

Gilbert Youmans
 Department of English
 University of Missouri-Columbia
 Columbia, MO 65211

The Vocabulary-Management Profile:
 Two Stories by William Faulkner

Abstract

The Vocabulary-Management Profile (VMP) is a graph of the moving average of the number of new vocabulary words introduced over successive intervals of text. Peaks and valleys on the VMP correlate closely with narrative structure: new episodes, new settings, new characters are signaled by an increase in new vocabulary, and hence by upturns on the curve. Conversely, downturns in the VMP normally signal a continuation of the episode, description, or characterization. Higher-level boundaries in discourse (such as those between paragraphs and narrative episodes) normally correspond with deeper valleys on VMPs than do lower-level boundaries (such as those between clauses and sentences). Hence, VMPs provide surprisingly accurate graphical representations of the hierarchical structure of discourse. In general, the VMPs for "A Rose for Emily" and "Dry September" confirm this close correspondence between VMPs and narrative structure. However, VMPs are formally equivalent to unlabeled tree diagrams, and like such diagrams, they can provide only partial, not complete analogues for the structure of discourse.

Most empirical studies of literature calculate statistics for entire texts or for groups of texts--the average length of sentences, the frequencies of words, phonemes, punctuation, and the like. Statistics such as these are often interesting and informative, but they reveal nothing about the constituent structure of discourse. By contrast, the Vocabulary-Management Profile (VMP), which was introduced for the first time in [1], provides a direct visual analogue for constituent structure: major valleys on the VMP normally occur near discourse boundaries--near divisions between paragraphs, episodes, and the like--and major peaks normally occur near the middle of paragraphs, episodes, and so on. In [1], this close correlation between constituent structure and VMPs is illustrated for five texts: James Joyce's short story "Eveline," and the first two thousand words of "The Dead," Finnegans Wake, and George Orwell's 1984 and "The Principles of Newspeak."

In [2], Olson found a similar correlation between VMPs and the constituent structure of ten oral discourses: ". . . five first person oral narratives, three third person oral narratives, one conversation interspersed with first person oral narratives, and one unplanned oral expository speech" [2, p. 69]. Applying principles of discourse analysis derived from contemporary linguistics, Olson divided each of these ten oral discourses into thematic paragraphs. He found a significant correlation between thematic breaks and valleys on VMPs: "89% of valleys correspond to thematic breaks, 66% of these occurring at major breaks and 23% at minor breaks. . . . Not only do most valleys mirror thematic breaks, but most thematic breaks (80%) show a slowdown in new vocabulary introduction that produces valleys in the VMPs. 85% of major thematic breaks and 69% of minor ones produce valleys" [2, pp. 65-66]. Hence, Olson's study supports the claim that ". . . VMPs correlate closely with constituent structure and information flow in discourse" [1, p. 788].

In this article, I will examine the correspondence between VMPs and higher-level constituents--narrative episodes rather than paragraphs--using "A Rose for Emily" and "Dry September," by William Faulkner, as illustrations [3]. I chose these stories partly because Faulkner explicitly marks boundaries between narrative episodes, dividing each story into five numbered sections. In addition, both stories present special challenges for VMP analysis because they depart from conventional, straight-ahead narration: "A Rose for Emily" frequently deviates from temporal sequence, including

several "flashbacks," whereas "Dry September" is an example of interlaced narration, with sections one, three, and five telling the story of one set of characters, and sections two and four telling the story of another. In the terminology of transformational–generative grammar, the narrative sequence of "A Rose for Emily" might be said to result from a series of movement transformations that alter the story's underlying temporal sequence. "Dry September," by contrast, illustrates discontinuous ("gapped") constituency, with sections one, three, and five forming one discontinuous constituent, and sections two and four forming another. In sentence-level syntax, discontinuous constituency is illustrated by conjunctions such as "either . . . or," by agreement rules, and by anaphora (represented by subscripted variables in generative grammar).

Later, I will argue that VMPs are formally equivalent to unlabeled surface-structure tree diagrams. As such, VMPs are usually successful at singling out constituents of discourse (often remarkably so), but VMPs cannot distinguish one type of constituent from another; for example, they cannot distinguish descriptive passages from narrative episodes, or dialogue from expository paragraphs, or whatever. Similarly, VMPs cannot identify narrative transformations (such as flashbacks) or discontinuous constituents. Hence, VMPs are not complete structural analyses of discourse any more than unlabeled tree diagrams are complete structural analyses of sentences. Nevertheless, unlabeled tree diagrams provide considerable information in themselves; therefore, VMPs are likely to prove useful tools in discourse analysis, particularly since they are generated automatically by computer and since most previous statistical analyses have ignored the constituent structure of discourse altogether.

