Case Contribution in Example-Based Verb Sense Disambiguation

Atsushi Fujii,† Kentaro Inui,† Takenobu Tokunaga†
and Hozumi Tanaka†

Word sense disambiguation has recently been utilized in corpus-based approaches, reflecting the growth in the number of machine readable texts. One category of approaches disambiguates an input verb sense based on the similarity between its governing case fillers and those in given examples. In this paper, we introduce the degree of case contribution to verb sense disambiguation into this existing method. In this, greater diversity of semantic range of case filler examples will lead to that case contributing to verb sense disambiguation more. We also report the result of a comparative experiment, in which the performance of disambiguation is improved by considering this notion of semantic contribution.

KeyWords: *Word Sense Disambiguation, Example-Based Approach, Case Contribution to Disambiguation*

1 Introduction

Word sense disambiguation is a crucial task in many kinds of natural language processing applications, such as word selection in machine translation (Brown, Pietra, & Pietra, 1991), pruning of syntactic structures in parsing (Lytinen, 1986; Nagao, 1994) and text retrieval (Krovets & Croft, 1992; Voorhees, 1993). Various researches on word sense disambiguation have recently been utilized in corpus-based approaches, reflecting the growth in the number of machine readable texts. Unlike rule-based approaches, corpus-based approaches free us from the task of generalizing observed phenomena to produce rules for word sense disambiguation, e.g. subcategorization rules. Corpus-based approaches are executed based on the intuitively feasible assumption that the higher the degree of similarity between the context of an input word and the context in which the word appears in a sense in a corpus, the more plausible it becomes that the word is used in the same sense. Corpus-based methods are classified into two approaches: example-based approaches (Kurohashi & Nagao, 1994; Uramoto, 1994) and statistics-based approaches (Brown et al., 1991; Dagan & Itai, 1994; Niwa & Nitta, 1994; Schütze, 1992; Yarowsky, 1995). We follow the example-based approach in explaining its effectivity for verb sense disambiguation in Japanese.

† Tokyo Institute of Technology, Department of Computer Science

A representative example-based method for verb sense disambiguation was proposed by Kurohashi and Nagao (Kurohashi's method) (Kurohashi & Nagao, 1994). Their method uses an example database containing examples of collocations as in figure 1.

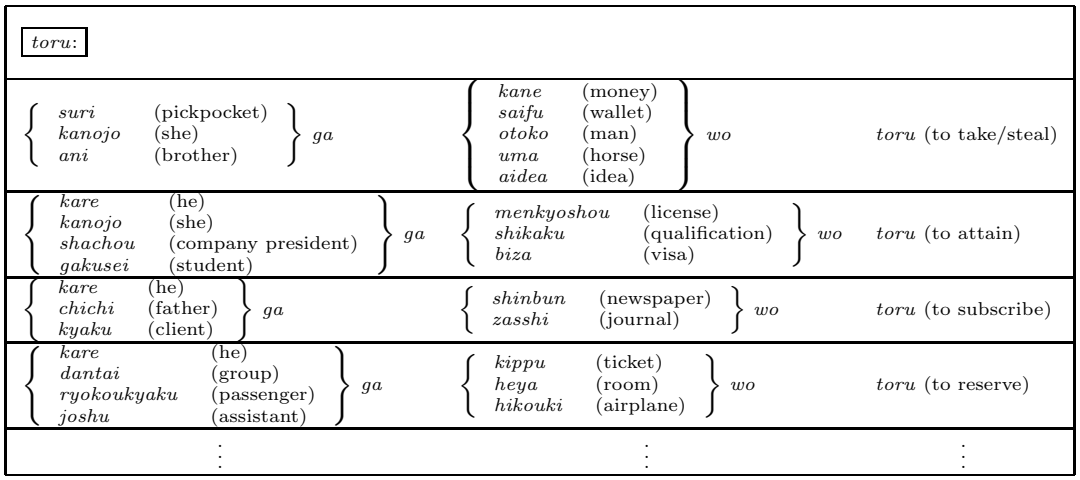

Fig. 1 A fragment of an example database, and the entry associated with Japanese verb *toru*

Figure 1 shows a fragment of the entry associated with the Japanese verb *toru*. As with most words, the verb *toru* has multiple senses, examples of which are "to take/steal", "to attain", "to subscribe" and "to reserve". The database gives one or more case frame(s) associated with the verbs for each of their senses. In Japanese, a complement of a verb, which is a constituent of the case frame of the verb, consists of a noun phrase (case filler) followed by a case marker suffix such as *ga* (nominative) or *wo* (accusative). The database has an example set of case fillers for each case. As shown in figure 1, examples of a complement can be considered as an extensional description of the selectional restriction on it.

