Original Paper _____ *Forma*, **21**, 259–273, 2006

Statistical Prosody: Rhyming Pattern Selection in Japanese Short Poetry

Kazuya HAYATA

Department of Socio-Informatics, Sapporo Gakuin University, Ebetsu 069-8555, Japan
E-mail address: hayata@edu.sgu.ac.jp

(Received August 5, 2005; Accepted August 2, 2006)

Keywords: Quantitative Poetics, Rhyme, HAIKU, TANKA, Bell Number

Abstract. Rhyme patterns of Japanese short poetry such as HAIKU, SENRYU, SEDOKAs, and TANKAs are analyzed by a statistical approach. Here HAIKU and SENRYU are poems composed of only seventeen syllables, which can be segmented into five, seven, and five syllables. As rhyming both head and end rhymes are considered. Analyses of sampled works of typical poets show that for the end rhyme composers prefere the avoided rhyming, whereas for the head rhyme they compose poems according to the stochastic law. Subsequently the statistical method is applied to a work of SEDOKAs as well as to those of TANKAs being written with three lines. Evaluation of the khi-square statistics shows that for a certain work of TANKAs the feature being identical to that of HAIKU is seen.

1. Introduction

Irrespective of languages, texts are categorized into proses and verses. Poems, in general, take the form of the latter. Conventional poetics has classified poems into a variety of forms such as a lyric, an epic, a prose, a long, and a short poem. One finds that in typical European poetry a sound on a site in a line is correlated to that on the same site in another line in an established form. Correlation among feet of lines is termed end rhyme in contrast to the head rhyme for the one among heads of lines (SAKAMOTO, 2002).

In Japan, in addition to modern poems being composed of many lines, several forms of short poems have traditionally enjoyed general popularity. (For poems belonging to the WAKA family, see Table 1.) For examples one enumerates TANKA [57577], SEDOKA [577•577], BUSSOKUSEKIKA [575777], HAIKU [575], SENRYU [575], KYOKA [57577], and DODOITSU [7775]. As is indicated in the bracket, basically these poems are composed of phrases with combination between five and seven syllables. For instance, a HAIKU includes seventeen syllables consisting of five, seven, and five syllables, while a TANKA has thirty-one syllables consisting of five, seven, five, seven, and seven syllables. On both types several rules are imposed; well known is that a seasonable word termed KIGO must be included in a sentence. Note that such restrictions are not put on SENRYU and DODOITSU, though the syllable form of the former is identical to that of HAIKU.

Table 1. The form of WAKA family with algebraic formula m [(5 + 7) n + 7].

Poem	m	n	Syllabic sequence
KATAUTA	1	1	5 + 7 + 7
SEDOKA	2	1	(5 + 7 + 7) + (5 + 7 + 7)
TANKA	1	2	5 + 7 + 5 + 7 + 7
CHOKA	1	≥3	(5 + 7) + (5 + 7) + ··· + (5 + 7) + 7

In this paper, rhyme patterns of Japanese short poetry such as HAIKU, SENRYU, SEDOKAs, and TANKAs are analyzed by a statistical approach. As rhymes both head and end rhymes are considered. Subsequently the statistical method is applied to a work of SEDOKAs as well as to those of TANKAs being partitioned into three lines. Here a SEDOKA, a form of WAKAs, consists of twin elements called KATAUTAs; a KATAUTA is segmented into five, seven, and seven syllables. Through evaluation of the khi-square statistics it is found that for a certain work of TANKAs the feature being identical to that of HAIKU is seen.

2. Analytical Method with Statistical Approach

In this section a statistical approach to the analysis of rhymes in Japanese short poems is first described through application to the well-known HAIKU work OKU NO HOSOMICHI (HAGIWARA, 1979), which can be translated as, for instance, *the Narrow Road to Oku* (KEENE, 1996) and includes sixty-three poems; the author is Basho Matsuo (1644–1694), who has been one of the most outstanding HAIKU poets; this work was published in 1702. It should be emphasized here that HAIKU is, along with SENRYU, one of the shortest Japanese poems of only seventeen syllables which contain a seasonable word, KIGO.

2.1. Phonetic system of Japanese syllables

With the Hepburn style in the ROMAJI expression of Japanese texts, twenty-two letters are available, which can be classified into five vowels {a, i, u, e, o}, eleven voiceless consonants {k, s, t, c, n, h, f, m, r, y, w}, five voiced consonants {g, z, j, d, b}, and a semivoiced consonant {p}. Here the n-sound in the voiceless consonants is also used as a flipping sound. A pronounced feature of Japanese is that, in comparison with European languages, the frequency of the vowels becomes considerably higher. In addition, the syllabic structure of Japanese is basically open, i.e., in Japanese texts, except the flipping sound, every consonant on a word accompanies a vowel. For these reasons, in what follows we will focus our attention on the five vowels in addition to the flipping sound; the alliteration between consonants is not considered.

