Preferential Presentation of Japanese Near-Synonyms Using Definition Statements

Hiroyuki OKAMOTO **Kengo SATO** **Hiroaki SAITO**
Department of Information and Computer Science
Keio University
3–14–1 Hiyoshi, Kouhoku-ku, Yokohama 223–8522, Japan
Tel: (+81–45)563–1151 (ex 43250), Fax: (+81–45)566–1747
{motch, satoken, hxs}@nak.ics.keio.ac.jp

Abstract

This paper proposes a new method of ranking near-synonyms ordered by their suitability of nuances in a particular context. Our method distincts near-synonyms by semantic features extracted from their definition statements in an ordinary dictionary, and ranks them by the types of features and a particular context. Our method is an initial step to achieve a semantic paraphrase system for authoring support.

1 Introduction

Most researches on automatic paraphrasing aim either at document modification for a wide range of NLP applications (Shirai et al., 1998; Tomuro and Lytinen, 2001), at reading comprehension support (Inui and Yamamoto, 2001), or at transformation based on external constraints (Dras, 1998). On the other hand, authoring / revision support is known as another type of paraphrasing which targets at texts in preparation. However, there are not so many researches of such paraphrasing.

Paraphrase systems which aim at revising documents can be classified into three types:

- **Syntactic suitability**
  This type of systems points out spelling or grammatical mistakes and corrects them, such as a grammar checker (Heidorn, 2000).
- **Readability**
  Similar to reading comprehension support, this type of paraphrase systems aims to simplify difficult / complicated sentences or phrases (Suganuma et al., 1990; Inui and Okada, 2000).
- **Semantic suitability**
  To reflect authors' intentions precisely, these paraphrase systems replace words, which are semantically ambiguous or inadequate, to ones which are suitable for their contexts.

Almost all known authoring / revision support systems aim at syntactic suitability or readability, while researches of the third type of paraphrasing, which handle semantics, are very rare.

Let us consider a kind of authoring support system, which first presents **near-synonyms** (words counted among the same semantic category) of a target word in an input sentence. Then, based on user's choise, the system paraphrases the target word to the selected one with keeping syntactic and semantic consistency through paraphrasing. Especially for semantic consistency, it is important to express semantic differences between paraphrased word pairs clearly. If fine-grained meanings of all near-synonyms (not only a paraphrased pair) can be extracted at a time, the system would be able to present semantically suitable near-synonyms. Based on this idea, this paper proposes a new method of ranking Japanese near-synonyms ordered by their suitability of nuances in a particular context. First, this paper describes an overview of the method in Section 2. Next, Section 3 shows the classification of fine-grained meanings of a word and a method of extracting those fine-grained meanings from a definition statement of the word, to identify semantic differences between near-synonyms. Then, Section 4 presents our method of ranking near-synonyms using fine-grained meanings described in Section 3. Finally, this paper shows conclusion and further works in Section 5.

2 Overview of our method of preferential presentation

Though some word processing applications (e.g. Microsoft Word) have a function of showing near-synonyms of a word, it is not easy to choose the most adequate word from the near-synonyms because they are not ordered by their semantic similarity or suitability. Also, a simple replacement from a word to one of its near-synonyms is very dangerous, because there are some differences between the words in their modification rules and in their fine-grained meanings.

Against these semantic problems, we propose a new method of presenting near-synonyms ordered by their semantic suitability in a particular context. When a target word is given from an input sentence, first our method obtains all near-synonyms of the target word from an existing thesaurus, and differentiates them semantically by features extracted from their definition statements. Next, our method ranks those near-synonyms by relations between the type of features and the context of the input sentence. Finally, the ranking of near-synonyms are presented with information of variation in the original sentence for each near-synonym. This process enables the user to choose a word suitable for the input context, and helps prevention of semantic variation (or redundancy / loss) in paraphrasing.

3 Semantic differentiation between near-synonyms

As the first step to realize the preferential suggestion of near-synonyms, we identify fine-grained word senses of near-synonyms in order to differentiate them semantically, by using sentences written in an ordinal dictionary (**definition statements**) and word co-occurrence information extracted from large corpora.

