*Glottometrics 15, 2007, 28-44*

# On distributions of sentence lengths in Japanese writing

*Motohiro Ishida, Tokushima*
*Kazue Ishida, Tokushima*

**Abstract.** The lognormal distribution had long been thought to be the most appropriate probability distribution for Japanese sentence length distributions. Yet this view had been supported only by few researches with sparse sampling data and reasoning contradicting language reality. In order to show a more realistic approach, we analyzed a substantial number of samples. At first, 150 essays and short stories were drawn as a random sample, out of which any pieces of writing whose length was either less than 100 or more than 1000 sentences were excluded. As a result, 113 pieces remained as sample texts. We also paid attention to the kinds of sentences, separating those of dialogue from narrative ones. From each one of these 113 sample texts, three sentence length frequency distributions were acquired – the first one for a complete text, the second one for the collection of direct speech in the same text, and the third one for all the narrative parts excluding direct speech above. The results completely overturn the long-standing belief, proving that a lognormal distribution – which has been computed but will not be shown here – can never be well applied to Japanese sentence length distributions. Our new findings indicate that in place of this lognormal distribution, the Hyperpascal distribution maintains an excellent goodness of fit.

*Keywords: Sentence length, Japanese*

## 1 Introduction

It has already been forty years since Yasumoto (1965, 1966) analyzed twenty sentences from each of 100 Japanese novels, judging that Japanese sentence lengths correspond either with a lognormal distribution or with a gamma distribution. Sasaki (1976) also examined 1500 sentences in total which were evenly extracted from three Japanese novels. The result was to corroborate Yasumoto's conclusion, with one of the three novels following a gamma distribution and the other two following a lognormal distribution. There is one more article in which Arai (2001) argues, with some of the literary works by Ryunosuke Akutagawa and Osamu Dazai as samples, that Japanese sentence lengths follow a lognormal distribution. In Europe and America, on the other hand, studies of the same kind have been conducted (Yule 1939; Williams 1940; Fucks 1968; Sichel 1974, 1975; Sigurd and Eeg-Olosson 2004; Kjetsaa 1978; Altmann 1988, 1992; Grotjahn and Altmann 1993; Niehaus 1997; Strehlow 1997; Wittek 1995; Kelih and Grzybek 2004, 2005; for more literature see http://lql.uni-trier.de). They have attempted to find models of sentence lengths in English, German, Chinese, Russian, Classical Greek, and Slovak, and used the negative binomial distribution, the Hyperpascal distribution, the Hyperpoisson distribution, a modified positive Poisson distribution, a compound Poisson distribution, and the lognormal distribution respectively. On the basis of this preceding research, we have analyzed all the sentences of 113 works by thirty-six Japanese writers. The result of our investigation into Japanese sentence length distributions follows.

Before we present the data, some theoretical preliminaries should be reviewed. The lognormal distribution has been introduced into linguistics on physical grounds. Since in the nature many phenomena are normally distributed, the first researchers supposed the same would hold for language. But "normality" contradicts the self-organisatory character of language, and in most cases also its self-regulatory character. The *speakers* try to render every entity as easy for them as possible (memory effort, coding effort, production effort, etc.). They try to adapt the language to their own needs. Hence everything must be skewed, deviating from "normality." It is the self-regulation (exerted by the *hearer*) that stops great deviations and "pulls" them back again, but never to the "normal" state, because language must develop. Thus non-normality is the natural state of any linguistic phenomenon. The first researchers realized this fact but in an attempt to maintain the connection to physics, they modified the normal distribution in a way which is very popular in many sciences: they performed a logarithmic transformation yielding a skew distribution which could hold for many different data. But so far, this has no linguistic foundation. Besides, in linguistics one tries to fit discrete distributions to discrete data, but this is no great problem because parallel discrete and continuous distributions can be converted into one another (cf. for instance, Mačutek and Altmann 2007). We see the same endeavor with the gamma distribution, which represents a sum of squared normal distributions. This is, however, a special case of Pearson's Type III distribution.

A slightly better way is to consider sentence length to be arising from a Poisson process with a constant coefficient leading to the Poisson distribution, regarding the coefficient *a posteriori* as a variable. However, the last step is not completely arbitrary. Sichel considered the parameter of the Poisson distribution to be following a generalized inverse Gaussian distribution (containing a very flexible Bessel function) but never gave reasons for this decision. It is more realistic to use a very simple distribution, namely the gamma distribution – remembering the skewed normality – and obtain

$$\text{Poisson d. } (\lambda) \underset{\lambda}{\wedge} \text{ gamma d. } (k, q/p)$$

yielding the usual negative binomial distribution, which is an acceptable result because it can be substantiated in different ways.

In this study we shall try to apply the synergetic way of modelling sentence length.

# 2 Sample Texts and Analytical Methods

## 2.1 A measurement unit of sentence lengths

In present-day Japanese, what is called "kuten" is usually used to mark the end of a sentence, in the same way as a full stop or period in English. This "kuten," or the Japanese equivalent of a period, can be omitted in dialogue or in the written form of a conversation, where the second quotation mark of a pair is to terminate a sentence. This quotation mark is also used as a way of emphasizing a word, as with "kuten" in the first and second line of this paragraph. If we regard the word as the counting unit of sentence length, in languages using Latin or Cyrillic script, the analysis is much easier than in Japanese, because in these languages the word is separated from the next one by a single space. One can easily extract the number of words in a sentence, and the number thus obtained is directly equal to the length of that sentence. Japanese, on the other hand, uses two syllabic and one logographic script, in which words are never separated by whitespace within a sentence, and several individual morphemes are intricately linked with strict rules. In this case, the morpheme (instead of the word) could be regarded as a secondary unit of sentence length, and a sentence should be resolved into morphemes in the first place. One of the outstanding application programs for morphological analysis of Japanese is "ChaSen."[1] Figure 1 illustrates how "ChaSen" morphemically analyzes a Japanese sentence, "Watashi-ha-sono-hon-wo-yonda," meaning *I have read the book*. The output consists of four columns, the first one on the left showing the exact forms of morphemes that appear on the paper, the next one showing their pronunciations in Roman letters, the third one showing the basic forms of their morphemes, and the last one showing what part of speech each morpheme belongs to. "ChaSen" enjoys high accuracy in its analysis, but it is not always wholly reliable and from the linguistic point of view it is a hybrid analysis. For instance, note that "sono" in fact consists of two morphemes from a diachronic viewpoint (cf. *sono, kono, ano* in which the morphemes can be separated in the same way as in German articles *d-er, d-ie, d-a*s); and "yonda" consists of the verb "yom(u)" and the past affix. But even if we accept the given analysis, sometimes it can provide false results. To make matters worse, Japanese has more compounds than many European languages. For example, Japanese equivalent of "lognormal distribution" is "taisu-seiki-bumpu." This compound consists of three words, "taisu" meaning "logarithm" or "log," "seiki," "normality," and "bumpu," "distribution." This compound, a little controversial, can be regarded either as one word representing one concept (cf. the German "Lognormal-verteilung") or as three words (as in Slavic languages), or even as two, "taisu," "log" and "seiki-bumpu," "normal distribution." Incidentally, "ChaSen" treats this compound as three words.

| | | | |
|---|---|---|---|
| 私 | watashi | 私 | noun (I) |
| は | ha | は | particle |
| その | sono | その | definite article (the) |
| 本 | hon | 本 | noun (book) |
| を | wo | を | particle |
| 読ん | yom | 読む | verb (read) |
| だ | da | だ | aux. verb (have) |

Figure 1. The output of ChaSen's morphemic analysis: "I have read the book"

We are aware that the length of a linguistic unit should be measured in terms of the number of its immediate constituents, in our case, clauses. But for Japanese there are no programs of this kind and computational linguists do not care for this aspect of sentence structure. Thus the complete analysis of all texts must be done with pencil and paper. On understandable grounds we shall evade such a procedure. The other way to get the length of a Japanese sentence mechanically is to count up the number of characters, or letters, instead of morphemes, using a character as the minimum constituent of a sentence. In this way, we can avoid the possible mistakes and ambiguities of a morphemic analysis. Naturally enough, previous studies have adopted the number of characters in their analyses. Here we must take note of the fact that even this approach has two problematic aspects. One is that a Japanese character can be either

[1] http://chasen.naist.jp/hiki/ChaSen/

an independent morpheme in itself, or a mere mora as in most cases. The second one is that Japanese has three different writing systems, *hiragana*, *katakana*, and *kanji*. An English word, "horse," can be written in three ways. In Figure 2, all three symbols adjacent to "horse" have the same pronunciation, "uma," conveying the same concept of "horse." The first two forms of the *kana* writing system have the same number of characters, but the third one has only one character. These equivalents for "horse" which are to be differently distributed might appear in one and the same Japanese sentence. Sentence length should be a fixed (invariant) quantity. In this sense, the fact that a choice of a writing system would possibly have a significant influence on the sentence length distribution simply casts doubts on the validity of the character as a counting unit. There is one more possibility to be taken into consideration: the phoneme. But phonemes are not immediate constituents of sentence and the support of the random variable "length" would contain so many values that many of them would have the frequency zero. This automatically leads to a senseless multimodal distribution having no relevance to the analysis.

**horse** うま ウマ 馬

Figure 2. The three ways of writing "horse" in Japanese

In collecting data on Japanese sentence lengths, there is one further question other than the selection of a measurement unit: It is the question of whether dialogue can be analyzed in the exactly same way as narrative. Mizutani (1957) pointed out that both description and dialogue in literary works have their own distributions and parameters, for example, the former following a normal distribution, and the latter, a gamma distribution. This is, however, Mizutani's mere speculation, not verified by any further experiments and, as said above, the first approximation of a kind.

