Sense Classification of Verbal Polysemy based-on Bilingual Class/Class Association*

Takehito Utsuro
Graduate School of Information Science, Nara Institute of Science and Technology
8916-5, Takayama-cho, Ikoma-shi, Nara, 630-01, JAPAN
utsuro@is.aist-nara.ac.jp

Abstract

In the field of statistical analysis of natural language data, the measure of word/class association has proved to be quite useful for discovering a meaningful sense cluster in an arbitrary level of the thesaurus. In this paper, we apply its idea to the sense classification of Japanese verbal polysemy in case frame acquisition from Japanese-English parallel corpora. Measures of bilingual class/class association and bilingual class/frame association are introduced and used for discovering sense clusters in the sense distribution of English predicates and Japanese case element nouns. In a small experiment, 93.3% of the discovered clusters are correct in that none of them contains examples of more than one hand-classified senses.

1 Introduction

In corpus-based NLP, acquisition of lexical knowledge has become one of the major research topics. Among several research topics in this field, acquisition from parallel corpora is quite attractive (e.g. Dagan et al. (1991)). The reason is that parallel sentences are useful for resolving both syntactic and lexical ambiguities in the monolingual sentences. Especially if the two languages have different syntactic structures and word meanings (such as English and Japanese), this approach has proved to be most effective in disambiguation (Matsumoto et al., 1993; Utsuro et al., 1993).

Utsuro et al. (1993) proposed a method for acquiring surface case frames of Japanese verbs from Japanese-English parallel corpora. In this method, translated English verbs and case labels are used to classify senses of Japanese polysemous verbs. Clues to sense classification are found using English verbs and case labels, as well as the sense distribution of the Japanese case element nouns. Then, a human instructor judges whether the clues are correct. One of the major disadvantages of the method is that the use of English information and sense distribution of Japanese case element nouns is restricted. Only surface forms of English verbs and case labels are used and sense distribution of English verbs is not used. Also, the threshold of deciding a distinction in the sense distribution of Japanese case element nouns is predetermined on a fixed level in a Japanese thesaurus. As a result, the human instructor is frequently asked to judge the correctness of the clue.

In the field of statistical analysis of natural language data, it is common to use measures of lexical association, such as the information-theoretic measure of mutual information, to extract useful relationships between words (e.g. Church and Hanks (1990)). Lexical association has its limits, however, since often either the data is insufficient to provide reliable word/word correspondences, or the task requires more abstraction than word/word correspondences permit. Thus, Resnik (1992) proposed a useful measure of word/class association by generalizing information-theoretic measure of word/word association. The proposed measure addresses the limitations of lexical association by facilitating statistical discovery of facts involving word classes rather than individual words.

We find the measure of word/class association of Resnik (1992) is quite attractive, since it is possible to discover a meaningful sense cluster in an arbitrary level of the thesaurus. We thus expect that the restrictions of the previous method of Utsuro et al. (1993) can be overcome by employing the idea of the measure of word/class association. In this paper, we describe how this idea can be applied to the sense classification of Japanese verbal polysemy in case frame acquisition from Japanese-English parallel corpora. First, sense distribution of English predicates and Japanese case element nouns is represented using monolingual English and Japanese thesaurus, respectively (sections 2 and 3). Then, the measure of the association of classes of English predicates and Japanese case element nouns, i.e., a measure of bilingual class/class association, is introduced, and extended into a measure of bilingual class/frame association (section 4).

*The author would like to thank Prof. Yuji MATSUMOTO for his valuable comments on this research. This work is partly supported by the Grants from the Ministry of Education, Science, and Culture, Japan, #07780326.
Using these measures, sense clusters are discovered in the sense distribution of English predicates and Japanese case element nouns. Finally, examples of a Japanese polysemous verb collected from Japanese-English parallel corpora are divided into disjoint clusters according to those discovered sense clusters (section 5). The results of a small experiment are presented and the proposed measure is evaluated (section 6).

