

The Cross Plainsman

A Journal of Robert E. Howard Studies by Frank Coffman
2002 / no. 2 / April
A Member Journal of REHupa / for Official Mailing 174

Stylometric and Thematic Approaches and Initial Hypotheses for Application to the Literature of Robert E. Howard

The study usually called “Stylometrics” and the methods of Stylometry evolved from the now-validated assumption that writers create distinctive patterns of language (vocabulary, grammatical construction, rhetorical device, average sentence length, etc.) as distinctive and unique as fingerprints. Initially, this study was used in the process of attempting to identify the authorship of unattributed or debated texts and to inquire into such scholarly questions as: “Was there a single poet involved in the compositions of *The Iliad* and *The Odyssey*, and, if so, was it the same poet?” [Most stylometrists answered yes to both of these, by the way].

Another example of the use of stylometry in this fashion was the verification (at least the very great likelihood, which is really all what I’ll call “Identification Stylometrics” can offer without other evidence) of a poem discovered with the signature “W. S.” as, indeed, the work of William Shakespeare.

But the kind of stylometry that I’m interested in as it might be applied to the work of Robert E. Howard (or any other known author) would not be primarily used to try to prove that he

wrote the stories attributed to him (except in possible rare cases of found fragments, etc. after major parameters of REH’s style are fully delineated), but, rather, in trying to discover, by means of what I’ll call “Investigative” or “Exploratory” or “Critical Stylometrics” the narrative and poetic methods used and even themes inherent in his literary compositions (which moves into the area that I call “Thematics”). I hope to discover nuances and formulae and some objectifiably demonstrable truths about both the palate and the brush strokes of this artist with words — perhaps something also about the man behind the methods and the meanings.

In this installment of what will be an ongoing and slowly-developing study, my plan is to do little more than present my own somewhat divergent brand of stylometric theory and to offer some initial hypotheses that are almost wholly subjective, based simply upon the normal critical methods of approach used to study the work of REH and also on simple “gut feelings” based upon reading his poems and stories.

First, it should be understood from the onset that I disagree with most of the mechanical

“tagging” and sorting of texts along purely grammatical lines. There is no doubt that KWIC (“Key Word In Context”) concordance sorting of a text and vocabulary study over a single story or over an entire body of work are significant. But vocabulary and use of words in context are, chiefly, “Content-significant.” The Identification Stylometry of analysis of grammar, is, by its very nature, “Form-significant.”

Having said this, I see a great potential in the use of the computer as a critical tool. By using an idiosyncratic brand of Stylometry and some initial premises about the possibility of Themetry, I believe that CALC (“Computer-Assisted Literary Criticism”) is a pathway into profitable discoveries about the methods, meaning, and mind of Robert Ervin Howard.

What needs to be done to any text for this sort of close analysis includes the following:

1) the creation of a KWIC Concordance, inclusive of all words used in their context and inclusive of a word frequency list (how often each word is used);

2) the creation of a Letter Concordance to show the frequency of letter distributions in a text (how many “As,” how many “Bs,” etc.) [While this is an important aspect of Identification Stylometry, the sounds represented by the letters have more content-significance, especially when approached from the perspectives of Psycholinguistics];

3) the division of the text into sentences (prose) or lines (poetry) for close analysis;

4) the objective and subjective “tagging” and indexing [spreadsheet and/or other custom-designed software] of fictional narrative and/or poetic attributes through a close investigation of each sentence or line;

5) the comparison of anything discovered with reference to a particular story against a group (i.e. the Conan stories) or the whole

body of work (i.e. all REH’s prose fiction) to find common denominators (what I call “DNA” [“Distinctive Narrative Attributes”] or, with reference to poetry, the comparison of any given poem to, say, other sonnets, or to the whole body of REH’s poetic output, thus discovering “CPA” [Common Poetic Attributes];

6) the comparison of results with results found from the study of other authors—especially ones who might have been influential upon REH (London, Mundy, Burroughs, Chesterton [poetics], Lovecraft, Smith, etc.), but possibly also on those who are consciously derivative of him (the pastiche writers and other Sword & Sorcery followers, for example)

7) the comparison of results from various stages of REH’s own work to see if and what aspects of his methods and messages changed over time;

8) the more subjective assessment, by pretty much any critical methods already in use or ones yet to emerge, of the messages or themes of the text, the author’s tone as seen through the text, etc.—these too can be analyzed by the imposition of an objective/numerical schema upon the subjective observations (this last point deals with the extension of method into Themetrics).

Examples of Objective Criteria

Among the objective criteria that might be “tagged” or noted for commentary are such numerically definite parameters as:

1) *average number of syllables per word*;

2) *average number of words per sentence*

3) *average number of sentence per paragraph* (differentiated between dialogue and narration);

4) *average number of rhetorical devices per sentence/per paragraph/per story, etc.;*

5) *kinds of and frequencies of rhetorical devices* (similes, metaphors [lots of cats here, I suspect, etc.], alliteration, parallel structures; etc.);

6) *the ratio of Mimesis to Digesis* (Aristotle's words in *The Poetics* for "Scene" [dialogue and real time action] vs. "Summary" [words of narration and summary of events and description, etc.]); a side by side comparison of "Mimetic Density" and "Diegetic Density" as I call them;