The simplest possible binary distinction between words in a text is the contrast between new and repeated vocabulary. This is the distinction that underlies the VMP, which is constructed as follows. First, an interval, say fifty-five words, is selected.¹ Next, in the first interval (words 1-55) we count the number of words that appear for the first time in the discourse, and we plot that number at the midpoint of the interval (the 27th word). Then we count the number of new words in the next interval, 2-56, and plot the number at the midpoint of that interval (the 28th word), and so on for the remainder of the text. This procedure generates a curve, the VMP, which is the moving average of the number of new words introduced over successive fifty-five–word intervals [1, pp. 771-76].²

The VMP for a concordance or a similar list of words is a straight line parallel to the x-axis, with the number of new words remaining constant at fifty-five; that is, every word is new, and no words are repeated. At the opposite extreme, the VMP for an endless repetition of words or nonsense syllables, such as "Fa la la la la, la la la la . . . ," is a straight line that coincides with the x-axis; that is, after the first repetition, the number of new words remains constant at zero.

VMPs for more typical discourses fluctuate between these two extremes, rising and falling to clearcut peaks and valleys much the way stock market averages do. These fluctuations correlate with the flow of information in discourse.³ When authors introduce a new topic, the incidence of new vocabulary tends to increase (and repetitions decrease); conversely, when authors elaborate upon an old topic, the incidence of new vocabulary tends to decrease (and repetitions increase). Hence, major upturns in the VMP tend to occur near major constituent boundaries of a discourse; for example, near the beginning of a new episode in a story or with the description of a new setting or the introduction of a new character. Conversely, downturns in the VMP normally signal a continuation of the episode, description, or characterization.

If our goal is to track the flow of information in discourse, various refinements in the VMP suggest themselves. For example, semantic content words such as love, time, beauty define narrower topics than do syntactic function words such as the, of, and to. Consequently, VMPs will correlate more closely with information flow if we

count the first appearances of syntactic function words such as the as repeated rather than as new vocabulary. Similarly, when tracking the flow of information, we might want to treat inflected forms such as love, loves, loved, and loving as variants of the same word rather than as four different words. Also, we might want to treat synonyms such as trip and journey as though they were one word rather than two.

In practice, however, "deleting affixes and conflating synonyms appear to be an unnecessary refinement in VMPs if their purpose is to provide graphical signals for major shifts in the flow of information in English discourses" [1, p. 776]. The reason for this is that inflectional variants and synonyms occur relatively infrequently and their first occurrences tend to be distributed more or less evenly throughout a text. Consequently, conflating inflectional variants and synonyms reduces the overall height of the VMP but has little effect upon the major turning points in the curve. That is, the revised version of the VMP looks very much like the original one except that the revised version is lower.

On the other hand, "distinguishing between function words and content words does have a significant effect [on VMPs], particularly for the first 500 words of a text" [1, p. 776]. The reason for this is that function words are the most frequently occurring words in the language; consequently, a disproportionately large part of the new vocabulary at the beginning of a text is likely to be composed of words such as the, of, and, a. This clustering effect tends to obscure information peaks and valleys in the early part of the VMP. (After 500 words, new function words are introduced much less frequently, and they have only minor effects on the VMP.) The computer program that was used to generate VMPs for [1] dilutes early clustering effects by counting the 200 most common function words as repeated rather than as new vocabulary. This results in a "skeletal" VMP that correlates more closely with information flow, especially at the beginnings of texts. Hereafter, I will focus upon this skeletal version of the VMP.

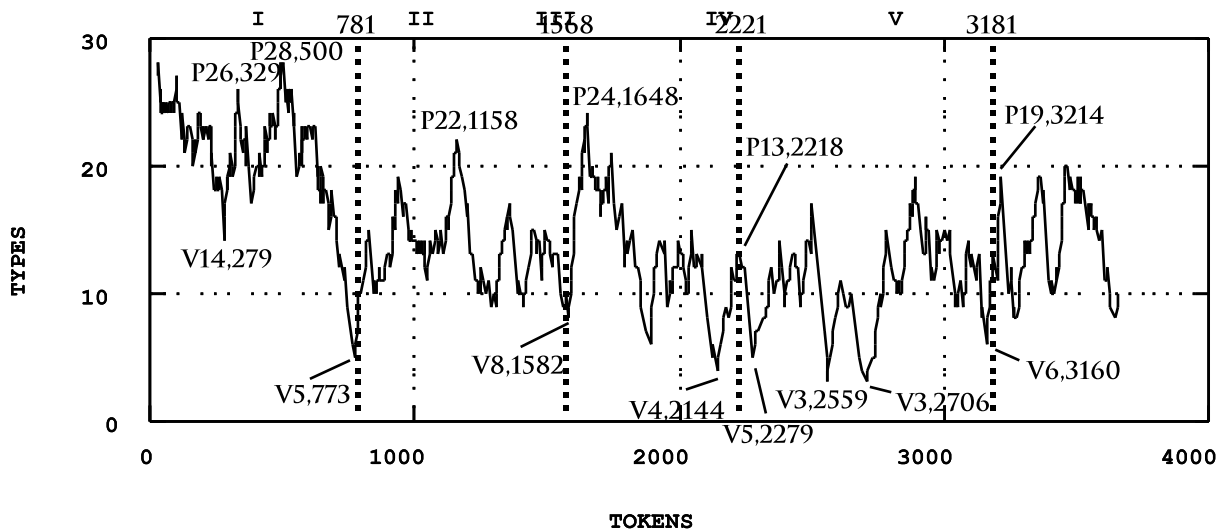
The typical characteristics of VMPs for narratives can be illustrated by the VMP for "A Rose for Emily" in figure 1. The horizontal axis, labeled "Tokens," is the total number of words of text that have accumulated by the midpoint of a moving interval fifty-five words long. The vertical axis, labeled "Types," is the number of new vocabulary words introduced over successive fifty-five-word intervals (ignoring the 200 most common function words). The vertical dotted lines at 781, 1568, 2221, and 3181 tokens indicate Faulkner's dividing lines between the story's five sections, I-V.

