The Acquisition and Expansion of Knowledge Data By Analyzing Natural Language —Using Five-Character Kanji (Chinese character) strings—

YASUHITO TANAKA
Aichi Shukutoku University
Aichi—Shukutoku University 9 Katahira Nagakute Nagakute Aichi 〒480—11 Japan
TEL +81-561-62-4111 FAX +81-561-62-3007

Abstract

Knowledge data are indispensable for the comprehension and context analysis of natural language. The author describes the ways of acquiring and expanding such knowledge data. Kanji (Chinese character) strings are frequently used in the Japanese language. The author attached importance to five-character Kanji strings and decided to extract the five-character strings which can be divided into two character ⊕ three-character or three character ⊕ two-character combinations. A large quantity of such data were collected and knowledge data were further expanded by combining them with postpositive particles and auxiliary verbs.

Five-character strings were extracted from the Asahi Shimbun, and about 76,000 items of knowledge data were obtained by sorting them out. The knowledge data thus obtained could be further expanded.

1. Introduction

Knowledge data are necessary for the comprehension and context analysis of natural language. How can a large quantity of such knowledge data be prepared?

The methods employed in, this study to collect and expand such knowledge data are described below. To begin with, the method of collecting the data will be described. Five-character strings were used in collecting the data, and they were collected from one-year old issues of the Asahi Shimbun.

2. Why were five-character strings selected?

Five-character strings were chosen for the study for the following reasons.

(1) Five-character strings could be selected mechanically, and five-character strings are numerous following four-character strings. The number of two-character strings is the largest among all types of character strings.

(2) Five-character strings can be divided into two-character ⊕ three-character or three-character ⊕ two-character combinations, or combined words. Two-character and, three-character are basic terms and occur very frequently.

(3) The category of basic terms can be expanded into phrases with the addition of particles and other words inserted between the two component parts.

Example 1.

経営・多角化	Keiei takakuka
↓	
経営を多角化する	Keiei wo takakukasuru (to diversify management)
経営の多角化	Keiei no takakuka (diversification of management)

Example 2.

自主的・判断	Jishuteki handan
↓	
自主的な判断	Jishutekina handan (autonomous judgment)
自主的に判断する	Jishutekini handansuru (to judge autonomously)

Some strings form phrases by reversing the order of the two parts.

Example 3.

自動車輸出	Jidosha yushutsu (automobile export)
↓	
自動車を輸出する	Jidosha wo yushutsusuru (to export automobiles)
自動車の輸出	Jidosha no yushutsu (export of automobiles)
輸出した自動車	Yushutsushita jidosha (exported automobiles)
輸出する自動車	Yushutsusuru jidosha (automobiles to be exported)

(4) It is necessary to translate the same term in a variety of ways. Therefore, it may be studied as a technical term or as a translation selected for machine translation.

(5) It is easy to expand knowledge by utilizing two-character ⊕ three-character or three-character ⊕ two-character combinations collected from five-character kanji strings. Detailed explanations will be made in this article.

3. Acquisition of knowledge data through partition of five-character kanji strings

3.1 Collection of five-character kanji strings

Five-character kanji strings can be extracted mechanically from Japanese text.

Five-character kanji strings can be classified into two types of combinations of basic words; namely, a two-character ⊕ three-character combination or a three-character ⊕ two-character combination. The two component words can be changed into sentences or phrases.

In this study, 76,000 different five-character kanji strings were extracted from data contained in one-year old issues of the Asahi Shimbun. The total number of five-character Chinese character strings was 210,000.

The number of different five-character kanji strings divided by the total number of five-character Chinese character strings that occurred was 0.36. Of these, 39,000 strings were usable as knowledge data of concurrence relations.

The data were sorted out according to the following procedures.

1) Extraction of five-character kanji strings from Corpus

2) Compression of same-character strings and analysis of their frequency

3) Reference to already sorted out knowledge data (This procedure was omitted this time.)