3.1 Fine-grained word senses

There are some researches which deal with fine-grained word senses for a lexical choice in language generation (DiMarco et al., 1993; Edmonds, 1999). Edmonds roughly classified semantic differences between near-synonyms into four categories: denotational (difference in nuances of near-synonyms), expressive (in attitudes or emotions), stylistic (in formalities or dialects), and collocational (as idioms or in co-occurrence restrictions). In addition, he classified them into 35 types and proposed an ontology for describing their differences formally.

Edmonds implemented I-Saurus, a prototype implementation of this ontology, to achieve a lexical choice in machine translation and denoted the effectiveness of differences between near-synonyms for a lexical choice. Though, there is a crucial problem that he did not mention how to obtain those differences automatically. Against this problem, our method extracts such differences by using definition statements for each near-synonym. Although (Fujita and Inui, 2001) has already focused on using definition statements in order to determine a pair of near-synonyms whether one can be paraphrased to the other or not, it was only a kind of matching between two statements and did not identify individual features in each statement. Therefore, this paper defines three types of semantic features as follows, which can be extracted from definition statements:

- **Core meaning** indicates the basic sense of a word. All near-synonyms in a category must always have the same core meaning, such as the name of the category which they belong to.
- **Denotation**, which can be paraphrased to 'nuance', is defined as "the thing that is actually described by a word rather than the feelings or ideas it suggests" in *Longman web dictionary*[1]. In this paper, this feature is defined as a meaning included in a word, which partially qualify the core meaning. It is similar to a denotational constraint in (Edmonds, 1999).
- **Lexical restriction** of a word is a constraint on the range of co-occurrence of the word. This feature is almost the same as a collocational constraint in (Edmonds, 1999).

An example of these features is shown in Figure 1.

We divide our method into two steps to extract each feature from a definition statement. First, we extract a word defined as a core meaning and all other content words (in Section 3.2). Then, the extracted words except the core meaning are classified into denotations or lexical restrictions by using each co-occurrence information obtained from large corpora (in Section 3.3).

[1] `http://www.longmanwebdict.com/`

Word: さいこん
再建 *saikon*
(rebuilding of shrines / temples)

Definition statement:
「神社・仏閣を建て直すこと。」
jinja (shrine) *bukkaku* (temple) *wo* (OBJ) *tate* (to build) *naosu* (to repair) *koto* (matter)
(To build a shrine or a temple to repair.

Core meaning:	建て *tate* (build)
Denotation:	直す *naosu* (repair)
Lexical restriction:	神社 *jinja* (shrine)
	仏閣 *bukkaku* (temple)

Figure 1: Features in a definition statement

3.2 Extraction of fine-grained word senses

In this paper, we assume that a definition statement of a word (hereafter an **entry**) in a dictionary consists of four types of materials as follows:

- **Core meaning** is a word which exactly describes a particular semantic category which the entry belongs to.
- **Fine-grained meaning** semantically differenciates the entry from its near-synonyms. It is defined as a core meaning of some content words in the definition statement. Fine-grained meaning can be divided into "denotation" or "lexical restriction".
- **Stop word** indicates a content word which commonly and frequently appears in any definition statement.
- **Others** include function words and symbols.

According to this assumption, the "core meaning" and "fine-grained meanings" of an entry are extracted from a definition statement, using of *Kadokawa thesaurus* (Ohno and Hamanishi, 1981)[2]. A procedure of this method is given as follows:

Step 1. For each morpheme in the morpheme dictionary of *ChaSen* (Matsumoto et al., 2002), a Japanese morphological analyzer, add a label of a semantic category in *Kadokawa Thesaurus*, which the morpheme belongs to.

Step 2. Assign semantic labels to each morpheme in a definition statement of an entry e, by applying *ChaSen* to the statement.

Step 3. Give a word c as a "provisional" core meaning if c is classified into the same semantic category as e.

Step 4. Extract all semantic labels, which are assigned to all content words except c, as fine-grained meanings.

Step 5. Recursively apply Step 2–4 to the definition statement of c until no core meaning is extracted from the definition statement.