2.2. Form of rhyming

As an example the opening poem of the HAIKU work is chosen

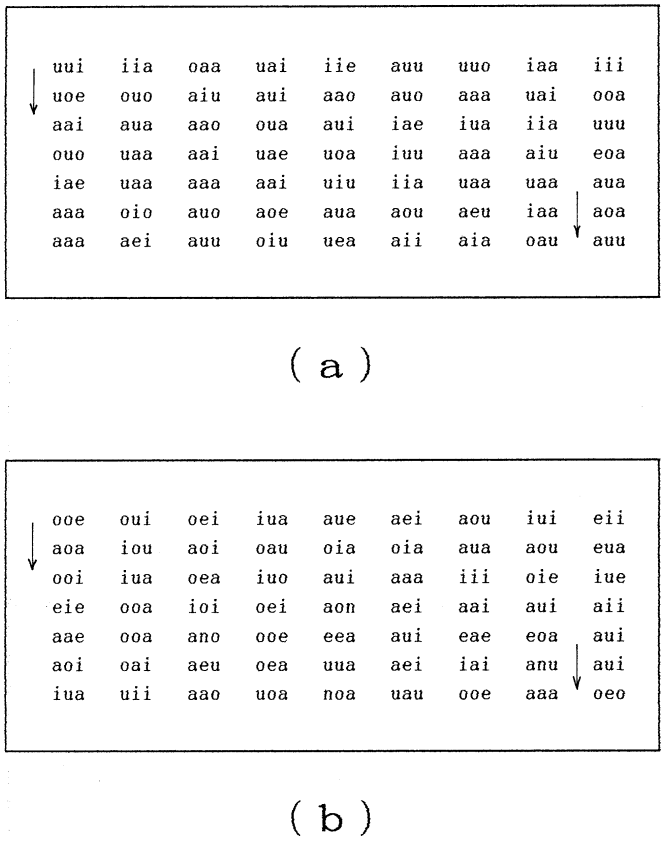

Fig. 1. Surveyed results of rhyme patterns in the HAIKU work of Basho Matsuo, entitled the *Narrow Road to Oku*. (a) For head rhyme; (b) For end rhyme.

Kusa no to mo
Sumikawaru yo zo (1)
Hina no ie,

where each line is composed of five, seven, and five syllables, respectively. A translation of this poem into English can be seen in Appendix A. Taking notice of rhyming between vowels, one finds in this poem 'uui' for the head rhyme and 'ooe' for the end rhyme. Results of the sixty-three poems in the work are juxtaposed in Fig. 1. Herein one finds that there are five patterns in rhyming

$$\{\mathrm{ABC}, \mathrm{AAB}, \mathrm{ABA}, \mathrm{ABB}, \mathrm{AAA}\}, \tag{2}$$

where ABC, AAB, ABA, ABB, and AAA represent, e.g., 'iuo', 'iiu', 'iui', 'iuu', and 'iii', respectively. In composing Japanese short poems the rhyme pattern is important because

Table 2. Frequency distribution of rhyme pattern in the *Narrow Road to Oku*.

	(a) For head rhyme		(b) For end rhyme	
Pattern	Surveyed	Expected	Surveyed	Expected
ABC	23	22.55	37	29.70
AAB	12	11.58	11	10.12
ABA	9	11.58	9	10.12
ABB	12	11.58	3	10.12
AAA	7	5.71	3	2.95 < 5
Sum	63	63.00	63	63.01

it would be responsible for making a poem sonorous. By analogy with English poetry, ABC, (AAB, ABA, ABB), and AAA correspond, respectively, to a blank verse, an imperfect rhyme, and a perfect rhyme. In other words, for ABC the three phrases of a poem are anticorrelated, whereas for AAA these are perfectly correlated; the intermediate, (AAB, ABA, ABB), could be regarded as being partially correlated. It should be stressed here that, because of their shortest length, for HAIKU the effect of rhyming might be more significant than that for, e.g., sonnets being a typical style of modern European poetry.

The number of the rhyming patterns in Eq. (2) can be determined with third-order (n = 2) Bell number (BELL, 1934; for a review see, e.g., GARDNER, 1992). Here the number can be generated recursively:

$$B(n+1) = \sum_{k=0}^{n} {}_nC_k B(k). \tag{3}$$

With $B(0) = 1$ being assumed, one obtains $B(1) = 1$, $B(2) = 2$, $B(3) = 5$, $B(4) = 15$, $B(5) = 52$, ⋯. It would be interesting to note that $B(5)$ yields the number of the KOUZU (GENJIKOU) patterns (TAKAKI, 2003). Frequency distribution of the present rhyming pattern is shown in Table 2. In imitation of KOUZU the five patterns of rhyming are illustrated in Fig. 2. Herein, for HAIKU, the three horizontal lines stand for the three phrases of a poem; lines with the same rhyme are connected with a vertical line. More symbolic expression termed Bell number diagrams has recently been presented by DICKAU (1996).

