Integration of Hand-Crafted and Statistical Resources in Measuring Word Similarity

Atsushi FUJII Toshihiro HASEGAWA Takenobu TOKUNAGA Hozumi TANAKA
Department of Computer Science
Tokyo Institute of Technology
{fujii,take,tanaka}@cs.titech.ac.jp

Abstract

This paper proposes a new approach for word similarity measurement. The statistics-based computation of word similarity has been popular in recent research, but is associated with a significant computational cost. On the other hand, the use of hand-crafted thesauri as semantic resources is simple to implement, but lacks mathematical rigor. To integrate the advantages of these two approaches, we aim at calculating a statistical weight for each branch of a thesaurus, so that we can measure word similarity simply based on the length of the path between two words in the thesaurus. Our experiment on Japanese nouns shows that this framework upheld the inequality of statistics-based word similarity with an accuracy of more than 70%. We also report on the effectivity of our framework in the task of word sense disambiguation.

1 Introduction

This paper proposes a new approach for word similarity measurement, as has been variously used in such NLP applications as smoothing [Dagan *et al.*, 1994; Grishman and Sterling, 1994] and word clustering [Charniak, 1993; Hindle, 1990; Pereira *et al.*, 1993; Tokunaga *et al.*, 1995].

Previous methods for word similarity measurement can be divided into two categories: statistics-based approaches and hand-crafted thesaurus-based approaches. In statistics-based approaches, and namely the "vector space model", each word is generally represented by a vector consisting of co-occurrence statistics (such as frequency) with respect to other words [Charniak, 1993]. The similarity between two given words is then computationally measured using two vectors representing those words. One typical implementation computes the relative similarity as the cosine of the angle between two vectors, a method which is also commonly used in information retrieval and text categorization systems to measure the similarity between documents [Frankes and Baeza-Yates, 1992]. Since it is based on mathematical methods, this type of similarity measurement has been popular. Besides this, since the similarity is computed based on given co-occurrence data, word similarity can easily be adjusted according to the domain. However, data sparseness is an inherent problem. This fact was observed in our preliminary experiment, despite using statistical information taken from news articles as many as 4 years. Furthermore, in this approach, vectors require $O(N^2)$ memory space, given that N is the number of words, and therefore, large data sizes can prove prohibitive. Note that even if one statically stores possible word similarity combinations, $O(N^2)$ space is required.

The other category of word similarity approaches uses semantic resources, that is, hand-crafted thesauri (such as the Roget's thesaurus [Chapman, 1984] or WordNet [Miller *et al.*, 1993] in the case of English, and *Bunruigoihyo* [National Language Research Institute, 1996] or EDR [EDR, 1995] in the case of Japanese), based on the intuitively feasible assumption that words located near each other within the structure of a thesaurus have similar meaning. Therefore, the similarity between two given words is represented by the length of the path between them in the thesaurus structure [Kurohashi and Nagao, 1994; Li *et al.*, 1995; Uramoto, 1994]. Unlike the former approach, the required memory space can be restricted to $O(N)$ because only a list of semantic codes for each word is required. For example, the commonly used Japanese *Bunruigoihyo* thesaurus [National Language Research Institute, 1996] represents each semantic code with only 8 digits. However, computationally speaking, the relation between the similarity (namely the semantic length of the path), and the physical length of the path is not clear[1]. Furthermore, since most thesauri aim at a *general* word hierarchy, the similarity between words used in specific domains (technical terms) cannot be measured to the desired level of accuracy.

[1]Most researchers heuristically define functions between the similarity and physical path length [Kurohashi and Nagao, 1994; Li *et al.*, 1995; Uramoto, 1994].

