

Elements of Style in Science Fiction: Andre Norton Compared with Others

JUDITH E. BOSS

■ Although statistical studies of style are usually used to solve attribution problems, to identify the particular characteristics of an individual author's style, or to gather information about language generally, such a study can also be used to determine a work's genre. Sentence length, sentence complexity, percentage of dialogue, and vocabulary differ from genre to genre as well as from individual to individual, and these distinctions may help to resolve a dispute over Andre Norton's science fiction. Though she won the 1977 Gandalf Award for lifetime achievement in fantasy and the 1984 Nebula Grand Master Award for lifetime achievement in science fiction, her work has been neglected by science fiction critics and remains virtually unexamined.¹

Among the reasons for this neglect is the dismissal of Norton's work as juvenile science fiction.² While thematic content and characters' ages often determine whether a work is juvenile, the work's style is another determinant. Commonly, critics and reviewers judge stylistic factors subjectively, following their particular biases. Some elements of style, however, can be measured statistically and thus objectively. Statistics on sentence length, sentence complexity, and vocabulary diversity can provide an objective measure of whether Norton's work is, in these respects, more juvenile than other so-called mature science fiction.

She has written juvenile novels, of course, and has received awards for them. Robert Heinlein and Anne McCaffrey also have written juvenile work but have not been stigmatized as Norton has. Norton has said that she experiences more prejudice against her as an author of juvenile fiction than as a female writer in a male-dominated field. As an example of the prejudice against her, Al Jackson's statement sounds blatant: "Ninety percent of her novels have been sold as juveniles and one should keep this in mind. One should tone down

his expectations." True, her publishers' decision to market much of her science fiction under the "juvenile" rubric undoubtedly misled many critics. Norton's own statement that she has thought of herself as "writing for young people" adds to the confusion. Nevertheless, neither her publisher's decision nor her own statement decisively marks her work as merely juvenile. Donald Wollheim demonstrated the marketing label's arbitrariness when he published *Daybreak 2250 A.D.* under a new title without the juvenile label and found that it sold just as well as any other science fiction. Both adults and younger readers continue to read Norton's books, but in a genre traditionally popular among young people, it is odd that Norton should be singled out as juvenile.³

Such dogmatism as Jackson's, coupled with dismissal of her work as naive or lacking seriousness invites an examination of her work for evidence for these assessments.⁴ Let us assume that (1) juvenile literature's syntax and vocabulary are more restricted than those of adult literature, and (2) that some basic aspects of style can be analyzed objectively through counting sentences, words, syllables, and syntactic elements in a text sample and describing them statistically. A comparison, then, of certain features in samples of Norton's work with the same features in samples from other science fiction should reveal whether her work is, in these respects, more juvenile than that of other science fiction authors. Because statistics deal with probability rather than certainty, the details of that comparison provide no definitive resolution. Nevertheless, we shall see that they certainly suggest that Norton has encountered more prejudice than objectivity and that her work deserves more thorough examination.

As samples of Norton's work I took her first two published science fiction stories, "The People of the Crater" (1947), which was later reprinted as "Garin of Tav," and "The Gifts of Asti" (1948), and two later stories, "Legacy from Sorn Fen" (1972) and "The Toads of Grimmerdale" (1973).⁵ The decision to work with short stories only was intentional, as was the choice of these four stories; I wanted samples separated by enough time to indicate whether Norton's style had changed in any significant way over time. "Garin" and "Toads" are each divided into five chapters, and by using each chapter as a separate sample, I obtained from these four stories twelve samples of her writing, six early ones and six about twenty-five years later.