2.3. Calculation of theoretical distribution

The probability for each type of rhyming is derivable in the form

$$\begin{aligned} P(\mathrm{ABC}) = 3!\,[&p_a p_i(p_u + p_e + p_o + p_n) + p_a p_u(p_e + p_o + p_n) + p_a p_e(p_o + p_n) + p_a p_o p_n \\ &+ p_i p_u(p_e + p_o + p_n) + p_i p_e(p_o + p_n) + p_i p_o p_n + p_u p_e(p_o + p_n) + (p_u + p_e)p_o p_n] \end{aligned} \tag{4a}$$

$$\begin{aligned} P(\mathrm{AAB}) &= P(\mathrm{ABA}) = P(\mathrm{ABB}) \\ &= p_a{}^2(1 - p_a) + p_i{}^2(1 - p_i) + p_u{}^2(1 - p_u) + p_e{}^2(1 - p_e) + p_o{}^2(1 - p_o) + p_n{}^2(1 - p_n) \end{aligned} \tag{4b}$$

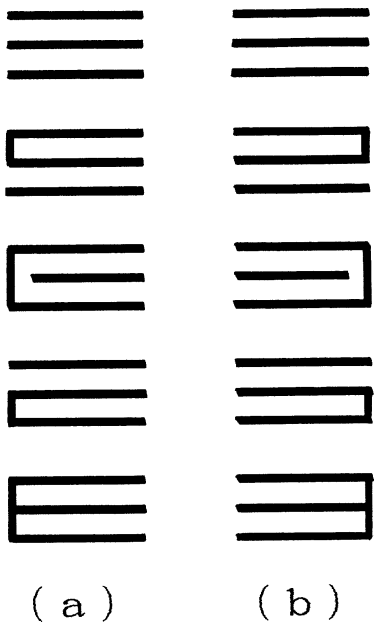

Fig. 2. Schematic illustrations for the five patterns of rhyming. The method follows that of KOUZU (GENJIKOU). (a) For head rhyme; (b) For end rhyme.

$$P(\mathrm{AAA}) = p_{\mathrm{a}}{}^3 + p_{\mathrm{i}}{}^3 + p_{\mathrm{u}}{}^3 + p_{\mathrm{e}}{}^3 + p_{\mathrm{o}}{}^3 + p_{\mathrm{n}}{}^3, \tag{4c}$$

where p_x (x = a, i, u, e, o, n) stands for the statistical probability of finding the letter x in the series of rhyming patterns. Derivation of Eq. (4) is described in Appendix B. In the sample shown in Fig. 1 they become

$$(p_{\mathrm{a}}, p_{\mathrm{i}}, p_{\mathrm{u}}, p_{\mathrm{e}}, p_{\mathrm{o}}, p_{\mathrm{n}}) = (78/189, 34/189, 44/189, 10/189, 23/189, 0) \tag{5a}$$

for the analysis of head rhyme (Fig. 1(a)) and

$$(p_{\mathrm{a}}, p_{\mathrm{i}}, p_{\mathrm{u}}, p_{\mathrm{e}}, p_{\mathrm{o}}, p_{\mathrm{n}}) = (54/189, 41/189, 27/189, 25/189, 38/189, 4/189) \tag{5b}$$

for the analysis of end rhyme (Fig. 1(b)).

2.4. Hypothesis test

In order to examine if the observed distribution of rhyme patterns is determined stochastically we will perform a test with the khi-square statistics (e.g., MUTO, 1995). First we set a null hypothesis

$$H\text{: The rhyme distribution is determined stochastically.} \tag{6}$$

The surveyed and the expected frequencies are listed in Table 2. Here with N being the number of poems ($N = 63$ for the present example) the latter is derivable from Eq. (4)

$$F_1 = N\,P(\mathrm{ABC}),\ F_2 = N\,P(\mathrm{AAB}),\ F_3 = N\,P(\mathrm{ABA}),\ F_4 = N\,P(\mathrm{ABB}),\ F_5 = N\,P(\mathrm{AAA}), \quad (7)$$

with $(F_1, F_2, F_3, F_4, F_5) = (22.55, 11.58, 11.58, 11.58, 5.71)$ for head rhyme, and $(F_1, F_2, F_3, F_4, F_5) = (29.70, 10.12, 10.12, 10.12, 2.95)$ for end rhyme. Note here that, in order to avoid singularity of the khi-square statistics, in the testing none of them must be smaller than 5 (MUTO, 1995). With the theoretical distribution being determined the khi-square statistics can be calculated by

$$ch^2 = \sum_{k=1}^{5} \left(f_k - F_k\right)^2 \Big/ F_k, \quad (8)$$

where f_k ($k = 1$–5) is the surveyed frequency for the respective rhyme pattern.Specifically, in Table 2 $(f_1, f_2, f_3, f_4, f_5) = (23, 12, 9, 12, 7)$ for head rhyme, and $(f_1, f_2, f_3, f_4, f_5) = (37, 11, 9, 3, 3)$ for end rhyme. On substitution of the data in Table 2(a) into Eq. (8) one obtains

$$ch^2 = 0.90. \quad (9)$$

For level 5% test and for the degree of freedom being $(5 - 1) - 1 = 3$, from the numerical table available one obtains as the critical value of the khi-square distribution (MUTO, 1995)

$$ch_3^2(0.05) = 7.81. \quad (10)$$

Comparison between Eqs. (9) and (10) indicates that $ch^2 < ch_3^2(0.05)$. Therefore the hypothesis (6) cannot be rejected. From this result one can conclude that the process of selecting the head rhyme pattern is stochastic, i.e., the selection is governed by the laws of probability. It should be noticed that, because $F_5 < 5$, for the end rhyme data in Table 2(b) one cannot use the khi-square test.

