,[12] it is only the idiosyncrasies of individual authors that might cause stylistic inconsistencies in a collaboratively written text. Therefore, I am limiting my discussion of statistical stylistics to work that investigates style markers of individuals. The underlying theory of this research is *basic* or *individualist* theory, which holds that a writer's style is individual, and that the mature style is stable over time. The strong view of basic theory does not admit that writers can significantly alter their mature style, nor that they can develop more than one style (Milic, 1991). Despite my focus on individuality, I do not mean to imply that the genre of a text and the era in which it was written do not affect writing style. I believe that each of these factors contributes to the overall style of any piece of writing. This view is perhaps most clearly stated by Cluett (1990): "Despite constraints of time, place, tenor, and genre, the style of any given writer *tends* to remain distinctive in crucial and identifiable respects. Though a writer may be girt round with the bonds of the language as given, yet will the writer's own identity work against those bonds at a number of conscious and unconscious points" (p. 18). The implication is not that writers are constrained to one mode of expression, however. A good writer is able draw on a variety of styles, in order to select the one that is most appropriate to the situation, but this is not to say that these various styles will differ radically from one another. Rather, there are certain characteristics that are preserved, marking the range of styles as unique to that writer (Corbett, 1971). Winter (1969) uses the analogy of identifying a writer's style, much as one identifies a speaker's dialect, by looking at its characteristic isoglosses.

Milic (1991) divides statistical stylistic studies concerned with the style of an author or group of authors into the following categories: *authorship attribution* (authenticity); *chronology of an author's writings*;[13] *character definition; imagery, theme* (content analysis); and *lexis* or *syntax* (aside from those studies that use lexis or syntax to investigate one of the other categories). I will focus on authorship attribution (also called *author fingerprinting*[14]) studies.

3.4 Authorial style

I have focused on authorship attribution because the aim of such research is to identify significant differences between the writing styles of different authors, and to discover style markers of particular authors. Presumably, stylistic inconsistencies are present in a document exactly to the extent that an author identification technique could (at least in principle) determine that different parts of the document have different authors. The difficulty faced by collaborative writers when trying to merge writing styles may be aided by the identification of the writers' stylistic differences, or style markers. Therefore, the designs of author attribution studies provide a starting point for a stylometric study of collaborative writing and style merging.

---

[12] Of course, one text could *deliberately* contain more than one genre or sub-genre, but in such a case, style merging of the different genres would simply not be done.

[13] Chronological studies of a single author seem contradictory to the individualist theory (section 3.3) since they assume that an author's unconscious stylistic features develop over time. However, the two ideas are compatible if it is the case that certain features develop, whereas others are stable (Laan, 1995).

[14] The fingerprint analogy should be used with care. It must be kept in mind that, despite the mathematical approach, there is no definitive authorship test. Even when stylometry advances enough to be highly reliable, the possibility of imitation will not allow the certainty that accompanies fingerprint identification (Kenny, 1982).

I will first discuss why there is a need for authorship attribution, then describe some specific studies.

3.4.1 Authorship attribution

Authorship attribution studies are mainly concerned with settling cases of disputed authorship. There are several reasons why the authorship of a text may be in question.

First, authors have not always had ownership of their writing; rather, these rights fell to the owners or publishers of the text. In ancient Greece, it was common practice to write speeches for other people to use in court cases. The person who paid for the speech to be written gained proprietary rights of the text. In England, up until the eighteenth century, publishers' rights, but not authors', were protected. Therefore, authors' names have in many cases been lost or deliberately left out of texts, since authorship was not at the time of writing deemed important (Morton, 1978).

Second, famous authors' names have been applied to the works of others, sometimes out of deliberate deception, sometimes out of hopefulness that one was in possession of an important text, sometimes because it was accepted practice to "borrow" famous names to ensure being read (e.g., in ancient Greece). Similarly, pieces of writing done in institutions of learning have sometimes been published under the name of the school's founder, despite the fact that the founder may have only supervised the work, or simply permitted it to be carried out in the school. In most such cases, the prime motive for attributing the authorship of the work to a famous person was a desire to profit from the credibility of the association (Morton, 1978).

