*Ecological Modelling*, 32 (1986) 29–41
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands



# ANALYSIS OF COGNITIVE STRUCTURES OF ENVIRONMENT OF LOCAL RESIDENTS THROUGH WORD ASSOCIATION METHODS

KO OI [1], SADAAKI MIYAMOTO [2], OSAMU ABE [3], ATSUO KATSUYA [4] and KAZUHIKO NAKAYAMA [3]

[1] *Environmental Information Division, National Institute for Environmental Studies, Tsukuba, Ibaraki 305 (Japan)*
[2] *Institute of Electronics and Information Sciences, University of Tsukuba, Sakura, Niihari, Ibaraki 305 (Japan)*
[3] *Science Information Processing Center, University of Tsukuba, Sakura, Niihari, Ibaraki 305 (Japan)*
[4] *Kyoto Sangyo University, Motoyama, Kamigamo, Kita-ku, Kyoto 603 (Japan)*

ABSTRACT

Oi, K., Miyamoto, S., Abe, O., Katsuya, A. and Nakayama, K., 1986. Analysis of cognitive structures of environment of local residents through word association methods. *Ecol. Modelling*, 32: 29–41.

The way through which people conceive images of their living conditions is examined in this paper in order to develop an evaluation method of the environment. Ordinary residents seem to recognize their living condition in its entirety. Therefore, the usual survey questionnaire, where a respondent finds questions on individual items, would not give satisfactory information of his/her cognitive structure of the environment.

To find the inseparable cognitive structure as a whole, the authors employ the association method. In the authors' questionnaire, a free association test asking subjects what they associate with a concept "ease of living" is followed by a controlled association test which deals with some 30 stimulus words and possibly associated words. These words were chosen to cover almost all environmental aspects including amenities, human relations in the neighborhood, culture, and administrative services. Surveys were conducted in a district in downtown Tokyo and in a small city near the metropolis. Researchers of environmental studies and pollution control also responded to the survey.

Data of the frequency of association between the words in the controlled association test was scrutinized through cluster analysis, and on the other hand the words involved were displayed into a directed graph due to the probability of association between them. By integrating these results into a diagram, the cognitive structures of environment are uncovered. Replies to the free association test are dealt with a two-way cluster analysis method to reveal the relation between groups of the associated words and groups of the subjects. This analysis makes it easy to know what aspect of the environment is of prime interest to a particular group of residents.

0304-3800/86/$03.50 

## INTRODUCTION

The knowledge of the way through which people in a community conceive images of their living condition is important in developing an evaluation method for the environment and therefore in examining the desirable characteristics to be given to an environmental indicator.

Ordinary residents seem to perceive their living condition in its entirety. Therefore, the usual survey questionnaire, where a respondent finds questions on items concerning living conditions one by one, would not give satisfactory information of his/her cognitive structure of the environment. Moreover, we have to know the variety and the extent of residents' concerns relating to their living conditions.

To find the inseparable cognitive structure as a whole and describe it in a model, the authors employ the association method in psychology. In the authors' questionnaire, a free association test asking subjects to report what they associate with the concepts "ease of living" and "happiness in living" is followed by a controlled association test which deals with some 30 stimulus words and possibly associated words. These words were chosen to cover almost all environmental aspects. The survey was carried out in two urban districts. Researchers on environmental studies and pollution control also responded to the survey.

Replies to the free association test are dealt with by two-way cluster analysis to reveal the relation between groups of the associated words and groups of subjects. This analysis makes it easy to know what aspect of the environment is of prime interest to a particular group of residents.

Data on the frequency of association between the words in the controlled association test are scrutinized through cluster analysis, and on the other hand, a graph of the words involved is developed on the basis of the direction and the probability of association between them. By integrating these results into a diagram, the cognitive structure of the environment is uncovered, and described in a model.

## FREE ASSOCIATION TEST

To know the range and the variety of subjects' interests concerning living conditions, the following free association method is employed in this study. In the questionnaire a subject is urged to write down what they associate with "ease of living" and "happiness in living" in as many words, phrases or sentences as possible. The pair of stimulus phrases "ease of living" and "difficulty of living" was not employed, because they are not only an antonym of each other, but also have quite different kinds of implications, the pair would cause psychological annoyance to a respondent.