4) Examination of content and storage as knowledge data

A book explaining the content of these procedures is scheduled to be published in the near future. Provided in the book will be data classified in the order of prepositional and postpositive particles.

Classification code

23…Data classifiable into 2 character ⊕ three character combinations
(仏人研究者) (Futsujin kenkyusha)
(French researcher)

32…Data classifiable into three character ⊕ two character combinations
(日本人気質) (Nihonjin katagi)
(Japanese trait)

70…Those classifiable into other than 23 and 32
(悲観主義者) (Hikan shugisha)
(pessimist)

80…Names of persons, enterprises and other proper names
(〜市議会、〜営業所、〜役場)
(〜shigikai (municipal assembly),
〜eigyosho (business establishment),
〜yakuba (local government office))

90…Place names

99…(unintelligible strings, strings requiring explanatory particles)
(同日朝現在、連勝中中国)
(Dojitsu asa genzai, renshochu chugoku)

Classification code	Types	Total number of combinations
23	17,705	57,271
32	22,076	60,092
70	8,823	24,316
80	10,101	41,988
90	1,244	2,744
99	16,293	25,381
Total	76,242	211,793

Data extracted from newspapers are characterized by a relatively high frequency of names of persons, enterprises and places.

A further examination of data classified into the 99 category will serve to improve the technique of collecting terms and analyzing them into form elements.

3 . 2 Analysis of component words

1) Results of analysis of five-character kanji strings classifiable into two-character ⊕ three-character combinations

	2-character strings	3-character strings
Code 23	4,199	7,467

2) Results of analysis of five-character strings classifiable into three-character ⊕ two-character combinations

	3-character strings	2-character strings
Code 32	8,253	3,914

3) Totals of two-character words and three-character strings

Two-character strings

4,199 + 3,914 = 8,113 ⇒ 6,417

Three-character strings

(duplications deleted)

7,467 + 8,253 = 15,720 ⇒ 13,527

These two-character and three-character string terms are basic and numerous in occurrence.

In the analysis of five-character strings, the reader's comprehension will be facilitated if the strings are written with a space between the component words. However, since all of these strings cannot be processed by machine, human interference is required.

If five-character strings are classified according to three-character and two-character component words, the following four combinations are possible.

1	○	○
2	○	×
3	×	○
4	×	×

(However, three-character words are processed on a priority basis.)

It is necessary to classify data largely in this way before it is examined by humans.

3 . 3 Simplification of coding work

Many five-character strings can be classified into three-character + two-character and two-character + three-character combinations. If we extract only three-character words, and analyze the final characters that appear within them, special characters can be extracted.

Therefore, it is possible to classify five-character strings into three-character ⊕ two-character and two-character ⊕ three-character combinations quickly by mechanically extracting five-character strings and analyzing the data by utilizing these words.

Examples : ～ teki, ～ ka, ～ sha, ～ sho

～的、 ～化、 ～者、 ～所

The following are three-character kanji strings in which teki and ka appear as the final characters.

～化			～的		
	種類	延件数		種類	延件数
民営化	35	421	政治的	195	580
自由化	33	93	国際的	115	344
合理化	23	102	具体的	110	409
近代化	19	58	社会的	106	362
民主化	16	78	経済的	92	454
情報化	14	93	歴史的	85	240
国際化	14	69	基本的	78	384
工業化	13	21	軍事的	53	126
実用化	10	17	本格的	49	83
活性化	10	13	技術的	47	76
一本化	9	22	個人的	42	98
機械化	9	12	世界的	42	80
自動化	9	9	戦略的	42	67
商品化	7	8	精神的	41	103
高齢化	6	229	国民的	39	169
軍事化	6	10	積極的	38	104
多角的	5	15	科学的	36	82
砂漠化	5	14	一方的	32	85
国産化	4	20	代表的	29	60
空洞化	4	16	比較的	29	41
保守化	4	16	総合的	28	45
孤立化	4	7	心理的	27	57
省力化	4	7	国家的	25	44

Furthermore, coding work can be simplified by classifying five character kanji strings in the following way.

