Applicability Analysis of Corpus-derived Paraphrases toward Example-based Paraphrasing

Kiyonori Ohtake[†]

†ATR Spoken Language Translation
Research Laboratories
2-2-2 Hikaridai, Seika-cho,
Soraku-gun, Kyoto 619-0288 Japan
kiyonori.ohtake@atr.co.jp

Kazuhide Yamamoto[†‡]

‡Department of Electrical Engineering,
Nagaoka University of Technology
1603-1 Kamitomioka, Nagaoka,
Niigata 940-2188 Japan
yamamoto@fw.ipsj.or.jp

Abstract

Two kinds of paraphrases extracted from a bilingual parallel corpus were analyzed. One is from an adjectival predicate sentence to a non-adjectival one. The other is from a passive form to a non-passive form. The ability to extract paraphrases is strongly desired for paraphrasing studies. Although extracting paraphrases from multi-lingual parallel corpora is possible, the type of paraphrases extracted is unknown. We discovered what types of examples can be obtained, and what types of paraphrasing will be available for the two kinds of paraphrases.

1 Introduction

Paraphrasing is a very important form of processing for natural language processing (NLP). A characteristic property of natural language is that various expressions can exist to express a single concept. So far, ad hoc paraphrasing methods have been applied to many NLP applications, such as machine translation, information retrieval, automatic summarization, and so on. Nowadays, however, many researchers recognize automatic paraphrasing as a very important technique for many areas of NLP, and work in this field has recently become more active.

We view paraphrasing as a translation where the source language and the target language are the same. Thus, we can take a similar approach to that of machine translation for paraphrasing. In this study, we use an example-based approach to control paraphrasing.

So far, a number of new methods for paraphrase extraction have emerged (e.g., (Lin and Pantel, 2001; Yamamoto, 2002; Barzilay and McKeown, 2001; Barzilay and Lee, 2003; Shinyama and Sekine, 2003; Pang et al., 2003)). There are several extracting methods: for example, one extracts paraphrases from a monolingual corpus, and another extracts from a multilingual parallel corpus. However, most of these methods have focused on efficiency or how many paraphrases would be extracted, and there have been few discussions on the types of paraphrases that were extracted.

Meanwhile, a rule-based approach seems reasonable in terms of the application of paraphrasing (Takahashi et al., 2001). Therefore, an investigation of how many paraphrasing examples are valuable, or difficult-to-write paraphrasing rules that cover the extracted examples are important. In this paper, we thus discuss the types of paraphrases that are extracted from a bilingual parallel corpus.

We analyze paraphrasing examples extracted from a Japanese-English parallel corpus of travel conversations. The corpus is made for Japanese speakers traveling to English-speaking countries. We can extract paraphrases from a parallel corpus (Barzilay and McKeown, 2001).

The basic idea of this method is that one expression can be translated into several expressions, and the translated expressions are paraphrases of each other.

In this paper, we introduce a method that extracts paraphrasing examples from a Japanese-English parallel corpus. Then, we analyze the extracted examples and discuss several issues about them. The purpose of this study is to answer the following questions:

(a) How many of the paraphrasing examples that are automatically extracted from a parallel corpus can be directly applied to paraphrasing?

(b) What types of paraphrasing will be achieved by the extracted examples?

2   Targets of Analysis

We analyzed two kinds of paraphrases. One is from an adjectival predicate sentence to a non-adjectival one. The other is from a passive formed sentence to a non-passive formed one. We focused on these two kinds of paraphrases because adjectival predicate sentences are used in various situations with various meanings, and passive formed sentences are also frequently used in Japanese to express several meanings.

Adjectival predicate expressions and passive formed expressions are frequently used in spoken language because they can express several meanings with one expression, and they are easy to use. On the contrary, such expressions are difficult for computers to handle correctly because they can convey several meanings. Thus, there are great advantages in being able to paraphrase adjectival predicate expressions and passive formed expressions into non-adjectival predicate expressions and non-passive formed expressions, respectively.

2.1   Adjectival predicates

The paraphrasing examples for the target of this analysis are paraphrases from sentences that have an adjectival predicate at the sentence-end to sentences that have a non-adjectival predicate at the sentence-end. In this paper, a sentence that has an adjectival predicate at the sentence-end is called an adjectival predicate sentence. An adjectival predicate sentence includes such expressions as "*ie kekkou desu* (no thank you)", which uses the adjectival noun (*kekkou*) + copula (*desu*).