2 Bilingual Surface Case Structure

In the framework of verbal case frame acquisition from parallel corpora, bilingually matched surface case structures (Matsumoto et al., 1993) are collected and surface case frames of Japanese verbs are acquired from the collection. In this paper, each bilingually matched surface case structure is called a bilingual surface case structure, and represented as a feature structure:

\[
\begin{array}{l}
\text{pred}: v_J \\
\text{sem}_B: \text{SEM}_B \\
p_1: \begin{cases} 
\text{pred}: n_{j1} \\
\text{sem}: \text{SEM}_{j1} \\
\vdots \\
\end{cases} \\
p_m: \begin{cases} 
\text{pred}: n_{jm} \\
\text{sem}: \text{SEM}_{jm} \\
\end{cases}
\end{array}
\]

\(v_J\) indicates the verb in the Japanese sentence, \(p_1, \ldots, p_m\) denote the Japanese case markers, and \(n_{j1}, \ldots, n_{jm}\) denote the Japanese case element nouns. When a Japanese noun \(n_{ji}\) has several senses, it may appear in several leaf classes in the Japanese thesaurus. Thus, \(\text{SEM}_{ji}\) is represented as a set of classes in the corresponding English predicate, i.e., a set of classes in the English thesaurus:

\(\text{SEM}_B = \{c_{b1}, \ldots, c_{bk}\}, \text{SEM}_{j1} = \{c_{j1}, \ldots, c_{jI}\}\)

\(c_{b1}, \ldots, c_{bk}\) and \(c_{j1}, \ldots, c_{jI}\) indicate the classes in the English and Japanese thesaurus, respectively.

By structurally matching the Japanese-English parallel sentences in Example 1, the following bilingual surface case structure is obtained:

**Example 1**

J: Watashi-ha uwagi-wo kagi-ni kaketa.

I hung my coat on the hook.

E: I hung my coat on the hook.

**Word/Class Association Score**

The measure of word/class association of Resnik (1992) can be illustrated by the problem of finding the prototypical object classes for verbs. Let \(\mathcal{V}\) and \(\mathcal{N}\) be the sets of all verbs and nouns, respectively. Given a verb \(v \in \mathcal{V}\) and a noun class \(c \in \mathcal{N}\), the joint probability of \(v\) and \(c\) is estimated as

\[
\text{Pr}(v, c) = \frac{\sum_{n \in c} \text{count}(v, n)}{\sum_{c' \in \mathcal{N}} \sum_{v' \in \mathcal{V}} \text{count}(v', n')}
\]

The association score \(A(v, c)\) of a verb \(v\) and a noun class \(c\) is defined as

\[
A(v, c) = \frac{\text{Pr}(v, c)}{\text{Pr}(c)} = \frac{\text{Pr}(v \mid c)}{\text{Pr}(c)}
\]

The association score takes the mutual information between the verb and a noun class, and scales

1. This process corresponds to sense disambiguation by maximizing the association score.

2. Five classes are allocated at the next level from the root node: abstract-relations, agents-of-human-activities, human-activities, products, and natural-objects-and-natural-phenomena.

3. At the next level from the root node, it has six classes: abstract-relations, space, matter, intellect, volition, and affections.
it according to the likelihood that a member of the class will actually appear as the object of the verb. The first term of the conditional probability measures the generality of the association, while the second term of the mutual information measures the co-occurrence of the association.

4.2 Bilingual Class/Class Association Score

We now apply the word/class association score to the task of measuring the association of classes of English predicates and Japanese case element nouns in the collection of bilingual surface case structures. First, we assume that for any polysemous Japanese verb \( v_j \), there exists a case marker \( p \) which is most effective for sense classification of \( v_j \).

Given the collection of bilingual surface case structures for \( v_j \), we introduce the bilingual class/class association score for measuring the association of a class \( c_E \) of English predicates and a class \( c_j \) of Japanese case element nouns for a case marker \( p \).