In this paper, we aim at intergrating the advantages of the two above methodological types, or more precisely, realizing statistics-based word similarity based on the length of the thesaurus path. The crucial concern in this process is how to determine the statistics-based length of each branch in a thesaurus. We tentatively use the *Bunruigoihyo* thesaurus, in which each word corresponds to a leaf in the tree structure. Let us take figure 1, which shows a fragment of the thesaurus. In this figure, w_i's denote words and x_i's denote the statistics-based length (SBL, for short) of each branch i. Let the statistics-based (vector space model) word similarity between w_1 and w_2 be $vsm(w_1, w_2)$. We hope to estimate this similarity by the length of the path through branches 3 and 4, and derive an equation "$x_3 + x_4 = sim(w_1, w_2)$". Intuitively speaking, any combination of x_3 and x_4 which satisfies this equation can constitute the SBLs for branches 3 and 4. Formalizing equations for other pairs of words in the same manner, we can derive the simultaneous equation shown in figure 2. That is, we can assign the SBL for each branch by way of finding answers for each x_i. This method is expected to excel in the following aspects.

First, this method allows us to measure the statistics-based word similarity, while retaining the optimal required memory space ($O(N)$). One may argue that statistics-based automatic thesaurus construction (for example, the method proposed by Tokunaga et al. [Tokunaga *et al.*, 1995]) can provide the same advantage, besides which there is no human overhead. However, it has been empirically observed that the topology of the structure (especially at higher levels) is not necessarily reasonable when based solely on statistics [Frankes and Baeza-Yates, 1992]. To avoid this problem, we would like to introduce hand-crafted thesauri into our framework because the topology (such as MAMMAL is a hyper class of HUMAN) allows for higher levels of sophistication based on human knowledge.

Second, since each SBL reflects the statistics taken from co-occurrence data of the *whole* word set, statistics of each word can complement each other, and thus, the data sparseness problem tends to be minimized. Let us take figure 1 again, and assume that the statistics for w_4 are sparse or completely missing. In previous statistics-based approaches, the similarity between w_4 and other words cannot be reasonably measured, or not measured at all. However, in our method, similarity value such as $vsm(w_1, w_4)$ can be reasonably measured because SBLs x_1, x_2 and x_3 can be well-defined with sufficient statistics.

In section 2, we elaborate on the methodology of our word similarity measurement. We then evaluate our method by way of an experiment in section 3 and applied this method to the task of word sense disambiguation in section 4.

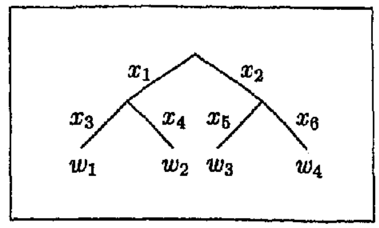

Figure 1: A fragment of the thesaurus

$$
\begin{aligned}
x_1 + x_2 + x_3 + x_5 &= vsm(w_1, w_3) \\
x_1 + x_2 + x_3 + x_6 &= vsm(w_1, w_4) \\
x_1 + x_2 + x_4 + x_5 &= vsm(w_2, w_3) \\
&\vdots
\end{aligned}
$$

Figure 2: A fragment of the simultaneous equation associated with figure 1

2 Methodology

2.1 Overview

Our word similarity measurement proceeds in the following way:

1. compute the statistics-based similarity of every combination of given words,
2. set up a simultaneous equation through use of the thesaurus and previously computed word similarity, and find solutions for the statistics-based length (SBL) of the corresponding thesaurus branch (see figures 1 and 2),
3. the similarity between two given words is measured by the sum of SBLs included in the path between those words.

We will elaborate on each step in the following sections.