7) *the distinctive uses of attribution* ("tagging" in the presentation of dialogue, according to the following distinction into varieties:

a) *unattributed* (where the dialogue alone is given, with no attribution wording whatsoever)—

b) *regular* (use of attribution through the "normal" words: "said," and "asked"—

c) *irregular* (use of attribution through any variant words: "shouted," "queried," etc.—

d) *simple* (the only words breaking up the dialogue or used with the dialogue are the "tag words" to identify the speaker)—

e) *complex* (something else is done along with the dialogue, such as delivery of narrative content ("Follow me!" he shouted, meanwhile cleaving a foeman's skull to the teeth with his scimitar.) or descriptive detail ("This is the Lost Valley of Yazdik lore," Nanok said, as they beheld a luxuriantly green expanse below them, divided by a narrow but rapid river, so different from the barren and bleak wastelands through which they had wandered the past weeks.) or character development ("You must be Nanok," she said, moving, he knew, intentionally between him and the great glowing hearth to reveal the curves of her young, lithe body through the gossamer-woven stuff of her gown.).

Clearly, combinations of the above occur, so that dialogue tag-

ging may be called, for example: "regular complex," "irregular simple," etc.

8) *the frequency of and position of modification* (adjectives and adverbs) in relation to the words modified (nouns, verbs, adjectives, other adverbs) [I've stated earlier in these REHupa musings and elsewhere {REHupa, etc} my belief that REH makes frequent — and, I believe what will turn out to be distinctive — us of what I've called "hypermmodification" in order to compress as much action and descriptive detail as possible into his narratives;

9) *the examination of story openings, closings, and transitions* to see if there are structural formulae at work (something like what Doyle succeeds with in the Sherlock Holmes stories — lots of variant plots built upon the framework of motifs that work again and again). [I believe, for example, that REH uses most often what I'll call the "Action Opening" {as opposed to the "Dialogue Opening" or the "Establishment of Setting Opening" or "Characterization Opening" or other possible varieties. I believe that REH tends to get right into Action and Character. Transitions are often Descriptive. Endings often are Dialogic;

10) *the comparison of results against the great theories already developed regarding mythic and folkloric and traditional material* (Rank, Raglan, Frazer, Campbell, Frye, Todorov, Aarne, Propp, Thompson etc.) to see how much REH makes use of traditional material and how closely his work — instinctively or purposefully — follows the mythic-traditional bases which I believe are the foundations for all popular imaginative literature seen as the extension of the mythic, legendary, folkloric, and traditional into the Age of Print and the Age of Mass Literacy (1440 on and 19th century on, respectively).

There are, of course, other objectifiable items, including sentence types according to various

principles of classification: Simple, Compound, Complex, Compound-Complex; Loose, Cumulative, Periodic, Inverted; Declarative, Interogative, Exclamatory; etc. and other aspects undiscussed in this installment and also unimagined as of yet.

Examples of Subjective Criteria

Moving into the realm of “Themetrics” — as I’m calling attempts at a scientific, statistical study for the purpose of interpretation of literature (as opposed to the study of its style and structure — means moving into the area of “Subjective Tagging” of sentences or lines.

The concept of subjective tagging of a text allows for thematic investigation and interpretation from many/any angles of approach or schools of literary appreciation and criticism. The critic may bring any method or agenda to bear upon the material. The key element and key requirement in any case is the consistency of criteria and the uniform application of the tags.

To give an example (and something I’m working on) of an imposed statistical method upon the “action sentences” of a piece of narrative fiction (the “action” part of Mimesis/Scene as opposed to the “dialogue” part; the what characters are doing as opposed to what they are saying), let us assign numbers in increasing order to the intensity of action depicted:

- 0 = stasis, rest, inactivity, sleep, trance
- 1 = awake and aware but stationary, watching, talking, etc.
- 2 = moving at a relatively slow pace, walking, journeying, traveling
- 3 = moving with haste, chasing, being chased, rushing to battle, etc.
- 4 = actually engaged in battle, fight for survival, etc.
- 5 = slaying, being slain, etc.

While the example numbers are entirely arbitrary and might serve only a particular story or type of story, they can be used, nonetheless to establish frequency and distribution of action throughout a tale. If patterns emerge across stories, the numbers would indicate such. Note that the numbers need not relate to intensity or relative significance (although that makes some sense, when possible). What is important is that they be applied consistently — whatever they represent. When compared against graphs of distribution of occurrence with other stories, consistencies and structural formulae might be perceived.

The same sort of “imposed tagging” and arbitrary designation — so long as it is consistently applied — could be done with almost any aspect of the theme or symbolism of the tale, or when a significant believed-to-be-distinctive trait is observed. Wherever a thematic or possibly symbolic portion of a story or a distinctive auctorial trait might be subjectively discovered or observed by the critic, a number could be assigned to that theme or symbol or trait. After that, what was initially subjective, can be numerically analyzed and — at least to a greater degree than before thought possible — objectified (or at least defended).

Of course the Feminist critic will apply different tags from the Structuralist, and they both will, quite likely, differ in tag choice and numbering system used from the Myth critic, and so on — but the results can be quantified. And that is the very heart of the study of literature via the computer.

Of course all of this requires the examination and often the initial creation of digital texts true to the original print texts. Obvious advantages of the digital text begin with the wonderful ability to search even lengthy texts for key words and verbal constructions.

More next time.

