Unsupervised Word Sense Disambiguation Using Bilingual Comparable Corpora

Hiroyuki Kaji and Yasutsugu Morimoto

Central Research Laboratory, Hitachi, Ltd.
1-280 Higashi-Koigakubo, Kokubunji-shi, Tokyo 185-8601, Japan
{kaji, y-morimo}@crl.hitachi.co.jp

Abstract

An unsupervised method for word sense disambiguation using a bilingual comparable corpus was developed. First, it extracts statistically significant pairs of related words from the corpus of each language. Then, aligning pairs of related words translingually, it calculates the correlation between the senses of a first-language polysemous word and the words related to the polysemous word, which can be regarded as clues for determining the most suitable sense. Finally, for each instance of the polysemous word, it selects the sense that maximizes the score, i.e., the sum of the correlations between each sense and the clues appearing in the context of the instance. To overcome both the problem of ambiguity in the translingual alignment of pairs of related words and that of disparity of topical coverage between corpora of different languages, an algorithm for calculating the correlation between senses and clues iteratively was devised. An experiment using Wall Street Journal and Nihon Keizai Shimbun corpora showed that the new method has promising performance; namely, the applicability and precision of its sense selection are 88.5% and 77.7%, respectively, averaged over 60 test polysemous words.

1 Introduction

Word sense disambiguation (WSD) is an "intermediate" task that is necessary for accomplishing most natural language processing tasks, especially machine translation and information retrieval. A variety of WSD methods have been proposed over the last decade; however, such methods are still immature. In response to this situation, we have developed an unsupervised WSD method using bilingual comparable corpora.

With the growing amount of texts available in electronic form, data-driven or corpus-based WSD has become popular. The knowledge useful for WSD can be learned from corpora. However, supervised learning methods suffer from the high cost of manually tagging the sense onto each instance of a polysemous word in a training corpus. A number of bootstrapping methods have been proposed to reduce the sense-tagging cost (Hearst 1991; Basili 1997). A variety of unsupervised WSD methods, which use a machine-readable dictionary or thesaurus in addition to a corpus, have also been proposed (Yarowsky 1992; Yarowsky 1995; Karov and Edelman 1998). Bilingual parallel corpora, in which the senses of words in the text of one language are indicated by their counterparts in the text of another language, have also been used in order to avoid manually sense-tagging training data (Brown, et al. 1991).

Unlike the previous methods using bilingual corpora, our method does not require parallel corpora. The availability of large parallel corpora is extremely limited. In contrast, comparable corpora are available in many domains. The comparability required by our method is very weak: any combination of corpora of different languages in the same domain is acceptable as a comparable corpus.

Several types of information are useful for WSD (Ide and Veronis 1998). Three major types are the grammatical characteristics of the polysemous word to be disambiguated, words that are syntactically related to the polysemous word, and words that are topically related to the polysemous word. Among these types, use of grammatical characteristics, which are language-dependent, is not compatible with the approach using bilingual corpora. On the other hand, since a topical relation is language-independent, use of topically related words is most compatible with the approach using bilingual corpora. Accordingly, we focused on using topically related words as clues for determining the most suitable sense of a polysemous word.

2 Approach

2.1 Framework

A comparable corpus consists of a first-language corpus and a second-language corpus of the same domain. Unlike a parallel corpus, we cannot align sentences or instances of words translingually. Therefore, we extract a collection of statistically significant pairs of related words from each language corpus independently of the other language, and then align the pairs of related words translingually with the assistance of a bilingual dictionary. The underlying assumption is that translations of words that are related in one language are also related in the other language (Rapp 1995).

Translingual alignment of pairs of related words enables us to acquire knowledge useful for WSD (i.e., sense-clue pair). For example, the alignment of (tank, gasoline) with (タンク<*TANKU*>, ガソリン<*GASORIN*>) implies that “gasoline” is a clue for selecting the “container” sense of “tank”, which is translated as “タンク<*TANKU*>”, and the alignment of (tank, soldier) with (戦車<*SENSYA*>, 兵士<*HEISI*>) implies that “soldier” is a clue for selecting the “military vehicle” sense of “tank”, which is translated as “戦車<*SENSYA*>”.