Having taken into account all problematic elements, we have examined a considerable number of Japanese writers' works. This time we have relied upon "ChaSen" automatically measuring all the sentence lengths of each entire text, and the number of each-sentence morphemes resulting from its analysis has been employed without any alteration. Using the theoretical approaches of G.K. Zipf (1949) and his followers in later years, we conjectured that there should be a kind of self-regulation of sentence lengths connecting the neighbouring classes by a proportionality function (cf. Altmann and Köhler 1996), i.e.

(1) $$P_x = g(x)P_{x-1}$$

Here $g(x) = f(x)/h(x)$. Now, $f(x)$ can be interpreted as the (diversification) force of the speaker, his subconscious endeavour to make his speech production as easy as possible. However, taken to an extreme, this would destroy any communication. Thus this self-organizing force must be controlled by the hearer (or the community), by a self-regulating function $h(x)$. Both must be constructed in such a way that the probability distribution converges. In a simple and very general case we let $f(x) = a + bx$ and $h(x) = c + dx$. Inserting them in (1) we obtain

(2) $$P_x = \frac{f(x)}{h(x)}P_{x-1} = \frac{a+bx}{c+dx}P_{x-1} = \frac{(a/b+x)}{(c/d+x)}\frac{b}{d}P_{x-1}\,.$$

Replacing $a/b = k$-1, $c/d = m$-1 and $b/d = q$ $(0 < q < 1)$ we obtain ($k$ and $m$ must fulfil some special conditions)

(3) $$P_x = \frac{k+x-1}{m+x-1} q P_{x-1} \quad .$$

Solving this simple difference equation we obtain the Hyperpascal distribution

(4) $$P_x = \frac{\binom{k+x-1}{x}}{\binom{m+x-1}{x}} q^x P_0 , \qquad x = 0,1,2,\ldots$$

where $P_0^{-1} = {}_2F_1(k,1;m,q)$ and $F(.)$ is the hypergeometric function. It has been shown (cf. Altmann 1988) that this distribution is adequate if sentence length is not measured in terms of the number of immediate constituents, whereas in terms of that of clauses, the negative binomial is adequate. The Hyperpascal distribution builds a family, some members of which (Poisson, geometric, Katz family, shifted logarithmic, Hyperpoisson, Waring, Yule etc.) are employed in different domains of linguistics. And it is interesting to see that its continuous counterpart is Pearson's Type III distribution, i.e. the generalized gamma distribution (cf. Mačutek and Altmann 2007). Thus using special kinds of gamma distribution is a continuous approximation to the solution of the problem. In both cases (discrete or continuous), we must perform an *a priori* pooling of classes for shorter texts because many classes are represented very insufficiently. Pearson Type III would require numerical integration with optimization, while a ready made software will be available, if one decides to work with the Hyperpascal.

Consider the data "aitobi3 [Osamu Dazai's Ai to Bi nitsuite]" presented in Table 1. As can be seen, the individual classes are not sufficiently occupied, thus a number of modes are present and the fitting of any discrete distribution would be rather a very raw approximation. The data is presented graphically in Figure 3. In a situation like this, one usually pools some classes in order to obtain expected values at least greater than 1. This can be done either before or after applying a theoretical distribution. We shall do it before the analysis. We first choose an interval of three values, i.e. pool the classes 1-2-3, 4-5-6, 7-8-9, …and let the class be represented by its mean, i.e. X = 2, 5, 8, 11, 14,…. Then we define a new variable, x = (X – 2)/3 whose support is x = 0,1,2,3,…. Filling the frequency values in the given intervals, we obtain a smoothed distribution presented in Table 2. The expected values of the Hyperpascal distribution are shown in the third column of Table 2 and a graphic picture of the fit is shown in Figure 4. It can be inferred that greater smoothing intervals would lead to a still better fitting.

Table 1
The raw data of the file "aitobi3"

| X | f(X) | X | f(X) | X | f(X) | X | f(X) |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 26 | 1 | 51 | 0 | 76 | 0 |
| 2 | 0 | 27 | 0 | 52 | 0 | 77 | 0 |
| 3 | 8 | 28 | 4 | 53 | 1 | 78 | 0 |
| 4 | 9 | 29 | 1 | 54 | 0 | 79 | 0 |
| 5 | 3 | 30 | 1 | 55 | 0 | 80 | 0 |
| 6 | 21 | 31 | 1 | 56 | 1 | 81 | 0 |
| 7 | 10 | 32 | 4 | 57 | 0 | 82 | 0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 8 | 14 | 33 | 1 | 58 | 0 | 83 | 0 |
| 9 | 6 | 34 | 1 | 59 | 0 | 84 | 0 |
| 10 | 7 | 35 | 0 | 60 | 0 | 85 | 0 |
| 11 | 8 | 36 | 0 | 61 | 0 | 86 | 0 |
| 12 | 6 | 37 | 0 | 62 | 0 | 87 | 0 |
| 13 | 5 | 38 | 0 | 63 | 0 | 88 | 0 |
| 14 | 6 | 39 | 1 | 64 | 0 | 89 | 0 |
| 15 | 6 | 40 | 0 | 65 | 0 | 90 | 0 |
| 16 | 3 | 41 | 1 | 66 | 1 | 91 | 0 |
| 17 | 5 | 42 | 1 | 67 | 0 | 92 | 0 |
| 18 | 1 | 43 | 1 | 68 | 0 | 93 | 0 |
| 19 | 3 | 44 | 0 | 69 | 0 | 94 | 0 |
| 20 | 1 | 45 | 0 | 70 | 0 | 95 | 0 |
| 21 | 0 | 46 | 0 | 71 | 0 | 96 | 0 |
| 22 | 3 | 47 | 1 | 72 | 0 | 97 | 0 |
| 23 | 1 | 48 | 0 | 73 | 0 | 98 | 0 |
| 24 | 6 | 49 | 0 | 74 | 0 | 99 | 0 |
| 25 | 0 | 50 | 0 | 75 | 0 | 100 | 1 |

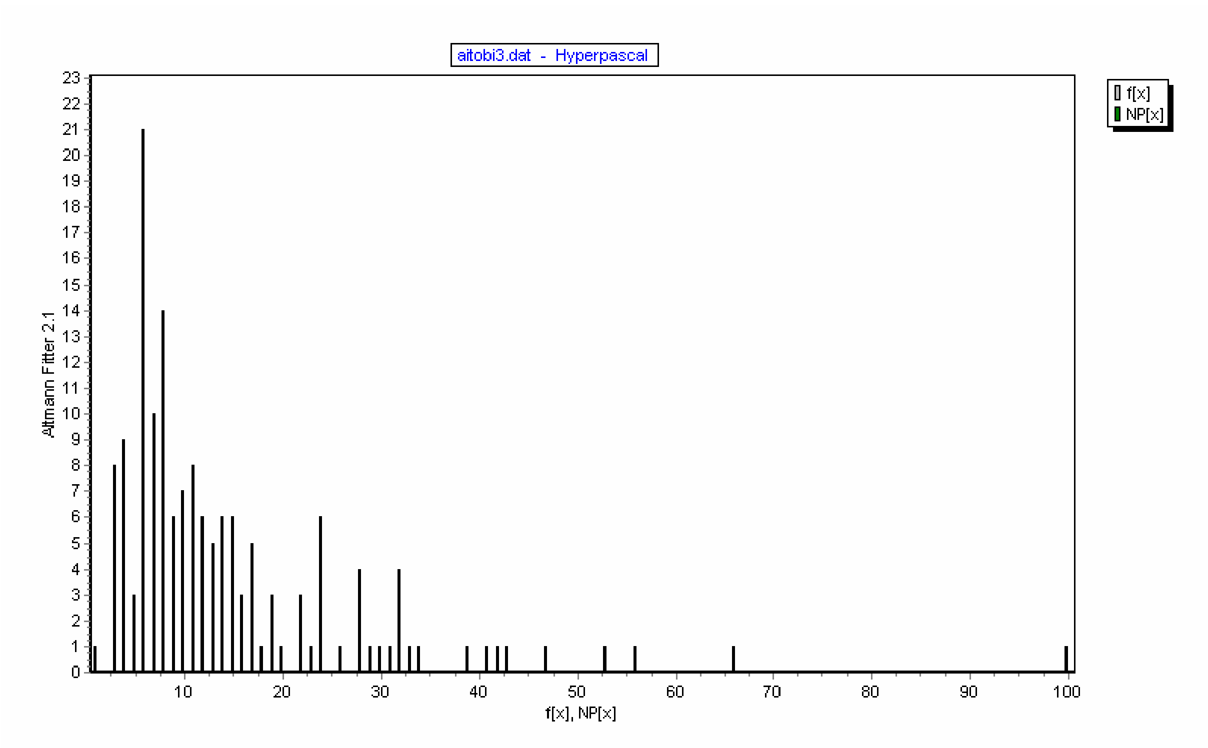


Figure 3. Raw data of sentence lengths in the file „aitobi3 [Ai to Bi nitsuite]“

Table 2
Smoothed data and the theoretical values

| x | f(x) | NP(x) | x | f(x) | NP(x) | x | f(x) | NP(x) |
|---|---|---|---|---|---|---|---|---|
| 0 | 9 | 8.48 | 12 | 1 | 2.09 | 24 | 0 | 0.08 |
| 1 | 33 | 29.86 | 13 | 2 | 1.60 | 25 | 0 | 0.06 |
| 2 | 30 | 25.94 | 14 | 1 | 1.22 | 26 | 0 | 0.05 |
| 3 | 21 | 21.07 | 15 | 1 | 0.93 | 27 | 0 | 0.03 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 4 | 17 | 16.71 | 16 | 0 | 0.71 | 28 | 0 | 0.03 |
| 5 | 9 | 13.09 | 17 | 1 | 0.54 | 29 | 0 | 0.02 |
| 6 | 7 | 10.17 | 18 | 1 | 0.41 | 30 | 0 | 0.02 |
| 7 | 10 | 7.87 | 19 | 0 | 0.31 | 31 | 0 | 0.01 |
| 8 | 1 | 6.06 | 20 | 0 | 0.24 | 32 | 0 | 0.01 |
| 9 | 6 | 4.66 | 21 | 0 | 0.18 | 33 | 1 | 0.03 |
| 10 | 6 | 3.57 | 22 | 1 | 0.14 | | | |
| 11 | 1 | 2.74 | 23 | 0 | 0.10 | | | |
| k = 0.2026, m = 0.0434, q = 0.7536, $X^2$ = 12.35, DF = 13, P = 0.50 | | | | | | | | |