The task considered in this paper is "to interpret" a verb in an input sentence, i.e. to choose one sense from a set of candidate senses of the verb. Given an input sentence, Kurohashi's method interprets the verb in the input by computing semantic similarity between the input and examples. For this computation, Kurohashi's method experimentally uses the Japanese word thesaurus *Bunruigoihyo* (National-Language Research Institute, 1964). As with most thesauruses, the length of the path between two words in *Bunruigoihyo* is expected to reflect the similarity between them. Figure 2 illustrates a fragment of *Bunruigoihyo* including nouns associated with the nominative and the accusative in figure 1 respectively.

Let us take the example sentence (1).

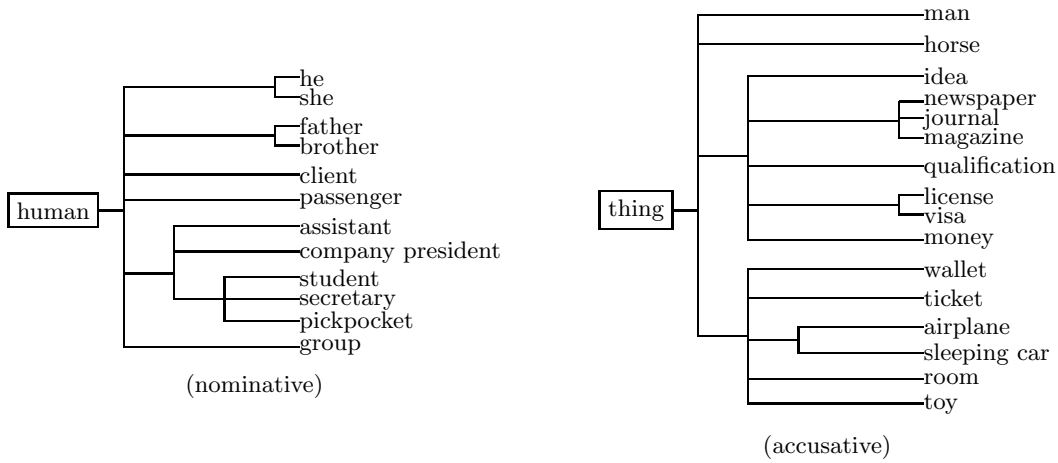

Fig. 2 Showing a fragment of *Bunruigoihyo*

(1) *hisho ga* *shindaisha wo* *toru.*
(secretary-NOM) (sleeping car-ACC) (?)

In this example, it may be judged according to figure 2 that *hisho* ("secretary") and *shindaisha* ("sleeping car") in (1) are semantically similar to *joshu* ("assistant") and *hikouki* ("airplane"), respectively, which are examples that collocate with *toru* ("to reserve"). As such, the sense of *toru* in (1) can be interpreted as "to reserve". However, in Kurohashi's method, several useful properties for verb disambiguation are missing:

(1) Intuitively speaking, the contribution of the accusative to verb sense disambiguation is greater than that of the nominative with the case of verb "toru".

(2) The selectional restriction of a certain case is stronger than those of others. For example, in the accusative, the selectional restriction of "to subscribe" is stronger than that of "to take/steal" which allows various kinds of objects as its case filler.

In this paper, we improve on Kurohashi's method by introducing a formalization of these notions, and report the result of a comparative experiment.

2 Motivation

Property 1 in section 1 is exemplified by the input sentence (2).

(2) *shachou ga* *shuukanshi o* *toru.*
(company president-NOM) (magazine-ACC) (?)

