Original Paper *Forma*, **18**, 97–117, 2003

A Comparative Study of Translated Texts through the Analysis of Their Word Spectra: Application to a Text in *Botchan*

Kazuya HAYATA

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

(Received February 21, 2003; Accepted April 9, 2003)

Keywords: Word Spectrum, Statistical Linguistics, Translated Texts, Hellinger Distance, Spiral Mapping

Abstract. Frequency distributions of word-length data are presented for the translated texts of paragraphs sampled from the novel *Botchan* by Soseki Natsume. Languages of the translations are English, German, French, Spanish, Russian, Filipino, Malay, and Indonesian. Divergence between the different spectra for the same language is measured by calculating the Hellinger distance, D_H^2. The results show that irrespective of the languages it maintains the same order of magnitude, specifically $D_H^2 \sim 10^{-2}$. In addition, a method for visualizing the evolution of the data is proposed.

1. Introduction

Statistical approaches to linguistic problems have led to several findings for universal properties of natural languages (CRYSTAL, 1987). For instance, letter frequency analyses of general English texts show that the most frequent letter is ‘e’ whereas the least frequent one is ‘z’ or ‘q’. The rank-ordered statistics for vocabulary in a paper demonstrates the existence of a statistical law, which is currently referred to as Zipf’s law. In these studies statistical data are, respectively, letters and words, both of which belong to the categorial data. In contrast, there are several methods that use the measurable data, one of which is the word-spectrum analysis. Here the term *word spectrum*, which might be borrowed from the terminology of physics, can be defined by the frequency versus the length of words in a text. With these spectra, one can obtain a stylistically important quality of texts, because their profile would depend significantly on the writer’s personality as well as the language. This method was initiated by MENDENHALL (1901), who was a geophysicist. For all texts written by Shakespeare and by Bacon he analyzed their spectra and compared those of the two authors. The main conclusion was that the most frequent word length (i.e., the mode) of the former texts is four, in sharp constrast to three being the mode of the latter. With this analysis the conjecture that Shakespeare might be none other than Bacon was rejected. Indeed the word-spectrum analysis has allowed one to make a comparative study of the statistical structure of texts and has subsequently been applied to a wide range of literary texts (BRINEGAR, 1963; WILLIAMS, 1975).

In this paper the word spectra are obtained for the translated texts of paragraphs sampled from the *Botchan*, which is one of the representative novels written by Soseki Natsume. There are two reasons why this novel was chosen. First, it had been translated into exceptionally many languages. Second, for several languages among them, there are different translations being available. For instance, for English, three translations have been published so far. The languages include five Indo-European ones: English, German, French, Spanish, Russian, and three Austronesian ones: Filipino, Malay, and Indonesian. Divergence between the spectra for the same language is measured by calculating the Hellinger distance, D_{H}^2, which is introduced in the next section. The more are the two spectral distributions different, the larger is the magnitude of D_{H}^2, and $D_{\mathrm{H}}^2 = 0$ solely when they coincide each other. Analyzed results show that irrespective of the languages it maintains the same order of magnitude. In addition, a method for visualizing the evolution of the data is described. With this method one can trace the process of creating the word spectrum.

It should be mentioned that both the mapping procedure and the combined use of the word spectra and the two-dimensional projection are, to the best of the author's knowledge, novel and thus proposed in this paper.

2. Outline of Analysis Method

2.1. Word spectrum

As an original text to be translated the opening two paragraphs of the *Botchan* (NATSUME, 1929) are selected; it is composed of 492 Japanese syllabographs (being called KANAs in Japanese). For this Japanese text three translations into English are currently available. As an example the beginning sentence of the text from the translation by MORRI (1918) is presented:

{Because of an hereditary recklessness,
I have been playing always a losing game since my childhood.}.

For this sentence a series of the word-length data become

$$\{7, 2, 2, 10, 12, 1, 4, 4, 7, 6, 1, 6, 4, 5, 2, 9\}, \tag{1}$$

where the word length (L) is defined by the number of letters composing a word. Sentences of other translations are shown in Fig. 1. Subsequent counting the frequency of each word length over the entire text consisting of 260 words in the Morri's translation yields the word spectrum as shown in Fig. 2(a).

2.2. Hellinger distance

The divergence between spectral distributions is measurable with the Hellinger distance, $D_{\mathrm{H}}^2(p|q)$, which is given by

$$D_{\mathrm{H}}^2\left(p|q\right) = \sum_{i=1}^{n}\left(p_i^{1/2} - q_i^{1/2}\right)^2 \tag{2}$$

＜坊っちゃん＞親譲りの無鉄砲で子供の時から損ばかりしている。

<Bottyan> Oyayudurino Muteppô de Kodomo no Toki kara Son bakari site iru.

<Botchan> Oyayuzurino Muteppô de Kodomo no Toki kara Son bakari shite iru.

<Botchan: Morri> Because of an hereditary recklessness, I have been playing always a losing game since my childhood.

<Botchan: Sasaki> A great loser have I been ever since a child, having a rash, daring spirit, a spirit I inherited from my ancestors.

<Botchan: Turney> Ever since I was a child, my inherent recklessness has brought me nothing but trouble.

<Botchan: Spann> Aus anererbter Unbesonnenheit spiele ich seit meinen Kindheitstagen eine gar klägliche Rolle.

<Der Tor Aus Tokio: Berndt & Shinohara> Da ich von Natur aus unbesonnen bin, habe ich seit meiner Kindheit stets ein verlorenes Spiel gespielt.

<Botchan: Ogata> Dès ma naissance j'étais destiné à une vie desavantageuse à cause de mon caractère irréfléchi comme celui de mon père.

<Botchan: Morita> Pour tout héritage, j'ai reçu une nature impulsive et risque-tout qui me vaut depuis ma petite enfance de perpétuelles mésaventures.

<Botchan: Valles> Por haber heredado de mis padres la temeridad, desde la niñez no he tenido más que contratiempos.

<Botchan: Izquierdo> Este afán temerario que me marca como herencia paterna viene dándome problemas desde mi niñez.

＜Барчук：Григорева＞С детских лет я только и делаю, что вrежу самому себе. А все иэ-эа своего беэрассудства, унаследованного от родного папаши.

<Botchan: Antonio> Sapul sa pagkabata, ang likas kong kawalan ng ingat ay walang ibinigay sa akin kundi problema.

