Dear Dr. Babak Hamidzadeh:

Enclosed please find a postscript file of our paper titled "Interactive Generalization of a Translation Example Using Queries Based on a Semantic Hierarchy", which is intended for submission to ICTAI–2000.

The information on the contact author etc is as follows:

1) Contact Author: Yausuhiro Akiba

Email: akiba@cslab.kecl.ntt.co.jp

Postal mail: NTT Communication Science Laboratories.

2-4 Hikaridai Seika-cho Souraku-gun

Kyoto 619-0237, Japan.

2) Appropriate subject areas: Natural Language Processing, Machine Translation, Knowledge Acquisition, Translation Knowledge, Interactive System.

Thank you for your attention.

Best regards,

Yasuhiro Akiba

Interactive Generalization of a Translation Example Using Queries Based on a Semantic Hierarchy

Yasuhiro Akiba Hiromi Nakaiwa Satoshi Shirai Yoshifumi Ooyama

NTT Communication Science Labs.

2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237, JAPAN

{akiba,nakaiwa,shirai,ooyama}@cslab.kecl.ntt.co.jp

Abstract

This paper addresses the issue of acquiring translation rules for machine translation (MT) systems that adopt a transfer approach. These rules are semantic pattern pairs (SPPs) of source and target languages. Practical MT systems must additionally contain a huge number of SPPs corresponding to rarely-used predicates and predicate usages. Such SPPs are difficult to automatically acquire with corpus-based methods [1, 9, 10]. To solve this difficulty, this paper proposes a method to acquire SPPs by using queries based on a semantic hierarchy. The proposed method asks the lexicographer for the necessary information in order to generalize the conditions of SPPs and then gradually generalizes these conditions. Experimental results show that the proposed method allows the acquisition of more plausible conditions within almost the same time spent for manual generalization.

Subject Areas: Natural Language Processing, Machine Translation, Knowledge Acquisition, Translation Knowledge, Interactive System.

1 Introduction

The World Wide Web (WWW) has become very popular throughout the world. Due to this popularity, more and more people have come to realize that the opportunity is growing for them to read documents written in unfamiliar languages. Accordingly, there are higher hopes for more efficient machine translation (MT) systems.

One attempt to achieve MT systems is the use of a semantic-based transfer approach [5, 3, 11]. This approach focuses on the collocation between verbs and nouns. Semantic-based transfer MT systems provide if-then rules (called **semantic pattern pairs** or **SPPs**), whose if-parts and then-parts are collocation patterns in the source language (SL) and the target language (TL), respectively.

Figure 1 shows a simplified example [4] of an SPP in a Japanese-English MT system, ALT-J/E [5]. The if-part is a Japanese pattern, and the then-part is an English equivalent. This SPP indicates that if the Japanese verb is *yaku*, its sub-

IF	J-Verb	$=$ *yaku*	THEN	E-Subj	$= N_1$
	J-Subj	$= N_1$:[`People`]		E-Verb	$=$ *bake*
	J-Obj	$= N_2$:[`Bread` or `Confectionery`]		E-Obj	$= N_2$

Figure 1: SPP for Japanese verb *yaku*.

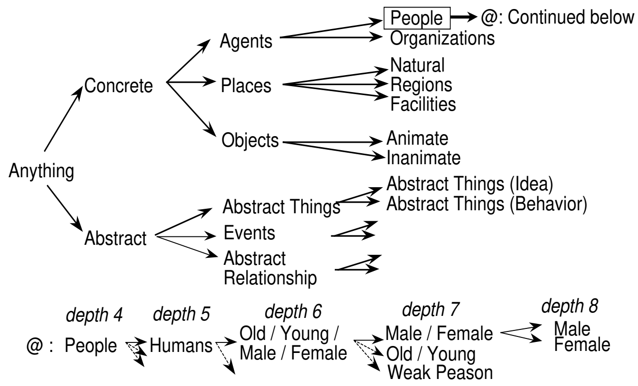

Figure 2: Upper Levels of the Semantic Hierarchy in ALT-J/E.

jective noun is categorized as `People` and its objective noun is categorized as `Bread` or `Confectionery`; accordingly, the corresponding English verb is *bake* and its subject and object are the translation into English of the Japanese subject and object, respectively. Each slot such as J-SUBJ or J-OBJ contains the meanings of nouns (**semantic categories**) like `People`, `Bread`, or `Confectionery`. [1]

ALT-J/E has about 2,800 semantic categories [4], which constitute a hierarchy with a maximum depth of 12 (Figure 2). Each of about 400,000 nouns in the ALT-J/E lexicon has one or more semantic categories as its meanings.

