Abstract: The number of computerized writing aids has grown dramatically in the past decade. One of the newest and most controversial of these aids is computerized text analysis. This paper reports a study which examined a widely available text analyzer and tested its adaptability for use with a specific population of ESL writers. Though the study found that the text analyzer could be customized to address certain problems ESL writers encounter, important questions remain about the cost and the gains involved in adapting text analysis for the ESL composition classroom.

INTRODUCTION

The number and variety of computerized writing aids available for use in the ESL composition classroom has grown exponentially in the past decade. Word processors, outliners, idea generators, and text analyzers, along with a number of other computerized writing aids, have influenced the way composition is taught and the process student writers follow in ESL classrooms around the world. Though word processing is undoubtedly the most widely used computerized writing aid in composition classrooms, Collins (1989) claims that text analysis is now running a close second. A brief consideration of some of the research examining these two very different approaches to computerized writing will serve as a foundation for examining the possibilities and problems of customizing a widely available disk-based text analyzer, Wang Laboratories' Grammatik IV, for use with specific populations of ESL writers.
RESEARCH ON WORD PROCESSING AND TEXT ANALYSIS

As word processing has been widely available much longer than computerized analysis, the number of research reports examining its use and viability is the ESL composition classroom is much greater than those examining computerized text analyzers. Much of the research on word processing has examined its effect on the revision behaviors and strategies of student writers. While results are somewhat mixed, a number of studies have found that the scope and amount of revision ESL writers perform on a piece of writing increases when using word processing as compared to using pen and paper (for detailed review of the research, see, e.g., Hawisher, 1989, and Phinney, 1989). As Hawisher (1987) has pointed out, at the very least researchers seem agreed that the act of revision itself is facilitated by word processing. When using word processing, writers can easily move or delete portions of text or carry out any number of revisions more easily and readily than with pen and paper.

In her review of the positive findings of research of word processing, Phinney (1989) discovered several effects that seem consistent with and applicable to a more process-oriented approach to writing and writing instruction. Among these are that writers spend more time writing and revising when using word processing than when using more traditional modes of composing. Additionally, attitudes towards writing among students using word processing improve over time more than those of students using pen and paper. Phinney also claims that writers using word processing are more likely to develop a piece of writing in stages, over time and over several drafts, than those using traditional means of composing. There is evidence gathered in a large-scale study at Indiana University of Pennsylvania that when word processing is combined with process-oriented instruction, several significant benefits result (see Etchison, 1987 and Williamson and Pence, 1989 for a detailed discussion).

While word processing seems to have the potential of enhancing process-oriented composition instruction, the place of computerized text analysis in the modern ESL composition classroom is more tenuous and controversial. Computerized text analysis, as its name implies, employs computer software technology (and sometimes highly powerful and sophisticated hardware) to analyze written text then provide the writer suggestions of how that text can be altered and improved. These suggestions are based on the computer's analysis,
which is arrived at through several different processing techniques (Smith, 1989). The most common of these is pattern matching, in which words and phrases in a text are compared with the word and phrase dictionaries of a text analysis program. Another technique provides a statistical summary of a text by counting words, sentences, paragraphs, etc. in the text. The most complex processing technique utilized by text analyzers is parsing. Text analyzers parse text by assigning words to parts of speech then analyzing the syntactic structures of the parse with programmed decoding rules.

The number of studies examining the use of text analysis in ESL composition instruction is small but growing. The earliest studies were conducted at Colorado State University where the English Department had access to Bell Lab’s Writer’s Workbench. In one study, Reid (1986) reported generally positive reactions of the ESL writers using the text analysis program, though no measures of writing quality examined in the study showed significant improvements attributable to Writer’s Workbench.

More recent research has been conducted at the University of Hawaii with IBM’s Critique, the most advanced text analyzer developed to date. In a small case study (Pennington and Brock, 1989, forthcoming) several compositions written by ESL writers using word processing and receiving process-oriented feedback from a tutor were compared with compositions written by ESL writers receiving feedback solely from Critique. In this study writers receiving feedback from the text analysis program wrote a significantly larger number of shorter sentences (less than 10 words) than those writers using word processing and receiving process-oriented feedback. They also wrote shorter drafts, and they made fewer revisions and significantly fewer revisions that changed meaning.

Several researchers (Brock, forthcoming; Collins, 1989; Gerrard, 1989; Pennington and Brock, 1989, forthcoming) have noted problems inherent in text analysis and its use in the ESL composition classroom. These include: the danger of ESL student writers growing dependent on text analysis; the fact that text analyzers often provide incorrect analyses of text; the tendency of ESL students to accept the computer’s analysis and incorporate the program’s suggestions into their texts, even when they are inappropriate; and the product-oriented, surface- and sentence-level focus of extant text analysis program. Another common criticism of computerized text analysis is that most programs are tailored for a particular group of writers (primarily business and technical
writers) and thus are ill-suited for many of the writing tasks ESL student writers are required to complete. In addition, few text analysis programs allow teachers the option of customizing the program to better suit their students’ particular needs.