For example, the functional role of "... *ii*[1] *desu ka*$_{(question\ particle)}$" may be classified as follows:

confirmation: *kore*$_{(this)}$ *de*$_{(particle)}$ ***ii desu ka***. (Is this all right?)

asking for permission: *tabako*$_{(cigarette)}$ *o*$_{(particle)}$ *sutte*$_{(smoke)}$ *mo*$_{(particle)}$ ***ii desu ka***. (May I smoke?)

asking a way: *donoyouni*$_{(how)}$ *kake*$_{(write)}$ *ba*$_{(particle)}$ ***ii desu ka***. (How do I write this?[2])

asking one's favorite/taste: *donna*$_{(what)}$ *iro*$_{(color)}$ *ga*$_{(particle)}$ ***ii desu ka***. (What color do you like?)

The definition of an adjectival predicate sentence is as follows:

> a sentence that has a morpheme where the part of speech (POS) is an adjective or adjectival noun in its last predicate phrase.

---

[1] The fundamental meaning of *ii* is good.

[2] A more correct translation might be 'How do I fill in this form?"

2.2 Passive forms

In Japanese, passive forms convey several meanings. From only surface information, it is difficult to determine whether a passive form is used as an honorific term or truly as a passive form. In addition, passive forms sometimes express the possibility of an action. For example, "*mirareru* (to be able to see, to see, or to be seen) " expresses the following three types of meaning:

- to express possibility
  "*yama wa **mirare** masita ka* (**Could** you **see** the mountain?),"

- to express an honorific
  "*yama o **mirareta** no desu ka* (**Did** you **see** the mountain?)," and

- the passive voice
  "*dareka ni **mirare** masita* (I **was seen** by someone.)."

We previously discussed the fact that it is hard to paraphrase passive forms by manually written rules (Ohtake and Yamamoto, 2001). That study suggested that example-based paraphrasing is suitable for paraphrasing passive forms. Thus, in this paper, we also collect paraphrases from passive forms to non-passive forms, and analyze them.

The definition of a passive formed sentence in this paper is as follows:

> a sentence that has a morpheme that is "*rareru*" or "*reru*" and a POS that is "*dousi-setubi* (verb ending[3])" in its last predicate phrase.

Some examples of passive formed sentences are as follows: "*iie, kinsi sareteimasu.* (No, you can't.)", "*okane o nusumare masita.* (My money was stolen.)", "*goyoyaku wa sarete imasuka.* (Did you make a reservation?)".

3 Extracting Paraphrases

We extracted paraphrasing examples from a Japanese-English parallel corpus that consists of travel conversations and that has been sentence aligned. The corpus is made for Japanese speakers traveling to English-speaking countries. The parallel corpus has 162,320 sentences. The number of English sentences is 98,290 and the number of Japanese sentences is 102,664. This corpus provides many instances of paraphrasing, since translations preserve the meaning of the original source but may use different words to convey the meaning.

The method used for collecting paraphrases is a simple one. The method we employed is fundamentally the same as that used by Barzilay and McKeown (Barzilay and McKeown, 2001). This method treats multiple Japanese translations corresponding to one English sentence as instances of paraphrasing. Unlike the situation of Barzilay and McKeown, we do not have to address the alignment or the differences in styles, since the parallel corpus has been sentence aligned.

We employed ChaSen[4] as a morphological analyzer and CaboCha [5] as a parser to obtain the two kinds of paraphrases.

The collecting method is as follows:

---

[3]This can also be described as an auxiliary verb.
[4]`http://chasen.aist-nara.ac.jp/`
[5]`http://cl.aist-nara.ac.jp/˜taku-ku/software/cabocha/`

Step 1: If there are some Japanese translations $J_{i1}, \cdots, J_{im}$ corresponding to an English sentence E_i, then we form all combinations of the two translations J_{ij} and J_{ik}, where $1 \leq j, k \leq m$, $j \neq k$, and the combined translations are paraphrases. (This operation is illustrated in Figure 1.)

