ORIGINAL RESEARCH:

Mining text data to analyze students' portfolios on team-teaching for language education

Takako Unetani[1]

Abstract

The purpose of this research was to examine the potential benefits of text mining in the field of language education and its effectiveness in extracting meaningful information from student portfolios in order to provide useful feedback to students and the educational community. The analysis was focused on an examination of the similarities and differences in awareness on teamwork concepts between Japanese and non-Japanese students in an international university located in Japan, especially in terms of their recognition of merits and demerits of team teaching and their attitude towards problem solving. Correspondence analysis was used along with a series of text mining tools. The results of the analysis are discussed in regards with possible future applications of data mining in language education.

Keywords: Correspondence analysis, Data mining, Japanese language, Language education, Team-teaching, Text mining.

Introduction

Using of portfolio is a well-known educational tool for keeping record of individuals' learning efforts and processes. However, it is difficult to gain meaningful information from such textual data especially in cases where the amount of data to be examined is massive. To overcome such a problem, application of text mining may prove to be useful because it enables us to extract quantitative indices from textual data that can later be analyzed statistically. The information obtained may provide meaningful information back to the educational community, as well as to the students and the instructors.

Therefore, I decided to focus on one of the important questions regarding student preferences in teamwork. To achieve this goal, I created online questionnaires and requested students to review their group performances from the perspective of communication. The data I collected were written by two groups of students, Japanese and non-Japanese, and consisted of the following three types of answers to three questions in a questionnaire:

1. The positive aspects of team-teaching (G),
2. The negative or/and problematic aspects of team-teaching (B),
3. The efforts that students made to solve the problems and the difficulty that they encountered during team-teaching (P).

By collecting the described data, I attempted to answer the following questions concerning the experiences of the students with team-teaching:

1. What do Japanese and non-Japanese students think are the merits and demerits of team-teaching?
2. What efforts and improvements did they make to lessen the demerits of team-teaching?
3. Are there any similarities and differences between Japanese and non-Japanese students, in recognition of the merits and demerits and in their attitude toward problem-solving?
4. Can text mining on student portfolios be a useful tool to extract meaningful information from such unstructured textual files, and in this case, can it provide feedback to students and instructors as well as educational institution as a whole?

[1]Ritsumeikan Asia Pacific University, Beppu city, Oita 874-8577, Japan e-mail: tut07764@apu.ac.jp

Methodology

The subjects for the analysis were students who had enrolled in the course titled 'Teaching Japanese as a Second Language' at Ritsumeikan Asia Pacific University (APU) located in Beppu city, Oita, Japan. One hundred and two students had registered for this class between the fall semester of 2010 and the spring semester of 2012. These students were required to submit two questionnaires per semester, and during this period, the total number of questionnaires collected was 254, with a response rate of 78.4% (Table 1).

Of the total respondents, 62.4% were Japanese and 37.4% were international students who had an advanced knowledge of Japanese and were mainly from Asian countries (e.g., Korea, China, Thailand and Vietnam). The average number of Japanese characters per questionnaire was 406 for the Japanese students and 305 for other nationalities. While the average numbers differed based on the question, i.e. G (merits), B (demerits) or P (improvements), it can be said that non-Japanese students in general wrote about 25% less than Japanese students (Table 1).

Table 1: The average number of Japanese characters written for each questionnaire

Students	***Number***	***Average number of Japanese characters***			***Average number of characters per questionnaire***	***Total number of characters of the data***
		Text G	Text B	Text P		
Japanese	159 (62.6%)	173	106	126	406	64,544
Non-Japanese	95 (37.4%)	135 (78%)	80 (75%)	90 (71%)	305 (75%)	28,962
Total	254					93,506

I used text mining to explore the data. By applying the text mining method, I could divide the text within the sentences into parts-of-speech, such as verbs, adjectives, nouns, and adverbs. Once this was completed, numerical indices could be extracted from the text and used for statistical analyses.