For comparison, I used three samples from *Isaac Asimov Presents the Great SF Stories 9* (1947) and another three samples from *Isaac Asimov Presents the Great SF Stories 10* (1948).⁶ A random number generator provided several page numbers for each volume at which a sample might begin, and, with two exceptions, I began the sample at a full paragraph beginning on that page and continued it sequentially to the end of the full paragraph containing the one-hundredth sentence from the sample's origin. The sample size of approximately one hundred sentences was used because, generally, one thousand words is the minimum text size for valid statistical determinations

about words, and one hundred sentences is the minimum text size for determinations about sentences. The first exception to purely random sampling was to limit each sample to a different story and a different author; this provided heterogeneity in the comparison samples rather than some weight toward a particular author or a particular kind of story. The second exception was to exclude any sample with fifty percent or more dialogue. I did this both because dialogue tends to have shorter sentences and a smaller vocabulary and because dialogue is a conscious attempt to imitate someone else's oral style, which thus distorts computations of sentence length and vocabulary diversity. Since I made no attempt to choose Norton's best work, whereas editors Asimov and Greenberg preselected the *Great SF* samples for their greatness, these choices probably bias the comparison somewhat against Norton. The inclusion of "Garin of Tav," her earliest and most imitative science fiction work, also tells against her. In addition to the *Great SF* control samples, where the data permit I have used random samples of science fiction from three short stories and three novels published in 1961. The samples, totaling about twelve thousand words, were collected and analyzed as part of the massive stylistic study of various genres by Kucera and Francis.⁷

In counting, I followed the word and sentence definitions used by Kucera and Francis (p. 369). A word is "a string of alphanumeric characters with space on either side; may include hyphens and apostrophes but no other punctuation marks." Thus, "uh" or "by-ways" were counted as single words just as "today" was. I did make an exception, however, of contractions: for all the samples, contractions were expanded into two words. A sentence is a string of words ended by a period, question mark, exclamation point, ellipses, or a long dash, and followed by a capital letter. Therefore, "Go! he shouted." is only one sentence, but "Go! He shouted." is two sentences. In addition, "the end of a paragraph was automatically considered the end of a sentence, no matter what punctuation marks appeared" (p. 369). After a computer count of sentences and words in each of the samples, I hand-counted syllables and then verified each count. I then had the necessary raw data, shown in table 1. A few simple calculations to obtain the average length of sentences in words and the average length of words in syllables converted the figures in table 1 into something meaningful.

TABLE 1.
Total Number of Sentences, Words, and Syllables

Sample	Sentences	Words	Syllables
Norton	3,314	45,128	58,122
<i>Great SF</i>	617	8,273	12,193
1961	899	11,633	—

Table 2 indicates that Norton's style may be as mature and sophisticated as the style of such writers as Poul Anderson, H. Beam Piper, Murray Leinster, or William Tenn. Readability specialists have long recognized sentence length as a statistically valid measure of sentence complexity, and Norton's sentences are longer than those of the *Great SF* samples, although the standard deviation indicates that the difference is insignificant. Variety, too, is commonly thought a sign of stylistic skill, and the range shows that Norton has produced a greater variety of sentence lengths than exists in the other samples. The striking feature in table 2 is the significant difference between word length in the Norton samples and the *Great SF* samples. Norton's monosyllabism is in the same extreme range as Ernest Hemingway's.⁸ Such a degree of monosyllabism is unusual, difficult to achieve, and could indicate either a severely restricted vocabulary or a stringently controlled one.

TABLE 2.
Average Sentence Length and Word Length,
with Range and Standard Deviation

Sample	Avg. SL	Range	S.D.	Avg. WL	Range	S.D.
Norton	13.62	1-63	1.85	1.29	1-6	0.04
<i>Great SF</i>	13.41	1-54	1.08	1.47	1-7	0.03
1961	12.94	—	—	—	—	—

Francis and Kucera indicate, however, that sentence length is an inadequate measure of complexity; short sentences may in fact be more complex than longer ones (pp. 550-51). More direct measures of syntactic complexity and vocabulary diversity are necessary. Again I counted, this time the number of sentences whose structures were simple, compound, complex, or compound-complex. While this traditional grammar may make linguists wince, the grammar used has less significance than its consistency across samples (Cluett, p. 22). A calculation of the percentage of total sentences whose structures were complex or compound-complex brought surprising results, as shown in table 3: the Norton samples contained a greater percentage of such sentence structures than the *Great SF* samples, confirming the suggestion made by sentence length data. Because the standard deviations mean that the difference is not significant, one can conclude only that, by this measure, the Norton samples are at least as great as the *Great SF* samples.

