6. Future work

In this thesis, I have analyzed the problem of deleterious inconsistencies of style in collaborative writing, and laid out an approach to research on the topic. My work was intended to be exploratory. I have described an experiment aimed at collecting data for the research, and some of the limitations and problems that arose. I have shown that some stylometric tests can match up different parts of one writer's text fairly well. Moreover, some of these tests flag inconsistencies that are likely to occur when different sections of a document are written by different people. For example, inconsistencies in verb tense, percentage of modifiers, level of vocabulary, type of nominal preference, placement of adverbials, and use of coordinate conjunctions were revealed by the stylostatistical tests. Part-of-speech distribution provided the most information about stylistic inconsistencies, but the relationship between some of the inconsistencies detected and how this translates into advice for the user requires considerable linguistic analysis which has yet to be done. The most promising tests at this time are ones that have a straightforward relationship between the inconsistency they reveal and the interpretation of this information. Tests that reveal inconsistencies in verb tense and word-length distribution are two that can be understood by writers without the need for an intermediary.

Many questions have been raised by this investigation, and much research remains to be done before we will know whether adapting stylostatistical techniques to the problem of stylistic inconsistency will result in an effective computer aid for writers. In this chapter, I will discuss interesting questions that arose during the course of this work and suggest some approaches that might begin to answer them.

6.1 Correlation with human perception

I was first interested in finding tests which were computationally tractable, but it is obviously important to find out what people notice when reading different styles, since subjective judgements of the stylistic qualities of a text are not always borne out by empirical tests (Bailey, 1969), and likely the opposite is also true. As anyone who has used an automatic writer's aid can attest, it is annoying and sometimes detrimental to the editing process to be presented with false positives, particularly when the system is slow and one is trying to meet an imminent deadline. False negatives are equally bothersome, especially since copyediting accuracy can actually drop if there are few errors—which may well be the case after using a good writer's aid. How, then, can both types of errors be reduced to optimize a writing tool that uses stylostatistical techniques? Unfortunately, unlike spelling and grammar, there is no agreed upon set of rules for stylistic consistency. Instead, we need to find out what people attend to when they judge a document to be stylistically consistent or inconsistent.

6.1.1 How can we ensure that stylometric tests identify only those documents that people perceive to be stylistically inconsistent?

In other words, how can we avoid false positives? In my investigation, most comparisons that were flagged as inconsistent were also perceived by me as inconsistent. However, some of the comparisons that were flagged showed no obvious inconsistency upon examination. Although there are several possible explanations for such results (see section 5.2.3), they might simply have been false positives. To find out whether there is a high correlation between stylostatistical tests and human judgements, experiments are needed in which people are presented with texts that have been flagged as stylistically inconsistent and asked for their judgements as to whether or not the alleged inconsistencies are present. Since many people are poor at consciously recognizing inconsistent style, experts, such as editors, would be the best subjects for such an investigation.

6.1.2 If people perceive a document as inconsistent, do stylostatistical tests identify it as inconsistent?

The flip side of the previous question is whether stylometric tests flag texts that people consider to be inconsistent. To investigate this question, researchers would first need a set of texts that had been judged to be stylistically inconsistent by a large sample of readers. Then, analysis of what aspects affected the perception of differences needs to be done (either by expert subjects or the researchers). Finally, the stylistic tests would be performed on the texts to find out whether they flagged the same inconsistencies that people perceived, and if they differed, what might account for the discrepancy.

6.1.3 Are previously-defined significance levels the best measures?

Related to both of the above questions is the problem of a cut-off point. Do standard significance levels identify stylistic inconsistencies in texts that people also perceive to be stylistically inconsistent? If not, what criteria can be used to uncover the presence of perceptible inconsistencies? One approach to answering these questions is to give different groups of readers texts that had been found to be stylistically inconsistent at different "significance" levels to find out which level correlates best with readers' perceptions. Since the level may not be the same for all stylostatistical tests, investigations using a variety of tests is necessary.

6.1.4 Do salient stylistic factors need to be more consistent?

Certain genres and text types are strongly associated with certain stylistic factors. Indeed, good parodists exploit the salient text characteristics of the style they are imitating, as do less scrupulous impostors. Because of their saliency, the consistency of such stylistic factors might need to be more consistent than other stylistic factors. One way of investigating such a question would be to compare more than one imitation with the real author's work. After agreement is reached about which imitation is better, stylostatistical analysis could then be done to find out whether salient aspects of style are closer to the original in the best imitation. Such comparisons may provide information about what clues people attend to in their perception of style and style similarity.