Step 6. Define c extracted at last as the "true" core meaning of e.

According to this procedure, some fine-grained meanings could be extracted from stop words. Thus, we give a semantic weight to each fine-grained meaning, by the reciprocal of its occurrence probability in all definition statements. These weights can distinct true fine-grained meanings from ones extracted from stop words.

A result of this method is shown in Figure 2, where the bold numbers show their categories and the italics show their weights.

Word: **[394]** さいこん
再建 *saikon*
(rebuilding of shrines / temples)

Core meaning:
[394] 建てる *tateru* (to build)

Fine-grained meaning:
[727a] 神社 *jinja* (shrine: *5687*)
[940c] 仏閣 *bukkaku* (temple: *6184*)
[277b] 直す *naosu* (to alter: *1441*)
[277c] 直す *naosu* (to recover: *2359*)
[392] 直す *naosu* (to repair: *7494*)
[417a] 直す *naosu* (to get right: *3703*)
[811] こと *koto* (matter: *30*)

Figure 2: Example of extraction of core-meaning and fine-grained meanings

3.3 Classification of fine-grained word senses

After obtaining features in Section 3.2, our method classifies fine-grained meanings into denotations and lexical restrictions, according to the following heuristics:

- If a word w includes a denotation d, w seldom co-occurs with any word whose core meaning is d. For example, one possible paraphrase of a sentence

  He is extremely *angry*.

[2] *Kadokawa thesaurus* semantically categorizes 57,130 entries into 2,924 categories and each entry has a definition statement.

is

He is *enraged*.

where the word *extremely* is deleted, because *enraged* has a denotation "*extremely*" if *angry* is defined as the core meaning of *enraged*.

- If w involves a lexical restriction l, w often co-occurs with words whose core meaning is l. For example, "a *rancid* butter" is more appropriate than "a *rotten* butter", because *rancid* has a lexical restriction "*oily or fatty food*", while *rotten* does not.

Based on these heuristics, our method classifies fine-grained meanings of an entry as follows:

Step 1. Assign semantic labels to all words in corpora (consisting of 1.93 million sentences, including newspapers [3] and novels [4]).

Step 2. Obtain co-occurrence frequencies of all pairs between a word and a semantic label of a neighbor word from the corpora.

Step 3. Delete the entry e from the thesaurus if e does not appear in the corpora at all.

Step 4. For each fine-grained meaning f of e which belongs to a semantic category C, compute co-occurrence probabilities

$$P(f, C) = \frac{\sum_i n_{s_i f}}{\sum_i N_{s_i}} \quad (1)$$

$$P(f, e) = \frac{n_{ef}}{N_e} \quad (2)$$

where s_i is a near-synonym of e, n_{ab} is the co-occurrence frequency between a word a and a label b, and N_a is the frequency of a.

Step 5. Remove f if $P(f, C) = 0$.

Step 6. Define f as a denotation if $P(f, e) = 0$. The weight of the denotation is the product of $P(f, C)$ and the weight of f.

Step 7. Define f as a lexical restriction if $P(f, e) \neq 0$. The weight of the lexical restriction is the product of $\frac{P(f,e)}{P(f,C)}$ and the weight of f.

Figure 3 shows an example of classification about the word '*saikon* (再建)'. In Figure 3, under-lined features are the results of word sense disambiguation and elimination of stop words.

[3] *Mainichi Shimbun CD-ROM* `http://cl.aist-nara.ac.jp/lab/resource/cdrom/Mainichi/MS.html`

[4] *Aozora Bunko* `http://www.aozora.gr.jp/`

Word: **[394]** 再建（さいこん） *saikon*
(rebuilding of shrines / temples)

Denotation:
[277b] 直す *naosu* (to alter: *1.45*)
[392] 直す *naosu* (to repair: *4.19*)

Lexical restriction:
[727a] 神社 *jinja* (shrine: *8518*)
[940c] 仏閣 *bukkaku* (temple: *5859*)
[277c] 直す *naosu* (to recover: *3504*)
[417a] 直す *naosu* (to get right: *2135*)
[811] こと *koto* (matter: *15*)

Figure 3: Classification example of fine-grained meanings

3.4 Evaluation and discussions

We applied these procedures to all 57,130 entries in *Kadokawa thesaurus* (2,924 categories). As a result, 36,434 entries, which consist of one core meaning and 0 or more fine-grained meanings, and 1,857 entries, which has no core meaning but is refered as a core meaning to other entries, were obtained. One entry has 4.7 denotations and 5.1 lexical restrictions on average.