2.2 Statistics-based word similarity

In the vector space model, each word w_i is represented by a vector comprising statistical factors of co-occurrence. This can be expressed by equation (1), where $\vec{w_i}$ is the vector for the word in question, and t_{ij} is the co-occurrence statistics of w_i and w_j.

$$\vec{w_i} = < t_{i1},\ t_{i2},\ \ldots,\ t_{ij},\ \ldots > \tag{1}$$

With regard to t_{ij}, we adopted TF·IDF, commonly used in information retrieval systems [Frankes and Baeza-Yates, 1992]. Based on this notion, t_{ij} is calculated as in equation (2), where f_{ij} is the frequency of w_i collocating with w_j, f_j is the frequency of w_j, and T is the total number of collocations within the overall co-occurrence data.

$$t_{ij} = f_{ij} \cdot \log\left(\frac{T}{f_j}\right) \tag{2}$$

We then compute the similarity between words a and b by the cosine of the angle between the two vectors $\vec{a}$ and $\vec{b}$. This is realized by equation (3), where vsm is the similarity between a and b, based on the vector space model.

$$vsm(a,b) = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|} \tag{3}$$

It should be noted that our framework is independent of the implementation of the similarity computation, which has been variously proposed by different researchers [Charniak, 1993; Frankes and Baeza-Yates, 1992].

2.3 Resolution of the simultaneous equation

The simultaneous equation used in our method is expressed by equation (4), where A is a matrix comprising only the values 0 and 1, and B is a list of vsm's (see equation (3)) for any possible combinations of given words. X is a list of variables, which represents the statistics-based length (SBL) for the corresponding branch in the thesaurus.

$$AX = B \tag{4}$$

Here, let the i-th similarity in B be $vsm(a,b)$, and let $path(a,b)$ denote the path between words a and b in the thesaurus. Each equation contained in the simultaneous equation is represented by equation (5), where x_j is the statistics-based length (SBL) for branch j, and α_{ij} is either 0 or 1 as in equation (6).

$$[\alpha_{i1}\ \alpha_{i2}\ \ldots\ \alpha_{ij}\ \ldots]\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_j \\ \vdots \end{bmatrix} = vsm(a,b) \tag{5}$$

$$\alpha_{ij} = \begin{cases} 1 & \text{if } j \in path(a,b) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

By finding the solutions for X, we can assign SBLs to branches. However, the set of similarity values outnumbers the variables. For example, the *Bunruigoihyo* thesaurus contains about 55,000 noun entries, and therefore, the number of similarity values for those nouns becomes about 1.5×10^9 (${}_{55,000}C_2$). On the other hand, the number of the branches is only about 53,000. As such, overly many equations are redundant, and the time complexity to solve the simultaneous equation becomes a crucial problem. To counter this problem, we randomly divide the overall equation set into equal parts, which can be solved reasonably. Thereafter we approximate the solution for x by averaging the solutions for x derived from each subset. Let us take figure 3, in which the number of subsets is given as two without loss of generality. In this figure, x_{i1} and x_{i2} denote the answers for branch i individually derived from subsets 1 and 2, and x_i is approximated by the average of x_{i1} and x_{i2} (that is, $\frac{x_{i1}+x_{i2}}{2}$). To generalize this notion, let x_{ij} denote the solution associated with branch i in subset j. The approximate solution for branch i is given by equation (7), where n is the number of divisions of the equation set.

$$x_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \tag{7}$$

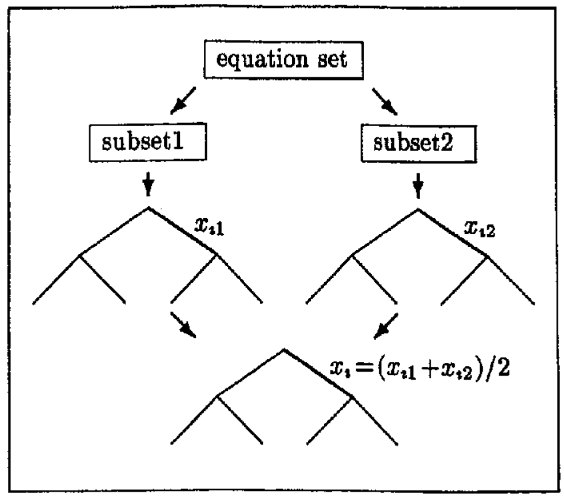

Figure 3: Approximation of the statistics-based length x_i

2.4 Word similarity using SBL

Let us reconsider figure 1. In this figure, the similarity between w_1 and w_2, for example, is measured by the sum of x_3 and x_4. In general, the similarity between words a and b using SBL ($sbl(a,b)$, hereafter) is realized by equation (8), where x_i is the SBL for branch i, and $path(a,b)$ is the path that includes thesaurus branches located between a and b.