Figure 1 shows an overview of our proposed method for acquiring knowledge for WSD. In the framework of translingually aligning pairs of related words, we encounter two major problems: the ambiguity in alignment, and the disparity of topical coverage between the corpora of the two languages. The following sections discuss how to overcome these problems.

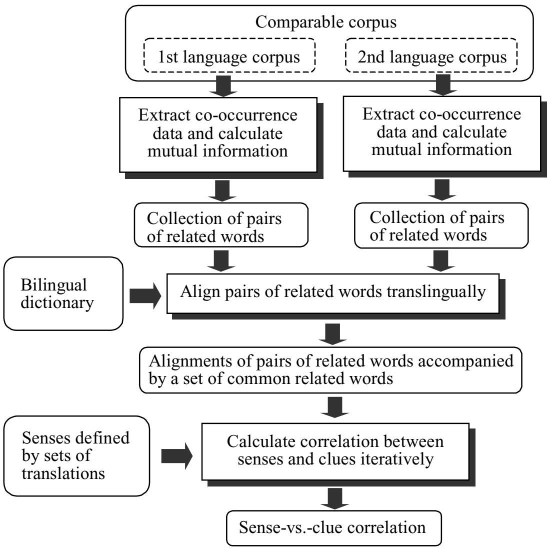
Comparable corpus
1st language corpus
2nd language corpus
Extract co-occurrence data and calculate mutual information
Extract co-occurrence data and calculate mutual information
Collection of pairs of related words
Collection of pairs of related words
Bilingual dictionary
Align pairs of related words translingually
Alignments of pairs of related words accompanied by a set of common related words
Senses defined by sets of translations
Calculate correlation between senses and clues iteratively
Sense-vs.-clue correlation

Fig. 1 Overview of the proposed method for acquiring knowledge for WSD

(a) Common related words of (tank, troop)

Army, Bosnian, Bosnian government, Chechen, Chechnya, Force, Grozny, Israel, Moscow, Mr. Yeltsin, Mr. Yeltsin's, NATO, Pentagon, Republican, Russia, Russian, Secretary, Serb, U.N., Yeltsin, Yeltsin's, air, area, army, assault, battle, bomb, carry, civilian, commander, control, defense, fight, fire, force, government, helicopter, military, missile, rebel, soldier, weapon

(b1) Common related words of (tank, troop) that can be aligned with common related words of (水槽<*SUISOU*>, 群れ<*MURE*>)

air, area, fire, government

(b2) Common related words of (tank, troop) that can be aligned with common related words of (槽<*SOU*>, 多数<*TASUU*>)

area, army, control, force

(b3) Common related words of (tank, troop) that can be aligned with common related words of (戦車<*SENSYA*>, 群<*GUN*>)

area, army, battle, commander, force, government

(b4) Common related words of (tank, troop) that can be aligned with common related words of (戦車<*SENSYA*>, 多数<*TASUU*>)

Serb, area, army, battle, force, government

(b5) Common related words of (tank, troop) that can be aligned with common related words of (戦車<*SENSYA*>, 隊<*TAI*>)

Russia, Serb, air, area, army, battle, commander, defense, fight, fire, force, government, helicopter, soldier

Fig. 2 Example of common related words

2.2 Coping with ambiguity in alignment

Matching of pairs of related words via a bilingual dictionary often suggests that a pair in one language can be aligned with two or more pairs in the other language. For example, an English pair (tank, troop) can be aligned with Japanese pairs (水槽<*SUISOU*>, 群れ<*MURE*>), (槽<*SOU*>, 多数<*TASUU*>), (戦車<*SENSYA*>, 群<*GUN*>), (戦車<*SENSYA*>, 多数<*TASUU*>), and (戦車<*SENSYA*>, 隊<*TAI*>). We resolve this ambiguity on the assumption that correct alignments are accompanied by a lot of common related words that can be aligned with each other. In the above example, a lot of words related to both “tank” and “troop” can be aligned with words related to both “戦車<*SENSYA*>” and “隊<*TAI*>” (see Figure 2(b5)).

