Development of a New Support System for English Composition and its Performance Evaluation for International Communication

Hiroki Matsuyama
Faculty of Management and Governance, Shumei University, 1-1 Daigaku-cho, Yachiyo City, Chiba 276-0003, Japan

Mitsuyuki Miyazaki
Faculty of English and IT Management, Shumei University, 1-1 Daigaku-cho, Yachiyo City, Chiba 276-0003, Japan

Isamu Okada
Faculty of Business Administration, Soka University, 1-236 Tangi-machi, Hachioji City, Tokyo 192-8577, Japan

Terumasa Ehara
Ehara Natural Language Processing Research Laboratory, Setagaya-ku, Tokyo, Japan

Dawn Lavelle Miyazaki
Global Education Center, Waseda University, 1-104 Totsukamachi, Shinjuku-ku, Tokyo 169-8050, Japan

and

Shinichiro Miyazawa
Faculty of English and IT Management, Shumei University, 1-1 Daigaku-cho, Yachiyo City, Chiba 276-0003, Japan

ABSTRACT

English proficiency has become essential for Japanese people in today's globalized society. However, since the structure of the Japanese language is very different from that of English, it has proven difficult for Japanese people to create natural and fluent English sentences without specialized training. We developed a support system for English composition using a new method which addresses this issue. The main characteristic of the system is the use of a dictionary of similar sentence patterns. This dictionary was developed by defining a new distance measurement between sentences that emphasize expressions at the end of a sentence in a Japanese text due to the head finality of Japanese. Our experiment revealed that in terms of fluency of translation, higher scores were obtained with this system, in comparison with the singular use of a word dictionary. Also, in terms of both adequacy and fluency, the average values with this system exceeded those with traditional support systems for English composition considered in this study.

Keywords: Support System for English Composition, Corpus, Similarity, Clustering, Patent Sentences.

1. INTRODUCTION

Systems to support English composition or to translate Japanese sentences into English have already been put into practical application [1]; however these systems still present many issues. The template type support systems for English composition are widely commercialized and are suitable to handle stylized documents such as business letters,[1] but their field is rather narrow and applicability is considered limited [2]. Corpus-based translation systems have been extensively studied in recent years in order to solve this shortcoming [3-8]. This method involves extracting example sentences or similar sentences from the corpus by entering a keyword or key sentence. It is possible with this method to extract similar sentences using a large volume of adequate corpus; therefore, there are numerous studies on how to create a large volume of corpus and find similar sentences using this method. However, because of the high volume of similar sentences, this necessitates employing a skilled translator to authenticate the precise meaning of similar expressions. Thus, the task is to present similar sentences of higher quality and exactitude which reflects the intentions of the user. It follows from the above that the crucial factor in extracting adequate sentences from the corpus in the translation memory is the appropriate type of algorithm employed for this process.

In a related study, Wang and Ikeda [9] examined sentence patterns and structures. This study did take into consideration the structural characteristics of Japanese sentences; however, it is not applicable to the English language, because it only focused on the existential sentences using "aru" and "iru," dealing with the translation between Japanese and Chinese. In addition, in the study by Ikehara et al. [10] a dictionary of sentence patterns targeting compound and complex sentences was constructed, but it did not have the function of a translation support system. Furthermore, Taniguchi et al. [11] focused on head finalization, a characteristic of Japanese sentence structure. Their study employed the method of shifting the word order of an original language into that of the target language before conducting statistical machine translations, and therefore it did not create a dictionary. Finally, Amano et al. [12] investigated

[1] http://pf.toshiba-sol.co.jp/prod/hon_yaku/business/index_j.htm

on retrieving sets of English sentences for English composition. This study proposed the generalization method to display simplified sample sentences, thereby not paying attention to the sentence-end expressions of Japanese.

In this study, an algorithm was developed to define a new distance measurement between sentences by focusing on sentence patterns and to present similar sentences according to the distance. This algorithm is characterized by its focus on the end of a sentence. This is considered significant since Japanese is a head-final language in which the head phrase follows the dependent. To evaluate the effectiveness of this algorithm, experiments on test subjects were conducted twice, to verify the difference in average values using the data obtained. As a result, it has become evident that in terms of fluency of translation 1% significance level was obtained in this support system, compared with the case where subjects only used a word dictionary. Although a difference of 5% significance level was not attained in adequacy and fluency in comparison with SCOPE[2] [13] by Sakai, et al., the average values were proved to be higher.

The remainder of this paper will address the following topics: Chapter 2 describes the development of a dictionary of similar sentence patterns. Chapter 3 illustrates the evaluation experiment with a prototype and resulting discussion. Chapter 4 discusses characteristics, evaluation and results of this system. Finally, Chapter 5 provides summary of the paper and future tasks.