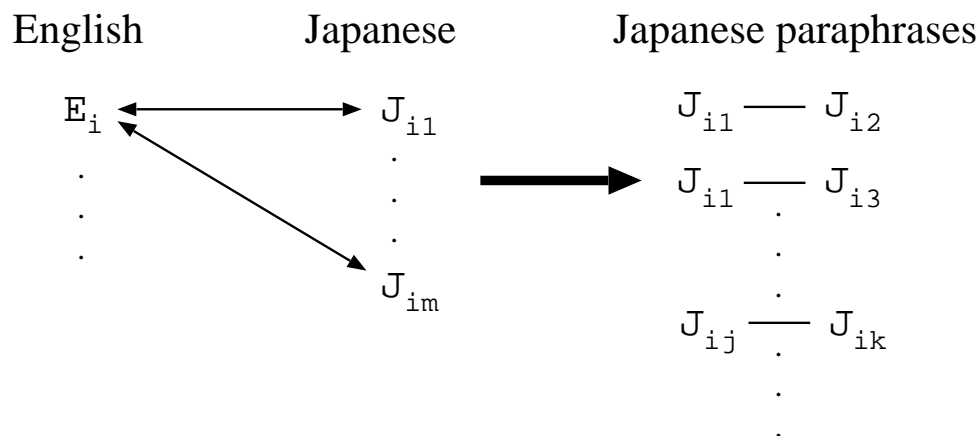**Step 2:** Do Step 1 for all English sentences.

Figure 1: Extracting Paraphrases

In addition, we can gather more paraphrases by the following operations:

1. Collect English translations corresponding to a Japanese sentence in the collected paraphrases.

2. Merge the Japanese translations of the collected English sentences into the collected paraphrases.

These operations are illustrated in Figure 2.

Of course, we can do this operation recursively, but, at the same time, the gathered paraphrases may be useless. Thus, in this paper, we reject this operation. A total of 102,747 pairs of Japanese paraphrases were obtained by doing Steps 1 and 2 above. Table 1 shows some examples of pairs for English sentence and its translated sentences of Japanese.

4 Analysis of Context Dependency

We analyzed how many extracted paraphrases are directly applicable for use in an example-based paraphrasing method. A directly applicable example means that the paraphrase is context-independent.

From the 102,747 pairs of collected paraphrasing examples, we extracted paraphrases from adjectival predicate sentences to non-adjectival predicate sentences and from passive formed sentences to non-passive formed sentences. We removed examples that had more than five paraphrases because they appeared likely to cause the following two problems.

First, in these cases, the differences between $P_{i1}, \cdots,$ and P_{im} tend to be very small, such as local changes of functional words. Therefore, these paraphrases are subsidiary examples. Second, the paraphrasing would be very complicated if we tried to apply these cases.
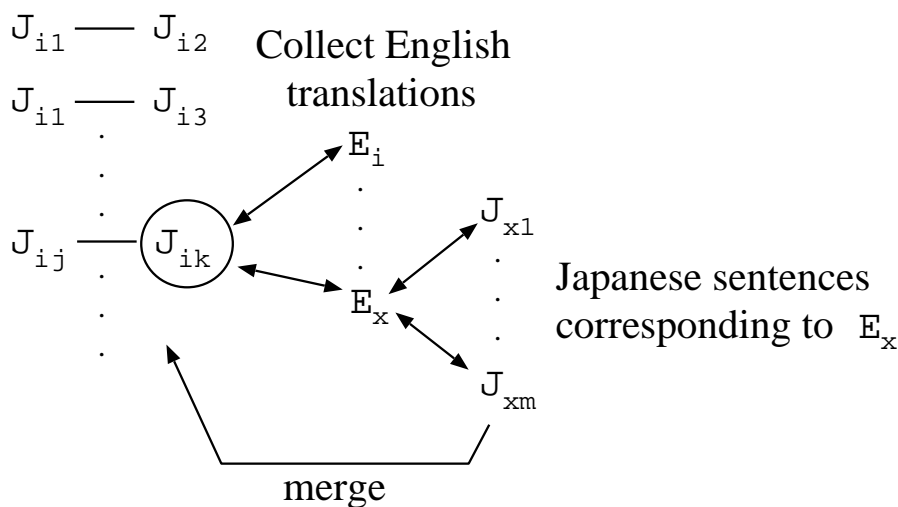
Japanese paraphrases