The plausibility of alignment is evaluated according to the set of first-language common related words that can be aligned with second-language common related words. Then, using the plausibility of alignment, the correlation between the senses of a polysemous word and the clues for selecting the most suitable sense is calculated. To precisely evaluate the plausibility of alignment, we define it as the sum of the correlations between the sense suggested by the alignment and the common related words accompanying the alignment.

2.3 Coping with disparity between corpora

Given the disparity of topical coverage between the corpora of two languages as well as the insufficient coverage of the bilingual dictionary, the method described in the preceding section seems too strict. As exemplified in Figure 2, even for a correct alignment of a first-language pair of related words with a second-language pair of related words, only a small part of the first-language common related words can be aligned with second-language common related words. To improve the robustness of the method, instead of the set of first-language common related words that can be aligned with second-language common related words, we use a weighted set consisting of all the first-language common related words, where those aligned with second-language common related words are given larger weights than the others.

The disparity of topical coverage between the corpora of two languages and the insufficient coverage of the bilingual dictionary also cause a lot of pairs of related words not to be aligned with any pair of related words. To recover the failure in alignment, we introduce a "wild card" pair, with which every first-language pair of related words is aligned compulsorily. The alignment with the wild-card pair suggests all senses of the first-language polysemous word, and it is accompanied by a set consisting of the first-language common related words with the same weight.

3 Proposed method

3.1 Defining word senses

We define each sense of a polysemous word x of the first language by a synonym set consisting of x itself and one or more of its translations y_1, y_2, ... into the second language. The synonym set is similar to that in WordNet (Miller 1990) except that it is bilingual, not monolingual. Examples of some sets are given below.

{tank, タンク<*TANKU*>, 水槽<*SUISOU*>, 槽<*SOU*>}

{tank, 戦車<*SENSYA*>}

These synonym sets define the "container" sense and the "military vehicle" sense of "tank" respectively.

Translations that preserve the ambiguity are preferably eliminated from the synonym sets defining senses because they are useless for distinguishing the senses. An example is given below.

{title, 肩書き<*KATAGAKI*>, 称号<*SYOUGOU*>, ~~タイトル <*TAITORU*>~~, 敬称<*KEISYOU*>}

{title, 題名<*DAIMEI*>, 題目<*DAIMOKU*>, 表題 <*HYOUDAI*>, 書名<*SYOMEI*>, ~~タイトル<*TAITORU*>~~}

{title, ~~タイトル<*TAITORU*>~~, 選手権<*SENSYUKEN*>}

These synonym sets define the "person's rank or profession" sense, the "name of a book or play" sense, and the "championship" sense of "title". A Japanese word "タイトル<*TAITORU*>", which represents all these senses, is preferably eliminated from all these synonym sets.

3.2 Extraction of pairs of related words

The corpus of each language is statistically processed in order to extract a collection of pairs of related words in the language (Kaji et al. 2000). First, we extract words from the corpus and count the occurrence frequencies of each word. We reject words whose frequencies are less than a certain threshold. We also extract pairs of words co-occurring in a window and count the co-occurrence frequency of each pair of words. In the present implementation, the words are restricted to nouns and unknown words, which are probably nouns, and the window size is set to 25 words excluding function words.

Next, we calculate mutual information $MI(x, x')$ between each pair of words x and x'. $MI(x, x')$ is defined by the following formula:

$$MI(x, x') = \log \frac{Pr(x, x')}{Pr(x) \cdot Pr(x')},$$

where $Pr(x)$ is the occurrence probability of x, and $Pr(x, x')$ is the co-occurrence probability of x and x'. Finally, we select pairs of words whose mutual information value is larger than a certain threshold and at the same time whose relation is judged to be statistically significant through a log-likelihood ratio test.

3.3 Alignment of pairs of related words

In this section, R_X and R_Y denote the collections of pairs of related words extracted from the corpora of the first language and the second language, respectively. D denotes a bilingual dictionary, that is, a collection of pairs consisting of a first-language word and a second-language word that are translations of each other.

Let $X(x)$ be the set of clues for determining the sense of a first-language polysemous word x, i.e.,

$$X(x)=\{x' | (x, x') \in R_X\}.$$

Henceforth, the j-th clue for determining the sense of x is denoted as $x'(j)$.