The nominative, *shachou* ("company president"), in (2) is found in the "to attain" case frame of *toru* and there is no other co-occurrence in any other sense of *toru*; therefore, the nominative

supports an interpretation "to attain". On the other hand, the accusative, *shuukanshi* ("magazine"), is most similar to the examples included in the accusative of the "to subscribe" and therefore the accusative supports another interpretation "to subscribe". Although the most plausible interpretation here is actually the latter, Kurohashi's method would choose the former since (a) the degree in which the nominative supports "to attain" happens to be stronger than the degree in which the accusative supports "to subscribe", and (b) their method always relies equally on the similarity in the nominative and the accusative. However, in the case of *toru*, since the semantic range of nouns collocating with the verb in the nominative does not seem to have a strong delinearization in a semantic sense, it would be difficult, or even risky, to properly interpret the verb sense based on the similarity in the nominative. In contrast, since the ranges are diverse in the accusative, it would be feasible to rely more strongly on the similarity in the accusative. This argument can be illustrated as in figure 3, in which the symbols "1" and "2" denote example case fillers of different case frames respectively, and an input sentence includes two case fillers denoted by "x" and "y".

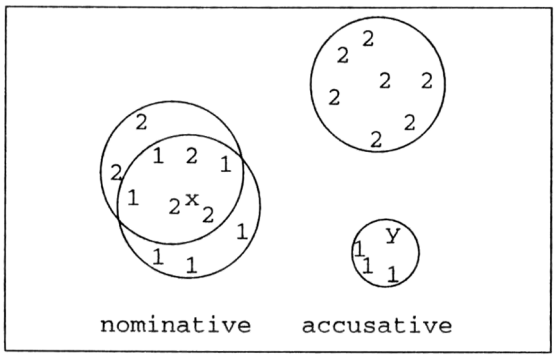

Fig. 3 The semantic ranges of the nominative and accusative with verb *toru*

The figure shows the distribution of example case fillers denoted by those symbols in a semantic space, where the semantic similarity between two case fillers is represented by the physical distance between two symbols. In the nominative, since "x" happens to be much closer to a "2" than any "1", "x" may be estimated to belong to the range of "2"s although "x" actually belongs to both sets of "1"s and "2"s. In the accusative, however, "y" would be properly estimated to belong to "1"s due to the mutual independence of the two accusative case filler sets, even though examples did not fully cover each of the ranges of "1"s and "2"s. Note that this difference would be critical if example data were sparse. This argument suggests that we introduce the degree of case contribution to verb sense disambiguation. One

may argue that this property can be generalized as the notion that the system always relies only on the similarity in the accusative for verb sense disambiguation. Although some typical verbs show this general notion, it is not guaranteed for any kind of verb. Our approach, which computes the degree of contribution for each verb respectively, can handle exceptional cases as well as typical ones.

Property 2 is exemplified by the input sentence (3).

(3) *oniisan ga* *omocha wo* *toru.*
(brother-NOM) (toy-ACC) (?)

In (3) the most plausible interpretation of *toru* is "to steal". The nominative does not give much information for interpreting the verb for the same reason as example (2). In the accusative, the database in figure 1 has two example case fillers that are equally similar to *omocha* ("toy"): *saifu* ("wallet") and *hikouki* ("airplane"). These examples equally support two different interpretations: "to steal" and "to reserve", which means that the verb sense ambiguity still remains. Here, one may notice that since the accusative examples in the case frame of *toru* ("to reserve") are less diverse in meaning than the other case frames, the selectional restriction on the accusative of *toru* ("to reserve") is relatively strong, and thus that it can be estimated to be relatively implausible for *omocha* ("toy") to satisfy it. If such reasoning is correct, given that the examples in the accusative of *toru* ("to steal") are most widely distributed, the input verb can be interpreted as "to steal". The consideration above motivated us to introduce the notion of relative strength of selectional restriction into our example-based verb sense disambiguation method.

3 Algorithm

We assume that inputs are simple sentences, each one of which consists of a sequence of cases followed by their governing verb. The task is to identify the sense of each input verb. The set of verb senses we use are those defined in the existing machine readable dictionary "IPAL" (IPA, 1987), which also contains example case fillers as shown in figure 1. As well as Kurohashi's method the similarity between two case fillers, or more precisely the semantic-head nouns of them, is computed by using *Bunruigoihyo* (National-Language Research Institute, 1964). Following Kurohashi's method, we define $sim(X, Y)$, which stands for the similarity between words X and Y, as in table 1. It should be noted here that both methods are theoretically independent of what resources are used.