Figure 1. The VMP for "A Rose for Emily" (the moving average of new vocabulary words in successive fifty-five-word intervals of text).

I have adopted the following convention for identifying peaks and valleys in figure 1: P28,500 (in section one) is a peak of twenty-eight new vocabulary words found in a fifty-five-word interval centered at the 500th token (word) of the text; whereas V3,2559 (in section four) is a valley of just three new vocabulary words found in a fifty-five-word interval centered at the 2559th token. In theory, the number of new vocabulary words in a fifty-five-word interval could range from fifty-five to zero. For "Emily," however, the actual variation is from a high of twenty-eight (at P28,500) to a low of three (at V3,2559 and V3,2706). This sort of variation is typical for a story about this long (3683 words).

One striking characteristic of the VMP for "Emily" is a cyclical alternation "between peaks and valleys (that is, between new and repeated vocabulary), reflecting two competing principles that necessarily underlie all normal discourse: innovation and coherence" [1, p. 763]. These peaks and valleys are interpreted as follows: "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" [1, p. 765]. These upturns and downturns tend to correlate closely with the ebb and flow of new topics in discourse.

The VMP for "Emily" also illustrates another typical characteristic of VMPs: they tend to drift downward over time. The total vocabularies of authors are finite; and genre, subject-matter, and audience impose additional limitations upon vocabulary.



Consequently, with every new word in a discourse, authors reduce their stock of available unused vocabulary, thereby lowering the probability of introducing new words. We should take this downward drift into account when interpreting VMPs. For example, in figure 1, the early peak at P26,329 is actually less significant than the later, lower one at P24,1648, for three reasons: (1) the later peak is only slightly lower (by just two types) than the earlier one; (2) the later peak at P24,1648 is preceded by a lower valley on the VMP (V8,1582), yielding a difference between peak and valley of sixteen types; whereas the earlier peak at P26,329 is preceded by a higher valley (V14,279), yielding a difference between peak and valley of just twelve types. For VMPs, relative heights (the differences between peaks and valleys) are more significant than absolute heights. Hence, relatively speaking, the peak at P24,1648 is "higher" than the one at P26,329, and this greater height is all the more significant for a peak occurring later in the VMP (where differences between peaks and valleys tend to be smaller). And, finally: (3) the peak at P26,329 is clearly subordinate to the following peak at P28,500, which is the highest peak in section one; whereas P24,1648 is the highest peak in section three, and it is superordinate to all the other peaks in its section.

This reference to subordinate and superordinate peaks implies that there is a structural analogy between VMPs and sentences, suggesting in turn that there ought to be a formal relationship between VMPs and the tree diagrams drawn by contemporary syntactic theorists. Figure 2 illustrates this formal equivalence between VMPs and trees by demonstrating how subordinate, superordinate, and coordinate peaks on VMPs can be translated into the tree notation that was introduced by Lerdahl and Jackendoff to represent the hierarchical structure of tonal music [8] and extended by Jackendoff to represent phonological structure and syntactic structure in human language [9]:

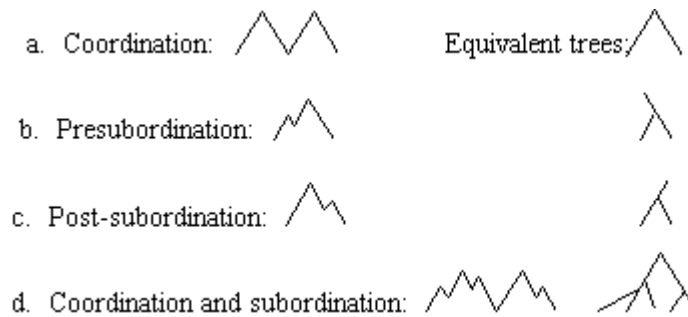


Figure 2. Equivalence between VMPs and tree diagrams.

Lerdahl and Jackendoff's tree notation embodies their Reduction Hypothesis for tonal music:

Reduction Hypothesis: The pitch-events of a piece are heard in a hierarchy of relative importance; structurally less important events are heard as ornamentations or elaborations of events of greater importance [9, p. 24].

This principle is called the Reduction Hypothesis because omitting less important musical notes has little effect on the ability of listeners to recognize a musical piece, whereas omitting more important notes makes the piece unrecognizable.

The Reduction Hypothesis can be applied to sentences as well as to music: some words, phrases, and clauses can be omitted from sentences more easily than others; and less important words, phrases, and clauses typically attach themselves to more important ones (often as modifiers). Similarly, at the level of discourse, some sentences, paragraphs, and episodes can be omitted more easily than others, and the less important constituents typically subordinate themselves to the more important ones. Theoretically, then, the Reduction Hypothesis provides a basis for constructing a tree diagram for any coherent verbal discourse, even one as long as *War and Peace*, although such a diagram would necessarily embody an interpretation of the text, and different analysts would be likely to construct different trees.