$$sbl(a,b) = \sum_{i\in path(a,b)} x_i \tag{8}$$

3 Experimentation

We conducted experiments on noun entries in the *Bunruigoihyo* thesaurus. Co-occurrence data was extracted from the RWC text base RWC-DB-TEXT-95-1 [Real World Computing Partnership, 1995]. This text base consists of 4 years worth of Mainichi Shimbun [Mainichi

Shimbun, 1991-1994] newspaper articles, which were automatically annotated with morphological tags. The total number of morphemes is about 100 million. Instead of conducting full parsing on the texts, several heuristics were used in order to obtain dependencies between nouns and verbs in the form of tuples (frequency, noun, postposition, verb). Among these tuples, only those which included the postposition *wo* (typically marking the accusative case) were used. Further, tuples with nouns appearing in the *Bunruigoihyo* thesaurus were selected. When the noun comprised a compound noun, it was transformed into the maximal leftmost substring contained in the *Bunruigoihyo* thesaurus. As a result, 419,132 tuples remained, consisting of 23,223 noun types and 9,151 verb types. In regard to resolving the simultaneous equations, we used the mathematical analysis tool "MATLAB"[2].

What we evaluated here is the degree to which the simultaneous equation was successfully approximated through the use of the technique described in section 2. In other words, to what extent the (original) statistics-based word similarity can be realized by our framework. We conducted this evaluation in the following way. Let the statistics-based similarity between words a and b be $vsm(a,b)$, and the similarity based on SBL be $sbl(a,b)$. Here, let us assume the inequality "$vsm(a,b) > vsm(c,d)$" for words a, b, c and d. If this inequality can be maintained for our method, that is, "$sbl(a,b) > sbl(c,d)$", the similarity measurement is taken to be successful. The accuracy is then estimated by the ratio between the number of successful measurements and the total number of trials. Since resolution of equations is time-consuming, we tentatively generalized 23,223 nouns into 303 semantic classes (represented by the first 4 digits of the semantic code given in the *Bunruigoihyo* thesaurus), reducing the total number of equations to 45,753. Figure 4 shows the relation between the number of equations used and the accuracy: we divided the overall equation set into n equal subsets[3] (see section 2.3), and progressively increased the number of subsets used in the computation. When the whole set of equations was provided, the accuracy became about 72%. We also estimated the lower bound of this evaluation, that is, we also conducted the same trials using the *Bunruigoihyo* thesaurus. In this case, if word a is more closely located to b than c is to d and "$vsm(a,b) > vsm(c,d)$", that trial measurement is taken to be successful. We found that the lower bound was roughly 56%, and therefore, our framework outperformed this method.

[2]Cybernet System, Inc.

[3]We arbitrarily set $n = 15$ so as to be able to resolve equations reasonably.