The software 'Tiny Text Miner' (hereafter TTM) was used to analyze Japanese textual data, an open-source software that can be easily downloaded from the Internet. Additionally, 'Mecab', a parts-of-speech and morphological analyzer, also open-source, was used for processing the text data. After applying 'TTM/Mecab' to the data, the following 6 output files were obtained:

ttm_1: Word frequency distribution
ttm_2: Word frequency distribution（maximum one occurrence per sentence）
ttm_3: Cross tabulation of the occurrences of words against grouping variables
ttm_4: Cross tabulation of the occurrences of words against grouping variables (maximum one occurrence per sentence)
ttm_5: Cross tabulation of words against words
ttm_6: Cross tabulation of words against texts

The second tool was 'correspondence analysis' (hereafter CR analysis), a statistical visualization method for examining the associations in a two-way or multi-way contingency table. According to Statsoft, in "How to analyze simple two-way and multi-way table, correspondence analysis":

> "The goal of a typical analysis is to represent the entries in the table of relative frequencies in terms of the distances between individual rows, and/or columns in a low-dimensional space."

The CR analysis was done using a module within 'R', another open-source software for various statistical computations. The next section will provide detailed procedures on how I applied TTM and CR analysis to each file (G, B and P data) along with the results obtained from the analyses.

Findings and Results

To collect data for the research, I had created online questionnaires using a software for conducting surveys. Students were asked to respond to the following three questions after each team-teaching session was over:

(1) What do you think are the merits of team-teaching? (G)

(2) What do you think are the demerits of, and/or the problems of your team-teaching? (B)

(3) What did you do to lessen the difficulty and/or to solve the problems? (P)

An example of the textual data collected after each survey is given in Table 2.

Table 2: A sample text excerpted from the online survey

	A	B	C	D	E	F	G	H
1	J	誰かが意見を述べると、必ず誰かがそれに対しての意見を言っていた						
2	J	お互いが打ち解けられて、仲良く意見を交換できたところ。初めての						
3	J	打ち合わせの時、無駄話をせずに、効率よく話せたと思う。						
4	J	グループワークで最も良い点は協力による相乗効果だと思います。私						
5	J	1番は、1人1人が本当に協力し合えた模擬授業であったことです。時間						
6	J	全員が積極的にグループワークに参加していた点です。２人づつでペ						
7	J	みんな一様に自分の意見をもってミーティングに参加していました。						
8	J	チーム全員が非常に協力的で、話し合いも意見を出し合いながら、全						
9	J	自分の担当パートを決めることで、みんな責任を持たないといけなく						
10	J	グループでよかった点として５つあげることができます。それは各ペ						
11	J	良かった点は、３つあります。一つは、授業中に役割分担をしてパー						
12	J	早いうちからメンバーそれぞれに役わりをあてて取り組んでいたので						
13	J	それぞれの役割に別れて分担して作業をやった分効率よくは出来たと						

The TTM was applied and the results were recorded, from which the most frequently used words were extracted. First, I prepared three types of input data (G, B, P) by saving each file separately in a 'csv format' (comma separated values). Then, I processed each file using TTM and used the following output format:

Output format: ttm_3

Parts of speech for output: nouns (hereafter N), adjectives (hereafter A) and verbs (V)

Optional files: none

After processing, a list of verbs, adjective and nouns was obtained. Among the three lists, file G had the highest number of words extracted (Table 3).

Table 3: Number of the words extracted from each file

File name	*Total number of words extracted*
File G (Merits)	1540
File B (Demerits)	842
File P (Improvements)	1334

As an example, the word frequency distribution for file B is provided in Table 4:

