Metrical Linguistic Analysis of English Books on Environmentology

○Hiromi Ban1 and Takashi Oyabu2
1Fukui University of Technology
3-6-1, Gakuen, Fukui-shi, Fukui, 930-8505, Japan, ban@fukui-ut.ac.jp
2Kanazawa Seiryo University
10-1, Ushi, Gosho-machi, Kanazawa-shi, Ishikawa, 920-8620, Japan, oyabu@seiryou-u.ac.jp

Abstract: To confront environmental problems, a system of “environmentology” is trying to be constructed. In order to study environmentology, reading materials in English that can be considered to be indispensable. In this paper, we investigated several English books on environmentology, comparing with journalism in terms of metrical linguistics. In short, frequency characteristics of character- and word-appearance were investigated. These characteristics were approximated by an exponential function. Furthermore, we calculated the percentage of Japanese junior high school required vocabulary and American basic vocabulary to obtain the difficulty-level. As a result, it was clearly shown that the English books on environmentology have a similar tendency to literary writings in the characteristics of character-appearance, and some books are more difficult than TIME magazine.

1. INTRODUCTION
In recent years, disasters arising from extreme weather have grown both in scale and frequency. To confront environmental problems, a system of “environmentology” is trying to be constructed[1].

In order to study environmentology, reading materials in English is considered to be indispensable. If we have beforehand enough knowledge of the features of English in this field, reading of the texts will become easier.

In this study, we investigated several English books on environmentology, comparing with journalism in terms of metrical linguistics.

2. METHOD OF ANALYSIS & MATERIALS
The materials analyzed here are as follows:
Material 1: Carson, Rachel,
Silent Spring
Material 2: DesJardins, Joseph R.,
Environmental Ethics
Material 3: Friedman, Thomas L.,
Hot, Flat, and Crowded
Material 4: Gore, Albert,
Earth in the Balance
Material 5: Hansen, James,
Storms of My Grandchildren
Material 6: Levin, Simon,
Fragile Dominion
Material 7: Lomborg, Bjorn,
The Skeptical Environmentalist
Material 8: Lovelock, James,
The Revenge of Gaia
Material 9: Nordhaus, William D.,
A Question of Balance
Material 10: Stern, Nicholas,
Blueprint for a Safer Planet

We examined the first three chapters of each material. For comparison, we analyzed the “TIME” magazine published on January 11 in 2010. The computer program for this analysis is composed of C++[2].

3. RESULTS
First, the most frequently used characters in each material and their frequency were derived. The frequencies of the 50 most frequently used characters were plotted on a descending scale. The vertical shaft is scaled with a logarithm. This characteristic curve was approximated by the following exponential function:

\[ y = c \cdot \exp(-bx) \]  

The distribution of coefficients \( c \) and \( b \) extracted from each material is shown in Fig. 1. There is a linear relationship between \( c \) and \( b \) for all the 11 materials. The values of coefficients \( c \) and \( b \) for Materials 1 to 10 are higher than those for TIME magazine. Previously, we analyzed various English writings and reported that the more journalistic the material is, the lower the values of \( c \) and \( b \) are, and the more literary, the higher the values of \( c \) and \( b \)[2]. Thus, the materials for environmentology have a similar tendency to literary writings.

![Fig. 1 Dispersions of coefficients \( c \) and \( b \) for character-appearance.](image)

Just as in the case of characters, the frequencies of the 50 most frequently used words in each material were plotted. Each characteristic curve was approximated by the same exponential function. In this case, we can see a weak positive correlation between coefficients \( c \) and \( b \) for all the materials for
environmentology. Besides, the values for Materials 1, 2, 6 and 7 are relatively similar, and we might be able to regard them as a cluster.

In order to show how difficult the materials for readers are, we derived the degree of difficulty for each material through the variety of words and their frequency[2]. That is, we came up with two parameters to measure difficulty: one is for word-type or word-sort ($D_w$), and the other is for the frequency or the number of words ($D_n$). The equation for each parameter is as follows:

$$D_w = 1 - \left( \frac{n_w}{n} \right)$$  \hspace{1cm} (2)

$$D_n = \left\{ 1 - \left( \frac{1}{n} \times \Sigma n(i) \right) \right\}$$  \hspace{1cm} (3)

where $n$ means the total number of words, $n_w$ means the total number of word-sort, $n_v$ means the required English vocabulary in Japanese junior high schools or American basic vocabulary by The American Heritage Picture Dictionary (American Heritage Dictionaries, Houghton Mifflin, 2003), and $n(i)$ means the respective number of each required or basic word. Then, in order to make the judgments of difficulty easier for the general public, we derived the principal component scores. The results are shown in Fig. 2.