(1)	X X X X X └┘ └┘ 2 1 3 Order of classification	Effective for analyzing three-character ⊕ two-character combinations
(2)	X X X X X └┘ 4 3 2 1 Order of classification	Effective for analyzing three-character ⊕ two-character and two-character ⊕three-character combinations
(3)	X X X X X └───────┘ 1 Order of classification	Effective for analyzing three-character ⊕ two-character and two-character ⊕three-character combinations

It is necessary to begin the coding task with the component that can be coded most easily, and to change the order of classification so that the coding can be accomplished accurately and quickly.

3 . 4 Detailed coding work

Data falling in the 70 category can be divided among 14 different detailed classifications. (For the purpose of eliminating 2, 3 ; 3, 2) However, the number of classifications becomes larger if the connection relations and parallel relations are considered in detail.

Partition patterns of five-character kanji strings

			Seq. No.
1)	5		1
2)	1, 4		2
	4, 1		3
3)	3, 2	Coded	4
	2, 3		5
	3, 1, 1		6
	1, 3, 1		7
	1, 1, 3		8
4)	2, 2, 1		9
	2, 1, 2		10
	1, 2, 2		11
	1, 1, 1, 2		12
	1, 1, 2, 1		13
	1, 2, 1, 1		14
	2, 1, 1, 1		15
5)	1, 1, 1, 1, 1		16

16 patterns of partitions

Generally speaking, the number of partitions is few in five-character strings, and in many cases the number of characters in partitioned words is the same. This is considered to be due to a phenomenon of optimization of the labeling of concepts or word expressions.

On this basis, it is necessary to perform detailed coding work on about 8,800 items of data coded in the 70 category.

4 . Expansion Method of Knowledge Data

4 . 1 Word-to-word relations in text

If we examine how five-character kanji strings are arranged in Japanese text, we will see what verbs and auxiliary verbs make up the phrases contained in the basic conceptual words of two-character ⊕ three-character and three-character ⊕ two-character combinations. This can be seen by analyzing data into phrases. It is also possi ble to examine this by forming actual sentences by making KWIC.

● 1 character ●、 ● 2 characters ●
● 3 characters ●、 ● 4 characters ●
● 5 characters ●、 …………

Here, ● represents a component element of five-character kanji strings.

It is possible to use the above relationship to make KWIC and collect new data in a concentrated way.

As for word-to-word relations and connectives, refer to Table 1.

4 . 2 Expansion of knowledge data

An attempt is made to expand knowledge data on the basis of word-to-word relations and con nectives.

Example:

新幹線建設 ↓	Shinkansen kensetsu	With/without relation
新幹線の建設	Shinkansen no kensetsu	O
新幹線が建設	Shinkansen ga kensetsu	X
新幹線を建設	Shinkansen wo kensetsu	O
新幹線に建設	Shinkansen ni kensetsu	X
新幹線で建設	Shinkansen de kensetsu	X

Only correct relations are marked and extracted.

● ・ ●

If the preceding word is a noun forming a Sa-hen verb,

● する ●　　● suru ●
● した ●　　● shita ●

are formed.

Example:

情報化・社会→情報化した社会
Johoka shakai → informationalized society

Furthermore, if it is the root of an adjective, −teki na is added to form the following.

● 的な ●　　● tekina ●
● 敵に ● する　　● tekini ● suru
When ● is a Sa-hen verb.

Example:

実質・成長率→実質的な成長率
Jisshitsu seichoritsu → Jisshitsuteki na seichoritsu
(real-term growth rate)

However, such combinations are required to develop the components into phrases for examination.

As there are cases in which the component words of two-character ⊕ three-character and three-character ⊕ two-character combinations are used independently, it is necessary to see if such component words are found in the dictionary and if they are not, it is necessary to add them.