Table 4: Excerpt from the word frequency distribution for file B

	A	B	C	D	E
1	語 "Word"	品詞 "Parts of speech"	品詞細分類 "Detailed classification"	Foreign	Japanese
2	する "do"	動詞 V	自立 "independent"	50	102
3	思う "think"	動詞 V	自立	42	41
4	時間 "time"	名詞 N	副詞可能 "could be a	30	52
5	できる "be able to do"	動詞 V	自立	18	61
6	ある "there be (for things)"	動詞 V	自立	22	47
7	ない "not"	形容詞 A	自立	22	33
8	なる "become"	動詞 V	自立	16	34
9	難しい "difficult"	形容詞 A	自立	17	32
10	人 "person/ people"	名詞 N	一般 "regular"	15	33
11	問題点 "problem"	複合名詞 CN	複合名詞	17	28
12	授業 "simulation class"	名詞 N	サ変接続 "can be use	9	35
13	ミーティング "meeting"	名詞 N	一般	9	29
14	グループワーク "group work"	複合名詞 CN	複合名詞	13	25
15	グループ "group"	名詞 N	一般	12	25
16	全員 "all members"	名詞 N	一般	5	31
17	メンバー "members"	名詞 N	一般	13	21
18	問題 "problem"	名詞 N	ナイ形容詞語幹 "can	13	18
19	集まる "come together"	動詞 V	自立	10	20
20	いる "there be (for people)"	動詞 V	自立	12	18
21	意見 "opinion"	名詞 N	サ変接続	8	21
22	みんな "all members"	名詞 N	代名詞 "pronoun"	10	17

In the next stage, lists were made from synonyms and unnecessary words. Based on a dictionary of synonyms, a synonym list was made for each G, B and P file, based on 'A Japanese Lexicon (CD-ROM)' (*Nihongo Goi Taikei*). The list of synonyms for file G is shown in Table 5.

Table 5: A list of synonyms for file G

		Word	Synonym 1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
1	J	文法	敬語	尊敬	謙譲	他動詞	自動詞	使役	受身	受け身	もらう	あげる	くれる	説明
	E	grammar	polite expression	honorific	humble	transitive verb	intransitive verb	causative	passive	passive	receive	give	give (to me)	explanation
2	J	皆	人	メンバー	メンバ	皆さん	みんな	グループ	全員	チーム	クラス	~たち	~連	仲間
	E	everyone	people	member	member	everyone	everyone	group	all members	team	class	we	we	fellow
3	J	意見	考え	発言	アドバイス	コメント	アイディア	アイデア	案	策	提案	発想		
	E	opinion	thought	statement	advice	comment	idea	idea	draft	policy	suggestion	idea		
4	J	ミーティング	話合い	話し合い	話	打ち合わせ	相談	話し合う	言い合う	meeting	集まる			
	E	meeting	discussion	discussion	talk	meeting	consultation	discussing	discuss	meeting	getting together			
5	J	協力	協調	助け合える	助け合う	助け合える	サポート	フォロー	手伝い	お手伝い	助ける	ヘルプ	手伝う	
	E	cooperation	coordination	can help each other	help each other	can help each other	support	follow	help	help	help	help	help	
6	J	自分	私	個人	各自	一人	ひとり	自身						
	E	myself	I	individual	each one	one person	one person	myself						
7	J	役割	部分	分担	担当	パート	セクション	分ける						
	E	role	part	role sharing	responsibility	part	section	sharing						
8	J	出す	だす	出る	でる	出し合う	言う							
	E	contribute	contribute	come out	come out	contribute jointly	talk							
9	J	グループワーク	チムワーク	作業	仕事	ワーク								
	E	group work	team work	working	work	work								
10	J	発表	本番	プレゼンテーション	プレゼン									
	E	presentation	presentation	presentation	presentation									
11	J	積極	自発	熱意	やる気									
	E	positive	spontaneous	enthusiasm	motivation									
12	J	理解	把握	わかる	分かる									
	E	understanding	grasp	understand	understand									
13	J	授業	模擬授業	模擬										
	E	lesson	trial lesson	trial lesson										
14	J	内容	課題	構成										
	E	content	task	organization										
15	J	効率	効率的	スムーズ										
	E	efficiency	efficient	smooth										
16	J	よい	良い	いい										
	E	good	good	good										
17	J	進める	進む	進め方										
	E	proceeding	proceed	how to proceed										
18	J	思う	考える											
	E	I think	think											
19	J	準備	予定											
	E	preparation	schedule											
20	J	練習	活動											
	E	practice	activity											

As shown in the first raw from the top in Table 5, words such as *bunpo* (grammar), *keigo* (polite expression), *sonkei* (honirific), *kenjou* (humble expression) and others are grouped together as synonyms, for they all belong to grammatical concepts taught in the class. Similarly, words like *mina, minasan, minna* (everyone), *hito* (people), *memba, membaa* (member), *guruupu* (group) and others in the second raw were also put together in the same group. Additionally, variants of a word were grouped together. For example, a noun in *hiragana* was grouped with its *Kanji* variant.

