**UESAKA — Quantitative**

<4 figures; 3 tables>

**A Quantitative Analysis for the Authorship of Saikaku’s Posthumous Works Compared with Dansui’s Works**

Uesaka, A. and Murakami, M.

Saikaku Ihara (c. 1642–1693**)** is one of the most famous writers of the Edo period (1603–1868) in Japan.1 After publishing the maiden works of *Kousyoku ichidai otoko* (The Life of an Amorous Man, 1682), he became the leading author of *Ukiyozoushi*,2 which was a realistic literature from the Edo period. Saikaku’s works are known for their significance in developing Japanese novels today (Emoto and Taniwaki, 1996).

It is said that he wrote 24 works in 10 years. However, with the exception of *Kousyoku ichidai otoko*, those achievements have not been fully verified, due to some doubts in their authorship. For instance, Saikaku only wrote *Kousyoku ichidai otoko*, while the other works were written by either Saikaku’s student Dansui Houjyou (1663–1711) or a collaboration of Dansui and Saikaku (Mori, 1955).

Saikaku researchers have tried to identify his works by investigating their history, content, format, and so on. However, it remains unclear which works are really written by Saikaku. Accordingly, we decided to use a quantitative approach to inspect Saikaku’s authorship problems, because the potential of quantitative analysis of textual data has dramatically advanced. That method can provide new knowledge about the authorship problem of Saikaku’s works. Moreover, this research will be a good example in using a quantitative approach for the Japanese classical literature research domain because the quantitative approach is not common in that domain.

**Purpose of This Study**

In this paper, we focus on Saikaku’s posthumous works because many of Saikaku’s researchers have raised questions about their authorship. Saikaku’s posthumous works were edited and published from 1693 to 1699 by his student Dansui (Table 1). Therefore, there are claims that Dansui may have modified Saikaku’s work.

We have compared Saikaku’s posthumous works and other Saikaku works for differences (Uesaka and Murakami, 2013; 2014). If we try to resolve the authorship problems of Saikaku, Dansui is the most suspect writer of Saikaku’s work; therefore, Dansui’s text should be analyzed also.

**Database of Saikaku’s Works**

Since Japanese morphological analyzers are not applicable for early modern Japanese texts (Ogiso et al., 2013), we developed a database of Saikaku’s works with his researchers, who are editors of *Shinpen Saikaku Zenshu* (Shinpen Saikaku Zenshu Henshu Inkai, 2000). Figures 1 and 2 show a page from the book. Moreover, we used Dansui’s database for an analysis, which was developed by Banno and Mizutani, who are Saikaku and Dansui researchers, based on *Shinpen Saikaku Zenshu*. In this research, we use *Shikidou otsuzumi* (1687), *Chuya youjin ki* (1707), and *Budou hariai okagami* (1709)*,* because these works’ digital text and database were finished developing.



Figure 1. Saikaku’s publication.



Table 1. Saikaku’s posthumous works.



Figure 2. Modern form of Japanese language.

Table 2 shows a list of works in our database and the number of words in each work. According to our database, there are 583,934 words contained in 24 of Saikaku’s works and 55,504 words contained in three of Dansui’s works.



Table 2. Work name and the number of words.

Table 3 is a part of the database from Saikaku’s works used for this analysis. Since Japanese sentences are not separated by spaces, we added spaces between the words in all of the sentences. In addition, information was added for the analysis.



Table 3. Database of Saikaku’s works.

**Analysis and Results**

We compared Saikaku’s works (*Kousyoku ichidai otoko*, which has been verified to be a work of Saikaku, and five posthumous works) to Dansui’s three works (*Shikidou otsuzumi*, *Chuya youjin ki*,and *Budou hariai okagami*) using principal component analysis (PCA). PCA reduces the dimensionality of a dataset consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the dataset (Jolliffe, 2002). When applied to the frequencies of high-frequency items in texts, PCA often successfully reveals the authorial structure in a dataset (Kestemont et al., 2013).

At first, we examined the appearance rate of the seven principal grammatical categories: nouns, particles, verbs, auxiliary verbs, adjectives, adverbs, and adnominal adjectives. Grammatical categories were the basic information for authorship attribution. The study of *Tale of Genji* (Murakami, 2002) and *Sandai hiho bonsyoji* (Itou and Murakami, 1992) adopted the stylometry research using grammatical categories, and these help to identify the author.