<Botchan: Ahmad> Sejak aku masih kecil lagi, sifat cuaiku sentiasa saja mendatangkan kesulitan kepada diriku.

<Botchan: Sarana> Sejak masih kecil, watakku yang selalu nekad itu hanya menyulitkan diriku sendiri saja.

Fig. 1. Polyglot description for the opening sentence of the *Botchan*.

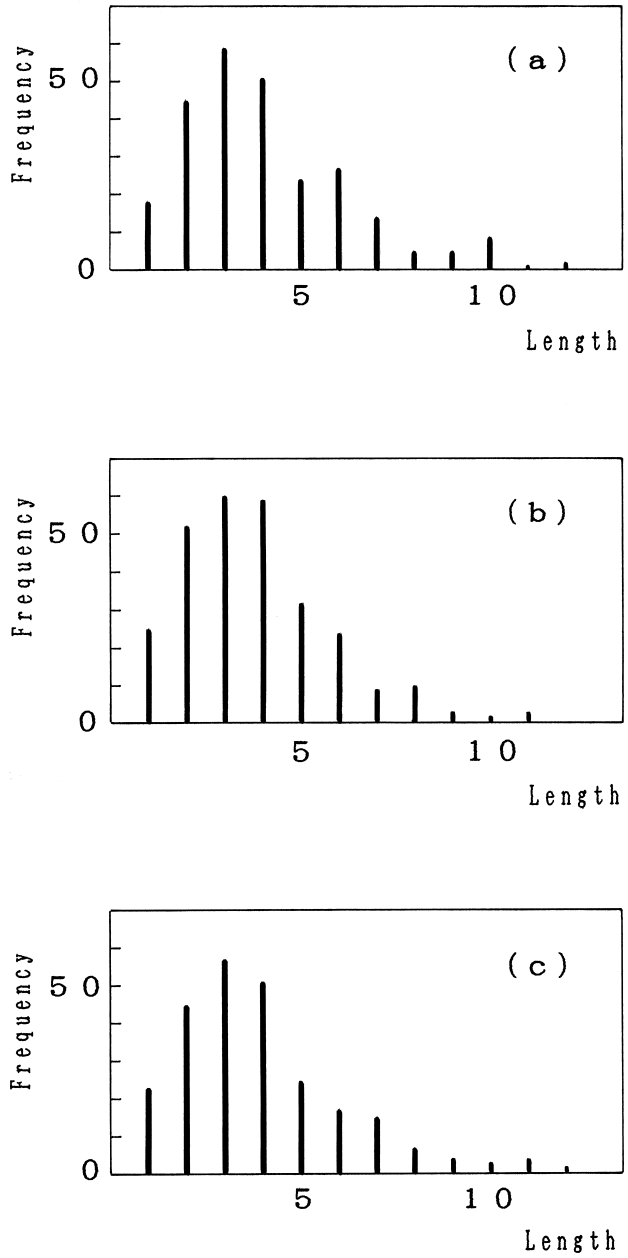


Fig. 2. Word spectra of the English texts by (a) Morri, (b) Sasaki, and (c) Turney.

with

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} q_i = 1. \tag{3}$$

Here p_i and q_i (i = 1, 2, 3, ..., n) represent the relative frequencies for the length $L = i$, and n is the maximum word-length (n = 12 for the English texts).

2.3. Spiral mapping

In order to visualize the sequence of word-length data a method that uses a spiral mapping technique is proposed, the prototypal concept of which was presented previously by the author (HAYATA, 1999). In this method, from the center (0,0) to the outermost orbit, a spiral pattern with the counterclockwise rotation is depicted in accordance with the direction of a sequence. For instance, for the sequence given by Series (1) the trajectory is determined as

$$\begin{aligned}&(0,0) \to (7,0) \to (7,2) \to (5,2) \to (5,12) \to (-7,12) \to (-7,11)\\ &\to (-11,11) \to (-11,\ 7) \to (-18,7) \to (-18,1) \to (-19,1)\\ &\to (-19,-5) \to (-15,-5) \to (-15,-10) \to (-13,-10) \to (-13,\ -19).\end{aligned}$$

3. Results and Discussion

To date *Botchan* has been translated into eight languages, which are English (3), German (2), French (2), Spanish (2), and Russian (1); more recently translations into Filipino (1), Malay (1), and Indonesian (1) have been made. Here the number in the bracket indicates that of translation(s) published so far. For comparative purposes their opening sentences are presented in Fig. 1. In this section the word spectra for these translations are shown and compared through calculation of the Hellinger distance between the spectral distributions.

3.1. English texts

There are three translations currently available. First MORRI (1918) has translated the *Botchan* into English. Subsequently the translation by SASAKI (1968) has been published (note that its original version was in 1922). More recently translation has been made by TURNEY (1972). The spectra of these translated texts are shown in Fig. 2. In the plots, first it can be seen that the three distributions exhibit a common feature with a single peak at L = 3 beyond which the frequency decays with a relatively long tail; with the terminology in descriptive statistics their shape can be identified by a positively distorted distribution because of the positive skewness (normalized third-order moment). Typical characteristic values of the data are listed in Table 1, where Σ, $\overline{L}$, Me, and Mo are, respectively, the summation of words in each text, the mean, the median, and the mode of the word length; R, s, and CV are, respectively, the range, the standard deviation, and the coefficient of variation. It is found from this table that the macroscopic feature of the spectra is qualitatively the same and thus that there exists no remarkable difference between the three

Table 1. Characteristic values for the word-length data of the three English texts. Here Σ, $\bar{L}$, Me, Mo, R, s, and CV indicate, respectively, the summation of words, the mean, the median, the mode, the range, the standard deviation, and the coefficient of variation (CV = s/ $\bar{L}$) of the data.

	Morri	Sasaki	Turney
Σ	260	280	253
$\bar{L}$	4.15	3.88	3.99
Me	4	4	4
Mo	3	3	3
R	11	12	11
s	2.29	2.10	2.28
CV	0.551	0.542	0.571

Table 2. Parity distribution (%) for the word-length data of the English texts.

	Morri	Sasaki	Turney
Odd	46.5	47.5	50.6
Even	53.5	52.5	49.4

Table 3. The characteristic values for the word-length data of the two styles with Roman letters.