TABLE 3.
Percentage of Complex and Compound-Complex Sentences
and Standard Deviations

Sample	Percentage of Complex and Compound-Complex Sentences	S.D.
Norton	36.93	5.93
<i>Great SF</i>	30.79	6.67

Two other more linguistically sophisticated measures confirm that conclusion. The number of predictions per sentence, a measure of "total complexity" used by Francis and Kucera (pp. 550-51) and based upon the number of both finite and nonfinite verbs in each sentence, again shows that the Norton samples are slightly more complex than the *Great SF* samples. So does Cluett's "density of rank shifting" (p. 143), which uses the total number of subordinators, relatives, and verbals per sentence as a measure of total complexity. These measures attempt to quantify apposition and subordination—or embedding, interruption, and suspension—as indicators of sentence complexity. Once again, as shown in table 4, the differences between the Norton and the *Great SF* samples are insignificant, although both are significantly higher in total complexity than Kucera and Francis' randomly chosen 1961 samples.

TABLE 4.
Total Complexity and Standard Deviations

<i>Sample</i>	<i>Predictions per Sentence</i>	<i>S.D.</i>	<i>Density of Rank Shifting</i>	<i>S.D.</i>
Norton	3.07	0.49	1.08	0.25
<i>Great SF</i>	2.75	0.22	0.97	0.12
1961	2.23	—	—	—

Of the three measures above which deal with sentence complexity (tables 2-4), none reveal any significant difference between the Norton samples and the *Great SF* samples. Because of the bias against Norton built in by my differing methods of selection, we would expect her work to suffer in the comparison, yet the data rank her an equal among the *Great SF* writers. More specific data might reveal significant differences among all three sample groups, or between the first two and the random 1961 samples, which uniformly rank lower.

Two further measures of sentence style in the samples are as revealing as the previous ones. Variety of sentence length and vocabulary appear more commonly in a mature style than in a juvenile one. When Kucera and Francis analyzed the Brown University Corpus, they determined "the variety of different sentence lengths occurring in the various" samples using Yule's *K* Factor as an "indicator of the range of grammatical stylistic devices allowed within the confines of a genre pattern." "A low *K*-factor value indicates a low rate of repetition of a sentence length and thus a greater diversity of sentence lengths present" (pp. 372-75), or a more mature style. To calculate the *K* Factor, one must prepare a frequency distribution table showing in one column the sentence lengths (*X*) and in another column the number of times each length occurs (*f*). Each *X*-value is multiplied by the corresponding *f*-value, and the results are added together to form S_1 . Multiplying each *f*-value by its corresponding *X*-value-squared and adding the results forms S_2 . Yule's formula for the *K* factor is: $K = 10,000 \times (S_2 - S_1^2)/(S_1)^2$ with the number

10,000 used to avoid very small decimals.⁹ The *K* Factor is expressed as a simple number, such as 10.2 or 131.36.

Kucera and Francis also ranked the samples by the "amount of homogeneity present" using the coefficient of variation to measure how closely the sentence lengths clustered around the mean sentence length for the sample, or how widely they dispersed from the mean (pp. 372-75). To calculate the coefficient of variation, they determined the mean sentence length for each sample, a mean of the means, and a standard deviation; the coefficient of variation they derived "from the formula standard deviation divided by the mean." A low coefficient of variation indicates a high degree of homogeneity.

In table 5, each of the twelve Norton samples shows a greater diversity of sentence lengths than any of the six *Great SF* samples, which in turn show greater diversity than the six 1961 samples, although as shown by the high coefficient of variation, the least homogeneous group is 1961, followed by Norton, then by *Great SF* as the most homogeneous.