Let $Y(x, x'(j))$ be the set of counterparts of a pair of first-language related words $(x, x'(j))$, i.e.,

$$Y(x, x'(j))= \{(y, y') \mid (y, y') \in R_Y, (x, y) \in D, (x'(j), y') \in D\}.$$

(1) Each pair of first-language related words $(x, x'(j))$ is aligned with each counterpart (y, y') $(\in Y(x, x'(j)))$, and a weighted set of common related words $Z((x, x'(j)), (y, y'))$ is constructed as follows:

$$Z((x, x'(j)), (y, y')) = \{x'' / w(x'') \mid (x, x'') \in R_X, (x'(j), x'') \in R_X\},$$

where $w(x'')$, which denotes the weight of x'', is set as follows:

- $w(x'') = 1 + \alpha \cdot MI(y, y')$ when $\exists y''\ (x'', y'') \in D$, $(y, y'') \in R_Y$, and $(y', y'') \in R_Y$.
- $w(x'') = 1$ otherwise.

The mutual information of the counterpart, $MI(y, y')$, was incorporated into the weight according to the assumption that alignments with pairs of strongly related words are more plausible than those with pairs of weakly related words. The coefficient α was set to 5 experimentally.

(2) Each pair of first-language related words $(x, x'(j))$ is aligned with the wild-card pair (y_0, y_0'), and a weighted set of common related words $Z((x, x'(j)), (y_0, y_0'))$ is constructed as follows:

$$Z((x, x'(j)), (y_0, y_0')) = \{x'' / w(x'') \mid (x, x'') \in R_X, (x'(j), x'') \in R_X\},$$

where $w(x'') = 1$ for all x''.

3.4 Calculation of correlation between senses and clues

We define the correlation between the i-th sense $S(i)$ and the j-th clue $x'(j)$ of a polysemous word x as follows:

$$C\big(S(i),x'(j)\big)=MI\big(x,x'(j)\big)\cdot\frac{\displaystyle\max_{\substack{(y,y')\in Y(x,x'(j))\\ \cup\{(y_0,y'_0)\}\\ y\in S(i)\cup\{y_0\}}} A\big((x,x'(j)),(y,y'),S(i)\big)}{\displaystyle\max_{k}\left\{\max_{\substack{(y,y')\in Y(x,x'(j))\\ \cup\{(y_0,y'_0)\}\\ y\in S(k)\cup\{y_0\}}} A\big((x,x'(j)),(y,y'),S(k)\big)\right\}},$$

where *A((x, x'(j)), (y, y), S(i))* denotes the plausibility of alignment of *(x, x'(j))* with *(y, y)* suggesting *S(i)*.

The first factor in the above formula, i.e., the mutual information between the polysemous word and the *j*-th clue, is the base of the correlation. The numerator of the second factor is the maximum plausibility of alignments that suggest the *i*-th sense of the polysemous word. The denominator of the second factor has been introduced to normalize the plausibility.

We define the plausibility of alignment suggesting a sense as the weighted sum of the correlations between the sense and the common related words, i.e.,

$$A\big((x,x'(j)),(y,y'),S(i)\big)=\sum_{x''\in Z((x,x'(j)),(y,y'))} w(x'')\cdot C\big(S(i),x''\big).$$

As the definition of the correlation between senses and clues is recursive, we calculate it iteratively with the following initial values: $C_0(S(i),\ x'(j))=MI(x,\ x'(j))$. The number of iteration was set at 6 experimentally.

Figure 3 shows how the correlation values converge. "Troop" demonstrates a typical pattern of convergence; namely, while the correlation with the relevant sense is kept constant, that with the irrelevant sense decreases as the iteration proceeds. "Ozone" demonstrates the effect of the wild-card pair. Note that the correlation values due to an alignment with the wild-card pair begin to diverge in the second cycle of iteration. The alignment with the wild-card pair, which is shared by all senses, does not produce any distinction among the senses in the first cycle of iteration; the divergence is caused by the difference in correlation values between the senses and the common related words.

