STATISTICAL DECODING OF CIPHER MOTIFS IN CLASSICAL DETECTIVE PROSE

Bazaralieva Aziza Ruslanovna
Lecturer, Department of practical English
Navoi State University
Navoi, Uzbekistan
Bazaralieva@nspi.uz

Abstract. *This article studies how quantitative procedures guide the decoding of cryptic signs in classical detective prose. The material is drawn from Edgar Allan Poe's «The Gold Bug» and Arthur Conan Doyle's «The Adventure of the Dancing Men». The analysis shows that counting, recurrence tracking and positional comparison do not merely support interpretation. They organize it, reduce uncertainty and convert an obscure sign sequence into a controlled route toward semantic recovery.*

Keywords*. quantitative analysis, cipher, detective prose, frequency, recurrence, sign sequence, decoding, Poe, Doyle, stylometry, literary method.*

Introduction. Ciphered writing enters detective prose not as decoration but as a mechanism that delays access to meaning and compels the reader to watch the procedure of inference itself. In Poe's «The Gold Bug» the cryptogram is materially present on the page as a string of signs that resists immediate reading. In Doyle's «The Adventure of the Dancing Men» the coded message is first perceived as childish drawing, yet it soon becomes the hinge of the whole investigation. In both cases the reader meets not an abstract mystery but a measurable sequence of visible marks.

What makes these texts especially productive for scholarship is that quantitative reasoning is already built into their inner logic. Legrand explicitly

counts symbols and ranks them by frequency, while Holmes reconstructs letter values through repetition, word length and terminal markers. This means that the literary clue is not only solved inside the story but also offers a compact model of how numerical regularities can shape interpretation in prose. That model later becomes methodologically legible in work by Burrows, Hoover, Moretti, Jockers, Evert and Underwood, who treat recurring verbal or graphic features as evidence rather than ornament.

Methods and literature review. The study combines close reading with manual counting of salient signs, comparison of repeated groups, observation of boundary markers, reconstruction of probable letter values and parallel reading of the two stories against work in information theory and quantitative literary studies. The procedure focuses on visible recurrence, local probability, explicit authorial counting and the movement from graphic anomaly to lexical resolution. This design makes it possible to describe not only what the code means in each text but also how the text disciplines the reader into accepting one solution rather than another.

The theoretical background joins several lines of inquiry. C.E.Shannon defines information as selection from a set of possibilities, which is crucial for any account of coded messages. G.K.Zipf links linguistic form to recurrent frequency effects. T.Todorov clarifies the formal organization of detective fiction. J.Burrows and D.L.Hoover show how numerical comparison of frequent elements can support literary inference. F.Moretti, M.L.Jockers and T.Underwood expand the scale of quantitative reading, while S.Evert and coauthors refine stylometric distance measures. The primary literary material is supplied by Poe and Doyle, whose stories present decoding as an observable sequence of operations rather than a sudden intuition.

Results. In Poe the code is not decoded by inspiration but by ordered reduction of uncertainty. Legrand states that my first step was to ascertain the predominant letters and this line is methodologically decisive because it moves

the story from wonder to countable regularity. The cryptogram is first treated as a visible distribution of marks, then as a ranked series. Poe makes this procedure fully explicit when Legrand reports that the sign 8 occurs 33 times, the semicolon 26 times and the sign 4 occurs 19 times. The reader therefore watches a transition from graphic clutter to probability table before any treasure instruction becomes legible.

The next step in Poe is not simple frequency but frequency constrained by language specific expectation. Legrand writes that we will commence by assuming it as the e of the natural alphabet, after which he checks doubled forms and short recurring groups. The procedure is strengthened by another local decision. Because the signature hints at Kidd, the cryptogram is treated as English rather than Spanish or French. Here the code becomes solvable only when numerical dominance, doubling and lexical plausibility are combined. The story shows with unusual clarity that counting alone does not yield meaning, yet without counting the correct lexical path does not even begin.

In Doyle the same logic appears in a more compact and more dramatic form. Holmes first notes that the figures seem like the mere random sketches of children, but he immediately reverses that impression by treating them as a constrained writing system. He reminds Watson and the inspector that he is familiar with many forms of secret writing and explains that the inventor wanted to disguise message bearing signs as harmless play. The crucial turn occurs when recurrence begins to matter. Holmes compares short messages, observes the presence or absence of a flag, infers word boundaries, and then uses repeated patterns to recover the probable word NEVER. From there he derives letter values and extends the same reasoning to ELSIE and COME ELSIE.