J_{i1} —— J_{i2} Collect English
J_{i1} —— J_{i3} translations
.
. E_i
.   .
J_{ij} —— (J_{ik}) . J_{x1}
.   .   .   Japanese sentences
. E_x .   corresponding to E_x
.   .
. J_{xm}

merge

Figure 2: Expansion of Extracting Paraphrases

Table 1: Examples of pairs for English sentence and Japanese sentences

English sentence	Japanese sentences
I'd like a refund.	harai modosi sitai no desuga.
	henkin site kudasai.
	henkin o onegai sitai no desuga.
I lost my wallet.	saifu o otosite simattan desu.
	saifu o nakusite simai masita.
	saifu o funsitu simasita.
I have a fever.	netu ga arun desu.
	netu ga takai no desu.
	netuppoi desu.
	hatunetu desu.

When we removed the examples that had more than five paraphrases, the average number of paraphrases to one expression both for the adjectival predicate sentences and the passive formed sentences was less than 2.0.

As a result of this restriction, we obtained 3,660 paraphrasing examples of adjectival predicate sentences and 485 paraphrasing examples of passive formed sentences.

Examples of paraphrases extracted from a bilingual corpus are shown in Table 2. The table shows the original English sentence and its Japanese paraphrases by the formulation J_s/J_p. The class and the feature will be introduced in this section and Section 5, respectively.

4.1 Analysis of paraphrases of adjectival predicate sentences

We analyzed the 3,660 paraphrases of adjectival predicate sentences and classified these examples into the following three classes:

(A) Examples that do not depend on context and are applicable at all times.

Table 2: Examples of extracted paraphrases

English	Japanese paraphrase	class	feature
Are you free tomorrow?	*asita wa hima desuka. /asita wa hima.*	(A)	(L)
I'd like a room with a bath.	*ofurotuki no heya ga ii no desuga. / basutuki no heya o onegaisimasu.*	(A)	(L)(S)
When was it built?	*itu taterareta no desuka. / taterata no wa itu desuka.*	(A)	(S)
I miss you.	*anata ni aenakute sabisii. / anata ni aitai.*	(A)	(E)
How can I get to this address?	*koko e wa dou ittara iinodesyou. / kono jyuusyo ni ikitai no desuga.*	(B)	—
That's good.	*kore wa ii. /kasikomari masita.*	(C)	—

(B) Examples that depend on context. In other words, there are some contexts in which applying these examples is improper.

(C) Examples that are completely improper.

The result of this classification is shown in Table 3. When we classified the examples, we did not view each original English sentence.

Table 3: Classification for paraphrases of adjectival predicate sentences

class	(A)	(B)	(C)
examples	2066	1233	361
proportion (%)	56.45	33.69	9.86

From Table 3, we can conclude that more than half of the paraphrases are directly applicable to paraphrasing. However, it is hard to automatically distinguish class-(A) from class-(B) or (C). To achieve example-based paraphrasing, we must determine how to eliminate class-(B) or (C) examples from the examples automatically extracted.

We can also conclude that almost 10% of the paraphrases extracted are included in class-(C). One reason is that we extracted paraphrases without considering contextual information. In the parallel corpus that we employed, contextual information is written with each sentence in some way. For example, some keywords, such as "at the airport" or "shopping" express context. However, these contextual keywords are not machine readable. We could not employ this contextual information, therefore, and thus improper paraphrases were extracted. If we were to employ this contextual information, we could not only automatically distinguish class-(A) from class-(B) or (C), but we could also achieve paraphrasing with examples of class-(B).