Practical MT systems based on a semantic-based transfer approach must addi-

[1] Some nouns like *Zeri* "jelly" are categorized as `Confectionery`; consequently, semantically strange sentences as *watasi-wa zeri-wo yaku* "I bake jelly" can meet the SPP shown in Figure 1. It does not matter that sentences like this meet an SPP since they are not inputted into MT systems.

tionally contain a huge number of SPPs corresponding to rarely-used predicates and predicate usages. In the case of ALT-J/E , Shirai et al. [8] reported that about 10,000 SPPs, corresponding to rarely-used predicates and predicate usages, have to be generated for ALT-J/E to cover nearly all of the predicates in Japanese.

The existing approaches to acquire SPPs include corpus-based approaches [1, 9, 10]. As Church indicates in [2], many possible co-occurrences cannot be observed even in a very large corpus. That is, a sufficient number of illustrative sentences cannot be prepared for the above kinds of SPPs. These approaches are, therefore, of limited applicability.

To date, the above kinds of SPPs have been acquired by manual approaches like those in [8]. The task of generating SPPs involves inputting the appropriate (not over-general and not over-specific) semantic categories into each slot. It is, however, difficult for lexicographers to manually generate SPPs due to the following reasons.

Large number of candidates: Searching for appropriate categories requires a lot of effort on the part of lexicographers. This is because the number of candidates for appropriate categories ranges from several thousands [4] to several tens of thousands [6], depending on the semantic hierarchy used.

Necessity of experience: To specify appropriate semantic categories, lexicographers must be very familiar with all of the semantic categories and the lexicon on an MT system.

Different quality of generalization: Lexicographers treat minor translation examples differently according to their own generalization standards. Some lexicographers specify more general semantic categories so as to be met by minor examples. Other lexicographers ignore minor examples.

To overcome the difficulty of generating SPPs and the sheer quantity of SPPs, methods to support lexicographers are required.

This paper proposes a new method that searches for appropriate semantic categories to be inputted into each slot by using queries based on a semantic hierarchy. The proposed method adopts three different approaches to searching for appropriate semantic categories. The approaches for the searching generate translation examples at each search point and ask whether the noun corresponding to the target slot can be semantically collocated with the verb of the SL side in the acquired SPP and whether that noun can also give a context where the most plausible equivalent for the verb of the SL side is the verb of the TL side. The only task that lexicographers have to do is to answer each query. Consequently, the resulting categories follow a specific generalization standard.

The authors experimentally evaluated the proposed method by acquiring the semantic categories for ten SPPs [4] with ALT-J/E . Experimental results showed that the proposed method was able to acquire, in five cases, the same categories as, or in the other cases, more plausible categories than, those specified manually within almost the same time spent for the manual generalization.

The next section describes the acquisition task in this study. The proposed method is presented in Section 3. Experimental results are shown and a discussion is provided in Section 4. Finally, our conclusions are presented in Section 5.

2 Acquisition Task

This section characterizes the appropriate semantic categories that the proposed method should input into each slot in SPPs and describes the acquisition task that the proposed method handles.

Let us assume that lexicographers attempt to acquire the appropriate semantic categories for the slots of the SPP shown in Figure 1 while they derive (1) as a sample sentence that meets the conditions of the acquired SPP.

(1) Taro-ga appurupai-wo yaku
taro-SUBJ apple pies-OBJ bake-PRESENT
'Taro bakes apple pies'

where the nouns such as *Taro* "a typical first name of Japanese males" for the J-SUBJ slot or *appurupai* "apple pies" for the J-OBJ slot are called **the sample nouns**, and the semantic categories of the sample nouns are called **the sample categories**. Let the sample categories of *Taro* and *appurupai* be `Male` and `Confectionery`, respectively.

