Text Mining Analysis of the Reconstruction Plans after the Great East Japan Earthquake

Minami Ito, Akihiro Iijima

Abstract—On March 11, 2011, the Great East Japan Earthquake occurred off the coast of Sanriku, Japan. It is important to build a sustainable society through the reconstruction process rather than simply restoring the infrastructure. To compare the goals of reconstruction plans of quake-stricken municipalities, Japanese language morphological analysis was performed by using text mining techniques. Frequently-used nouns were sorted into four main categories of "life", "disaster prevention", "economy", and "harmony with environment". Because Soma City is affected by nuclear accident, sentences tagged to "harmony with environment" tended to be frequent compared to the other municipalities. Results from cluster analysis and principle component analysis clearly indicated that the local government reinforces the efforts to reduce risks from radiation exposure as a top priority.

Keywords—Eco-friendly reconstruction, harmony with environment, decontamination, nuclear disaster.

I. Introduction

ON March 11, 2011, the Great East Japan Earthquake (the magnitude of the earthquake was 9.0) occurred off the coast of Sanriku, Japan. A massive tsunami hit the coastal area in the Tohoku region, causing catastrophic damage to human life and infrastructure. The tsunami also hit the Fukushima Daiichi Nuclear Power Plant and destroyed reactor cooling systems. This serious nuclear accident certainly obstructs the efforts to recover from the crisis. To eliminate the effects of the so-called triple disaster (strong earthquake, devastating tsunami, and nuclear accident) will be a great challenge in Japan. Therefore, it is important to build a sustainable society through the reconstruction process rather than simply restoring the infrastructure.

In our previous study [1], we discussed the trends in eco-friendly reconstruction projects and the emerging pro-environmental behavior in victims of the Great East Japan Earthquake. To accelerate efforts in the field of eco-friendly reconstruction project, the government should provide an incentive for creating a sustainable society based on a long-term perspective. The current study is intended to compare the goals of reconstruction plans of quake-stricken municipalities from the environmental viewpoint. The reconstruction plan is positioned as a master plan not only to restore the infrastructure but also to design a new society. It has also an important role to share the vision, system, and

M. Ito is with the Faculty of Regional Policy, Takasaki City University of Economics, Takasaki, Gunma 3700801 Japan (e-mail: ts10043im@tcue.ac.jp).

A. Iijima is with the Faculty of Regional Policy, Takasaki City University of Economics, Takasaki, Gunma 3700801 Japan (corresponding author; phone: +81-27-343-5417; fax: +81-27-343-4830; e-mail: a-iijima@tcue.ac.jp).

procedure with citizens to achieve a goal. By using a text mining technique, Japanese language morphological analysis was performed. Briefly, a text mining technique enables quantitative analysis in terms of the syntax and the semantics by breaking a massive text dataset down into words or phrases [2], [3]. In this study, we discuss the difference of environmental policies described in the reconstruction plans of three quake-stricken municipalities. Moreover, we consider the future direction of eco-friendly reconstruction projects which will be implemented in the disaster area.

II. Method

A. Overview of the Cities for Analysis

To date, 43 municipalities in quake-stricken area have developed reconstruction plans. In this study, we focus on three reconstruction plans of Miyako City, Ishinomaki City, and Soma City as typical cases (see Fig. 1).
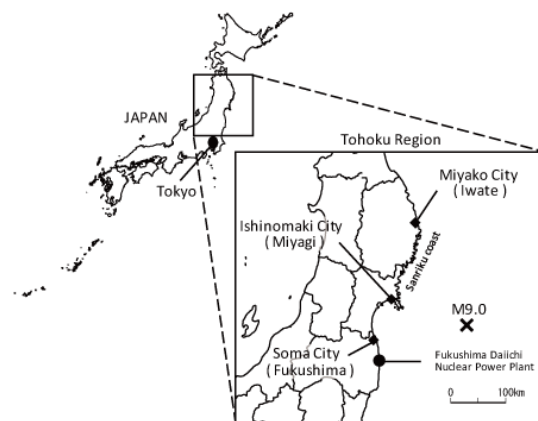