Text	Initial coded object	First measurable cue	Main analytic move	Immediate result
Poe, The Gold Bug	String of nonalphabetic signs	explicit symbol count	ranking, double sign check, probable letter assignment	entry into the treasure instruction
Doyle, The Adventure of the Dancing Men	repeated human figures	repeated groups and flag marker	segmentation into words, recovery of recurrent names and responses	identification of sender and target

A broader theoretical frame helps explain why these literary procedures remain persuasive. Shannon writes that the actual message is one selected from a set of possible messages. Zipf, from another direction, formulates behavior as an attempt to minimize the probable average rate of his work expenditure over time. Moretti later generalizes the value of counting for literary study when he argues that a field this large cannot be understood by stitching together separate bits of knowledge about individual cases. Read together, these propositions clarify what Poe and Doyle stage in miniature. A coded text becomes readable when the field of possibilities is narrowed, when the next trial is made cheaper than the previous one and when scattered particulars are converted into a pattern that can be tested. The two stories nevertheless differ in where they place the decisive burden of proof. Poe foregrounds explicit counting and displays the ranked table directly inside the story. Doyle foregrounds serial comparison across several small messages and lets recurrence perform the same work with fewer numbers on the page. Holmes can finally state that the symbols stand respectively for N, V, and R, because the message stream has already been reduced to a manageable set of

alternatives. In Poe the reader sees a table before reading the hidden sentence. In Doyle the reader sees a chain of reversible but increasingly constrained guesses.

Quantitative procedure	Poe	Doyle	Interpretive payoff
Recurrence tracking	repeated symbol counts are made explicit	repeated figure groups are compared across notes	stable candidates for letter values appear
Boundary detection	word breaks are absent and must be inferred indirectly	the flag functions as a boundary cue	segmentation becomes possible
Local probability	English letter frequency guides the first hypothesis	short answer pattern supports NEVER	false options are reduced
Proper name recovery	Kidd directs the language choice	ELSIE and ELRIGES connect code to persons and place	code enters the social world of the story

In place of extended reproductions, this version uses short verified excerpts and analytical paraphrase. That limitation does not weaken the result. On the contrary, it makes clear that the real scholarly object is not quotation length but the sequence of formal operations. Poe's cipher requires ranked frequency, doubled forms and language choice. Doyle's figure code requires repetition mapping, terminal marking and lexical reconstruction from short units. Both texts confirm that quantitative attention is not foreign to literary reading. It is already embedded in the very structure of the clue.

Discussion. The comparison shows that cipher motifs in detective prose are especially suitable for quantitative analysis because they externalize the hidden mechanism of interpretation. Ordinary clues often depend on tacit background knowledge, whereas coded clues force the story to display counting, ranking, matching and revision in overt form. For that reason these texts are useful not only for literary history but also for method. They provide compact demonstrations of how visible repetition can discipline inference before semantic certainty is reached. A second point concerns scale. The material examined here is small, yet the analytic logic aligns with larger quantitative work in literary studies. Moretti, Jockers and Underwood argue that formal repetition becomes readable when one attends to recurrence and distribution rather than isolated brilliance. Poe and Doyle confirm the same principle at a different level. Their coded passages condense the shift from singular sign to patterned system and show why numerical discipline can coexist with philological precision.

Conclusion. Quantitative analysis enters ciphered detective prose at the very point where meaning becomes uncertain. In Poe and Doyle, decoding advances when the text is turned into a finite set of repeatable signs, measured relations and testable lexical options. The role of counting here is neither decorative nor secondary. It organizes proof, narrows interpretation and converts the literary clue into an explicit procedure of reading.

REFERENCES

1. Sobirovich S. A. Robust Uzbek Asr AND Tts for Dialectal AND Noisy Settings //American Journal of Research in Humanities and Social Sciences. – 2025. – T. 40. – C. 1-4.

2. Khamidovna N. L. et al. An Online Platform for Uzbek-Russian and Russian-Uzbek Parallel Corpora:: Linguistic Challenges and Prospects Exemplified by

A. Kadyri's Novel “Bygone Days” //2024 9th International Conference on Computer Science and Engineering (UBMK). – IEEE, 2024. – C. 11-16.

3. Burrows J. Delta A Measure of Stylistic Difference and a Guide to Likely Authorship / J. Burrows // Literary and Linguistic Computing. 2002. Vol. 17. No. 3. P. 267-287.

4. Doyle A. C. The Adventure of the Dancing Men / A. C. Doyle // The Return of Sherlock Holmes. Project Gutenberg eBook No. 108.

5. Hoover D. L. Quantitative Analysis and Literary Studies / D. L. Hoover // A Companion to Digital Literary Studies / ed. R. Siemens, S. Schreibman. Oxford, Blackwell, 2008. P. 517-533.

6. Jockers M. L. Macroanalysis Digital Methods and Literary History / M. L. Jockers. Urbana, University of Illinois Press, 2013. 328 p.

7. Poe E. A. The Gold Bug / E. A. Poe // The Works of Edgar Allan Poe. Volume 1. Project Gutenberg eBook No. 2147.

8. Zipf G. K. Human Behavior and the Principle of Least Effort An Introduction to Human Ecology / G. K. Zipf. Cambridge, Mass., Addison Wesley Press, 1949. XI + 573 p.