Paragraph analysis has typically proceeded by doing autopsies on polished final texts. In a fuller analysis, however, we must consider the choices the writer made before arriving at the final text. In this paper we examine a college student's paper, first examining an analysis based on vocabulary changes (Vocabulary Management Profile) and second referring to an interview with the writer about her paragraphing choices.

Since the fifteenth century, the practice of indenting the first sentence of some unit of text has been standard (Moran 1984:425). Thus we might expect that there is some accepted sense of what these units we call paragraphs are, and of how they fit into larger units of text—whole discourse.

Yet there seems to be no generally accepted definition of the paragraph at all. Even the simplest definition based on written text—"text occurring between two indentations," or, in other words, a paragraph is whatever the writer or editor marks as a paragraph—is not automatically accepted as useful or appropriate. And although rhetoricians might agree that we don't know as much as we should about paragraphs, there is no clear sense of how to proceed towards a deeper understanding.

We would like to propose that a richer understanding will result from looking at paragraphs from two more angles (besides the traditional approaches briefly summarized below). One is to apply a recent discourse tool, the Vocabulary Management Profile, to our understanding of paragraphs; the other is to look at paragraphs throughout the revising process and to interview writers about the role paragraphs play in their writing.

Rhetorical Theory from Bain to the Present

The history of paragraphs in rhetoric has been well described in Ned Shearer's article, "Alexander Bain and the genesis of paragraph theory," and in Paul Rodgers, "Alexander Bain and the rise of the organic paragraph." This traditional view centered on a prescriptive notion of paragraphs, although as Rodgers argues, Bain (writing in 1866) did not consider actual paragraphs at all, but developed his ideas from applying sentence principles to the paragraph.... All evidence suggests he formed his model of the paragraph deductively, first by assuming a close organic similarity between paragraph and sentence, then by applying to the paragraph the classical, sentence-oriented rhetoric he had inherited. (Rodgers 1965:406)

In the 1960's several articles in College Composition and Communication marked a shift in a focus, away from prescriptivism. Although some of these articles (such as Becker's) did focus on structural patterns in paragraph development, a more unusual point was made by Paul Rodgers: "Structure precedes... the indentation that marks [a paragraph] physical limits.... Paragraphs are not composed; they are discovered" (Rodgers 1966:4-6). More radically, Leo Rockas believed that an indentation in the text was not determined by a unit, or organism, called the paragraph but could rather come before "almost any sentence of sophisticated prose" (Rockas 1964:6), and he used an essay by E. B. White to illustrate this assertion. Calling into question the notion of paragraph as a structural entity marked a shift away from a subject-centered view.

Another strand in paragraph analysis has focused on the role of indentations in providing cues to readers. This was noted in the earliest textbook devoted to the paragraph, Fred N. Scott
and Joseph V. Denney's *Paragraph-writing* of 1891: "The indented lines serve as landmarks for the reader's eye, enabling him to find his place again if he should happen to turn aside for a moment" (Scott and Denney 1891:69). But the importance of this perspective was not developed; Scott and Denney in fact rejected it, calling it "manifestly inadequate" either "as a practical rule of composition or as an explanation of the phenomena of paragraphing" (p. 69).¹ Probably the most thorough attempt at reformulating paragraph theory around the expectations of readers, however, can be found in Eden and Mitchell's "Paragraphing for the Reader," which appeared in 1986. Stating that "The teaching of paragraphs needs a revolution," Eden and Mitchell argue for a "reader-oriented theory of the paragraph," based on "how indentions affect the reader's perception of prose discourse" (Eden and Mitchell 1986:416). They believe that most paragraph theories err in focusing solely on the text, ignoring the reader's experience. Thus, their theory of the paragraph stresses that "good paragraphs will have initial sentences which effectively orient the reader" (p. 428) and urges students to find "their own ways of meeting and working with the reader's expectations" (p. 429).