Associated phrases and sentences are at first decomposed into words. From the resulting words, those considered meaningless as associated ones, e.g., by the analogy in English, prepositions and articles, are eliminated after the decomposition. Then a merger of the remaining words and those given as words from the beginning by a respondent makes the sequence of associated words used by a subject.

To carry out cluster analysis, the similarities between subjects, as well as between the associated words, are defined. Consider a word $w_i$ which is found at least $s$ times in the sequences of associated words by all the subjects, and designate the set of such words by:

$$W = \{w_1, \ldots, w_n\}$$

Now specify the vector of association of a subject by:

$$A = (a_1, \ldots, a_n)$$

where $a_i$ is the frequency of association of the $i$th word $w_i$ of the set $W$ by the subject. We consider that two respondents who share more common associated words with one another than other pairs are more similar. Suppose $A$ is the vector of association by a subject $u$ and that the vector of association by a subject $v$ is:

$$B = (b_1, \ldots, b_n)$$

The similarity between the two subjects $u$ and $v$ (Miyamoto et al., 1983) is:

$$r_{uv} = \sum_{k=1}^{n} \min(a_k, b_k) / \sum_{k=1}^{n} \max(a_k, b_k)$$

On the other hand, the number of respondents whose association vector is not zero is assumed to be $m$. As for the set of these respondents denoted by:

$$R = \{r_1, \ldots, r_m\}$$

the vector of respondents of an associated word $u'$ in $W$ is defined by:

$$A' = (a'_1, \ldots, a'_m)$$

where $a'_i$ is the number of appearances of the word $u'$ in the sequence of associated words by the $i$th respondent.

The similarity between the word $u'$ and the word $v'$ and the vector of the respondents:

$$B' = (b'_1, \ldots, b'_m)$$

is defined by:

$$r_{u'v'} = \sum_{k=1}^{m} \min(a'_k, b'_k) / \sum_{k=1}^{m} \max(a'_k, b'_k)$$

That is, two words associated by more common respondents are more similar.