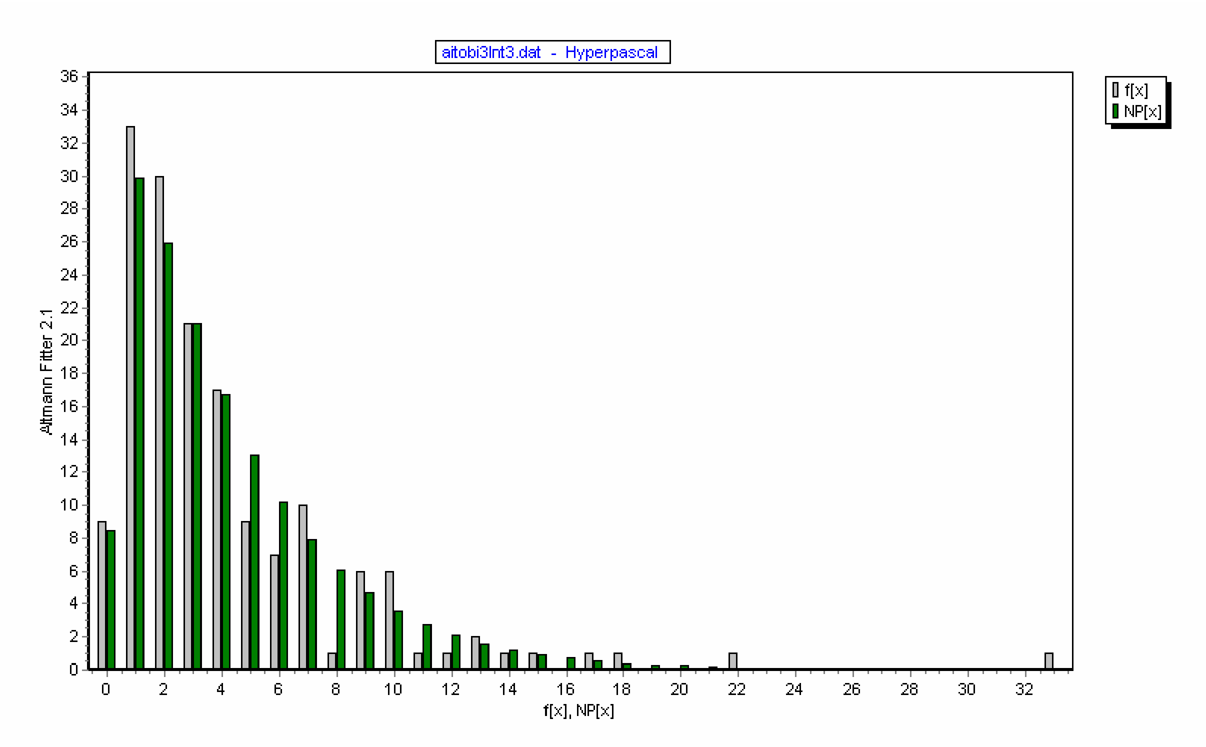


Figure 4. Fitting the Hyperpascal to smoothed data

The results of fitting the Hyperpascal to all data analysed in morphemes are presented in Table 3. We preferred three kinds of intervals: 3, 5, and 7, though using other ones perhaps a still better fitting could be achieved. Actually we tried 9 and 11 besides, in a few particular cases of either direct speech or narrative as we shall see later in Table 4 and 5. But our aim was not to show exactly how a given text must be prepared in order to obtain a perfect fit. Rather we try to show that Japanese sentence lengths measured in the traditional way accord with Sherman's law resulting in the Hyperpascal distribution (Altmann 1988). The significance level has been fixed at $\alpha = 0.01$.

Table 3
Fitting the Hyperpascal to sentence lengths measured morphemically: complete texts (Intervals 3, 5, 7) (FF – fitting failed)

