Table 3. Hellinger distance, D_{H^2} ($\times 10^{-3}$), between word spectra of English texts. To make a comparison the Poisson distribution with $\lambda = 3.8464$ is also included.

	Poissonian	Sasaki	Turney	Cohn
Poissonian	0	45.5	38.8	40.7
Sasaki	45.5	0	1.02	1.06
Turney	38.8	1.02	0	0.608
Cohn	40.7	1.06	0.608	0

frequent word length (i.e., the mode) of the former texts is four, in contrast 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 property of texts and has subsequently been applied to a wide range of literary texts (Brinegar, 1963; Williams, 1975). In Fig. 6 the word spectra, i.e., the statistical probabilities of words with length x, are shown of the famous Japanese novel *Botchan* (Work #N1) that was translated into English (Sasaki, 1968; Turney, 1972; Cohn, 2005). Here the length of a word is defined with the number of letters in it. First, it can be seen that the overall profile of the English spectrum bears a resemblance to the Poisson distribution already plotted in Fig. 2(b). For this reason, in Figs. 6(a)–(c) the spectrum of the Poisson distribution that meets Eq. (2) ($\lambda = 3.8464$) is juxtaposed with fine lines. The divergence between two spectra can be quantified through calculation of the Hellinger distance D_{H^2} (≥ 0; equality holds for the perfect similarity)

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

with

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

Here p_i and q_i ($i = 1, 2, 3, \ldots, n$) represent the relative frequencies for the length $x = i$, and n is the maximum word-length. For the results shown in Fig. 6, the distances have been calculated for all combinations of the spectra (Table 3), where the smallest divergence is seen between the spectrum of Turney (Fig. 6(b)) and that of Cohn (Fig. 6(c)). In contrast to this case, the largest divergence can be seen between the two spectra drawn with the bold and the fine lines in Fig. 6(a), namely

$$D_{H^2} = 4.55 \times 10^{-2}.$$

It has been confirmed that this value is comparable to that between the Spanish and the Filipino text of the same novel, which becomes $D_{H^2} = 4.09 \times 10^{-2}$. Here we should remember the linguistic fact that for historical reasons Filipino has a considerable part of the vocabulary in common with that of Spanish. Characteristic values of the word-length data as well as the results of r_F are summarized in Table 4, along with those of the Poisson distribution, where the relative difference between r_F and $1/\phi$ is defined by

$$\delta = \phi \left| r_F - \phi^{-1} \right|.$$

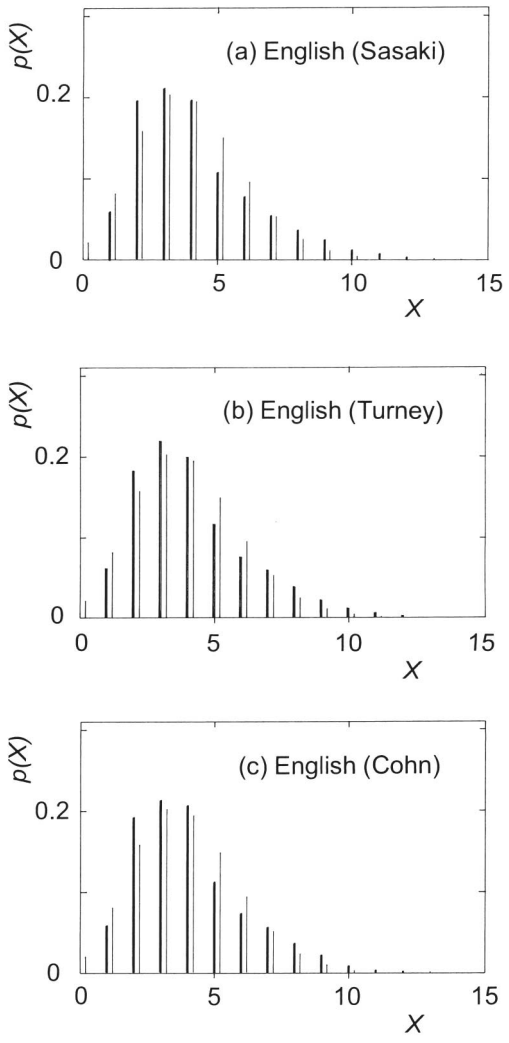

Fig. 6. Word spectrum of the famous Japanese novel *Botchan* (Work #N1) that was translated into English by (a) Sasaki (1968; first printed in 1922), (b) Turney (1972), and (c) Cohn (2005). Irrespective of the translators, a single peak with a positively distorted shape (i.e., positive skewness) is seen. There are steep walls between $x = 1$ and $x = 2$ as well as between $x = 4$ and $x = 5$. To make a comparison, the spectrum of the Poisson distribution with $r_F = 1/\phi$ ($\lambda = 3.8464$) is superimposed with fine lines.

In Table 4 one will notice the interesting fact that the magnitude of r_F for all the English texts are extremely close to $1/\phi$. In particular, it would be surprising that for the text translated by Sasaki the magnitude of δ is no more than 0.03%.

There are two reasons why Work #N1 was chosen. First, it had been translated into exceptionally many languages. Second, for several languages among them, there are different translations being available. Calculation has been made also for non-English texts currently available. In Fig. 7 the word spectra are shown of *Botchan* (#N1) that was translated into (a) Italian (Pastore, 2007), (b) Polish (Murakami, 2009), (c) Hungarian (Judit, 2003), and (d) Indonesian (Haryono, 1992); their characteristic values are listed in Table 5. Here we notice that for the Italian text (Fig. 7(a)) the magnitude of r_F would be close to $1/\phi$. In addition to the five languages listed in Tables 4 and 5, analyses