To illustrate the overall algorithm, we replace the illustrative cases mentioned in section 1 with a slightly more general case as in figure 4. The input is $\{n_{c_1}\text{-}m_{c_1}, n_{c_2}\text{-}m_{c_2}, n_{c_3}\text{-}m_{c_3},$

Table 1 The relation between the length of path between two nouns X and Y ($len(X,Y)$) in *Bunruigoihyo* and the similarity between them ($sim(X,Y)$)

$len(X,Y)$	0	2	4	6	8	10	12
$sim(X,Y)$	11	10	9	8	7	5	0

v}, where n_{c_i} denotes the case filler in the case c_i, and m_{c_i} denotes the case maker of c_i. The candidates of interpretation for v, which are s_1, s_2 and s_3, are derived from the database. The database also gives a set $\mathcal{E}_{s_i,c_j}$ of case filler examples for each case c_j of each sense s_i. "—" denotes that the corresponding case is not allowed.

input	n_{c_1}-m_{c_1}	n_{c_2}-m_{c_2}	n_{c_3}-m_{c_3}		v (?)
database	$\mathcal{E}_{s_1,c_1}$	$\mathcal{E}_{s_1,c_2}$	$\mathcal{E}_{s_1,c_3}$	—	v (s_1)
	$\mathcal{E}_{s_2,c_1}$	$\mathcal{E}_{s_2,c_2}$	$\mathcal{E}_{s_2,c_3}$	$\mathcal{E}_{s_2,c_4}$	v (s_2)
	—	$\mathcal{E}_{s_3,c_2}$	$\mathcal{E}_{s_3,c_3}$	—	v (s_3)

Fig. 4 An input and the database

In the course of the verb sense disambiguation process, the system first discards the candidates whose case frame constraint is grammatically violated by the input (this parallels Kurohashi's method). In the case of figure 4, s_3 is discarded because the case frame of v (s_3) does not subcategorize the case c_1[1]. In contrast, s_2 will not be rejected at this step. This is based on the fact that in Japanese, cases can be easily omitted if they are inferable from the given context.

Thereafter, the system computes the plausibility of the remaining candidates of interpretation and chooses the most plausible interpretation as its output. In Kurohashi's method, the plausibility of an interpretation is computed by averaging the degree of similarity between the input complement and the example complements[2] for each case as in equation (1), where $P(s)$ is the plausibility of interpreting the input verb as sense s, and $SIM(n_c, \mathcal{E}_{s,c})$ is the degree of the similarity between the input complement n_c and example complements $\mathcal{E}_{s,c}$.

$$P(s) = \sum_c SIM(n_c, \mathcal{E}_{s,c}) \tag{1}$$

1 Since IPAL does not necessarily enumerate all the possible optional cases, the absence of case c_1 from v (s_3) in the figure may denote that c_1 is optional. If so, the interpretation s_3 should not be discarded in this stage. To avoid this problem, we use the same technique as used in Kurohashi's method. That is, we define several particular cases beforehand, such as the nominative, the accusative and the dative, to be obligatory, and impose the grammatical case frame constraint as above only in those obligatory cases. Optionality of case needs to be further explored.

2 $\mathcal{E}_{s_2,c_4}$ is not taken into consideration in the computation since c_4 does not appear in the input.

$SIM(n_C, \mathcal{E}_{s,c})$ is the maximum degree of similarity between n_C and each of $\mathcal{E}_{s,c}$ as in equation (2).

$$SIM(n_C, \mathcal{E}_{s,c}) = \max_{e \in \mathcal{E}_{s,c}} sim(n_C, e) \tag{2}$$

In our method, on the other hand, for the reason indicated in section 1, we introduce two new factors:

- case contribution to disambiguation (CCD),
- relative strength of selectional restriction (RSSR).

First, in regard to CCD, we compute the plausibility of an interpretation by the *weighted* average of the degree of similarity for each case as in equation (3), replacing equation (1).

$$P(s) = \frac{\sum_c SIM(n_C, \mathcal{E}_{s,c}) \cdot CCD(c)}{\sum_c CCD(c)} \tag{3}$$

Here, $CCD(c)$ is a newly introduced weight, such that $CCD(c)$ is greater when the degree of case c's contribution is higher.