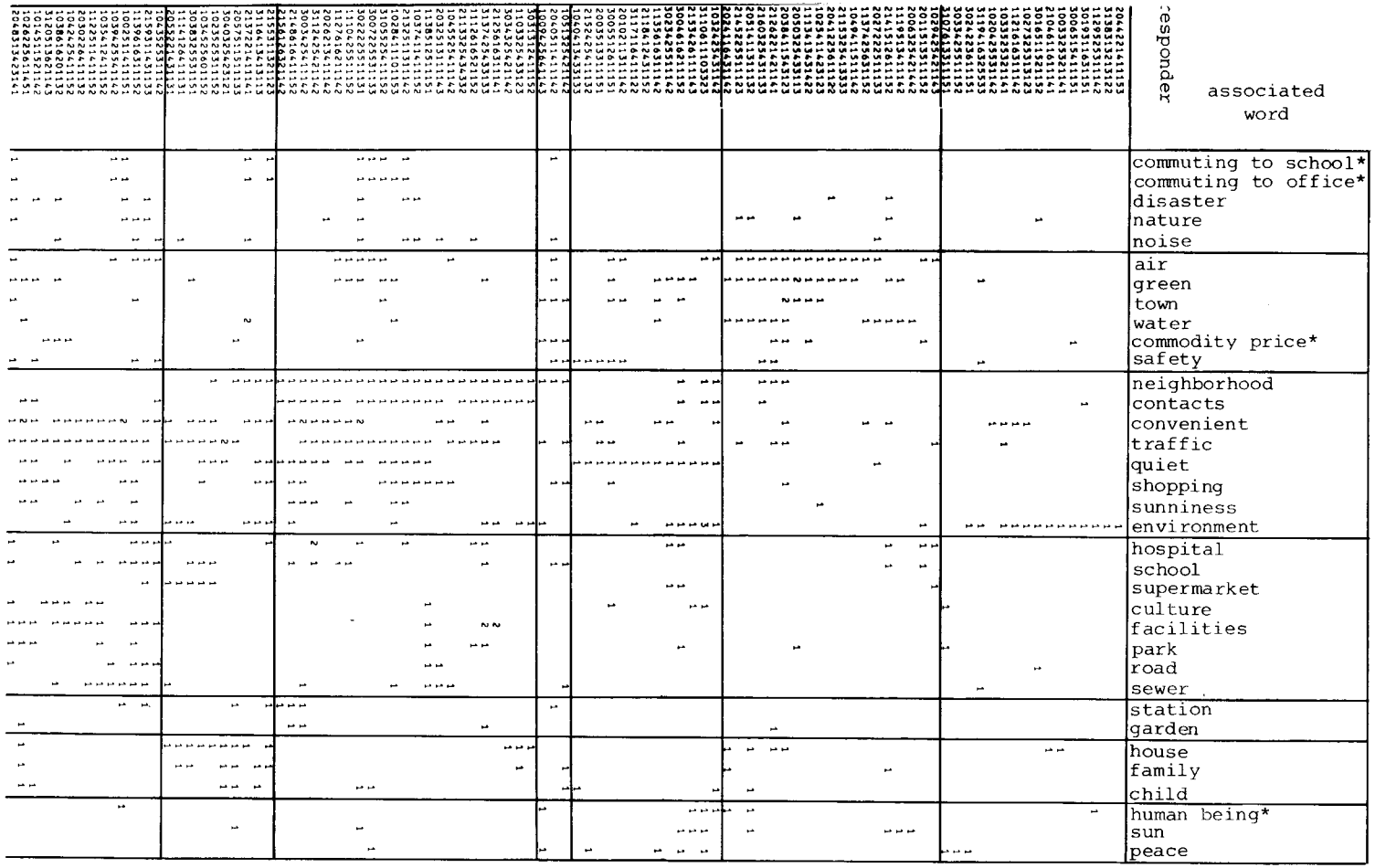


Fig. 1. Two-way cluster analysis for the free association test in ISHIOKA.

