

The Third Map of the Asia-Pacific Region: A New Approach to Cognitive Structure

Kazuya YAMAMOTO

Introduction

This article analyzes the manner in which people perceive a region, using the retrieval log of an Internet database, and illuminates the difference between the perceived region and the actual geography. In particular, this study uses the retrieval log data of the “Japanese Politics and International Relations” database, which is a collection of various documents regarding Japanese politics and diplomacy. This database also includes numerous important documents pertaining to countries around the world, with focus on East Asian countries and their regional organizations.⁽¹⁾ I apply a data mining algorithm to the retrieval log and extract a cognitive map of the document readers. In doing so, I detect significant differences between the cognitive and actual maps and derive important implications for international relations.

This paper proceeds in four parts. First, I conduct a short literature review and argue the distinctiveness of this study compared to the existing researches. In the second section, I provide an overview of the database from which the retrieval log was collected. Then, this section describes the statistical characteristics of the log data. Third, I apply a data mining algorithm to the log and derive the implications for actual international relations from the analysis. Finally, the concluding section refers to several notes for further investigation.

New Approach to the Cognitive Structure

To iterate, this study uses Internet retrieval log data to delineate the cognitive map of people who search for documents, particularly those pertaining to politics and diplomacy in East Asian countries. The document-based objective and systematic identification for the cognitive structure of people was traditionally conducted using the documents *per se*. A series of content analyses pertaining to political/diplomatic documents and newspapers in the East Asian region have been conducted from the middle of the twentieth century (e.g., Okabe, 1971; Watanabe, 1974; Inoguchi, 1972, 1981; Hyun, 2006: ch.2).⁽²⁾ These studies have clarified many aspects of international relations in the region, for example, illuminating the way in which politicians and diplomats perceived the disputes between themselves and conceived the relationships of amity and enmity in the region, or elucidating the manner in which the mass media describes these conflicts and relationships. In brief, the traditional content analysis has attempted to explain the writers' (speakers') side of the cognition of reality.

In contrast to these researches, this study focuses on another side of the perception, that is, that of the readers or analysts of the documents. The Internet-based retrieval system enables us to gather the users' retrieval words. The collection of these words allows us to depict the aggregate map of their perceptions. Although, hitherto, surveys of the general public have been generally limited to opinion polls, recent technological development has made it possible to conduct this new type of research.⁽³⁾

Since international affairs are first managed by politicians, diplomats, and other members of governments, there is no doubt that the cognition of the politicians and diplomats is of prime importance. However, it is equally evident that the cognition of the general public has a strong

impact on politics. Therefore, this side of the perception is as significant as that of administrative officials for understanding political and international issues in the contemporary world. The research on the readers' side of cognition will pave the way for the development of content analysis and form the third map of the world, following the first (geographical) and second (writers'/speakers') map.

Data

This section outlines the composition of the database itself and summarizes the statistics of the log gathered through the retrieval system implemented in the database.

As mentioned earlier, the database used for this study is called "Japanese Politics and International Relations." As of September 15, 2007, this database comprised 3,580 texts. It contains documents such as treaties, agreements, charters, addresses and speeches by Japan's ministers, and joint statements and exchanged notes between countries. Regarding the documents that are related to more than two countries, many of the texts involve Japan as a party; however, some of the texts are documents between countries other than Japan (e.g., between ASEAN countries or between China and South Korea). In terms of the focus area, many of the texts are those between Japan and the Asia-Pacific countries (including the US), although the documents between Japan and the countries in regions such as Europe and Africa, as well as the important documents that do not involve Japan (e.g., Warsaw Pact), are also included. Based on the composition of the database, we expect that many of the database users might be interested in Asia-Pacific regional relations and the cognitive map of this region would be derived from the retrieval log.