3.5 Selection of the sense of a polysemous word

Consulting sense-vs.-clue correlation data acquired by the method described in the preceding sections, we select a sense for each instance of a polysemous word *x* in a text. The score of each sense of the polysemous word is defined as the sum of the correlations between the sense and clues appearing in the context, i.e.,

$$Score\big(S(i)\big)=\sum_{x'(j)\in Context(x)} C\big(S(i),x'(j)\big).$$

A window of 51 words (25 words before the polysemous word and 25 words after it) is used as the context.

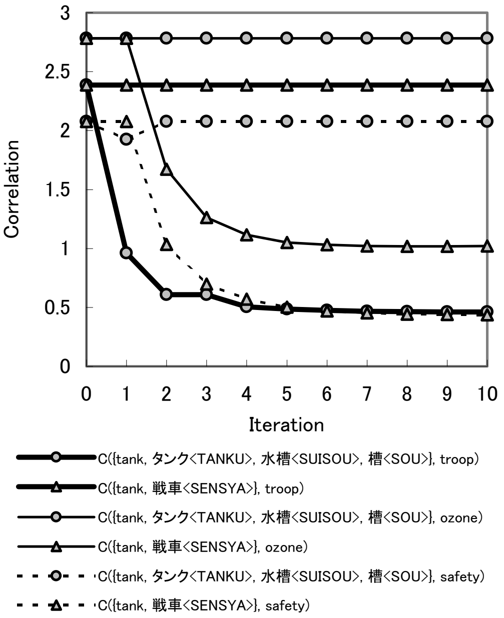

Fig. 3 Convergence of correlation between senses and clues

Scores of all senses of a polysemous word are calculated, and the sense whose score is largest is selected as the sense of the instance of the polysemous word. When all scores are zero, no sense can be selected; the case is called "inapplicable".

4 Experiment

4.1 Experimental method

We evaluated our method through an experiment using corpora of English and Japanese newspaper articles. The first language was English and the second language was Japanese. A Wall Street Journal corpus (July, 1994 to Dec., 1995; 189 Mbytes) and a Nihon Keizai Shimbun corpus (Dec., 1993 to Nov., 1994; 275 Mbytes) were used as the training comparable corpus. EDR (Japan Electronic Dictionary Research Institute) English-to-Japanese and Japanese-to-English dictionaries were merged for the experiment. The resulting dictionary included 269,000 English nouns and 276,000 Japanese nouns. Pairs of related words were extracted from the corpus of each language under the following parameter settings:

- threshold for occurrence frequencies of words: 10
- threshold for mutual information: 0.0

These settings were common to the English and Japanese corpora.

We selected 60 English polysemous nouns as the test words. Words whose different senses appear in newspapers were preferred. The frequencies of the test words in the training corpus ranged from 39,140 ("share", the third noun in descending order of frequency) to 106 ("appreciation", the 2,914th noun).

We defined the senses of each test word. The number of senses per test word ranged from 2 to 8, and the average was 3.4. For each test word, sense-vs.-clue correlation data were acquired by the method described in Sections 3.2 through 3.4. 175 clues on average were acquired for each test word.

For evaluation, we selected 100 test passages per test word from a Wall Street Journal corpus (Jan., 1996 to Dec. 1996) whose publishing period was different from that of the training corpus. The instances of test words positioned in the center of each test passage were disambiguated by the method described in Section 3.5, and the results were compared with the manually selected senses.

4.2 Results and evaluation

We used two measurements, applicability and precision (Dagan and Itai 1994), to evaluate the performance of our method. The applicability is the proportion of instances of the test word(s) that the method could disambiguate. The precision is the proportion of disambiguated instances of the test word(s) that the method disambiguated correctly. The applicability and precision of the proposed method, averaged over the 60 test polysemous words, were 88.5% and 77.7%, respectively.

The performance of our method on six out of the 60 test words is summarized in Table 1. That is, the instances are classified according to the correct sense and the sense selected by our method. These results show that the performance varies according to the test words, that our method is better in the case of frequent senses, but worse in the case of infrequent senses, and that our method can easily distinguish topic-specific senses, but not generic senses.