| Writer | Text | Size | K | M | Q | DF | $X^2$ | Prob | Interval |
|---|---|---|---|---|---|---|---|---|---|
| akutagawa | giwaku | 276 | 0.1035 | 0.0567 | 0.8013 | 18 | 31.75 | 0.02 | 5 |
| akutagawa | jigokuhen | 526 | 0.6193 | 0.1744 | 0.8656 | 34 | 50.86 | 0.03 | 3 |
| akutagawa | kaikano | 298 | FF | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| akutagawa | kage | 333 | FF | | | | | | |
| akutagawa | kataki | 285 | 9.4345 | 2.9191 | 0.5083 | 15 | 18.28 | 0.25 | 3 |
| akutagawa | kuranosuke | 218 | 7.5058 | 1.1591 | 0.3611 | 9 | 5.13 | 0.82 | 5 |
| akutagawa | oritsuto | 965 | 2.8834 | 0.8834 | 0.4260 | 9 | 12.74 | 0.17 | 5 |
| akutagawa | rashom | 160 | FF | | | | | | |
| akutagawa | umano | 374 | 1.8397 | 0.2441 | 0.4299 | 7 | 16.33 | 0.02 | 5 |
| ango | ishino | 322 | 0.1787 | 0.0515 | 0.7347 | 13 | 14.75 | 0.32 | 7 |
| arishima | chiisaki | 300 | 2.6558 | 0.7654 | 0.6784 | 18 | 20.73 | 0.29 | 3 |
| arishima | hitofusano | 177 | 1.3649 | 0.3644 | 0.6086 | 10 | 8.18 | 0.61 | 5 |
| arishima | kajito | 307 | 0.1623 | 0.0467 | 0.6583 | 10 | 5.11 | 0.88 | 5 |
| arishima | oyako | 560 | 1.3034 | 0.2893 | 0.6048 | 13 | 10.16 | 0.68 | 5 |
| arishima | kankan | 303 | 1.7041 | 0.4752 | 0.6154 | 13 | 11.81 | 0.54 | 5 |
| dazai | aitobi | 454 | 0.5679 | 0.4472 | 0.5542 | 7 | 13.56 | 0.06 | 7 |
| dazai | joseito | 988 | 0.2281 | 0.1188 | 0.8016 | 25 | 39.86 | 0.03 | 3 |
| dazai | kamome | 460 | 0.0384 | 0.0223 | 0.5159 | 6 | 8.61 | 0.20 | 7 |
| dazai | kirigirisu | 301 | 0.3302 | 0.0738 | 0.7120 | 14 | 12.03 | 0.60 | 5 |
| dazai | kotenfu | 444 | 0.3418 | 0.1386 | 0.7251 | 15 | 27.93 | 0.02 | 3 |
| dazai | merosu | 471 | 0.1721 | 0.0532 | 0.5046 | 6 | 3.59 | 0.73 | 5 |
| dazai | oto | 159 | 0.8497 | 0.5280 | 0.7471 | 13 | 17.81 | 0.17 | 3 |
| dazai | sado | 560 | 0.1395 | 0.0489 | 0.5015 | 6 | 9.99 | 0.13 | 5 |
| dazai | ubasute | 608 | FF | | | | | | |
| dazai | osan | 212 | FF | | | | | | |
| dazai | kashoku | 696 | FF | | | | | | |
| hashimoto | chizu | 378 | 1.0635 | 0.03384 | 0.5919 | 10 | 13.48 | 0.20 | 7 |
| hirabayashi | yamabuki | 437 | 0.9106 | 0.3076 | 0.6612 | 14 | 5.86 | 0.97 | 5 |
| hojo | gantai | 192 | 2.2442 | 1.0185 | 0.6434 | 12 | 7.80 | 0.80 | 5 |
| hori | seikazoku | 435 | 1.1506 | 0.3388 | 0.7521 | 21 | 34.53 | 0.03 | 3 |
| hori | tabino | 225 | 0.2257 | 0.0668 | 0.7530 | 15 | 15.73 | 0.40 | 7 |
| hori | hanao | 382 | FF | | | | | | |
| hori | banka | 312 | 1.6980 | 0.6295 | 0.6597 | 14 | 16.00 | 0.31 | 5 |
| hori | hono | 242 | 0.2175 | 0.0609 | 0.8409 | 21 | 17.89 | 0.66 | 5 |
| itakura | goshiki | 568 | 1.7728 | 0.3777 | 0.5854 | 13 | 14.76 | 0.32 | 3 |
| itakura | haruno | 217 | 6.9692 | 0.6865 | 0.2053 | 4 | 1.55 | 0.82 | 5 |
| itakura | yamato | 331 | 0.5303 | 0.1434 | 0.6014 | 9 | 12.33 | 0.20 | 5 |
| ito | hamagiku | 342 | 2.1256 | 0.3238 | 0.4731 | 9 | 4.02 | 0.91 | 5 |
| ito | koroku | 281 | 2.0073 | 0.3486 | 0.6579 | 16 | 13.87 | 0.61 | 3 |
| ito | kyonen | 466 | FF | | | | | | |
| ito | nanako | 283 | 1.3859 | 0.1809 | 0.5288 | 9 | 3.91 | 0.92 | 5 |
| ito | nogiku | 975 | 0.5316 | 0.1260 | 0.5578 | 10 | 8.10 | 0.62 | 7 |
| kajii | deinei | 242 | 2.3732 | 0.4015 | 0.4458 | 7 | 8.77 | 0.27 | 5 |
| kajii | fuyuhae | 330 | 2.3847 | 0.5663 | 0.6403 | 16 | 12.28 | 0.72 | 5 |
| kajii | koubi | 193 | 11.9507 | 1.4538 | 0.2232 | 6 | 13.27 | 0.04 | 5 |
| kajii | remon | 142 | 0.6119 | 0.0727 | 0.7874 | 17 | 19.75 | 0.29 | 3 |
| kajii | setsugo | 277 | 0.2736 | 0.1301 | 0.7295 | 14 | 17.01 | 0.26 | 3 |
| katai | ippei | 511 | 1.9371 | 0.4552 | 0.3933 | 6 | 4.96 | 0.55 | 5 |
| katai | shojo | 266 | 0.4783 | 0.2250 | 0.7607 | 16 | 9.71 | 0.88 | 5 |
| katai | tokoyo | 635 | 0.8703 | 0.2359 | 0.7332 | 20 | 18.55 | 0.55 | 3 |
| kikuchi | emu | 268 | 2.6557 | 1.0129 | 0.4489 | 6 | 2.97 | 0.81 | 7 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| kikuchi | seni | 297 | 5.1167 | 0.3295 | 0.2304 | 5 | 6.65 | 0.25 | 7 |
| kikuchi | shimabara | 404 | 0.0394 | 0.0204 | 0.7450 | 15 | 18.34 | 0.25 | 5 |
| kikuchi | shusse | 228 | 0.9690 | 0.1952 | 0.5241 | 7 | 6.24 | 0.51 | 7 |
| kikuchi | wakasugi | 221 | 17.7068 | 1.0198 | 0.1526 | 6 | 5.30 | 0.51 | 7 |
| koda | shonen | 114 | 0.1233 | 0.0110 | 0.8329 | 17 | 20.18 | 0.27 | 5 |
| kuroshima | ana | 435 | 2.0282 | 0.5573 | 0.6166 | 14 | 14.24 | 0.43 | 3 |
| kuroshima | dempo | 216 | 0.5852 | 0.2358 | 0.6399 | 10 | 9.13 | 0.52 | 5 |
| kuroshima | mogura | 654 | 4.3585 | 0.7056 | 0.3205 | 7 | 13.70 | 0.06 | 5 |
| kuroshima | mon | 166 | 0.9769 | 0.1241 | 0.3403 | 3 | 3.79 | 0.28 | 7 |
| kuroshima | nusumu | 364 | 2.7890 | 0.3142 | 0.2367 | 3 | 1.11 | 0.77 | 7 |
| kuroshima | sato | 167 | 0.3650 | 0.1321 | 0.5824 | 7 | 1.68 | 0.98 | 5 |
| kuroshima | tongun | 252 | 1.6990 | 0.2883 | 0.4819 | 8 | 9.23 | 0.32 | 5 |
| kuroshima | zensho | 316 | 1.9027 | 0.5389 | 0.3649 | 5 | 1.99 | 0.85 | 7 |
| makino | tsurube | 361 | FF | | | | | | |
| makino | kinada | 279 | 1.7881 | 0.7121 | 0.6357 | 12 | 10.28 | 0.59 | 7 |
| minakami | yamanote | 387 | 0.5858 | 0.3683 | 0.7696 | 18 | 13.90 | 0.74 | 5 |
| mishima | hashi | 359 | 0.8802 | 0.3009 | 0.5449 | 8 | 8.83 | 0.36 | 7 |
| miyamoto | akarui | 297 | 1.6922 | 1.3023 | 0.7606 | 17 | 28.31 | 0.04 | 3 |
| miyazawa | karasu | 165 | 0.2216 | 0.1312 | 0.7984 | 16 | 18.01 | 0.32 | 3 |
| murai | sobano | 179 | 1.6458 | 0.1299 | 0.7071 | 18 | 33.89 | 0.01 | 5 |
| oda | osaka | 164 | 0.2717 | 0.0734 | 0.7978 | 16 | 29.21 | 0.02 | 5 |
| oda | keiba | 231 | 0.1659 | 0.0387 | 0.7543 | 15 | 24.77 | 0.05 | 7 |
| ogai | futarino | 412 | 34.5950 | 1.1263 | 0.0548 | 4 | 8.57 | 0.07 | 7 |
| ogai | asobi | 373 | 2.9731 | 0.4692 | 0.2801 | 4 | 1.77 | 0.78 | 7 |
| ogai | shokudo | 193 | 53.72 | 4.3131 | 0.0817 | 5 | 5.05 | 0.41 | 7 |
| ogai | niwatori | 762 | 6.6721 | 1.1193 | 0.1941 | 4 | 4.07 | 0.40 | 7 |
| ogai | kazui | 271 | 0.4945 | 0.1427 | 0.5723 | 8 | 6.69 | 0.57 | 7 |
| okamoto | karei | 286 | 1.1520 | 0.3446 | 0.7406 | 18 | 12.74 | 0.81 | 3 |
| okamoto | kingyo | 830 | 1.5109 | 0.6276 | 0.5896 | 12 | 20.43 | 0.06 | 7 |
| okamoto | rigyo | 171 | 0.6448 | 0.2775 | 0.7991 | 19 | 19.11 | 0.45 | 3 |
| okamoto | sushi | 366 | 0.8861 | 0.2568 | 0.6546 | 13 | 14.82 | 0.32 | 5 |
| okamoto | tokaido | 378 | 0.0820 | 0.0438 | 0.6949 | 12 | 11.01 | 0.53 | 7 |
| sasa | kikansha | 217 | 20.0666 | 0.5214 | 0.0542 | 2 | 0.73 | 0.70 | 7 |
| sasa | midori | 127 | 1.3497 | 0.3012 | 0.6031 | 10 | 12.79 | 0.24 | 5 |
| shimazaki | shishu | 316 | 0.8158 | 0.1699 | 0.7545 | 20 | 21.55 | 0.37 | 3 |
| shimazaki | namiki | 373 | 0.2725 | 0.0881 | 0.6094 | 9 | 6.85 | 0.65 | 5 |
| shimazaki | fune | 354 | 2.9653 | 0.5659 | 0.4660 | 9 | 12.08 | 0.21 | 5 |
| shimazaki | mebae | 881 | 2.1641 | 0.4502 | 0.4824 | 9 | 3.56 | 0.94 | 5 |
| shimazaki | bumpai | 497 | 0.8753 | 0.3177 | 0.5588 | 9 | 7.76 | 0.56 | 7 |
| shiraki | saite | 236 | 0.6287 | 0.1199 | 0.7862 | 20 | 23.21 | 0.28 | 3 |
| soseki | buncho | 411 | 0.8648 | 0.0818 | 0.4942 | 8 | 4.31 | 0.83 | 5 |
| soseki | hennaoto | 110 | 0.0692 | 0.0158 | 0.7292 | 10 | 6.76 | 0.75 | 5 |
| soseki | sakubutsu | 183 | 0.0463 | 0.0070 | 0.7325 | 12 | 6.37 | 0.90 | 5 |
| soseki | kotonone | 776 | 1.2049 | 0.3660 | 0.4216 | 7 | 1.96 | 0.96 | 7 |
| soseki | tegami | 263 | 0.6802 | 0.2308 | 0.6347 | 10 | 10.31 | 0.41 | 7 |
| suzu | ogon | 439 | 20.3908 | 1.7829 | 0.1431 | 6 | 11.73 | 0.07 | 5 |
| takiji | haha | 294 | 0.2951 | 0.1829 | 0.7342 | 14 | 10.90 | 0.69 | 5 |
| takiji | yukino | 739 | 19.5497 | 8.2619 | 0.3997 | 12 | 9.88 | 0.63 | 3 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| unno | daino | 489 | 0.9564 | 0.2851 | 0.5884 | 10 | 10.41 | 0.41 | 5 |
| unno | kagaku | 114 | 0.6669 | 0.1175 | 0.5556 | 6 | 1.81 | 0.94 | 7 |
| unno | kibutsu | 466 | 1.2499 | 0.4762 | 0.4860 | 7 | 7.58 | 0.37 | 7 |
| unno | neon | 408 | 1.1334 | 0.4731 | 0.5099 | 7 | 9.91 | 0.19 | 7 |
| unno | tsuki | 505 | 0.0921 | 0.0394 | 0.8025 | 22 | 28.00 | 0.18 | 3 |
| watanabe | uso | 300 | 0.2824 | 0.1313 | 0.6904 | 12 | 7.11 | 0.85 | 5 |
| watanabe | akai | 197 | 2.0690 | 0.5612 | 0.5421 | 9 | 7.80 | 0.55 | 5 |
| watanabe | aruhaha | 193 | 2.8232 | 1.0305 | 0.4449 | 6 | 9.30 | 0.16 | 7 |
| watanabe | shohai | 298 | 0.5499 | 0.1965 | 0.6890 | 12 | 9.22 | 0.68 | 5 |
| yamada | musashino | 351 | 0.3253 | 0.1736 | 0.8413 | 26 | 21.01 | 0.74 | 3 |
| yamashita | ruten | 116 | 1.2913 | 0.0707 | 0.6484 | 12 | 16.08 | 0.19 | 5 |
| yokomitsu | kikaiy | 244 | FF | | | | | | |
| yokomitsu | jikan | 150 | 3.6477 | 0.6335 | 0.6743 | 18 | 19.59 | 0.36 | 7 |
| yumeno | koko | 439 | 1.3923 | 0.4764 | 0.6191 | 12 | 14.53 | 0.27 | 5 |