Fig. 1 Area affected by the Great East Japan Earthquake

Here, we briefly summarize the damage from the disaster in respective cities (data was cited from publicly available materials released by local governments). In Miyako City, 463 people are dead (data as of July 31, 2013). The tsunami inflicted heavy damage on harbor facilities and agricultural lands on the Sanriku coast. Many fishing ports and vessels were destroyed, and most coastal agricultural lands were submerged as a result of the tsunami, with deep repercussions for agriculture. In Ishinomaki City, 3,510 people are dead (data as of July 31, 2013). Marine product processing industry promoted by high fishery yield was the key industry in Ishinomaki City. However, the major fishing ports and factories were completely destroyed. In Soma City, 479 people are dead (data as of August 16, 2013).

Soma City is located in the northern area approximately 40 km from the Fukushima Daiichi Nuclear Power Plant. In addition to the damage from tsunami, citizens are now fairly worried about the effects of nuclear accident.

B. Text Mining Analysis of the Reconstruction Plans

Text mining analysis was performed by using open-source software named Tiny Text Miner (TTM) coupled with syntax analysis engines called MeCab and CaboCha [4]-[6]. This technique enables Japanese language morphological analysis for the large text dataset in simple way. Table I summarizes the basic statistics of reconstruction plans of respective cities.

TABLE I
BASIC STATISTICS OF RECONSTRUCTION PLANS

City	Number of sentences	Number of words
Miyako	832	43,713
Ishinomaki	1,567	58,273
Soma	800	30,299

Based on the results from morphological analysis, frequently-used nouns were simply sorted into "damage" and "reconstruction". To compare the difference in the goals of reconstruction plans, the nouns in "reconstruction" were further sorted into four main categories of "life", "disaster prevention", "economy", and "harmony with environment" as illustrated in Fig. 2. Each category was consisted with two sub-categories as shown in Fig. 2. Table II shows typical nouns sorted into respective categories. All the sentences in reconstruction plans were tagged to respective categories by using IBM SPSS Text Analytics for Surveys. Moreover, to reveal the attitude of "harmony with environment", cluster analysis and principle component analysis were carried out by using IBM SPSS Statistics.
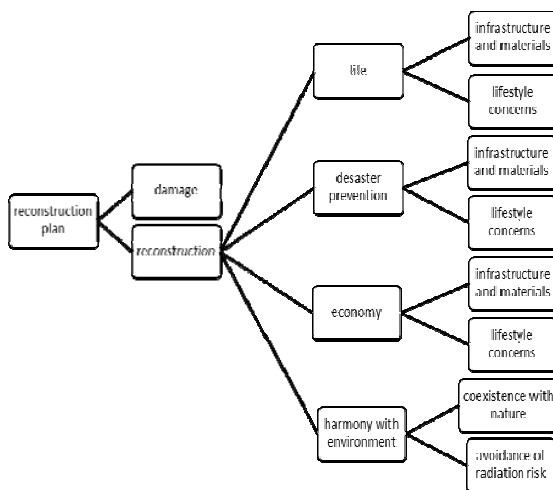

Fig. 2 Hierarchical framework for text mining analysis

TABLE II
CATEGORIZATION OF NOUNS FOR TEXT MINING ANALYSIS

Main category	Sub-category	Nouns (example)
life	infrastructure and materials	house, school, urban area, road, hospital
	lifestyle concerns	children, community, health, residence, medical care
disaster prevention	infrastructure and materials	escape route, evacuation site, breakwater, high ground
	lifestyle concerns	safety, evacuation, defense, disaster prevention measure
economy	infrastructure and materials	fishing port, farmland, fish market, shopping street
	lifestyle concerns	fishery, agriculture, company, employment, sightseeing
harmony with environment	coexistence with nature	forest, river, nature, biomass, renewable energy
	avoidance of radiation risk	radiation, measurement, decontamination