2. DEVELOPMENT OF A DICTIONARY OF SIMILAR SENTENCE PATTERNS

We develop a dictionary of similar sentence patterns by focusing on sentence-end expressions in the Japanese language. Here, "similar sentence patterns" means a set of sentences obtained by the clustering methods described in subsection 2.1.2 and 2.1.3. Also, we call the closest sentence to the centroid of the cluster the "representative sentence." By gathering all representative sentences of the clusters, we can construct a dictionary of similar sentence patterns.
This dictionary uses distance between sentences, newly defined to extract similar sentences. The distance between sentences was traditionally defined with similarity in words used in a source Japanese sentence, or similarity in the length of the word count in general. While translation of a word itself can be easily obtained, it is relatively difficult to specify the sentence structure or sentence pattern. Furthermore, the meaning is often determined by the end of a sentence, previously referred to as a head final characteristic. For example, an interrogative and subtle nuance of a sentence in Japanese are determined largely by the last phrase. Consequently, if the distance between sentences can be defined that focuses on sentence-end expressions, it is easier to extract sentences with similar sentence patterns.

To focus on sentence-end expressions, dependency structure of a sentence is analyzed and the distance between sentences is defined to minimize the cost of correspondence between phrases for which the depth from the root phrase is deep. A root phrase is at the highest destination as a result of dependency structure analysis, and comes at the very end of a sentence. This relates to the fact that Japanese is a typical head final language. Similarity of sentence patterns as a major framework can be captured by defining with this policy, rather than exceptional sentences including inversion. Focusing on the information on the depth from the root phrase is the characteristic of this system.

2.1. Flow to Prepare the Dictionary of Similar Sentence Patterns

The procedure to prepare the dictionary of similar sentence patterns by using the distance between sentences is overviewed in Figure 1.

Japanese sentences in a Japanese-English bilingual corpus are analyzed by morphological and dependency analyzers and then converted into dependency trees. A cluster analysis is conducted on a set of dependency trees and a data aggregate is created by the two steps described below.

Clustering in Figure 1 is a method to collect sentences with a close distance with the furthest neighbor method, and a dictionary can be developed by building the aggregate (cluster) of these sentences and extracting "representative sentences[3]" by each sentence pattern. Clustering should be performed in accordance with the distance matrix created with the distance between every sentence; however, it is performed with two steps to reduce the cost for calculation. First, we adopt a policy to keep the calculation time realistic by collecting sentences with corresponding sentence-end expressions to create a temporary cluster of sentences (clustering by sentence end) and performing clustering only with many sentences in this temporary cluster of sentences based on the distance between sentences. The cluster created as a result of clustering becomes the candidate of sentence patterns. The centroid sentence is selected as the representative sentence for the relevant cluster by using the distance matrix of the cluster. By collecting these sentences, it is possible to create an index based on the dictionary of similar sentence patterns. Sentences in the cluster are indicated for each index. Each process is explained in detail below.

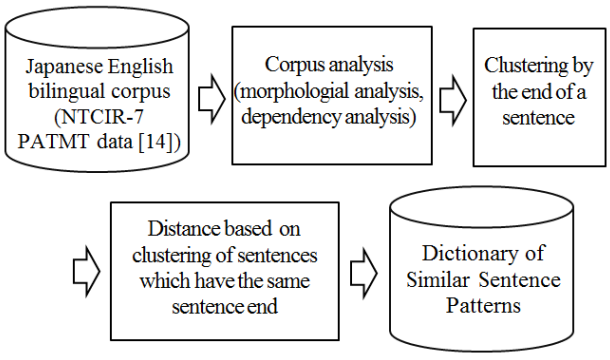

Figure 1: Process to prepare the dictionary of similar sentence patterns

2.1.1. Corpus Analysis: The method to analyze entered sentences is described in this section. Morphological analysis and dependency analysis are conducted in connection with the Japanese-English bilingual corpus used for training

[2] http://scope.itc.nagoya-u.ac.jp/

[3] Here, "representative sentence of a cluster" is defined by the closest sentence to the centroid of the cluster.

professional translators consisting of patent documents with approximately 1,800,000 model sentence pairs (PATMT) provided by NTCIR−7 [14], to convert them into a phrase dependency format by each morpheme unit. ChaSen [15] and CaboCha [16] are used for morphological analysis and dependency analysis, respectively. Next, the CaboCha output file is converted into a dependency format by each morpheme unit, which is further converted into a file in a phrase dependency format by each phrase unit. ("Phrase unit" means a chunk in the Japanese language (bunsetsu).) So that it can be processed by each phrase. PATMT corpus was used because patent documents are similar to scientific and technical sentences and this was the only large-scale bilingual corpus available for this purpose at the present time.

Since it is necessary to add phrase features in subsequent processing, we defined heading and depending features (Table 1) attached to each phrase to create an extracted file of phrase features.

Table 1: Phrase features

Heading Features

MEANING	SYMBOL
Noun	N
Predicate	V
Verbal Noun	NV
Predicate Noun	NV
Adverb, Adnominal	E
Conjunction	NV

Depending Features

MEANING	SYMBOL
Adverbial Modification	y
Adnominal Modification	t
Adverbial or Adnominal Modification	ty
Biding Particle "Ha"	h
Final	s

The file in a phrase dependency format by each phrase unit prepared in the above processing is merged into the extract file of phrase features to create a merge file of dependency and phrase features. Sentence-end expressions are extracted from the merged file to create a file of sentence-end expressions. A file is created by adding the data on the dependency depth that represents the dependency count to the root phrase (phrase at the end of a sentence). Regarding the dependency depth, the depth from the root phrase is defined as the path length for each phrase to reach the root phrase along depending features.