Though the problems inherent in computerized text analysis are not insignificant, at least one commercially available disk-based text analyzer does give ESL teachers the option of customizing the rules the program employs in providing analyses of text as well as the advice the program gives to writers based on those analyses. It is to this program and the potential of customizing it to fit the needs of particular populations of ESL writers that I now turn.

*Grammarik IV: An Introduction*

By employing programming techniques developed through research in Artificial Intelligence, the authors of *Grammarik IV* have developed a disk-based text analyzer with parsing capabilities previously available only via mainframe-based text analyzers, such as IBM’s *Critique*. *Grammarik IV* runs on IBM-compatible personal computers equipped with a hard disk or a high density disk and at least 512K RAM.

In providing analyses of texts, *Grammarik IV* utilizes two basic programming techniques, in addition to reporting the baseline statistical data and readability scores that most other text analyzers supply. The first, and more rudimentary programming technique, is pattern matching, through which the program compares the words and phrases in its dictionary with the words and phrases in a text. When it locates a correspondence or match in the two, *Grammarik IV* offers programmed advice on a preferred style or usage of the matched words and phrases. For example, *Grammarik IV* will "flag" all passives it find in a text and advise the writer to use active voice where possible. Certain easily confused words or phrases, such as "its/it's" are also marked, the writer is given advice concerning which is appropriate in specific contexts. Other phrases, for example "continue on" and "in my opinion," are marked as being redundant or long-winded, and the writer is advised to revise or delete them.

The second programming technique employed by *Grammarik IV* in analyzing text and providing the writer advice is parsing rules. While pattern matching is quite straightforward and teachers can easily add their own pattern matching rules and accompanying advice to *Grammarik IV*, parsing rules are much more complex and writing parsing rules is much like using a programming language. Parts of speech indicators and several programming rule operators are
used in creating parsing rules. Examples of parsing rules and pattern matching rules will be illustrated below.

CUSTOMIZING Grammatik IV

In order to evaluate how effectively Grammatik IV addresses some of the sentence-level, grammatical problems ESL writers encounter and to measure how readily it can be customized for a specific population of students, a frequency count of the most common errors persisting in the writing of intermediate-level students enrolled in an ESL composition course in the Junior College of the University of East Asia, Macau, was tabulated over one semester. Ten of the most frequently occurring errors were chosen for analysis, and they, along with example sentences, are listed in Table 1.

TABLE 1

1. Verb error following a modal
   "I think that you should chose Guangzhou."

2. Sentence fragments beginning with "such as," "for example," and "for instance"
   "For example, a shirt, a sweater, and some socks."

3. Verb infinitive errors following "to"
   "We didn't know how to went anywhere in Taiwan."

4. Errors in subject-verb agreement
   "She like to go there every Saturday."

5. Errors in passive and perfect verb forms
   "I've never forget that day."
   "The bus was drove by his brother."

6. Confusing "in that time" for "at that time"
   "I wanted him to come to my home in that time."

7. Sentence fragments beginning with "if," "when," and "because"
   "Because the airport is very important to Macau."
8. Superfluous prepositions used before "here" and "there"
   "I went to there yesterday."

9. Using "arrive" without the required preposition
   "When we arrive Hong Kong went to the hotel."

10. Errors with superfluous BE + "very" + Verb
    "I am very enjoy this place."

The example sentences 1 to 5 in Table 1 contain errors that Grammatik IV consistently flags, alerting the writer to the error as well as providing advice on how it can be corrected. Over 75 sentences containing the types of errors listed in sentences 1 to 5 were analyzed with Grammatik IV. Though the program was not faultless in its analysis, it correctly marked and analyzed the majority of these. The types of errors found in sentences 6 to 10 in Table 1 were not marked by Grammatik IV, and they will provide the focus for discussing attempts to customize the program.

Grammatik IV was developed for use by native writers of English, and therefore it is not surprising that the program fails to mark a number of grammatical errors that are peculiar to second language writers of English. Some of the errors listed in Table 1 are common to many different populations of second language learners of English. Others are more common among Cantonese learners of English, the population from which the examples were drawn. The potential of Grammatik IV is that ESL teachers can customize the program to address some of the specific problems their students encounter in writing. Two questions, however, need to precede any attempt at doing so. First, how easily and efficiently can the program be altered and customized? Second, when the program has been customized for a specific group of ESL students, what benefits, if any, do they gain in accessing it?