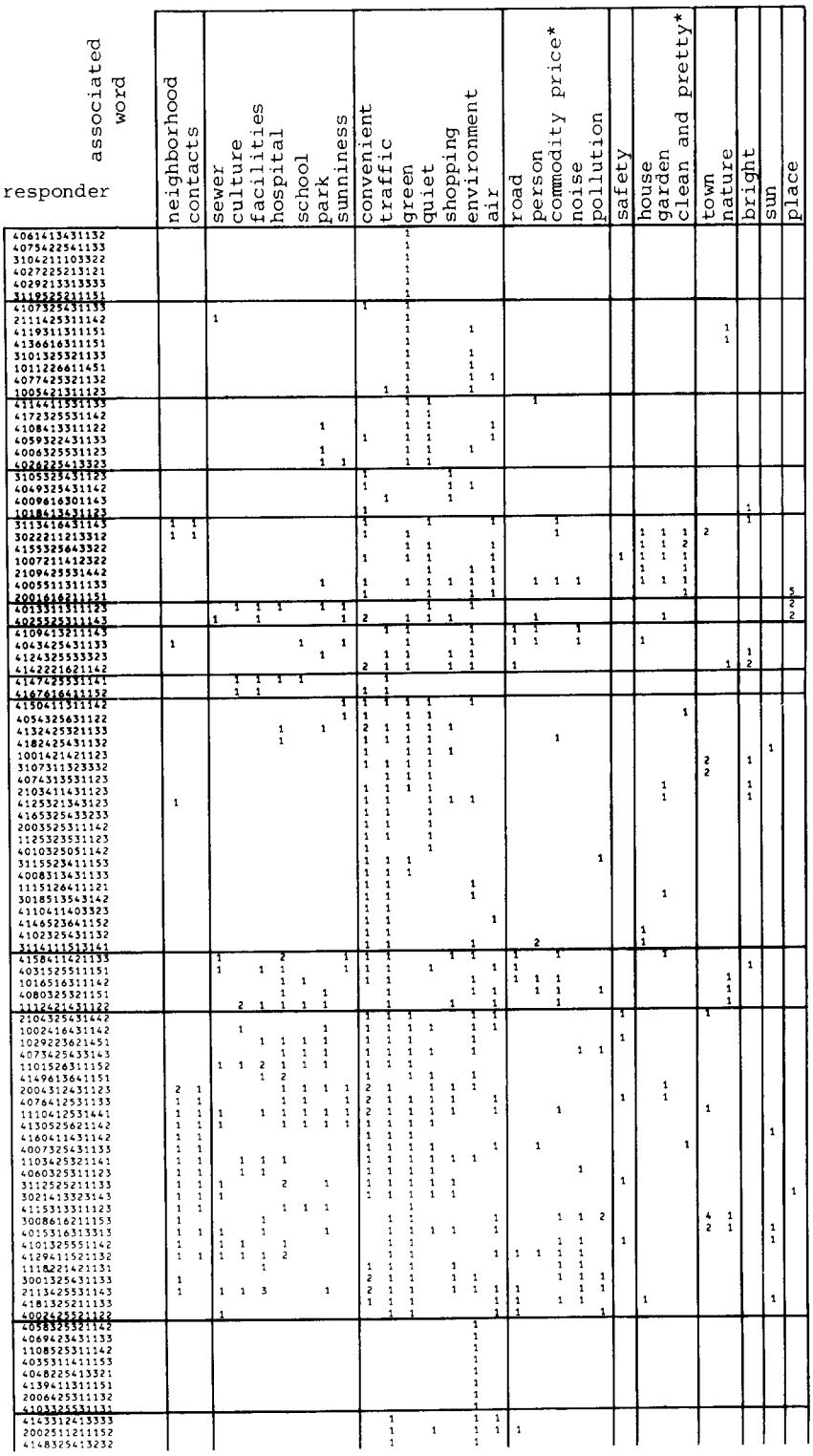


Fig. 2. Two-way cluster analysis for the free association test in HANAHATA.

With these similarities, cluster analysis is employed to divide the associated words in $W$ as well as the respondents in $R$ into groups. Then the frequency of association of some word by some subject is displayed in a matrix whose column and row correspond to an associated word and a respondent, respectively. The order of each of the words and responders in the matrix is determined to be identical with its order in the dendrogram (Figs. 1 and 2).

## CONTROLLED ASSOCIATION TEST

To examine the structure of cognition of the living environment, or, from the wider viewpoint, living circumstances, and to build its model, a controlled association test is employed. That is, in the test several stimulus words are shown to a subject one by one, and, in a given word list, the respondent marks words which are associated with the stimulus word. The order of association of words is not taken into consideration. The free association test with many successive stimulus words could impose a heavy psychological burden on a subject, so the controlled test is employed.

In the authors' method, $N$ words concerning living conditions are chosen. In the questionnaire one after the other of them is given as a stimulus word, and the remaining $N - 1$ words are shown as possible associated words.

The $N$ words were determined through a study on: (a) a list of items about the safety, health and amenities of a resident's life (S. Nishioka, National Institute for Environmental studies, personal communication, 1983); (b) discussions on living circumstances by the authors; and (c) results of a preliminary free association test as mentioned in the previous section. The words were selected to cover almost all the aspects of living except family, house, and private garden. Too large an $N$ is not acceptable to the subjects. In a pretest carried out after the preliminary one in an urban residential area near Tsukuba, the number $N$ was chosen as 30. In the 65% of the questionnaires recovered, no respondent was found who gave up answering amid the sequence of questions of this test. Therefore, it was decided that the same word list should be used in the main survey.

For analysis of the results of this test, a method presented by Miyamoto and Nakayama (1980) for research on a representation of citation relationships between scientific journals is used.

The bases of the method are two concepts: the relatedness between words and the probability of association from a stimulus word to an associated word.

*Relatedness between words*: Consider the set of the $N$ words chosen as stimulus words and possible associated words:

$$W = \{ w_1, \ldots, w_N \}$$

The frequency with which word $w_j$ is associated with $w_i$ by all the subjects is designated by $f_{ij}$. The sum of these frequencies over all the associated words for a stimulus word $w_i$ is denoted by $f_i$. That is:

$$f_i = \sum_{k=1}^{N} f_{ij}$$

The relatedness $r_{ij}$ between the words $w_i$ and $w_j$ is defined by:

$$r_{ij} = (f_{ij} + f_{ji})/(f_i + f_j)$$

The relatedness between two words means how easily each of them is associated with its counterpart.

To divide the words in $W$ into some groups of words more easily associated with each other, cluster analysis is used, where the measure of similarity between two words is their relatedness.