The log was collected from March 12 to September 15, 2007. During this period, a total of 1,271 retrievals were conducted. The retrieval system

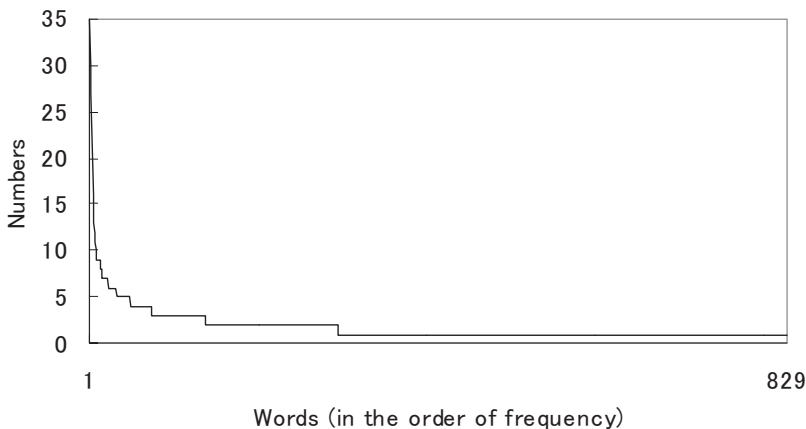
enables the users to search for documents/speeches by indicating some of the words that are expected to be included in the texts and/or by specifying the date at which they were published. After excluding records thrown up due to incorrect searches, system and network errors, and searches conducted by specifying date/period only, a total of 1,204 searches were obtained as the sample data, in which at least one word was entered as a keyword. Table I describes the most frequently retrieved words in the log. As expected, many high ranking keywords are related to Japanese diplomacy toward the East Asian countries and the US.

Table I. Rank of Appeared Words

<i>Words</i>	<i>Number</i>
diplomacy (gaikou)	35
Japan-South Korea (Nikkan)	30
Japan (Nihon)	27
child (jidou)	16
independent Foreign Policy (jishu gaikou)	13
independence (jishu)	12
independency (jishu-sei)	11
Plaza Accord (Puraza Goui)	10
Southeast Asia (Tounan Ajia)	9
Nobusuke Kishi	9
South Korea (Kankoku)	9
the US-Japan Mutual Defense Assistance Agreement (MSA Kyotei)	9
Shigeru Yoshida	8
USA (Amerika)	8

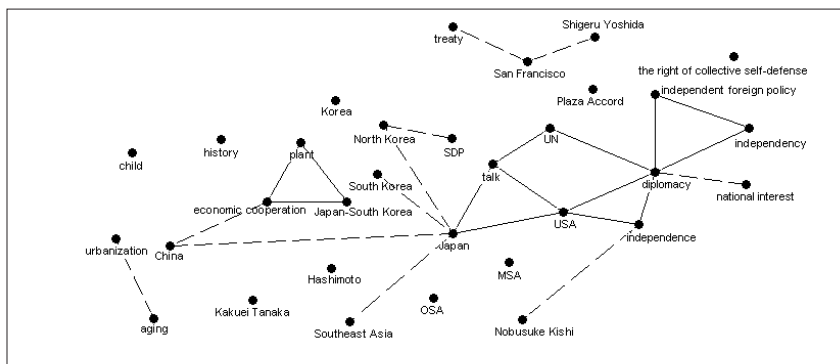
Figure 1 depicts the frequency distribution of all words that appeared in the log. As indicated by the steep slope on the left-hand side of this figure, there is a cluster of very frequently used words. On the other hand, 829 words were used for all the 1,204 retrievals. Therefore, the result suggests that the document readers have very diverse interests.

Figure 1. All Appeared Words vs. Numbers



30 most frequent words (circles) and the 25 highest co-occurrences between these words (lines).⁽⁵⁾ This figure indicates the *relationships* between words. Therefore, the location of each word has no implications. The simply connected links between two words are indicated with dashed lines.⁽⁶⁾ On the other hand, solid lines are drawn between words that can access each other through other paths, even if their direct link is cut off. These connected subgraphs with solid lines imply stronger and more complicated connections between words. They are referred to as *bases* or *islands* in the KeyGraph.⁽⁷⁾

Figure 2. Islands and their Relationships



Note: See Appendix C for each word.