Third, one of the most common methods of stylistic instruction since Greek times has been to require students to imitate the style of well-known writers. Although modern imitations are unlikely to fool present scholars, ancient exercises written by gifted students in a dead language might mislead them, since today's scholars cannot have the same appreciation as a native speaker for the nuances of that language (Morton, 1978).

Fourth, some writers have deliberately published anonymously, or under one or more pseudonyms. There are a variety of reasons for doing so: to avoid personal criticism due to the views being espoused; to conceal an already-established reputation in a different occupation or writing genre; because it was common practice to publish anonymously at certain times, in certain places, in certain kinds of publications (e.g., in American newspapers of the eighteenth century, Mosteller & Wallace, 1964).

3.4.2 Forensic linguistics

Another area in which stylometric techniques for authorship attribution have been applied is in forensic linguistics. Forensic linguistics is "the use of linguistic techniques to investigate crimes in which language data forms part of the evidence" (Crystal, 1992, p. 142). Most of the work that has been done in forensic linguistics has been concerned with issues that were of little relevance to the present investigation. Indeed, a large part of forensic linguistics, *forensic phonetics*, is concerned mainly with voice identification. Therefore, I will briefly discuss stylometric applications, and problems associated with using stylometry in court, but will not include any of them in the summary of authorship attribution studies (below). The reader who is interested in forensic linguistics is directed to the references mentioned in the following paragraph.

Stylometric techniques have been used to try to distinguish genuine from fabricated police confessions (e.g., Morton, 1978), and to determine the authorship of anonymous letters (e.g., extortion or ransom letters, Perret, 1986), and wills and other legal documents (e.g., Miron, 1990). Although stylometric evidence has been presented in court in a number of countries, this application is controversial, for several reasons. First, there is often not enough writing, particularly commensurable texts, to allow reliable tests (Kenny, 1982). Second, the credibility attributed to a forensic linguistics expert witness by the jurors may far exceed the reliability of the test (Kenny, 1982). Third, it is difficult for an expert witness to convey findings to the court in such a way that non-experts can fully understand (Morton, 1978). Finally, even when the best conditions exist, stylometric tests do not provide conclusive evidence of authorship. When the matter is a literary question, the lack of conclusiveness is unfortunate, but not critical; in a court case, however, the conclusions being drawn will affect people's lives.[15]

3.4.3 Authorship attribution studies

Some authorship attribution studies explore texts that were or are believed to have been written by more than one author. Such studies are most relevant to my investigation because they involve trying to find stylistic differences within one, possibly collaboratively written, document. Therefore, I will briefly summarize several of these studies to provide more information about authorship studies, and to describe some of the potential strategies for finding significant stylistic differences within a single text.

Both Morton (1978) and Smith (1988) studied the play *Pericles*, which is alleged to have been written by two different playwrights. It is generally accepted that Acts III, IV, and V of this play were written by Shakespeare; however, Acts I and II have been variously attributed to Marlowe, Bacon (Morton, 1978), Wilkins, Rowley, Heywood, and Chapman (Smith, 1988). Morton's (1978) study found no significant differences in the preferred position of frequently occurring words, the occurrence of common collocations, or proportionate pairs of words (e.g., ratio of *not* to *no*) between the first and second parts of *Pericles*. However, he did find significant differences between *Pericles*, selected essays of Bacon, and several plays by Marlowe. Morton therefore concluded that *Pericles* was in fact written by one author—Shakespeare. Smith (1988), however, believed that Morton's study was deficient in several respects, and that the tests he used lacked the necessary sensitivity to resolve the identification of authorship from among playwrights of the same period. Smith therefore conducted a study that compared the rates of usage of the first words of speeches (excluding proper names) that occurred at least ten times per thousand in one or more of the plays under investigation. He separately compared both parts of *Pericles* with plays by Shakespeare, Chapman, Jonson, Middleton, Tourneur, Webster, and Wilkins. The rates of occurrence of first words of speeches were often similar among Shakespeare and his contemporaries, but groups of words could be used to distinguish one playwright from another. Whereas the second half of *Pericles* was most similar to Shakespeare's other works, the first (disputed) half was most similar to Wilkins's play.