To evaluate our methods, we compared the results of automatic extraction against manually extracted ones for randomly selected 50 entries. Table 1 shows the result of extracting core meanings, and the result of the classification is shown in Table 2.

	number of entries
corrects	40
errors	10
(direct)	(4)
(indirect)	(6)
precision	80 %

Table 1: Result of extracting core meanings

Failure results of extractions of core meanings appeared in the following cases; a core meaning in a definition statement does not belong to the same semantic category as the entry; the correct core meaning involves negative expressions in a definition statement; or two or more near-synonyms are appeared in one definition statement. Therefore, the extraction of core meanings needs to be estimated without relying on their semantic categories, that is, with other information such as modification re-

		result		recall [%]
		denotation	lexical restriction	
answer	denotation	56	13	81.2
	lexical restriction	22	20	47.6
precision [%]		71.8	60.6	

Table 2: Result of classification

lations of a definition statement.

Table 2 shows that both the precision and the recall of the classification into lexical restrictions are worse than the ones of denotations. A sparse data problems could cause it. In our classification method, if a feature of an entry does not co-occur with the entry, the feature is classified into a denotation or deleted, even though it is expected to be defined as a lexical restriction. It would be improved by increasing domains and the size of corpora, or by using information of modification relations just as the extraction of core meanings.

4 Preferential presentation of near-synonyms

We secondly propose a method of ranking near-synonyms by using information derived in Section 3. Though (Edmonds, 1999) proposed a ranking method for lexical choice by using information of fine-grained meanings in I-Saurus, it requires more detailed information than the one which can be extracted from a definition statement. Thus, this paper proposes a ranking method as follows: when a target word in a sentence is given, our method obtains all near-synonyms[5] of the target word and their semantic features. Then, our method ranks the near-synonyms with respect to their suitability between the input context and features of each near-synonym. Additionally, if a paraphrase to a near-synonym causes neighbor words in the input sentence to arrange in order to keep semantic consistency, our method adds such information to the near-synonym when the ranking is presented.

4.1 Comparison between denotations and contexts

"Denotations" can appear in any word, including a target word in an input sentence. Therefore, all denotations of each near-synonym have to be compared not only with the input context but with denotations of a target word. Our method determines the propriety of paraphrasing between a target word w and its near-synonym s_i for each denotation d_{ij} of s_i, with the following cases:

Case 1. No denotation appears in neither w nor s_i:
$\Rightarrow$ w can be directly paraphrased to s_i.

Case 2. w has a denotation d_w equivalent to d_{ij}:
$\Rightarrow$ w can be paraphrased to s_i on the sense of d_{ij}.

Case 3. d_w does not match with any d_{ij}:
$\Rightarrow$ w can be paraphrased to s_i with adding d_w to the input sentence.

Case 4. d_{ij} does not match with any d_w:
- (a) if d_{ij} can be covered with a neighbor word w' of w in the input sentence:
$\Rightarrow$ w can paraphrase to s_i with deleting w' from the input sentence.
- (b) if d_{ij} can not be covered with any words in the input sentence:
$\Rightarrow$ w can not be paraphrased to s_i.

In Case 3 and Case 4a, some arrangements (addition / deletion of words) to the input sentence are needed. Our method presents these information with the presentation of near-synonyms rankings (in Section 4.3).

According to these cases, the total denotational score S_d of s_i is defined by

$$S_d = \sum_j pW_j \tag{3}$$

where W_j is the weight of d_ij (one of the denotations of s_i) and

$$p = \begin{cases} 1 & \text{(in Case 1, 2, 4a)} \\ 0 & \text{(in Case 3)} \\ -1 & \text{(in Case 4b)} \end{cases}$$

Note that Case 3 gives no weight, because the case does not consider any denotation of s_i but compares only between d_w and its context.