When specifying different semantic categories from sample categories into slots so that we can investigate what kinds of linguistic phenomena can be observed, let us replace one of the sample nouns with another noun as follows.

(2) Meari-ga appurupai-wo yaku
mary-SUBJ apple pies-OBJ bake-PRESENT
'Mary bakes apple pies'

(3) Mangetsu-ga appurupai-wo yaku
a full moon-SUBJ apple pies-OBJ bake-PRESENT
'A full moon bakes apple pies'

(4) *Taro-ga sake-wo yaku*
taro-SUBJ a salmon-OBJ grill-PRESENT
'Taro grills a salmon'

In the case of (2), the sentence is natural as a Japanese sentence, so the substituted noun *Meari* "Mary" can be semantically collocated with *Yaku* "bake". In the case of (3), the description of the sentence *Mangetsu-ga appurupai-wo yaku* is not realistic and cannot happen normally. In such a sense, *Mangetsu* "a full moon" cannot be semantically collocated with *yaku.* In the case of (4), the sentence is natural as a Japanese sentence so the substituted noun *sake* "a salmon" can be semantically collocated with *Yaku.* On the other hand, the most plausible equivalent for Japanese verb *yaku* is *grill* rather than *bake.* This is because, although the source sentence does not describe definite clues in order to select the English equivalent, the most plausible situation of the source is *Taro grills a salmon* as long as one takes account of the culture of SL; in this case, Japanese culture.

As seen above, the following linguistic phenomena can be observed through the replacement of nouns,

(P1) Some nouns (or semantic categories) can be semantically collocated with the verb of the SL side in the acquired SPP.

(P2) Out of all of the semantically collocated nouns (or semantic categories), some nouns (or semantic categories) can give a context where the most plausible equivalent for the verb of the SL side is the target verb of the TL side.

The phenomena is a good indicator for finding the appropriate categories.

As a semantic category specified to a slot processed by lexicographers (**the target slot**) can change from the sample category for the target slot to the root of the semantic hierarchy step by step, the number of nouns able to meet an acquired SPP can increase gradually. When C_i denotes the i-th semantic category on the path from the root of the semantic hierarchy to the sample category for the target slot (Figure 3), and when C_i is specified to the target slot instead of C_{i+1}, the ratio of the collocated nouns in (P1), to the additionally covered nouns, is called **the acceptable rate** of C_i, which is related to the occurrence probabilities of the SL sentences generated by the replacement. The ratio of the nouns that can give the context in (P2), to all collocated nouns in (P1), is called **the translatable rate** of

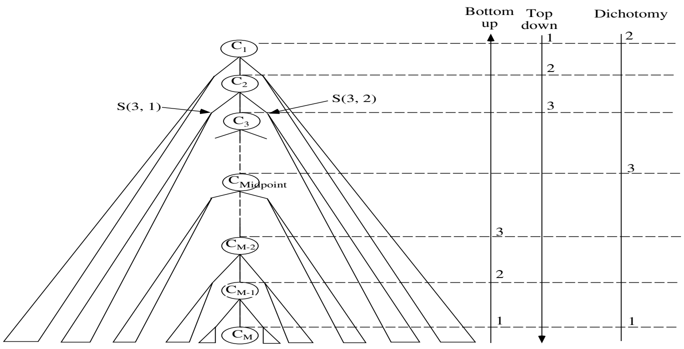

Figure 3: Proposed method: M denotes the depth of the sample category for the target slot.

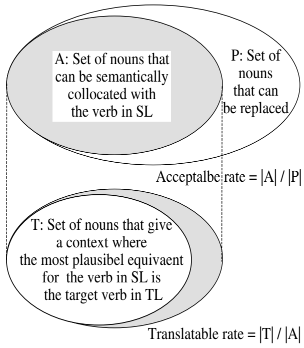

Figure 4: Acceptable rate and translatable rate of C_i

C_i (Figure 4), which is related to the probabilities that such substituted nouns can give the context in (P2). Because both rates of the appropriate semantic categories specified to the target slot should be sufficiently large, let us focus on the lower thresholds of both rates and call them **the minimal acceptable rate** and **the minimal translatable rate**, respectively. A semantic category whose acceptable rate is the minimal acceptable rate or greater and whose translatable rate is the minimal translatable rate or greater is called an **ok-category**. On the other hand, a category that is not an ok-category is called an **ng-category**.

