

A Comparison of Texts for Writing and Original Speaking Texts by Correspondence Analysis

Koji UENISHI

University Education Center, Fukuyama University

ABSTRACT

This paper explores the differences of the four texts used in speaking and writing classes. Specifically, the texts were analyzed using a readability tool and Correspondence Analysis (hereafter, CA). An analysis of the data suggests that, in terms of readability and content, there were both similarities and differences among the four texts. More precisely, regarding readability, the two texts for the speaking class are closer to the level of English textbooks which junior high school students (seniors) use, compared to the other two texts. In addition, concerning the content including vocabulary, CA shows that the two texts, *Writing 1* and *Writing 2*, are similar to each other, but those texts are different from the other two texts, *Speaking 1* and *Speaking 2*.

Keyword: Correspondence analysis, Readability, Teaching texts

INTRODUCTION: EARLIER RESEARCH

With regard to readability, English educators or researchers tend to utilize Flesch Reading Ease or Flesch Kincaid Grade Level, both of which are very commonly used as useful indices to measure the difficulty of English texts. However, these readability tests were produced mainly on the basis of the number of words and length of sentences, and they need to be adapted for use with English textbooks approved by the Japanese Ministry of Education, Culture, Sports, Science, and Technology because such textbooks feature specially sequenced grammar and vocabulary content. Therefore, by taking into consideration the content of English textbooks, the development of useful readability tools began in Japan, (Fukui and Ozasa, 2017; Uenishi, Watanabe, and Ozasa, 2017). The most recent readability test, Fukui-Ozasa Year Level (Ver. 3.5nhnc1-6), was developed in 2016, based upon textual data of EFL textbooks for junior and senior high schools in Japan.

Regarding the linguistic research using Correspondence Analysis (CA), there have been only a few research papers in which the obtained data were analyzed and discussed based upon statistical analysis. One important research paper, Uenishi et. al (2017), serves to quantitatively re-evaluate EFL experts' judgment from a different angle on three sets of current EFL textbooks by using CA. Results of the analysis confirm the significance of the qualitative empirical judgment made by the experts in the English education field. It was found that on the first dimension of a CA analysis map, *Discovery* was distant from *Prominence* while *Crown* was somewhat close to *Prominence*, which is explained by the category of 'natural vs. controlled structure.' On the second dimension of the CA map, it was also shown that *Crown* was distant from *Prominence*, while *Discovery* was close to *Prominence*, which is explained by the category of 'focused vs. less-focused discourse'. These findings support the experts' intuitive judgment

that *Discovery* is different from the other two textbooks and that it should be excluded from the source of criterion sentences.

Another CA-related study analyzes five Japanese EFL textbooks using CA (Ozasa, Kawamura, Umamoto, and Matsuoka, 2018). The results of the analysis revealed that the five textbooks could be explained by the four dimensions of the discourse: (1) natural vs. drilled-centered, (2) easy vs. difficult, (3) focused vs. unfocused and (4) varied vs. single practice. All of their features were made evident in radar-graph consisting of all the four dimensions.

Furthermore, Uenishi (2018) analyzes four texts from original teaching materials by using CA. As a result, the CA map shows that the four texts could be explained by the two dimensions: (1) interactive vs. less-interactive, and (2) academic vs. less-academic. Also, it is shown that in the intermediate English level class, the results of the CA are rather fit for student awareness of the texts. Likewise, Uenishi (2019) analyzes the other four texts from the same teaching materials by use of CA and discusses the similarity and difference among them. The results show that the four texts could be explained by the two dimensions: (1) nature-created vs. human-created, and (2) business negotiating vs. less-business negotiating.

When it comes to text analysis, only recently Fukui and Watanabe (2019) show the significance of the specific method to classify the texts, using CA. In order to distinguish between normal CA and their CA tool to analyze texts, they call it Textual Analyses with Correspondence Analysis.