2.1.2. Creation of Cluster by Sentence-End Clustering: A cluster is prepared by collecting sentence-end expressions. For this purpose, sentence-end expressions in the file of sentence-end expressions created earlier are divided by phrase count. The phrase count is divided into one phrase, two phrases and three phrases, which are then filed. Four or more phrases are ignored due to the calculation volume. Sentence-end expressions are then collected by phrase count to create a cluster of sentence-end expressions. Since many sentence ends have one to three phrases, we used three types of phrase counts. For example, the sentence-end expression *〈得る(eru) ／ ことが(kotoga) ／ できる(dekiru) ／ "can be obtained"〉* consists of three phrases, and the sentence-end expression *〈説明する(setsumeisuru) ／"explain"〉* consists of one phrase.

2.1.3. Creation of a file in which sentences with the same ending are clustered by distance: The distance of sentences with the same sentence end is calculated for clustering. For this purpose, three steps are taken as follows:

(1) Creation of a file with clustering by distance

A file by sentence-end expression is created from the dependency depth file as well as from the data with collection of sentence-end expressions. For realistic calculation of the distance, only phrases with a depth level 3 or less are kept to calculate the distance between each sentence within the file and create a file of calculation results of the distance between sentences (the calculation method of the distance between sentences is explained in 2.2). Furthermore, the data on distance calculation is clustered to create a file with clustering by distance. An example of dependency depth file is shown below (Figure 2).

|S|	
|P|0|1|3|流体圧シリンダ３１|N|の|t|/P|	the fluid pressure cylinder 31
|P|1|5|2|場合|N|は|h|/P|	when…is used
|P|2|4|3|流体|N|が|y|/P|	fluid
|P|3|4|3|徐々に|E||y|/P|	gradually
|P|4|5|2|排出される|NV||t|/P|	applied
|P|5|6|1|こと|N|と|y|/P|	(There is no English equivalent expression)
|P|6|-1|0|なる|V|。|s|/P|	is
|/S|	

Figure 2: Example of dependency depth file

The following indicates the makeup of Figure 2.

| P | Dependent Phrase No. | Head Phrase No. | Depth | Heading Word | Heading Feature | Depending Word | Depending Feature | /P |

Also, the following examples illustrate dependency (Figure 3) and depth (Figure 4).

Dependent Phrase No.	Head Phrase No.	Heading Word / Depending Word
0	1	流体圧シリンダ３１｜の the fluid pressure cylinder 31
1	5	場合｜は when…is used
2	4	流体｜が fluid
3	4	徐々に gradually
4	5	排出される applied
5	6	こと｜と (There is no English equivalent expression)
6	-1	なる｜。 is

Figure 3: Example of dependency

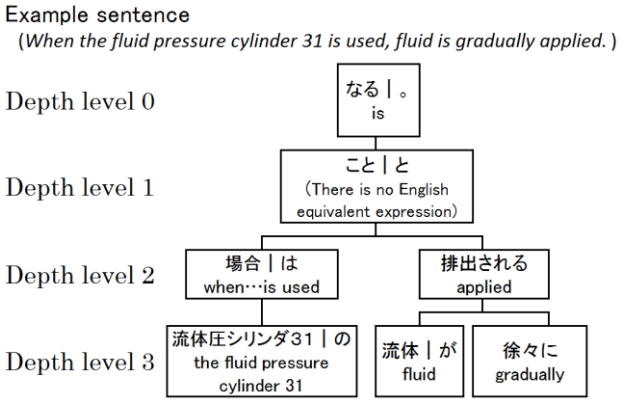

Figure 4: Example of depth

(2) Creation of an extract file of sentence ID containing sentence-end expressions
From both the dependency depth file created and the data with the collection of sentence-end expressions, sentence ID with sentence-end expressions described in the data with the collection of sentence-end expressions are extracted by each sentence end to create an extract file of sentence ID containing sentence-end expressions.

(3) Creation of a file in which sentences with the same ending are clustered by distance
Three files, including the one with sentences targeted for analysis (sets of sentences extracted from NTCIR-7 PATMT as the target for analysis), the one with sentences clustered by distance, and the one from which the sentence ID was extracted are merged to create the dictionary of similar sentence patterns (or a file in which the same sentence endings are clustered by distance). Sentences representing the cluster explained in 2.1 are added as indices to each cluster in the dictionary of similar sentence patterns.

2.2. Algorithm to Calculate Distance between Sentences

The algorithm is defined to calculate the distance between sentence a and sentence b in this section. In the case of this support system for English composition, sentence a is a Japanese sentence as the target of translation and sentence b is a sentence taken from the database. For preprocessing, morpheme analysis with ChaSen and CaboCha, phrase analysis, and dependency structure analysis are conducted for both sentences a and b to decompose them into phrases. After that, heading features and depending features of phrases indicated in Table 1 are automatically attached to each phrase to extract the dependency depth data. The following algorithm is performed after the above preprocessing. Values for each type of parameters are empirically determined at this time.