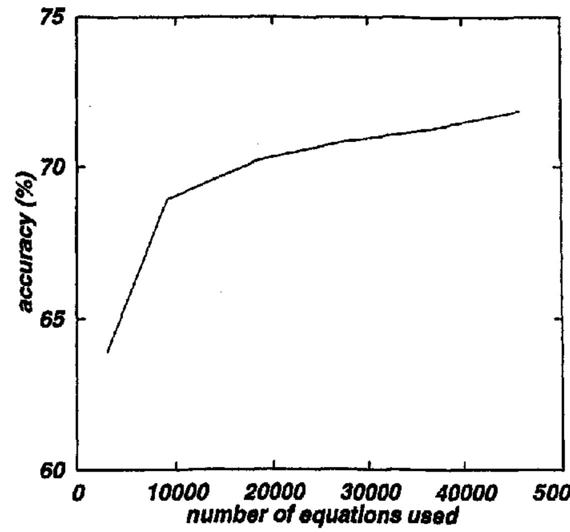

Figure 4: The relation between the number of equations used and the accuracy

4 An application

We further evaluated our word similarity technique in the task of word sense disambiguation (WSD). In this task, the system is inputted with sentences containing sense ambiguous words, and interprets them by choosing the most plausible meaning for them based on the context[4]. The WSD technique used in this paper has been proposed by Kurohashi et al. [Kurohashi and Nagao, 1994] and enhanced by Fujii et al. [Fujii *et al.*, 1996], and disambiguates Japanese sense ambiguous verbs by use of an example-database[5]. Figure 5 shows a fragment of the database associated with the Japanese verb *tsukau*, some of which senses are "to employ", "to operate" and "to spend". The database specifies the case frame(s) associated with each verb sense. In Japanese, a complement of a verb consists of a noun phrase (case filler) and its case marker suffix, for example *ga* (nominative), *ni* (dative) or *wo* (accusative). The database lists several case filler examples for each case. Given an input, the system identifies the verb sense on the basis of the similarity between the input and examples for each verb sense contained in the database. Let us take the following input:

enjinia ga	*fakkusu wo*	*tsukau.*
(engineer-NOM)	(facsimile-ACC)	(?)

In this example, one may consider *enjinia* ("engineer") and *fakkusu* ("facsimile") to be semantically similar to

[4]In most WSD systems, candidates of word sense are predefined in a dictionary.

[5]There have been different approaches proposed for this task, based on statistics [Charniak, 1993].

gakusei ("student") and *konpyuutaa* ("computer"), respectively, from the "to operate" sense of *tsukau*. As a result, *tsukau* is interpreted as "to operate". To formalize this notion, the system computes the plausibility score for each verb sense candidate, and chooses the sense that maximizes the score. The score is computed by considering the weighted average of the similarity of the input case fillers with respect to each of the corresponding example case fillers listed in the database for the sense under evaluation. Formally, this is expressed by equation (9), where $Score(s)$ is the score for verb sense s. n_c denotes the case filler for case c, and $\mathcal{E}_{s,c}$ denotes a set of case filler examples for each case c of sense s (for example, $\mathcal{E} = \{kare, kigyou\}$ for the *ga* case in the "to employ" sense in figure 5). $sim(n_c, e)$ stands for the similarity between n_c and an example case filler e.

$$Score(s) = \sum_{c} CCD(c) \cdot \max_{e \in \mathcal{E}_{s,c}} sim(n_c, e) \qquad (9)$$

$CCD(c)$ expresses the weight factor of case c using the notion of case contribution to verb sense disambiguation (CCD) proposed by Fujii et al [Fujii *et al.*, 1996]. Intuitively, the CCD of a case becomes greater when example sets of the case fillers are disjunctive over different verb senses. In the case fillers of figure 5, for example, $CCD(ACC)$ is greater than $CCD(NOM)$ (see Fujii et al's paper for details).

One may notice that the critical content of this task is the computation of the similarity between case fillers (nouns) in equation (9). This is exactly where our word similarity measurement can be applied. In this experiment, we compared the following three methods for word similarity measure:

- the *Bunruigoihyo* thesaurus (BGH): the similarity between case fillers is measured by a function between the length of the path and the similarity. In this experiment, we used the function proposed by Kurohashi et al. [Kurohashi and Nagao, 1994] as shown in table 1.
- vector space model (VSM): we replace $sim(n_c, e)$ equation (9) with $vsm(n_c, e)$ computed by equation (3)
- our method base on statistics-based length (SBL): we simply replace $sim(n_c, e)$ in equation (9) with $sbl(n_c, e)$ computed by equation (8).