We consider the reason for the poor performance concerning "measure" [Table 1(a)] and "race" [Table 1(c)] as follows. The second sense of "measure", {measure, 対策<*TAISAKU*>, 手段<*SYUDAN*>, 処置<*SYOTI*>}, is a very generic sense; therefore effective clues for identifying the sense could not be acquired. The first sense of "race", {race, レース<*REESU*>, 競争<*KYOUSOU*>, 競走<*KYOUSOU*>, 争い<*ARASOI*>, 戦<*SEN*>}, is specific to the "race for the presidency" topic and the second sense of "race", {race, 人種<*ZINSYU*>, 民族<*MINZOKU*>, 種属<*SYUZOKU*>}, is specific to the "racial discrimination" topic; however, both topics are related to "politics" and, therefore, many clues were shared by these two senses.

Comparison with a baseline method, which selects the most frequent sense of each polysemous word independently of contexts, was also done. Since large sense-tagged corpora were not available, we simulated the baseline method with a modified version of the proposed method; namely, for each polysemous word, the sense that maximizes the sum of correlations with all clues was selected as the most frequent sense. The applicability of the baseline method is 100%, while that of the proposed method is less than 100%. To compare with the baseline method, the proposed method was substituted with the proposed method + baseline method; namely, the baseline method was applied when the proposed method was inapplicable.

The average precisions of the baseline method and the proposed method + baseline method, both of which attained 100% applicability, were 62.8% and 73.4% respectively. Figure 4 visualizes the superiority of the proposed method + baseline method; the 60 test polysemous words are scattered on a plane whose horizontal and vertical coordinates represent the precision of the baseline method and that of the proposed method + baseline method, respectively.

5 Discussion

Although it has produced promising results, the developed WSD method has a few problems. These limitations, along with future extensions, are discussed below.

(1) Multilingual distinction of senses

The developed method is based on the premise that the senses of a polysemous word in a language are lexicalized differently in another language. However, the premise is not always true; that is, the ambiguity of a word may be preserved by its translations. As described in Section 3.1, we preferably use translations that do not preserve the ambiguity. However, doing so is useless unless such translations are frequently used words. An essential approach to solving this problem is to use two or more second languages (Resnik and Yarowsky 2000).

(2) Use of syntactic relations

The developed method extracts clues for WSD according to co-occurrence in a window. However, it is obvious that doing this is not suitable for all polysemous words. Syntactic co-occurrence is more useful for disambiguating some sorts of polysemous words. It is an important and interesting research issue to extend our method so that it can acquire clues according to syntactic co-occurrence. This extended method does not replace the present method; however, we

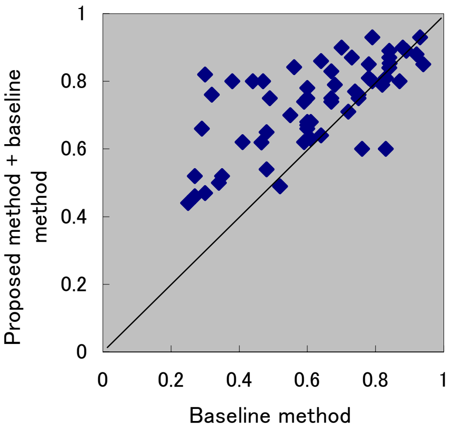

Fig. 4 Precision of sense selection

Table 1 Results of sense selection for six polysemous words

(a) Polysemous word “measure” (applicability=91.0%; precision=48.4%)

Results Correct sense	S1	S2	S3	?	Total
S1={measure, 寸法, 大きさ, 量, 程度, 尺度, 指数, 基準, 測定, 計測}	20	0	13	4	37
S2={measure, 対策, 手段, 処置}	4	0	29	5	38
S3={measure, 法案, 議案, 法令}	1	0	24	0	25
Total	25	0	66	9	100

[Note]
S1: a system or instrument for calculating amount, size, weight, etc.
S2: an action taken to gain a certain end
S3: a law suggested in Parliament

(b) Polysemous word “promotion” (applicability=96.0%; precision=89.6%)

Results Correct sense	S1	S2	S3	?	Total
S1={promotion, 宣伝, 売り込み, 販売促進, プロモーション}	71	1	0	1	73
S2={promotion, 昇格, 昇進, 昇任, 就任, 登用, 進級}	6	15	0	3	24
S3={promotion, 奨励, 振興, 促進, 増進, 助長}	2	1	0	0	3
Total	79	17	0	4	100