3. Results of Statistical Analysis

In this section typical results are shown for several Japanese short poems. Throughout hypothesis tests the level is fixed on 5%. Thus, as the critical value of the khi-square one may use Eq. (10).

3.1. HAIKU

Table 3 shows the observed and the calculated results for a work of Basho Matsuo; therein 202 poems concerning the new year to the spring are contained (OHTANI *et al.*, 1962). Note that the present sample text has nothing to do with the HAIKU work considered in the preceding section. The values of the khi-square statistics yield $ch^2 = 2.88 < 7.81$ for head rhyme (Table 3(a)) and $ch^2 = 37.75 > 7.81$ for end rhyme (Table 3(b)), which verifies that for the head rhyme the hypothesis (6) is not rejectable whereas for the end rhyme it is rejected.

Table 3. Frequency distribution of rhyme pattern in a work of Basho Matsuo, in which HAIKU concerning new year and spring are compiled. Sample data necessary for the calculation of Eq. (7) with Eq. (4) are (p_a, p_i, p_u, p_e, p_o, p_n) = (261/606, 103/606, 99/606, 27/606, 116/606, 0) for head rhyme and (p_a, p_i, p_u, p_e, p_o, p_n) = (174/606, 128/606, 102/606, 54/606, 132/606, 16/606) for end rhyme.

	(a) For head rhyme		(b) For end rhyme	
Pattern	Surveyed	Expected	Surveyed	Expected
ABC	68	71.39	135	93.61
AAB	46	37.05	15	32.83
ABA	37	37.05	28	32.83
ABB	34	37.05	21	32.83
AAA	17	19.45	3	9.88
Sum	202	201.99	202	201.98

Table 4. Frequency distribution of rhyme pattern in a work of Buson Taniguchi-Yosa, in which HAIKU concerning summer are collected. Here (p_a, p_i, p_u, p_e, p_o, p_n) = (252/699, 139/699, 133/699, 35/699, 140/699, 0) for head rhyme and (p_a, p_i, p_u, p_e, p_o, p_n) = (201/699, 125/699, 131/699, 82/699, 147/699, 13/699) for end rhyme.

	(a) For head rhyme		(b) For end rhyme	
Pattern	Surveyed	Expected	Surveyed	Expected
ABC	98	91.92	158	109.42
AAB	36	41.61	26	37.54
ABA	35	41.61	29	37.54
ABB	46	41.61	19	37.54
AAA	18	16.26	1	10.95
Sum	233	233.01	233	232.99

Table 4 shows the surveyed and the expected results for a work of Buson Taniguchi-Yosa (1716–1783); therein 233 poems concerning the summer are collected (OGATA, 1989). The values of the khi-square statistics yield $ch^2 = 2.86 < 7.81$ for head rhyme (Table 4(a)) and $ch^2 = 45.26 > 7.81$ for end rhyme (Table 4(b)), which indicates that for the head rhyme the hypothesis (6) is not rejectable whereas for the end rhyme it is rejected.

Table 5 lists the observed and the calculated results for a work of Issa Kobayashi (1763-1827); therein 136 poems published in the Kansei period (1789–1801) are selected (MARUYAMA, 1990). The values of the khi-square statistics yield $ch^2 = 6.95 < 7.81$ for head rhyme (Table 5(a)) and $ch^2 = 20.43 > 7.81$ for end rhyme (Table 5(b)), which indicates that for the head rhyme the hypothesis (6) is not rejectable whereas for the end rhyme it is rejected.

Table 6 shows the surveyed and the expected results for KATSUSHIKA (1930), a representative work of Shuoshi Mizuhara (1892–1981); therein 539 poems are enumerated (MATSUNE *et al.*, 1967). The values of the khi-square statistics are $ch^2 = 2.27 < 7.81$ for head rhyme (Table 6(a)) and $ch^2 = 23.79 > 7.81$ for end rhyme (Table 6(b)), which indicates that for the head rhyme the hypothesis (6) cannot be rejected whereas for the end rhyme it is rejected.

Table 5. Frequency distribution of rhyme pattern in a work of Issa Kobayashi, in which HAIKU composed in the Kansei period are collected. Here $(p_a, p_i, p_u, p_e, p_o, p_n) = (147/408, 90/408, 69/408, 16/408, 86/408, 0)$ for head rhyme and $(p_a, p_i, p_u, p_e, p_o, p_n) = (130/408, 93/408, 75/408, 37/408, 70/408, 3/408)$ for end rhyme.