III. RESULT AND DISCUSSION

A. Comparison of Category Classification

According to the framework of the text mining analysis (see Fig. 2 and Table II), all the sentences in respective plans were tagged to the categories. Fig. 3 shows the frequency distribution in four main categories. Here, if one sentence has contained multiple nouns sorted into different categories, we redundantly have counted the frequency of respective categories.
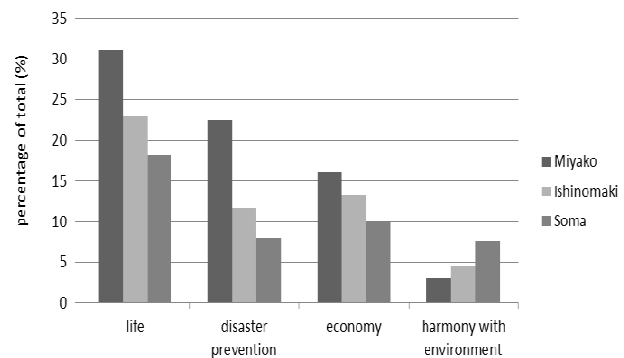

Fig. 3 Frequency distribution in four main categories

First, we have a quick view of the frequency in four main categories. Fig. 3 clearly shows that "life" is the most important issue in the reconstruction plans of three municipalities. The local governments may aim to improve the living environment as a top priority. From the viewpoint of infrastructure and materials, for instance, financial support for rebuilding destroyed homes has been emphasized in Miyako City. In terms of the lifestyle concerns, care of mental and physical health has been positioned as a key issue in Ishinomaki City. On the other hand, the frequency of sentences tagged to "harmony with environment" was the lowest among the four main categories. The local governments take a stance of placing importance on "harmony with environment" through the reconstruction process. However, it is not very high on governments' list of priorities.

Next, we compare the frequency distribution by municipality.

In Miyako City, the sentences tagged to "disaster prevention" were the second frequent topic, followed by "economy" and "harmony with environment". As mentioned above, Miyako City has suffered serious damage from tsunami. Therefore, no wonder the local government reinforces the efforts to prevent from disaster. On the other hand, in Ishinomaki City, the sentences tagged to "economy" were the second frequent topic. Economic revival will be an urgent issue in Ishinomaki City in which the main industry has been destroyed. In Soma City, the sentences tagged to "economy" were the second frequent topic, followed by "disaster prevention" and "harmony with environment". It is noteworthy that the frequency of "harmony with environment" was significantly high (7.6%) compared to the other two municipalities. As described previously, Soma city is now facing a nuclear disaster caused by the accident in the Fukushima Daiichi Nuclear Power Plant. Hence, avoidance of radiation risk will become a strong awareness in the reconstruction plan. The following section discusses the attitude of "harmony with environment" in more detail.

B. Cluster Analysis for the Category of "Harmony with Environment"

To elucidate the major attitude of "harmony with environment" in respective reconstruction plans, co-occurrence relation among the words was analyzed based on a cluster analysis. First, we extracted all the sentences tagged to "harmony with environment" from the reconstruction plan. Next, morphological analysis was performed by using TTM, and the dataset for further multivariable analysis was created with frequently-used (more than 3 times) nouns. Then, a cluster analysis was performed to the created dataset. The statistical distance between the variables was calculated as a Euclidean distance, and a Ward method was used for clustering. Here, we mainly discuss about the case of Soma City in which the high frequency of the sentences tagged to "harmony with environment" was observed. Fig. 4 illustrates the dendrogram of Soma City.