OBJECTIVES

The present paper has two aims: The first aim is to explore the readability of four texts: *Speaking 1*, *Speaking 2*, *Writing 1*, and *Writing 2*; the second aim is to explore the similarity / dissimilarity of the four texts by using CA. In particular, focusing on the four texts introduced in class, the readability and the results of CA are compared and discussed. The following are the research questions in the present paper:

- (1) Are there differences among the four texts regarding readability?
- (2) What kind of dimension explains the similarity / difference among the four texts?

METHOD

The present study employed the readability test and a one-way CA in order to explore the relationships among the four nominal variables, i.e., *Speaking 1*, *Speaking 2*, *Writing 1*, and *Writing 2*. As describe in the above, in this research an original readability tool, Fukusi-Ozasa Year Level (Ver. 3.5nhnc1-6) developed by Japanese researchers, is utilized to explore the difficulties of the four texts. Therefore, the commonly known readability tools, Flesch Reading Ease and Flesch Kincaid Grade Level, are not used in this research.

Also, in order to clarify the relationships among the four texts, CA is utilized in this research. CA is a data reduction procedure like factor analysis, and basically describes the relationships among the four nominal variables (texts) while simultaneously describing the relationships between the nominal categories (dimensions) for each variable.

In this analysis, first, a cross tabulation table of the four texts was prepared as a basic datum for the present CA, using four text corpora for *Speaking* and *Writing* classes. This is essentially a cross-textbook word-frequency table made by using the word-frequency data of the four text

corpora computed, using AntConc, a vocabulary processing tool. In this frequency table the most frequent 100 content words were picked up from the frequency list after deleting function words and proper nouns, and their frequency values were placed for each of the four texts. (Since four kinds of 100 word lists were incorporated into one cross-text word-frequency table, the number of words of the cross-tabulation table was more than 100.)

TEXT CONTENT

The original speaking textbook, *English for World Travel*, was designed to encourage students to become more motivated to learn English, and enhance their English proficiency, especially speaking ability, through use of the texts, referring to the teachers' various traveling experiences. On the other hand, the texts for writing classes are from practical business English and the contents tend to focus on business-related matters. Below are the specific texts covered in this research:

Speaking 1

The main character, Yuta, visits several countries / areas for sightseeing. He flies from Singapore to London, via Amsterdam. In the UK, the way you buy a train ticket affects the price you pay. Then Yuta goes to a pub and feels the difference between a British pub and a Japanese *izakaya*. Further, he arrives in Vancouver, Canada and has a good time for a homestay. One day he crosses the Canada-US border. Immigration officials usually want concise clear answers to fairly direct questions: the purpose of Yuta's visit and his plan to stay. Also, they want to see the evidence of his plan to leave US and enough money for his trip. Finally, he visits the Museum of Flight in Seattle.

Speaking 2

The main character, Hana, visits India for sightseeing and goes to Thailand to see her brother working at his company's local headquarters in Bangkok. In China she is meeting a Chinese doctor who is organizing some events for her concerning medicine in Shanghai and becomes sick on her way to Germany to join the international conference. Finally, Hana stops by Paris and France, and finally flies to Berlin in Germany to participate in the conference there. Then she meets Dr. Ritter at breakfast, and they sort out a computer problem in the presentation room.

Writing 1

There were three contents in Writing 1. In the first content a businessperson was going to go on business on the west coast of USA and got a handout of 'Outline of Secure Air Travel' from the executives. The memos of the handout were described precisely to travel abroad safely, such as where to sit while waiting for boarding at the departure gate. In the second content the CEO got upset about office members using digital gadgets in the corporate meetings and announced a new company wide policy; that is, banning laptops, smartphones and other digital gadgets in the official meetings. The policy led to recognition of significance of the meetings. The third one is about the story in which staff who took charge of job interviewing started discussion to select new members for the company. Almost all of them were middle-aged or senior people and various types of applicants such as people who retired or unemployed or just quit a job.