Let \( E_g(v_j,p) \) be the set of bilingual surface case structures collected from the Japanese-English parallel corpora, each element of which has a Japanese verb \( v_j \) and a Japanese case marker \( p \). Among the elements of \( E_g(v_j,p) \), let \( E_g(v_j,p,c_E) \) be the set of those whose semantic label \( \text{SEM} \_E \) of the English predicate satisfies the class \( c_E \), i.e., \( \text{SEM} \_E \leq c_E \), and \( E_g(v_j,p,c_j) \) be the set of those whose semantic label \( \text{SEM} \_J \) of the Japanese case element noun for the case marker \( p \) satisfies the class \( c_j \), i.e., \( \text{SEM} \_J \leq c_j \). Let \( E_g(v_j,c_E,p,c_j) \) be the intersection of \( E_g(v_j,p,c_E) \) and \( E_g(v_j,p,c_j) \). Then, conditional probabilities \( \Pr(c_E | v_j,p) \), \( \Pr(c_j | v_j,p) \), and \( \Pr(c_E \cap c_j | v_j,p) \) are defined as the ratios of the numbers of the elements of those sets:

\[
\Pr(c_E | v_j,p) = \frac{\Pr(E_g(v_j,p,c_E))}{\Pr(E_g(v_j,p))}
\]
\[
\Pr(c_j | v_j,p) = \frac{\Pr(E_g(v_j,p,c_j))}{\Pr(E_g(v_j,p))}
\]
\[
\Pr(c_E \cap c_j | v_j,p) = \frac{\Pr(E_g(v_j,c_E,p,c_j))}{\Pr(E_g(v_j,p))}
\]

Then, given \( v_j \) and \( p \), the association score \( A(c_E,c_j | v_j,p) \) of \( c_E \) and \( c_j \) is defined as

\[
A(c_E,c_j | v_j,p) = \Pr(c_E \cap c_j | v_j,p) \log \frac{\Pr(c_E \cap c_j | v_j,p)}{\Pr(c_E | v_j,p) \Pr(c_j | v_j,p)}
\]

This definition is slightly different from that of the word/class association score in that it only needs the set \( E_g(v_j,p) \), for a Japanese verb \( v_j \) and a Japanese case marker \( p \), but not the whole Japanese-English parallel corpora. This is because our task is to discover strong association of an English class and a Japanese class in \( E_g(v_j,p) \), rather than in the whole Japanese-English parallel corpora. Besides, as the first term for measuring the generality of the association, we use \( \Pr(c_E \cap c_j | v_j,p) \) instead of \( \Pr(c_j | v_j,p,c_E) \) or \( \Pr(c_E | v_j,p,c_j) \) below:

\[
\Pr(c_j | v_j,p,c_E) = \frac{\Pr(E_g(v_j,c_E,p,c_j))}{\Pr(E_g(v_j,p,c_j))}
\]
\[
\Pr(c_E | v_j,p,c_j) = \frac{\Pr(E_g(v_j,c_E,p,c_j))}{\Pr(E_g(v_j,p,c_j))}
\]

4.3 Bilingual Class/Frame Association Score

In the previous section, we assume that for any polysemous Japanese verb \( v_j \), there exists a case marker \( p \) which is most effective for sense classification of verbal polysemy \( v_j \). However, it can happen that a combination of more than one case marker characterizes a sense of the verbal polysemy \( v_j \). Even if there exists exactly one case marker which is most effective for sense classification, it is necessary to select the most effective case marker automatically by some measure. For example, using some measure, it is desirable to automatically discover the fact that, for the task of sense classification of verbal polysemy, subject nouns are usually most effective for intransitive verbs, while object nouns are usually most effective for transitive verbs.

This section generalizes the previous definition of bilingual class/class association score, and introduces the bilingual class/frame association score. In the new definition, we consider every possible set of pairs of a Japanese case marker \( p \) and a Japanese noun class \( c_j \), instead of predetermined the most effective case marker. The bilingual class/frame association score measures the association of an English class \( c_E \) and a set of pairs of a Japanese case marker \( p \) and a Japanese noun class \( c_j \) marked by \( p \). By searching for a large association score, it becomes possible to find any combination of case markers which characterizes a sense of the verbal polysemy \( v_j \).