As shown in Fig. 4, variables (frequently-used nouns in the sentences tagged to "harmony with environment") are typically associated with nuclear accident. The dendrogram can be roughly divided into three clusters. Cluster 1 consists of the words indicating the vulnerable people who need special care from radiation exposure, such as "pregnant woman", "student", "glass badge", "radiation exposure", and "measures". Cluster 2 seems to consist of several sub-clusters. For instance, we can identify the cluster indicating the establishment of decontamination method and its implementation, and the cluster regarding the construction of renewable energy. Cluster 3 consists of the words corresponding to the determination of radiation dose and radioactive nuclides.
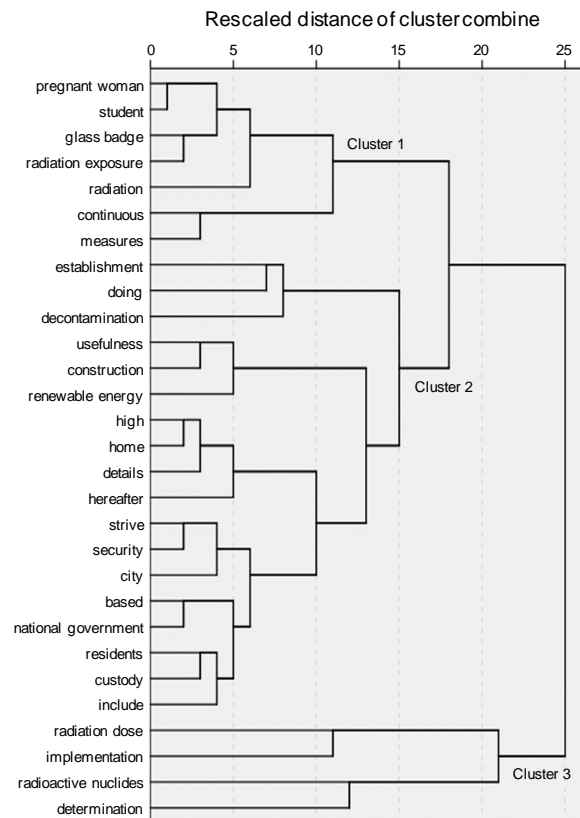

Fig. 4 Dendrogram of frequently-used words in the sentences tagged to "harmony with environment" in the reconstruction plan of Soma City

On the other hand, the compositions of frequently-used words and their co-occurrence relation in the other two plans from Miyako City and Ishinomaki City were completely different with that in Soma City. Since the effect of nuclear accident is not so serious in Miyako City and Ishinomaki City, there are few words relating to radiation risk. Although we did not show in the figures, a few words indicating natural riches, forest conservation, and constructing a renewable energy were found in the plan of Miyako City. The reconstruction plan of Ishinomaki City highlights the words relating to the development of eco-friendly society, the development of disaster-proof community, and the promotion of renewable energy industries. Consequently, text mining analysis can identify a clear diversity in the major attitudes of "harmony with environment" in respective reconstruction plans.

C. Principal Component Analysis for the Category of "Harmony with Environment"

To determine the dominant components of "harmony with environment" in respective reconstruction plans, principle component analysis was conducted to the same dataset applied to a cluster analysis. The varimax rotated matrix of the three dominant components was summarized in Table III.

TABLE III
VARIMAX ROTATED MATRIX OF PRINCIPLE COMPONENTS FOR SOMA CITY

Words	Components		
	1	2	3
hereafter	0.102	0.240	-0.026
radioactive nuclides	-0.159	-0.119	0.023
continuous	0.079	-0.139	-0.066
city	-0.050	-0.123	-0.052
based	-0.039	-0.053	0.911
glass badge	0.601	-0.024	-0.028
radiation exposure	0.816	0.013	-0.010
renewable energy	-0.106	-0.119	-0.103
radiation	0.581	-0.082	-0.135
strive	-0.160	-0.005	-0.009
measures	0.581	-0.063	-0.057
pregnant woman	0.933	-0.072	-0.022
details	-0.034	0.768	-0.035
establishment	-0.133	-0.230	-0.048
residents	-0.019	0.001	0.446
security	-0.080	-0.092	-0.085
custody	-0.074	-0.021	0.046
student	0.933	-0.072	-0.022
usefulness	-0.043	-0.092	-0.091
high	-0.040	0.898	-0.022
determination	0.379	0.108	-0.194
decontamination	-0.114	-0.160	0.587
construction	-0.029	-0.008	0.024
include	-0.086	-0.108	0.035
home	-0.047	0.725	-0.058
doing	0.258	0.268	0.072
radiation dose	-0.148	0.481	-0.117
national government	-0.051	-0.053	0.798
implementation	0.180	0.141	0.441