Writing 2

There were three contents in Writing 2. First, the main character talked to other business people about brushing up business communication skills. He has made strenuous efforts to improve the skills and joined a lot of networking parties. He was impressed by American colleagues because they started with small talk and then moved on. Second, the content was about the decline of America's pet population. One of the causes was that fewer Americans live in families. Another reason was economics and people tended to scrimp and save money. Then, they talked about values of pets to humans, such as enrichment of children's lives. Third, in the skit they started talking about money and happiness in life because the trip to Bhutan sounded like a life-changing experience. Bhutan is the only country in the world that includes its citizens' happiness in the equation of the national wealth.

RESULTS AND DISCUSSION

Research Question 1

In this section Research Question 1: *Are there differences among the four texts regarding readability?* is discussed. By using the readability tool, Fukui-Ozasa Year Level, Ver.3.5nhnc1-6, the readability of the four texts is measured and discussed.

As shown in Table 1, the year-level readability of the text, *Writing 2*, is the highest among all four texts. This means that *Writing 2* is the most difficult text of all these texts. The second most difficult text is *Writing 1*, and the other two texts, *Speaking 1* and *Speaking 2* have almost the same difficulty regarding their readability. Judging from the year level corresponding to the readability, the texts for speaking classes are equivalent to the level of English textbooks which junior high school students (seniors) use, whereas the other two texts, *Writing 1* and *Writing 2*, are closer to the level of textbooks suitable for freshmen and seniors at high school, respectively.

Table 1: Readability of Each Text

Text	Mean	Y. level
Speaking 1	2.75	3.26
Speaking 2	2.98	3.43
Writing 1	3.83	4.35
Writing 2	4.41	5.95

Research Question 2

In this section, Research Question 2: *What kind of dimension explains the similarity / difference among the four texts?* is discussed. As mentioned in the former section, the target texts are the following: *Speaking 1*, *Speaking 2*, *Writing 1*, and *Writing 2*. The texts are analyzed using CA, and the data results are discussed.

Table 2 shows the statistics of the present CA, i.e., the eigenvalues, coefficient correlations, contribution rates and cumulative contribution rates of the analysis. As the contribution rates indicate in Table 2, the relationships among the four nominal variants (texts) are explained 60.2% by the 1st dimension and 25.6% by the 2nd dimension. This means that the first dimension is much more powerful in explanatory power than the second one.

Table 2: Eigenvalues, Contribution Rates, Etc

	<i>1st Dimension</i>	<i>2nd Dimension</i>	<i>3rd Dimension</i>
<i>Eigenvalue</i>	0.149	0.063	0.035
<i>Correlation</i>	0.386	0.252	0.188
<i>Contribution rate</i>	0.602	0.256	0.142
<i>Cumulative contribution rate</i>	0.602	0.858	1

Table 3 shows the points of the three dimensions as a summary of the relationships among the four variants. As shown in Table 3 and Figure 1, on the first dimension, *Speaking 1* and *Speaking 2* are located on the high plus side of the first dimension (0.436 and 0.307) respectively, while *Writing 1* (-0.412) and *Writing 2* (-0.362) are located on the minus side of its dimension.

Concerning the description of dimensions that best explains the similarity / difference among the four texts, on the first dimension, *Speaking 1* (0.436) is higher in points than *Speaking 2* (0.307) only by 0.129 points, while *Writing 1* (-0.412) is close to *Writing 2* (-0.362), the difference being only 0.05 points.

Table 3: Weighted Points of Four Texts

Text	<i>1st Dimension</i>	<i>2nd Dimension</i>	<i>3rd Dimension</i>
Speaking 1	0.436	-0.286	0.006
Speaking 2	0.307	0.421	-0.03
Writing 1	-0.412	-0.055	-0.252
Writing 2	-0.362	-0.002	0.28

The difference between the *Speaking* group and the *Writing* group could best be explained by the category of 'travel-centered vs. business-centered'; in *Speaking 1* and *Speaking 2* the main characters travel some good places in foreign countries and enjoy talking to local people in each pace and feeling local culture. On the other hand, in *Writing 1* and *Writing 2*, as far as the texts are concerned, mainly the contents of them are business-related matters, such as use of laptops in the formal meetings, job interviewing and honing business communication skills. That is to say, in *Speaking texts* the main characters enjoy traveling in foreign countries and talking to local people without any business matters, except for joining the conference, while in *Writing texts* business-related things tend to be focused on. For this reason, the first dimension was termed 'travel-centered vs. business-centered'.