In general, analysts have not mentioned the role that a writer's own goals may play in paragraphing, seeming instead to assume what Robin Bell Markels states overtly in *A New Perspective on Cohesion in Expository Paragraphs*: the alternative to indentations based on subject matter is simply statements of "authorial whimsy" (Bell 1984:2). Those who have attempted to look at the writer's role include Michael Hoey, in a 1983 paper entitled "The Paragraph boundary as a marker of relations between the parts of discourse," where he stated that "Decisions on where to paragraph and interpretations of existing paragraph boundaries are made... depending on the rhetorical needs of the writer." And Rodgers argued that a writer's "decision to indent may be taken for any one (or more) of at least half a dozen different reasons" (Rodgers 1966:5).

Rockas and Rodgers did explore the possibility that authors had reasons for their own paragraphing, but they did not interview actual writers. Instead, working from finished texts, they could only offer guesses about why White and, in Rodgers's analysis, the essayist Walter Pater indented as they did. As Eden and Mitchell attempt to bring both reader and writer into the picture, they state that "strategically paragraphed prose... molds and shapes [a message] to achieve the writer's purpose" (1986:416). They too go on to hypothesize about a writer's decisions, looking at a passage from the historian Barbara Tuchman and offering guesses such as, "It is possible that Tuchman... goes through a... process of tinkering with order and placement" (p. 425). Thus in their thesis that "Paragraphing is not part of the composing but of the editing process" (p. 417), it is unclear whether this is a description of what successful writers do or whether this merely represents their ideas of what unsuccessful writers ought to do in order to become successful.

### Discourse Studies and Written Paragraphs

The term *paragraph* is also used in linguistic studies on discourse, carried out by discourse analysts who have been working largely independently of rhetoricians. In these circles, there has been a tendency to dismiss the written paragraph as not relevant, or not necessarily relevant, to their analyses. Michael Hoey, for example, states in *On the surface of discourse*, "Historically, the paragraphs developed as a punctuation device. There is therefore no self-evident reason why sentences should join together into systematically organised units that coincide with the orthographic paragraph.... We cannot assume the natural divisions in a discourse to be those ortho-

¹ This focus on readers led to research using actual readers to determine paragraph boundaries; the most famous experiments are those of Koen et al. 1969, Stern 1976, and Bond and Haynes 1984. These results, however, did not lead to a unified view on analyzing paragraphs from the reader's perspective.
graphically signalled" (Hoey 1983a:9-12). And Longacre offers this argument for not accepting indented paragraphs as discourse units:

"Paragraph" is taken here to designate a structural rather than an orthographic unit. The paragraph indentations of a given writer are often partially dictated by eye appeal.... Conversely, a writer may put together several paragraphs as an indentation unit in order to show the unity of a comparatively short embedded discourse. Finally, the orthographic rule in English composition that we must indent for each change of speaker in a dialogue obscures the unity of dialogue paragraphs. (Longacre 1979:115-16)

This does not mean that there have been no discourse analyses which have looked at indented paragraphs. Even in these cases, however, the indentation is seen as a secondary signal; it merely reinforces the discourse unit which is already present. Brown and Yule, for example, after presenting a structural analysis of a passage of William Wharton's novel Birdy, then reveal that Wharton's indentations correspond exactly to their analysis (Brown and Yule 1983:97-99). Their conclusion is that the writer's indentations are one means of indicating topic-shift—but are typically reinforced by other linguistic features as well. Similarly, Youmans, in his study of several passages from James Joyce, deliberately limited himself to paragraphs which were "validate[d]... by joint evidence: paragraph breaks AND the principles" of various discourse analysts (Youmans 1992:763).

This mistrust of orthographically indicated paragraphs can probably be traced to the long-standing principle in linguistics that language is primarily spoken. Otto Jespersen, writing in 1924, states this principle clearly at the beginning of this classic The Philosophy of Grammar: "...the spoken and heard word is the primary form for language, and of far greater importance than the secondary form used in writing (printing) and reading" (Jespersen 1924:2). Because of this insistence on spoken language as being far more important than the written language, linguists have treated indented paragraphs in the same way that they have treated spelling and other conventions of written text: of some marginal interest, but not really relevant to the structure of the language.