[Note]
S1: an activity intended to help sell a product
S2: advancement in rank or position
S3: action to help something develop or succeed

(c) Polysemous word “race” (applicability=79.0%; precision=57.0%)

Results Correct sense	S1	S2	S3	?	Total
S1={race, レース, 競争, 競走, 争い, 戦}	28	33	0	15	76
S2={race, 人種, 民族, 種属}	1	17	0	6	24
S3={race, 水路, 用水}	0	0	0	0	0
Total	29	50	0	21	100

[Note]
S1: any competition, or a contest of speed
S2: one of the groups that humans can be divided into according to physical features, history, language, etc.
S3: a channel for a current of water

(d) Polysemous word “tank” (applicability=89.0%; precision=89.9%)

Results Correct sense	S1	S2	?	Total
S1={tank, タンク, 水槽, 槽}	57	1	6	64
S2={tank, 戦車}	8	23	5	36
Total	65	24	11	100

[Note]
S1: a large container for storing liquid or gas
S2: an enclosed heavily armed, armored vehicle

(e) Polysemous word “title” (applicability=92.0%; precision=81.5%)

Results Correct sense	S1	S2	S3	S4	?	Total
S1={title, 肩書き, 称号, 敬称}	43	1	0	0	2	46
S2={title, 題名, 題目, 表題, 書名}	6	26	0	1	5	38
S3={title, 権利, 資格, 所有権}	1	1	0	1	1	4
S4={title, 選手権}	3	3	0	6	0	12
Total	53	31	0	8	8	100

[Note]
S1: a word or name given to a person to be used before his/her name as a sign rank, profession, etc.
S2: a name given to a book, play, etc.
S3: the legal right to own something
S4: the position of being the winner of an sports competition

(f) Polysemous word “trial” (applicability=92.0%; precision=92.4%)

Results Correct sense	S1	S2	S3	S4	S5	?	Total
S1={trial, 裁判, 公判, 審理}	62	3	0	0	0	5	70
S2={trial, 試し, 試み, 試験, 実験, 試用}	4	23	0	0	0	2	29
S3={trial, 予選}	0	0	0	0	0	1	1
S4={trial, 厄介, 困り者}	0	0	0	0	0	0	0
S5={trial, 苦難, 試練, 辛苦}	0	0	0	0	0	0	0
Total	66	26	0	0	0	8	100

[Note]
S1: a legal process in which a court examines a case
S2: a process of testing to determine quality, value, usefulness, etc.
S3: a sports competition that tests a player’s ability
S4: annoying thing or person
S5: difficulties and troubles

should combine both methods or use the one suitable for each polysemous word. It should be noted that this extension also enables disambiguation of polysemous verbs.

The framework of the method is compatible with syntactic co-occurrence. Basically, we only have to incorporate a parser into the step of extracting pairs of related words. A parser of the first language is indispensable, but a parser of the second language is not. As for the second language, we can use co-occurrence in a small-sized window instead of syntactic co-occurrence.

6 Comparison with other methods

While our method aligns pairs of related words that are statistically extracted, WSD using parallel corpora aligns instances of words (Brown, et al. 1991). Both alignment techniques are quite different. Actually, from the technological viewpoint, our method is close to WSD using a second-language monolingual corpus

(Dagan and Itai 1994; Kikui 1998), where instances of co-occurrence in a first-language text are aligned with co-occurrences statistically extracted from the second-language corpus. A comparison of our method with WSD using a second-language monolingual corpus is given below.

First, our method performs alignment during the acquisition phase, and transforms word-word correlation data into sense-clue correlation data, which is far more informative than the original word-word correlation data. In contrast, a method using a second-language monolingual corpus uses original word-word correlation data during the disambiguation phase. This difference results in a difference in the performance of WSD, particularly in a poor-context situation (e.g., query translation).