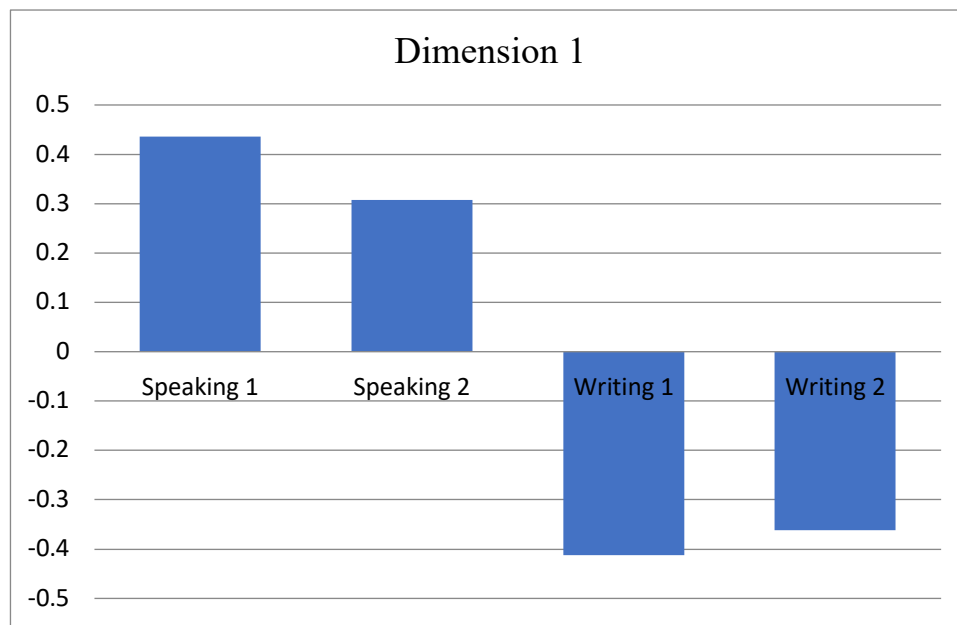


Figure 1: Weighted Points on Dimension 1

On the second dimension, as is clear in Table 3 and Figure 2, the difference between *Speaking 1* and *Speaking 2* on its dimension could best be explained by the category of 'academic vs. less academic'. In *Speaking 2*, as far as the content is concerned, the main character first visits the countries, India and Thailand to see his brother working there. Then she goes to China to see a Chinese doctor and moves to Germany via France to participate in the international conference. She has some trouble when she gives a presentation at the conference. On the other hand, in *Speaking 1* the main character travels around the several countries such as Singapore, UK and Canada to do the sightseeing and feels the cultural differences. Also, he has some trouble with homestay and crossing the border, etc. In short, it means that in *Speaking 2* not only traveling content but also more academic vocabulary and sentences are presented, while in *Speaking 1* the contents tend to focus on the matters related to traveling, not academic things. The other two texts (*Writing 1* and *Writing 2*) seem to be in the middle of the graph. Therefore, the second dimension was termed 'academic vs. less academic'.