*Probability of association between words*: With a stimulus word $w_i$ the probability to associate a word $w_j$ among various other words is:

$$p_{ij} = f_{ij}/f_i$$

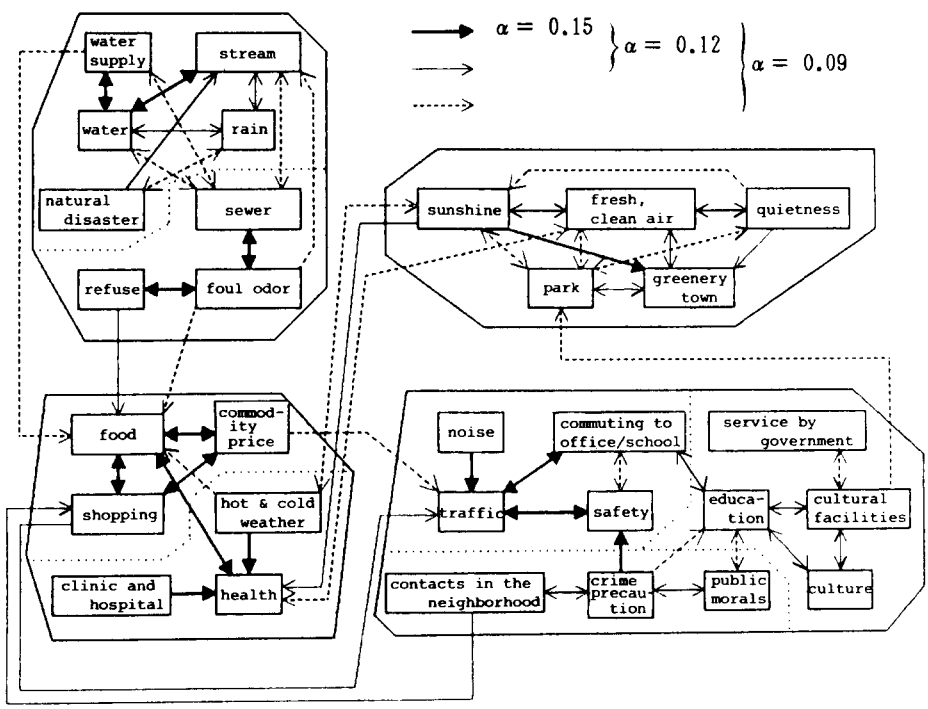


Fig. 3. Diagram of the cognitive structure by residents in ISHIOKA.

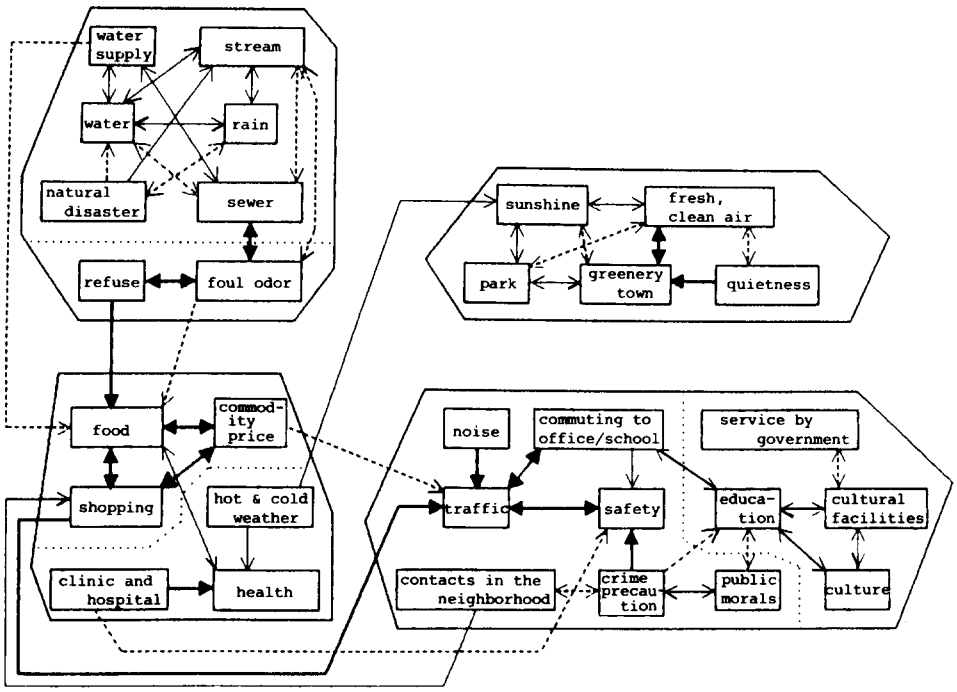


Fig. 4. Diagram of the cognitive structure by residents in HANAHATA.