Also a list of unnecessary words was made. In table 4, verbs *suru* (to do), *dekiru* (to be able to), *aru* (to be, to exist; inanimate), *iru* (to be, to exist; animate), *nai* (not), *naru* (to become) are not necessary for the analysis since they are not meaningful words compared to, for example, a verb like *atsumaru* (to come together). After processing, the unnecessary words in the list were deleted from the output. Similarly, demonstratives such as *sore* (that one) and question words such as *dare* (who) were also not necessary and therefore deleted.

Finally, TTM was applied with the lists of synonyms and unnecessary words. The format adopted for this last process by TTM was as follows:

Output format: ttm_4

Parts of speech for output: nouns, adjectives and verbs

Optional files: synonyms and unnecessary files

After processing each file G, B and P as input, three outputs were obtained; one was the output from file G (merits of team-teaching). An excerpt of the output after processing the text file G by TTM is shown in Table 6. It is a cross tabulation of the occurrence of the top 20 words that appeared most frequently against the grouping variable, Japanese and non-Japanese students. One adjective *yoi* (good), 3 verbs *omou* (think), *dasu* (expressing opinions), and *susumeru* (advance), and 16 nouns, such as *minna* (all), *iken* (opinion), *jibun* (self), *guruupu waaku* (group work), *jugyou* (simulation class), and *miitingu* (meeting), were extracted.

Table 6: The top 20 words that appeared most frequently in file G

Word	English translation	Foreign	Japanese
皆	all	87	132
よい	good	60	98
意見	opinion	54	91
自分	self	48	85
思う	think	48	78
役割	role	40	77
出す	express	30	70
グループワーク	G work	31	68
授業	class	27	62
ミーティング	meeting	28	60
文法	grammar	30	57
協力	cooperation	19	45
進める	advance	13	47
理解	understanding	18	41
練習	practice	12	46
積極	active	21	36
発表	presentation	15	38
効率	efficiency	6	45
内容	content	14	36
準備	preparation	11	32

Next was the output from file B (demerits of the team-teaching). Similarly, the text file B was processed by TTM; an output is shown in Table 7. The top 20 words that appeared most frequently and were extracted from the file included 2 adjectives, *nai* (not) and *muzukashii* (difficult), 1 verb *omou* (to think), and 17 nouns such as *minna* (all), *miitingu* (meeting), *jikan* (time), *jugyou* (class), and *mondaiten* (problem).

Table 7: The top 20 words that appeared most frequently in file B

Word	English translation	Foreign	Japanese
皆	all of them	63	113
ない	not	56	115
ミーティング	meeting	32	72
時間	time	38	62
思う	think	42	41
授業	class (=simulation class)	20	59
問題点	problem	29	47
意見	opinion	20	45
自分	self	16	44
発表	presentation	13	39
文法	grammar	17	34
難しい	difficult	18	32
役割	role	12	36
参加	participation	13	31
作業	work	13	25
練習	practice	10	24
準備	preparation	15	16
内容	content	8	19
予定	schedule	5	20
連絡	contact	6	19

Finally, there was the output from file P (attitudes toward solving problems); the text file P was processed by TTM in a similar manner. The top 20 words that appeared most frequently for file P are provided in Table 8, and include 7 verbs such as *atsumaru* (get together), *hanashiau* (discuss), *kimeru* (to decide), *iku* (to go), and *toru* (to take), and 13 nouns such as *minna* (all), *jibun* (self), *yakuwari* (role), *miitingu* (meeting), and *iken* (opinion).