In this way it is possible to collect five-character strings exhaustively, without any omission, and to increase knowledge data with some assurance. Furthermore, it is also possible to obtain printouts and distribute them to a large number of people, and as a result, to obtain and expand knowledge data.

There are connections with special verbs between two base words. It is all right if such connections are input as proper base-word-base word connections.

5. Other Methods of Expanding Knowledge Data

5.1 Expansion of knowledge by association

所得税・減税　Shotokuzei • genzei
↓
所得税を減税する　Shotokuzei wo genzeisuru
(to reduce income tax)
↓

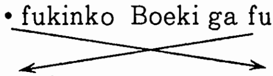
(i)所得税を減らす　(i) Shotokuzei wo herasu
(to reduce income tax)
(ii)税金を減らす　(ii) Zeikin wo herasu
(to reduce tax)

It is possible to expand knowledge by utilizing a thesaurus and synonyms.

5.2 It is also important to reverse the order of two-word combinations and examine if there are relations between the reversed words.

Example:

貿易・不均等　　貿易が不均等
Boeki • fukinko　Boeki ga fukinko (trade imbalance)



不均等・貿易　　不均等な貿易
Fukinko • boek　Fukinkona boeki (unbalanced trade)

Knowledge data can be expanded by utilizing the existing 390,000 items of knowledge data. All word combinations are not reversible.

5.3 Generation of compound words

Two-word combinations have the following relations with one word as the center.

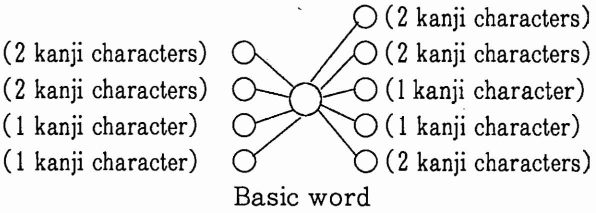


Fig. 1 Compositions of words

It is possible to examine the generation conditions of compound words by producing a composition diagram such as the one shown above.

6. Translation of Compound Words

6.1 English translations of five-character kanji strings

Some five-character kanji strings were selected and given English translations. The following became clear from these translations.

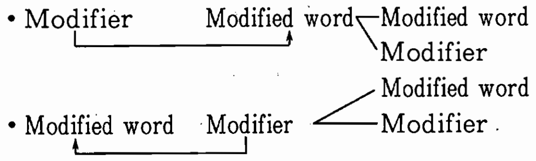
(i) English translations had diverse meanings
(ii) The joining of translated words is as follows.
Translation of base word ⊗ translation

of base word—resulting translation

Noun		noun		noun
Verb	ⓧ	verb	→	verb
adjective		adjective		adjective
adjectival ver		badjectival ver		badjectival verb

(iii) Translation of compound words

The translations of compound words differ according to whether they are modifiers or modified words.



(iv) Others

- Japanese base words do not necessarily correspond to English words word by word.
- This question is related very much to the quality of translators. Attempts should be made to automate or semiautomate translation.
- Translation work is costly, so it is desirable to reduce translation costs.
- In view of the large quantities of translation work needed, it is necessary to develop a system to improve the speed and ensure the accuracy of translation.

6 . 2 Translation of ordinary compound words

Translation of compound words is an important task. While it is necessary to use generally accepted translations of technical terms, etc., it is also necessary, after considering the details described in 6.1, to select translations according to the following procedures.

(i) To partition compound words into base words

(ii) To relate base words and partition them structurally

(iii) To give proper translations

This procedure is shown in the following example.