Towards a More Realistic View

Based on this survey of discourse analysis and of rhetorical studies, we believe that composition teachers can move towards a clearer understanding of the paragraph in two important areas. First, we can become more aware of linguistic research in discourse structure; and second, we ought to consider more carefully the role of the writer in producing the text, seeking to understand how the indentations that exist in finished prose have come to be there.

To illustrate how these insights can give us a fuller understanding of paragraphs, we here present two analyses of a student final draft. The student, Margie, was enrolled in a freshman writing course at Illinois State University and wrote this paper in response to the first assignment of the semester. All students of this course were interviewed immediately after completing the assignment, using a stimulated recall method. We chose Margie for this analysis because she was the only student who spontaneously began talking about paragraphs, and throughout the interview she referred to paragraphs several times. In fact, the last thing she said as the interview was drawing to a close, was, "At least I know how to make paragraphs!"

Margie's Text

The assignment students were given was to write about a significant person in their lives. The assignment was made on Monday, the first day of the semester; a rough draft was due on

---

2 For more information on the research methodology and results, see Meyer 1987.
Wednesday (but was not collected), and the final was due on Friday. Here is Margie's final draft:

"One o'clock, the ghost isn't out. Two o'clock, the ghost isn't out. Three o'clock, the ghost isn't out... Twelve o'clock, midnight, the ghost is out tonight." When I think back at all the good times I had with Cathy, I remember how Cathy would get us, the neighborhood kids and I, all involved in the games: Ghost, Kick-the-Can, and Red Lion. The games gave me a sense of belonging and being part of a team. One lesson I got out of the games was the desire to be needed. Cathy has made me think about the direction my life will take through her teaching, her advice, and her leadership.

In High School, I kept my distance from the people I didn't like because I was afraid they would make fun of me if I did something wrong. Later, when I went to visit Cathy, I saw her accomplishments: Pugwash at University of Illinois, great grades, and a job that she really enjoyed. At the beginning of the fall semester at Illinois State University, I became motivated from seeing her accomplishments to join some clubs: ARH, TTSA, and College Republicans. I believe a great part of my drive to success came from Cathy.

Even though we are now far apart, she still finds a way to give me advice. Recently I received a letter from her. She told me that knowledge is learned, but wisdom is the processing of knowledge to make an educated opinion. Her advice relates to the problems I have with arguments. People can get into some bad arguments if they don't know what they are talking about. I enjoy hearing from her because I can learn a lot from her advice.

When I was younger, Cathy would babysit us, my brother and I, on New Year's Eve. First, we climbed into bed about three hours before midnight. Then, Cathy woke us five minutes before midnight so we could welcome in the New Year together.

Cathy is a very important part of my life. Most of my strength to become what I want to comes from her teachings, her advice, and her leadership.

A Vocabulary Management Profile Analysis

The discourse analysis technique which we present here is relatively new and is based quite closely on the vocabulary of a text. This technique, developed by Gilbert Youmans, is based on the assumption that found that "the simplest possible binary distinction between words in a text is the contrast between new and repeated vocabulary" (Youmans 1992:129). From this base Youmans has created the Vocabulary Management Profile (VMP), a computer program designed to analyze the occurrence of new words in a given text.

Youmans has produced several VMP programs, but this article will make use of one we will refer to as VMP Content. Taking one 35-word interval of text at a time, VMP Content counts the number of "content words" in the interval which occur for the first time in the text, and plots that number at the midpoint of the interval. "Content words" are considered any words excluding the 200 most common function words (and, but, or, etc.—these are counted as repeated rather than new as they occur in the text). The first interval occurs in words 1-35, the second in 2-36, and so on. As Youmans explains, "This procedure generates a curve, the VMP, which is a moving average of the number of new (content) words introduced over successive 35 word intervals" (Youmans 1992:129).

Increases in new words cause "peaks" on the VMP graph, while decreases in new words occurring correspond to "valleys" in the graph. Youmans notes that "an upturn in the curve... signals an increase in new vocabulary at the end of the interval, whereas a downturn signals an increase in repetitions" (Youmans 1991:765). As the text progresses, writers will generally present ideas, expound on them and then move on to something new. Most VMP graphs reflect this varied word usage, producing a curve which can rise and fall drastically at times; at the end
of a paragraph, relatively few new words are introduced, producing a valley in the VMP graph; at the beginning of a new discourse unit, or paragraph, the rate at which the author uses new words increases, producing a peak. In Youmans's analyses of an essay by George Orwell and fiction by James Joyce, he found that "VMPs correlate closely with constituent structure and information flow in discourse" (Youmans 1991:788).