As can be seen, out of 113 texts 103 could be captured by the Hyperpascal. Even the residual 10 data could be captured if we attempted a different priori pooling, but this is no more than a question of principle. Sometimes we even left smaller intervals if the fitting was significant. We did not try to achieve “the best fit.” We rather scrutinize the problem of distinguishing between dialogue and narrative parts. Table 4 contains the same texts, showing the result of direct speech, and followed by Table 5 showing that of narrative. However, there is a problem associated with direct speech. If it is dialogue, i.e. if there are at least two speaking persons, the author must differentiate them in some way, for example, in their sentence length. Hence dialogue has two independent parts which must not be mixed. If there are *more* speaking persons, then each speaker’s words must be evaluated separately. This kind of research might be of great interest and use for the purposes of analyzing a single work from a literary point of view. For our particular purposes, namely for corroborating a variant of Sherman’s law, it would play a merely subordinate role. The results of dialogue and narrative parts are given below consecutively. In respect of direct speech, there are a dozen texts that contain no dialogue data (ND) and as many texts that do not have enough data to fit the Hyperpascal (NED), and there are four texts where the fitting failed (FF).

Table 4
Direct speech (Intervals 3, 5, 7, 9, 11)

| Writer | Text | Size | K | M | Q | DF | $X^2$ | P | Interv. |
|---|---|---|---|---|---|---|---|---|---|
| akutagawa | giwaku | 44 | FF | | | | | | |
| akutagawa | jigokuhen | 114 | 0.1816 | 0.1033 | 0.4837 | 4 | 1.27 | 0.87 | 7 |
| akutagawa | kage | 116 | 0.1826 | 0.1846 | 0.5044 | 3 | 2.25 | 0.52 | 7 |
| akutagawa | kaikano | 188 | 20.2672 | 3.6874 | 0.2500 | 9 | 15.87 | 0.07 | 9 |
| akutagawa | kataki | 36 | 6.2749 | 0.4248 | 0.1614 | 2 | 0.40 | 0.82 | 5 |
| akutagawa | kuranosuke | 79 | 22.9651 | 1.448374 | 0.1061 | 3 | 0.47 | 0.93 | 7 |
| akutagawa | oritsuto | 462 | 0.2653 | 0.121514 | 0.3137 | 2 | 2.09 | 0.35 | 7 |
| akutagawa | rashom | 26 | 1.3298 | 0.393276 | 0.5471 | 4 | 5.08 | 0.28 | 5 |
| akutagawa | umano | 132 | 0.1818 | 0.049125 | 0.4942 | 4 | 10.92 | 0.03 | 5 |
| ango | ishino | NED | | | | | | | |
| arishima | chiisaki | NED | | | | | | | |
| arishima | hitofusano | 35 | FF | | | | | | |
| arishima | kajito | 53 | 2.2645 | 1.3530 | 0.3170 | 1 | 0.30 | 0.59 | 7 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| arishima | kankan | 37 | 1.5100 | 1.2894 | 0.6497 | 5 | 5.14 | 0.40 | 5 |
| arishima | oyako | 199 | 0.4421 | 0.2238 | 0.5669 | 6 | 2.07 | 0.91 | 7 |
| dazai | aitobi | 298 | 0.4143 | 0.3265 | 0.6619 | 10 | 12.99 | 0.22 | 5 |
| dazai | joseito | 58 | 2.5983 | 1.8394 | 0.5290 | 4 | 3.86 | 0.43 | 3 |
| dazai | kamome | 122 | 1.1527 | 1.2311 | 0.5092 | 3 | 4.83 | 0.18 | 7 |
| dazai | kashoku | 209 | 0.1433 | 0.0780 | 0.4593 | 4 | 9.23 | 0.06 | 5 |
| dazai | kirigirisu | NED | | | | | | | |
| dazai | kotenfu | 187 | 0.2690 | 0.1160 | 0.7351 | 13 | 18.14 | 0.15 | 3 |
| dazai | merosu | 168 | 0.4932 | 0.1373 | 0.3484 | 3 | 0.77 | 0.86 | 5 |
| dazai | osan | 96 | 1.7210 | 3.2101 | 0.5548 | 1 | 2.95 | 0.086 | 9 |
| dazai | oto | 38 | 0.7478 | 0.7053 | 0.6608 | 4 | 5.18 | 0.27 | 3 |
| dazai | sado | 102 | 1.1612 | 1.2352 | 0.4033 | 1 | 0.76 | 0.38 | 7 |
| dazai | ubasute | 208 | 0.0723 | 0.0428 | 0.6776 | 10 | 14.39 | 0.16 | 3 |
| hashimoto | chizu | 117 | 4.1328 | 1.9360 | 0.5181 | 7 | 4.30 | 0.75 | 5 |
| hirabayashi | yamabuki | 184 | 1.1526 | 0.5655 | 0.5887 | 8 | 1.03 | 0.998 | 5 |
| hojo | gantai | 39 | 0.6168 | 0.4457 | 0.5251 | 3 | 1.52 | 0.68 | 5 |
| hori | banka | 54 | 3.1609 | 0.9597 | 0.2349 | 1 | 0.13 | 0.72 | 7 |
| hori | hanao | 31 | 0.3769 | 0.1168 | 0.6588 | 6 | 5.96 | 0.43 | 3 |
| hori | hono | 152 | 1.5339 | 0.7042 | 0.7066 | 13 | 9.58 | 0.73 | 7 |
| hori | seikazoku | 57 | 1.7604 | 0.4201 | 0.4760 | 5 | 2.21 | 0.82 | 3 |
| hori | tabino | NED | | | | | | | |
| itakura | goshiki | NED | | | | | | | |
| itakura | haruno | NED | | | | | | | |
| itakura | yamato | NED | | | | | | | |
| ito | hamagiku | 84 | 1.5591 | 0.3012 | 0.4500 | 5 | 3.24 | 0.66 | 5 |
| ito | koroku | 111 | 0.0454 | 0.0180 | 0.4796 | 3 | 1.60 | 0.66 | 7 |
| ito | kyonen | 35 | 0.7758 | 0.1511 | 0.4300 | 2 | 0.75 | 0.69 | 5 |
| ito | nanako | 29 | 15.9699 | 1.2484 | 0.1563 | 3 | 1.16 | 0.76 | 3 |
| ito | nogiku | 393 | 0.1374 | 0.0317 | 0.6467 | 10 | 18.21 | 0.05 | 5 |
| kajii | deinei | ND | | | | | | | |
| kajii | fuyuhae | ND | | | | | | | |
| kajii | koubi | ND | | | | | | | |
| kajii | remon | ND | | | | | | | |
| kajii | setsugo | ND | | | | | | | |
| katai | ippei | 95 | 0.5172 | 0.1831 | 0.4502 | 3 | 5.89 | 0.12 | 3 |
| katai | shojo | 57 | 0.1548 | 0.0763 | 0.6293 | 5 | 1.62 | 0.90 | 5 |
| katai | tokoyo | 192 | 0.3829 | 0.1211 | 0.6609 | 10 | 10.72 | 0.38 | 3 |
| kikuchi | emu | 83 | 0.3832 | 0.1026 | 0.4789 | 3 | 2.715 | 0.44 | 5 |
| kikuchi | seni | 85 | 1.715 | 0.2063 | 0.1747 | 1 | 1.29 | 0.26 | 11 |
| kikuchi | shimabara | 214 | 0.3771 | 0.1867 | 0.5792 | 7 | 12.26 | 0.09 | 5 |
| kikuchi | shusse | 34 | 0.0994 | 0.0165 | 0.4302 | 2 | 0.34 | 0.84 | 5 |
| kikuchi | wakasugi | 11 | FF | | | | | | |
| koda | shonen | ND | | | | | | | |
| kuroshima | ana | 108 | 0.4544 | 0.1806 | 0.3687 | 1 | 4.18 | 0.04 | 5 |
| kuroshima | dempo | 80 | 0.6002 | 0.5297 | 0.6574 | 6 | 5.59 | 0.47 | 5 |
| kuroshima | mogura | 153 | 38.946 | 9.1844 | 0.2144 | 7 | 2.38 | 0.94 | 3 |
| kuroshima | mon | 62 | 0.1098 | 0.0206 | 0.5401 | 3 | 5.23 | 0.16 | 5 |
| kuroshima | nusumu | 96 | 0.1706 | 0.1061 | 0.6348 | 6 | 12.08 | 0.06 | 3 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| kuroshima | sato | 68 | 0.2488 | 0.0922 | 0.6740 | 6 | 14.66 | 0.02 | 3 |
| kuroshima | tongun | 61 | 1.4061 | 0.2862 | 0.2784 | 1 | 0.96 | 0.33 | 5 |
| kuroshima | zensho | 74 | 2.3327 | 2.6170 | 0.7587 | 8 | 3.02 | 0.93 | 3 |
| makino | kinada | 89 | 0.5892 | 0.1421 | 0.7448 | 12 | 14.78 | 0.25 | 3 |
| makino | tsurube | 157 | 1.5685 | 1.3012 | 0.5458 | 5 | 7.30 | 0.20 | 7 |
| minakami | yamanote | 102 | 0.1527 | 0.0598 | 0.2744 | 1 | 0.40 | 0.53 | 7 |
| mishima | hashi | 74 | 0.3204 | 0.1001 | 0.3467 | 1 | 4.76 | 0.03 | 5 |
| miyamoto | akarui | 113 | 1.7018 | 1.5909 | 0.6519 | 6 | 2.97 | 0.81 | 3 |
| miyazawa | karasu | 62 | 0.4617 | 0.2967 | 0.5867 | 4 | 2.86 | 0.58 | 3 |
| murai | sobano | ND | | | | | | | |
| oda | keiba | NED | | | | | | | |
| oda | osaka | NED | | | | | | | |
| ogai | asobi | 65 | 4.1224 | 5.8304 | 0.7282 | 3 | 6.05 | 0.11 | 5 |
| ogai | futarino | 56 | 1.7175 | 1.4188 | 0.6062 | 4 | 4.63 | 0.33 | 5 |
| ogai | kazui | 65 | 0.2298 | 0.1528 | 0.5261 | 3 | 3.02 | 0.39 | 5 |
| ogai | niwatori | 175 | 0.9699 | 1.0521 | 0.4804 | 3 | 8.79 | 0.03 | 7 |
| ogai | shokudo | 102 | 1.6770 | 0.5108 | 0.4607 | 5 | 3.40 | 0.64 | 7 |
| okamoto | karei | 92 | 0.0462 | 0.0101 | 0.7534 | 10 | 8.29 | 0.60 | 3 |
| okamoto | kingyo | 198 | 0.0338 | 0.0243 | 0.7854 | 13 | 22.56 | 0.05 | 3 |
| okamoto | rigyo | 52 | 2.0128 | 1.6204 | 0.5564 | 2 | 3.00 | 0.22 | 5 |
| okamoto | sushi | 78 | 0.3798 | 0.1484 | 0.4783 | 3 | 5.49 | 0.14 | 5 |
| okamoto | tokaido | 102 | 0.0672 | 0.0225 | 0.8015 | 14 | 7.20 | 0.93 | 3 |
| sasa | kikansha | 102 | FF | | | | | | |
| sasa | midori | 53 | 3.3945 | 0.7119 | 0.3239 | 3 | 0.74 | 0.86 | 7 |
| shimazaki | bumpai | 120 | 1.6281 | 0.2471 | 0.2613 | 2 | 1.54 | 0.46 | 5 |
| shimazaki | fune | 34 | 0.3927 | 0.1668 | 0.5610 | 3 | 2.87 | 0.41 | 5 |
| shimazaki | mebae | 142 | 1.4990 | 0.8369 | 0.6176 | 8 | 11.57 | 0.17 | 3 |
| shimazaki | namiki | 156 | 0.4712 | 0.1523 | 0.4678 | 5 | 4.20 | 0.52 | 5 |
| shimazaki | shishu | 33 | 2.9693 | 0.5955 | 0.3710 | 3 | 1.93 | 0.59 | 5 |
| shiraki | saite | 38 | 0.3329 | 0.0478 | 0.5619 | 4 | 4.63 | 0.33 | 3 |
| soseki | buncho | NED | | | | | | | 7 |
| soseki | hennaoto | 28 | 0.3773 | 0.2522 | 0.5495 | 2 | 0.56 | 0.76 | 5 |
| soseki | kotonone | 375 | 0.0813 | 0.0264 | 0.6194 | 9 | 3.38 | 0.95 | 5 |
| soseki | sakubutsu | ND | | | | | | | 7 |
| soseki | tegami | 52 | 35.634 | 2.1977 | 0.0910 | 4 | 3.59 | 0.46 | 3 |
| suzu | ogon | 115 | 2.4353 | 0.1166 | 0.2720 | 3 | 1.84 | 0.61 | 5 |
| takiji | haha | 83 | 0.6368 | 0.3612 | 0.6280 | 6 | 10.18 | 0.12 | 5 |
| takiji | yukino | 174 | 0.6376 | 0.5780 | 0.4857 | 4 | 4.2 | 0.38 | 3 |
| unno | daino | 200 | 0.8300 | 0.3943 | 0.5754 | 7 | 2.3 | 0.94 | 5 |
| unno | kagaku | NED | | | | | | | 7 |
| unno | kibutsu | 216 | 1.3625 | 1.0254 | 0.7349 | 14 | 9.89 | 0.77 | 3 |
| unno | neon | 226 | 0.2116 | 0.1524 | 0.6677 | 9 | 4.54 | 0.87 | 5 |
| unno | tsuki | 245 | 0.1203 | 0.0556 | 0.6774 | 10 | 6.17 | 0.80 | 3 |
| watanabe | akai | NED | | | | | | | |
| watanabe | aruhaha | ND | | | | | | | |
| watanabe | shohai | ND | | | | | | | |
| watanabe | uso | 164 | 0.6077 | 0.5130 | 0.5353 | 5 | 3.28 | 0.66 | 7 |
| yamada | musashino | 185 | 0.4815 | 0.5999 | 0.6410 | 6 | 5.24 | 0.51 | 7 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| yamashita | ruten | 90 | 3.2763 | 0.3158 | 0.3972 | 6 | 7.10 | 0.31 | 7 |
| yokomitsu | jikan | ND | | | | | | | |
| yokomitsu | kikaiy | ND | | | | | | | |
| yumeno | koko | 137 | 0.3294 | 0.2449 | 0.8120 | 16 | 16.59 | 0.41 | 3 |