In addition, some other points resulted in the class-(B) or (C) examples. One is the quality of the translations. The target corpus consists of travel conversations for Japanese speakers traveling to English-speaking countries. Thus, the Japanese examples tend to be unnatural expressions. This is because the writer who wrote the unnatural Japanese expressions is not a native Japanese speaker. These unnatural Japanese expressions are related in part to class-(B) or (C) examples.
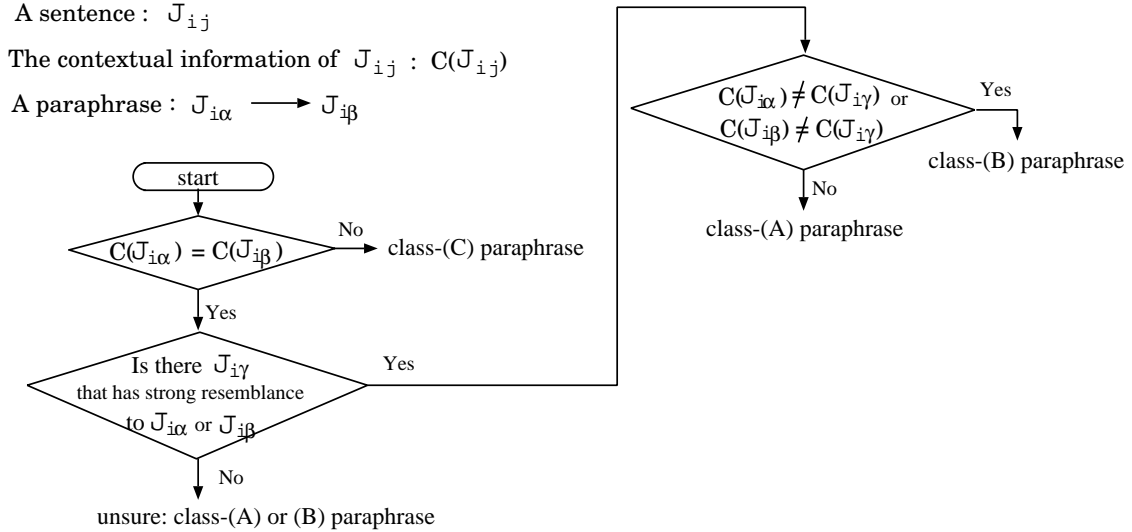
A sentence : J_{ij}

The contextual information of J_{ij} : $C(J_{ij})$

A paraphrase : $J_{i\alpha} \longrightarrow J_{i\beta}$

start

$C(J_{i\alpha}) = C(J_{i\beta})$ — No → class-(C) paraphrase

Yes

Is there $J_{i\gamma}$ that has strong resemblance to $J_{i\alpha}$ or $J_{i\beta}$ — Yes →

No

unsure: class-(A) or (B) paraphrase

$C(J_{i\alpha}) \neq C(J_{i\gamma})$ or $C(J_{i\beta}) \neq C(J_{i\gamma})$ — Yes → class-(B) paraphrase

No

class-(A) paraphrase

Figure 3: Recognizing the classes of examples

Let us consider a method of recognizing the classes of examples. The contextual information $C(J_{ij})$ of a sentence J_{ij} is given by words (e.g., at restaurant, in an airplane, and so on). Paraphrases are given in the form $J_{i\alpha} \rightarrow J_{i\beta}$. When $C(J_{i\alpha}) = C(J_{i\beta})$ and a sentence $J_{l\gamma}$ that has a strong resemblance to $J_{i\alpha}$ or $J_{i\beta}$ is given, if $C(J_{i\alpha}) \neq C(J_{l\gamma})$ or $C(J_{i\beta}) \neq C(J_{l\gamma})$, then the example $J_{i\alpha} \rightarrow J_{i\beta}$ is included in class-(B). On the other hand, if $C(J_{i\alpha}) \neq C(J_{i\beta})$, then the example is included in class-(C). This flow is shown in Figure 3.

Applying class-(B) examples to paraphrasing requires the contextual information $C(K)$ of the target sentence K. If $C(K)$ is given properly, we can estimate whether a paraphrase $K \rightarrow J_{i\alpha}$ that is applied to sentence K is applicable or not. This is because, if $C(K) \neq C(J_{i\alpha})$, then the paraphrase is not applicable.

To give the context of a sentence, we can use the topic as an approximation of context (Florian and Yarowsky, 1999). This seems to be suitable for traveling conversations, because the target domain is not very broad, and several topic detection methods are available. Fortunately, the corpus that was we employed has such contextual information. The contextual information is given by several topic words. However, currently available topic words are very rough and there are many inconsistencies. Thus the methods shown in above will be very promising if we prepare proper topic words and remove the inconsistencies.