This paper characterizes each appropriate semantic category as an ok-category that is located at the highest level on the semantic hierarchy when the minimal acceptable rate and the minimal translatable rate are given.

Consequently, given the minimal acceptable rate and the minimal translatable

Input: (A) the skeleton of a sentence that meets the acquired SPP,
(B) a combination of nouns such that, for each slot in the if-part of the SPP, one of the nouns should meet the condition of the slot (**the sample nouns**),
(C) the semantic category for each sample noun (**the sample categories**).

Output: the highest ok-category in the semantic categories on the path between the root in the semantic hierarchy and the sample category for the target slot.

Figure 5: Acquisition task

rate, the acquisition task that the proposed method handles is to search for the highest ok-category for the target slot of the SPP acquired by the method (Figure 5).

For example, to make the proposed method acquire the appropriate semantic category for the J-SUBJ slot of the SPP shown in Figure 1, the inputted skeleton of the sentence that meets the acquired SPP is N_1-*ga* N_2-*wo yaku* "N_1 bake N_2", where N_i is a variable. *Taro* and `Male` are inputted as the sample noun for the J-SUBJ slot and the semantic category, respectively. At the same time, *appurupai* and `Confectionery` are inputted as the sample noun for the J-OBJ slot and the semantic category, respectively. Lexicographers can easily input this information. The output is the highest ok-category in the eight semantic categories on the path between the root in Figure 2, `Anything` (C_1), and the leaf, `Male` (C_8).

3 Proposed Method

This section describes the proposed method that adopts three approaches to search for the highest ok-category described in Section 2. The approaches use the same strategy for generating queries and presenting them to lexicographers in order to estimate the acceptable rate and the translatable rate of the current search point.

3.1 Generating Sentences for Queries

When the current search point is C_i, the approaches for the searching generate sentences in the following way: (i) initially generate a sentence by filling each variable N_i in the skeleton with the corresponding sample noun; (ii) then, generate some sentences by replacing the sample noun for the target slot with other nouns in $\mathbf{Cluster}_i$, which hereafter denotes the set of nouns categorized as C_i or descendants of C_i but not categorized as C_{i+1} or descendants of C_{i+1}. For example, assume that, in order to acquire the SPP shown in Figure 1, the input to the proposed method is the same as presented at the end of Section 2: N_1-*ga* N_2-*wo yaku* as the skeleton, *Taro* and *appurupai* as the sample nouns, and `Male` and `Confectionery` as the semantic categories; that the target slot is the J-SUBJ one; and that the current search point is C_2 (`Concrete` in Figure 2), then the substituted nouns are categorized as either `Palaces` or `Object` or descendants of them.

The approaches use the generated sentences for estimating the acceptable rate and the translatable rate of the current search point C_i. The main issues are that, in order to estimate both rates within significantly small errors by using only the limited number of generated sentences, how do the approaches select the nouns substituted for the sample noun and how do the approaches estimate both rates.

To resolve the two issues, the approaches employ stratified sampling [7], a sampling survey technique in statistics, in the following. $\mathbf{Cluster}_i$ is separated into subsets of nouns, from which some substituted nouns are collected, where the number of substituted nouns from each subset is decided according to the total number of nouns in the subset as will be seen later. Then, the acceptable rate of C_i is estimated as the weighted average of the acceptable rates for the subsets, where the weight for each subset is the number of substituted nouns. For example, assume that C_i is separated into two subsets, that the ratio of the substituted nouns for the subsets is 3:1, and that the acceptable rates for the subsets are 100% and 50%, respectively. Then, the acceptable rate of C_i is estimated as $(100*3+50*1)/(3+1)$. The translatable rate of C_i is also estimated in the same way.