4.3.1 Japanese Case-Class Frame

First, we introduce a data structure which represents a set of pairs of Japanese case marker \( p \) and a Japanese noun class \( c_j \) marked by \( p \), and call it Japanese case-class frame. A Japanese case-class frame can be represented as a feature structure:

\[
[ p_1 : c_{j_1} \]
\[
\vdots
\]
\[
p_m : c_{j_m}
\]

\( ^4 \text{Pr}(c_j | v_j,p,c_E) \) and \( \Pr(c_E | v_j,p,c_j) \) are too large in lower parts of the thesaurus, since we focus on examples which have a Japanese verb \( v_j \) and a Japanese case marker \( p \). When we used the average of \( \Pr(c_j | v_j,p,c_E) \) and \( \Pr(c_E | v_j,p,c_j) \) instead of \( \Pr(c_E \cap c_j | v_j,p) \) in the experiment of section 6, most discovered clusters consisted of only one example.
4.3.2 Subsumption Relation

Next, we introduce subsumption relation $\leq_f$ of a bilingual surface case structure $e$ and a Japanese case-class frame $f_j$:
\[ e \leq_f f_j \iff \text{for each case marker } p \text{ in } f_j \text{ and its noun class } c_j, \text{there exists the same case marker } p \text{ in } e \text{ and its semantic label } SEM_j \text{ is subordinate to } c_j, \text{i.e. } SEM_j \leq c_j \]

This definition can be easily extended into a subsumption relation of Japanese case-class frames.

4.3.3 Bilingual Class/Frame Association Score

Let $Eg(v_j)$ be the set of bilingual surface case structures collected from the Japanese-English parallel corpora, each element of which has a Japanese verb $v_j$. Among the elements $e$ of $Eg(v_j)$, let $Eg(v_j, c_B)$ be the set of those whose semantic label $SEM_B$ of the English predicate satisfies the class $c_B$, i.e., $SEM_B \leq c_B$, and $Eg(v_j, f_j)$ be the set of those which satisfy the Japanese case-class frame $f_j$, i.e., $e \leq_f f_j$. Let $Eg(v_j, c_B, f_j)$ be the intersection of $Eg(v_j, c_B)$ and $Eg(v_j, f_j)$. Then, conditional probabilities $Pr(c_B | v_j)$, $Pr(f_j | v_j)$, and $Pr(c_B, f_j | v_j)$ are defined as the ratios of the numbers of the elements of those sets:
\[ Pr(c_B | v_j) = \frac{|Eg(v_j, c_B)|}{|Eg(v_j)|} \]
\[ Pr(f_j | v_j) = \frac{|Eg(v_j, f_j)|}{|Eg(v_j)|} \]
\[ Pr(c_B, f_j | v_j) = \frac{|Eg(v_j, c_B, f_j)|}{|Eg(v_j)|} \]

Then, given $v_j$, the association score $A(c_B, f_j | v_j)$ of $c_B$ and $f_j$ is defined as
\[ A(c_B, f_j | v_j) = \frac{Pr(c_B, f_j | v_j)}{Pr(c_B | v_j)Pr(f_j | v_j)} \log \frac{Pr(c_B, f_j | v_j)}{Pr(c_B | v_j)Pr(f_j | v_j)} \]

As well as the case of the bilingual class/class association score, this definition only needs the set $Eg(v_j)$ for a Japanese verb $v_j$, not the whole Japanese-English parallel corpora.

5 Sense Classification of Verbal Polysemy

This section explains how to classify the elements of the set $Eg(v_j)$ of bilingual surface case structures according to the sense of the verbal polysemy $v_j$, with the bilingual class/frame association score defined in the previous section. In this classification process, pairs of an English class $c_B$ and a Japanese case-class frame $f_j$ which give large association score $A(c_B, f_j | v_j)$ are searched for. It is desirable that the set $Eg(v_j)$ be divided into disjoint subsets by the discovered pairs of $c_B$ and $f_j$. The classification process proceeds according to the following steps:

1. First, the index $i$ and the set of examples $Eg$ are initialized as $i \leftarrow 1$ and $Eg \leftarrow Eg(v_j)$.

2. For the $i$-th iteration, let $c_B$ and $f_j$ be a pair of an English class and a Japanese case-class frame which satisfy the following constraint for all the pairs of $c_{Ej}$ and $f_j$ ($1 \leq j \leq i - 1$): $c_B$ is not subordinate nor superordinate to $c_{Ej}$ (i.e., $c_B \notin c_{Ej}$ and $c_{Ej} \notin c_B$), or $f_j$ is not subordinate nor superordinate to $f_{Ej}$ (i.e., $f_j \notin f_{Ej}$ and $f_{Ej} \notin f_j$). Then, among those pairs of $c_B$ and $f_j$, search for a pair $c_{B1}$ and $f_{j1}$ which gives maximum association score $A(c_{B1}, f_{j1} | v_j)$,\(^6\) and collect the elements of $Eg$ which satisfy the restrictions of $c_{B1}$ and $f_{j1}$ into the set $Eg(v_j, c_{B1}, f_{j1})$.