TABLE 5.
Range, Mean, and Standard Deviation of *K* Factor for Sentences, and Coefficient of Variation for Sentences

<i>Sample</i>	<i>Range</i>	<i>K Factor Mean</i>	<i>S.D.</i>	<i>Coefficient of Variation</i>
Norton	35.22-61.81	47.01	7.74	0.1364
<i>Great SF</i>	120.23-147.13	132.67	8.32	0.0805
1961	—	368.08	—	0.1738

The significance of the diversity and homogeneity shown by these figures lies in the greater variety shown by Norton as a single author, from whom one would expect a certain limitation of self-consistency. Norton, in other words, shows a greater diversity in her sentence lengths than either group of authors. If it is true that "whatever visible changes a writer is able to make in his style, he is likely, on balance, to write more like himself than like any 13 contemporaries" (Cluett, p. 269), then diversity is inherent in Norton's prose. The greater homogeneity of the *Great SF* samples, as shown by the low coefficient of variation, may be caused by editors Asimov and Greenberg's preselection on the basis of some conscious or unconscious choice, but it is more likely to be a product of the respective authors' close adherence to the conventions of the genre (Kucera and Francis, p. 375). Since I deviated from purely random selection by insisting on different authors and different stories to achieve heterogeneity in these samples, their homogeneity in sentence length is anomalous except as a genre convention.

Vocabulary diversity is another accurate test of sophistication or complexity, and a necessary test for this comparison because Norton's monosyllabic prose suggests a restricted vocabulary. The Type-Token Ratio (TTR), the number

of different words, or types, compared to the total number of words, or tokens, in a sample measures such diversity. The TTR is dependent upon sample size, however, because the total number of words in a given text increases much faster than the number of different words; the longer the text, the more vocabulary repetition. To eliminate this difficulty, Yule devised the *K* Factor—which Kucera and Francis used to measure sentence diversity—as a measure of vocabulary diversity that would be independent of sample size.¹⁰ The *K* Factor actually measures the repeat rate of types and hence measures their diversity. As with sentences in table 5, a low *K* Factor for vocabulary means a lower repeat rate of words, that is, greater diversity of vocabulary. Although the Norton samples do have a slightly higher repeat rate for vocabulary than the *Great SF* samples, as the *K* Factor in table 6 reveals, the difference is not significant. Further, the coefficient of variation indicates once more that Norton's work is less homogeneous, that is, shows a greater variety of repeat rates among the samples than the *Great SF* does. The standard deviation and the range confirm the greater variety in the Norton samples.

TABLE 6.
Range, Mean, and Standard Deviation of *K* Factor for
Vocabulary, and Coefficient of Variation for Vocabulary

<i>Sample</i>	<i>Range</i>	<i>K Factor</i> <i>Mean</i>	<i>S.D.</i>	<i>Coefficient</i> <i>of Variation</i>
Norton	84.36–156.11	110.63	17.89	0.1617
<i>Great SF</i>	90.69–133.48	108.02	12.82	0.1187

The statistical results summarized in tables 1–6, in fact, show that these samples of Norton's writing contain diversities of sentence length, sentence complexity, and vocabulary equivalent to those in the *Great SF* samples and greater than those (where measured) in the 1961 samples. Norton's unique feature, among those elements measured, is her monosyllabism. In these samples 76.17 percent of the words are monosyllables compared to 68.11 percent for the *Great SF* samples. Since the norm for English is 50 percent one-syllable words,¹¹ science fiction as a genre may deviate from the norm, and Norton may be an extreme variant of this deviation. Only a great deal more data and analysis can resolve the question. Monosyllabism, however, broadens Norton's audience, making her writing accessible to younger readers without otherwise affecting it. Further, the Norton samples exceed the others in range and variety suggesting greater flexibility in her writing style and less rigid adherence to the genre's conventions. Such variations and flexibility usually count as indicators of increased maturity in style.