Second, in regard to RSSR, the stronger the selectional restriction on a case of a case frame is, the less plausible an input complement satisfies that restriction as mentioned in section 1. Note here that the plausibility of an interpretation of an input verb can be regarded as the plausibility that the input complements satisfy the selectional restriction associated with that interpretation. This leads us to replace $SIM(n_C, \mathcal{E}_{s,c})$ in equation (3) with $PSS(n_C, \mathcal{E}_{s,c})$, which denotes the plausibility that the case filler n_C satisfies the selectional restriction described by the example case fillers $\mathcal{E}_{s,c}$.

$$P(s) = \frac{\sum_c PSS(n_C, \mathcal{E}_{s,c}) \cdot CCD(c)}{\sum_c CCD(c)} \tag{4}$$

From the assumption that $PSS(n_C, \mathcal{E}_{s,c})$ should be greater for a larger $SIM(n_C, \mathcal{E}_{s,c})$ and lesser relative strength of the selectional restriction described by $\mathcal{E}_{s,c}$, we can derive equation (5).

$$PSS(n_C, \mathcal{E}_{s,c}) = SIM(n_C, \mathcal{E}_{s,c}) - RSSR(s, c) \tag{5}$$

Here, $RSSR(s, c)$ denotes the relative strength of the selectional restriction on a case c associated with a sense s.

4 Computation of CCD and RSSR

The degree of case contribution to verb sense disambiguation (CCD) is computed in the following way. The degree of contribution of a case should be high if the semantic range of the example case fillers in that case is diverse in the case frame (see figure 3). Let a certain verb

have n senses $(s_1, s_2, \ldots, s_n)$ and the set of example case fillers of a case c associated with s_i be $\mathcal{E}_{s_i,c}$. Then, the degree of c's contribution to disambiguation, $CCD(c)$, is expected to be higher if the example case filler sets $\{\mathcal{E}_{s_i,c} \mid i = 1, \ldots, n\}$ share less elements. This can be realized by equation (6).

$$CCD(c) = \left(\frac{1}{{}_nC_2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{|\mathcal{E}_{s_i,c}| + |\mathcal{E}_{s_j,c}| - 2|\mathcal{E}_{s_i,c} \cap \mathcal{E}_{s_j,c}|}{|\mathcal{E}_{s_i,c}| + |\mathcal{E}_{s_j,c}|} \right)^{\alpha} \tag{6}$$

α is the constant for parameterizing to what extent CCD influences verb sense disambiguation. When α is larger, CCD more strongly influences the system's output. Considering the data sparseness problem, we do not distinguish two nouns X and Y in equation (6) if X and Y are similar enough, as in equation (7).

$$\{X\} \cup \{Y\} = \{X\} \quad \text{if } sim(X, Y) >= 9 \tag{7}$$

Relative strength of selectional restriction (RSSR) is computed in the following way. The selectional restriction on a case of a case frame is expected to be strong if the example case fillers of the case are similar to each other. Given a set of example case fillers in a case associated with a verb sense, the strength of the selectional restriction on that case (SSR) can be estimated by averaging the similarity between any combination of two elements of that set. Thus, given a set $\mathcal{E}_{s,c}$ of example case fillers in a case c associated with a verb sense s, the SSR of c associated with s can be estimated by equation (8), where $\mathcal{E}^i_{s,c}$ is an i-th element of $\mathcal{E}_{s,c}$, and m is the number of elements in $\mathcal{E}_{s,c}$, i.e. $m = |\mathcal{E}_{s,c}|$.

$$SSR(s, c) = \begin{cases} \dfrac{\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} sim(\mathcal{E}^i_{s,c}, \mathcal{E}^j_{s,c})}{{}_mC_2} & \text{if } m > 1 \\ \text{maximum} & \text{otherwise} \end{cases} \tag{8}$$

In the case $m = 1$, that is, the case has only one example case filler, the SSR becomes maximum, because the selectional constraint associated with the case is highest (following table 1, we assign 11 as the maximum to SSR). The *relative* strength of selectional restriction (RSSR) of a case associated with a verb sense is estimated by the ratio of the SSR of the case to the summation of the SSRs of each case associated with the verb sense, as in equation (9) [3].

$$RSSR(s, c) = \frac{SSR(s, c)}{\sum_i SSR(s_i, c)} \tag{9}$$

5 Evaluation

Our experiment compared the performance of the following methods:

3 Note that, in equation (5), while SIM is an integer, RSSR ranges in its value from 0 to 1. Therefore, RSSR is influential only when several verb senses take the same value of SIM for a given case.