	Japanese	Hepburn
Σ	196	196
$\bar{L}$	4.42	4.52
Me	4	4
Mo	2	2
R	9	10
s	2.23	2.40
CV	0.505	0.531

distributions. Indeed, the spectral profiles shown in Fig. 2 are akin to that obtained by MENDENHALL (1901) for the Bacon's texts. However, careful observation of the results manifests a nontrivial difference in the variation of the frequency. In Fig. 2(c) the curve varies smoothly with increasing L, whereas in Fig. 2(a) it shows a considerable fluctuation with local maxima at L = 6 and 10; the result shown in Fig. 2(b) is intermediate between the two. Next, to extract an additional difference between the translated texts the parity distribution of the word lengths was calculated, where the even (odd) parity indicates the cummulative frequency of the even (odd) lengths. The results are summarized in Table 2. It is interesting to note that for the texts by Morri and Sasaki the parity distribution deflects toward the even side, in contrast to the deflection toward the odd side for the Turney's text. Here it should be remembered that both Morri and Sasaki are Japanese while Turney is a native speaker of English. To discuss this problem in more detail, word-spectrum analysis

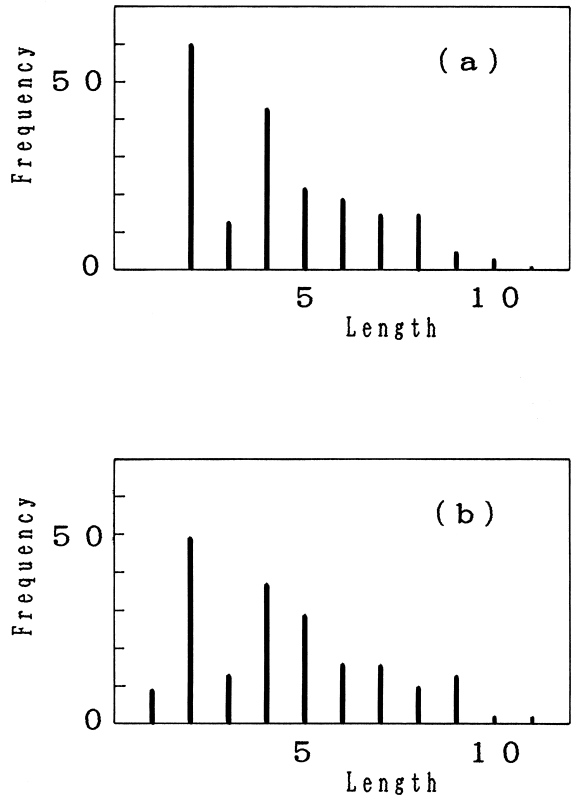


Fig. 3. Word spectra of the *Romaji* texts written with (a) the Japanese and (b) the Hepburn style.

was made for Japanese texts written in Roman letters. As specific expressions the Japanese and the Hepburn styles were chosen. Here, as an example text for the ROMAJI (an expression of Japanese texts by Roman letters) movement the former had been used for the previous publication (NATUME, 1922). The results of their word spectra, characteristic values, and parity distributions are presented in Fig. 3, Tables 3 and 4, respectively. It is evident from Table 4 that independent of the style the parity of the word lengths exhibits substantial deflection toward the even side; note that this feature is much more remarkable for the Japanese style. In comparison between Tables 2 and 4 a hypothesis could be posed that English of the Japanese translators is affected by the property of their native language.

The Hellinger distances obtained for the English texts are presented in Table 5. It can be seen that the distance between the Morri's and the Sasaki's spectra becomes approximately two times larger than that between the Morri's and the Turney's ones. Note that calculation for the two ROMAJI texts (Fig. 3) yields $D_H^2 = 6.66 \times 10^{-2}$, the value of which is found to be about two times larger than that between the Morri's and the Sasaki's distributions for the English texts.

Table 4. Parity distribution (%) for the word-length data of the texts written with Roman letters.

	Japanese	Hepburn
Odd	28.6	41.8
Even	71.4	58.2

Table 5. A matrix representation of the Hellinger distances ($\times 10^{-2}$) for the word-spectral distributions of the English texts.

	q		
	Morri	Sasaki	Turney
p			
Morri	0	3.38	1.67
Sasaki	3.38	0	1.97
Turney	1.67	1.97	0

Table 6. The characteristic values for the word-length data of the two German texts.

	Spann	Berndt
Σ	261	268
$\overline{L}$	4.98	4.94
Me	4	4
Mo	3	3
R	14	14
s	2.82	2.45
CV	0.566	0.496

Table 7. Parity distribution (%) for the word-length data of the German texts.

	Spann	Berndt
Odd	57.9	57.8
Even	42.1	42.2

In Fig. 4 the spiral patterns for the three English texts are depicted. Through comparing the three illustrations it can be concluded that, although the way of interference between adjacent orbits depends strongly on the statistical property of texts, the total areas of these maps do not exhibit a pronounced difference. For comparison, also shown in Fig. 5 are the maps for the ROMAJI texts.