3. Subtract the set $Eg(v_j, c_{B1}, f_{j1})$ from $Eg$ as $Eg \leftarrow Eg - Eg(v_j, c_{B1}, f_{j1})$. If $Eg \neq \emptyset$, then increment the index $i$ as $i \leftarrow i + 1$ and go to step 2. Otherwise, set the number $k$ of the subsets as $k \leftarrow i$ and terminate the classification process.

As the result of this classification process, the set $Eg(v_j)$ is divided into disjoint subsets $Eg(v_j, c_{B1}, f_{j1}), \ldots, Eg(v_j, c_{Bk}, f_{jk})$.\(^6\) For example, if a Japanese polysemous verb $v_j$ has both intransitive and transitive senses, pairs with the subject case like $(c_{B1}, \{\text{sub} : c_{J1}\}), \ldots, (c_{Bk-1}, \{\text{sub} : c_{Jk-1}\})$ will be discovered for intransitive senses, while pairs with the object case like $(c_{Bk-1}, \{\text{obj} : c_{Jk-1}\}), \ldots, (c_{Bk}, \{\text{obj} : c_{Jk}\})$ will be discovered for transitive senses.

Given the set $Eg(v_j)$, the iterations of the association score calculation is $O(|Eg(v_j)|)^5$. Since the classification process can be regarded as sorting the calculated association score, its computational complexity can be $O(|Eg(v_j)| \log |Eg(v_j)|)$ if efficient sorting algorithms such as quick sort are employed.

6 Experiment and Evaluation

This section gives the results of a small exper-
Table 1: Sense Classification of kau

| Hand-Classif. | Cluster No. | English Predicate Class (cE) | Japanese wo(ACC) Case Noun Class (cA) | Number of Egs. | Association Score |
|---------------|-------------|-----------------------------|---------------------------------------|----------------|-------------------|
| 1             | 1           | buy(Leaf)/131(Level3, kon(book)) | /          | 8              | 0.048            |
|               | 2           | buy(Leaf)/13220(Level5, e(picture)) | /          | 3              | 0.018            |
|               | 3           | Purchase(Leaf-1, buy, pay)/141(Level2, Products) | /          | 46             | 0.149            |
|               | 4           | treat oneself to(Leaf)/14650-6-80(Leaf, gaisha(foreign car)) | /          | 1              | 0.070            |
|               | 5           | treat oneself to(Leaf)/14280-3-10(Leaf, yubin(transfer)) | /          | 1              | 0.070            |
|               | 6           | purchase(Leaf)/11720-3-10(Leaf, disha(land)) | /          | 1              | 0.083            |
|               | 7           | bring(Leaf)/14010-4-40(Leaf, migaye(souvenir)) | /          | 1              | 0.062            |
|               | 8           | sell(Leaf)/14570-1-10(Leaf, omocha(toy)) | /          | 1              | 0.070            |
| 2             | 9           | incur(Leaf)/130(Level3, urami(enmity)) | /          | 5              | 0.185            |
|               | 10          | Motive(Leaf-1, rossie)/13020-5-50(Leaf, kankan(antipathy)) | /          | 3              | 0.169            |
|               | 11          | disgust(Leaf)/13010-1-50(Leaf, hinshuku(displeasure)) | /          | 1              | 0.083            |
| 3             | 12          | appreciate(Leaf)/13040-6-30(Leaf, doryoku(effort)) | /          | 1              | 0.083            |
|               | 13          | get an opinion of(Leaf)/12040-1-50(Leaf, oto(idea)) | /          | 1              | 0.083            |
|               | 14          | use(Leaf)/13421-6-50(Leaf, chewan(ability)) | /          | 1              | 0.083            |
| 4             | 15          | win(Leaf)/13010-6-300(Leaf, kanshin(favor)) | /          | 1              | 0.083            |