Figure 2 depicts two significant islands (i.e., subgraphs connected with solid lines).⁽⁸⁾ The first is the main island that comprises “Japan,” “the US,” “diplomacy,” and others. The second is a smaller island that includes “Japan-South Korea,” “plant,” and “economic cooperation.” This indicates that (a) many of the document readers are interested in relations between Japan-US and Japan-South Korea and, among these, (b) the readers are particularly interested in the general characteristics of Japanese diplomacy toward the US after World War II. This result may not be unexpected because many parts of the database have associations with

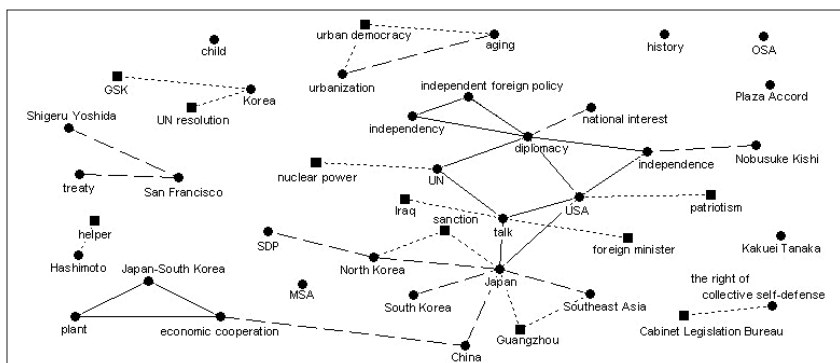
the US. However, considering the large number of documents in the database regarding other Asian countries such as China and Southeast Asian countries; the China-South Korea relationship; and multilateral relations such as ASEAN, APEC, and the G8, the focus on the nature of the post World War II Japanese diplomacy toward the US appears to be excessive.

Second, countries and regions in East Asia are connected to Japan only through the direct links (i.e., links connected by the dashed lines). This is interesting because, although these countries are often retrieved simultaneously with Japan, the result indicates that they appear bilaterally with Japan and do not co-occur with other Asian countries and regions. This result seems to reflect the standard image of the East Asian region that is in contrast with Europe—the Asian region comprises patches of separate bilateral relations between countries.

In sum, in the picture in Figure 2, the US occupies a large portion of the map and links with Japan, while Japan is bilaterally connected to other East Asian countries/regions. In other words, this map appears to be very similar to the Asia-Pacific region that is described as a porous system, strongly influenced by the US *imperium* (Katzenstein, 2005) or the hub-spoke system (Yamamoto, 2007).⁽⁹⁾ The analysis implies that the cognitive map of the Asia-Pacific region retained by the document readers is drawn in a similar manner.

Third, I investigate whether or not there are the words appearing with low frequency but having a strong connection with the islands and functioning as bridges between them. This analysis enables us to elucidate the significant relationships between the islands that would remain hidden if we focus only on the words with high frequency. Figure 3 indicates this result.⁽¹⁰⁾ The squares in Figure 3 represent the nodes that do not rank among the 30 highest words but have high coincidence with the islands.⁽¹¹⁾ Although many findings can be indicated, two relationships in the figure are particularly interesting.

Figure 3. Islands and Low Frequency Words



Note: See Appendix C for each word.

The first is the relationship between “sanction,” “Japan,” and “North Korea.”⁽¹²⁾ This result indicates that although sanction appears with a lower frequency, it co-occurs with Japan or North Korea with a high frequency, each time it appears. This triangle between Japan, North Korea, and sanction suggests that among many past and present arguments over the international sanctions against countries such as Afghanistan, Iraq, Iran, China, Myanmar, and so forth, the sanctions against North Korea are a matter of great concern for the document readers. The second is the relationship that connects “Iraq” and “patriotism” to “talk” and “USA,” respectively, in the main island. As argued earlier, the main island reflects the dominant concern about Japan’s relationship with the US in the post World War II era. However, Figure 3 suggests that the document readers’ interests are not confined to such long-term historical matters, and contemporary issues such as the Iraq war also appear as a concern.

Conclusion

This article offers a cognitive map of the document readers, particularly

regarding international relations in the Asia-Pacific region, using the retrieval log of an Internet-based database. The analysis clarifies that Japan's diplomatic relationship with the US after World War II constitutes the core interest of many document readers. On the other hand, the analysis also indicates that these readers have a bilateral perception of Japan's relationships with East Asian countries. These results suggest that the image of the US centered hub-spoke system in the Asia-Pacific region remains dominant for the document readers. While the possibility of forming an East Asian community is broadly discussed in current political debates, the perception of the East Asian region illustrated by this study is also indicative of this discourse. If an East Asian community expands in the future, the map depicted by this study would also change from the simple connections between two countries to the connected subgraph that is constituted by many countries and regional organizations. Therefore, such a research would serve as an indicator as well as other resources for assessing the degree to which the region is integrated. Moreover, by highlighting the less frequent words, this study clarifies that the readers are not only interested in historical questions but are also concerned about the topics in contemporary international relations.