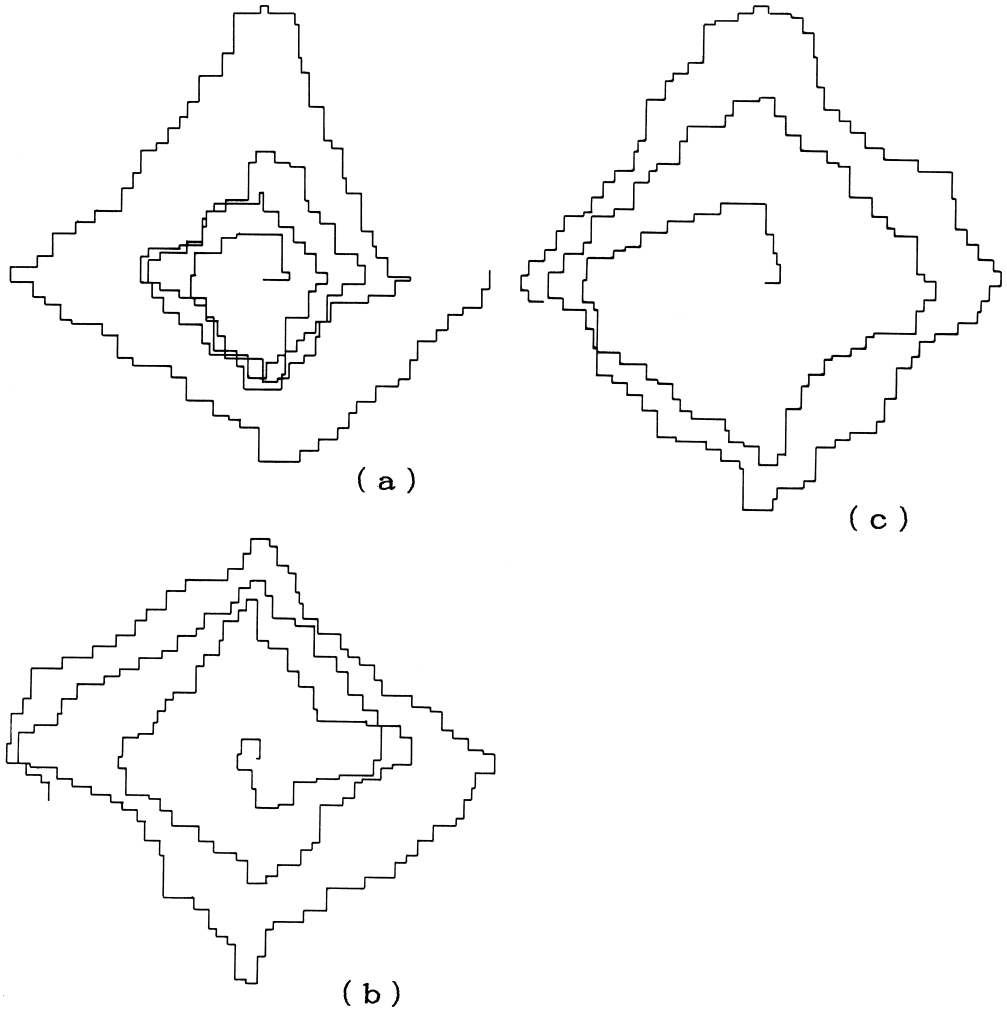


Fig. 4. Spiral orbit for Fig. 2(a), (b), and (c), respectively.

3.2. German texts

The translations of the *Botchan* into German have been made by SPANN (1925) and more recently by BERNDT and SHINOHARA (1990). The word spectra, the two-dimensional spiral maps, the characteristic values of the word-length data, and their parity distributions are shown in Figs. 6 and 7, Tables 6 and 7, respectively. The Hellinger distance between the two spectral distributions shown in Fig. 6 is $D_H^2 = 3.07 \times 10^{-2}$ with $n = 16$. Here it would be interesting to note that this value has the same order of magnitude as those obtained for the English texts (see Table 5). The widths of the spectral distributions, which can be represented with R, s, and CV, become larger than those observed for the English texts (Table 1). This is attributable to the multisyllable nature of German.

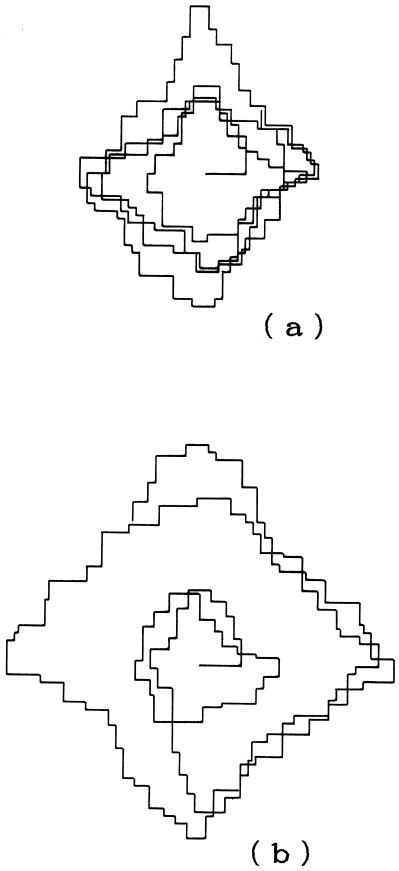


Fig. 5. Spiral orbit for Fig. 3(a) and (b), respectively.

3.3. French texts

The *Botchan* has been translated into French initially by OGATA (1923). Recently translation by MORITA (1993) has been published. The word spectra, the two-dimensional spiral maps, the characteristic values of the word-length data, and their parity distributions are shown in Figs. 8 and 9, Tables 8 and 9, respectively. The Hellinger distance between the two spectral distributions shown in Fig. 8 is $D_H^2 = 3.47 \times 10^{-2}$ with $n = 14$. Again, it would be worth noting that this value has the same order of magnitude as those obtained for the English texts (see Table 5) and is comparable to that obtained for the German texts. The widths of the spectral distributions are found to be larger than those observed for the English texts (Table 1).

3.4. Spanish texts

About a half century later from the first translation into English, *Botchan* has been translated into Spanish (VALLES, 1969). Only recently has another translation been

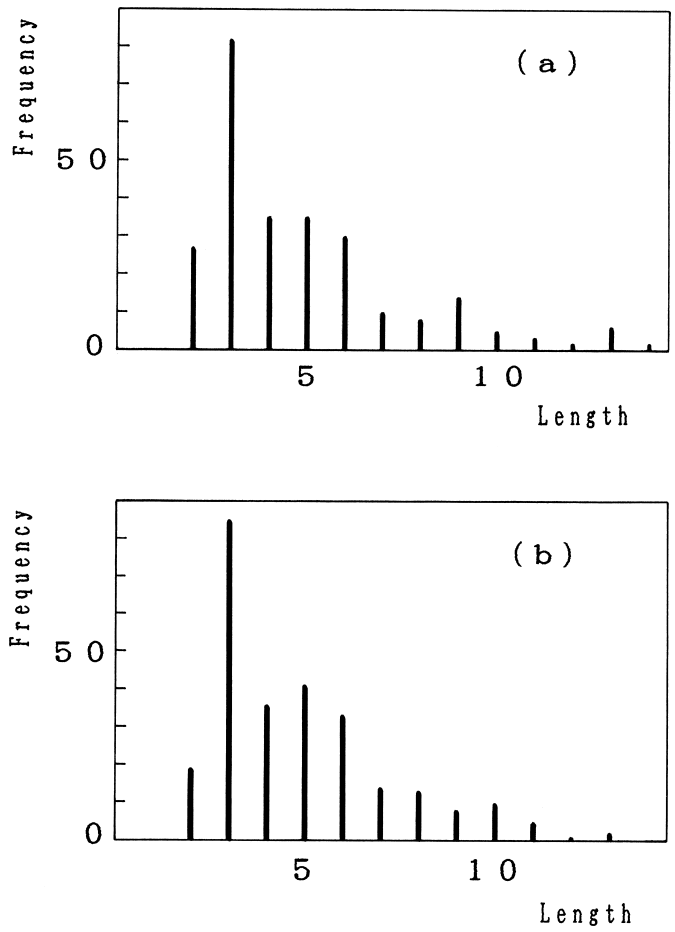


Fig. 6. Word spectra of the German texts by (a) Spann and (b) Berndt and Shinohara.