Classifying constituents of discourse as subordinate, superordinate, or coordinate requires interpretive judgment; consequently, any such classification is likely to be controversial. By contrast, a surface-level grouping of discourse constituents (ignoring the the issue of subordination) is likely to be less controversial, particularly in the case of written texts, which authors themselves have already segmented into sentences, paragraphs, episodes, chapters, and the like. For example, in "A Rose for Emily" and "Dry September," Faulkner groups words into phrases; phrases into clauses; clauses into sentences; sentences into paragraphs; and paragraphs into numbered sections. Of course, alternative groupings would be possible, but Faulkner's explicit divisions are literally authoritative, and it is interesting to examine how closely the VMPs for his stories correspond with his segmentation of the text.

Faulkner divides "Emily" into five numbered sections.⁴ In addition, his narrator (an anonymous member of the town) jumps backward and forward along the time-line at several points in his story. This narrative technique generates especially clearcut boundaries between discrete episodes (some of which resemble flashbacks in movies). The order of narration of the story is summarized below, with Faulkner's sections numbered I-V.

Order of narration in Faulkner's "A Rose for Emily"

- I (a) Paragraphs 1-2: Emily's funeral mentioned. Her house described.

- (b) 3: Colonel Sartoris remits her property taxes. 4-14: The next generation tries and fails to collect.
- II (a) 15-24: Thirty years earlier, Emily's neighbors complain that her house smells, but the authorities refuse to confront her.
- (b) 25-28: Her family history through her father's death.
- III (a) 29-32: Emily's convalescence and the beginning of her affair with Homer Barron, a Yankee foreman.
- (b) 33-42: She buys arsenic.
- IV (a) 43: Town's reaction to Emily's arsenic purchase.
- (b) 44-45: Relatives come to town to intervene in the Homer Barron affair.
- (c) 46-47: The last appearance of Homer Barron.
- (d) 48-53: Emily ages and dies. 48: Her hair turns gray.
- V (a) 54-55: Emily's funeral.
- (b) 56-60: The discovery of Homer's corpse in Emily's bed and a strand of her gray hair on the pillow beside him.

The "surface structure" of this story (its order of narration) deviates sharply from its "underlying structure" (the temporal sequence of the events recounted), summarized below:

- Order of experience: (from the town's viewpoint)
- II (b) Family history through the death of Emily's father.
- I (b) Emily's taxes remitted.
- III (a) The beginning of her affair with Homer Barron.
- IV (b) Relatives come to intervene in the affair.
- III (b) Arsenic purchase.
- IV (a) Reaction to arsenic purchase.
- (c) Homer Barron's disappearance.
- II (a) The smell.
- I (b) 4-14: Failed attempt to collect Emily's taxes.
- IV (d) Emily ages and dies.
- I (a) Emily's funeral.
- V (a) Emily's funeral.
- (b) The discovery of Homer Barron's corpse and Emily's gray hair on the pillow beside him.

One rhetorical effect of the narrative "movement transformations" implicit in the surface structure of this story is to clarify Faulkner's boundaries between episodes. Narrative transformations can apply only to narrative constituents; hence, "moving" an episode or event out of temporal sequence marks that episode or event as a discrete, separable unit. For example, the division between the first and second sections of "Emily" coincides not only with a paragraph boundary and the Roman numeral two, but also with an explicit regression on the time-line: ". . . just as she had vanquished their fathers thirty years before about the smell" [followed by ten paragraphs dramatizing this episode]. Any narrative boundary this explicit ought to show up clearly on the VMP, and figure 1 illustrates how remarkably accurate the VMP turns out to be in this case: the boundary between sections one and two coincides almost exactly with a sharp upturn after the valley at V5,773. This upturn begins with the word Tobe in the excerpt below:

"I have no taxes in Jefferson. [Upturn after V5,773] Tobe!" The Negro appeared. "Show these gentlemen out."

II

So she vanquished them, horse and foot, just as she had vanquished their fathers thirty years before . . . (121)

In this case, the VMP for "Emily" corresponds almost exactly with Faulkner's own explicit segmentation of the text. The boundary between sections one and two comes just eight words after V5,773, which is the deepest valley on the VMP in the first two sections of the story, implying that V5,773 marks the most significant boundary up to that point in the narrative. This an ideal result: if VMPs corresponded perfectly with the hierarchical tree structure of a discourse, then all the minor divisions in the text (such as those between clauses and sentences) would coincide with minor valleys on the VMP, and progressively more significant divisions (such as those between paragraphs and sections of the story) would coincide with progressively deeper valleys on the VMP. The valley at V5,773 conforms with this ideal exactly.

Another way to illustrate the remarkably close correspondence between the VMP for "Emily" and a hypothetical tree diagram for the first two sections of the story is to note that all the subsidiary peaks preceding V5,773 are subordinate to the central, most prominent peak (P28,500) in the first section; whereas, after V5,773, the subsidiary peaks in the second section are subordinate to the central, most prominent peak in that section (at P22,1158). Again, this is an ideal result.