1) i and I represent the phrase number and phrase count of sentence a, respectively $(i=1,\cdots,I)$. k and K represent the phrase number and phrase count of sentence b, respectively $(k=1,\cdots,K)$. The beginning dummy phrase $i=0$ is considered for sentence a, and the beginning dummy phrase $k=0$ is considered for sentence b as well.

2) The distance between sentences is considered as the upper limit (1.0 in this case) when the following conditions are satisfied.

(1) The phrase count for sentence a is less than 5 and the difference of the phrase count between sentences a and b is 2 or more.
(2) The phrase count for sentence a is less than 10 and the difference of the phrase count between sentences a and b is 3 or more.
(3) The phrase count for sentence a is 10 or more and the difference of the phrase count between sentences a and b is 4 or more.

3) If the above is not applicable, the cost of correspondence between phrases $C_{i,k}$ is calculated as follows in regards to all pairs of phrases. In this case, $cost_1$ is determined by the consistency level of the word forms, depending features and heading features between phrases. The smaller of the depth of the phrase i and phrase k from the root phrase is considered as $depth$, i.e., $cost_2 = 2^{-depth}$.

$$C_{i,k} = cost_1 \times cost_2$$

・The content word form and function word form of phrase i and phrase k are consistent. $\Rightarrow$ $cost_1 = 0.0$

・The heading feature and function word form of phrase i and phrase k are consistent. $\Rightarrow$ $cost_1 = 0.2$

・The function word form of phrase i and phrase k is consistent. $\Rightarrow$ $cost_1 = 0.4$

・The heading feature and depending feature of phrase i and phrase k are consistent. $\Rightarrow$ $cost_1 = 0.6$

・The depending feature of phrase i and phrase k is consistent. $\Rightarrow$ $cost_1 = 0.8$

・The heading feature of phrase i and phrase k is consistent. $\Rightarrow$ $cost_1 = 0.9$

・There is no consistency between the phrase i and phrase k. $\Rightarrow$ $cost_1 = 1.0$

4) The accumulated cost of correspondence between phrases $d_{i,k}$ is calculated as follows in regards to all pairs of phrases. The accumulated cost of correspondence between phrases is expressed as $d_{0,k} = k$ in the case of $i=0$ and $k=1,\cdots,K$ and $d_{i,0} = i$ in the case of $k=0$ and $i=1,\cdots,I$ for the purpose of initialization.

$$d_{i,k} = \min\{d_{i-1,k-1} + c_{i,k}, d_{i,k-1} + c_{*,k}, d_{i-1,k} + c_{i,*}\}$$
$$c_{*,k} = 2^{-depth(k)}$$
$$c_{i,*} = 2^{-depth(i)}$$

5) The distance between sentences is obtained with $d_{I,K}$.

Consistency in the root phrase (phrase at the end of a sentence) is the most important for the distance between sentences calculated in this way, and the level of importance decreases as the depth from the root phrase becomes deeper by definition. Similarity in sentence patterns is determined with a definition that focuses on sentence-end expressions. It is confirmed with this calculation procedure that the distance between sentences can be calculated as 0.263 when sentences a and b are described

as *〈これにより、雌コンタクト長をさらに短くすることができる。"In this way, the length of the female contact can be made even shorter."〉* and *〈溶存オゾン濃度が高ければ注入時間を短くすることができる。 "The higher the dissolved ozone density is, the shorter the injection time might be made."〉*, respectively. The cost and the accumulated cost of correspondence between phrases are indicated in Tables 7 and in Table 8, respectively.

Heading feature value of a head phrase and depending feature value of a dependent phrase must correspond. For example, if a head phrase has the heading feature value "Noun," the dependent phrase should have the depending feature value "Adnominal Modification" or "Adverbial or Adnominal Modification." Thus, heading feature value and depending feature value are crucial in determining structural similarity of sentences.

3. EVALUATION EXPERIMENT WITH A PROTOTYPE

To evaluate the effectiveness of the dictionary of similar sentence patterns developed with the method in the above section, a prototype of the support system for English composition was prepared by incorporating this dictionary for the purpose of an evaluation experiment**.**

3.1. Overview of the Support System for English Composition

First, an overview of the support system for English composition using the dictionary of similar sentence patterns is given below. Figure 6 illustrates the user interface of the support system for English composition, consisting of the area to enter Japanese sentences and to indicate sentences with similar sentence patterns. Relevant sentences with similar sentence patterns are indicated in the order of the distance between sentences (similarity level) by entering Japanese texts and clicking the search button. Figure 5 is the structure of the support system for English composition. First, the Japanese sentence to be translated is entered in the area designated for user input in the interface. Next, by clicking the search button, morpheme analysis with ChaSen and dependency structure analysis with CaboCha is performed on the entered sentence. Then, pattern matching is performed in the dictionary where sentences with similar patterns were stored. Distance between the input sentence and representative sentences of the clusters in the sentence pattern dictionary is calculated and the data in the nearest cluster are displayed to the user. The distance focusing on sentence ends (Section 2.2) is calculated at this time. Pairs of Japanese and English sentences are displayed in the area to indicate sentences with similar sentence patterns in the order of closer distance out of all indices. All sentences within the relevant cluster can be viewed from each pair, based on which the translator creates English sentences.