We collected sentences (as test/training data) from the EDR Japanese corpus [EDR, 1995][6]. Since Japanese sentences have no lexical segmentation, the input has to be both morphologically and syntactically analyzed prior to the sense disambiguation process. We experimentally used the Japanese morph/syntax parser "QJP" [Kameda, 1996] for this process. Based on analysis by the QJP parser, we removed sentences with missing verb complements (in most cases, due to ellipsis or zero anaphora). The EDR corpus also provides sense information for each The EDR corpus provides sense information for each word based on the EDR dictionary, which we used as a means of checking the correct interpretation. Our derived corpus contains ten verbs frequently appearing in the EDR corpus, which are summarized in table 2. In table 2, the column of "English gloss" describes typical English translations of the Japanese verbs. the column of "# of sentences" denotes the number of sentences in the corpus, while "# of senses" denotes the number of verb senses, based on the EDR dictionary. For each of the ten verbs, we conducted four-fold cross validation: that is, we divided the corpus into four equal parts, and conducted four trials, in each of which a different one of the four parts was used as test data and the remaining parts were used as training data (the database). Table 2 also shows the precision of each method. The precision is the ratio of the number of correct interpretations, to the number of outputs. The column of "control" denotes the precision of a naive WSD technique, in which the system systematically chooses the verb sense appearing most frequently in the database [Gale *et al.*, 1992].

[6]The EDR corpus was originally collected from news articles.

Table 1: The relation between the length of the path between two nouns n_1 and n_2 in the *Bunruigoihyo* thesaurus ($len(n_1, n_2)$) and their similarity ($sim(n_1, n_2)$)

$len(n_1, n_2)$	0	2	4	6	8	10	12	14
$sim(n_1, n_2)$	12	11	10	9	8	7	5	0

The precision for each similarity calculation method did not differ greatly, and the use of the length of the path in the *Bunruigoihyo* thesaurus (BGH) slightly outperformed other method on the whole. However, since the overall precision is biased by frequently appeared verbs (such as *tsukau* and *ukeru*), our word similarity measurement is not necessarily inferior to other methods. In fact, disambiguation of verbs such as *motomeru*, in which BGH is surpassed by VSM, SBL maintains a precision level relatively equivalent to that for VSM. Besides this, as we pointed out in section 1, SBL allows us to reduce the data size from $O(N^2)$ to $O(N)$ in our framework, given that N is the number of word entries.

5 Conclusion

In this paper, we proposed a new method for the measurement of word similarity. Our method integrates the statistics-based and thesaurus-based approaches. By this, we can realize the statistical computation of word similarity based on a thesaurus, with optimal computation cost. We showed the effectivity of our method by

kare (he) *kigyou* (company)	*ga*	*kikaku* (project)	*ni*	*jyuugyouin* (employee) *sotsugyousei* (graduate)	*wo*	*tsukau* (to employ)
kanojo (she) *gakusei* (student)	*ga*	*shigoto* (work) *kenkyuu* (research)	*ni*	*konpyuutaa* (computer) *kikai* (machine)	*wo*	*tsukau* (to operate)
kare (he) *seifu* (government)	*ga*	*kuruma* (car) *fukushi* (welfare)	*ni*	*nenryou* (fuel) *shigen* (resource) *zeikin* (tax)	*wo*	*tsukau* (to spend)

Figure 5: A fragment of the database associated with the Japanese verb *tsukau*

Table 2: Precision of word sense disambiguation (the highest precision is typed in boldface)

verb	English gloss	# of sentences	# of senses	precision (%) BGH	VSM	SBL	control
tsukau	spend	1729	7	**58.8**	55.0	52.8	27.8
ukeru	receive	1573	10	80.2	**80.9**	75.5	38.4
motsu	hold	1471	12	**72.1**	70.1	71.3	37.5
miru	see	1096	17	49.1	46.5	**49.8**	22.7
motomeru	request	1025	5	67.4	**71.4**	71.0	48.8
dasu	evict	872	5	**65.9**	63.4	63.4	42.3
kuwaeru	add	467	4	68.7	67.7	**69.4**	58.5
okuru	send	387	9	**58.4**	56.8	**58.4**	28.9
kaku	write	382	2	**74.5**	73.0	73.3	48.7
moukeru	establish	343	3	**67.1**	65.6	64.7	51.0
total	—	9345	—	**66.4**	65.2	64.5	37.4

way of an experiment, and demonstrated its application to word sense disambiguation. Future work will include how to decrease the number of equations without degrading the performance, and application of our framework to other NLP tasks for the further evaluation.