published (RODRÍGUEZ-IZQUIERDO, 1997). The analyzed results for these texts are shown in Figs. 10 and 11 and in Tables 10 and 11. The Hellinger distance between the two spectra shown in Fig. 10 is $D_H^2 = 4.00 \times 10^{-2}$ ($n = 15$), which is comparable to those calculated for the German and the French texts. Comparison between the Spanish (Fig. 10) and the French (Fig. 8) spectra shows that in both cases they peak steeply at $L = 2$ (Mo = 2) and that their profiles have some common features. Such similarity as seen in the French and the Spanish spectra is due to the linguistically established fact that both languages bifurcated from a common trunk of the Indo-European family, termed Romance in modern linguistics.

3.5. Russian text

For this language, only the translation by GRIGORYEV (1943) is available. Thus, the comparative study as has been done for the other languages is impossible. The results

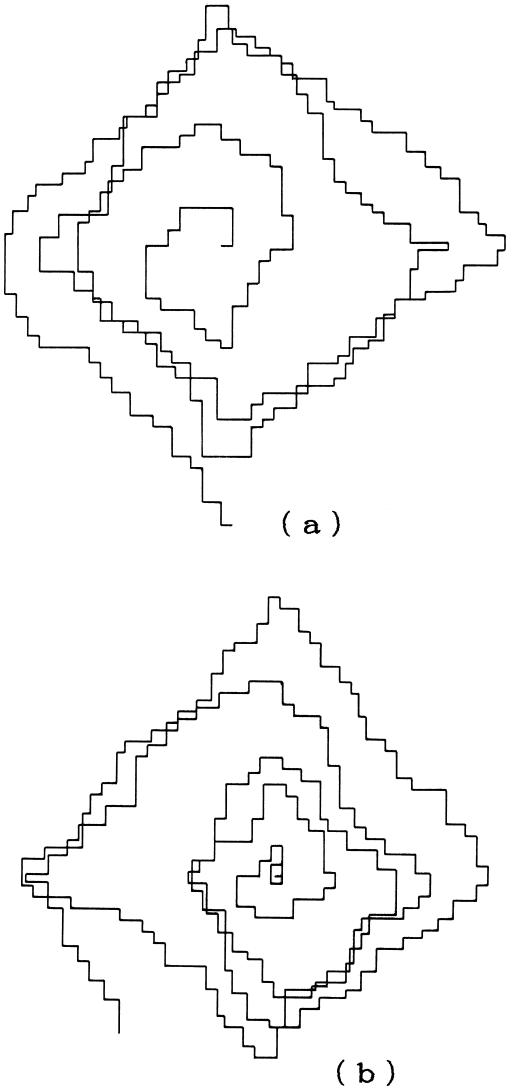


Fig. 7. Spiral orbit for Fig. 6(a) and (b), respectively.

obtained for the Russian text are presented in Figs. 12 and 13. The spectral distribution in Fig. 12 attracts our attention, because it has a curious shape with twin humps observed at $L = 2$ and 6.

3.6. Filipino, Malay, and Indonesian texts

In recent years translations of the *Botchan* into the three kinds of non-European languages have been successively published (AHMAD, 1989; ANTONIO, 1991; SARANA,

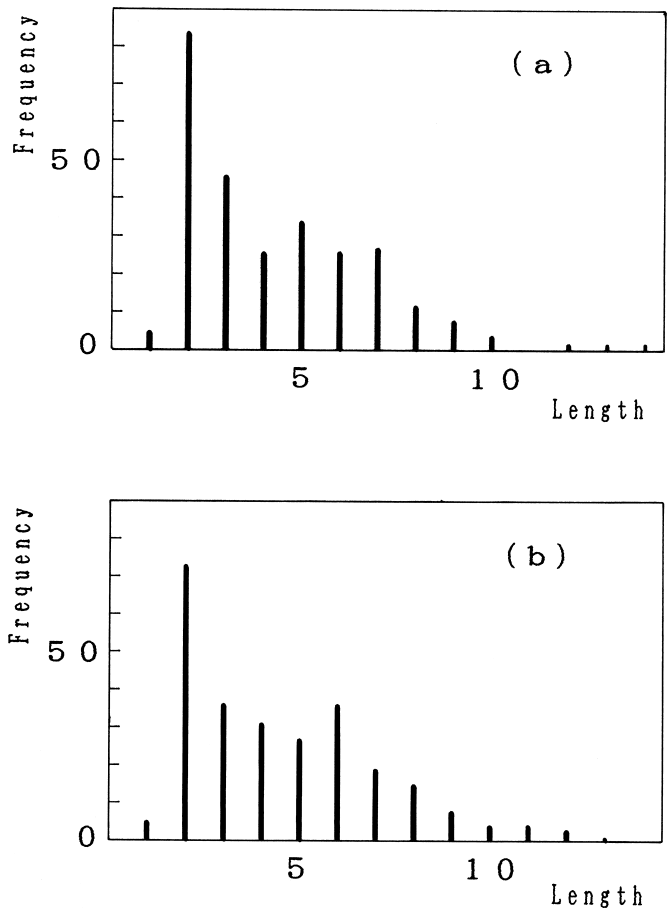


Fig. 8. Word spectra of the French texts by (a) Ogata and (b) Morita.

