

The Values in Numbers: Reading Japanese Literature in a Global Information Age

By Hoyt Long. New York: Columbia University Press, 2021. 368 pp. \$95.00.

**Reviewed by
George T. Sipos**

Hoyt Long's second book, *The Values in Numbers*, focuses on ways in which numerical and computational methodologies in the study of literature can reveal additional facets to canonical texts or entire genres. Organized in seven chapters, an introduction, five content chapters and an epilogue, *The Values in Numbers* progresses from an overview of the genealogy of computational and statistical methods in the study of literature in Japan by Japanese writers and researchers to specific study cases of genres such as *shishōsetsu*, writing methods such as stream-of-consciousness, racial discourse in literature, as well as the works of specific writers.

The introduction, "Uncertainty in Numbers," provides a justification for the choice of quantitative approaches to the study of literary texts. The author identifies two theoretical frameworks that underline the research interests in the book: the history behind the reluctance and outright resistance of the scholarly establishment against the quantitative method in the study of literature, and ways in which it can be overcome thus adding computational research to the array of recognized and vetted analysis tools customarily employed in the critical reading of literary texts. One of the theoretical frameworks is constituted by a historical (genealogical) approach of century-old attempts to employ numbers in literary studies, while the other is the way in which location, be it disciplinary, historical, or social, informs the use of numbers in said studies. When it comes to the latter, Long is particularly invested in bringing his contribution to the body of research that is specifically focused on area studies. Foregrounded in Franco Moretti's "distant reading," the author posits the core of the research in his book at the interstices between the conceptual divisions of numbers and literature on one hand, and literature and literatures on the other.

The introduction proceeds to explain the conceptual frameworks and the research behind each of the following five chapters. The organization of the book, as revealed by Long, is informed by five themes: Facts,

Archive, Genre, Influence, and Discourse, all with their own specific subthemes: Difference, Sample, Repetition, Judgment, and Character. The themes and subthemes work in tandem to create a conceptual bridge for the perceived (and much feared) “epistemological divide between numbers and literature” (8) and apply it to the specific conditions of modern Japanese literature, particularly to discussions around national canon, transnational novelistic forms, and colonial literature.

Chapter one, entitled “Facts and Difference,” follows a genealogy of literary study approaches through numbers, from Natsume Sōseki’s 1907 *Bungakuron* (Theory of literature) through to the late 1980s works on the literary style of scholars such as Morioka Kenji and Komori Yōichi, and the latter’s 1988 *Buntai toshite no monogatari* (Narrative as style). Contemporary with early twentieth-century attempts to establish universal aesthetic truths with the aid of numerical analysis, Sōseki’s work represented the logical continuation of early Meiji intellectuals’ eagerness to embrace statistical methodology in all realms of knowledge, as a way to process the ontological gap between Japan and the “civilized” world of Western science and technology and quickly assess and devise ways to close that gap. Insufficient as it was, Sōseki’s work was ahead of its time, and won’t exercise a direct influence on future literary studies, as Long shows, until 1935 with psychologist Hatano Kanji’s *Bunshō shinrigaku: Nihongo no hyōgen kachi* (The psychology of style: evaluating Japanese expression), a work that is part of a larger preoccupation of the time with writing and style. Hatano would attempt to create a veritable technical science capable to measure how the author’s individual language impacts the writing of literary texts.

In the mid-1950s, a new wave of studies focused mostly on empirical data (numbers) on their own, and devoid of direct interpretation, brought to the fore new research methods such as sampling, statistical modeling, and multidimensional analysis, which Hoyt Long delves into through the work of postwar linguists, psychologists, and literature scholars. The more these scholars’ work shifted away from literary studies, however, the wider the gap between them and literature experts became and, as such, the smaller their impact on the latter.

In chapter two: “Archive and Sample,” Long directs his attention to the more recent phenomenon of digitization and archiving of Japanese literature and of translations of “classics” of Western literature into Japanese. Through an analysis of databases such as Aozora and its randomly self-selected array of texts—largely digitized through the effort

and enthusiasm of volunteers—Long reveals the very limitation of such archival endeavors, inevitably guided by subjective criteria for selection and further skirted by legal and copyright limitations. In Long’s words, “Aozora functions as a sample of the archive in the way a climate scientist might think of an ice core sample, its variegated layers of particles and dust recording, from a localized context, the effects of complex processes and systems driving global climate conditions over time” (71).

To what extent is, then, Aozora, as well as other such archival attempts, representative for a category as vast as “modern Japanese literature”? Several pages of numerical evidence and statistical analysis prove with irrefutable mathematical precision what researchers of modern Japanese literature who use the texts in the database with frequency can attest: not truly representative. Long also looks at other exhaustive archival attempts, coming from established publishers, and at their efforts to capture the canon of foreign literature translations in Japan, such as the 1927 *Sekai bungaku zenshū* (Anthology of world literature), published by Shinchōsha, or the 1950s *Gendai Nihon bungaku zenshū* (Collected works of contemporary Japanese literature) published by Chikuma shobō, with similar conclusions. The chapter includes an impressive number of numerical analyses captured in charts and tables that provide a wealth of data.