Table 8: The top 20 words that appeared most frequently in file P

Word	English translation	Foreign	Japanese
皆	all of them	56	116
自分	self	34	69
役割	role	32	54
ミーティング	meeting	21	61
意見	opinion	24	45
文法	grammar	24	37
時間	time	20	41
問題	problem	17	36
練習	practice	23	28
連絡	contact	11	38
集まる	get together	7	35
話し合う	discuss	7	30
確認	confirmation	8	28
決める	decide	14	21
行く	go	5	29
とる	take	5	29
内容	content	9	24
準備	preparation	14	19
聞く	ask	12	20
考える	think	9	19

In the next stage, the results of correspondence (CR) analyses were examined. An analysis of all the two-way tables 6, 7 and 8 by the CR analysis software, provided the following results. Figure 1 shows the scatter plot of data in Table 9, for the row coordinates and in Table 10, for the column coordinates. Please note that the result is plotted in a single dimension scatter plot. This is because the cumulative contribution ratio of the analysis was 100% as shown below, which means that 100% of the inertia can be explained in a single dimension.

Eigenvalues: [1] 0.017 0.000 Cumulative contribution ratios: [1] 100 0

A clear distinction is observed in the recognition of the merits of team-teaching between non-Japanese and Japanese students (Table 10). Also Japanese students are likely to value the process in a simulation class. This is expressed by a group of words as *naiyou* (content), *happyou* (presentation), *rikai* (understanding), *kyouryoku* (cooperation), *jugyou* (class) and *dasu* (expressing an opinion), which are based on similar scores of the Japanese student coordinates scores (Figure 1). That is, the process of working cooperatively, understanding contents better, and making good presentations seems valuable to them. They evaluate these as merits.

In contrast, non-Japanese students are likely to recognize the merits of team-teaching on the fact that all the members of the group expressed their opinions and participated actively. This is expressed in words such as *minna* (all), *yoi* (good), *omou* (to think), *iken* (opinion) and *sekkyoku* (active), which have coordinate values close to that of the non-Japanese students (Fig. 1).

Table 9: Row coordinates

Row Coordinates		
Word (J)	Word (E)	Coordin. Dim.1
皆	all	−1.1037811
よい	good	−0.817476
意見	opinion	−0.6975974
自分	self	−0.5094098
思う	think	−0.8371839
役割	role	−0.1984446
出す	express	0.4862042
グループワーク	G work	0.271537
授業	class	0.4310995
ミーティング	meeting	0.1889727
文法	grammar	−0.2466253
協力	cooperation	0.5372909
進める	advance	1.8485155
理解	understanding	0.4030801
練習	practice	2.0082348
積極	active	−0.6323251
発表	presentation	0.7638073
効率	efficiency	3.4672619
内容	content	0.8131589
準備	preparation	1.208546

Table 10: Column coordinates

Column Coordinates		
Tags (J)	Tags (E)	Coordin. Dim.1
留学生	Foreign	−1.4257208
日本人学生	Japanese	0.7013996

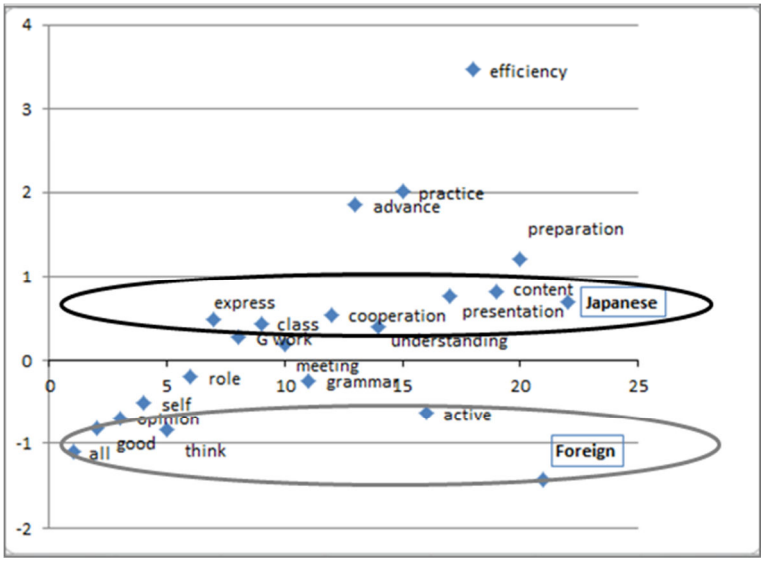

Figure 1: Scatter plot for Table 9 and 10.