In addition to accurately signaling the boundary between sections one and two, the VMP reflects the complex internal structure within each section. For example, the deepest valley before V5,773 in section one occurs at V14,279. The upturn after this valley begins with the word modern, which occurs in the first sentence of the episode in which the town fails to collect Emily's taxes:

Colonel Sartoris . . . remitted her taxes . . .

When the next generation, with its more [Upturn after V14,279] modern ideas, became mayors and aldermen, this arrangement created some little dissatisfaction. On the first of the year they mailed her a tax notice. [Followed by eleven paragraphs about the attempt to collect Miss Emily's taxes.] (119-20)

As the VMP demonstrates, the bulk of section one is devoted to this failed attempt to collect Emily's taxes--the first of several episodes illustrating that she is an outlaw (whose insanity would make her immune to prosecution). Hence, in this case, the VMP correctly signals the most important subsidiary boundary within section one as well as the higher-level boundary between sections one and two.

The boundary between sections two and three is also clearly marked on the VMP, by the valley at V8,1582, which is lower than any other valley between it and V5,773. V8,1582 is followed by a sharp upturn beginning with the word hair in the second sentence of section three; thereafter, the VMP rises rapidly to a peak at P24,1648 with the word Yankee:

III

She was sick for a long time. When we saw her again, her [Upturn after V8,1582] hair was cut short, making her look like a girl, with a vague resemblance to those angels in colored church windows--sort of tragic and serene.

The town had just let the contracts for paving the sidewalks, and in the summer after her father's death they began to work. The construction company came with niggers and mules and machinery, and a foreman named Homer Barron, a Yankee [P24,1648]--a big, dark, ready man . . . (124)

Relatively speaking, the peak at P24,1648 (which is sixteen types higher than the preceding valley at V8,1582) is the most prominent one on the VMP. Appropriately, this peak coincides with the introduction of Homer Barron, who becomes the major figure in Emily's life for the rest of the story.

As the foregoing discussion illustrates, the boundaries of sections one and two coincide almost exactly with clearcut valleys on the VMP, and the VMP also mirrors the internal structure of these sections. However, the division between sections three and four, indicated in figure 2 by the vertical dotted line at 2221 tokens, coincides with a peak rather than a valley on the VMP. Hence, in this case, the VMP is out of phase with Faulkner's explicit segmentation of the story. The reason for this is that the division within the narrative is less clearcut: there is no shift in the time-line, and Faulkner continues his discussion of Emily's arsenic purchase beyond section three into section four, a fact which is highlighted by his use of the subordinating conjunction so to begin the new section. As a result of this continuation, the upturn in the VMP signaling a new topic is delayed until the third sentence of section four:

The Negro delivery boy brought her the package; the druggist didn't come back. When she opened the package at home there was written on the box, under the skull and bones: "For rats."

IV

[Downturn after P13,2221] So the next day we all said, "She will kill herself"; and we said it would be the best thing. When she had first begun to be seen with Homer Barron, we had said, "She will marry him." Then we said, "She will persuade him yet," because Homer himself had remarked--he liked men, and it was [Upturn after V5,2279] known that he drank with the younger men in the Elk's Club--that he was not a marrying man. (126)

Unlike the boundary between sections three and four, the boundary between four and five is once again in phase with the VMP. A clearcut valley at V6,3160 is followed quickly by a peak at P19,3214. The upturn begins with the word walnut in the predicate of the last sentence of section four:

She died in one of the downstairs rooms, in a heavy [Upturn after V6,3160] walnut bed with a curtain, her gray head propped on a pillow yellow and moldy with age and lack of sunlight.

V

The Negro met the first of the ladies at the front door and let them in with their hushed, sibilant voices . . . (129)

The final section of the story has three major peaks, and these, too, correlate with significant subdivisions within the narrative.

Overall, then, the VMP for "Emily" provides a remarkably accurate visual analogue for the story's narrative structure. The VMP mirrors both major and minor narrative constituents quite successfully, although it cannot identify "moved" constituents. For example, by itself, the VMP provides no evidence that section two describes events that precede rather than follow the events described in section one. However, apart from this inherent blindness to narrative transformations, the VMP has no difficulty with moved constituents; on the contrary, their boundaries tend to be marked even more clearly on VMPs than are the boundaries of "unmoved" constituents. "Gapped" constituency, by contrast, does tend to obscure constituent boundaries on VMPs, as an analysis of "Dry September" will illustrate. The story's five sections are summarized below:

- I. Hawkshaw's barber shop. Hawkshaw tries to spike rumors that Will Mayes, a Negro, raped Minnie Cooper, a white woman. McLendon, a hot-headed war veteran, bursts in to recruit a vigilante posse.
- II. Minnie Cooper's biography. Minnie was popular briefly as a youth, but she is now a sexually repressed spinster nearing forty.
- III. The murder of Will Mayes. Hawkshaw accompanies the vigilantes to try to dissuade them. After Will Mayes is captured, Hawkshaw jumps from the car. He does not witness the actual murder.
- IV. Minnie Cooper's night on the town. Once again an object of lascivious interest, Minnie goes to a movie, where she gives way to hysterical laughter and must be led home.
- V. Coda. After the murder, McLendon returns to his house, where he "half struck, half flung" his worried wife across the chair because she waited up for him against his orders.