6.2 Generalizability

There are a number of variables in my experiment that can be altered. Doing so might change the results considerably, or might have no significant effect on the outcome. To find out whether the results of my investigation are generalizable to other situations, there are several questions that must be answered.

6.2.1 Does the text type affect the reliability of stylostatistical tests?

Writing style is influenced by factors other than personal style, such as genre and purpose. In my experiment, subjects wrote a retelling after viewing a television episode. Comparisons of the samples suggest that they are, in some respects, relatively homogenous, which might be due to this writing task. Bereiter and Scardamalia's (1987) cognitive research on writing suggests that writing style tends to show less variation when people are simply reporting what they know (*knowledge telling*) rather than intentionally reworking knowledge as they write (*knowledge transforming*) (Scardamalia, personal communication). Since the writing task in this research involved straight retelling and limited time was given to complete the task, which allowed little chance for revision, it is likely that most of the writing samples were examples of knowledge telling. Comparisons of texts resulting from tasks that encourage knowledge transforming might reveal a wider variety of, or more distinctive, stylistic differences. To find out whether the stylostatistical tests would be useful in uncovering inconsistencies not seen in the type of samples I used, but that might occur in other types of texts, experiments involving polished writing samples of various genres and purposes need to be done.

6.2.2 To what extent does sample size affect the sensitivity of stylostatistical tests?

The corpora traditionally used in stylostatistical investigations are huge. Moreover, many researchers in this area claim that the stylistic properties of a text can only be adequately determined when a great deal of text is available (Bailey, 1969). Cluett (1976) suggests a 700-word minimum, but points out that the appropriate sample size depends on which aspects of the text are under investigation; if the sentence is being analyzed, for example, the minimum would have to be measured in sentences, rather than words. This observation is consistent with my finding that most tests that had considerably fewer data points (e.g., comparisons of sentence-initial and final tags) were noticeably less reliable than tests that had many. Of course, in any study, larger sample sizes contribute to more reliable results because they are less affected by small errors and outliers, but huge sample sizes are not practical for my desired application. Although the problem of inconsistency usually emerges over longer stretches of text, stylistic inconsistency can occur within a sentence. Therefore, stylostatistical tests would ideally be of use even in the analysis of very short texts. Research on the effects of sample size on the reliability of stylostatistical tests would provide information about the optimal minimal sample size for the aspect of text under investigation, and perhaps suggest ways to adapt the tests when the text is shorter than is ideal.

6.2.3 Is my linguistic analysis valid for other writing situations?

In my analysis of the inconsistencies that had been flagged, I found correlations between the quantitative inconsistencies that were detected and the qualitative inconsistencies that I observed. In some cases, the correlations were relatively straightforward (e.g., tense inconsistency), and the interpretation of the inconsistency is likely to be the same in other writing situations. However, some of the inconsistencies (e.g., rate of pronominalization) did not have a single cause, but rather resulted from one or more of several factors that were not always immediately obvious. Further, the fact that the mapping of inconsistencies is not one-to-one suggests that the patterns of differences associated with certain part-of-speech inconsistencies will not necessarily be the same in other cases. Finally, my investigation could not be exhaustive because of the limitations of the samples. Before people can be given stylistic feedback from stylostatistical tests, aside from the information that an inconsistency in the use of a specific tag has been detected, a great deal of linguistic analysis of the patterns revealed in many writing contexts is needed to determine which are valid.

6.3 Applicability of tests

6.3.1 Which tests are most useful?

My work did not include an exhaustive list of currently possible automatic stylostatistical tests, so further research that investigates other stylostatistical tests is needed. Since some of the tests I used revealed more useful information (e.g., overall tags) than did others (e.g., two- and three-letter words), it is likely that gradations of effectiveness would be revealed when the various tests are compared. Effectiveness would have to be measured on a number of counts, such as: how much information is revealed; how consistent the results are; and how well the results correlate with human perceptions.

6.3.2 Which inconsistencies matter?

As mentioned earlier, stylistic consistency is not always necessary or desirable. Given information about stylistic inconsistencies, users will have to decide whether each inconsistency is detrimental in the context of the writing, and if so, how to fix it. However, there may be inconsistencies that are generally undesirable, and other that are generally innocuous. Studies in which subjects read texts with a variety of stylistic inconsistencies, then are asked for their judgements as to which are noticeably taxing might identify inconsistencies that do not tend to burden readers and others that interfere with the reading process. This information might provide a focus for future software development.