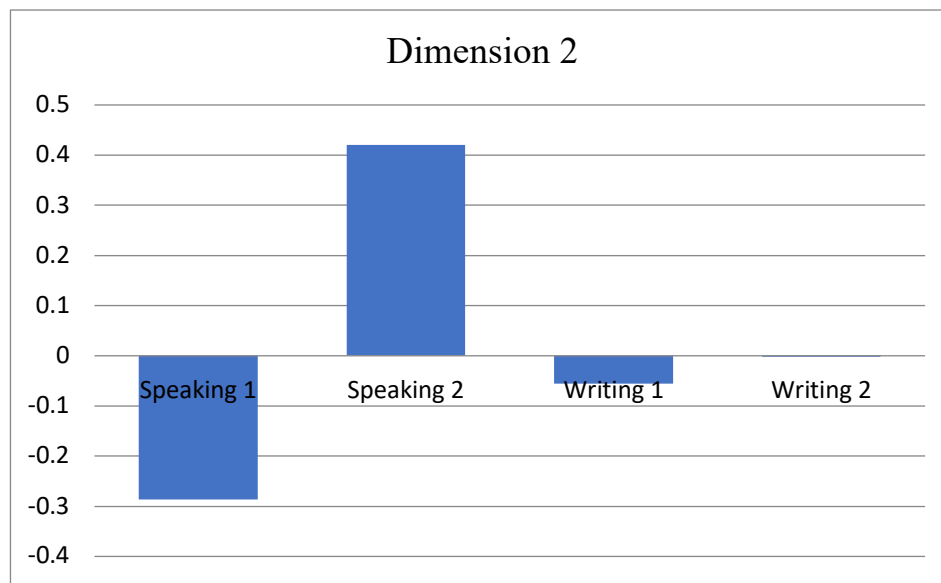
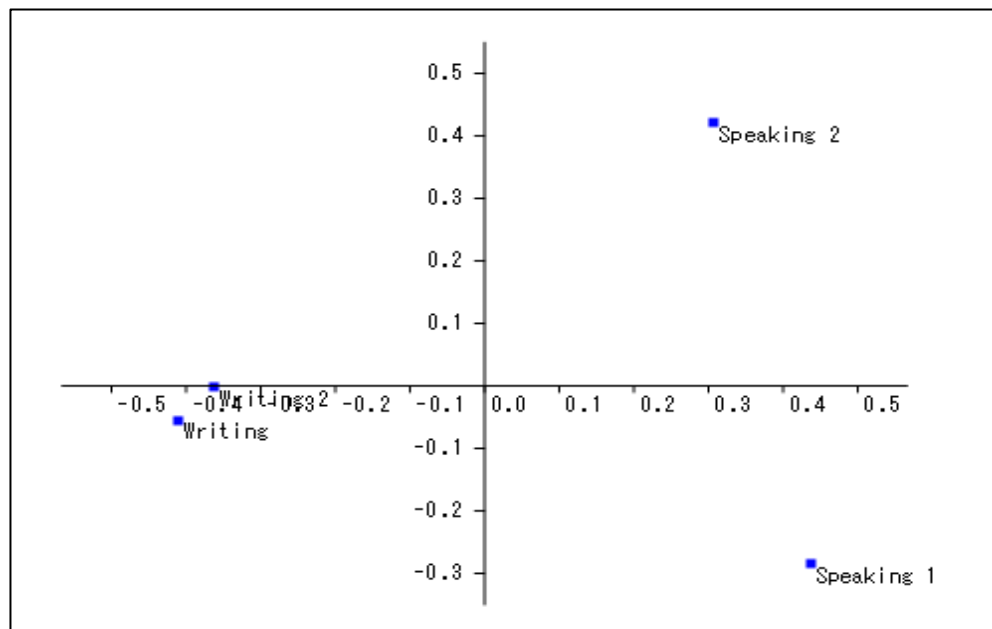


Figure 2: Weighted Points on Dimension 2

It will be impossible to clarify the meaning of the dimensions (the x-axis and y-axis) in Figure 3, without a detailed and critical examination of their contents concerning the similarity / difference of the four texts. Therefore, based upon the results gained through the CA and the text contents, it is necessary to confirm the meaning of each dimension.



x-axis: travel-centered vs. business-centered
y-axis: academic vs. less academic
Figure 3: Biplots on CA Map of Four Texts

Figure 3 illustrates the spatial relationships among the four nominal variants or texts in focus, in which the x-axis represents the first dimension, or the category of 'travel-centered vs. business-centered' and the y-axis shows the second dimension, or the category of 'academic vs. less academic'. As is clear in Figures 1 & 3 and Tables 3 & 4, *Speaking 1* and *Speaking 2* are

distant from *Writing 1* and *Writing 2* on the *x*-axis (the first dimension) at least by 0.669 points, and *Speaking 2* is distant from *Speaking 1* on the *y*-axis (the second dimension) by 0.707 points. This means that *Speaking texts* are different from both *Writing texts*, each having their own characteristic features, and that those texts vary by differences in perspective, as shown by the coordinate axes in the figure.