Generally, for $N$ stimulus words

$$\sum_{j=1}^{N} p_{ij} \leq 1$$

In the present survey method, the equality is always valid, because the sets of stimulus words and of possible associated words are identical.

To represent the direction of association between any two words in $W$, the arc of association ($\rightarrow$, $\leftrightarrow$) is defined by the probability of association. That is:

(1) $w_i \rightarrow w_j \Leftrightarrow p_{ij} \geq \alpha, \quad p_{ij} \geq \beta p_{ji}$

(2) $w_i \leftrightarrow w_j \Leftrightarrow p_{ij} \geq \alpha, \quad p_{ji} \geq \alpha, \quad (1/\beta) \leq (p_{ij}/p_{ji}) \leq \beta$

(3) Otherwise, these is no arc between the two words.

*Diagram representation of cognitive structure*: The relatedness between the words in $W$ and the direction of association among them is depicted in a diagram of cognitive structure in the following way (Figs. 3–5). First, from the result of cluster analysis based on the relatedness between the words, words belonging to a cluster are displayed in a block encircled by line

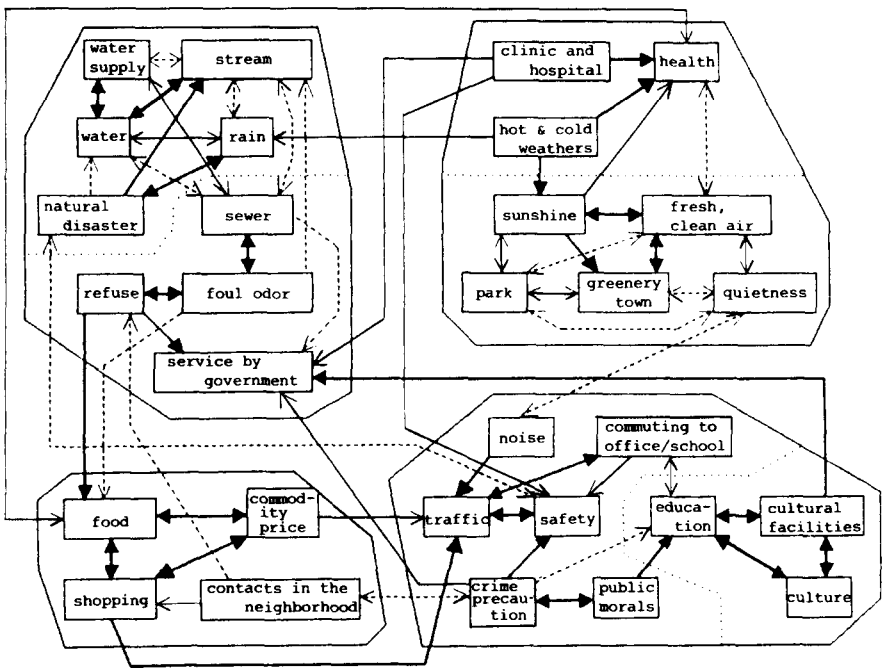


Fig. 5. Diagram of the cognitive structure by the employees in NIES.

segments. Then the arcs of association are drawn between the words. If necessary, word blocks and words in a block are relocated to diminish the number of criss-crossing and stretched arcs for easy intuitive understanding.

## QUESTIONNAIRE

In the eleven-page questionnaire of the main survey, the free association test is followed by the controlled association test. The third part of the questionnaire is face sheets. In the sheets the age, sex and vocation of a subject are to be answered first. Then, the questions about the number of family members, and of children attending nursery school, kindergarten, primary school or junior high school appear. Inquiries as to the ownership and the architectural style of his/her house and the duration of the respondent's occupation in the present place follow. Finally, the place of his/her longest dwelling is to be listed. The last sheet is prepared to be filled in with the respondent's opinions and comments about the survey.