This summary illustrates that "Dry September" is not a simple, straight-ahead narrative. Instead, Faulkner shifts back and forth between two related stories: between the Hawkshaw-Mayes-McLendon episodes (sections one, three, and five) and the Minnie Cooper episodes (sections two and four).

The VMP for "Dry September" is shown in figure 3. (The vertical dotted lines indicate Faulkner's boundaries between the story's five sections.)

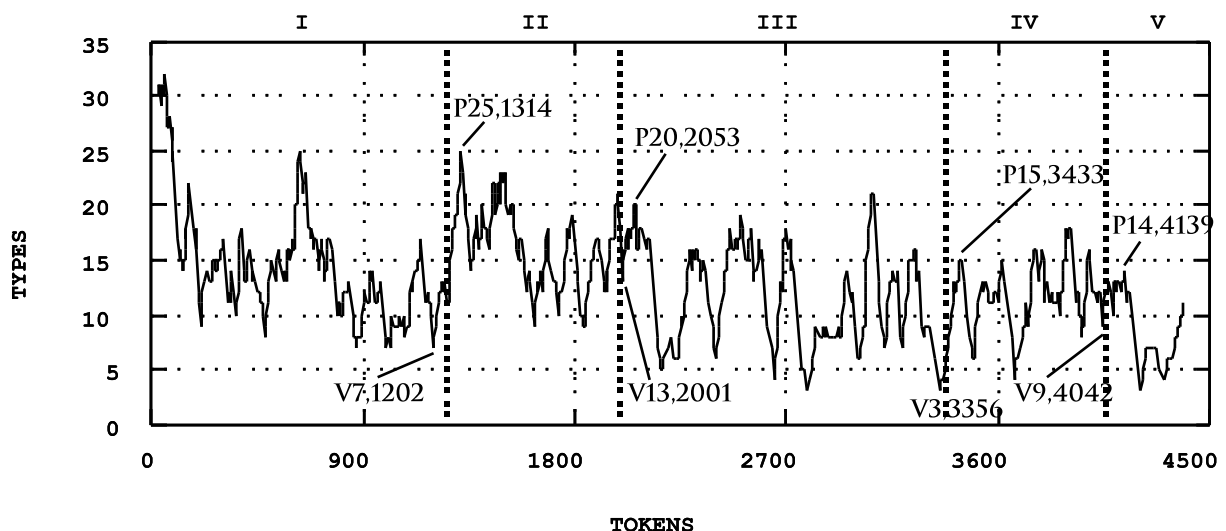


Figure 3. The VMP for "Dry September" (fifty-five-word intervals).

As figure 3 illustrates, the bulk of the story is devoted to section one (Hawkshaw's barber shop) and section three (the murder of Will Mayes). These two episodes, which are about the same length (1288 and 1382 words), are considerably longer than the Minnie Cooper episodes, which are also similar in length (730 and 685 words). In addition, the Hawkshaw-Mayes-McLendon episodes are more vivid in the sense that they rely almost exclusively upon dialogue and dramatized narrative, whereas the Minnie Cooper episodes include summarized and generalized narrative. With respect

to both length and vividness, then, the Hawkshaw–Mayes–McLendon episodes can be considered the "main" story and the Minnie Cooper episodes "interpolations." VMPs cannot distinguish interpolated constituents from any other kind of subordinate constituent. Correlatively, VMPs cannot identify narrative bridges across intervening constituents. That is, VMPs cannot represent discontinuous constituency. For example, the VMP for "Dry September" provides no signal that section three of the story is a continuation of section one rather than a continuation of section two, or that section five is a continuation of section three rather than section four.

One reason VMPs cannot represent discontinuous constituency is that they treat all repetition the same way, whether it is immediate (as in "And miles to go before I sleep, / And miles to go before I sleep") or delayed (as in a verse refrain). Hence, VMPs have no way of indicating whether repeated constituents are adjacent to each other or separated by gaps. Partly because of this limitation, the VMP in figure 3 is successful at signaling the onset of each interpolated Minnie Cooper episode (where the narrative departs from the main story-line), but less successful at signaling the endpoint of each interpolation (where the narrative returns to the main story-line). That is, the beginnings of sections two and four are preceded by major valleys on the VMP (at V7,1202 and V3,3356) and followed by major peaks (at P25,1314 and P15,3433), but the endings of these episodes are signaled only by minor valleys (at V13,2001 and V9,4042), which, being relatively high, tend to obscure the subsequent peaks (at P20,2053 and P14,4139).

The masking effect on the VMP of the interpolated Minnie Cooper episodes can be illustrated by deleting sections two and four from the story and constructing a new VMP for the Hawkshaw-Mayes-McLendon episodes, which form a complete story in themselves. The result is shown in figure 4.