The scatter plot in Figure 2 helps analyze the demerits, with the data of Table 11 as the row coordinates, and data of Table 12 as the column coordinates. Please note that Figure 2 is also single dimensional. The reason for this is the same as that for Figure1. Since the cumulative distribution ratio for the data in Tables 11 and 12 was 100% as shown below, 100% of the inertia can be explained in a single dimension.

Eigenvalues: [1] 0.021 0.000 Cumulative contribution ratios: [1] 100 0

It demonstrates that the Japanese students have unique features in recognizing the demerits (Table 11). Japanese students are likely to feel that meeting one another and expressing opinions is difficult and problematic. This is expressed by a group of words such as *sanka* (participation), *iken* (opinion) and *miitingu* (meeting), which have coordinate values close to that of the Japanese students. They also think that *naiyou* (content), especially *renshuu* (practice or activity) are difficult.

Table 11: Row coordinates

Row Coordinates		
Word (J)	Word (E)	Coordin. Dim.1
皆	all	−0.36793
ない	not	0.082832
ミーティング	meeting	0.375651
時間	time	−0.69407
思う	think	−2.55847
授業	class (=simulation class)	1.182334
問題点	problem	−0.71743
意見	opinion	0.375651
自分	self	0.982584
発表	presentation	1.229151
文法	grammar	−0.00368
難しい	difficult	−0.39819
役割	role	1.229151
参加	participation	0.556696
作業	work	−0.13345
練習	practice	0.576474
準備	preparation	−2.23074
内容	content	0.544243
予定	schedule	1.968851
連絡	contact	1.377091

Table 12: Column coordinates

Column Cordinates		
Tags (J)	Tags (E)	Coordin. Dim.1
留学生	Foreign	−1.41501
日本人	Japaese	0.706711

Non-Japanese students, on the other hand, are likely to recognize the duration of work as a problem. Such an interpretation is based on a group of words such as *jikan* (time) and *mondai* (problem), which have coordinate values relatively close to the non-Japanese students' coordinates (Figure 2). Both Japanese and non-Japanese students think that grammar explanation is difficult; this is expressed by words such as *muzukashii* (difficult), *bunpou* (grammar), and *sagyou* (work), and are plotted in the middle of the two student coordinates (Figure 2).

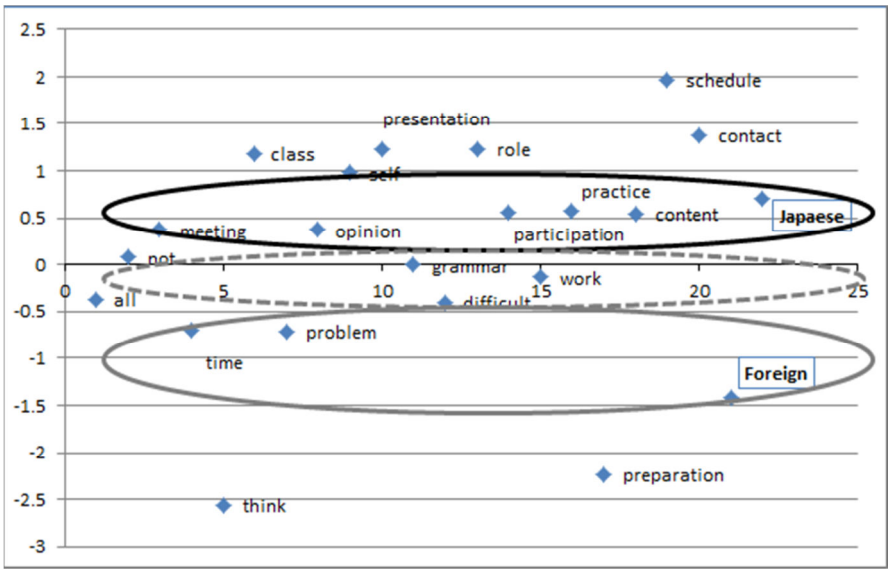

Figure 2: Scatter plot for Table 11 and 12.

Finally, improvement was interpreted through the scatter plot in Figure 3 with Table 13, for the row coordinates and Table 14, for the column coordinates. Note that Figure 3 is also single dimensional. This is because the cumulative contribution ratio of the data in Tables 13 and 14 as shown below was 100%. Therefore, the result is plotted in a single dimension.