Table 5
Narrative (Intervals 3, 5, 7, 9, 11)

| Writer | Text | Size | K | M | Q | DF | $X^2$ | P | |
|---|---|---|---|---|---|---|---|---|---|
| akutagawa | giwaku | 232 | 3.2508 | 0.7474 | 0.4755 | 9 | 11.81 | 0.22 | 7 |
| akutagawa | jigokuhen | 412 | 1.7096 | 0.3189 | 0.6290 | 13 | 26.62 | 0.01 | 7 |
| akutagawa | kage | 217 | 12.0019 | 0.8216 | 0.2260 | 7 | 11.47 | 0.12 | 5 |
| akutagawa | kaikano | 110 | 1.4427 | 0.2013 | 0.5610 | 7 | 9.97 | 0.19 | 7 |
| akutagawa | kataki | 249 | 527.7755 | 1.9709 | 0.0087 | 7 | 13.24 | 0.07 | 5 |
| akutagawa | kuranosuke | 139 | 5.9010 | 0.3954 | 0.3701 | 9 | 2.28 | 0.99 | 5 |
| akutagawa | oritsuto | 503 | FF | | | | | | |
| akutagawa | rashom | 134 | FF | | | | | | |
| akutagawa | umano | 242 | 11.2705 | 0.2983 | 0.15839 | 5 | 7.36 | 0.20 | 5 |
| ango | ishino | 318 | 0.1412 | 0.0388 | 0.7412 | 13 | 14.82 | 0.32 | 7 |
| arishima | chiisaki | 297 | 3.9009 | 1.1110 | 0.6215 | 17 | 18.79 | 0.34 | 3 |
| arishima | hitofusano | 142 | 2.0363 | 0.2829 | 0.5584 | 10 | 4.81 | 0.90 | 5 |
| arishima | kajito | 254 | 0.5221 | 0.1362 | 0.5169 | 6 | 3.56 | 0.74 | 7 |
| arishima | kankan | 266 | 0.3565 | 0.0729 | 0.6264 | 9 | 6.06 | 0.73 | 7 |
| arishima | oyako | 361 | 4.0829 | 0.9981 | 0.6000 | 17 | 9.78 | 0.91 | 3 |
| dazai | aitobi | 156 | 0.5603 | 0.3690 | 0.5256 | 5 | 4.85 | 0.43 | 7 |
| dazai | joseito | 930 | 0.2664 | 0.1216 | 0.8011 | 25 | 38.53 | 0.04 | 3 |
| dazai | kamome | 338 | 0.1819 | 0.0760 | 0.4907 | 5 | 7.05 | 0.22 | 7 |
| dazai | kashoku | 487 | 0.0963 | 0.0275 | 0.6607 | 11 | 17.94 | 0.08 | 5 |
| dazai | kirigirisu | 299 | 0.3799 | 0.0815 | 0.7082 | 14 | 12.30 | 0.58 | 5 |
| dazai | kotenfu | 257 | 0.4089 | 0.1571 | 0.6982 | 12 | 14.66 | 0.26 | 3 |
| dazai | merosu | 303 | 0.3224 | 0.0806 | 0.5184 | 6 | 4.00 | 0.68 | 5 |
| dazai | osan | 116 | 0.3899 | 0.1703 | 0.8125 | 17 | 33.73 | 0.01 | 7 |
| dazai | oto | 121 | 0.2934 | 0.1345 | 0.5840 | 6 | 8.65 | 0.19 | 7 |
| dazai | sado | 458 | 0.2872 | 0.0790 | 0.4953 | 6 | 11.47 | 0.07 | 5 |
| dazai | ubasute | 400 | 0.4038 | 0.2527 | 0.4846 | 5 | 14.88 | 0.01 | 7 |
| hashimoto | chizu | 261 | 0.6289 | 0.1289 | 0.6293 | 10 | 11.99 | 0.29 | 7 |
| hirabayashi | yamabuki | 253 | 0.8716 | 0.0803 | 0.6629 | 13 | 5.46 | 0.96 | 5 |
| hojo | gantai | 153 | 2.0267 | 0.2963 | 0.5915 | 11 | 7.67 | 0.74 | 5 |
| hori | banka | 258 | 6.6566 | 2.5129 | 0.6290 | 19 | 14.07 | 0.78 | 3 |
| hori | hanao | 351 | 1.1630 | 0.1029 | 0.8032 | 29 | 35.43 | 0.19 | 3 |
| hori | hono | 90 | 0.1972 | 0.0369 | 0.8306 | 17 | 14.45 | 0.63 | 5 |
| hori | seikazoku | 378 | 1.7554 | 0.3774 | 0.7199 | 20 | 35.53 | 0.02 | 3 |
| hori | tabino | 217 | 0.2795 | 0.0572 | 0.7479 | 15 | 16.16 | 0.37 | 7 |
| itakura | goshiki | 532 | 0.5177 | 0.0946 | 0.3713 | 4 | 3.66 | 0.45 | 7 |
| itakura | haruno | 214 | 9.7424 | 0.7743 | 0.1661 | 4 | 1.47 | 0.83 | 5 |
| itakura | yamato | 309 | 0.6158 | 0.1083 | 0.5859 | 9 | 11.63 | 0.23 | 5 |
| ito | hamagiku | 258 | 2.5754 | 0.3032 | 0.4484 | 8 | 2.97 | 0.94 | 5 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ito | koroku | 170 | 1.5610 | 0.0512 | 0.4003 | 6 | 3.13 | 0.79 | 7 |
| ito | kyonen | 431 | 1.6990 | 0.1969 | 0.3002 | 4 | 9.27 | 0.05 | 9 |
| ito | nanako | 254 | 2.8372 | 0.2986 | 0.4302 | 8 | 7.76 | 0.46 | 5 |
| ito | nogiku | 582 | 1.8452 | 0.2851 | 0.4589 | 8 | 9.78 | 0.28 | 7 |
| kajii | deinei | 222 | 5.1011 | 0.4674 | 0.3167 | 7 | 10.22 | 0.18 | 5 |
| kajii | fuyuhae | 305 | 2.3021 | 0.3570 | 0.6201 | 15 | 9.47 | 0.85 | 3 |
| kajii | koubi | 187 | 3.7351 | 0.5448 | 0.3925 | 7 | 14.88 | 0.04 | 5 |
| kajii | remon | 134 | 1.360 | 0.1093 | 0.7077 | 15 | 18.87 | 0.22 | 3 |
| kajii | setsugo | 180 | 0.6015 | 0.1389 | 0.7129 | 13 | 16.01 | 0.25 | 3 |
| katai | ippei | 416 | 0.5105 | 0.0858 | 0.3799 | 4 | 3.87 | 0.42 | 7 |
| katai | shojo | 209 | 0.5526 | 0.1179 | 0.8481 | 27 | 22.16 | 0.73 | 3 |
| katai | tokoyo | 443 | 1.9931 | 0.2240 | 0.6600 | 18 | 13.79 | 0.74 | 3 |
| kikuchi | emu | 185 | 0.2209 | 0.0580 | 0.7372 | 9 | 13.21 | 0.15 | 5 |
| kikuchi | seni | 212 | 2.3085 | 0.1345 | 0.3451 | 5 | 4.99 | 0.42 | 7 |
| kikuchi | shimabara | 190 | FF | | | | | | |
| kikuchi | shusse | 194 | 1.5669 | 0.1836 | 0.4743 | 7 | 6.802 | 0.45 | 7 |
| kikuchi | wakasugi | 210 | FF | | | | | | |
| koda | shonen | 114 | 0.1233 | 0.0110 | 0.8329 | 17 | 20.18 | 0.27 | 5 |