6.4 Benefits to users

6.4.1 Does the information about where the inconsistencies are located in the text improve people's speed or ability to merge different writing styles?

Methods and terms for explaining stylistic problems to users and helping them with improvements are not yet available. However, simply presenting information about the presence of stylistic inconsistencies might help people merge inconsistent writing styles, first, by making style an object of conversation, thus engendering discussion of style, and second, by providing a focus for such discussion (i.e., which inconsistencies have been detected). To find out whether such feedback is useful, experiments which involve a writing task to make two pieces of writing more consistent would provide some answers. One group of subjects would be given only the writing, while another group of subjects would be provided with additional information about stylistic inconsistency. Differences in length of time spent merging, perceived ease of the merging and the quality of the finished product can be investigated.

6.4.2 What kind of stylistic advice is helpful to people?

Even if simply providing information about where inconsistencies are located in a text is useful to people, a good style tool needs to explain each stylistic problem and suggest to users how the problems can be corrected in terms that users can understand. This is no easy undertaking. Sanford and Moxey (1989) point to the lack of clear psychological evidence as to whether stylistic instructions given to users can be followed, in what way they are followed, and what side effects may result. For example, in an experiment involving the use of **Writer's Workbench**, several people complained about not being sure how to apply the computer's stylistic advice (Gingrich, 1983). Once clear correlations between quantitative measures and qualitative aspects of style had been established, this information could be used in studies to investigate whether people can use stylistic advice to improve the consistency of their documents. Comparisons between subjects that are given the advice and those that are only told that certain inconsistencies are present could be made to find out whether the stylistic advice contributes to the quality, pace and perceived ease of style merging. To further compound the difficult task of providing appropriate stylistic advice, developers will also have to take into account that users will not all have the same background knowledge about or understanding of style. Therefore, explanations that might be appropriate for experts, such as academics who have studied style and are writing in their native language, will not be comprehensible to novices, such as high school students who are attempting to write in their second language. This diversity suggests that there is a need for a user model to avoid presenting people with explanations that are too complex or too simplistic for their understanding.

6.5 Technological advances

6.5.1 What improvements in existing technology might provide better results?

Existing tools, such as taggers, although robust, are not as accurate as they could be. Parsers remain too slow, fragile, and inaccurate to be used for unrestricted text. However, some promising improvements are now under investigation. Taggers are currently being developed as front ends to parsers, allowing parsers to work at the tag level, thus increasing their accuracy (Charniak, 1994). The resulting tool would allow a wider investigation of stylistic inconsistencies, since many important stylistic variables which cannot currently be computed from a text, such as distribution of phrase structures, would be subject to automatic analysis. Additional stylistic information would potentially lower the number of problems missed by a stylistic tool.

6.6 Collaborative writing strategies

6.6.1 Do certain methods of collaboration result in more consistent text?

Although not directly related to my main question, one of the questions that initially interested me was whether certain collaborative writing strategies produce more stylistically consistent text. I was interested in whether collaborative writers who pass the document from writer to writer (*relay*), rather than partitioning the document (*independent*), might create fewer stylistic inconsistencies. Due to the poor response rate, I did not have enough subjects to contrast the relay method of collaborative writing with the independent method, which was simulated by comparing two parts of the same writing task written separately by different writers. Assigning collaborative strategies to subjects in a collaborative writing experiment to study the effect of different collaborative writing situations on writing consistency would be an interesting experiment that might provide some useful information on the pros and cons of the various methods.

6.7 Pre-defined styles

6.7.1 Can stylostatistical information be incorporated into automatic style mergers?

Perhaps in the farther future, computer style aids will actually be capable of performing the style merging independently. However, even if the computer itself performs the style merging, there will have to be user input, namely the goals the writers are trying to meet. The problem of style merging would thus become similar to the work involved in style generation, in which goals are specified, the information on the topic is available, audience characteristics are provided, and the software must generate an appropriate text. Perhaps in the nearer future, tools that recognize a "canned" style will provide feedback to writers who are writing in heavily genre-influenced areas.

6.8 Conclusion

Despite the many unresolved issues related to this work, the application of stylostatistical techniques to the problem of automatically detecting stylistic inconsistency appears promising. Given the current state of most writing support software, the development of an automatic style merger seems very far off. However, as editing devices evolve in the direction of Dale and Douglas's (1992) language sensitivity, text base retrieval becomes more automated, and understanding of stylistics progresses, it seems possible that the writing tools of today may eventually be developed into a sophisticated literary assistant, which will further extend the definition of collaborative writing.