自然言語処理…複合語　Shizen gengo shori…Compound word

↓

自然・言語・処理…単語分割　Shizen • gengo • shori…Partition into words

↓

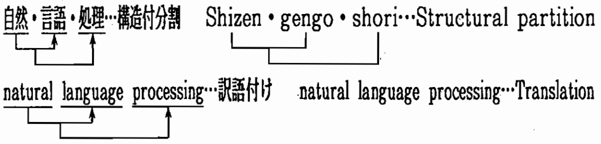

自然・言語・処理…構造付分割
Shizen • gengo • shori…Structural partition
natural language processing…訳語付け
natural language processing…Translation

Fig. 2 Process of translation

In order to complete a system of this kind, it is necessary to devise a system by which corresponding translations to base words are collected and selected for combination.

7 . Reference to Thesaurus

A thesaurus collects and sorts words according to similar concepts and systematically develops them into upper-ranking concepts.

As words are grouped according to similar concepts, it is possible to systematically grasp the characteristics of words. With the aid of other reference sources, the thesaurus serves to resolve the problems presented by the diverse meanings of the same words.

Word-to-word relations in this study should be referred to the Thesaurus and their concepts should be analyzed in detail. The thesaurus plays an important role in this respect.

7 . 1 Reference to Thesaurus

The reference of hundreds of thousands of items of knowledge data to the Thesaurus makes, it possible to supplement deficient knowledge data and extend the classification of concepts in the Thesaurus according to the meanings of words. It also enables machine translation to select accurate translations for words with different meanings. It is expected that a large scale machine-readable thesaurus will be provided.

Word-to-word relations can be obtained by the partition of five-character kanji strings and through additions of verbs or auxiliary verbs to them. The number of word-to-word relations will be tremendously large, and the task of collecting them and sorting them will pose a difficult problem.

If we consider a number of words (n), combinations multiplied by itself (n2) are possible, and with the addition of verbs, auxiliary verbs and other words, the number is increased to kn2. Very few of these Kn2 combinations make sense, but they must be examined.

Since this involves a tremendous amount of work, it may be possible to refer to the Thesaurus system and focus on the examination of groups of combinations which are supposed to make sense.

In this way, it is possible to keep a nearly infinite number of relations down to a limited number.

1) Meaning of reference to Thesaurus (i)

委員長就任 → 委員長が就任　Iincho shunin → Iincho ga shunin
　　　　　↘ 委員長を就任　　　　　　　↘ Iincho wo shunin

In this word-to-word relation, are Iincho (chairman), gicho (chairman of a meeting), kaicho (board chairman), shacho (president), officers, etc., on the same semantic marker?

To confirm the expansion of knowledge data and expanded semantic markers through reference to the Thesaurus

2) Significance of reference to Thesaurus (ii)

Reference of word-to-word concurrence relations of five-character kanji strings to the Thesaurus serves the following purposes.

- To facilitate the verification of the accuracy of the Thesaurus
- To facilitate the subdivision and integration of semantic markers
- To facilitate judgment in the selection of separate translations and in the extraction of exceptions

3) Meaning of reference to Thesaurus (iii)

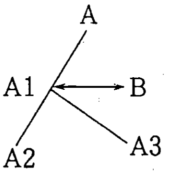
By combining knowledge data on word-to-word relations in five-character kanji strings with the Thesaurus, it is possible to know with what concepts verbs are combined. If it is known that there is a connection relation between A1 and the verb B, it is possible to examine whether all the words in Group A2 to which A1 belongs can be combined with the verb B.

If the connection with A1 makes sense, it may be considered a combination of A1 and B.

If it is known that the words in Group A2 can be combined with B, it is necessary to examine whether the same translation can be used.

Furthermore, the connection is expanded to A3, and its connection with B is examined. Similarly, connections are further developed to an upper-ranking concept of the Thesaurus. It is possible to save a tremendous amount of labor in this way.

It is because of this, also, that prepositional words are classified as basic conceptual words in knowledge data on word-to-word relations. The partitioning of five-character kanji strings is of great significance for this purpose, too.



7 . 2 Thesaurus and long-unit words

Compound words and long-unit words are used very frequently in text. There is a way of extracting base words from long-unit words, but for this purpose, it is necessary for many long-unit words to be incorporated in the Thesaurus.