The third chapter of the book, “Genre and Repetition,” is the first to deal with a concrete analysis in literary studies that employs numerical evidence. Long chose as the object of his research the long-debated and controversial genre of *shishōsetsu* (I-novel), as prevalent and necessary for the understanding of modern Japanese literature as it has been elusive to generations of scholars trying to define it. Long posits himself on the side of the scholars who, rather than contesting the very existence of *shishōsetsu* as a genre, look for ways to identify textual commonalities that could help define it. To do so he focuses on lexical repetitiveness, or redundancy, which he employs to find elements of shared mentality between texts usually identified as belonging to the genre. For his research, Long uses the notion of type-token ratio (TTR), the number of unique words in a text divided by the total number of words, as developed by John B. Carroll. Looking contrastively at popular works of fiction and comparing their TTR with that of some of the best-known prewar I-novels, the author identifies not only an overall lexical paucity, but also similarities between I-novels belonging to writers from different generations and their texts in terms of the ways in which they express

certain feelings. The chapter ends with a brief look at some of the iterations of the I-novel in modern Chinese literature.

The following chapter, “Influence and Judgment,” expands the analysis from a largely nationally bound genre such as the *shishōsetsu* to the circulation of a writing technique that achieved worldwide popularity throughout the 1920s: stream-of-consciousness (SOC). The transnational expansion and ready adoption of SOC by writers from Japan to the United States raises the question of literary influence across languages and cultural limitations. Long sees the SOC technique as the perfect vehicle for the exploration of the “world of world literature” (173) through computational models. To achieve his goal, he builds a computational model for SOC in English fiction which he then employs in building a similar one for Japanese. The two models demonstrate remarkable formal similarities between English and Japanese works, albeit the significant features are not all the same.

In the final content chapter, “Discourse and Character,” Long directs his attention to the study of racial discourse in modern Japanese literature and the ways in which Japanese writers navigated the pervasive racialized discourse of imperial Japan. Informed by Edward Said’s *Orientalism*, the author is looking at ways to utilize numbers to mitigate individual utterances of racial discourse. Thousands of texts covering the rise and fall of the Japanese empire (delineated here as covering the period 1890-1960) are used to extract a “semantics of race,” a repetitive vocabulary or racialized otherization. From general racial identifiers to specific literary characters, the chapter traces stereotypical language in close to two thousand literary texts and thousands of magazine articles. The final pages of the epilogue, “Difference in Numbers,” combine the conclusive remarks of the value and necessity of utilizing numerical analyses with the lucid acknowledgement of the limitations that they might impose on the study of literature.

Not at all surprisingly, Hoyt Long’s book makes for an interesting and engaging read not only because of the extraordinary number of literary texts that the numerical method allows him to analyze, but also because of the refreshing perspective he brings on the study of modern Japanese literature. Three conclusions emerge from *The Values in Numbers*: numbers and statistical data can (and should) play a significant role in the analysis of large number of texts, especially when it comes to identifying stylistic, linguistic, or thematic commonalities that define a certain genre or literary category (the three content chapters on *shishōsetsu*, SOC, and

racial discourse represent a great argument in that direction); numerical analysis can reveal—in combination with other, more traditional ways of literary analysis—facets of the text that might otherwise remain hidden and/or in-accessible; and, finally, the ever increasing access to digital databases and archives not only makes numerical analysis accessible and necessary, but also creates the opportunity for a new type of literary research (“batch text research”?) that might in turn produce additional venues for thematic, semantic, linguistic (among others) interpretations of intersectional and transnational studies of literature. From that perspective, Hoyt Long’s research in *The Values in Numbers* is a path-opener and it safely places Japanese literature and its study as an integral part of future numerical studies of world literature.

A couple of comments are perhaps necessary in the end. One has to do with the absence of the Japanese original text in some of the analyses that would have warranted direct access to it for the benefit of the Japanese literature scholar. The Japanese word tables included in the Appendix, while relevant, seem insufficient when it comes to discussions of style or numbers of characters in Chapter One. On the other hand, there must have been plenty of editorial and publishing limitations and restrictions preventing the inclusion of texts in the Japanese original. One other comment concerns the almost apologetic (justificatory?) tone of the Introduction, where it seems like Hoyt Long is almost worried that his research would face criticism for bringing up (again!) numerical analysis as a method to study literature. The author tries to emphasize repeatedly the fact that *The Values in Numbers* is merely offering the methodology as an *additional* tool for the study of literature, and not as *the* tool. While there must be plenty of critics out there who object to this way of looking at the study of literary texts, critics whom the author must have encountered and with whom he must have dealt, the thorough and informed research, the depth of the analysis, and the informed and productive results presented in *The Values in Numbers* speak for themselves and need no additional justification. Numerical analysis is (and has been) a valid and relevant method to study literature, and it is entirely up to literary scholars whether to utilize it in their own research or not.

From the genealogy of numerical analysis in Japanese literature to the importance of digital databases and archives, with their importance and limitations, to the in-depth case studies of the *shishōsetsu* (possibly the most important recent contribution to the study of the genre), the stream-of-consciousness method, and the racial discourse in modern Japanese

literature, *The Values in Numbers* is an original and refreshing contribution to the field, and a must read for students and scholars alike.