1992), though all of the three texts are not the translations from the original text but are those from the English text translated by Turney. Linguistically these languages are categorized into the same family termed Austronesian. In this subsection, analyzed results for the three texts are presented and compared through calculations of the Hellinger distances as well as through depictions of the two-dimensional spiral patterns. First, the word spectra of the translated texts are illustrated in Fig. 14 with their two-dimensional maps, Fig. 15. Subsequently, for each spectral distribution the characteristic values of the word-length data, the parity distributions, and the Hellinger-distance matrix (n = 17), respectively, are shown in Tables 12–14. In order to discuss differences between the spectra, attention is focused on Table 14. Herein it is interesting to note that the distance between the Malay (AHMAD, 1989) and the Indonesian (SARANA, 1992) becomes minimum, the order-of-magnitude of which coincides with those obtained for the European texts. In

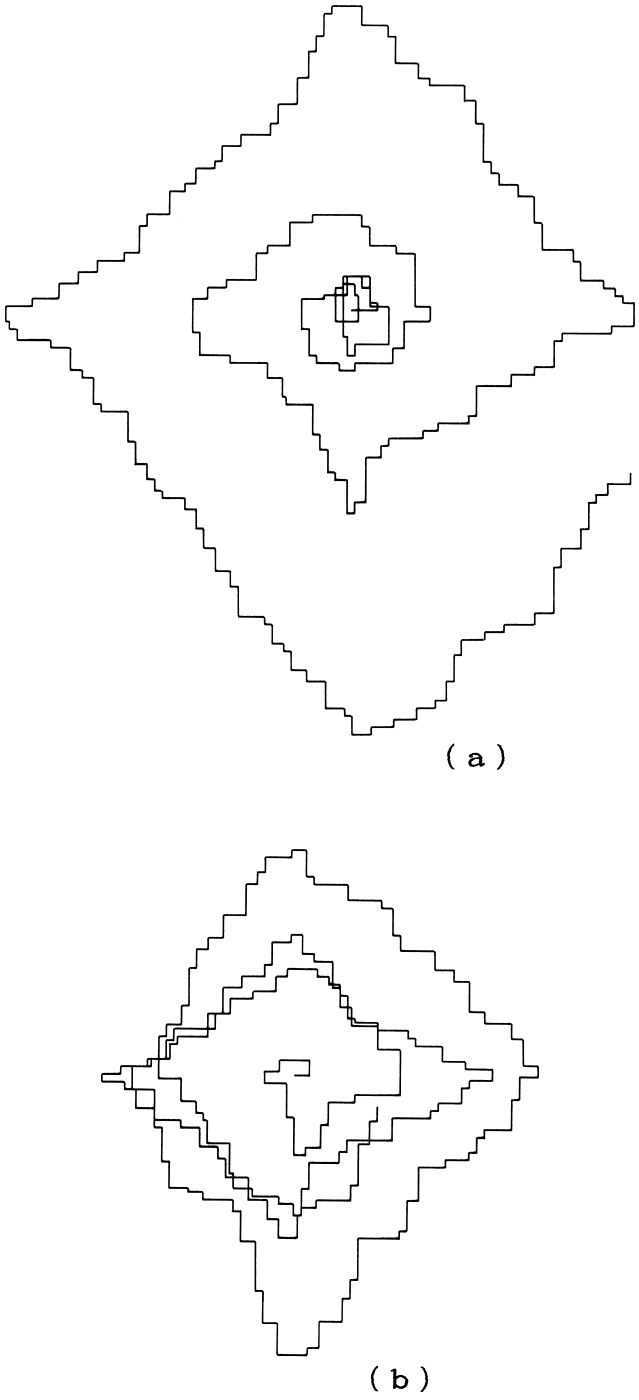


Fig. 9. Spiral orbit for Fig. 8(a) and (b), respectively.

Table 8. The characteristic values for the word-length data of the two French texts.

	Ogata	Morita
Σ	278	262
$\bar{L}$	4.37	4.55
Me	4	4
Mo	2	2
R	13	12
s	2.54	2.54
CV	0.581	0.558

Table 9. Parity distribution (%) for the word-length data of the French texts.

	Ogata	Morita
Odd	43.9	38.2
Even	56.1	61.8

Table 10. The characteristic values for the word-length data of the two Spanish texts.

	Valles	lzquierdo
Σ	241	300
$\bar{L}$	4.47	4.31
Me	4	4
Mo	2	2
R	14	12
s	2.69	2.38
CV	0.602	0.552

Table 11. Parity distribution (%) for the word-length data of the Spanish texts.

	Valles	Izquierdo
Odd	47.3	45.7
Even	52.7	54.3

contrast, other combinations, i.e., the distance between the Filipino (ANTONIO, 1991) and the Malay and that between the Filipino and the Indonesian, show that $D_H^2 \sim 10^{-1}$, which is found to be an order of magnitude larger than those presented in the preceding subsections. The reason is described in what follows: first, it should be noted that linguistically the Malay is much akin to the Indonesian. Indeed, they say that the latter may be included in the former. Because of this property the distance between the two spectra

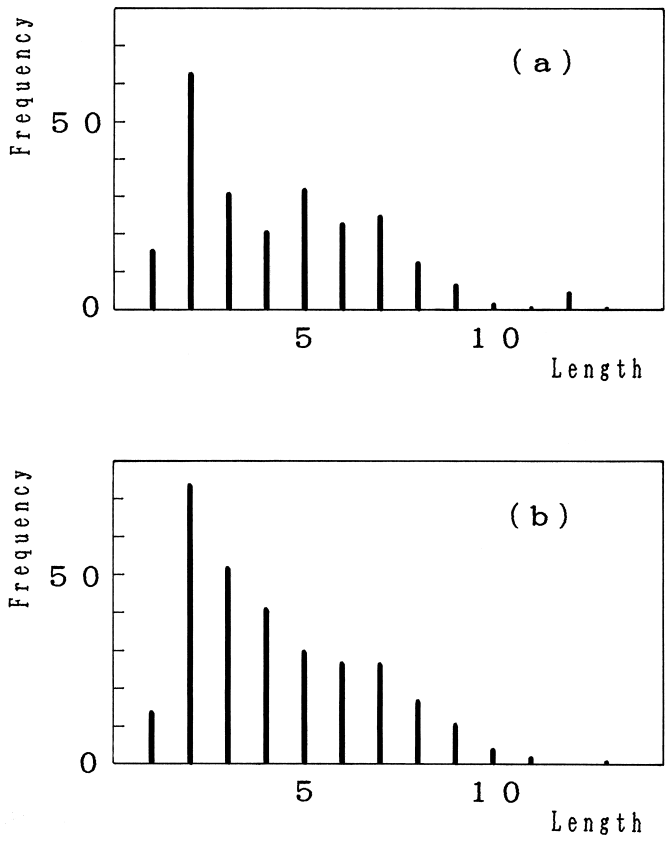
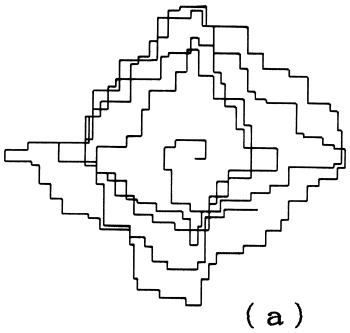


Fig. 10. Word spectra of the Spanish texts by (a) Valles and (b) Rodríguez-lzquierdo.

