

ELEMENTS OF STYLE IN SCIENCE FICTION

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.

Dogmatism coupled with dismissal of her work as naive or lacking seriousness invites an examination of her work for evidence for these assessments [4]. We 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 sentence 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.

Norton's first two published science fiction stories were taken as samples of her work, "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).

For comparison, 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) were used [6]. A random number generator provided several page numbers for each volume at which a sample might begin, and with two exceptions, we 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 approximately one hundred sentences was used because, generally, 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. Since no attempt to choose Norton's best work was made, whereas editors Asimov and Greenberg preselected the *Great SF* samples for their greatness, these choices probably bias the comparison some-

what 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 we 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, the word and sentence definitions used by Kucera and Francis was followed [7, 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. An exception of contractions, however, was made: 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" [7, 369]. All the necessary raw data, shown in table 1, described the figures.

Table 1.

Total Number of Sentences, Words, and Syllables

Sample	Sentences	Words	Syllables
Norton	3,314	45,128	58,122
Great SF	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
Great SF	13.41	1-54	1.08	1.47	1-7	0.03
1961	12.94	-	-	-	-	-

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 by sentence length data.

Table 3.

Percentage of Complex and Compound-Complex Sentences and Standard Deviations

Sample	Percentage of Complex and Compound-Complex Sentences	Standard Deviations
Norton	36,93	5,93
Great SF	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 [7, 550-551] and based upon the number of both finite/nonfinite verbs in each sentence, again shows that the Norton samples slightly more complex than the *Great SF* samples. So does Cluett's "density rank shifting" [7, 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 Deviation

Sample	Predictions per sentence	S.D.	Density of Rank Shifting	S.D.
Norton	3.07	0.49	1.08	0.25
Great SF	2.75	0.22	0.97	0.12
1961	2.23	-	-	-

Of the three measures above, which deal with sentence complexity (tables 1-4), none reveal any significant difference between the Norton samples and the *Great SF* samples.

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 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. 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 5 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 5.

Range, Mean, and Standard Deviation of K Factor for Vocabulary, and Coefficient of Variation for Vocabulary

Sample	Range	K Factor Mean	S.D.	Coefficient of Variation
Norton	84,36-156.11	110,63	17.89	0.1617
Great SF	90,69-133.48	108.02	12.82	0.1187

The statistical results summarized in tables 1-5, 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 monosyllables. 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 out otherwise affecting it. Further, the Norton samples exceed the others in range and variety suggesting greater flexibility in her writing style and less adherence to the genre's conventions. Such variations and flexibility usually count as indicators of increased maturity in style.

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 maturing those skills over her writing career; (3) she has, in addition, a talent for writing clear, predominantly monosyllabic prose accessible to any reader.

Works Cited

1. Henry Kucera and W. Nelson Francis, *Computational Analysis of Present-Day American English* (Providence, R.I.: Brown Univ. Press, 1967); Sandra Miesel, *Introd., Witch World*, by Andre Norton (Boston: Gregg Press, 1977), p. v.
2. Donald Wollheim, *Introd., 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, WS), 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. 111-120.
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, 101-118, 294-326; Murray Leinster, "The Strange Case of John Kingman"; William Tenn, "Brooklyn iip reject"; 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, 139-148, 179-194.
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. 277-280; 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 Guslav 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