![Fig. 2 Principal component scores of difficulty.](image)

According to Fig. 2, in the case of the required vocabulary, TIME is by far the most difficult of all the materials. The most difficult of the environmentology materials is Material 9, and the second most is Material 2. As for the case of the basic vocabulary, Materials 9 and 2 are much more difficult than other 9 materials. TIME is the fifth most difficult, whose difficulty is very similar to Materials 6, 7 and 10. Therefore, we might say that while the materials for environmentology are easier to read than TIME for Japanese, some environmentology materials are more difficult than TIME for Americans.

Other metrical characteristics of each material were compared. The results are shown together in Table 1. As for the “mean word length” for 10 materials for environmentology, it varies from 5.722 to 6.271 letters. 7 materials are a little longer than TIME. It seems that this is because the materials for environmentology contain many long-length technical terms for environmentology such as CONTAMINATION, DEFORESTATION, ENVIRONMENTAL and PRESERVATIONIST.

The “number of words per sentence” for Material 2 (20.375 words) is the fewest of 10 materials. This is the only material that is fewer than TIME. Other 9 materials are 22.554 to 28.347 words. From this point of view, the materials for environmentology seem to be rather difficult to read.

The “frequency of auxiliaries” of 10 materials for environmentology varies from 1.048% to 2.398%. All 10 materials contain more auxiliaries than TIME. Therefore, it might be said that while the writers of the books on environmentology tend to communicate their subtle thoughts and feelings with auxiliary verbs, the style of TIME magazine can be called more assertive.

4. CONCLUSIONS

We investigated some characteristics of character- and word-appearance of some famous English books on environmentology. As a result, it was clearly shown that English materials for environmentology have the same tendency as English literature in the character-appearance, and some books are more difficult than TIME. In the future, we plan to apply these results to education.

REFERENCES


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total num. of characters</td>
<td>60,825</td>
<td>170,456</td>
<td>138,038</td>
<td>127,594</td>
<td>126,675</td>
<td>123,980</td>
<td>133,737</td>
<td>101,152</td>
<td>96,905</td>
<td>105,399</td>
</tr>
<tr>
<td>Total num. of character-type</td>
<td>73</td>
<td>76</td>
<td>82</td>
<td>75</td>
<td>76</td>
<td>74</td>
<td>78</td>
<td>70</td>
<td>77</td>
<td>75</td>
</tr>
<tr>
<td>Total num. of words</td>
<td>10,221</td>
<td>27,180</td>
<td>23,643</td>
<td>21,402</td>
<td>20,935</td>
<td>19,803</td>
<td>25,864</td>
<td>17,678</td>
<td>15,664</td>
<td>17,835</td>
</tr>
<tr>
<td>Total num. of word-type</td>
<td>2,542</td>
<td>3,553</td>
<td>4,331</td>
<td>4,081</td>
<td>3,546</td>
<td>3,469</td>
<td>4,019</td>
<td>3,485</td>
<td>3,646</td>
<td>3,828</td>
</tr>
<tr>
<td>Total num. of sentences</td>
<td>437</td>
<td>1,334</td>
<td>956</td>
<td>755</td>
<td>929</td>
<td>746</td>
<td>1,064</td>
<td>839</td>
<td>644</td>
<td>690</td>
</tr>
<tr>
<td>Total num. of paragraphs</td>
<td>99</td>
<td>257</td>
<td>165</td>
<td>163</td>
<td>237</td>
<td>130</td>
<td>261</td>
<td>108</td>
<td>133</td>
<td>146</td>
</tr>
<tr>
<td>Mean word length</td>
<td>5.95</td>
<td>6.271</td>
<td>5.838</td>
<td>5.962</td>
<td>6.045</td>
<td>6.261</td>
<td>5.905</td>
<td>5.722</td>
<td>6.186</td>
<td>5.974</td>
</tr>
<tr>
<td>Repetition of a word</td>
<td>0.023</td>
<td>7.698</td>
<td>5.459</td>
<td>5.244</td>
<td>5.006</td>
<td>5.709</td>
<td>6.435</td>
<td>5.073</td>
<td>6.576</td>
<td>6.184</td>
</tr>
<tr>
<td>Comma/sentence</td>
<td>1.156</td>
<td>1.112</td>
<td>1.584</td>
<td>1.470</td>
<td>1.268</td>
<td>1.643</td>
<td>1.157</td>
<td>1.271</td>
<td>1.107</td>
<td>1.335</td>
</tr>
<tr>
<td>Freq. of relatives (%)</td>
<td>2.389</td>
<td>3.563</td>
<td>3.072</td>
<td>2.999</td>
<td>2.865</td>
<td>3.616</td>
<td>3.076</td>
<td>2.749</td>
<td>2.119</td>
<td>2.092</td>
</tr>
<tr>
<td>Freq. of auxiliaries (%)</td>
<td>0.075</td>
<td>1.932</td>
<td>1.216</td>
<td>1.046</td>
<td>1.407</td>
<td>1.398</td>
<td>1.659</td>
<td>1.431</td>
<td>1.303</td>
<td>2.398</td>
</tr>
</tbody>
</table>