Table 12. The characteristic values for the word-length data of the three Austronesian texts.

	Filipino	Malay	Indonesian
Σ	241	259	229
$\bar{L}$	4.83	5.66	5.68
Me	4	5	6
Mo	2	6	4
R	11	15	10
s	2.49	2.31	2.00
CV	0.515	0.408	0.352

results in the same order of magnitude that has been obtained for two different texts written in an identical language. The results in Fig. 15 are consistent with those presented in Table 14; visually the evolutional pattern for the Malay (Fig. 15(b)) is akin to that for the Indonesian (Fig. 15(c)).



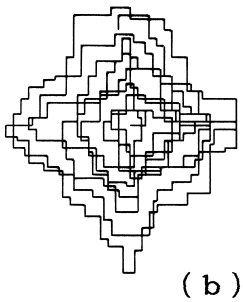


Fig. 11. Spiral orbit for Fig. 10(a) and (b), respectively.

Table 13. Parity distribution (%) for the word-length data of the Austronesian texts.

	Filipino	Malay	Indonesian
Odd	45.2	49.4	47.2
Even	54.8	50.6	52.8

Table 14. The Hellinger-distance matrix ($\times 10^{-2}$) for the word-spectral distributions of the Austronesian texts.

	q		
	Filipino	Malay	Indonesian
p			
Filipino	0	17.3	15.4
Malay	17.3	0	2.51
Indonesian	15.4	2.51	0

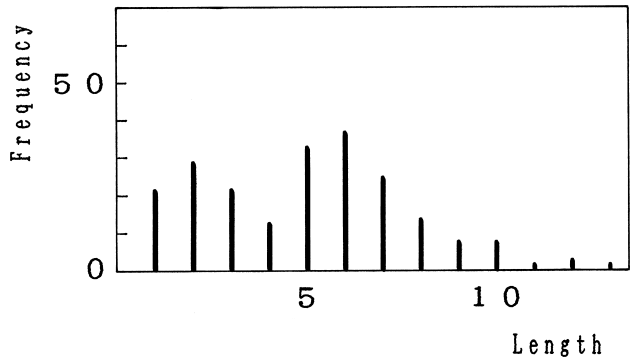


Fig. 12. Word spectra of the Russian text. The characteristic values are $\Sigma = 220$, $\bar{L} = 5.13$, Me = 5, Mo = 6, R = 14, s = 2.82, and CV = 0.550; the parity distribution (%) becomes (Odd, Even) = (52.7, 47.3).

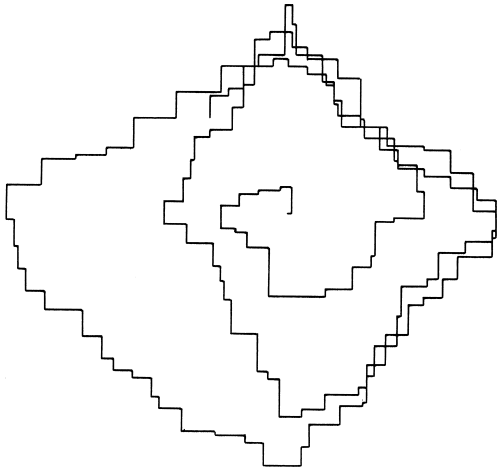


Fig. 13. Spiral orbit for Fig. 12.

4. Conclusions

The frequency distributions of word-length data have been presented for the translated texts of paragraphs sampled from the *Botchan*. All the translations being currently available have been considered. Divergence between the spectra for the same language has been measured by means of the Hellinger distance, D_H^2. Analyzed results have shown that for the identical language the spectral profile does not exhibit substantial difference between the translations. This indicates that translators are bound unconsciously by a statistical 'grammer' of the language into which they translated the original text. Furthermore,

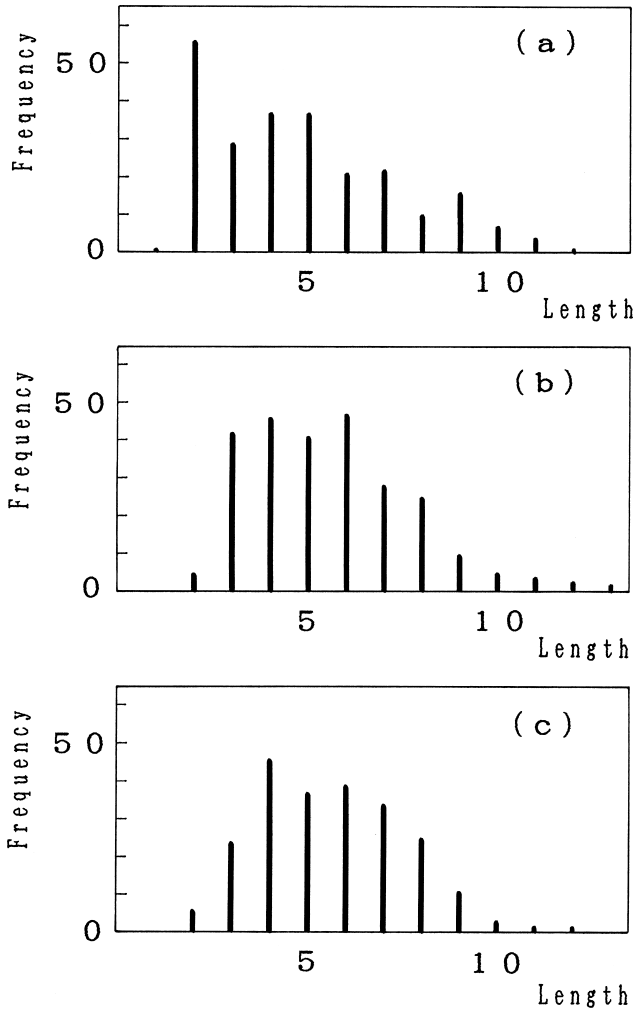


Fig. 14. Word spectra of the texts written in (a) Filipino, (b) Malay, and (c) Indonesian.