	(a) For head rhyme		(b) For end rhyme	
Pattern	Surveyed	Expected	Surveyed	Expected
ABC	60	52.28	85	59.49
AAB	27	24.65	14	22.95
ABA	26	24.65	18	22.95
ABB	13	24.65	16	22.95
AAA	10	9.76	3	7.64
Sum	136	135.99	136	135.98

Table 6. Frequency distribution of rhyme pattern in the work, KATSUSHIKA, of Shuoshi Mizuhara. Here $(p_a, p_i, p_u, p_e, p_o, p_n) = (621/1617, 343/1617, 290/1617, 64/1617, 299/1617, 0)$ for head rhyme and $(p_a, p_i, p_u, p_e, p_o, p_n) = (407/1617, 442/1617, 312/1617, 137/1617, 299/1617, 20/1617)$ for end rhyme.

	(a) For head rhyme		(b) For end rhyme	
Pattern	Surveyed	Expected	Surveyed	Expected
ABC	215	202.37	286	242.82
AAB	91	98.13	58	89.66
ABA	100	98.13	96	89.66
ABB	97	98.13	82	89.66
AAA	36	42.23	17	27.21
Sum	539	538.99	539	539.01

3.2. SENRYU: Japanese satirical verses

Although the syllabic structure of senryu coincides with that of HAIKU, this poetry is not bound by the stylistic requirements concerning the KIGO. Table 7 lists the observed and the calculated results for the SENRYU work entitled HAIFU YANAGIDARU: Vol. 1 (1765); therein 316 poems are seen (YAMAZAWA, 1995). The values of the khi-square statistics become $ch^2 = 2.24 < 7.81$ for head rhyme (Table 7(a)) and $ch^2 = 20.75 > 7.81$ for end rhyme (Table 7(b)), which indicates that for the head rhyme the hypothesis (6) is not rejectable whereas for the end rhyme it is rejected.

3.3. SEDOKAs

An iterative form of twin KATAUTAs is termed SEDOKA, which is composed of thirty-eight syllables with six phrases (Table 1). Table 8 summarizes the observed and the calculated results for a work of SEDOKAs; therein 64 poems, i.e., 128 KATAUTAs, are edited (KAMIYA, 1906). The values of the khi-square statistics yield $ch^2 = 10.98 > 7.81$ for head rhyme (Table 8(a)) and $ch^2 = 1.65 < 7.81$ for end rhyme (Table 8(b)), which indicates that for the head rhyme the hypothesis (6) is rejected whereas for the end rhyme it cannot be rejected. It is interesting to note that this conclusion exhibits a striking contrast to those concluded for HAIKU and SENRYU.

Table 7. Frequency distribution of rhyme pattern in the work, HAIFU YANAGIDARU: Vol. 1, where (p_a, p_i, p_u, p_e, p_o, p_n) = (297/948, 227/948, 161/948, 77/948, 186/948, 0) for head rhyme and (p_a, p_i, p_u, p_e, p_o, p_n) = (163/948, 282/948, 121/948, 153/948, 218/948, 11/948) for end rhyme.

	(a) For head rhyme		(b) For end rhyme	
Pattern	Surveyed	Expected	Surveyed	Expected
ABC	132	134.83	179	145.20
AAB	47	54.34	45	51.68
ABA	58	54.34	39	51.68
ABB	57	54.34	49	51.68
AAA	22	18.16	4	15.75
Sum	316	316.01	316	315.99

Table 8. Frequency distribution of rhyme pattern in an Annotated Book on SEDOKAs, where (p_a, p_i, p_u, p_e, p_o, p_n) = (164/384, 92/384, 59/384, 8/384, 61/384, 0) for head rhyme and (p_a, p_i, p_u, p_e, p_o, p_n) = (56/384, 94/384, 40/384, 36/384, 141/384, 17/384) for end rhyme.

	(a) For head rhyme		(b) For end rhyme	
Pattern	Surveyed	Expected	Surveyed	Expected
ABC	53	42.41	56	54.50
AAB	13	24.29	26	21.54
ABA	30	24.29	19	21.54
ABB	24	24.29	19	21.54
AAA	8	12.71	8	8.87
Sum	128	127.99	128	127.99

3.4. TANKAs written in three lines

The short poem called TANKA is the most popular form in the WAKA family, which is composed of thirty-one syllables with five phrases (Table 1). There once were poets who attempted to write TANKAs with three lines. For instance, in his representative works such as *a Handful of Sand* (1910) and *Sad Toys* (1912), Takuboku Ishikawa (1886–1912) adopted this style of writing (ISHIKAWA, 2001). From the former a poem will be sampled

Ningen no tsukawanu kotoba
Hyottoshite (11)
Ware nomi shireru gotoku omou hi.