(1) the control (Kurohashi's method): equation (1)

(2) our method (considering CCD): equation (3)

(3) our method (considering both CCD and RSSR): equation (4)

In method (2) and (3), the influence of CCD, i.e. α in equation (6), was extremely large. We will show the relation between the variation of α and the performance of the system later in this section.

The training/test data used in the experiment contained over one thousand simple Japanese sentences collected from news articles. The examples given by IPAL were also used as training data[4]. Each of the sentences in the training/test data used in our experiment consisted of one or more complement(s) followed by one of the ten verbs enumerated in table 2. For each of the ten verbs, we conducted six-fold cross validation; that is, we divided the training/test data into six equal parts, and conducted six trials in each of which a different one of the six parts was used as test data and the rest was used as training data. We shall call the former the "test set" and the latter the "training set", in each case.

When more than one interpretation of an input verb is assigned the highest plausibility score, any of the above methods will choose as its output the one that appears most frequently in the training data. Therefore, the applicability in each method is 100%, given that the applicability is the ratio of the number of the cases where the system gives only one interpretation, to the number of inputs. Thus, in the experiment, we compared the precision of each method, which is in our case equal to the ratio of the number of correct outputs, to the number of inputs.

Since the performance of any corpus-based method depends on the size of training data, we first investigated how the precision of each method was improved as the training data increased. In this, we initially used only the examples given by IPAL, and progressively increased the size of the training data used, by considering an extra part of the training set (five parts of the total six data portions used) at each iteration, until finally taking all five parts in the training of our system. The results are shown in figure 5, in which the x-axis denotes the ratio of the data used from the training set, to the total size of the training set.

What can be derived from figure 5 are the following. First, as more training data was considered, the precision got higher for each method. Second, the consideration of CCD, i.e. case contribution to verb sense disambiguation, improved on Kurohashi's method regardless of the size of training data. Given the whole training set, the precision improved from 75.2% to 82.4% (7.2% gain). Third, the introduction of the notion of RSSR did not further improve

4 The number of examples given by IPAL was, on average, 3.7 for each case of each case frame.

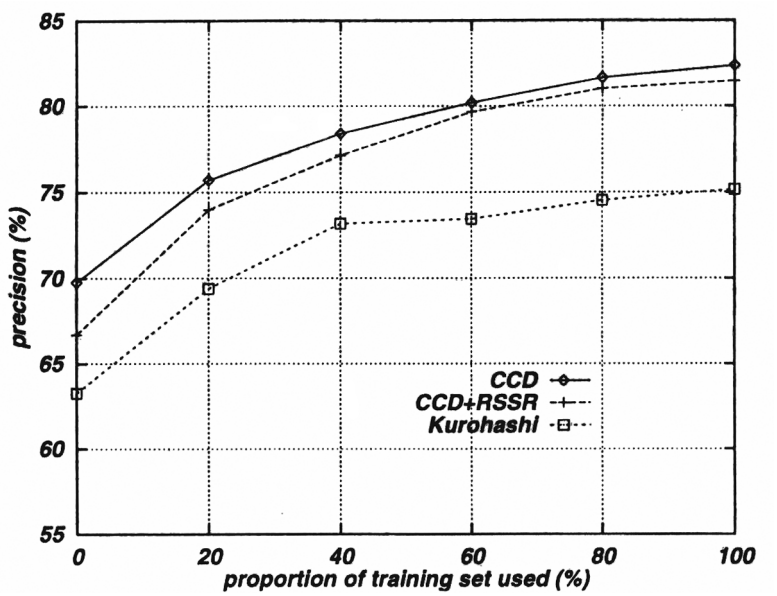

Fig. 5 The precision of each method, for each size of training data

on the method using only CCD.

Table 2 shows the performance for each verb on using the whole training set. The column of "lower bound" denotes the precision gained in a naive method such that the system always chooses the interpretation most frequently appearing in the training data (Gale, Church, & Yarowsky, 1992). The column of "two highest CCDs" gives the two highest CCD values from the cases for each verb, which are calculated using whole training set.