the parity analysis of the English texts has posed the interesting hypothesis that translators would also be bound by the property of their native language. Calculations of the Hellinger distances between the spectral distributions of different translations have shown that irrespective of the languages it maintains the same order of magnitude, specifically $D_H^2 \sim 10^{-2}$. In addition, a method for visualizing the evolution of the data has been proposed and applied to analyzing statistical properties of the texts. Through comparison among spiral orbits it has been found that their main features are not dependent pronouncedly on the translations but are determined by the property of each individual language, though the way of interference between orbits is strongly dependent on the texts.

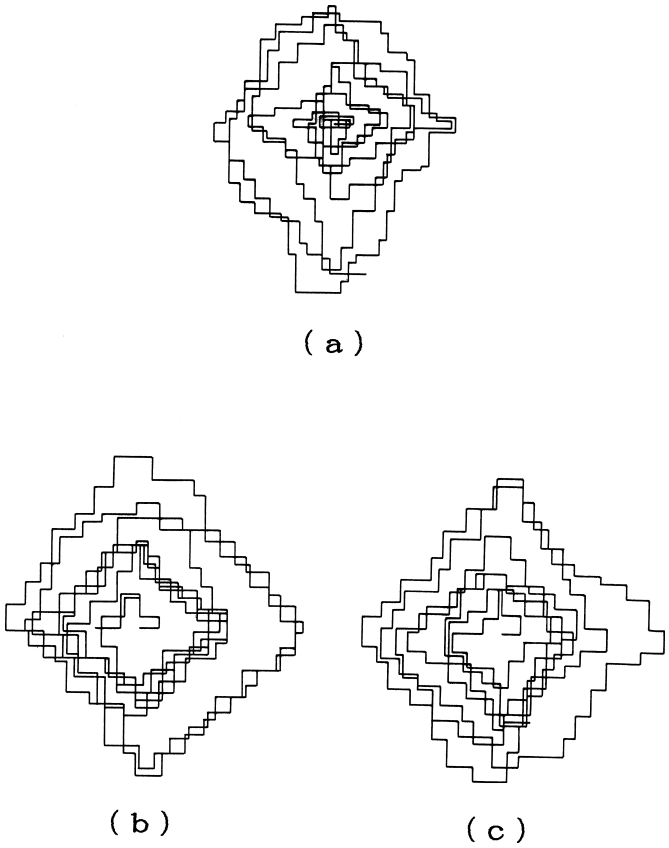


Fig. 15. Spiral orbit for Fig. 14(a), (b), and (c), respectively.

This work was supported in part by the Sapporo Gakuin University (SGU) Research Support Grant (No. SGUS0119600512) and a Scientific Research Project Grant-in-Aid from the Department of Socio-Informatics, SGU.

REFERENCES

AHMAD, M. (1989) *Natsume Soseki Botchan,* Dewan Bahasa dan Pustaka, Kuala Lumpur, pp. 1–2.
ANTONIO, L. F. (1991) *Natsume Soseki Botchan,* Solidarity Foundation, Manila, p. 11.
BERNDT, J. and SHINOHARA, S. (1990) *Natsume Soseki Der Tor aus Tokio,* Theseus Verlag, München, p. 5.
BRINEGAR, C. S. (1963) Mark Twain and the Quintus Curtius Snodgrass Letters: a statistical test of authorship, *J. Amer. Statist. Assoc.*, **58**, 85–96.
CRYSTAL, D. (1987) *The Cambridge Encyclopedia of Language,* Cambridge UP, Cambridge.
GRIGORYEV, M. P. (1943) *Sooseki Natume Bartsuk (Bottyan),* Mantetu, Dairen, pp. 11–12 (in Russian).
HAYATA, K. (1999) Visualizing musicality in a sequence of titles: a critical region between disorder and order in cases of *the Tale of Genji* and the Films of Akira Kurosawa, *Bulletin of the Society for Science on Form,* **14**, No. 1, 42–43 (in Japanese).

MENDENHALL, T. C. (1901) A mechanical solution of a literary problem, *Popular Sci. Monthly*, **60**, 96–105.
MORITA, H. (1993) *Natsumé Sôseki Botchan*, Le Serpent a Plumes, Paris, pp. 7–8.
MORRI, Y. (1918) *Botchan (Master Darling), by the Late Mr. Kin-nosuke Natsume*, Ogawa Seibundo, Tokyo, pp. 1–2.
NATSUME, S. (1929) *Botchan*, Iwanami Shoten, Tokyo, p. 3 (in Japanese).
NATUME, S. (1922) *Bottyan*, Iwanami Syoten, Tookyoo, pp. 1–2.
OGATA, N. (1923) *Natsumé Sôseki Botchan (Jeune Homme Irréfléchi)*, Maruzen, Tookyoo, pp. 1–2.
RODRÍGUEZ-IZQUIERDO, F. (1997) *Botchan, Soseki Natsume*, Luna Books, Kamakura, pp. 9–10.
SARANA, Y. K. (1992) *Natsume Soseki Botchan*, Universitas Katolik Soegijapranata, Semarang, p. 5.
SASAKI, U. (1968) *Botchan, by Soseki Natsume*, Charles E. Tuttle, Tokyo, pp. 13–14.
SPANN, A. (1925) *Botchan (Ein Reiner Tor) von Kin-no-suke Natsume*, Kyodo-Verlag, Osaka, pp. 1–2.
TURNEY, A. (1972) *Botchan, Natsume Soseki*, Kodansha International, Tokyo, p. 9.
VALLES, J. G. (1969) *Botchan (El Joven Mimado), Soseki Natsume*, Sociedad Latino-Americana, Tokyo, pp. 1–2.
WILLIAMS, C. B. (1975) Mendenhall's studies of word-length distribution in the works of Shakespeare and Bacon, *Biometrika*, **62-1**, 207–212.