Two of the most common readability measures show the positive effect of Norton's monosyllabism: the Flesch Reading Ease formula and the Fog Index which test a sample's accessibility to younger readers. Readability tests attempt

to measure sentence difficulty by sentence length and vocabulary difficulty by the number of hard words or the number of syllables per word or the number of words three syllables or more in length. The formulas for the Flesch Reading Ease test and the Fog Index are: Flesch, $R = 206.835 - 84.6 \times S/W - 1.015 \times W/T$; Fog, $G = 3.068 + 9.84 \times P/W + .0877 \times W/T$; where W represents the total number of words, T is the total number of sentences, S equals total number of syllables, P is total number of words with three syllables or more (McCallum and Peterson, pp. 45-46).

The Reading Ease score ranges from 0 (very hard) to 100 (very easy), and the Fog score provides a grade level for which the text is thought to be appropriate. The results in table 7 indicate that Norton's writing, although equally complex in syntax and diverse in vocabulary, as shown in tables 3-6, is accessible to readers considerably younger or less skilled than the readers of *Great SF*. The Reading Ease scores differ from "easy" to read and understand for Norton samples to "standard" for the *Great SF* samples, and the Fog Index differential is nearly a full grade level. Norton has increased her audience by a careful choice of words, rather than a total restriction of vocabulary, another indicator of the skillful stylist.

TABLE 7.
Readability Scores and Standard Deviations

Sample	Flesch		Fog	
	Reading Ease	S.D.	Grade Level	S.D.
Norton	83.75	3.26	4.67	0.18
<i>Great SF</i>	68.68	3.82	5.30	0.20

A linear regression shows that this increased readability results from her monosyllabism. The word difficulty component has a high correlation for the Norton samples ($r = 0.82$), whereas sentence difficulty shows a much lower correlation ($r = 0.34$). In contrast, the word difficulty component for the *Great SF* samples shows an even higher correlation ($r = 0.96$), but so does the sentence difficulty component ($r = 0.75$). I interpret this to mean that Norton's monosyllabism is so abnormally great that it outweighs the sentence difficulty component almost entirely in measuring the readability of her prose.

Educators and publishers rely heavily on readability scores for determining the suitability of texts for given age groups, and they have created in the minds of others the misconception that what is suitable for, say, fifth-grade students would be too simplistic for college readers. This is an oversimplification, at best. McCallum and Peterson applied ten of the most commonly used readability tests to their article "Computer-Based Readability Indexes," an article addressed to educators and computer scientists, and the ten scores for the same article range from grade level 6.5 to grade level 13.6, from "easy" to "hard" (p. 47). Thus, while the readability scores for the Norton and the *Great SF*

samples tell us reliably that Norton's prose is easier to read than the *Great SF*, they do not provide a reliable index to grade level or age group for either of the samples.

The remaining question regarding Norton's work is whether her style has matured or improved in some measurable way over the years. Comparing the 1947-48 stories to the 1972-73 stories, using the same elements summarized in tables 2-7, reveals a change, and significance testing indicates that the results in table 8 are statistically significant at the 99 percent confidence level.

TABLE 8.
Norton 1947-48 Stories Compared to Norton 1972-73 Stories

Measure	1947-48	Norton Samples		S.D.
		S.D.	1972-73	
Sentence Length	12.70	1.81	14.43	1.43
Word Length	1.32	0.02	1.26	0.02
Complex and Compound- Complex Sentences	32.54%	5.34	40.65%	2.98
Predications per Sentence	2.90	0.60	3.24	0.24
Density of Rank Shifting	0.91	0.17	1.24	0.21
K Factor, Sentences	48.28	8.83	45.74	6.21
Coefficient of Variation, Sentences	0.1425	—	0.0991	—
K Factor, Vocabulary	123.45	15.73	97.81	8.00
Coefficient of Variation, Vocabulary	0.1274	—	0.0818	—
Flesch Reading Ease	82.17	2.56	85.33	3.11
Fog Index	4.69	0.21	4.72	0.16