Total 75

Table 2: Examples of Intransitive/Transitive Distinction

| Japanese Verb | English Predicate Class (cE)/Japanese Case-Class Frame (fJ) | Number of Egs. | Association Score |
|---------------|---------------------------------------------------------|----------------|-------------------|
| haru          | expensive(Leaf)/ga(NOM):ne(price)(Leaf) | 3              | 0.299            |
|               | Special Sensation(Leaf-3, freeze)/        | 3              | 0.237            |
|               | ga(NOM):15130-11-10(Leaf, koors site) | /              | 0.459            |
|               | Acts(Leaf-2, persist, stick to)/        | /              | 0.459            |
|               | wo(ACC):13040(Level5, gyoujin(obstination)) | /              | 0.459            |
| hiku          | Decrease(Leaf-1, suside)/ga(NOM):151(Level3, konzui(floods)) | 2              | 0.109            |
|               | Results of Reasoning(Leaf-2, catch, have)/ | /              | 0.421            |
|               | wo(ACC):15860-11(Level6, kaze(cold)) | /              | 0.421            |
| hiraku        | Intellect(Leaf1, open)/ga(NOM):14(Products)(Level2, to(open)) | 12             | 0.339            |
|               | hold(Leaf)/wo(ACC):13510-11(Level6, kaigou(meeting)) | 3              | 0.114            |
| kannau        | Completion(Leaf-1, realize)/ga(NOM):1304(Leve14, negai(desire)) | 8              | 0.460            |
|               | Quantity(Leaf-3, equal)/ni(DAT):12000-3-10(Leaf, kare(kg)) | 8              | 0.504            |

6.1 Example of kau

First, we show the result of classifying 75 examples (represented as bilingual surface case structures) of the Japanese polysemous verb kau.

As the result of searching for pairs of an English class and a Japanese case-class frame with a large association score, the wo case (the accusative case) is preferred as the most effective case for sense classification. 15 pairs of an English class and a Japanese case-class frame are found and the set of the 75 examples are divided into 15 disjoint clusters (Table 1). Each cluster is represented as a pair of the class cE of the English predicates and the class cA of the Japanese case element nouns of wo case, along with the level of the class in the thesaurus and the example word. English classes are taken from Roget's Thesaurus and Japanese classes from BGH. In both thesauri, leaf classes correspond to one word.

For the evaluation of the results, we hand-classified the 15 clusters into four groups, each of which corresponds to only one sense of kau. Most hand-classified clusters for kau consist of more than one clusters found by maximizing the association score. However, these clusters are correct in that none of them contains examples of more than one hand-classified senses of kau.

6.2 Examples of Intransitive/Transitive Distinction

For four Japanese verbs haru, hiku, hiraku, and kannau, Table 2 shows examples of classifying intransitive/transitive senses by the proposed sense codes, in which each digit denotes the choice of the branch in the thesaurus. The classes starting with '11', '12', '13', '14', and '15' are subordinate to abstract-relations, agents-of-human-activities, human-activities, products and natural-objects-and-natural-phenomena, respectively.

*The criterion of this hand-classification is taken from the existing Japanese dictionaries for human use and the hand-compiled Japanese case frame dictionary IPAL (IPA, 1987).
classification method. Clusters of intransitive senses are discovered with the Japanese case-class frames which contain the ga case (the nominative case), while those of transitive senses are discovered with the Japanese case-class frames which contain the wo case (the accusative case) and ni case (the dative case).

### 6.3 Evaluation

For 9 verbs, we made an experiment on sense classification of verbal polysemy. We compared the result with the hand-classification and checked whether each cluster contained examples of only one hand-classified sense (Table 3). In the table, ‘Cl.’ and ‘Eg.’ indicate the numbers of clusters and examples, respectively. The column ‘One Sense Cluster’ means that each cluster contains examples of only one hand-classified sense, and the sub-columns ‘Cl.’ and ‘Eg.’ list the number of such clusters and the sum of examples contained in such clusters, respectively. We evaluated the accuracy of the method as the rate of the number of examples contained in one sense clusters as in the ‘Eg.’ sub-column. This achieved 100% accuracy for four verbs out of the 9 verbs, and 93.3% in average. The column ‘Total Cl./Hand-Classif.’ indicates the ratio of the total number of clusters to the number of hand-classified senses, corresponding to the average number of clusters into which one hand-classified sense is divided. Its average, median, and standard deviation are 2.46, 1.80, and 1.06, respectively.