Eigenvalues: [1] 0.029 0.000 Cumulative contribution ratios: [1] 100 0

It demonstrates a clear distinction between non-Japanese and Japanese students in their approach and attitude towards solving problems (Table 13). As pointed out before, Japanese students recognized meeting and participation as a serious problem. They seemed to focus on this point and tried to provide solutions for it. The interpretation is based on a group of words such as *miitingu* (meeting) and *naiyou* (content) that is the closest to the Japanese students' coordinates in Fig. 3. Although such words as *renraku* (contact), *atsumaru* (to get together), *hanashiau* (to discuss) and *kakunin* (confirmation) are not as close as the previous two words, they are still relatively close to the Japanese students. Thus, it can be interpreted that Japanese students in general felt that contacting, meeting and discussing were the best solution to the problem.

Non-Japanese students seem to have taken a different approach to solving problems. As pointed out before, they recognized the duration of the actual work time as a problem. To solve this, they seem to have tried to prepare themselves better. Such an interpretation is based on words like *junbi* (preparation) and *renshuu* (practice), which are close to the non-Japanese students coordinate value.

Both Japanese and non-Japanese students felt difficulties with the grammar explanation, which is a part of the content. To ease the difficulty, they seemed to have taken the same approach of asking (questions),

(expressing) opinions and deciding together with the members. Such an interpretation is based on words such as *kiku* (to ask/listen), *kimeru* (to decide), and *iken* (opinion), *minna* (all), and *jibun* (self) being plotted in between Japanese and non-Japanese students' coordinates.

Table 13: Row coordinates

Row Coordinates		
Word (E)	Word (J)	Coordin. Dim.1
皆	all	−0.182997
自分	self	−0.2405733
役割	role	−0.7760331
ミーティング	meeting	0.7029411
意見	opinion	−0.4666229
文法	grammar	−1.0482464
時間	time	−0.2121627
問題	problem	−0.1214555
練習	practice	−1.7818688
連絡	contact	1.1059488
集まる	get together	1.8432097
話し合う	discuss	1.5560413
確認	confirmation	1.134861
決める	decide	−1.1318547
行く	go	2.0932151
とる	take	2.0932151
内容	content	0.4909077
準備	preparation	−1.4409523
聞く	ask	−0.8130978
考える	think	−0.1300473

Table 14: Column coordinates

Column Coordinates		
Tags (J)	Tags (E)	Coordi. Dim.1
留学生	Foreign	−1.48764
日本人学生	Japanese	0.672206