Morton (1978) also analyzed *Sanditon*, a novel that Jane Austen did not have time to complete before her death. Using a summary of *Sanditon* that Austen had written, "Another Lady" finished the book for publication. A great admirer of Austen's writing, she deliberately imitated Austen's style to try to produce a stylistically consistent novel. Morton was interested in whether stylistic differences could be detected between the two writers, despite the latter's attempt at imitation. He compared characteristic writing habits of Austen's, culling them from *Sense and Sensibility*, *Emma* and the first part of *Sanditon,* to the second part of *Sanditon*. The Other Lady was able to reproduce relatively mechanical habits such as the use of *and* following commas, semicolons, and colons. However, less conscious habits, such as the ratio of *with* to *without*, were not successfully imitated.

---

[15] This issue is becoming more crucial, as stylometric techniques are being used more and more in actual court cases. One well-publicized controversy concerns the validity of Morton and Michaelson's QSUM authorship test. It has been used in England to settle court cases involving alleged confessions, and university complaints involving alleged plagiarism and the writing of defamatory pamphlets (Morgan, 1991), despite the fact that critics of the test claim that it is not valid for authorship attribution (e.g., Hilton & Holmes, 1993). The debate continues to rage.

Irizarry's (1991) computer analysis of *Infortunios de Alonso Ramírez* (IAR) attempted to discover whether the novel was collaboratively written, or has a single author. The novel purports to be the description of an illiterate sailor's life adventures written by an amanuensis, the writer Carlos de Sigüenza y Góngora, but it is believed by some to be a complete work of fiction. Irizarry investigated the plausibility of the collaboration by comparing IAR to three other narrative works of Góngora, all of which were written within three years of IAR. The analysis of type-token ratio (both lemmatized and unlemmatized), hapax legomena (words that occur only once in a text), sentence length, syntactical differences in sentence beginnings, and various expressions and constructions (e.g., superlative adjectives) revealed significant divergences in style between IAR and the other works. Variation in word length was the only test Irizarry tried that was not useful in distinguishing the works. She therefore concluded that the novel was a collaborative effort.

McColly (1987) investigated the style and structure of the Middle English poem, *Cleanness* or *Purity*. Rather than asking who actually wrote this poem, or whether the first and second parts were written by same person, the question he posed is whether the two parts are halves of the same whole, or whether they form two distinct texts. He compiled function-word frequencies, as well as frequencies of certain modifiers (e.g., *many*) and pronouns (e.g., *all*), discarding frequencies of less than one per thousand, for a total of fifty-nine words. He then compared the relative frequencies of these words in the two halves of the poems, as well as in random samples from each half. The difference between the halves was significant, particularly the use of conjunctions and some verb tenses. He concludes that these differences reflect a lack of structural unity in the poem.

3.4.4 Work-internal studies

A slightly different type of study that is also relevant to the present research is the *work-internal study*. In such studies, a single work that is known to have been written by a single author is examined primarily to catalogue his or her writing style. Such a detailed description can then be used as an authority in studies such as those on authorship attribution listed above (Laan, 1995). When stylistic inconsistencies are found within a work, they are examined to discover what kinds of text features are associated with the differences. For example, in Laan's (1995) analysis of metre in Euripedes's *Orestes*, she discovered that significantly high incidences of vowel elision were associated with dramatic intensity, and, similarly, low incidences of elision were associated with non-excited passages. She concludes that stylometrists need to discover text aspects that are reliably associated with certain stylistic features, so that they can be taken into account and not allowed to bias studies which rely on writer's individual styles, such as authorship attribution studies.

Conclusion 3.4.5

These studies, both those that study stylistic variations of a single author and those that seek to attribute authorship of sections within a text, had the common goal of detecting differences within a single text. They therefore suggest techniques that might be appropriate for finding stylistic differences among collaborative writers.