Finally, I refer to two limitations in this study. First, as mentioned in the first section, the random sampling method poses a problem. Unlike controlled opinion polls, this type of log data is not collected randomly. Therefore, the result of the analysis should not be considered to strictly reflect the images of the general public. Even so, the results obtained in this study can be interpreted as an image of the Asia-Pacific region and would benefit our understanding of international relations in this region. Second, the log was collected for a period of approximately six months. This period might be rather short. Due to this, the events that occurred in the period could distort the map. A longer period for gathering the log would decrease this sensitivity toward specific events. A reexamination is

expected to be conducted in the future.

Appendix A. Data Format

One line of the data (log file) comprises one search. Since the retrieval system can accept a maximum of five words, the data is formatted as follows:

$$D = \left| \begin{array}{ccccc} w_{11} & w_{12} & w_{13} & \phi & \phi \\ w_{21} & \phi & \phi & \phi & w_{25} \\ & & \vdots & & \\ w_{1204\ 1} & w_{1204\ 2} & \phi & \phi & w_{1204\ 5} \end{array} \right|$$

where D : dataset; w : word; ϕ : no entry.

As mentioned in the main text, the incorrect searches, system errors, and the searches that contained no words (i.e., the searches for which only the dates/periods were entered) are excluded at this stage (or modified, rather than excluded, if that is appropriate to the study).

Appendix B. Data Mining Algorithm

This section formally describes the algorithm employed in this study. The KeyGraph algorithm was originally developed by Ohsawa et al. (1999), Ohsawa (2003: ch.9), and Ohsawa (2006). There are variants of this algorithm and its analytical options. Therefore, this appendix outlines the basic logic that is largely shared by the variants of the KeyGraph algorithm, based on Ohsawa et al. (1999), Kita et al. (2002), and Ohsawa (2003: ch.9, 2006: ch.4). In addition, the author has slightly modified the description used in these papers for the purpose of adjusting them to this research. Among these variants, the result of this analysis is obtained by

The Third Map of the Asia-Pacific Region: A New Approach to Cognitive Structure using Polaris version 0.19 α . For further details regarding Polaris and its KeyGraph algorithm, refer to Okazaki and Ohsawa (2003) and <http://www.chokkan.org/software/polaris/>.

1. Extraction of Islands

- a. Extract the specific number (e.g., 10 or 30) of words that appear most frequently.
- b. Calculate the co-occurrence frequency $c_0(w_i, w_j)$ between each pair of the words extracted above (where $i \neq j$). This study employs the Jaccard index (coefficient) for calculating $c_0(w_i, w_j)$. This index is one of the most popular indices to measure the similarity between two items (in this study, co-occurrence). This is defined as follows:

$$J(w_i, w_j) = N(w_i \cap w_j) / N(w_i \cup w_j) \quad (A1)$$

where

$N(w_i \cap w_j)$: The number of lines in D that contain both w_i and w_j .

$N(w_i \cup w_j)$: The number of lines in D that contain either w_i or w_j , ($i \neq j$).

- c. Select the highest specific number of $c_0(w_i, w_j)$ value pairs. Thereafter, link w_i and w_j with a line and create graph G .
- d. Cut off all simply connected links in G and retain the connected subgraphs. Refer to these subgraphs (and each word that is not connected with other words and stands alone) as island g ($\forall g \in G$): These subgraphs comprise circles and solid lines in the figures. (Note: In this study, the simply connected links between words that were extracted at step a are not cut off and delineated with dashed lines.)

2. Extraction of Significant Words with a Low Frequency (and Keywords)

a. Calculate the co-occurrence frequency $c_1(w_i, w_g)$ between each word w_i in D ($\forall w_i \in D$) and each island g , where all words in g are replaced with w_g based on the following formula:

$$w_g = \begin{cases} \forall w \in g & \text{if } w_i \notin g \\ \forall w^{-i} \in g & \text{if } w_i \in g \text{ (} w^{-i} \text{ indicates that } w_i \text{ is excluded).} \end{cases}$$

$c_1(w_i, w_g)$ is defined as the Jaccard index in this study.

b. Select the highest specific number of $c_1(w_i, w_g)$ value pairs. Link w_i to g with dotted lines if they have not been connected with solid lines (in this study, with lines). The squares in Figure 3 represent the nodes that are not included in the most frequent words selected in section 1 in this appendix but have high co-occurrence frequencies with the islands.

c. Calculate the sum of the co-occurrence frequency between words w_i and each island g_k , $c_1(w_i, g_k)$. This sum $key(w_i)$ is not normal summation and is defined as follows:

$$key(w_i) = 1 - \prod_k (1 - c_1(w_i, g_k) / f_0).$$

f_0 is a standardization coefficient. It is used if nonstandardized $c_1(w_i, w_g)$ is employed in analyses.

d. Select the highest specific number of $key(w_i)$ value. The words w_i that are selected in this process are termed as the “keywords.”