4.2 Analysis of paraphrases of passive formed sentences

We also analyzed 485 paraphrases of passive formed sentences and classified these examples into three classes: (A), (B), and (C). Table 4 shows the results of this classification.

Table 4: Classification for paraphrases of passive formed sentences

class	(A)	(B)	(C)
examples	294	182	9
proportion (%)	60.62	37.53	1.86

From Tables 3 and 4, we can conclude that the difference in the type of classified paraphrase

has little effect on the distribution of the classes. Almost 60% of the paraphrases were directly applicable to example-based paraphrasing.

5 Classifying Context-independent Examples

More than half of both the adjectival predicate sentences and the passive formed sentences were context-independent. We analyzed the types of paraphrasing that such examples belonged to.

We divided the paraphrases in the context-independent paraphrasing examples into the following three types: (L) lexical paraphrasing, (S) syntactical paraphrasing, and (E) semantic paraphrasing.

(L) Lexical paraphrasing handles differences in notation (such as differences in characters, e.g., the Chinese character vs the Japanese syllabary), or differences in verbal suffixes (such as "X *site kudasai* (Please, do X)" to "X *site kudasai masenka*"). (S) Syntactical paraphrasing handles the changing of word order (such as "*donokuraino takasa* (how tall is ...)" to " *takasaha donokurai*"), or paraphrasing predicate parts with changing particles (such as "*katto dake de iidesu* (Just a haircut, please)" to "*katto dake o onegai simasu*"). (E) Semantic paraphrasing handles paraphrases that are not covered by (L) or (S).

We have already showed that manually written rules are practical for local changed paraphrases in (Yamamoto, 2001; Ohtake and Yamamoto, 2001). Thus, we expect that the examples classified into class-(A) include not local changed paraphrases but substantial changed paraphrases. To clarify what types of paraphrasing are included in these examples, we assigned the following three features to the class-(A) examples:

(**L**) **L**exical paraphrasing achieves the example. By using the changed part and its fore-and-after morphemes, the obtained paraphrases have this feature.

Examples: *pasupôto o nakusite simai masita.* (I have lost my passport.)
  → *pasupôto o funsitusite simai masita.*

otearai wa doko desyouka. (Where's the restroom?)
  → *toire wa doko desuka.*

(**S**) **S**yntactical paraphrasing achieves the example. Changing a word order or its voice and changing a predicate by exchanging particles obtain this feature. Some examples are as follows, where roman letters show the original English sentence, and italic letters show a direct translation:

kôhî ga hosii no desuga. (I'll have some coffee.)
  → *kôhî o kudasai.* (*May I have coffee, please?*)

dono kurai fukai desuka. (How deep is it?)
  → *fukasa ha donokurai desuka.* (*How is the depth of it?*)

(**E**) **S**emantic paraphrasing achieves the example. In other words, examples that cannot be obtained by combining feature-(L) and (S) paraphrasing have this feature.

Examples:

sore ijyou wa yasuku nari masenka. (Is it able to go down in price more?)
  → *sore ga saisyuutekina nedan desuka.* (*Is that your last price?*)

doko de haraeba ii desuka. (Where should I pay?)
  → *siharai basyo wa doko desuka.* (Where is the paying place?)

We classified 2,066 paraphrases into the following four classes: having only (L), having only (S), having both (L) and (S), and having only (E). The classification result is shown in Table 5.

Table 5: Classification of class-(A) examples of adjectival predicate sentences

feature	(L)	(S)	(L)+(S)	(E)
examples	624	355	321	766
proportion (%)	30.20	17.18	15.54	37.08

Through observations of examples that have feature-(L), or self-examination, it seems easy to extract paraphrasing rules from these paraphrases. In addition, from the examples that have feature-(S), although we need dictionaries, such as verbal case frames, we can extract paraphrasing rules. On the other hand, from the examples that have feature-(E), it is difficult to extract paraphrasing rules. If we collect these feature-(E) examples, broad paraphrasing is possible. Unfortunately, almost 60% of the collected examples have feature-(L) or (S). Thus, we cannot expect many paraphrasing examples having feature-(E) to be collected from the parallel corpus we employed.