Second, our method can acquire sense-clue correlation even from a pair of related words for which alignment results in failure [e.g., C({tank, タンク<*TANKU*>, 水槽<*SUISOU*>, 槽<*SOU*>}, ozone) in Figure 3]. On the contrary, a conventional WSD method using a second-language monolingual corpus uses only pairs of related words for which alignment results in success. Thus, our method can elicit more information than the conventional method.

Tanaka and Iwasaki (1996) exploited the idea of translingually aligning word co-occurrences to extract pairs consisting of a word and its translation form a non-aligned (comparable) corpus. The essence of their method is to obtain a translation matrix that maximizes the distance between the co-occurrence matrix of the first language and that of the second language. Their method is useful for extracting corpus-dependent translations; however, it does not extract knowledge for WSD, i.e., which co-occurring word suggests which sense or translation.

7 Conclusion

A method for word sense disambiguation using a bilingual comparable corpus together with sense definitions by translations into another language was developed. In this method, knowledge for WSD, i.e., sense-vs.-clue correlation, is acquired in an unsupervised fashion as follows. First, statistically significant pairs of related words are extracted from the corpus of each language. Then, aligning pairs of related words translingually, the correlation between the senses of a polysemous word and the clues, i.e., the words related to the polysemous word, is calculated. In order to overcome both the problem of ambiguity in the translingual alignment of pairs of related words and that of disparity of topical coverage between corpora of different languages, an iterative algorithm for calculating the correlation was developed.

WSD for each instance of the polysemous word is done by selecting the sense that maximizes the score, i.e., the sum of the correlations between each sense and the clues appearing in the context of the instance. An experiment using corpora of English and Japanese newspaper articles showed that the performance of the new method is promising: the applicability and precision of sense selection were 88.5% and 77.7%, respectively, averaged over 60 test polysemous words.

Acknowledgments: This research was sponsored in part by the Telecommunications Advancement Organization of Japan.

References

Basili, Roberto, Michelangelo Della Rocca, and Maria Tereza Pazienza. 1997. Towards a bootstrapping framework for corpus semantic tagging. In *Proceedings of the ACL-SIGLEX Workshop "Tagging Text with Lexical Semantics: Why, What, and How?"* pages 66-73.

Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1991. Word-sense disambiguation using statistical methods. In *Proceedings of the 29th Annual Meeting of the ACL*, pages 264-270.

Dagan, Ido and Alon Itai. 1994. Word sense disambiguation using a second language monolingual corpus. *Computational Linguistics*, 20(4): 563-596.

Hearst, Marti A. 1991. Noun homograph disambiguation using local context in large corpora. In *Proceedings of the 7th Annual Conference of the Centre for the New OED and Text Research: Using Corpora*, pages 1-22.

Ide, Nancy and Jean Veronis. 1998. Introduction to the special issue on word sense disambiguation: the state of the art. *Computational Linguistics*, 24(1): 1-40.

Kaji, Hiroyuki, Yasutsugu Morimoto, Toshiko Aizono, and Noriyuki Yamasaki. 2000. Corpus-dependent association thesaurus for information retrieval, In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 404-410.

Karov, Yael and Shimon Edelman. 1998. Similarity-based word sense disambiguation. *Computational Linguistics*, 24(1): 41-59.

Kikui, Genichiro. 1998. Term-list translation using monolingual word co-occurrence vectors. In *Proceedings of the 17th International Conference on Computational Linguistics*, pages 670-674.

Miller, George A. 1990. WordNet: an on-line lexical database. *International Journal of Lexicography*, 3(4): 235-312.

Rapp, Reinhard. 1995. Identifying word translations in non-parallel texts. In *Proceedings of the 33rd Annual Meeting of the ACL*, pages 320-322.

Resnik, Philip and David Yarowsky. 2000. Distinguishing systems and distinguishing senses: new evaluation methods for word sense disambiguation. *Natural Language Engineering*, 5(2): 113-133.

Tanaka, Kumiko and Hideya Iwasaki. 1996. Extraction of lexical translations from non-aligned corpora, In *Proceedings of the 16th International Conference on Computational Linguistics*, pages 580-585.

Yarowsky, David. 1992. Word sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proceedings of the 14th International Conference on Computational Linguistics*, pages 454-460.

Yarowsky, David. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting of the ACL*, pages 189-196.