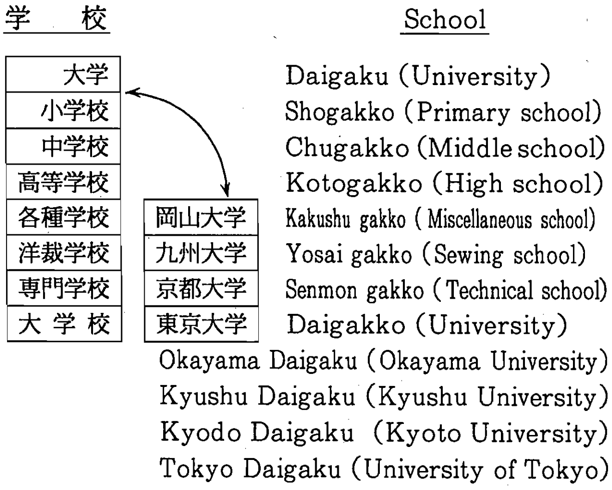

学　校
School
大学
小学校
中学校
高等学校
各種学校
洋裁学校
専門学校
大 学 校
岡山大学
九州大学
京都大学
東京大学
Daigaku (University)
Shogakko (Primary school)
Chugakko (Middle school)
Kotogakko (High school)
Kakushu gakko (Miscellaneous school)
Yosai gakko (Sewing school)
Senmon gakko (Technical school)
Daigakko (University)
Okayama Daigaku (Okayama University)
Kyushu Daigaku (Kyushu University)
Kyodo Daigaku (Kyoto University)
Tokyo Daigaku (University of Tokyo)

Fig. 4 Groups of words with lower level meanings are studied in order to further develop word-to-word relations, and upper and lower ranking groups are formed in relation to groups of words with lower level meanings.

7 . 3 Thesaurus and systematization of knowledge

In order to realize a high level machine translation system and sentence comprehension system, it is necessary to input knowledge in machines. This system of knowledge is represented by a thesaurus system and a system of concepts.

This system is illustrated as follows.

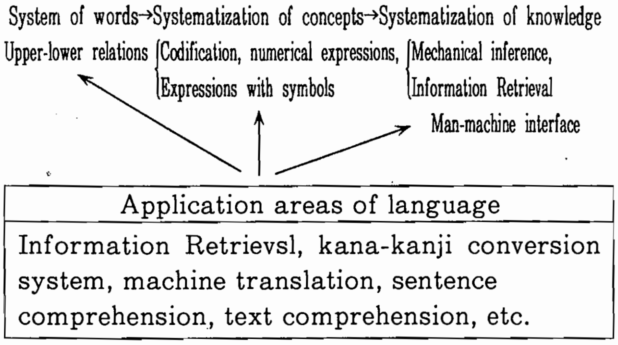

System of words→Systematization of concepts→Systematization of knowledge
Upper-lower relations
Codification, numerical expressions,
Expressions with symbols
Mechanical inference,
Information Retrieval
Man-machine interface
Application areas of language
Information Retrievsl, kana-kanji conversion system, machine translation, sentence comprehension, text comprehension, etc.

Fig. 5 Thesaurus of systematization of knowledge

It may be said that we use the Thesaurus to evolve a system of knowledge for incorporation into machines.

8 . Evaluation of Knowledge Data

A method has been established that enables us to collect large quantities of knowledge data. In the future it will be necessary for us to evaluate the knowledge data, i.e., to examine what data remains to be collected, to learn the extent to which the collected data is duplicated, and to find the answers to other questions.

It is also important to create an environment that permits additions and revisions to the collected knowledge data.

A step has just been taken toward the work of extracting large quantities of knowledge data. In order to incorporate knowledge data in machine translation systems, it is necessary for us to go through the following stages.