## PACKAGE FOR DATA ANALYSIS AND A CLUSTER ANALYSIS METHOD

To carry out the analysis of survey data through the tests, a computer package "PAB" (Miyamoto, 1984) is employed which has been developed for studies in bibliometrics.

Among several methods of cluster analysis available in the PAB, the method of the average linkage between the merged groups is used for the tests described.

## UNIVERSES OF THE SURVEY

The survey was carried out in two local areas and in a research institute on environmental affairs.

The first area (ISHIOKA) is a residential district in Ishioka, a rural city 70 km north of Tokyo. The history of the city can be traced through more than 10 centuries. An area spotted with farmland neighbors the business and shopping area in the center of the city around Ishioka station, and there is an agricultural area on the outskirts of the city. The second area (HANAHATA) is on the northeastern edge of downtown Tokyo, a part of Hanahata district in the Adachi Ward. The area is separated from Saitama prefecture by the Ayase River and its tributary, which are among the most polluted water systems in the country. The dark-color of the stream can be seen from the banks, and its odor is readily perceptible in the area.

Each area is occupied by some 700 families. Residents in apartment houses were excluded from the universe, for they seem to have a considerably different consciousness from those owning their own homes. The second area is far more densely populated than the first one.

The third universe is all the employees of The National Institute for Environmental Studies (NIES) in Tsukuba, which comprises both researchers and administrative and engineering personnel. It should be pointed out that Tsukuba is a newly planned and sparsely populated town for higher education and scientific research in the countryside 60 km north of Tokyo.

In both ISHIOKA and HANAHATA, the survey was carried out in two periods, namely the end of November, 1983 and the end of February through the beginning of March in 1984. Between the two periods, Japan experienced an unusually cold winter. In each period and area, 150 different subjects were chosen through the systematic sampling technique. Self-administration, in which a canvasser hands out a questionnaire to a subject, and some days later recovers it filled in, was employed. In ISHIOKA the recovery ratio of the questionnaires was 69% in the first period and 66% in the second. In HANAHATA these figures were 70% and 76%.

In NIES all the 237 employees were given a questionnaire in early November, 1983, and 92% of them returned it within a month. The ratio of recovery being higher than in ISHIOKA and HANAHATA may be due to the difference in the delivering and recovering procedure of questionnaires, and to the fact that many of the subjects are acquaintances of the present researchers.

## RESULTS OF THE FREE ASSOCIATION TEST AND DISCUSSIONS

In Figs. 1 and 2, the upper two-thirds of the matrices of two-way cluster analysis of the data obtained from the free associaton test in ISHIOKA and HANAHATA are shown. The threshold value $s$ of the frequency of association of a word is 10. Associated words as well as respondents are divided into cluster by solid lines.

A resulting rectangle, densely filled with positive matrix elements indicates the close relationship between the corresponding cluster of responders and that of associated words. The first four digits for a responder are the code number of a subject, and the remaining nine characters denote the answers on the face sheets. A phrase with an asterisk in the word row is written in one word in the original Japanese language.

Even though the entire matrix is not displayed because of lack of space, the two figures show the significance and the usefulness of two-way cluster analysis.

In Fig. 1 the cluster of responders second from the top corresponds to that of the words "air" through "safety". That is, these responders' prime interests are hinted at by those in the word cluster. It should be noted that some words in other clusters as well as those in the second cluster are naturally involved in making up the responders' cluster.

The third cluster of responders has two major word clusters, i.e., the second and third ones from the left. The fifth cluster from the top is closely related to the third word cluster. The sixth cluster of responders is formed chiefly due to the words in the clusters third and sixth from the left.

In the same manner, the matrix in Fig. 2 is examined. Responders in the first and twelfth clusters from the top wrote only "green" and "environment", respectively.

The most frequently associated words in ISHIOKA are "traffic", "convenience", "quiet", "green", and "environment" with frequencies of 70, 61, 44, 42, and 42, respectively. In HANAHATA they are "green", "traffic", "convenience", "environment", and "quiet" with the frequencies of 71, 70, 69, 58, and 48, respectively. As for the members of the first seven words, including "neighborhood" and "air", the two areas have shown no difference.