Acknowledgments

The authors would like to thank Mr. Timothy Baldwin (TITECH, Japan) for his comments on the earlier version of this paper, Mr. Masayuki Kameda (RICOH Co., Ltd., Japan) for his support with the QJP parser, and Mr. Akira Hirabayashi and Mr. Naoyuki Sakurai (TITECH, Japan) for aiding with experiments.

References

[Chapman, 1984] Chapman, R. L. *Roget's International Thesaurus (Fourth Edition).* Harper and Row.

[Charniak, 1993] Charniak, E. *Statistical Language Learning.* MIT Press.

[Dagan *et al.*, 1994] Dagan, I., Pereira, F., and Lee, L. Similarity-based estimation of word cooccurrence probabilities. In *Proceedings of ACL*, pp. 272–278.

[EDR, 1995] EDR. *EDR Electronic Dictionary Technical Guide.* (In Japanese).

[Frankes and Baeza-Yates, 1992] Frankes, W. B., and Baeza-Yates, R. *Information Retrieval: Data Structure & Algorithms.* PTR Prentice-Hall.

[Fujii *et al.*, 1996] Fujii, A., Inui, K., Tokunaga, T., and Tanaka, H. To what extent does case contribute to verb sense disambiguation? In *Proceedings of COLING*, pp. 59–64.

[Gale *et al.*, 1992] Gale, W., Church, K. W., and Yarowsky, D. Estimating upper and lower bounds on the performance of word-sense disambiguation programs. In *Proceedings of ACL*, pp. 249–256.

[Grishman and Sterling, 1994] Grishman, R., and Sterling, J. Generalizing automatically generated selectional patterns. In *Proceedings of COLING*, pp. 742–747.

[Hindle, 1990] Hindle, D. Noun classification from predicate-argument structures. In *Proceedings of ACL*, pp. 268–275.

[Kameda, 1996] Kameda, M. A portable & quick Japanese parser : QJP. In *Proceedings of COLING*, pp. 616–621.

[Kurohashi and Nagao, 1994] Kurohashi, S., and Nagao, M. 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.

[Li *et al.*, 1995] Li, X., Szpakowicz, S., and Matwin, S. A WordNet-based algorithm for word sense disambiguation. In *Proceedings of IJCAI*, pp. 1368–1374.

[Miller *et al.*, 1993] Miller, G. A., Bechwith, R., Fellbaum, C., Gross, D., Miller, K., and Tengi, R. Five Papers on WordNet. Tech. rep. CSL Report 43, Cognitive Science Laboratory, Princeton University. Revised version.

[National Language Research Institute, 1996] National Language Research Institute. *Bunruigoihyo* (revised and enlarged edition). (In Japanese).

[Pereira *et al.*, 1993] Pereira, F., Tishby, N., and Lee, L. Distributional clustering of English words. In *Proceedings of ACL*, pp. 183–190.

[Real World Computing Partnership, 1995] Real World Computing Partnership. RWC text database. `http://www.rwcp.or.jp/wswg.html`.

[Mainichi Shimbun, 1991-1994] Mainichi Shimbun CD-ROM '91-'94.

[Tokunaga *et al.*, 1995] Tokunaga, T., Iwayama, M., and Tanaka, H. Automatic thesaurus construction based on grammatical relations. In *Proceedings of IJCAI*, pp. 1308–1313.

[Uramoto, 1994] Uramoto, N. Example-based word-sense disambiguation. *IEICE TRANSACTIONS on Information and Systems*, *E77-D*(2), 240–246.