(Note: Although the extraction of keywords constitutes the core part of the KeyGraph, this study does not utilize this process (steps c and d) because no interesting finding has been detected.)

Appendix C. List of Words in Figures 2 and 3

Table A1. The List of Words

<i>Figures 2 and 3</i>	<i>Retrieval words</i>
aging	koureika
Cabinet Legislation Bureau	naikaku housei-kyoku
China	Chugoku
economic cooperation	keizai kyoryoku
foreign minister	gaisho
GSK	Daikanminkoku seifu (the government of South Korea)
Guangzhou	Koushu
helper	herupa
history	rekishi
Korea	Chosen
national interest	kokueki
North Korea	Kita-Chosen
nuclear power	genshiryoku
OSA	Okinawa mitsuyaku (Okinawa Secret Agreement)
patriotism	aikokushin
plant	puranto
sanction	seisai
SDP	Nihon Shakai-tou (Social Democratic Party of Japan)
talk	kaidan
the right of collective self-defense	shudanteki jieiken
treaty	joyaku
UN	kokuren
UN resolution	Kokusai rengou ketsugi
urban democracy	toshi demokurashi
urbanization	toshika

Note: The words listed in Table I are excluded.

1 The database is available at <http://www.ioc.u-tokyo.ac.jp/~worldjpn/documents/index.html>. Although many documents in the database are

available in English and/or Chinese, some documents are available only in Japanese due to the nature of the documents (e.g., the Japanese prime minister's addresses in the diet).

- 2 Since there are too many researches and textbooks concerning content analysis in general, it is impossible to describe the entire literature pertaining to this field in this article. Among the most important textbooks, see Krippendorff (2004 [1980]) and Holsti (1969).
- 3 The problem of random sampling will be discussed in the concluding section.
- 4 See Appendix B for more details regarding the KeyGraph algorithm. In order to conduct the analysis, this study uses Polaris, which implements the KeyGraph algorithm and visualizes the results into figures. As described in Appendix B, this algorithm has variants, and each differ slightly. The result of this research is obtained from Polaris version 0.19 α .
- 5 The number of words in the figure is larger than 30 due to the same rank. The co-occurrence frequency is calculated using the Jaccard index. See Appendix B for the definition of this index. The KeyGraph analysis was conducted using original words (many of the retrieval words were in Japanese). Consequently, all the words that appeared in the figures were in Japanese. Therefore, for the purpose of this paper, the author translated the words in the figures into English after the analysis. In addition to the change regarding the language, the figure is slightly modified from the original Polaris output due to technical reasons.
- 6 The “simply connected link” means that two circles can be reached only through the direct link between them.
- 7 Succinctly, a circle that is not linked to other circles with solid lines, and which stands alone (in this study, each circle that is connected with dashed lines only), is also regarded as an island that is composed of one node in the KeyGraph.
- 8 Although I changed the values of the most frequently occurring words and those with the highest co-occurrence, no distinctive or interesting results could be observed; two islands appear when the value of the highest co-occurrence is set to approximately more than 25, while no interesting results could be found changing the values of the most frequent words.

The Third Map of the Asia-Pacific Region: A New Approach to Cognitive Structure

- 9 For other recent arguments pertaining to the Asia-Pacific region, see Pempel (2005), Acharya and Goh (2007), and Acharya (2001).
- 10 The parameters are set to 30 (circle node), 30 (link between circle node), and 10 (square node).
- 11 See Appendix B for the calculation.
- 12 North Korea constitutes an isolated island. As mentioned in footnote 7, each circle unconnected by solid lines is also considered to be one island in the KeyGraph, although it may be likely that the connected subgraphs contain more significant implications for interpreting the data.