It is also vital to establish a method for evaluating a thesaurus, and to specify the parameters of a thesaurus.

The following can be suggested as the parameters of a thesaurus.

1) No. of words in the thesaurus

2) Application areas of the thesaurus

3) Contents of the thesaurus and its machine r eadability

4) Provision of various kinds of utilities for the use of the thesaurus

5) Relations between the succession of knowledge and inference systems, and other items.

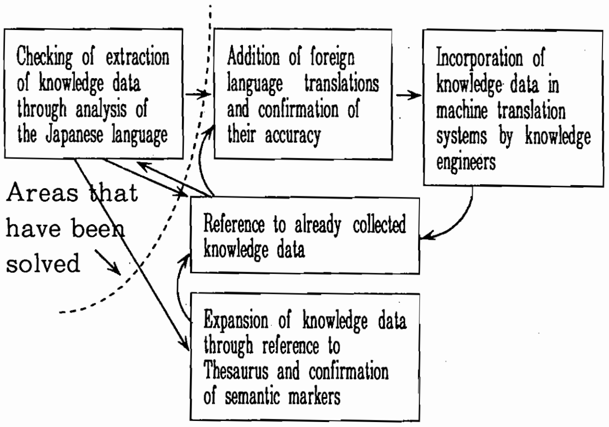

Checking of extraction of knowledge data through analysis of the Japanese language
Addition of foreign language translations and confirmation of their accuracy
Incorporation of knowledge data in machine translation systems by knowledge engineers
Areas that have been solved
Reference to already collected knowledge data
Expansion of knowledge data through reference to Thesaurus and confirmation of semantic markers

Fig. 6 Work processes until knowledge data on word-to-word relations are incorporated within a machine translation system

9 . Future Tasks

1) To translate all knowledge data in five-character kanji strings

- Translation and checking cost per case ¥500
- Cost for translating 39,000 cases ¥19.5 million
- Work volume per day (1 person) 100 cases／day
- Total number of man-days (about one year by 2 persons) 390 man-days

If the budget is limited, it is possible to begin the task with words that occur with a higher frequency, or to begin with a certain word and continue the work sequentially. Part of the contents of translations are to be shown upon completion of the task.

If the these expenses are to be paid, it is necessary to consider a proper distribution of costs to excellent translators and the costs for the examination and reference of data.

Furthermore, it is necessary to think of ways to improve speed and maintain the accuracy of translations.

2) Regarding data which are not collected in this experiment, it is necessary to study whether the lack of such data is due to the differences

of the areas covered or to the obsoleteness of the data itself, and other related matters.

3) It is necessary to make efforts to practically apply knowledge data on word-to-word relations to a machine translation system, kana-kanji conversion system, and voice and character recognition systems.

4) If large quantities of knowledge data on word-to-word relations in five-character kanji strings are obtained at low cost, studies on natural language will necessarily develop in a new direction. Just as it is said in philosophy that quantitative expansion leads to a qualitative change, studies on natural language must move onward to a new stage.

We are now in an era in which large quantities of knowledge data on word-to-word relations can be obtained at low cost.

The provision of large quantities of knowledge data

1) can systematize grammar (simplification, sophistication of grammar),
2) prevent the generation of a large number of structures in sentence structure analysis,
3) resolve the existence of many meanings of the same words in machine translation and improve the results of machine translations,
4) serve to improve the precision of character and voice recognition,
5) simplify the differentiation of homonyms and words of same the forms with different meaning, and
6) promote the development of semantic analysis in natural language processing.

10. Conclusion

It is desirable that knowledge on natural language be acquired automatically, but this poses difficult problems, due partly to the fact that knowledge data are needed in order to resolve the problem of the different meanings that exist for the same words in sentence structure analysis.

In this article, the author discussed five-character kanji strings obtained in a simple way, and methods of sorting them and expanding them. In addition, a study was made on giving translations which are necessary for machine translation.