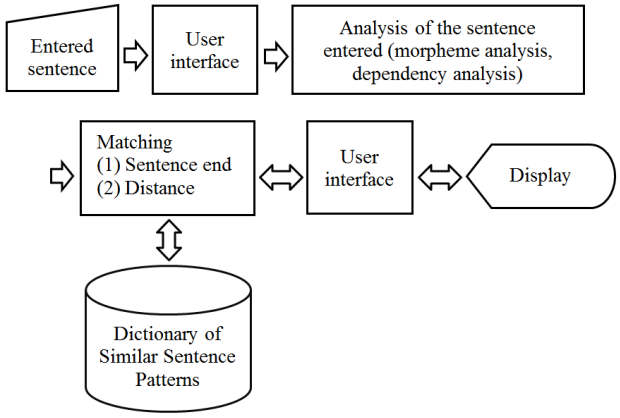

Figure 5: Structure of support system for English composition

3.2. Evaluation experiment

The evaluation experiment at this time is conducted by creating a prototype of the support system for English composition. Rather than analyzing entered sentences with ChaSen or CaboCha with the prototype, we adopted a method in which the distance for example sentences are calculated in advance and displayed in the area to indicate sentences with similar sentence patterns in the order of similarity level in regards to 24 Japanese sentences.

The experiment was conducted twice in 2011 and 2013. There were 12 subjects in the first experiment, where English sentences composed by using this system and a word dictionary were compared. Distribution of TOEIC scores of the subjects is indicated in Table 2. The subjects were undergraduate students, and scientific and technical documents (patent sentences) were targeted for translation.

Table 2: TOEIC scores for subjects in experiment 1

	TOEIC scores for subjects			
	500-599	600-699	700-799	800-899
Number of people	4	3	3	2

The experiment was conducted using a cross validation method to eliminate the influence of the subjects' ability to create English sentences, as well as fatigue due to the prolonged task of composing a composition in English.

First, 24 problems were divided into two groups to make them appear at a similar level of difficulty at first glance. These two groups are referred to as Problem X and Problem Y respectively. Next, the subjects were divided into four groups to average their English proficiency and named in order as the Groups 1 to 4. The experiment was conducted by each group as indicated in Table 3.

Table 3: Grouping and experimental procedure with the cross validation method

	Group 1	Group 2	Group 3	Group4
Session 1	Solve Problem X by using this system	Solve Problem Y by using this system	Solve Problem X by using a dictionary	Solve Problem Y by using a dictionary
Rest	15 minutes	15 minutes	15 minutes	15 minutes
Session 2	Solve Problem Y by using a dictionary	Solve Problem X by using a dictionary	Solve Problem Y by using this system	Solve Problem X by using this system

English sentences composed by the subjects in the experiment were manually evaluated by using adequacy and fluency [12]. Adequacy was evaluated by a native Japanese speaker teaching English at college and fluency was evaluated by a native English speaker teaching English at college. The highest and lowest scores are 5 and 1, respectively.

As an analysis method, the difference in average values for adequacy and fluency were tested on the implemented data.

Table 4: Test results of adequacy and fluency in experiment 1

		Adequacy	Fluency
	This system	2.681	3.084
	Dictionary	2.496	2.821
Testing of the difference in two population mean values	Significant probability assuming equal variances (two-tailed)	0.050	0.002
	Significant probability not assuming equal variances (two-tailed)	0.050	0.002

N=238+240

Evaluation results of statistical processing are indicated in Table 4. The table clearly indicates that the average values were higher for both adequacy and fluency when the support system for English composition was used in comparison with the singular use of a word dictionary. Regarding the testing of the difference in two population mean values, they are significant: 95% and 99% or higher, respectively. These results suggest that this system improve accuracy and is superior in consideration for improved fluency in English sentences composed.

For the second experiment, SCOPE[2] [13], a free system to search phrases provided by Nagoya University, was used instead of a dictionary following the same procedure as the first experiment. Although the system of SCOPE differs from ours, both aim to support composing scientific and technical sentences, while at the same time being able to output example sentences.

The subjects were comprised of 12 postgraduate students specializing in computer science with the task of translating material in their field of study. A situation closer to the one where students compose English sentences in the area of their specialty was created in this way. Distribution of TOEIC scores for the subjects is indicated in Table 5. Two subjects had TOEFL scores only, which were converted into TOEIC scores according to the score conversion guideline by IELTS NAVI[4].

[2] http://scope.itc.nagoya-u.ac.jp/

Table 5: TOEIC scores for subjects in experiment 2

	TOEIC scores for subjects					
	200-299	300-399	400-499	500-599	600-699	700-799
Number of people	1	2	2	5	1	1

In the same way as experiment 1, English sentences composed were manually evaluated by using previously established benchmarks relating to adequacy and fluency.

As an analysis method, testing similar to experiment 1 was conducted.

Table 6: Test results of adequacy and fluency in experiment 2

		Adequacy	Fluency
	This system	2.801	3.028
	SCOPE	2.743	2.917
Testing of the difference in two population mean values	Significant probability assuming equal variances (two-tailed)	0.603	0.393
	Significant probability not assuming equal variances (two-tailed)	0.603	0.393

N=144×2

Evaluation results of statistical processing are indicated in Table 6. As clearly indicated in the table, average values were higher for this system in comparison with the control system (SCOPE) for both adequacy and fluency, even though a difference at 5% significance level was not obtained.