References

- Acharya, Amitav, 2001. *Constructing a Security Community in Southeast Asia: ASEAN and the Problem of Regional Order*, London & New York: Routledge.
- Acharya, Amitav and Evelyn Goh, eds., 2007. *Reassessing Security Cooperation in the Asia-Pacific: Competition, Congruence, and Transformation*. Cambridge, MA: The MIT Press.
- Holsti, Ole R., 1969. *Content Analysis for the Social Sciences and Humanities*. Reading, Mass.; London: Addison-Wesley.
- Hyun, Daesong, 2006. *Ryodo Nashonarizumu no Tanjo: "Dokuto/Takeshima Mondai" no Seijigaku* [The Birth of Territorial Nationalism: the Politics of Dokuto/Takeshima Islands Territorial Dispute]. Kyoto: Minerva Publishing.
- Inoguchi, Takashi, 1972. 'Measuring Friendships and Hostility among Communists Powers: Some Unobtrusive Measures of Esoteric Communications', *Social Science Research* 1 (1): 79-105.
- Inoguchi, Takashi, 1981. 'Etsu-Chu-So Kankei, 1964-1980: Tansaku-teki Bunseki' [The Relationships between Vietnam, China, and Soviet, 1964-1980], *Toyo Bunka Kenkyujo Kiyō* [The Memoirs of Institute of Oriental Culture] 87: 197-224.
- Katzenstein, Peter J., 2005. *A World of Regions: Asia and Europe in the American Imperium*. Ithaca: Cornell University Press.
- Kita, Kenji, Kazuhiko Tsuda, and Masami Shishibori, 2002. *Joho Kensaku Arugorizumu* [Information Retrieval Algorithms]. Tokyo: Kyoritsu Publishing.
- Krippendorff, Klaus, 2004[1980]. *Content Analysis: an Introduction to Its*

- Methodology*, 2nd ed. Thousand Oaks, CA: Sage.
- Ohsawa, Yukio, ed., 2003. *Chansu Hakken no Joho Gijutsu* [The Information Technology of Chance Discovery]. Tokyo: Tokyo Denki University Press.
- Ohsawa, Yukio, 2006. *Chansu Hakken no Deita Bunseki* [The Data Analysis of Chance Discovery]. Tokyo: Tokyo Denki University Press.
- Ohsawa, Yukio, Nels E. Benson, and Masahiko Yachida, 1999. 'KeyGraph: Go no Kyoki Gurafu no Bunkatsu Tougou ni yoru Kiiwado Chushutsu' [KeyGraph: Automatic Indexing by Segmenting and Unifying Co-occurrence Graphs], *Denshi Joho Tsushin Gakkai Ronbun-shi* [The transactions of the Institute of Electronics, Information and Communication Engineers] J82-D-I (2): 391-400.
- Okabe, Tatumi, 1971. *Gendai Chugoku no Taigai Seisaku* [The Foreign Policy of Contemporary China]. Tokyo: University of Tokyo Press.
- Okazaki, Naoaki and Yukio Ohsawa. 2003. 'Polaris: An integrated data miner for chance discovery', in Workshop of Chance Discovery and Its Management (in conjunction with International Human Computer Interaction Conference (HCI2003)): 27-30.
- Pempel, T.J., ed., 2005. *Remapping East Asia : the construction of a region*. Ithaca: Cornell University Press.
- Watanabe, Akio 1974. 'Taigai Ishiki ni okeru "Senzen" to "Sengo": Syhushou/Gaishou no Gikai Enzetsu no Bunseki ni motozuku Jakkan no Kousatu' [The Images of International System before and after WWII: A Study Based on the Analysis of the addresses of the Prime and Foreign Ministers], in Seizaburo Sato and Roger Dingman, eds., *Kindai Nihon no Taigai Taido* [The Attitude toward International System of Japan in the Modern Times]. Tokyo: University of Tokyo Press: 225-274.
- Yamamoto, Yoshinobu, 2007. 'Chiiki Tougou Riron to "Higashi Ajia Kyoudou-tai"' [Regional Integration Theory and an East Asian Community], in Takehiko Yamamoto and Satoshi Amako, eds., *Aratana Chiiki Keisei: Higashi Ajia Kyoudou-tai no Kouchiku 1* [New Formation of Region: Designing as East Asian Community 1]. Tokyo: Iwanami Shoten.: 315-334