Using Youmans's VMP program, we attempted to discover how well the VMP graph would correspond to paragraph indentations in Margie's text. The text contains five paragraphs with a total length of 337 words. Paragraph endings occur at words 92, 185, 266, 310 and 337, respectively. Figure 1 is the VMP chart, which plots the number of new content words over a given 35-word interval versus the total number of content words in the text. A quick glance at the chart immediately reveals that the paragraph endings do not directly line up with distinct valleys, as we had hoped. Several prominent drops do occur, along with several shorter upswings, but a closer, more methodical look is needed to discover just what the VMP is showing in relation to the text.

Figure 1: VMP Curve of Margie's Text

Beginning with ¶1, an immediate dip in the curve down to 8 new words is, while unexpected, quite explainable given the opening format which Margie uses. Her first three sentences involve the repetition of the phrase "(one/two/three) o'clock, the ghost isn't out." This, while an unusual manner of beginning a paper, does explain the corresponding dip in the VMP curve. The program encounters more repeated words than normal as early as the second sentence, and the chart reflects it. The VMP curve for the rest of ¶1 gives more promising results. From the first valley—a valley where the vertical axis value is eight at word 22, henceforth V8,22—the curve rises rather smoothly to a peak of fifteen at word 62, or P15,62, then descends again to a shallow valley at the paragraph ending V12,92 (though it does hit V11 twice just before the paragraph ending).

Paragraph two presents a different scenario, because no prevalent peaks or valleys occur in the VMP curve. In fact, new content words show only a small increase, from 12 to 16 by word 152, where the curve plateaus over the last 30 words of the paragraph. To find an explanation, we again look to the actual text. The paragraph opens with Margie referring to her high school experience, and then describing some of the things she admired about her friend, Cathy. This corresponds to the gradual rise in the curve, as new content words reach 16 by the beginning of
the following sentence. However, instead of bringing the paragraph to some closure, perhaps by summarizing what Cathy meant to her, Margie decides to bring up further examples of how Cathy inspired her. She lists several previously unmentioned clubs Cathy belonged to while at the university. Not surprisingly, the curve plateaus as new content words hold between 15 and 16 for the remainder of the paragraph.

Probably the most characteristic aspect of ¶3 is the consistent repetition of certain words and phrases throughout. With this in mind, the fact that the VMP curve drops from P15,192 to V9,262 by the end of the paragraph, comes as no surprise. Margie ends sentence one with the phrase "she still finds a way to give me advice." She again refers to "her advice" three sentences later, and ends the paragraph with a phrase not so different from the first: "...because I can learn a lot from her advice." The final usage of advice comes only 4 words after the valley at V9,262. Other cases of ¶3 repetition are found with the words arguments (2 occurrences) and knowledge (2).

While reasons for the downward slope of the curve are evident, these examples still do not account for the lack of an upswing in the VMP curve at the beginning of the paragraph. One explanation is that even the majority of words occurring in the first sentence are previously repeated in the text--words such as advice. More obvious, perhaps, is the fact that the VMP curve is already at a peak to begin with. In fact, the curve never rises above P15 after the word apart (¶3, sentence 1). The highest new-content-word count in the text is 16, thus for the curve of ¶3 to rise any more would perhaps be more unusual than if it fell, given its starting position.

Starting at V9,271, the curve for ¶4 rapidly climbs to 13 new content words and remains there after briefly ascending to a peak at welcome (P14,305). It is interesting to note that Margie here is bringing up another example from her childhood, probably from roughly the same time as ¶1. However, the content is entirely different--she talks of "New Year's Eve" and how she "climbed into bed about three hours before midnight" to rest before welcoming in the new year. Neither the event of New Year's Eve nor any vocabulary associated with bed and nighttime occurs in the text before ¶4. This influx of new content words associated with a new textual example once again explains why the VMP rises. But the briefness of the example, only two sentences with no summary or any placing of the example into a larger context, means that there is no valley at the end of the paragraph.