Herein one finds ‘ioa’ (ABC type) for the heads and ‘aei’ (ABC type) for the feet. An example of translation into English is given by Appendix A. Table 9 shows the observed and the calculated results for this work; therein 551 poems are seen. The values of the khi-square statistics are $ch^2 = 1.02 < 7.81$ for the analysis of head rhyme (Table 9(a)) and $ch^2 = 9.72 > 7.81$ for the analysis of end rhyme (Table 9(b)), which indicates that for the head rhyme the hypothesis (6) is not rejectable whereas for the end rhyme it is rejected. It is

Table 9. Frequency distribution of rhyme pattern in *a Handful of Sand*, edited by Takuboku Ishikawa. Here $(p_a, p_i, p_u, p_e, p_o, p_n) = (634/1653, 380/1653, 204/1653, 71/1653, 364/1653, 0)$ for the analysis of head rhyme and $(p_a, p_i, p_u, p_e, p_o, p_n) = (275/1653, 505/1653, 312/1653, 168/1653, 363/1653, 30/1653)$ for the analysis of end rhyme.

	(a) For head rhyme		(b) For end rhyme	
Pattern	Surveyed	Expected	Surveyed	Expected
ABC	200	201.59	278	251.49
AAB	101	101.56	95	90.38
ABA	98	101.56	87	90.38
ABB	101	101.56	68	90.38
AAA	51	44.75	23	28.37
Sum	551	551.02	551	551.00

Table 10. Frequency distribution of rhyme pattern in *Sad Toys*, edited by Takuboku Ishikawa. Here $(p_a, p_i, p_u, p_e, p_o, p_n) = (198/582, 135/582, 77/582, 48/582, 124/582, 0)$ for the analysis of head rhyme and $(p_a, p_i, p_u, p_e, p_o, p_n) = (107/582, 161/582, 127/582, 60/582, 114/582, 13/582)$ for the analysis of end rhyme.

	(a) For head rhyme		(b) For end rhyme	
Pattern	Surveyed	Expected	Surveyed	Expected
ABC	83	79.75	97	91.27
AAB	28	33.92	32	31.24
ABA	42	33.92	36	31.24
ABB	31	33.92	26	31.24
AAA	10	12.49	3	9.00
Sum	194	194.00	194	193.99

worth noting that this conclusion coincides with those made for HAIKU (Subsec. 3.1) and SENRYU (Subsec. 3.2).

Table 10 lists the surveyed and the calculated results for another work of the same author, which is entitled *Sad Toys*; therein 194 poems are seen. The values of the khi-square statistics produce $ch^2 = 3.84 < 7.81$ for the analysis of head rhyme (Table 10(a)) and $ch^2 = 5.97 < 7.81$ for the analysis of end rhyme (Table 10(b)), which indicates that, in sharp contrast to the above test, for both cases the hypothesis (6) is not rejectable.

Lastly a three-lined writing by another poet will be considered. Table 11 shows the observed and the expected results for the TANKA work *NAKIWARAI* compiled by Zemmaro Toki (1885–1980); he has been known by his pen name Aika, and contracted a friendship with Takuboku. In this virgin work of Aika 146 poems are juxtaposed (YAMAZAKI, 1968). The values of the khi-square statistics produce $ch^2 = 2.63 < 7.81$ for the analysis of head rhyme (Table 11(a)) and $ch^2 = 0.77 < 7.81$ for the analysis of end rhyme (Table 11(b)), which indicates that for both cases the hypothesis (6) cannot be rejected. This judgment is identical to that for *Sad Toys*.

Table 11. Frequency distribution of rhyme pattern in NAKIWARAI compiled by Aika Toki. Here $(p_a, p_i, p_u, p_e, p_o, p_n) = (172/438, 79/438, 65/438, 15/438, 107/438, 0)$ for the analysis of head rhyme and $(p_a, p_i, p_u, p_e, p_o, p_n) = (86/438, 117/438, 78/438, 43/438, 107/438, 7/438)$ for the analysis of end rhyme.

	(a) For head rhyme		(b) For end rhyme	
Pattern	Surveyed	Expected	Surveyed	Expected
ABC	57	52.53	66	67.46
AAB	30	27.05	27	23.85
ABA	24	27.05	23	23.85
ABB	21	27.05	22	23.85
AAA	14	12.31	8	6.98
Sum	146	145.99	146	145.99

Table 12. Comparison among frequency distributions of partition patterns of three TANKA works, where all poems are partitioned into three lines. Here Texts I, II, and III represent, respectively, *Sad Toys* by Takuboku, *a Handful of Sand* by the same poet, and NAKIWARAI by Aika.