4. DISCUSSION

Evaluation of this system and results are discussed in this section.

Evaluation experiments were conducted in this study targeting scientific and technical documents (patent sentences) by using this proposed method first, and then by using a word dictionary. Adequacy and fluency were manually evaluated. As a result of evaluations with the testing of the difference in average values, adequacy was found to be higher and fluency was proved even higher with the use of this system. Therefore, it was clarified by the experiments that this system could support not only creating English sentences by which the meaning is correctly communicated, but also composing natural English sentences in terms of grammar and expression. We successfully approached the goal of this study which was to help create a better quality English composition compared with the use of a dictionary- based system.

Comparison with SCOPE was conducted in the second experiment. The results of this system showed that average values were superior for this system; however variance was too large to obtain 5% significance in the testing of the difference in average values. It is probably because there were only 12 subjects. Differences will be clarified when the amount of data increases. At least, this experiment indicated that quality of the same level as SCOPE was successfully maintained.

[4] http://ieltsnavi.com/score_conversion.html

5. SUMMARY AND FUTURE TASKS

To facilitate extraction of sentence patterns, considered crucial in Japanese-English translation, we proposed a method of focusing on the end of a sentence, in place of the traditional type that extracts candidate sentences with matching keywords only. With this method, sentences with similar sentence-end structure are grouped with a clustering method to develop a dictionary of similar sentence patterns with representative sentences as indices. By using this dictionary, it is possible to preferentially display multiple sentences with the same sentence-end structure as the sentence to be translated. This method is expected to demonstrate the effects of composing natural-sounding English sentences that cannot be created with machine translation which relies heavily on literal translation. To confirm this, a prototype of the support system for English composition was developed by using the dictionary of similar sentence patterns. With this system, English sentences that effectively incorporate information on sentence patterns can be presented by entering a Japanese sentence itself, rather than Japanese keywords. Using this system, evaluation experiments were conducted with subjects. The result proved that when this system was used in comparison with the use of a word dictionary only, both accuracy and fluency of composed English sentences improved and the improvement of fluency was particularly superior. The same level of quality was successfully maintained in the case of comparison with a similar system (SCOPE) as well.

Our future tasks include verification of adequacy in the distance between sentences defined, as well as implementation of the system by including corpora of other fields.

In addition, although experiments were conducted this time comparing our system with SCOPE, more comparative experiments may be performed involving different types of translation support system, other than SCOPE (e.g. QRedit[5] [18]).

It is necessary to conduct experiments with different subjects in order to obtain more accurate results. It is our hope to acquire new knowledge and realize further improvement of the system by repeating the experiments.

Other tasks include whether or not the sentence pattern prepared by defining the distance between sentences is consistent with the so-called linguistic sentence pattern. The distance between sentences is empirically defined at the present stage. Thus, linguistically adequate sentence patterns are not always extracted. There is still some room for discussion in this regard. It is also necessary to calculate the distance between all sentences after target sentences are presented. We propose collecting in advance similar sentences in the corpus into one cluster rather than by calculating the distance between all sentences. By narrowing the cluster, the effectiveness of the system will be improved. Furthermore, we hope to improve practical use by extracting and implementing sentences with similar sentence patterns from corpora in other fields, without limitation to patent sentences.

In any case, we hope to proceed with the implementation of a support system of higher quality for English composition taking into consideration the knowledge and tasks we obtained.

[5] http://trans-aid.jp/index.php/stat/aboutus#qredit

6. REFERENCES

[1] Isao Tominaga, Masayuki Sato, **The evolution and practical application of machine translation system (I) Details and present situation of development in Japan and overseas**, Journal of Information Processing and Management 33(7), 1990, pp.593-605.

[2] Takeo Igarashi, **Translingual Communication by Explicit Specification and Presentation of Sentence Structures**, 19th Workshop on Interactive Systems and Software, 2011.

[3] Takeshi SAKAI, **WEB service for Similar Sentences Search, Japanese-French, German, Spanish, Italian and Portuguese: Kotobasagashi**, 2012 [in Japanese].

[4] Shigeki Matsubara, Seiji Egawa, Yoshihide Kato, **ESCORT: English Sentence Retrieval System: Library Service using Article Database**, Preprints of the information professional symposium, 2007, pp. 125-129 [in Japanese].

[5] Hironori Oshika, Manabu Satou, Susumu Ando, Hayato Yamana, **An English Composotion Support System using Google**, 2005, 4B-i8 [in Japanese].

[6] Masumi Narita, **User Assessment of a Web-based English Abstract Writing Tool**, working papers of Grant-in-Aid for COE Research (Researching and Verifying an Advanced Theory of Human Language --Explanation of the human faculty for constructing and computing sentence structures on the basis of lexical conceptual features--), Kanda University of International Studies, 2001, pp.309-318 [in Japanese].

[7] Yasuo Miyoshi, Ryo Okamoto, **A Sample-Based Interactive Support System for English Writing Learning**, The 3rd Young Researchers Forum of JSiSE(Japanese Society for Information and Systems in Education)-W, 2000 [in Japanese].