The results obtained from ¶5 are probably the most easily explainable of all segments of the VMP curve for Margie's final draft, due to the method by which Margie chose to conclude her paper. The entire paragraph consists of only two sentences: "Cathy is a very important part of my life. Most of my strength to become what I want to comes from her teachings, her advice, and her leadership."

Noticeably, important is the only new word contained in the first sentence, and from there the curve plummets to a valley only eight words later at strength (9,323). The concluding sentence is primarily a repetition of the last sentence in ¶1, with the final phrase differing by only one letter--the s in teachings. No upswing in the curve occurs, yet given the content, none is to be expected.

What has the Vocabulary Management Profile shown about Margie's draft and her use of paragraph indentations? And what are its limitations?

It is clear that the VMP shows something helpful about the text it is based on. In the case of Margie's paper, for example, the lack of VMP valleys at the end of paragraphs corresponds to the lack of closure. In Margie's second paragraph, for example, rather than returning to the idea that she had problems with people she didn't like, she ends with a list of activities that she was involved in and attributes her success to Cathy's influence.
On the other hand, the VMP cannot distinguish between genuine development/closure and simple repetition. The fact that the last sentence of the first paragraph and the last sentence of the last paragraph are nearly identical causes the VMP curve to drop at the end of ¶5, but these results would obtain if any sentence, no matter what its content, were repeated.

It is apparent that the VMP cannot be used as the sole measure of rhetorical effectiveness. It is a simple program which looks at an individual word and decides whether it is new or repeated—and nothing more. It can tell nothing about how a word fits into the context of a sentence, or whether a sentence "flows" well stylistically. Looking again at Margie's draft, she ends both ¶1 and ¶5 talking about teaching and advice. Yet nowhere in the paper is any distinction made between the two. ¶3 is devoted in a roundabout way to her friend's advice, but the last sentence of the opening and closing paragraphs is the only place where anything about teaching is mentioned. Both references to teaching are either unnecessarily redundant or grossly underdeveloped, yet they still affect the shape of the VMP curve. Thus while the VMP will show where certain words are used in a text, it cannot say anything about how they are used.

From the text Margie seems to be using teaching and advice as two different characteristics of her friend. In another context, of course, these words might be treated as synonyms. However, the VMP cannot distinguish between these two cases.

If the effectiveness of the VMP program is marginal in relation to certain areas of text analysis, it can prove to be a helpful tool if used in the proper context. It provides insight into information flow—the use of new and known information as measured by vocabulary. It seems a promising area for discourse analysis and thus provides a view of paragraphs which complements traditional rhetorical perspectives.

The Writer's Perspective: Information from the Drafts and from an Interview

The second perspective that we propose for richer understanding of paragraphs is the perspective of the writer through the writing process. Simply looking at Margie's drafts (she wrote four, although only two were required) gives us additional insight into the role paragraphs play in her writing and into the factors that are uppermost in her mind as she revises.

During the interviews the first question we asked all students was simply, "How did this writing go?" Margie mentioned paragraphs right away: "It was kind of difficult at first.... I usually like to start things, I start like a paragraph and then I go on to another paragraph. I just like to work that out first." As we followed up on her answer, we asked her, "What's the first thing that you try to work out?" She answered, "The paragraphs."

Margie is also unusual in the group of writers because of the number of drafts she wrote. All students were required to write two drafts as part of the assignment, and most students wrote only two; Margie wrote four. Margie's first draft contained only two paragraphs. This was, she said, part of her strategy for writing; she wants to "work out" the paragraphs first: "I only started out with two paragraphs, and then I went on with my second draft a little further." These are her first two paragraphs, her entire first draft (the first lines are the refrain of a chant associated with a childhood game she remembers):

One o'clock, the ghost isn't out. Two o'clock, the ghost isn't out. Three o'clock, the ghost isn't out. Twelve o'clock, midnight, the ghost is out tonight. When I think back at all the good times I had with Cathy, I remember how Cathy would get us all involved in the games. In High School, I didn't get too active, keeping my distance from the people I didn't like. For the first time since then, I have the motivation to join some clubs. I believe a great part of my drive came from Cathy because I see what she accomplished: Pugwash at U of I, high grades, a good job after school.
Before break, I wrote Cathy a letter, explaining my reason for craving to learn about issues. After break, I found her letter in my mailbox. She responded, "You are right that you need knowledge and wisdom to tackle the type questions you are beginning to become interested in. Just don’t confuse the two concepts."

