1. Introduction

Computers have changed the way people write. Word processors have become the preferred writing tool in many offices, classrooms and homes. The popularity of the word processor can be attributed to the fact that it makes certain aspects of writing easier. However, it by no means solves all of the difficulties involved in writing, and indeed exacerbates some, as well as creating new ones. Hence, the proliferation of writing software, as designers attempt to support writers in more, and better, ways. To ensure the development of software that people will use, writing practices are being carefully examined to determine what types of tools could most benefit writers. One practice that has received increasing attention in recent years is *collaborative writing*. The current impetus to produce collaborative writing software comes from the desire to support a writing practice that has only recently begun to be regarded as prevalent and legitimate.

Since people tend to bring their individual practices to collaborative work, writers accustomed to using a word processor when writing alone will doubtless want to use one when writing with others. Although writers can (and do) collaborate using software designed for individual writers, collaborative writing tools would allow them to bring not only their preferred writing practices, but also their preferred collaboration practices to the collaborative writing situation. For writers who work collaboratively at a distance, a practice which is increasing among academics and within large corporations (Sharples, 1993), computer support for collaborative writing is crucial. The ultimate goal of computer support for collaborative writing, therefore, is to make it possible for writers to go about their task even when they are unable to physically meet. To achieve this goal, designers of collaborative writing support software need to be concerned not only with facilitating particular practices that occur only when writers collaborate, but also with alleviating particular difficulties that arise due to the collaborative situation.

One such difficulty is the merging of individual writing styles[1] to produce a multi-authored document that has a single voice. Although collaborative writers might individually produce exemplary pieces of writing to contribute to their collaborative document, the pieces might be *stylistically inconsistent*. An example of stylistic inconsistency can be seen in the following sentence, which is from a brochure given to hospital patients before they undergo cardiac catheterization.[2]

> (1) Once the determination for a cardiac catheterization has been made, (2) various tests will need to be performed (3) to properly assess your condition prior to the procedure.[3]

Clause 1 and, to a slightly lesser extent, clause 3 are in medical language, as if in a formal communication between physicians, whereas clause 2 is much less formal, and is expressed in ordinary lay language. The effect of these two styles mixed together in one sentence is a feeling of incongruity—which was presumably not intended by the authors of the brochure. This example, however, is unusual in its brevity. More often, the problem of inconsistency emerges only over longer stretches of text, especially where the granularity of the multiple authorship is at the paragraph, section, or chapter level. Further, although people might notice that something is wrong when

---

[1] *Style* has a number of different meanings, and thus the term *writing style* tends to be used in several ways: the method by which someone carries out a writing task; formatting guidelines (e.g., APA style); literary style. In my discussion of style, I am referring to the author's choice of words and syntactic constructions that gives a piece of writing its particular "feel". A more formal defunition will be given in section 2.4.10.

[2] The parenthesized numbers are mine, to refer to the individual clauses.

[3] Massachusetts General Hospital, Knight Cardiac Catheterization Laboratory (1993). "Your guide to cardiac catheterization." Page 1.

reading a stylistically inconsistent document, they are often unable to articulate exactly where the problems lie. If writers or editors are unable to perceive why a document is inconsistent, they will not know how to go about solving the problem.

The ultimate goal of this research is to build software that will help with the problem of stylistic inconsistency. There are two steps to this task: first, the system must identify stylistic inconsistencies in a document; second, the findings must be presented and changes suggested in a manner that is easily comprehensible to the average user. My immediate goal is to investigate the viability of the first step using a method adapted from *stylostatistics* that might help writers identify, and therefore more easily solve, instances of stylistic inconsistency. To accomplish this goal, I first gather writing samples by carrying out an experimental task in which subjects produce documents written in two parts. By pairing each first part with each of the second parts, I construct a set of "collaborative" documents, with possible stylistic inconsistencies, that are controlled for content. Next, I analyze these samples using relevant stylostatistical tests that can currently be carried out on text that has been processed automatically. I then manually examine the results of these tests to determine what they mean qualitatively. Finally, I conclude with suggestions for future work that might help answer the many questions raised by this exploratory study.

The contributions of this thesis are:

1) The application of stylostatistical techniques to a new problem: the identification of stylistically inconsistent writing.

2) The linguistic interpretation of the results of the stylostatistical tests.

3) The demonstration that the results of stylostatistical tests provide information about writing style that is potentially more useful to writers than the advice from existing style checkers.