The 1st component (accounts for 12% of total variance) consists of the words in Cluster 1 such as "glass badge", "radiation exposure", "radiation", "pregnant woman", and "student" (see Fig. 4). It clearly demonstrates that the care of vulnerable people from radiation exposure has been emphasized as a top priority issue in Soma City. The 2nd component (accounts for 8.6% of total variance) consists of the words such as "details", "high", "home", and "radiation dose". This indicates that Soma City has placed the careful field survey of contaminated residential area as one of the top priorities of the reconstruction. The 3rd component (accounts for 8.0% of total variance) consists of the words such as "based", "decontamination", and "national government". Therefore, the reconstruction plan calls for the implementation of decontamination based on the initiative of national government. In terms of "harmony with environment", reconstruction plan of Soma City has been mainly designed for taking measures to the nuclear disaster.

On the other hand, the results of the other two cities were completely different with that in Soma City. Similar to the outcomes from a cluster analysis, the words regarding the coexistence with nature dominated the components of plans in Miyako City and Ishinomaki City.

In summary, we can find a large diversity in attitudes of "harmony with environment" by municipalities. The diversity will be due to the difference in situation after the earthquake. Because Soma City is affected by a nuclear disaster, no wonder the local government reinforces the efforts to reduce risks from radiation exposure. In order to reconstruct a new society, decontamination of radioactive nuclides should be continuously implemented for a while. In contrast, the issues contributing to coexistence with nature and/or constructing a renewable energy seem to have less attention as compared to the other two cities. Reconstruction of the disaster areas should not be preoccupied with decontamination projects. To place the reconstruction process as a piece of the framework for developing a sustainable society, we need to propose a revise of attitude of "harmony with environment" in appropriate phases.

IV. CONCLUSION

Based on the reconstruction plans, victims of the Great East Japan Earthquake are moving ahead powerfully toward a reconstruction of new society. To date, the reconstruction plans have mainly focused on the issues of "life", "disaster prevention", and "economy". However, projects regarding "harmony with environment" seem to be placed on a lower position of governments' list of priorities whereas their importance has been clarified in the reconstruction plans. The attitudes of coexistence with nature and introduction of renewable energy are commonly highlighted, but we need more effort to promote them. In case of the area suffering nuclear disaster, future-oriented goals have not been clearly described so far. The reconstruction plans should be revised in appropriate phases. Subsequent to a progress in the field of "life", "disaster prevention", and "economy", local governments should provide an incentive for creating a sustainable society through environmental policies.

REFERENCES

[1] A. Kamiyama, A. Iijima, "Current trends in eco-friendly reconstruction after the Great East Japan Earthquake," World Academy of Science, Engineering and Technology, vol. 72, pp. 1232–1235, 2012.
[2] G. Miner, J. Elder, A. Fast, T. Hill, R. Nisbet, D. Delen, "The histry of text mining," in Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications, Amsterdam: Elsevier, 2012, ch. 1.
[3] W. He, S. Zha, L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry." International Journal of Information Management, vol. 33, pp. 464–472, 2013.
[4] T. Kudo, Y. Matsumoto, "Japanese dependency structure analysis based on support vector machines," in Proc. 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora, New York, 2000, pp. 18–25.
[5] A. Miura, Y. Kawaura, "What is the knowledge-sharing community? A content analysis of posted Q&A and users' perspectives on the community," Japanese Journal of Social Psychology, vol. 25, pp. 153–160, 2009.
[6] S. Iwasaki, H. Nishikubo, "A Study on Tendencies of the Required Characteristics for New University Graduates by the Business Categories: Message Analysis by the Company's Web Site Recruitment Pages," The Journal of Communication Studies, 35, pp. 179–207, 2012.