[8] **New Dictionary of English Composition for Scientists**, Newly Revised Edition (EPWING Version on CD-ROM), 2015, Ogura Shoten [in Japanese].

[9] Wang Yiou, Ikeda Takashi, **A solution for the problem of Existential Expressions in Japanese-Chinese Machine Translation**, 2007, Journal of Natural Language Processing, Vol. 14 (2007) No. 5, pp.65-105 [in Japanese].

[10] Satoru Ikehara, Satsuki Abe, Masato Tokuhisa, Jin'ichi Murakami, **Japanese to English Sentence Pattern Generations for Semantically Non-Linear Complex Sentences**, 2004, IPSJ SIG Technical Report, pp.49-56 [in Japanese].

[11] Masanori Taniguchi, Chenchen Ding, Mikio Yamamoto, **An Improvement in Quantifiers Movement of Head Finalization Reordering for E-J Machine Translation**, 2014, Forum on Information Technology, pp.251-252 [in Japanese].

[12] Tsubasa Amano, Takayuki Watabe, Shosaku Tanaka, Yoshinori Miyazaki, **An Approach to Simplify Retrieved English Sentences in Consideration of Their Structures**, 2015, Forum on Information Technology, pp.211-214 [in Japanese].

[13] Shigeki Matsubara, Yuta Sakai, Shunsuke Kozawa, Kenji Sugiki, **Automatic extraction of English useful expressions from scientific papers**, INFOPRO2010, 2010, pp.41-44 [in Japanese].

[14] Fujii, A., Utiyama, M., Yamamoto, M. and Utsuro, T., **Overview of the Patent Translation Task at the NTCIR-7 Workshop**, Proceedings of NTCIR-7 Workshop Meeting, http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings7/pdf/NTCIR7/C3/PATMT/01-NTCIR7-OV-PATMT-FujiiA.pdf, 2008.

[15] Yuji Matsumoto, **ChaSen--Morphological Analyzer**, http://chasen-legacy.sourceforge.jp/, 2015 [in Japanese].

[16] Yuji Matsumoto, **CaboCha: Yet Another Japanese Dependency Structure Analyzer**, http://taku910.github.io/cabocha/, 2015 [in Japanese]

[17] **Linguistic Data Annotation Specification**, Assessment of Fluency and Adequacy in Chinese-English Translations Revision 1.0, 2002, pp. 2-3, https://catalog.ldc.upenn.edu/docs/LDC2003T17/TransAssess02.pdf.

[18] Takeshi Abekawa and Kyo Kageura. **A translation aid system with a stratified lookup interface**, In ACL Demos and Posters, pp. 5–8, 2007.

Table 7: Calculation results of the cost of correspondence between phrases (example)

Cost of correspondence between phrases							0	1	2	3	4	5	6	Dependent Phrase No
								2	4	4	5	6	-1	Head Phrase No.
								4	3	3	2	1	0	Depth
							ダミー (Dummy)	溶存オゾン濃度 (The dissolved ozone density)	高けれ (The higher)	注入時間 (The injection time)	短くする (The shorter…be made)	こと There is no English equivalent	できる (Might)	Heading Word
								N	V	N	NV	N	NV	Heading Feature
								が (Ga)	ば (Ba)	を (Wo)		が (Ga)	。	Depending Word
								y	y	y	t	y	s	Depending Feature
0			ダミー (Dummy)				0.000	1.000	1.000	1.000	1.000	1.000	1.000	
1	2	4	これ (This)	N	に(Ni)	y	1.000	0.038	0.063	0.063	0.063	0.063	0.063	
2	5	3	より (Way)	V	、	y	1.000	0.063	0.075	0.100	0.125	0.125	0.125	
3	5	3	雌コンタクト長 (The female contact)	N	を(Wo)	y	1.000	0.063	0.100	0.025	0.125	0.125	0.125	
4	5	3	さらに (Even)	E		y	1.000	0.063	0.100	0.100	0.100	0.125	0.125	
5	6	2	短くする (Made shorter)	NV		t	1.000	0.063	0.125	0.125	0.000	0.250	0.500	
6	7	1	こと There is no English equivalent expression.	N	が(Ga)	y	1.000	0.063	0.125	0.125	0.250	0.000	0.500	
7	-1	0	できる (can)	NV	。	s	1.000	0.063	0.125	0.125	0.250	0.500	0.000	
Dependent Phrase No.	Head Phrase No.	Depth	Heading Word	Heading Feature	Depending Word	Depending Feature								

Table 8: Calculation results of the accumulated cost of correspondence (example)