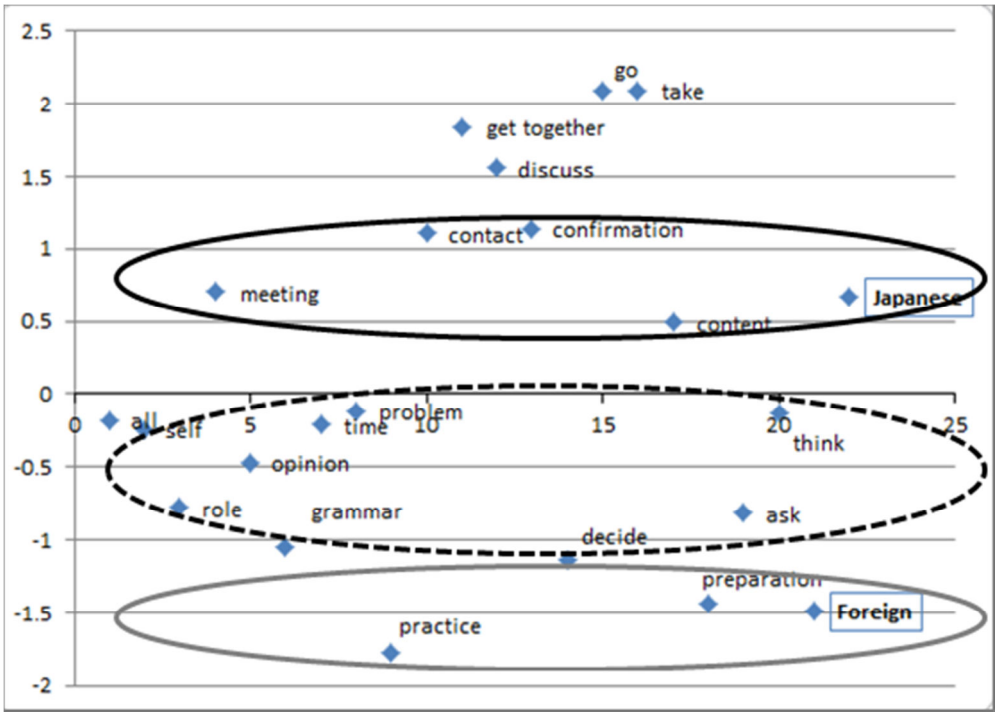

Figure 3: Scatter plot for Table 13 and 14

Conclusion

This study attempted to answer the following three questions:

1. What do Japanese and non-Japanese students think are the merits and demerits of team-teaching? What efforts and improvements did they make to lessen the demerits of team-teaching?
2. Are there any similarities and differences between Japanese and non-Japanese students in their recognition of the merits and demerits and in their approach and/or attitude towards problem-solving?
3. Can text mining of student portfolios be a useful tool for extracting meaningful information from unstructured textual files and provide feedback to students and instructors as well as educational institutions as a whole?

The first two questions have already been answered in the part on discussion and therefore will not be repeated here. As for the third question, a complete answer at this stage cannot be presented because this research has only been conducted as a primary trial. Nevertheless, the experiences gained through this trial may direct us to the needed improvements in the future. The advantages to this methodology include its quick feedback and low cost. The biggest merit of this method is speed. All the tools used in this trial are on the Internet (e.g., the survey, text mining tools, and the statistical tool R) and massive data sets can be processed with these tools in just a few minutes. Thus the information extracted from the text can be fed back very quickly to students as well as to the educational community.

In the cost perspective, all the software used for the research is open-source and hence no expenses are necessary. Therefore, for beginners who would like to experiment with this type of research, the use of such open-source software is quite economical.

The disadvantages of this methodology include its requirement of handling skills and skills for interpretation. Installation of the software is not difficult, but it takes time to get familiar with the TTM and the CR analysis, and to handle the data efficiently. Although a 'trial and error' method by individual researchers can improve their skills, a project consisting of experts of different academic fields should provide better results. The skills to interpret the scatter plots may be the most important that needs to be improved.

One limitation of this study is that the text has been analyzed in terms of the most interesting grouping variable, that is, Japanese versus non-Japanese students. However, other variables could be used for this research, such as male vs. female, freshmen vs. seniors, and good performers vs. ordinary ones. The more experiences we get in text mining analysis, the higher our skills will become, and this can be done next.

Acknowledgements

I would like to thank my students and colleagues for their time and the efforts they made for this research. The students filled in their self-evaluation sheets to reflect on their teaching performance. Professor Otake looked over the results and gave very valuable comments, and Professor Arii spent his valuable time to proof read my paper. At last I would like to thank Ms. Sato who helped me make the data sets. Without your kind consideration and assistance, I would never have completed this paper.

References

Hornik, Kurt. 2012. “The {R} {FAQ}.” http://CRAN.R-project.org/doc/FAQ/R-FAQ.html (accessed on October 10, 2012).

Matsumura, N and Miura, A. “TTM (Tiny Text Miner)”. 2011. http://mtmr.jp/ttm/ (accessed on September 16, 2011).

Matsumura, N. and Miura, A. 2009. *Jinbun & shakai kagaku no tame no tekisuto mainingu [Text mining for the researchers of humanistic and social sciences]*. Seishinshoboo.

Kudo, Taku. 2012. “Mecab (Parts-of-speech and morphological analyzer).” http://mecab.sourceforge.net/ (accessed on September 16, 2012).

NTT Communication Science Laboratories. 1999. Nihongo Goi Taikei “A Japanese Lexicon (CD-ROM)”. Iwanami Shoten.

Statsoft.. 2012. “How to analyze simple two-way and multi-way table, correspondence analysis.” http://www.statsoft.com/textbook/text-mining/ (accessed on October 10, 2012).