Figure 3 shows the results of the analysis on grammatical categories, using the PCA with a correlation matrix. The horizontal axis shows the importance of the first principal component, and the vertical axis shows the second. The proportion of variance of the first principal component is 0.4031, while it is 0.29088 for the second; the cumulative proportion up to the second principal component is 0.69398. In this figure, indicating differences are revealed by PCA, Saikaku’s works are on the lower left and Dansui’s works on the upper middle.



*Kousyoku ichidai otoko*

Saikaku’s works

Dansui’s works

Figure 3. The PCA results (these circles drawn on the figure are 95% confidence ellipse).

Next, we examined the appearance rate of the eight principal particles: ‘no’ (の), ‘ni’ (に), ‘wo’ (を), ‘te’ (て), ‘ha’ (は), ‘to’ (と), ‘mo’ (も), and ‘ba’ (ば). Particles have a high appearance frequency and do not relate to the contents of a work. That kind of information identifies an authorship attribution very well (Murakami, 1994).

Figure 4 shows the results of the analysis on the appearance rate of the eight principal particles using the PCA. The proportion of variance of the first principal component is 0.3688, while it is 0.29204 for the second; the cumulative proportion up to the second principal component is 0.66084. In this figure, indicating differences are revealed by PCA; Saikaku’s works are on the lower right, and Dansui’s works are on the upper left. From these results, the second principal component identifies Saikaku’s works and Dansui’s works.



*Kousyoku ichidai otoko*

Saikaku’s works

Dansui’s works

Figure 4. The PCA results (these circles drawn on the figure are 95% confidence ellipse).

**Conclusion**

We conduct the analysis of Saikaku’s works and Dansui’s works using a quantitative approach. This result revealed that Saikaku’s works and Dansui’s works differ in grammatical categories and particles (Figures 3 and 4).

However, the possibility remains that Dansui modified a greater or lesser proportion of the work. Thus, we need to consider this issue from other perspectives and using other works and variables.

**Acknowledgments**

We would like to thank Professor Banno Hidekatsu and Professor Mizutani Takayuki for their help on our research.

**Notes**

1. In the late 18th century there was a Saikaku revival, inspiring Santo Kyoden and other fiction writers. Saikaku is generally considered the greatest fiction writer of the Edo period, and his works have influenced many modern Japanese writers (Shirane, 2004).

2. The term *Ukiyozoushi* refers to a vernacular fictional genre that originated in the Kyoto-Osaka area and spanned a 100-year period from the publication in 1682 of Ihara Saikaku’s *Kousyoku ichidai otoko* to the late 18th century (Shirane, 2004).

**References**

**Emoto, Y. and Taniwaki, M.** (1996). *Saikaku Jiten*. Ouhu.

**Itou, Z. and Murakami, M.** (1992). *Sandai hihou bonjouji no keiryoubunkengaku teki shin kenkyu*. Osaki hakuhou, no. 148.

**Jolliffe, I. T.** (2002). *Principal Component Analysis.* Springer, New York.

**Kestemont, M., Moens, S. and Deploige, J.** (2013). Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux. *Digital Scholarship in the Humanities,* http://dx.doi.org/10.1093/llc/fqt063.

**Mori, S.** (1955). *Saikaku to Saikaku bon*. Motomoto sha.

**Murakami, M.** (1994). *Shingan no kagaku-keiryou bunkengaku nyumon*. Asakura shoten.

**Murakami, M.** (2002). *Bunka wo hakaru–bunka keiryougaku josetsu–*. Asakura shoten.

**Ogiso, T., Ichimura, T. and Kono, T.** (2013). Preliminary Study of Morphological Analysis of Early Modern Japanese. *The 4th Workshop of Corpus Japanese Language,* National Institute for Japanese Language and Linguistics, pp. 145–50.

**Shinpen Saikaku Zenshu Henshu Inkai.** (2000). *Shinpen Saikaku Zenshu*. Bensei shuppan.

**Shirane, H.** (2004). *Early Modern Japanese Literature: An Anthology, 1600–1900.* Columbia University Press, New York.

**Uesaka, A. and Murakami, M.** (2013). Authorship Problem of Japanese Early Modern Literatures in the Seventeenth Century. *Digital Humanities 2013,* pp. 449–51.

**Uesaka, A. and Murakami, M.** (2014). A Quantitative Analysis for the Authorship of Saikaku’s Posthumous Works in the Seventeenth Century. *Digital Humanities 2014*, 547–49.