| kuroshima | ana | 327 | 1.9025 | 0.2617 | 0.4512 | 8 | 4.22 | 0.84 | 5 |
| kuroshima | dempo | 136 | 1.8996 | 0.3230 | 0.6698 | 15 | 12.16 | 0.67 | 3 |
| kuroshima | mogura | 501 | 3.5743 | 0.2388 | 0.3218 | 7 | 12.05 | 0.10 | 5 |
| kuroshima | mon | 104 | 1.5992 | 0.1087 | 0.3088 | 3 | 3.06 | 0.38 | 7 |
| kuroshima | nusumu | 268 | 4.0326 | 0.2841 | 0.4515 | 11 | 8.64 | 0.66 | 3 |
| kuroshima | sato | 99 | 0.0074 | 0.0010 | 0.5192 | 4 | 1.76 | 0.78 | 7 |
| kuroshima | tongun | 191 | 2.4103 | 0.1772 | 0.4365 | 8 | 8.02 | 0.43 | 5 |
| kuroshima | zensho | 242 | 1.9764 | 0.1880 | 0.4369 | 7 | 3.49 | 0.84 | 5 |
| makino | kinada | 190 | 1.3305 | 0.5076 | 0.7135 | 12 | 21.60 | 0.04 | 7 |
| makino | tsurube | 204 | 0.5034 | 0.2017 | 0.5960 | 8 | 13.21 | 0.11 | 9 |
| minakami | yamanote | 285 | 3.1271 | 0.6449 | 0.4968 | 10 | 15.85 | 0.10 | 7 |
| mishima | hashi | 285 | 0.9655 | 0.1537 | 0.6460 | 12 | 10.05 | 0.61 | 5 |
| miyamoto | akarui | 184 | 11.9772 | 1.3814 | 0.1700 | 4 | 8.57 | 0.07 | 7 |
| miyazawa | karasu | 103 | 0.0022 | 0.0014 | 0.8347 | 17 | 28.06 | 0.04 | 3 |
| murai | sobano | 179 | 1.6458 | 0.1299 | 0.7071 | 18 | 33.89 | 0.01 | 5 |
| oda | keiba | 228 | 0.1002 | 0.0215 | 0.7602 | 15 | 25.33 | 0.05 | 7 |
| oda | osaka | 162 | 0.2893 | 0.0806 | 0.7980 | 16 | 28.82 | 0.03 | 5 |
| ogai | asobi | 308 | 3.2854 | 0.3025 | 0.2594 | 4 | 1.79 | 0.77 | 7 |
| ogai | futarino | 356 | 6.6858 | 1.4855 | 0.4836 | 14 | 12.42 | 0.57 | 3 |
| ogai | kazui | 206 | 1.3157 | 0.1693 | 0.4771 | 7 | 6.69 | 0.46 | 7 |
| ogai | niwatori | 587 | 13.1228 | 0.7404 | 0.1021 | 4 | 4.21 | 0.38 | 7 |
| ogai | shokudo | 91 | 0.2776 | 0.0832 | 0.7804 | 13 | 9.53 | 0.73 | 3 |
| okamoto | karei | 194 | 3.5711 | 0.6043 | 0.4678 | 9 | 12.45 | 0.19 | 5 |
| okamoto | kingyo | 632 | 3.4650 | 0.9013 | 0.5850 | 16 | 24.03 | 0.09 | 5 |
| okamoto | rigyo | 119 | 0.2577 | 0.0490 | 0.8499 | 22 | 21.25 | 0.51 | 3 |
| okamoto | sushi | 288 | 0.7983 | 0.1455 | 0.5572 | 8 | 8.21 | 0.41 | 7 |
| okamoto | tokaido | 276 | 0.1071 | 0.0433 | 0.7016 | 12 | 14.70 | 0.26 | 7 |
| sasa | kikansha | 115 | 20.8736 | 0.1578 | 0.0819 | 4 | 1.85 | 0.76 | 5 |
| sasa | midori | 74 | 1.4028 | 0.1152 | 0.4674 | 5 | 4.05 | 0.54 | 7 |
| shimazaki | bumpai | 377 | 1.7282 | 0.5855 | 0.6445 | 14 | 23.06 | 0.06 | 5 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| shimazaki | fune | 320 | 3.1831 | 0.4953 | 0.6151 | 17 | 19.83 | 0.28 | 3 |
| shimazaki | mebae | 739 | 1.4973 | 0.1784 | 0.3944 | 6 | 1.75 | 0.94 | 7 |
| shimazaki | namiki | 217 | 0.2989 | 0.0960 | 0.5413 | 6 | 3.64 | 0.73 | 7 |
| shimazaki | shishu | 283 | 1.1646 | 0.2866 | 0.4810 | 7 | 6.08 | 0.53 | 7 |
| shiraki | saite | 198 | 2.2885 | 0.2657 | 0.4015 | 6 | 3.51 | 0.74 | 7 |
| soseki | buncho | 409 | 0.8612 | 0.0815 | 0.4941 | 8 | 4.32 | 0.83 | 5 |
| soseki | hennaoto | 82 | 0.3567 | 0.2955 | 0.7203 | 8 | 7.60 | 0.47 | 5 |
| soseki | kotonone | 401 | 3.2625 | 0.6202 | 0.5749 | 15 | 17.52 | 0.29 | 3 |
| soseki | sakubutsu | 183 | 0.0463 | 0.0070 | 0.7325 | 12 | 6.37 | 0.90 | 5 |
| soseki | tegami | 211 | 2.2489 | 0.3751 | 0.5131 | 9 | 13.03 | 0.16 | 7 |
| suzu | ogon | 324 | 10.4524 | 0.5815 | 0.0638 | 2 | 0.89 | 0.64 | 11 |
| takiji | haha | 211 | 0.7726 | 0.3625 | 0.6113 | 8 | 16.23 | 0.04 | 7 |
| takiji | yukino | 565 | 0.2493 | 0.0695 | 0.4080 | 5 | 6.25 | 0.28 | 7 |
| unno | daino | 289 | 1.1157 | 0.1607 | 0.5731 | 10 | 11.29 | 0.34 | 5 |
| unno | kagaku | 108 | 1.0242 | 0.1328 | 0.5156 | 6 | 2.15 | 0.91 | 7 |
| unno | kibutsu | 250 | 1.7748 | 0.2114 | 0.4305 | 7 | 13.68 | 0.06 | 7 |
| unno | neon | 182 | 1.1837 | 0.1312 | 0.4246 | 5 | 7.58 | 0.18 | 7 |
| unno | tsuki | 260 | 1.7667 | 0.2264 | 0.4269 | 7 | 15.37 | 0.03 | 7 |
| watanabe | akai | 194 | 2.1726 | 0.5709 | 0.5182 | 8 | 5.25 | 0.73 | 5 |
| watanabe | aruhaha | 120 | 0.8454 | 0.0944 | 0.5616 | 7 | 8.00 | 0.33 | 7 |
| watanabe | shohai | 298 | 0.5499 | 0.1965 | 0.6890 | 12 | 9.22 | 0.68 | 5 |
| watanabe | uso | 136 | 0.7294 | 0.0809 | 0.5565 | 7 | 2.46 | 0.93 | 7 |
| yamada | musashino | 166 | 0.3080 | 0.0362 | 0.7765 | 15 | 9.88 | 0.83 | 5 |
| yamashita | ruten | 26 | 1.0755 | 0.2078 | 0.5730 | 5 | 8.74 | 0.12 | 9 |
| yokomitsu | jikan | 150 | 3.6477 | 0.6335 | 0.6743 | 18 | 19.55 | 0.36 | 7 |
| yokomitsu | kikaiy | 244 | FF | | | | | | |
| yumeno | koko | 302 | 2.6289 | 0.2347 | 0.4816 | 10 | 7.12 | 0.71 | 5 |