Accumulated cost of correspondence							0	1	2	3	4	5	6	Dependent Phrase No
								2	4	4	5	6	-1	Head Phrase No.
								4	3	3	2	1	0	Depth
							ダミー (Dummy)	溶存オゾン濃度 (The dissolved ozone density)	高けれ (The higher)	注入時間 (The injection time)	短くする (The shorter…be made)	There is no English equivalent こと	できる (Might)	Heading Word
								N	V	N	NV	N	NV	Heading Feature
								が (Ga)	ば (Ba)	を (Wo)		が (Ga)	。	Depending Word
								y	y	y	t	y	s	Depending Feature
0			ダミー (Dummy)				0.000	1.000	2.000	3.000	4.000	5.000	6.000	
1	2	4	これ (This)	N	に(Ni)	y	1.000	0.038	0.100	0.163	0.225	0.288	0.350	
2	5	3	より (Way)	V	、	y	2.000	0.100	0.113	0.200	0.325	0.450	0.575	
3	5	3	雌コンタクト長 (The female contact)	N	を(Wo)	y	3.000	0.163	0.200	0.138	0.263	0.388	0.513	
4	5	3	さらに (Even)	E		y	4.000	0.225	0.263	0.263	0.238	0.363	0.488	
5	6	2	短くする (Made shorter)	NV		t	5.000	0.288	0.388	0.388	0.263	0.513	0.763	
6	7	1	こと There is no English equivalent expression.	N	が(Ga)	y	6.000	0.350	0.513	0.513	0.513	0.263	0.763	
7	-1	0	できる (can)	NV	。	s	7.000	0.413	0.638	0.638	0.763	0.763	0.263	
Dependent Phrase No.	Head Phrase No.	Depth	Heading Word	Heading Feature	Depending Word	Depending Feature								

* Shaded cells are indicating correspondence between phrases at the shortest path.

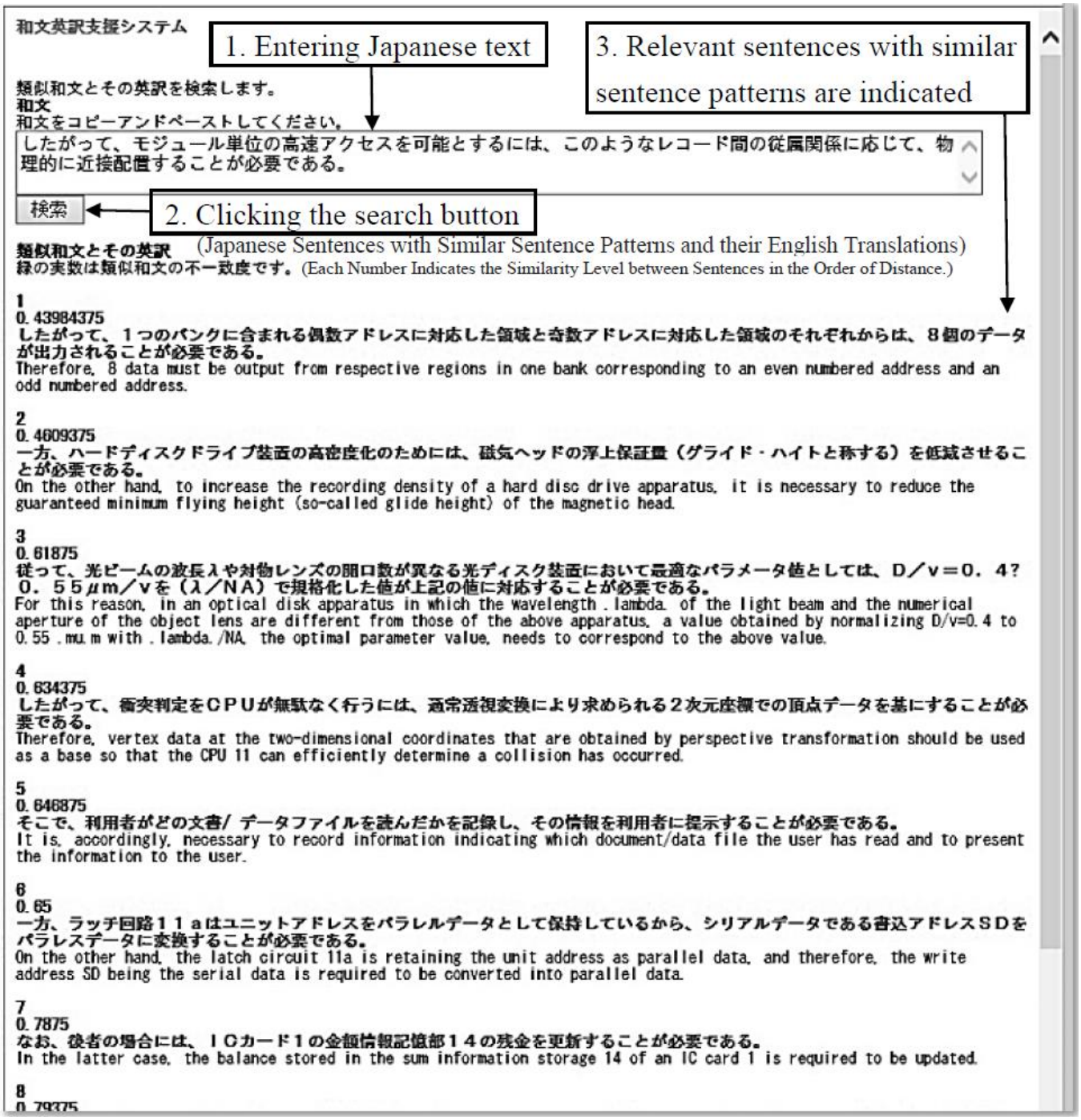

Figure 6: User interface of the support system for English composition