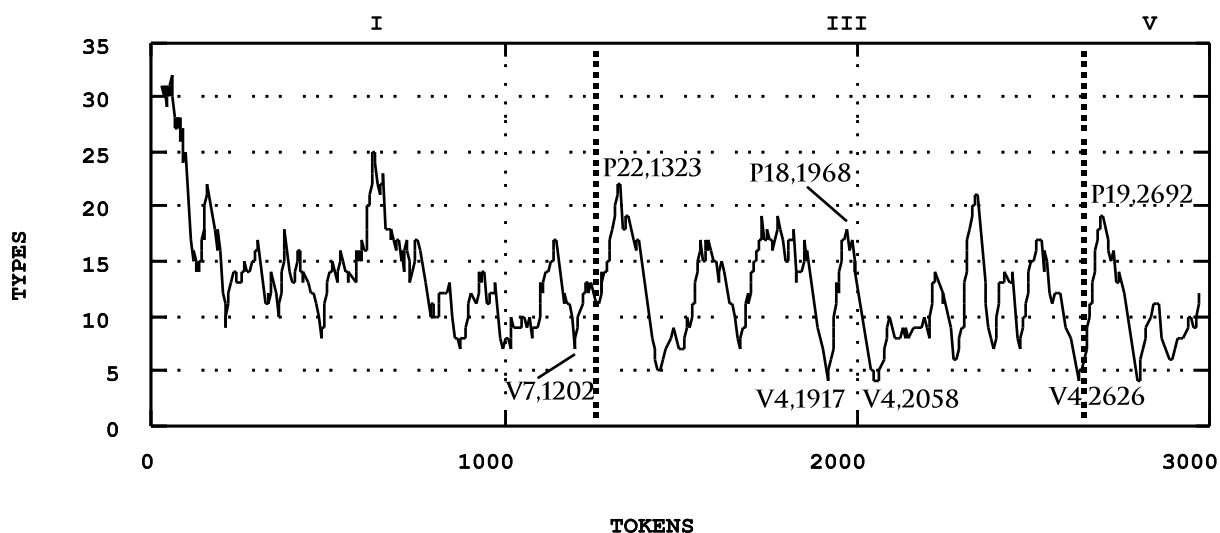


Figure 4. The VMP for the Hawkshaw-Mayes-McClendon sections of "Dry September" (with the Minnie Cooper sections, two and four, deleted).

Necessarily, the VMP for section one is the same in figure 4 as it is in figure 3 except for the difference in the horizontal scale. After section one, however, the VMP in figure

4 correlates more closely with Faulkner's division of the Hawkshaw-Mayes-McLendon story into three discrete sections. The onset of the final coda is especially clear, with a major valley at V4,2626 followed by a major peak at P19,2692. The turning point at V4,2626 occurs just seventeen words (and one sentence) before the beginning of the last section of the story.

The VMP does not correlate this closely with the boundary between the first and the middle sections of figure 4. The major turning point before the middle section occurs with the valley at V7,1202, which is followed by a major peak at P22,1323. This valley at V7,1202 anticipates the section boundary by sixty words, considerably more than the seventeen words between V4,2626 and the final coda. The reason for this earlier onset becomes clear when we examine the pivotal passage in question. The upturn after V7,1202 begins with the word *metallic* in the excerpt below, immediately after Hawkshaw runs out of the barber shop, trying to stop McLendon and his posse of vigilantes:

"Come on," McLendon said. . . . The screen door crashed behind them reverberant in the dead air.

The barber . . . ran to the rear "I'll be back as soon as I can," he said to the other barbers. "I cant let--" He went out, running.

The two other barbers followed him to the door and caught it on the rebound, leaning out and looking up the street after him. The air was flat and dead. It had a [UPTURN after V7,1202] metallic taste at the base of the tongue.

"What can he do?" the first said. The second one was saying "Jees Christ, Jees Christ" under his breath. "I'd just as lief be Will Mayes as Hawk, if he gets McLendon riled."

"Jees Christ, Jees Christ," the second whispered.

"You reckon he really done it to her?" the first said.

II

[Minnie Cooper's biography.]

III

The barber went swiftly up the street where the sparse lights, insect-swirled, glared in rigid and violent suspension in the lifeless air. [6, p. 173-75]

Structurally, the major turning point in the first part of the Hawkshaw-Mayes-McLendon story might be said to occur with the change in setting, when Hawkshaw leaves his barber shop, rather than with the beginning of section three, where the story returns to Hawkshaw after the brief dialogue between the other barbers. In effect, the end of section one leaves the reader hanging--wondering what will become of Hawkshaw on his mission of mercy (and justice)--and section three picks up this story where it left off. Insofar as this analysis is correct, the VMP signals the major turning point in the narrative more accurately than Faulkner's own boundary between sections does.

As the VMPs in figures 3 and 4 illustrate, "Dry September" includes many significant turning points within, as well as between, sections of the story. To illustrate that subsidiary peaks on the VMP tend to correlate closely with subdivisions in the narrative, I will cite just one example, the peak at P18,1968 in figure 4, which is flanked by two clearcut valleys. The onset of this peak begins after V4,1917 with the word *sick* in the passage below. After this valley, the VMP rises sharply to the peak at P18,1968 (with the word *dim*), then falls to another valley at V4,2058 (before the word *struck*). This sequence of valley-peak-valley coincides almost exactly with Will Mayes's first passage of dialogue in the story. After this dialogue ends, the VMP begins to rise again, signaling a new narrative constituent: the beating of Will Mayes.