Pattern	Text I (Surveyed)	Text II (Modified)	Text III (Modified)
113	5	5.99	6.64
122	35	42.25	39.86
131	18	17.25	22.59
212	56	60.56	23.92
221	45	32.74	54.48
311	35	35.21	46.51
Sum	194	194.00	194.00

4. Analyzing Pattern of Partition

In this section, instead of rhyming, for the three works of TANKAs the method of division into three lines is studied. The distribution of segmentation is seen in Table 12, where there are six patterns being abbreviated by 113, 122, ..., 311; for instance, 113 stands for partitioning the five phrases into one/one/three phrases. (Poem (11) is written in the form of 212.) Note that in this table the distributions for Texts II and III are modified in comparison with Text I. The null hypothesis becomes

H: The deviation of the distribution of Text II (or Text III) from that of Text I is not statistically significant. (12)

The khi-square statistics can be calculated by

$$ch^2 = \sum_{k=1}^{6} \left(f_k - F_k\right)^2 \Big/ F_k = \begin{cases} 6.38 & \text{for Text II} \\ 49.45 & \text{for Text III.} \end{cases} \tag{13}$$

Here $(f_1, f_2, f_3, f_4, f_5, f_6) = (5, 35, 18, 56, 45, 35)$; for Text II, $(F_1, F_2, F_3, F_4, F_5, F_6) = (5.99, 42.25, 17.25, 60.56, 32.74, 35.21)$, whereas for Text III, $(F_1, F_2, F_3, F_4, F_5, F_6) = (6.64, 39.86, 22.59, 23.92, 54.48, 46.51)$.

With level 5% test and with the degree of freedom being $6 - 1 = 5$, from the numerical table available one obtains as the critical value of the khi-square distribution (MUTO, 1995)

$$ch_5^2(0.05) = 11.07. \tag{14}$$

Comparison between Eqs. (13) and (14) verifies that $ch^2 < ch_5^2(0.05)$ for comparison between Texts I and II, while $ch^2 > ch_5^2(0.05)$ for comparison between Texts I and III. Hence for the former the hypothesis (12) is not rejectable, whereas for the latter it is rejected. In summary, it can be concluded that the segment distribution of Text I is (is not) consistent with that of Text II (Text III). Here it should be remembered that Texts I and II were authored and edited by the same poet, whereas Text III was composed by, though they had formed a friendship each other, a different poet.

5. Discussion

In Sec. 3, through a statistical approach it has been verified that for works of HAIKU and SENRYU as well as for a certain work of TANKAs the end rhyme is avoided. To explain this fact one should notice some of forbidden rules imposed on a string of HAIKU, which is termed HAIKAI NO RENGA. Among those we mention an important rule termed SARIKIRAI or KIRAIMONO, which means that in order to avoid monotony as well as to maintain variations, in addition to the same character, similar words must not be collocated in the neighborhood. The preferential use of avoided rhyming (selected rhymeless pattern) could arise from the avoided monotony and/or the enhanced variations. It should be stressed that this feature seems to reflect a poetic talent of a composer. The second reason for the nonstochastic nature of the end rhyme might be recognized in the existence of KIREJI (selected syllables for use in separating neighboring phrases) being used rhetorically. One could find their candidates in the imperative as well as the terminative form of specific JOSHIs and auxiliary verbs. Note that there exist main syllables for KIREJIs, termed KIREJI JUHACHIJI, which are reproduced by

kana mogana shi ji ya ran ka keri yo
zo tsu se zu re nu he ke ikani. (15)

Herein the syllables {se, re, he, ke} correspond to the ending of the imperative form of a verb, whereas {shi} stands for the ending of an adjective. In the second line of Poem (1) being cited in Subsec. 2.2, one can find {zo}. To conclude, preferential use of these eighteen syllables in the end of the HAIKU phrases would be responsible for finding the nonstochastic nature in the end rhyme.

6. Conclusions

Rhyme patterns of Japanese short poetry such as HAIKU, SENRYU, SEDOKAs, and

TANKAs have been analyzed by a statistical approach. As rhymes both head and end rhymes have been considered. Analyses of sampled poetic works of typical poets have shown that for the end rhyme they select avoided rhyming whereas for the head rhyme they compose poems according to the stochastic law. Subsequently the statistical method has been applied to a work of SEDOKAs (twin KATAUTAs) as well as to those of TANKAs being written with three lines. Through evaluation of the khi-square statistics it has been found that for a certain work of TANKAs the feature being identical to that of HAIKU is seen.

Finally, it should be mentioned that the methodology proposed in this paper can readily be extended to the study of other styles of short poetry such as DODOITSU as well as arbitrary TANKAs. Therein the combination of rhyming patterns is described in terms of $B(4) = 15$ and $B(5) = 52$, respectively.

Appendix A. Translation of Poems into English

Poem (1) in the text was translated by BRITTON (1980):

This rude hermit cell
Will be different now, knowing Dolls'
Festival as well.

More recently, the same poem was translated by KEENE (1996):

Even a thatched hut
May change with a new owner
Into a doll's house.

It might be interesting to remember that the trinity between HAIKU, Zen, and living was discussed by BLYTH (1947).

Poem (11) by Takuboku could be translated as

"Words which man does not use,
Possibly,
Only I could know," thinking today.

Appendix B. Derivation of Eq. (4)

For the rhyming pattern of ABC there are twenty combinations; specifically

aiu, aie, aio, ain, aue, auo, aun, aeo, aen, aon,
iue, iuo, iun, ieo, ien, ion, ueo, uen, uon, eon.

Note that each combination includes 3! permutations; for instance, for 'aiu', one obtains

aiu, aui, iau, iua, uai, uia.