[5] There are sometimes two or more core meanings in one semantic category. We treat whole core meanings as the exactly same meaning here.

4.2 Comparison between lexical restrictions and contexts

"Lexical restriction", the other fine-grained meaning, is the feature which notably often co-occur with its target word, as described in Section 3.3. In fact, however, a word which often co-occurs with a target word does not have to belong exactly to one of the lexical restrictions of the target word. They could be the "similar" words. Therefore, it is necessary to compute the similarity between a lexical restriction and a context in order to compare them.

The thesaurus used in our method has a tree structure and each entry belongs to the node at 4 or 5 in depth. The similarity can be defined by a heuristic approach that any two words are semantically independent if the depth of their root node is less than 3, such as the categories between [588] "rebels" and [506] "private and public". Hence, our method defines the similarity between a lexical restriction v_i and a semantic label q_i of a word in an input context as follows:

$$sim(v_i, q_i) = \log_2 \left(\frac{dep\,(root(v_i, q_i)) \times 4}{dep(v_i) + dep(q_i)} \right) \quad (4)$$

where $root(a, b)$ is the root node of the minimum subtree which includes both a and b, and $dep(a)$ is the depth of a in the thesaurus.

To determine the score of a lexical restriction, there is another problem. An input sentence has several content words outside of the target word, and some of them belong to several semantic categories because of their ambiguities. Also, the target word often has two or more lexical restrictions. Thus, each lexical restriction must select a semantic label which has the highest similarity with the lexical restriction from the input sentence. Against the problem, first, our method computes the similarities of all possible pairs which consist of a lexical restriction and a semantic label extracted from the sentence. Then, our method extracts pairs in descending order of the similarity with no overlap in any category or any lexical restriction.

Based on this process, we can compute the total score S_v of each near-synonym s_i of a target word w in an input sentence, with all extracted pairs of a lexical restriction v_j and a semantic label q_j in the input sentence by

$$S_v = \sum_j \left(W_j \cdot sim(v_j, q_j)\right) \quad (5)$$

where W_j is the weight of v_j.

4.3 Ranking method

This section describes our method of ranking near-synonyms with respect to the scores defined in Section 4.2 and Section 4.1, which is the aim of this paper. The criterion of ranking is simply the sum of normalized S_d and S_v[6]. Our method presents near-synonyms according to their ranking, and if necessary, information of arrangements to an input sentence (extracted in Section 4.1) are shown with each near-synonym.

4.4 An example

When an input sentence is

「寺を建て直す。」
tera (joss house) *wo* (OBJ) *tate* (to build) *naosu* (to repair)
(*Someone* rebuilds a joss house.)

and the word "建て(る) (*tate(ru)*, to build)" is given as a target, the semantic labels assigned to each content word in the sentence are

寺	*tera*	[727b] temple	
建て	*tate*	[394] to build	
直す	*naosu*	[277b] to alter	[277c] to recover
		[392] to repair	[417a] to get right

and 24 near-synonyms of *tateru* are extracted. Then, our method computes S_d and S_v for each near-synonym. For example, the scores of a word "再建(さいこん) (*saikon*, rebuilding of shrines / temples)", which includes features shown in Figure 3, are given as follows:

- S_d (the denotational score)
  For the denotations of *saikon*, [277b] (to alter: *1.45*) and [392] (to repair: *4.19*) could be obtained, where the italic numbers show their weight. They match to the labels in the word *naosu*, thus S_d of *saikon* is 5.64 and the word *naosu* is given as a deletion information.