Finally, let us see to what extent we should allow CCD to influence verb sense disambiguation. Figure 6 shows the performance with the parametric constant α in equation (6) set to various values. $\alpha = 0$ corresponds with Kurohashi's method, in which CCD is never considered. As shown in figure 6, the stronger influence we allow CCD to have, the better performance we gain.

6 Conclusion

In this paper, we proposed a new example-based method for verb sense disambiguation, which improved the performance of the existing methods by considering the degree of case contribution to verb sense disambiguation.

The performance of our method significantly depends on the method of assigning degree of

Table 2 Performance for each verb
(*ga*: nominative, *ni*: dative, *wo*: accusative, *kara*: locative, *de*: instrumental)

verb	data size	# of candidates	lower bound (%)	two highest CCDs	precision (%) Kurohashi	precision (%) CCD
ataeru	136	4	66.9	*wo* (0.98) *ga* (0.86)	77.2	86.0
kakeru	160	29	25.6	*wo* (0.99) *ni* (0.98)	66.3	76.9
kuwaeru	167	5	53.9	*wo* (0.98) *ni* (0.95)	82.6	88.0
noru	126	10	45.2	*ni* (0.96) *ga* (0.92)	82.5	81.0
osameru	108	8	25.0	*wo* (0.95) *ni* (0.94)	73.2	70.4
tsukuru	126	15	19.8	*de* (1.0) *wo* (0.98)	59.2	84.9
toru	84	29	26.2	*kara* (1.0) *wo* (0.99)	56.0	71.4
umu	90	2	81.1	*wo* (1.0) *ga* (0.94)	100	98.9
wakaru	60	5	48.3	*ga* (0.96) *ni* (0.70)	65.0	70.0
yameru	54	2	59.3	*wo* (1.0) *de* (0.71)	96.3	96.3
total	1111	—	43.7	—	75.2	82.4

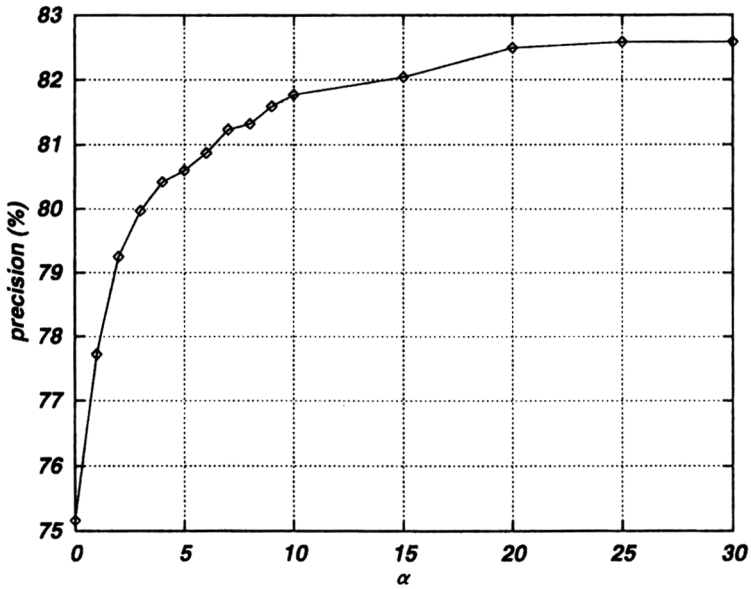

Fig. 6 The relation between the degree of CCD and precision

similarity to a pair of case fillers. Since *Bunruigoihyo* is fundamentally based on human intuition, it does not reflect the similarity between a pair of case fillers computationally. Proposed methods of word clustering (Tokunaga, Iwayama, & Tanaka, 1995, etc.) can potentially be used in conjunction with our method to overcome this human reliance.

In our current implementation, we consider the collocation between case fillers and verbs,

but ignore the combination of case fillers. Instead of a database as in figure 1, we could store a set of combinations of example case fillers, e.g. the combination of *suri* ("pickpocket") and *saifu* ("wallet"), but not that of *suri* and *otoko* ("man"). However, this way of data storage would require the collection of a much larger number of examples than the current method. This issue needs to be further investigated.