Stratified sampling does not provide a way to separate $\mathbf{Cluster}_i$. the proposed method, therefore, adopts an original technique separating $\mathbf{Cluster}_i$ into subsets, which is decided step by step as follows. For example, when the total number of substituted nouns in $\mathbf{Cluster}_2$ is 40 under the same input as seen at the beginning of this section, at first, allocate the number of substituted nouns (**the sample size**) in $\mathbf{Cluster}_2$ among all of the siblings of C_3 (`Agents` in Figure 2) according to the ratio of the number[2] of leave that are descendants of each sibling of C_3, for example, $3:1$. In this example, the sample sizes for `Palaces` and `Object` become 30 and 10,

[2]These numbers are used as convenient indicators of semantic diversity under each sibling.

respectively. For each sibling, allocate the sample size of the sibling among all of the children in the same way, until the sample size is too small, for example, equal to or less than 15. If the ratio of the number of leaf-level descendants of `Natural`, `Regions` and `Facilities` is $2:1:3$, then the sample sizes [3] for `Natural`, `Regions`, and `Facilities` become $10, 5,$ and 15. Since the sample size of `Object` is less than 15, the sample size of `Object` is not allocated among the children: `Animate` and `Inanimate`. After this, for each child, allocate the sample size of the child among all of the children of the child recursively, until the sample size is too small. Let $S(i,j), (1 \leq j \leq L_i)$, hereafter, denote the semantic categories whose sample sizes are not allocated to their children. In the example, $L_2 = 4$ and $S(2,j)(1 \leq j \leq L_2)$ correspond to `Natural`, `Regions`, `Facilities`, and `Object`.

$S(i,j), (1 \leq j \leq L_i)$ are used as the subsets used by the stratified sampling. The substituted nouns are selected from $S(i,j)$ or descendants of $S(i,j)$ in the order of frequency in use. The number of selected nouns is the sample size for $S(i,j)$.

3.2 Presenting Queries and Getting Answers

For each $S(i,j), (1 \leq j \leq L_i)$, the approaches for searching simultaneously present the generated sentences to lexicographers. This simultaneous presentation prevents the lexicographers from misunderstanding the meanings of the substituted nouns.

[3] When the sample size becomes a decimal as $13.\dot{3}$ $(= 20 * 2/3)$, the decimal is rounded up.

Since all of the substituted nouns in the presented sentences are categorized in a certain semantic category, the lexicographers can easily guess the correct meaning of a substituted noun.

The lexicographers judge whether each substituted noun can be semantically collocated with the verb of the SL side in the acquired SPP (Q1). If and only if this answer is positive, they also judge whether the substituted noun can give a context where the most plausible equivalent for the verb of the SL side is the verb of the TL side (Q2). The lexicographers make their determinations by answering queries: (Q1) and (Q2). For example, as explained in detail in Section 2, they answer positive to both (Q1) and (Q2) for (2) and negative to (Q1) for (3). Moreover, they answer positive to (Q1) and negative to (Q2) for (4).

3.3 Approaches for searching for an appropriate category

The three approaches for searching differ in the order in which they search for an appropriate semantic category (Figure 3). The first two approaches, the Bottom-up approach and the Top-down approach, are the same as a linear search. The last approach, the Dichotomy approach, is the same as a binary search. For convenience in the following explanation, let us define that $L_M = 1$ and $S(M, 1) = C_M$.

The Bottom-up approach applies the above query strategy to each semantic category in reverse order of depth, $C_M, C_{M-1}, \cdots$. When an ng-category is found, this approach stops searching and outputs the latest ok-category.

The Top-down approach applies the above query strategy to each semantic category in order of depth, $C_1, C_2, \cdots$. When an ok-category is found, this approach stops searching and outputs that ok-category.

The Dichotomy approach initially applies the above query strategy to the leaf and the root in order. Next, this approach prepares a candidate list, $(C_1,\ C_2,\ \cdots,\ C_i,\ \cdots,\ C_M)$ and applies the above query strategy to the semantic category in the middle of the candidate list or to the lower semantic category closest to the middle if a precisely central semantic category does not exist. According to whether the semantic category is an ok-category or not, the approach revises the candidate list in the same way as a binary search does. Then the above procedure by using the updated candidate list is repeated. Consequently, the first and last elements of the candidate list are always an ng-category and an ok-category, respectively, after the root and leaf semantic categories are processed. When the length of the candidate list becomes 2, the approach stops searching and outputs the last element.