The result of the experiment indicated that the proposed sense classification method has achieved almost pure classification, while the result seems a little finer than hand-classification. This is mainly caused by the fact that clusters which correspond to the same hand-classified sense are separately located in the human-made thesaurus, and it is not easy to find exactly one representative class in the thesaurus (Utsuro, 1996). It is necessary to further merge the clusters so that exactly one cluster corresponds to one hand-classified sense.

### Table 3: Evaluation of Sense Classification

| Japanese Verb | Japanese Case-Class Frame $g$ | Total | One Sense Cluster | Hand-Classif. |
|---------------|--------------------------------|-------|-------------------|---------------|
| agaru~a~ (rise) | ga(NOM) | 41 74 39 69 (93.2%) | 17 | 2.41 |
| agaru~a~ (raise) | wo(ACC) | 54 107 52 93 (86.9%) | 18 | 3.00 |
| aku~a~ (open, iv) | ga(NOM) | 12 29 12 29 (100%) | 8 | 1.50 |
| hara~a~ (spread, iv/tv) | ga(NOM)/wo(ACC) | 19 36 17 30 (83.3%) | 11 | 1.73 |
| hika~a~ (subside, pull) | ga(NOM) | 40 105 40 105 (100%) | 23 | 1.74 |
| hira~a~ (open, iv/tv) | ga(NOM)/wo(ACC) | 15 54 13 50 (92.6%) | 10 | 1.50 |
| kaken~a~ (hang) | wo(ACC) | 45 103 42 86 (83.5%) | 25 | 1.80 |
| kana~a~ (realize, conform to) | ga(NOM) | 14 31 14 31 (100%) | 3 | 4.67 |
| kana~a~ (realize, conform to) | /ni(DAT) | 45 103 42 86 (83.5%) | 25 | 1.80 |
| kana~a~ (realize, conform to) | wo(ACC) | 15 75 15 75 (100%) | 4 | 3.75 |
| Average | | | (93.3%) | 2.46 |

### 7 Conclusion

This paper proposed a bilingual class-based method for sense classification of verbal polysemy, which is based on the maximization of the bilingual class/frame association score. It achieved fairly high accuracy, although it is necessary to further merge the clusters so that exactly one cluster corresponds to one hand-classified sense. We are planning to make experiments on sense classification without bilingual information to evaluate the effectiveness of such bilingual information.

### References

K. W. Church and P. Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22-29.

I. Dagan, A. Itai, and U. Schwall. 1991. Two languages are more informative than one. In *Proc. of the 29th Annual Meeting of ACL*, pages 130-137.

IPA, (Information-technology Promotion Agency, Japan). 1987. IPA Lexicon of the Japanese Language for Computers IPAL. (Basic Verbs). (in Japanese).

Y. Matsumoto, H. Ishimoto, and T. Utsuro. 1993. Structural matching of parallel texts. In *Proc. of the 31st Annual Meeting of ACL*, pages 23–30.

NLRI, (National Language Research Institute). 1993. *Word List by Semantic Principles*. Syuei Syuppan. (in Japanese).

P. Resnik. 1992. WordNet and distributional analysis: A class-based approach to lexical discovery. In *Proc. of the AAAI-92 Workshop on Statistically-Based Natural Language Programming Techniques*, pages 48–56.

S. R. Roget. 1911. *Roget’s Thesaurus*. Crowell Co.

M. Shimizu and S. Narita, editors. 1979. *Japanese-English Dictionary*. Kodansha Gakujutsu Bunko.

T. Utsuro, Y. Matsumoto, and M. Nagao. 1993. Verbal case frame acquisition from bilingual corpora. In *Proc. of the 13th IJCAI*, pages 1150-1156.

T. Utsuro. 1995. Class-based sense classification of verbal polysemy in case frame acquisition from parallel corpora. In *Proc. of the 3rd Natural Language Processing Pacific Rim Symposium*, pages 671-677.