Acknowledgement

The authors would like to thank Dr. Manabu Okumura (JAIST, Japan), Dr. Michael Zock (LIMSI, France) and Mr. Timothy Baldwin (TITech, Japan) for their comments on the earlier version of this paper.

Reference

Brown, P. F., Pietra, S. A. D., & Pietra, V. J. D. (1991). Word-Sense Disambiguation Using Statistical Methods. In *Proceedings of ACL*, pp. 264–270.

Dagan, I., & Itai,A . (1994). Word Sense Disambiguation Using a Second Language Monolingual Corpus. *Computational Linguistics*, *20*(4), 563–596.

Gale, W., Church, K. W., & Yarowsky, D. (1992). Estimating Upper and Lower Bounds on the Performance of Word-Sense Disambiguation Programs. In *Proceedings of ACL*, pp. 249–256.

IPA (1987). *IPA Lexicon of the Japanese Language for computers IPAL (Basic Verbs).* (In Japanese).

Krovets, R., & Croft, W. B. (1992). Lexical Ambiguity and Information Retrieval *ACM Transactions on Information Systems*, *10*(2), 115–141.

Kurohashi, S., & Nagao, M. (1994). A Method of Case Structure Analysis for Japanese Sentences Based on Examples in Case Frame Dictionary. *IEICE TRANSACTIONS on Information and Systems*, *E77-D*(2), 227–239.

Lytinen, S. L. (1986). Dynamically Combining Syntax and Semantics in Natural Language Processing. In *Proceedings of AAAI*, pp. 574–578.

Nagao, K. (1994). A Preferential Constraint Satisfaction Technique for Natural Language Analysis. *IEICE TRANSACTIONS on Information and Systems*, *E77-D*(2), 161–170.

National-Language Research Institute (Ed.). (1964). *Bunruigoihyo.* Shuei publisher. (In Japanese).

Niwa, Y., & Nitta, Y. (1994). Co-occurrence Vectors from Corpora vs. Distance Vectors from Dictionaries. In *Proceedings of COLING*, pp. 304–309.

Schütze, H. (1992). Word Sense Disambiguation with Sublexical Representations. In *Workshop Notes, Statistically-Based NLP Techniques, AAAI*, pp. 109–113.

Tokunaga, T., Iwayama, M., and Tanaka, H. (1995). Automatic Thesaurus Construction Based on Grammatical Relations. In *Proceedings of IJCAI*, pp. 1308–1313.

Uramoto, N. (1994). Example-Based Word-Sense Disambiguation. *IEICE TRANSACTIONS on Information and Systems*, *E77-D*(2), 240–246.

Voorhees, E. M. (1993). Using WordNet to Disambiguate Word Senses for Text Retrieval. In *Proceedings of ACM SIGIR*, pp. 171–180.

Yarowsky, D. (1995). Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. In *Proceedings of ACL*, pp. 189–196.

Atsushi Fujii: He is a Ph.D. student of Department of Computer Science, Tokyo Institute of Technology. He received the B.S. and the M.S. degrees from Tokyo Institute of Technology in 1993 and 1995, respectively. His current interest is in natural language processing.

Kentaro Inui: He is a research associate of Department of Computer Science, Tokyo Institute of Technology. He received the B.S. degree in 1990, the M.S. degree in 1992, and the Dr. Eng. degree in 1995 from Tokyo Institute of Technology. His current interest is in natural language processing.

Takenobu Tokunaga: He is an associate professor of Department of Computer Science, Tokyo Institute of Technology. He received the B.S. degree in 1983 from Tokyo Institute of Technology, the M.S. and the Dr. Eng. degrees from Tokyo Institute of Technology in 1985 and 1991, respectively. His current interests are natural language processing, information retrieval.

Hozumi Tanaka: He is a professor of Department of Computer Science, Tokyo Institute of Technology. He received the B.S. degree in 1964 and the M.S. degree in 1966 from Tokyo Institute of Technology. In 1966 he joined in the Electro Technical Laboratories, Tsukuba. He received the Dr. Eng. degree in 1980. He joined in Tokyo Institute of Technology in 1983. He has been engaged in artificial intelligence and natural language processing research.

(Received September 10, 1996)

(Accepted November 11, 1996)