Therefore, the probability of observing the rhyming can be written as

$$P(\mathrm{ABC}) = 3!\,(p_a p_i p_u + p_a p_i p_e + p_a p_i p_o + p_a p_i p_n + p_a p_u p_e + p_a p_u p_o + p_a p_u p_n + p_a p_e p_o + p_a p_e p_n + p_a p_o p_n + p_i p_u p_e + p_i p_u p_o + p_i p_u p_n + p_i p_e p_o + p_i p_e p_n + p_i p_o p_n + p_u p_e p_o + p_u p_e p_n + p_u p_o p_n + p_e p_o p_n).$$

Grouping the terms in this equation, we obtain Eq. (4a) in the text.

For the rhyming pattern of AAB there exist thirty combinations; specifically

aai, aau, aae, aao, aan, iia, iiu, iie, iio, iin,
uua, uui, uue, uuo, uun, eea, eei, eeu, eeo, een,
ooa, ooi, oou, ooe, oon, nna, nni, nnu, nne, nno.

Therefore, the probability of observing the rhyming can be expressed as

$$\begin{aligned} P(\mathrm{AAB}) = {} & p_a^2 p_i + p_a^2 p_u + p_a^2 p_e + p_a^2 p_o + p_a^2 p_n + p_i^2 p_a + p_i^2 p_u + p_i^2 p_e + p_i^2 p_o + p_a^2 p_n + p_u^2 p_a \\ & + p_u^2 p_i + p_u^2 p_e + p_u^2 p_o + p_a^2 p_n + p_e^2 p_a + p_e^2 p_i + p_e^2 p_u + p_e^2 p_o + p_e^2 p_n + p_o^2 p_a \\ & + p_o^2 p_i + p_o^2 p_u + p_o^2 p_e + p_o^2 p_n + p_n^2 p_a + p_n^2 p_i + p_n^2 p_u + p_n^2 p_e + p_n^2 p_o. \end{aligned}$$

Using the normalization condition

$$p_a + p_i + p_u + p_e + p_o + p_n = 1,$$

we obtain Eq. (4b) in the text. The same procedure is applicable to the derivation of $P(\mathrm{ABA})$ and $P(\mathrm{ABB})$.

For the rhyming pattern of AAA there are six cases

aaa, iii, uuu, eee, ooo, nnn.

Consequently the probability of observing this pattern can be written in the form

$$P(\mathrm{AAA}) = p_a^3 + p_i^3 + p_u^3 + p_e^3 + p_o^3 + p_n^3,$$

which coincides with Eq. (4c) in the text.

REFERENCES

BELL, E. T. (1934) Exponential numbers, *Amer. Math. Monthly*, **41**, 411–419.

BLYTH, R. H. (1947) *HAIKU: Volume 1 Eastern Culture*, Hokuseido Shoten, Tokyo.

BRITTON, D. (1980) *A HAIKU Journey: Basho's Narrow Road to a Far Province*, Kodansha International, Tokyo.

DICKAU, R. M. (1996) Bell number diagrams, http://mathforum.org/advanced/robertd/bell.html

GARDNER, M. (1992) *Fractal Music, Hypercards and More ...*, pp. 24–38, Freeman, New York.

HAGIWARA, Y. (1979) *The Narrow Road to Oku*, pp. 7–82, Iwanami Shoten, Tokyo (in Japanese).

ISHIKAWA, T. (2001) *A Handful of Sand; Sad Toys*, Iwanami Shoten, Tokyo (in Japanese).

KAMIYA, Y. (1906) *An Annotated Book on SEDOKAs*, Meiji Shoin, Tokyo (in Japanese).

KEENE, D. (1996) *Oku no Hosomichi: The Narrow Road to Oku*, Kodansha International, Tokyo.

MARUYAMA, K. (1990) *The Collected HAIKU of Issa*, pp. 11–38, Iwanami Shoten, Tokyo (in Japanese).
MATSUNE, T. *et al*. (1967) *A Collection of Modern HAIKU*, pp. 110–119, Chikuma Shobo, Tokyo (in Japanese).
MUTO, S. (1995) *Handbook for Statistical Analysis*, pp. 418–419, Asakura Shoten, Tokyo (in Japanese).
OGATA, T. (1989) *The Collected HAIKU of Buson*, pp. 50–90, Iwanami Shoten, Tokyo (in Japanese).
OHTANI, T. *et al*. (1962) *The Collected HAIKU of Basho*, pp. 17–66, Iwanami Shoten, Tokyo (in Japanese).
SAKAMOTO, H. (2002) *Encyclopedia of Semiotics*, pp. 45–46, Kashiwa Shobo, Tokyo (in Japanese).
TAKAKI, R. (2003) *The Encyclopedia of Form*, pp. 253–254, Maruzen, Tokyo (in Japanese).
YAMAZAKI, T. (1968) *The Collected Poets in the Meiji Era*, pp. 118–128, Chikuma Shobo, Tokyo (in Japanese).
YAMAZAWA, H. (1995) *HAIFU YANAGIDARU*, Vol. 1, pp. 11–54, Iwanami Shoten, Tokyo (in Japanese).