- S_v (the score in lexical restriction)
  For the lexical restrictions of *saikon*, [277c] (to recover: *3504*), [417a] (to get right: *2135*), [727a] (shrine: *8518*), [811] (matter: *15*) and [940c] (temple: *5859*) could be obtained, then the extracted pairs and their similarity are calculated as follows:

[6] Each score has to be normalized because the place of S_d far differs from that of S_v.

lexical restriction		context	similarity
[277c]	$\Leftrightarrow$	[277c]	1.00
[417a]	$\Leftrightarrow$	[417a]	1.00
[727a]	$\Leftrightarrow$	[727b]	0.68
[811]	$\Leftrightarrow$	[392]	−1.00
[940c]	$\Leftrightarrow$	[277b]	−1.32

Therefore, S_v of *saikon* is calculated as 3682.

Finally, by computing S_d and S_v of all the other near-synonyms, our method ranks the near-synonyms and presents them as shown in Figure 4.

In Figure 4, the first 9 near-synonyms can be paraphrased from the target word appropriately. However, *saikon* is ranked next to *fushin* contrary to our expectation that it would be ranked as the first, because *saikon* and the fifth word *saiken* has the same orthography, and thus the co-occurrence information of *saikon* is imprecise by mixture with the information of *saiken*.

4.5 Evaluation and discussions

To evaluate our ranking method, we randomly extracted 40 sentences from corpora and applied our method to a certain word in each sentence. Also, for each case, we manually selected all near-synonyms which can be paraphrased[7]. We evaluated the ranking results of our method by the measure of non-interpolated average precision (NAP):

$$NAP = \frac{1}{R}\sum_{i=1}^{n}\frac{z_i}{i}\left(1+\sum_{k=1}^{i-1}z_k\right) \qquad (6)$$

where R is the number of near-synonyms which can be paraphrased, n is the number of presented near-synonyms, and

$$z_i = \begin{cases} 1 & \text{if a near synonym in rank } i \text{ can be paraphrased} \\ 0 & \text{otherwize} \end{cases}$$

Table 3 shows the result.

Table 3 shows that our method is remarkably effective for the judgement of semantic suitability of near-synonyms if a target word is not ambiguous. However, the average precision is worse for ambiguous words, thus it is important to disambiguate those target words before applying to our method.

[7] For the criterion if a word can paraphrase to another or not, we dissemble any addition / deletion informations. That is, we assume that a word can paraphrase if the paraphrased sentence has the same meaning as the original with some changes to their context.

ambiguity of target word (sentences)	NAP [%]			
	our method			non-ordered
	S_d	S_v	$S_d + S_v$	
distinct (21)	74.2	63.8	71.2	60.0
vague (19)	48.8	48.3	51.0	42.1
both (40)	62.8	56.9	62.2	52.0

Table 3: Average precision of ranking

Most of failure results are caused by the following cases; incorrect core meanings or fine-grained meanings were extracted in Section 3; adequate relations between a near-synonym and an input context could not be identified because of the ambiguity of neighbor words in the input sentence; or the semantic range of the label of a denotation or a lexical restriction is too wide to express the fine-grained meaning of the near-synonym clearly.

In addition, Table 3 shows that the average precision by only S_v is worse than the one by only S_d. It could be caused by the low precision of classification into lexical restrictions and by the inadequacy in the measure of similarity described in Section 4.2. To improve those problems, another measure such as semantical similarities without using a structure of a thesaurus is needed. Also, we would learn from a method of lexical choice with knowledge about collocational behavior (Inkpen and Hirst, 2002).

Though we have not discussed the evaluation of the propriety of arrangements to an input sentence, it seems that the information of addition often occurs imprecisely, against that the information of deletion appears infrequently but almost correctly, because, in our method, all denotations of a target word are given as the information of addition when they do not match with any denotation of a near-synonym. Therefore, we must define the importance of each addition information and to present selected ones.