They dragged the Negro to the car. The barber had waited beside the car. He could feel himself sweating and he

knew he was going to be [UPTURN after V4,1917] sick at the stomach.

"What is it, captains?" the Negro said. "I aint done nothing. 'Fore God, Mr John." Someone produced handcuffs. They worked busily about the Negro as though he were a post, quiet, intent, getting in one another's way. He submitted to the handcuffs, looking swiftly and constantly from dim [DOWNTURN after P18,1968] face to dim face. "Who's here, captains?" he said, leaning to peer into the faces until they could feel his breath and smell his sweaty reek. He spoke a name or two. "What you all say I done, Mr John?"

McLendon jerked the car door open. "Get in!" he said.

The Negro did not move. "What you all going to do with me, Mr John? I aint done nothing. White folks, captains, I aint done nothing: I swear 'fore God." He called another name.

"Get in!" McLendon said. He [UPTURN after V4,2058] struck the Negro. The others expelled their breath in a dry hissing and struck him with random blows . . . [6, p. 177-78]

In this case, then, the VMP correlates almost exactly with a clearcut narrative constituent.

VMPs have obvious limitations—for example, they provide no mechanism to represent narrative movement transformations or gapped constituency—but, in general, VMPs are remarkably accurate graphical analogues for the surface structure of narrative. They signal the complex ebb and flow of discourse more sensitively and more accurately than traditional graphical analogues such as Freytag's Pyramid can do. As such, VMPs are likely to prove useful tools in discourse analysis, especially since they can be generated automatically by computer.

From a statistical point of view, the VMP is simply a moving average, and, as such, it can be subjected to the usual statistical analyses—calculations of the mean, the standard deviation, and so on. Especially tantalizing are the cyclic alternations between peaks and valleys, which look almost periodic on many VMPs. However, autocorrelation statistics for the VMPs for "Emily" and "Dry September" provide no evidence of periodicity.⁵ In the future, other statistical analyses of VMPs may prove fruitful, but, for now, VMPs appear to be useful primarily as tools for analyzing constituent structure.

Notes

¹ In [1] I used intervals of thirty-five rather than fifty-five words. Shorter intervals work better for examining shorter constituents of discourse such as sentences and paragraphs; longer intervals, which result in smoother VMPs, work better for longer constituents such as narrative episodes.

² More precisely, I plotted points only for intervals of text that end with a new vocabulary word. This reduces the number of points that need to be plotted by about two-thirds while still signaling all turning points in the curve. The program is written in Turbo Pascal for the Macintosh SE. Points are plotted using the SYSTAT statistical package. Further information about the program may be obtained by writing the author.

³ The literature on information flow is extensive. See, for example, [4]-[7].

⁴ Part of this analysis of "A Rose for Emily" appeared first in [10].

⁵ The computer program that generated the VMPs in figures 1 and 3 calculates statistics only for intervals that end with a new vocabulary word, omitting roughly two-thirds of data points. (See note 2.) In order to compute meaningful autocorrelations, an additional program was required to restore these intervening data points. Autocorrelations for "Eveline" did show modest evidence of periodicity, but no such evidence was found for "Emily" or "Dry September."

REFERENCES

1. Gilbert Youmans, A New Tool for Discourse Analysis: the Vocabulary-Management Profile, Language, 67, pp. 763-89, 1991.
2. Daniel W. Olson, A Comparison of Thematic Paragraph Analysis and Vocabulary Management Profiles for an Oral Discourse, unpublished Master's thesis, University of North Dakota, Grand Forks, ND, 1992.
3. William Faulkner, Collected Stories of William Faulkner, Random House, New York, 1950.
4. Wallace L. Chafe, Language and Consciousness, Language, 50, pp. 111-33, 1974.
5. Wallace L. Chafe, Cognitive Constraints on Information Flow, in Coherence and Grounding in Discourse, Russell S. Tomlin (ed.), John Benjamins, Philadelphia, 1987.
6. M. A. K. Halliday and Ruqaiya Hasan, Cohesion in English, Longman, London, 1976.
7. Ellen F. Prince, Toward a Taxonomy of Given-New Information, in Radical Pragmatics, Peter Cole (ed.), Academic Press, New York, 1981.
8. Fred Lerdahl and Ray Jackendoff, A Generative Theory of Tonal Music, MIT Press, Cambridge, MA, 1983.
9. Ray Jackendoff, A Comparison of Rhythmic Structures in Music and Language, in Phonetics and Phonology: Rhythm and Meter, Paul Kiparsky and Gilbert Youmans (eds.), Academic Press, San Diego, pp. 15-44, 1989.
10. Gilbert Youmans, The Vocabulary-Management Profile in Narrative, in Proceedings of the 1992 Mid-America Linguistics Conference and Conference on Siouan-Caddoan Languages, Evan Smith and Flore Zéphir (eds.), the University of Missouri, Columbia, MO, 1993, pp. 129-36.