The two later stories (1972-73) show an increase in sentence length and an increase in all three measures of sentence complexity (percentage of complex and compound-complex sentences, number of predications per sentence, and density of rank shifting). They also show a reduction in the repeat rate for sentence lengths and for vocabulary (the lower *K* Factor). Norton's style in the later samples has become more complex, sophisticated, and varied. In fact, comparison with tables 1-7 reveals that Norton's later stories rank higher in complexity than the *Great SF* samples. As expected, however, the lower coefficient of variation shows that the homogeneity of sentence lengths and vocabulary has increased. Nevertheless, as the *K* Factor and coefficient of variation indicate, Norton's later stories contain a greater variety of sentence lengths than the six *Great SF* samples, although the comparable variety in vocabulary has diminished, as one would expect in a sample of only two stories rather than four. The surprise is that word length has decreased and readability thus has remained stable. Such monosyllabism is rare in conjunction with syntactic complexity and a large vocabulary.

This study cannot determine definitively and conclusively whether Norton's

work is more or less juvenile than other science fiction except in matters of syntax and vocabulary. Nevertheless, its results suggest that (1) Andre Norton's science fiction and fantasy exhibit as much stylistic skill and sophistication as most great science fiction; (2) she has been thoroughly professional in maturing those skills over her writing career; (3) she has, in addition, a rare talent for writing clear, predominantly monosyllabic prose accessible to almost any reader, which talent she has also developed over the years of her career. She follows her own dictum that "under-stating rather than over-stating . . . is the better course in writing" (Walker, p. 265). Because of her ability to write prose mature enough for the adult audience yet simple enough for the juvenile audience, she continues to outsell her "more acclaimed competitors" (Sparks, p. 56). This study's results indicate that, at least in the matters addressed here, Norton's work has met prejudice rather than honest evaluation, and a further examination of her work would be appropriate.

These results, and those of Francis and Kucera, suggest also that a further examination of science fiction works generally would be appropriate. Francis and Kucera sampled six genres of imaginative prose: general fiction, mystery and detective, science fiction, adventure and western, romance and love story, and humor (p. 552). In these 1961 samples, only the mystery and detective and the adventure and western genres had shorter sentences than the science fiction samples, and the science fiction samples had the lowest number of predications per sentence of all the genres, that is, the least complexity of syntax. However, the Norton and *Great SF* samples show a higher number of predications per sentence than any of the 1961 imaginative prose genres except humor. Francis and Kucera also discovered that the percentage of quoted material ranged from a low of 14.9 percent in general fiction to a high of 26.8 percent in science fiction (p. 549), a factor that explains, in part, the comparative simplicity of science fiction prose in their study. In their analysis, Kucera and Francis found that their science fiction samples possessed less diversity in sentence length than all but two of the other genres, a fact they linked to science fiction's pulp origins (pp. 378-79). The Norton and *Great SF* samples, though, revealed greater diversity in sentence length than any of the 1961 imaginative prose genres.

To date, the massive collection and analysis of data necessary to characterize science fiction's stylistic and syntactic conventions, either in themselves or in comparison to those of other genres, does not exist. Supporters claim that science fiction is real literature despite its pulp origins, despite their inability to define science fiction in any meaningful way, and despite their inability to identify with much accuracy those science fiction conventions of which we speak so glibly. Definitions, by their nature, limit and are inconsistent with science fiction's youthful vitality and virtually boundless eclecticism. Perhaps, as with other genres, only a few centuries of discriminating writers, publishers, readers, and critics can bring to science fiction the maturity associated with

mainstream fiction. But we could do worse than to examine a collection of the best science fiction, analyze its characteristics, and watch for patterns to emerge. This study is a small-scale effort to apply that dictum to an evaluation of Norton's work.