5 Conclusion and future work

This paper proposed a new method of preferential presentation of Japanese near-synonyms in order to treat with semantic suitability against contexts, as a first step of semantic paraphrase system for elaboration. We achieved the effectiveness of using definition statements for extracting fine-grained meanings, especially for denotations. Also, the experimental results showed that our method could rank near-synonyms of an unambiguous word for 71%

1.	普請 *fushin* (削除: 直す) (delete *naosu*) (Construct or repair a house / a temple / a road)	6.	築造 *chikuzo* (Build or construct)
2.	さいこん 再建 *saikon* (削除: 直す) (delete *naosu*) (Rebuild a shrine / a temple)	7.	建てる *tateru* (Build)
3.	修築 *shuchiku* (削除: 直す) (delete *naosu*) (Repair a house etc.)	8.	築く *kizuku* (Build)
4.	建立 *konryu* (Build a chapel / a tower of a temple)	9.	建造 *kenzo* (Construct a buildiing / a ship)
5.	さいけん 再建 *saiken* (Rebuild or Reconstruction)	10.	建て増し *tatemashi* (Add to a building)

Figure 4: Result of preferential presentation of "*tera wo <u>tate</u> naosu.*"

in accuracy by non-interpolated average precision, about 10 points higher than non-ordered.

We have discussed only the initial step of the elaboration system, thus one of our future work is to handle syntactic and semantic constraints on actual paraphrasings after applying this method.

Acknowledgements

We would like to thank Mainichi Shinbun-sha and Aozora Bunko for allowing us to use their corpora, and Kadokawa Sho-ten for providing us with their thesaurus. We are also grateful to our colleagues for helping our experiment.

References

Akira Suganuma, Masanori Kurata and Kazuo Ushijima. 1990. A textual Analysis Method to Extract Negative Expressions in writing Tools for Japanese Documents. *Journal of Information Processing Society of Japan*, 31(6):792–800. (In Japanese)

Atsushi Fujita and Kentaro Inui. 2001. Paraphrase of Common Nouns to Its Synonyms by Using Definition Statements. *The Seventh Annual Meeting of The Association for Natural Language Processing*, 331–334. (In Japanese)

Chrysanne DiMarco, Graeme Hirst and Manfred Stede. 1993. The semantic and stylistic differentiation of synonyms and near-synonyms. *AAAI Spring Symposium on Building Lexicons for Machine Translation*, 114–121.

Diana Zaiu Inkpen and Graeme Hirst. 2002. Acquiring Collocations for Lexical Choice between Near-Synonyms. *ACL 2002 Workshop on Unsupervised Lexical Acquisition*, Philadelphia.

George E. Heidorn. 2000. Intelligent Writing Assistance. In Robert Dale, Hermann Moisl and Harold Somers (eds.), *A Handbook of Natural Language Processing*, Marcel Dekker, New York. Chapter 8.

Hiroko Inui and Naoyuki Okada. 2000. Is a Long Sentence Always Incomprehensible?: A Structural Analysis of Readability Factors. *Information Processing Society of Japan SIGNotes Natural Language*, 135(9):63–70. (In Japanese)

Kentaro Inui and Satomi Yamamoto. 2001. Corpus-Based Acquisition of Sentence Readability Ranking Models for Deaf People. *Proceedings of the sixth Natural Language Processing Pacific Rim Symposium (NLPRS)*, 159–166, Tokyo.

Mark Dras. 1998. Search in Constraint-Based Paraphrasing. *Proceedings of the second International Conference on Natural Language Processing and Industrial Applications*, 213–219, Moncton.

Noriko Tomuro and Steven L. Lytinen. 2001. Selecting Features for Paraphrasing Question Sentences. *Proceedings of the Workshop on Automatic Paraphrasing at Natural Language Processing Pacific Rim Symposium (NLPRS)*, 55–62, Tokyo.

Philip Edmonds. 1999. Semantic Representations of Near-Synonyms for Automatic Lexical Choice. *Ph.D. thesis, Department of Computer Science, University of Toronto*.

Satoshi Shirai, Satoru Ikehara, Akio Yokoo and Yoshifumi Ooyama. 1998. Automatic Rewriting Method for Internal Expressions in Japanese to English MT and Its Effects. *Proceedings of the second International Workshop on Controlled Language Applications (CLAW-98)*, 62–75.

Shin Ohno and Masato Hamanishi. 1981. *New Synonym Dictionary*. Kadokawa Shoten, Tokyo.

Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, Kazuma Takaoka and Masayuki Asahara. 2002. *Morphological Analysis System ChaSen 2.2.9 Users Manual*. Nara Advanced Institute of Science and Technology, Nara.