In answering the question concerning the ease and efficiency of customizing Grammatik IV to address specific student problems, as attempt was made to develop rules that would mark the errors found in sentences 6 through 10 illustrated in Table 1. There are four components comprising rules written for Grammatik IV: the rule pattern, the advice accompanying the rule, the rule class, and the rule dictionary to which the rule is assigned. This report will focus on the first two of these. The first three error
types—confusing the phrase "in that time" for "at that time," sentence fragments beginning with "if," "when," or "because," and superfluous prepositions used before "here" and "there"—were fairly simple to address, as they required a rule composed primarily of word or phrase matching.

In developing a rule that would mark "in that time," a phrase students in the study often used when "at that time" would be appropriate, the following entries were made in the Grammatik IV rule/help editor:

**Rule 1** Pattern: in that time

Advice: If you are referring to a specific moment of time, use "at that time."

Note that the rule pattern consists solely of a phrase. After the rule is saved in the Grammatik IV dictionary, the program will flag any occurrence of "in that time" and offer the advice above.

The rule pattern for fragments beginning with "if," "when," or "because" is slightly more complicated to develop, as it consists partially of a parsing rule.

**Rule 2** Pattern: @# \because

Advice: Make certain that you have written a complete sentence: (e.g., "Because it is raining I will stay home.")

In this pattern, "@" signifies a parsing rule, while "#" indicates a sentence position, and "\" indicates the beginning of the sentence. This rule will match any sentences beginning with "because" and provide the given advice, whether the sentence is a fragment or not. Rules to mark sentences beginning with "if" and "when" would be written in the same fashion.

The third rule is more complicated still.

**Rule 3** Pattern: thee @ <> | P

Advice: Delete the preposition before "there."

The pattern for superfluous prepositions before "there" contains a shift operator (<> that moves the part of speech indicator, in this case "P" indicating prepositions, before the word "there." This is necessary as rules cannot begin with the parsing rule indicator (@) except in cases where the indicator is immediately followed by the sentence position indicator (#), as in Rule 2. One problem with Rule 3 is that it will mark every occurrence of "there" which is preceded by a preposition. In some cases this will cause the program to give erroneous advice for sentence that are correct, such as: "She told him to sit in there for a moment."
Rule 4 will mark every occurrence of the verb "arrive" which is not followed by a preposition.

**Rule 4** Pattern: arriv* @~P

Advice: A preposition is often required after the verb "arrive." Do you need one in this case?

This rule contains parsing rule indicators that require explanation. The first is a wildcard (*) which causes the program to mark all words beginning with "arriv" and followed by any letters. This precludes having to write several different rules for each form of the verb "arrive." The second indicator (~) signifies that a match between the rule and a text will occur only if the part of speech following "arriv*" is not a preposition. Rule 4 will flag every occurrence and form of the word "arrive" not followed by a preposition. This causes the rule to flag correct usages, such as "I didn't know that he had arrived already." and "Their time of arrival has been changed."

The final rule is the most complicated rule created for this study.

**Rule 5** Pattern very @ <> @|B @|V

Advice: Check to make sure you have it correct. Don't write, for example, "I am very enjoy this place." Instead, write "I enjoy this place very much."

Rule 5 will match any occurrence of the word "very" preceded by a BE-verb (signified by B) and followed immediately by a verb (signified by V).

Rules 1 through 5 were tested with more than 75 sentences containing errors similar to those listed in 6 through 10 in Table 1. The rules were found consistent in marking errors of these types.

**CONCLUSION**

The option of customizing *Grammatik IV* for specific populations of ESL writers creates opportunities previously unavailable to composition teachers who utilize computerized writing aids. Important questions remain, however, about the benefits resulting from the time and efforts required to develop rules that mark some of the sentence-level, grammatical errors found in ESL student writing. Though the pattern matching rules developed in this study were created quite readily, the creation of parsing rules required a great deal of trial and error. Additionally, the results are not completely satisfactory because of several exceptions these rules overlook.

In addition to misgivings about the utility of customizing *Grammatik IV*
or other text analyzers for ESL writers, teachers should consider the following questions before adopting a text analysis program for their ESL composition classrooms:
—For what purpose is the text analysis program to be used?
—To what extent will a grammar checking program, such as Grammatik IV, improve student writing?
—What messages about writing are teachers conveying when they encourage ESL students to use text analysis?
—Could the time students spend with a text analysis program be given to other writing tasks that offer greater returns?
—If text analysis is utilized, where in the writing process should this occur?

These questions and others should be addressed as teachers consider the costs and the gains of incorporating text analysis into ESL composition classrooms.

REFERENCES


**AUTHOR'S BIODATA**

Mark Brock is a lecturer at the City Polytechnic of Hong Kong where he teaches classes in technical communication. In addition to computerized composition, his research interests include teacher development and the use of literature in the ESL classroom.

**Author's Address**

Mark N. Brock  
Department of English  
City Polytechnic of Hong Kong  
88 Tat Chee Avenue  
Kowloon, Hong Kong