Notes

1. Henry Kucera and W. Nelson Francis, *Computational Analysis of Present-Day American English* (Providence, R.I.: Brown Univ. Press, 1967); Sandra Miesel, *Introduct., Witch World*, by Andre Norton (Boston: Gregg Press, 1977), p. v.
2. Donald Wollheim, *Introduct., The Book of Andre Norton*, by Andre Norton (New York: DAW Books, 1974), p. 7.
3. Paul Walker, "Andre Norton," in *Speaking of Science Fiction* (Oradell, N.J.: Luna Press, 1978), p. 269; Rev. of *Android at Arms*, by Andre Norton, *Luna Monthly* (Oct. - Nov. 1972), p. 53; Brian Aldiss, *The Billion Year Spree* (New York: Schocken Books, 1973), p. 263; Lester del Rey, *The World of Science Fiction, 1926-76* (New York: Ballantine Books, 1979), p. 204; John Nicholls, ed., *The Science Fiction Encyclopedia* (Garden City, N.Y.: Doubleday, 1979), pp. 206, 537, 661.
4. Miesel, pp. vi - vii; Elisa Kay Sparks, "Norton, Andre," in *Dictionary of Literary Biography Vol. 8: Twentieth Century American Science-Fiction Writers. Part 2: M-Z*, ed. David Cowart and Thomas L. Wymer (Detroit: Gale Research, 1981), p. 56.
5. Andre Norton, *Garan the Eternal* (New York: DAW Books, 1972), pp. 9-50; *The Book of Andre Norton* (New York: DAW Books, 1974), pp. 11-57, 110-23; and *Lore of the Witch World* (New York: DAW Books, 1980), pp. 111-20.
6. Poul Anderson, "Tomorrow's Children"; H. Beam Piper, "Time and Time Again"; Eric Frank Russell, "Hobbyist"; in *Isaac Asimov Presents the Great SF Stories 9 (1947)*, ed. Isaac Asimov and Martin Greenberg (New York: DAW Books, 1983), pp. 42-70, 101-18, 294-326; Murray Leinster, "The Strange Case of John Kingman"; William Tenn, "Brooklyn Project"; A. E. Van Vogt, "Dormant"; in *Isaac Asimov Presents the Great SF Stories 10 (1948)*, ed. Isaac Asimov and Martin Greenberg (New York: DAW Books, 1983), pp. 50-65, 139-48, 179-94.
7. Susan Hockey, "Stylistic Analysis and Authorship Studies," in *A Guide to Computer Applications in the Humanities*, ed. Susan Hockey (Baltimore: Johns Hopkins Univ. Press, 1980), p. 126; Robert Cluett, *Prose Style and Critical Reading* (New York: Teachers College Press, Columbia Univ., 1976), pp. 277-80; W. Nelson Francis and Henry Kucera, *Frequency Analysis of English Usage: Lexicon and Grammar* (Boston: Houghton Mifflin, 1982), continues their study of the 1961 Brown corpus.
8. Douglas R. McCallum and James L. Peterson, "Computer-Based Readability Indexes," in *Proceedings of the ACM '82 Conference, October 22-27, 1982* (Dallas: ACM, 1982), p. 45; Darrell Mansell, "The Old Man and the Sea and the Computer," *Computers and the Humanities*, 8 (1974), 199.
9. Paul E. Bennett, "The Statistical Measurement of a Stylistic Trait in *Julius Caesar* and *As You Like It*," in *Statistics and Style*, ed. Lubomir Dolezel and Richard W. Bailey (New York: American Elsevier, 1969), pp. 29-41.
10. G. Udny Yule, *The Statistical Study of Literary Vocabulary* (Cambridge: Cambridge Univ. Press, 1944; rpt. New York: Archon Books, 1968); see also Gustav Herdan, *The Calculus of Linguistic Observations* (The Hague: Mouton, 1962), pp. 36-40.
11. Zipf, G. K., *The Psycho-Biology of Language* (Boston: Houghton Mifflin, 1935), pp. 23-28.