As mentioned above, the only task that lexicographers have to do is to answer each query. Consequently, the resulting categories follow a specific generalization standard. After repeatedly applying one of the approaches for searching to each corresponding input, the proposed method can specify all of the appropriate semantic categories for each slot in the SPP.

Table 1: Skeletons and sample sentences for interactive generalization

No	Skeleton Sample Sentence	Category specified manually
1	*N1-ga N2-wo yomu* "N1 read N2" *Chichi-ga hon-wo yomu* "My father reads a book"	`Agents`
2	*N1-ga N2-wo yomu* "N1 read N2" *Chichi-ga hon-wo yomu* "My father reads a book"	`Abstract Thing (Idea)`
3	*N1-ga N2-wo yomu* "N1 read N2" *Chichi-ga houkokusho-wo yomu* "My father reads a report"	`Spirit/Soul/Mind`
4	*N1-ga N2-wo N3-ni erabu* "N1 elect N2 N3" *Juhmin-ga kare-wo kaichou-ni erabu* "Residents elect him their head"	`Chief/President/ Manager`
5	*N1-ga N2-de nyushou-suru* "N1 win a prize in N2" *Kare-ga Konkuhru-de nyushou-suru* "He wins a prize in the contest"	`Abstract Thing (Behavior)`
6	*N1-ga N2-wo tatamu* "N1 close N2" *Chichi-ga mise-wo tatamu* "My father closes his shop"	`Facilities`
7	*N2-ga N2-wo unten-suru* "N1 run N2" *Chichi-ga hatsudouki-wo unten-suru* "My father runs an electric dynamo"	`Machinery`
8	*N1-ga N2-wo nageru* "N1 throw N2" *Chichi-ga bohru-wo nageru* "MY father throws a ball"	`Objects`
9	*N1-ga hanpatsu-suru* "N1 rebound" *Kabusiki-ga hanpatsu-suru* "Shares rebound"	`Economic system`
10	*N1-ga N2-ni tassuru* "N1 rise to N2" *Doru-ga saitakane-ni tassuru* "The dollar rises to the highest level"	`Price/Cost`

4 Experimental Work

The authors evaluated the proposed method on the three following points.

- How different are the semantic categories acquired by the proposed method from those manually specified by lexicographers?
- Are the semantic categories acquired by the proposed method better than those manually specified by lexicographers?
- How is the proposed method able to contribute to the acquisition of SPPs?

Table 2: Experimental Results: the relative position of the acquired semantic category to that specified manually, the number of paired queries: (Q1) and (Q2) (See Section 3.2) and the time spent

No	Difference			# of paired Queries			Time (M.)		
	B	T	D	B	T	D	B	T	D
1	0	0	0	259	53	53	23	6	5
2	0	0	0	158	155	155	17	16	18
3	−1	+3	+3	119	51	51	24	10	5
4	0	0	0	55	262	158	9	34	15
5	−2	−2	−2	153	259	204	8	21	11
6	−1	−1	−1	147	208	157	10	13	12
7	−1	−1	−1	104	310	208	10	54	8
8	−2	−2	−2	208	207	156	55	75	13
9	0	0	0	158	212	160	7	12	8
10	0	0	0	50	316	155	16	4	10
Ave.	—	—	—	**141.1**	**203.3**	**145.7**	**17.9**	**24.5**	**10.5**

In order to evaluate the above points, the authors attempted to acquire semantic categories for SPPs whose if-parts corresponded to the skeletons in Table 1. The sample nouns inputted to the proposed method are shown under the skeletons. The target of the generalization is indicated by underline. The semantic hierarchy used was that of ALT-J/E as shown in Figure 2. The minimal acceptable rate and the minimal translatable rate were fixed at 2% and 80%, respectively. The sample size for C_i and the lower threshold of the sample sizes were 50 and 10, respectively. Two lexicographers, who generated SPPs of ALT-J/E , participated in the experiments. Each semantic category in the 3rd column indicates a semantic category specified manually for the SPP corresponding to the skeleton in the 2nd column.