In the two areas, the words "quiet" and "noise", which relate to the same issue in the desirable and undesirable aspects, do not belong to a cluster. In HANAHATA the cluster including "environment" does not have the word "pollution", whose frequency of association is 16, while in ISHIOKA "pollution", associated only eight times, is not found in Fig. 1.

Through the figures, a group of subjects deeply related to some clusters of words of examiner's interest can easily be found, and a further study in detail on their association and cognition can be carried out by closely reexamining only the questionnaires.

## RESULTS OF THE CONTROLLED ASSOCIATION TEST AND DISCUSSIONS

The results of the controlled association test of the survey are shown in Figs. 3–5, where the threshold $\alpha$ is chosen to be 0.09, 0.12, and 0.15, and $\beta$ is equal to 1.5. In the figures, a dotted line divides the four clusters surrounded by solid-line segments into seven or eight smaller clusters. For the larger four clusters the results in ISHIOKA and HANAHATA give the clusters with the same members, while the larger four clusters in NIES are different from those of the other two areas in "service by government", "contacts in the neighborhood", and the smaller cluster of "health", "clinic and hospital" and "hot and cold weather".

Despite the difference between ISHIOKA and HANAHATA, and NIES, each of the major four clusters stands for some concepts. The cluster including "sunshine" may be labeled as the "pleasantness". The cluster of "culture" may be named as the "social environment". Words intimately related to a household are found in the cluster including "food". The last cluster corresponds to words relating to water in its various aspects and "refuse".

In the following, the number of arcs in the structure diagram is examined, where $\leftrightarrow$ is two arcs. It should be noted that "service by government" in NIES has five arcs from others, while the phrase has only two arcs to and from "cultural facilities", with the threshold 0.09 in the other areas. In ISHIOKA 87, 55 and 25 arcs are found for $\alpha = 0.09$, 0.12 and 0.15, respectively. In HANAHATA the numbers are 81, 55, and 22, while in NIES they are 94, 66 and 44. For the larger threshold value, the diagram for NIES has nearly twice as many as each of the other diagrams. That is, the association by the employees of the NIES is less diverse than that of residents in ISHIOKA and HANAHATA.

For the case where 30 words are divided into the four larger clusters, in the diagram for ISHIOKA the number of arcs connecting two words belonging to different clusters is 12, 4 and 0 for the threshold values 0.09, 0.12 and 0.15, respectively. For HANAHATA the numbers are 8, 4 and 2, while for NIES they are 18, 10 and 3. The subjects in NIES show a higher probability of association between words in distinct clusters.

## DIFFICULTY IN TRANSLATION

A given word in a language cannot always have a one-to-one translation in another language, and even though a word may be found in a language as a suitable counterpart of a word in another language, what one of the pair stands for exactly cannot always be identical with what the other means (Fukuda, 1960). Furthermore, what is implied by a word is inevitably different from what is suggested by its translation. In the present study, it is

assumed and shown that words associated with a stimulus word differ from subject to subject even in the same language.

Therefore, it is impossible to find a translated word whose probable associated words are exactly identical with those of the original word in an original language.

## CONCLUDING REMARKS

Two association test methods have been proposed to examine the cognitive structure of the environment by local residents, and the extent and the variety of their interests concerning the living conditions. The method has been successfully applied in a survey in three universes of subjects. The results have shown the usefulness and the significance of the methods.

Further analysis of the results will reveal more detailed aspects of the cognition.

## ACKNOWLEDGMENT

The authors would like to express their gratitude to the subjects of the survey for their collaboration.

## REFERENCES

Fukuda, T., 1960. A pocketbook by a critic—some literary considerations of the functions of the language. Shincho-sha, Tokyo, 227 pp.

Miyamoto, S., 1984. Development of a computer program package for bibliometrics. Report of a research supported by the Grant in Aid for Fundamental Scientific Research of the Education Ministry in fiscal 1983, Tsukuba, 175 pp.

Miyamoto, S. and Nakayama, K., 1980. A hierarchical representation of citation relationships, IEEE Trans. Systems Man Cybern., SMC-10: 899–903.

Miyamoto, S., Miyake, T. and Nakayama, K., 1983. Generation of a pseudothesaurus for information retrieval based on coocurrences and fuzzy set operations, IEEE Trans. Systems Man Cybern., SMC-13: 62–70.