The author is indebted to Yoriko Inada for her cooperation in the data-sorting work in this study, to the staff members of the Aoe Research Laboratory at the University of Tokushima for their cooperation in extracting five-character kanji strings and inputting them in code, and finally to Mr. Yoshikawa and Mr. Kato of Brother Industries, Inc., for their cooperation in data analysis.

References:

(1) Yasuhito Tanaka and Masaru Yoshida: Acquisition of Knowledge Data by Analyzing Natural Language, 11th International Conference on Computational Linguistics COLING, '86, August 1986

(2) Yasuhito Tanaka and Masaru Yoshida: Knowledge Data (Word-to-Word Relations) and solution of Multivocal word, Natural Language Processing, Information Processing Society, 60-3, March 1987

(3) Yasuhito Tanaka: Data for Analysis of Word-to-Word Relations – Explanations and Materials with "wo" as the Center (1), (II)
The Summing-Up Group on "Sophistication of Language Information," a designated scientific research project subsidized by the Ministry of Education, March 1987

(4) Yasuhito Tanaka: On Knowledge Data Based on Word-to-Word relations, "Quantitative Linguistics," Collection of Articles on the Japanese Language, Akiyama Shoten, March 1987

Table 1
Method of Acquiring Knowledge Data
（Method of expanding knowledge within frameworks, acquisition of knowledge through verification）

1－1	〜の〜	〜 no 〜
1－2	〜を〜	〜 wo 〜
1－3	〜が〜	〜 ga 〜
1－4	〜な〜	〜 na 〜
1－5	〜に〜	〜 ni 〜
1－6	〜で〜	〜 de 〜
1－7	〜や〜	〜 ya 〜
1－8	〜も〜	〜 mo 〜
1－9	〜と〜	〜 to 〜
1－10	〜へ〜	〜 e 〜
1－11	〜的〜	〜 teki 〜
1－12	〜性〜	〜 sei 〜
1－13	〜化〜	〜 ka 〜
1－14	〜・〜	〜・〜
1－15	〜，〜	〜，〜
2－1	〜から〜	〜 kara 〜
2－2	〜され〜	〜 sare 〜
2－3	〜した〜	〜 shita 〜
2－4	〜する〜	〜 suru 〜
2－5	〜との〜	〜 tono 〜
2－6	〜とかへ〜	〜 toka 〜
2－7	〜ない〜	〜 nai 〜
2－8	〜なり〜	〜 nari 〜
2－9	〜にも〜	〜 nimo 〜
2－10	〜には〜	〜 niha 〜
2－11	〜まで〜	〜 made 〜
2－12	〜への〜	〜 eno 〜
2－13	〜より〜	〜 yori 〜
2－14	〜での〜	〜 deno 〜
2－15	〜では〜	〜 deha 〜
2－16	〜的に〜	〜 tekini 〜
2－17	〜的な〜	〜 tekina 〜
2－18	〜性の〜	〜 seino 〜
2－19	〜性を〜	〜 seiwo 〜
2－20	〜上の〜	〜 jono 〜
2－21	〜側に〜	〜 gawani 〜
2－22	〜化の〜	〜 kano 〜
2－23	〜内の〜	〜 naino 〜
3－1	〜および〜	〜 oyobi 〜
3－2	〜された〜	〜 sareta 〜
3－3	〜される〜	〜 sareru 〜
3－4	〜しうる〜	〜 shiuru 〜
3－5	〜すべき〜	〜 subeki 〜
3－6	〜すると〜	〜 suruto 〜
3－7	〜だけを〜	〜 dakewo 〜
3－8	〜ている〜	〜 teiru 〜
3－9	〜できる〜	〜 dekiru 〜
3－10	〜である〜	〜 dearu 〜
3－11	〜という〜	〜 toiu 〜
3－12	〜として〜	〜 toshite 〜
3－13	〜と同じ〜	〜 to onaji 〜
3－14	〜とかの〜	〜 tokano 〜