Table 2 reports experimental results. The 2nd to 4th columns show the relative position of an acquired semantic category in comparison to a semantic category manually specified. For example, $+1$ or -1 indicates that the acquired semantic category is one semantic category above the semantic category manually specified or below it, respectively. Each number in the 5th to 7th columns shows the number of paired queries, i.e., (Q1) and (Q2), presented to the lexicographers. Each number in the 8th to 10th columns shows the time spent for generalization of the sample noun to the target slot. B, T, and D on the 2nd line indicate that the approach for the searching is, Bottom-up, Top-down, and Dichotomy, respectively. Through this experiment, the following things could be found:

- Although the proposed method acquired different semantic categories in five rows out of ten, the required semantic categories whose relative positions were one or two below were judged by the lexicographers to be better than those specified manually. This is because manual specification to a translation example in their minds is very sensitive, even if the example belongs to a minority of translation examples; accordingly, semantic categories specified manually tend to become over general.

- The above consideration tells us that the Bottom-up approach is the best approach for searching. The average time spent was 17.9 minute. On the other hand, the lexicographers reported that the time spent by manual specification was 15 to 20 minutes on average. The efficiency of the proposed method is, therefore, almost equal to that of manual specification as far as one compares the time spent. Note that, although the trials using each approach were done in the following order: Top-down, Bottom-up, and Dichotomy, the ratio of the number of paired queries between Top-down and Bottom-up was almost equal to that of the time spent between them. The lexicographers, therefore, did not remember the Q&A in the previous trial: the Top-down approach, at lease when they moved to the next trial: the Bottom-up approach.

5 Conclusion

This paper proposed a method to acquire appropriate semantic categories to be inputted into each slot of an SPP by using queries based on a semantic hierarchy. The queries ask whether the noun corresponding to the target slot in presented sentences can be semantically collocated with the verb in SL and ask whether the noun can also give a context where the most plausible equivalent for the verb in SL is the verb of the TL side in the acquired SPP.

The method allows lexicographers to acquire more plausible semantic categories for SPPs by simply answering the queries presented by the method.

References

[1] H. Almuallim, Y. Akiba, and S. Kaneda. On handing tree-structured attributes in decision tree learning. In *Proc. ICML–95*, pages 12–20, 1995.

[2] K. W. Church and R. L. Mercer. Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, 19:1–24, 1993.

[3] M. Dorna and M. C. Emele. Efficient implementation of a semantic-based transfer approach. In *Proc. ECAI–96*, pages 567–571, 1996.

[4] S. Ikehara, M. Miyazaki, S. Shirai, A. Yokoo, H. Nakaiwa, K. Ogura, Y. Ooyama, and Y. Hayashi. *Goi-Taikei: A Japanese Lexicon (in Japanese)*. Iwanami Shoten Publisher, Tokyo, 1997.

[5] S. Ikehara, S. Shirai, A. Yokoo, and H. Nakaiwa. Toward an MT system without pre-editing-effects of new methods in ALT-J/E. In *Proc. MT Summit-3*, pages 101–106, 1991.

[6] G. A. Miller. Wordnet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

[7] R. L. Scheaffer, W. I. Mendenhall, and R. L. Ott. *Elementary Survey Sampling (5th ed.)*. Duxbury Press, California, 1996.

[8] S. Shirai, S. Ikehara, A. Yokoo, and H. Inoue. The quantity of valency pattern pairs required for Japanese to English MT and their compilation. In *Proc. NLPRS–95*, pages 443–448, 1995.

[9] H. Tanaka. Decision tree learning algorithm with structured application to verbal case-frame acquisition. In *Proc. Coling–96*, pages 943–948, 1996.

[10] T. Utsuro. Sense classification of verbal polysemy based on bilingual class/class association. In *Proc. COLING–96*, pages 968–973, 1996.

[11] J. Yang. Towards the automatic acquisition of lexical selection rules. In *Proc. MT Summit-7*, pages 397–403, 1999.