## Results

Though there are several texts (complete, dialogue, narrative) which do not accord with the Hyperpascal distribution in any case of our different pooling intervals, the overall result satisfactorily indicates that the law of sentence length distributions can also be effectively applied to Japanese. Each individual text which does not conform to the law must be studied separately. If there had been some kind of regulations or restrictive conditions in their creation, and if only these could be detected and allowed for in the analysis, it would be unlikely that such deviations should occur. In the considerable cases of direct speech, testing itself was impracticable simply because there was too little or no dialogue data. The first step in a new direction has been taken, and further development could be expected should sentence lengths be measured by counting the number of clauses. This project will follow shortly.

## Writers and Works

Ryunosuke Akutagawa: Giwaku, Jigokuhen, Kage, Kaika no Otto, Aru Katakiuchi no Hanashi, Aruhi no Oishi Kuranosuke, Oritsu to Kora, Rashomon, Uma no Ashi
Sakaguchi Ango: Ishi no Omohi
Takeo Arishima: Chiisaki mono he, Hitofusa no Budo, Kaji to Pochi, Kankan Mushi, Oyako

Osamu Dazai: Ai to Bi nitsuite, Joseito, Kamome, Kashoku, Kirigirisu, Koten Fu, Hashire Merosu, Osan, Oto nitsuite, Sado, Ubasute
Goro Hashimoto: Chizu ni nai Machi
Hatsunosuke Hirabayashi: Yamabuki Cho no Satsujin
Tamio Hojo: Gantai Ki
Tatsuo Hori: Banka, Hana wo moteru Onna, Hoo no saku Koro, Sei Kazoku, Tabi no E
Katsunobu Itakura: Goshiki Onsen Suki Nikki, Haru no Kamikochi he, Yama to Yuki no Nikki
Sachio Ito: Hamagiku, Kooroku, Kyonen, Nanako, Nogiku no Haka
Motojiro Kajii: Deinei, Fuyu no Hae, Kobi, Remon, Setsugo,
Katai Tayama: Ippeisotsu, Shojo Byo, Tokoyo Goyomi
Kan Kikuchi: M Koshaku to Shashinshi, Seni no Tachiba, Shimabara Shinju, Shusse, Wakasugi Saiban Cho
Rohan Koda: Shonen Jidai
Denji Kuroshima: Ana, Dempo, Mogura to Rakuban, Mon, Nusumu Onna, Sato Dorobo, Tongun, Zensho
Shinichi Makino: Kinada Mura, Tsurube to Gekko to
Takitaro Minakami: Yamanote no Ko
Yukio Mishima: Hashi Zukushi
Yuriko Miyamoto: Akarui Kaihin
Kenji Miyazawa: Karasu no Hokutoshichisei
Masayoshi Murai: Soba no Aji to Kuikata Mondai
Sakunosuke Oda: Keiba, Osaka Hakken
Ogai Mori: Asobi, Futari no Tomo, Kazuisuchika, Niwatori, Shokudo
Kanoko Okamoto: Karei, Kingyo Ryoran, Rigyo, Sushi, Tokaido Gojusantsugi
Toshiro Sasaki: Kikansha, Midori no Me
Toson Shimazaki: Bumpai, Fune, Mebae, Namiki, Shishu
Shizu Shiraki: Saite yuku Hana
Soseki Natsume: Buncho, Henna Oto, Koto no Sorane, Sakubutsu no Hihyo, Tegami
Miekichi Suzuki: Ogon Cho
Takiji Kobayashi: Hahatachi, Yuki no Yoru
Juza Unno: Daino Shujutsu, Kagaku ga Heso wo mageta Hanashi, Kibutsudo Jiken, Neon Yokocho Satsujin Jiken, Tsuki no Sekai Tankenki
On Watanabe: Akai Entotsu, Aru Haha no Hanashi, Shohai, Uso
Bimyo Yamada: Musashino
Risaburo Yamashita: Ruten
Riichi Yokomitsu: Jikan, Kikai
Kyusaku Yumeno: Kokonatto no Mi

(All the texts above were downloaded from http://www.aozora.gr.jp/index.html, with the exception of Yukio Mishima's "Hashi Zukushi." This one, still in copyright, was from *Shincho-Bunko no 100 Satsu: CD-ROM Version*, 1995.)

## References

**Altmann, Gabriel** (1988). Verteilungen der Satzlängen. *Glottometrika 9, 147–169.*
**Altmann, Gabriel; Köhler, Reinhard** (1996). „Language Forces“ and synergetic modelling of language phenomena. *Glottometrika 15, 62-76.*

**Arai, Hiroshi** (2001). On lognormal distributions of sentence length. *Hitotsubashi University Journal 125(3), 75–100.* (In Japanese).

**Cohen, Jacob** (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Lawrence Erlbaum Associates, 2nd edition.

**Fucks, Walther** (1968). *Nach allen Regeln der Kunst*. Stuttgart: Deutsche Verlagsanstalt.

**Grotjahn, Rüdiger; Altmann, Gabriel** (1993). Modelling the distribution of word length: some methodological problems. In: Reinhard Köhler; Burghard B. Rieger (eds.), *Contributions to Quantitative Linguistics: 141-153*. Dordrecht/Boston/London: Kluwer Academic Publishers.

**Kelih, Emmerich; Peter Grzybek** (2004). Häufigkeit von Satzlängen: Zum Faktor der Intervallgröße als Einflussvariable. *Glottometrics 8, 23–41.*

**Kelih, Emmerich; Peter Grzybek** (2005). Satzlänge: Definition, Häufigkeiten, Modelle. *Zeitschrift für Computerlinguistik und Sprachtechnologie 20(2), 31–51.*

**Kjetsaa, Geir** (1978). The Battle of The Quiet Don: Another Pilot Study. *Computers and the Humanties 11, 341–346.*

**Mačutek, Jan; Altmann, Gabriel** (2007). Discrete and continuous modelling in quantitative linguistics. *Journal of Quantitative Linguistics 14(1), 81-91.*

**Mizutani, Shizuo** (1957). Notes to distributions of sentence length. *Mathematical Linguistics, 2, 22–23.* (In Japanese).

**Niehaus, Brigitta** (1997). Untersuchung zur Satzlängenhäufigkeit im deutschen. *Glottometrika 16, 213-275.*

**Sasaki, Kazue** (1976). On distributions of sentence length. *Mathematical Linguistics 78, 13–22*. (In Japanese).

**Sichel, H.S.** (1974). On a distribution representing sentence-length in written prose. *Journal of the Royal Statistical Society. Vol. 1, No. 137, 25–34.*

**Sichel, H.S.** (1975). On a distribution law for word frequencies. *Journal of the American Statistical Association, Vol. 70, No. 351, 542–547.*

**Sigurd, Bengt; Mats Eeg-Olosson** (2004). Word length, sentence length and frequency – Zipf revisited. *Studia Linguistica, Vol. 1, No. 58, 37–52.*

**Strehlow, Michael** (1997*). Satzlängen in pädagogischen Fachartikeln des 19. Jahrhunderts*. Staatsexamensarbeit, Göttingen.

**Williams, C. B.** (1940). A note on the statistical analysis of sentence-length as a criterion of literary style. *Biometrika 31, 356–361.*

**Wittek, Martin** (2001). Zur Entwicklung der Satzklänge im gegenwärtigen Deutschen. In: Best, K.-H.. (ed.*), Häufigkeitsverteilungen in texten: 219-247.* Trier: WVT.

**Yasumoto, Biten** (1965). *Introduction to Sentence Psychology*, Tokyo: Seishin-Publisher.( In Japanese).

**Yasumoto, Biten** (1966). *Frontier of Sentence Psychology*. Tokyo: Seishin-Publisher. (In Japanese).

**Yule, George Udny** (1939). On sentence-length as a statistical characteristic of style in prose: with application to two cases of disputed authorship. *Biometrika 30, 363–390.*

**Zipf, George K.** (1949). *Human behavior and the principle of least effort.* Cambridge: Addison-Wesley.