If it is difficult to collect paraphrases having feature-(E), it appears reasonable to create such examples by hand. However, some examples having feature-(E) seem to be difficult to create manually in a short time. Making these examples by hand, therefore, may be a difficult task.

Meanwhile, in the collected examples, some adjectival predicate sentences have examples that have both feature-(E) and feature-(L+S)[6] at the same time. In other words, $J_{ij} \rightarrow J_{ik}$ has the feature-(E) and the other $J_{ij} \rightarrow J_{il}$ has the feature-(L+S). For example, "*suwatte mo ii desuka.*(May I sit here)$\rightarrow$ *kono seki wa aite imasuka.*[7]" has the feature-(E) and "*suwatte mo ii desuka* (May I sit here)$\rightarrow$ *suwatte mo kamai masenka*" has the feature-(L). From these examples, to satisfy a *paraphrasing factor*, the paraphrases having feature-(L) or (S) might adequately work for paraphrasing.

In addition, a part of feature-(E) examples is subordinate. For example, when an example $J_s \rightarrow J_p$ has feature-(E), and another example $J_s \rightarrow J'_p$ has also feature-(E), if J'_p can be obtained by applying paraphrases having feature-(L+S) to J_p, then these examples are subordinate. We estimate, thus, that the number of examples having feature-(E) is essentially smaller than the number shown in Table 5.

We also classified 294 paraphrases of passive formed sentences. The classification result is shown in Table 6.

Table 6: Classification of class-(A) examples of passive formed sentences

feature	(L)	(S)	(L)+(S)	(E)
examples	37	25	20	212
proportion (%)	12.59	8.50	6.80	72.11

The distribution is very different from that of adjectival predicate sentences. This difference tells us that the type of paraphrase exerts a strong influence on the distribution of the paraphrasing types.

---

[6]The feature-(L+S) means feature-(L) or (S), or both (L) and (S).

[7]The direct translation of this Japanese expression is 'Is this seat taken?'

On the other hand, an interesting fact exists that is not shown in Table 6. The 85 passive forms of the verb "*fukumu* (to include)" constituted the 212 examples that have feature-(E). The 212 examples consist of 36 verbs. The verb "*suru* (to do)" has 36 examples in the second place. Judging from these facts, it is very significan that the verb "*fukumu* (to include)" has 85 examples. The verb "*fukumu* (to include)" may be frequently used in traveling situations, such as renting a car (e.g., Does it include insurance?), reserving a hotel room, etc. (e.g., Are any meals included?), or shopping (e.g., Does it include tax?), and so on.

6   Related Works

To date, there has been no work directly related to analyzing paraphrases extracted from corpora. This is because the methods or techniques used to extract paraphrases from corpora have been premature, and few people have an interest in collecting paraphrases. However, there are some slightly related works.

We mentioned that an example-based paraphrasing method is suitable for passive formed sentences from the viewpoint of paraphrasing honorifics (Ohtake and Yamamoto, 2001). The results shown in this paper support this discussion.

The first attempt to derive paraphrasing rules from corpora was undertaken by Jacquemin et al., who investigated morphological and syntactic variants of technical terms (Jacquemin et al., 1997). Although, these rules achieve high accuracy in identifying term paraphrases, the techniques used have not been extended to other types of paraphrasing.

Lapata investigated polysemous adjectives whose meanings vary depending on the nouns they modify (Lapata, 2001). Lapata acquired the meanings of these adjectives from a large corpus and proposed a probabilistic model that provides a ranking of the set of possible interpretations. Statistical techniques were also successfully used by Lapata to identify paraphrases of adjective phrases.

Barzilay and McKeown presented an unsupervised learning algorithm for the identification of paraphrases from a corpus of multiple English translations of the same source text (Barzilay and McKeown, 2001). They employed literary texts written by foreign authors. Their algorithm produced 9,483 pairs of lexical paraphrases and 25 morpho-syntactic rules with very high accuracy. However, there was no discussion on extracted paraphrases, and therefore, what kinds of paraphrases are extracted and whether it is difficult to manually write those paraphrases remain undetermined.

From the viewpoint of machine translation, to paraphrase an expression into another expression that is easy to translate is closely related to designing a controlled language. Mitamura reported on designing a controlled language (Mitamura, 1999). Although, there is a language specific part, the discussion that touched on the points should be focused on provides helpful information to consider for paraphrasing.