Margie’s second draft shows her both working on these two paragraphs and going "a little further." These first two paragraphs are revised, but they are clearly the same paragraphs:

One o’clock, the ghost isn’t out. Two o’clock, the ghost isn’t out. Three o’clock, the ghost isn’t out. Twelve o’clock, midnight, the ghost is out tonight. When I think back at all the good times I had with Cathy, I remember how Cathy would get us, the neighborhood kids and I, all involved in the games: Ghost, Kick-the-Can, and Red Lion. The games gave me a sense of belonging and being part of a team. One lesson I got out of the games was the desire to be needed. In High School, I kept my distance from the people I didn’t like because I was afraid they would make fun of me if I did something wrong. For the first time since my graduation, I became motivated from seeing her accomplishments to join some clubs: A.R.H., T.T.S.A, and College Republicans. I believe a great part of my drive to success came from Cathy because when I went to visit Cathy, I see what she accomplished: Pugwash at U of I, great grades, and a good job after school.

Even though we are far apart, she still finds a way to give me advice. Recently, I received a letter from her. She told me that knowledge is learned, but wisdom is the process of knowledge to make a educated opinion. I enjoy hearing from her because I can learn a lot.

After she had revised these two paragraphs, Margie went on with two new paragraphs: one is about Cathy’s babysitting for her on New Year’s Eve, and the other is a conclusion.

Margie’s revision of the first paragraph included adding more description of the games she played when she was young, an explanation of the importance of these games, and more details of clubs Cathy is in now; she also added a transition to the second paragraph. The second paragraph shows a more significant change: Margie explained the difference between knowledge and wisdom in the second draft. These changes are apparently what Margie means by "working out" the paragraphs.

After Margie had written the second draft, she asked a friend to read it and comment on it. On the copy of the second draft a paragraph mark had been added between "Red Lion" and "In High School." When we asked Margie why she had decided to add an indentation at this point, she said, "That was advice." Her friend had pencilled in the paragraph mark as she was reading the draft.

We asked Margie why she had taken her friend’s advice. At first she seemed surprised by the question, but then answered, "Because I wanted to start on a new idea.... That was about games and this is about my success and things like that, my impressions and her impressions, they’re two different things really." And once Margie decided to take her friend’s advice and indent, she did not change this indentation again. She added two sentences to the end of the first paragraph, but in both the third draft and the fourth and final draft, the second paragraph begins with the sentence "In High School, I kept my distance from people I didn’t like."

Towards a Richer Understanding

For Margie, these new perspectives take us towards an understanding of why her paragraphs are unsatisfactory: the VMP shows that they tend to lack closure or to have artificial closure. The interview showed us that Margie did rely on reader feedback in paragraphing: there is some sense in which paragraphing is part of editing for her, as Eden and Mitchell argue it should be. We saw that some of Margie’s indentations stayed the same throughout the writing process, but
others were added as she revised. And we learned that she thought she knew everything there was to know about paragraphs: "At least I know how to make paragraphs!" If we as teachers are to help Margie, we must first understand that she needs to be convinced of her need for instruction.

To date, composition teachers' understanding of paragraphs has been based on traditional patterns of development and on reader expectations. We propose here, however, that increasing the lenses through which we look at paragraphs will lead to a deeper and richer understanding of paragraphs. Recent linguistic work in discourse analysis, using tools such as the Vocabulary Management Profile, and process-centered composition research, including examining drafts and interviewing writers—as well as other perspectives still to be applied to paragraphs—can only make our comprehension more complete.

References
Hoey, Michael. 1983b. The paragraph boundary as a marker of relations between the parts of discourse. Mimeographed.