Table 4. Top 20 Words of Dimensions 1 and 2

	Dimension 1 Word		Dimension 2 Word	
1	attendant	2.926	room	4.454
2	publican	2.888	receptionist	3.797
3	ok	2.784	will	2.528
4	please	2.595	I	2.235
5	museum	2.567	up	1.819
6	receptionist	2.28	morning	1.779
7	thank	2.224	no	1.745
8	here	2.036	he	1.474
9	room	2.024	also	1.212
10	oh	1.993	good	0.982
11	go	1.94	tomorrow	0.954
12	tomorrow	1.913	yes	0.813
13	morning	1.838	we	0.76
14	want	1.62	so	0.721
15	over	1.437	have	0.673
16	yes	1.336	from	0.639
17	get	1.209	thank	0.621
18	would	1.128	take	0.6
19	you	1.062	well	0.574
20	very	1.052	do	0.524

CONCLUSION

In this paper through CA the four texts from business English and the original speaking textbook were analyzed and discussed. First, the readability of the four texts was explored. It was found that *Writing 2* is the most difficult among all the texts, and the second most difficult one is *Writing 1*. As a result, it was found that those texts are closer to the level of textbooks at high school.

Second, the four teaching texts proved to be different from each other except for *Writing 1* and *Writing 2*. The CA map shows that on the first dimension, *Speaking 1* and *Speaking 2* were distant from *Writing 1* and *Writing 2*. The difference was described by the category of 'travel-centered vs. business-centered'. On the second dimension, *Speaking 2* was distant from *Speaking 1* while the other two texts were midway between both texts. The difference was explained by the category of 'academic vs. less academic'.

Therefore, further studies will be useful to modify the original texts or create new, better ones. Future research should focus on the analysis of more texts or textbooks, using CA in order to create better textbooks for English learners.

References

- Fukui, M., and Ozasa, T. (2017). Ozasa-Fukui Year Level, a Readability Measuring Tool: Its System and Measuring Program. *Journal of ICTATLL Japan*, Vol. 4, 1-12.
- Fukui, M., and Watanabe, K. (2019). Tekisuto CR bunseki no gosuu chouseihou to tango no senntaku—Senyou program no kaihatsu—[Number Adjustment and Selection of Words for Textual Analyses with Correspondence Analysis: Development of a special program]. A paper presented at Workshop of ICTATLL Japan.
- Ozasa, T., Kawamura, K., Umamoto, T., and Matsuoka, H. (2018). Correspondence Analysis of Four Japanese Primary School EFL Textbooks Published during 1900-1908 and a Current Junior High School Textbook. *Journal of ICTATLL JAPAN*, Vol. 5, 57-68.
- Tono, Y. (2000). A Corpus-based Analysis of Interlanguage Development: Analysing Part-of-speech Tag Sequences of EFL Learner Corpora. *Proceedings of PALC'99: Practical Applications in Language Corpora*, 15-18 April, 1999, 323-340.
- Uenishi, K. (2019). A Comparison of Original Teaching Texts by Correspondence Analysis. *International Journal of Management and Applied Science*, Vol. 5 (11), 54-58.
- Uenishi, K. (2018). Correspondence Analysis of Teaching Materials and Potential Relationships between the Materials and Student Awareness. *International Conference on Education, Psychology and Learning (ICEPL), Conference Proceedings*, 158-171.
- Uenishi, K., Watanabe, K., & Ozasa, T. (2017). Correspondence Analysis of Three Japanese EFL Textbooks: In Reference to the Improvement of a Criterion Measure for a Readability Tool. *International Conference on Education, Psychology and Learning (ICEPL), Conference Proceedings*, 49-60.