7   Conclusion

We analyzed two kinds of paraphrases automatically extracted from a bilingual parallel corpus. These two kinds are focused since the expressions may be ambiguous in their meanings, or they are frequently used in Japanese conversations. By paraphrasing these kinds to simpler and clearer expressions, easier processing becomes possible.

We investigated the following points:

(a) How many of the paraphrasing examples that are automatically extracted from a parallel corpus are directly applicable to paraphrasing?

(b) What types of paraphrasing will be achieved by the extracted examples?

To answer question (a), the results showed that almost 60% of the paraphrases extracted from the parallel corpus we employed were directly applicable for use in example-based paraphrasing. Moreover, without contextual information, it is hard to extract more examples that are directly applicable to paraphrasing. On the other hand, contextual processing enables us to use the class-(B) (contextually dependent) examples that account for almost 35% of extracted paraphrases.

To answer question (b), the results clarified that the two types of paraphrases extracted from a bilingual corpus showed different distributions. Almost 35% of the context-independent paraphrasing examples of adjectival predicate sentences and almost 70% of the context-independent paraphrasing examples of the passive formed sentences were desirable for example-based paraphrasing.

Although we analyzed two kinds of paraphrases, to analyze other kinds of paraphrases is left as a future topic for study.

Acknowledgement

References

Barzilay, Regina and Lillian Lee. 2003. Learning to paraphrase: An unsupervised approach using multiple-sequence alignment. In *Proceedings of HLT-NAACL 2003*, pages 16–23.

Barzilay, Regina and Kathleen R. McKeown. 2001. Extracting paraphrases from a parallel corpus. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 50–57.

Florian, Radu and David Yarowsky. 1999. Dynamic nonlocal language modeling via hierarchical topic-based adaptation. In *Proceedings of the Thirty-Seventh Annual Meeting of the Association for Computational Linguistics (ACL'99)*, pages 167–174.

Jacquemin, Christian, Judith L. Klavans, and Evelyne Tzoukermann. 1997. Expansion of multi-word terms for indexing and retrieval using morphology and syntax. In *Proceedings of the Thirty-Fifth Annual Meeting of the Association for Computational Linguistics and the Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pages 24–31.

Lapata, Maria. 2001. A corpus-based account of regular polysemy: The case of context-sensitive adjectives. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Lin, Dekang and Patrick Pantel. 2001. Discovery of inference rule for question-answering. *Natural Language Engineering*, 7(4):343–360.

Mitamura, Teruko. 1999. Controlled language for multilingual machine translation. In *Proceedings of MT Summit VII*, pages 46–52.

Ohtake, Kiyonori and Kazuhide Yamamoto. 2001. Paraphrasing honorifics. In *Workshop Proceedings of Automatic Paraphrasing: Theories and Applications (NLPRS2001 Post-Conference Workshop)*, pages 13–20.

Pang, Bo, Kevin Knight, and Daniel Marcu. 2003. Syntax-based alignment of multiple translations: Extracting paraphrases and generating new sentences. In *Proceedings of HLT-NAACL 2003*, pages 102–109.

Shinyama, Yusuke and Satoshi Sekine. 2003. Paraphrase acquisition for information extraction. In *Proceedings of the Second International Workshop on Paraphrasing: Paraphrase Acquisition and Applications*, pages 65–71.

Takahashi, Tetsuro, Tomoya Iwakura, Ryu Iida, Atsushi Fujida, and Kentaro Inui. 2001. Kura: A transfer-based lexicoo-structural paraphrasing engine. In *Proceedings of Automatic Paraphrasing: Theories and Applications (NLPRS2001 Workshop)*, pages 37–46.

Yamamoto, Kazuhide. 2001. Paraphrasing spoken Japanese for untangling bilingual transfer. In *Proceedings of the Sixth Natural Language Processing Pacific Rim Symposium (NLPRS2001)*, pages 203–210.

Yamamoto, Kazuhide. 2002. Acquisition of lexical paraphrases from texts. In *Proceedings of the 2nd International Workshop on Computational Terminology (Computerm 2002, in conjunction with Coling 2002)*, pages 22–28.