Evaluation of DMAX Criteria for Selecting Equivalent Translation based on Dual Corpora Statistics

Shinichi DOI and Kazunori MURAKI
NEC Corp. C&C Information Technology Research Laboratories
4-1-1, Miyazaki, Miyamae-ku, Kawasaki 216, JAPAN
e-mail: doi@mtl.cl.nec.co.jp, k-muraki@mtl.cl.nec.co.jp

Abstract

This paper describes a method for ambiguity resolution in translation using linguistic statistics extracted independently from dual corpora of source and target languages, which provides reasonable criteria for determining a suitable translation by making the dependency relation in the source language be reflected in the translated text. This method is tractable because the required statistics can be computed semi-automatically in advance from dual corpora, while a conventional corpus-based translation method needs a large volume of bilingual corpus of strict pairs of a sentence and its translation. We call this method for selecting equivalent translation Double MAXimize Criteria (DMAX Criteria) based on Dual Corpora Statistics.

To prove the effectiveness of our method, we have calculated co-occurrence frequencies —how frequently the expressions co-occur in the same sentence— between nouns in Japanese and English text and made some experiments on Japanese-to-English and English-to-Japanese translation. We compared the outputs of translation with DMAX Criteria and translation with the method using only the statistics of target language for translation. The result indicates that, with DMAX Criteria, we can select equivalent translation more correctly than with the latter method. We also confirmed the effectiveness to enlarge the size of the language unit where the two words co-occur also.

1 Introduction

Recently a lot of natural language processing systems like machine translation systems have been developed and put into practical use, but ambiguity resolution in translation is still one of the biggest problems in such systems. These systems have conventionally adopted a rule-based disambiguation method, using linguistic restrictions described logically in dictionary and grammar, but it is impossible to provide all the restrictions in advance. Furthermore, such systems have no reasonable means to select the most suitable translation if the input expression meets two or more restrictions, or if the input expression meets no restrictions.

In order to overcome these difficulties, following methods based on corpus-based knowledge have been proposed:

1. Example-Based Translation : This method is based on translation examples (pairs of source text and its translation) [Nagao 84, Sato 90, Sumita 92].
2. Structured-Corpus-Based Translation : This method uses structured bilingual corpora coupled together by cross-coding translation units [Sadler 90],
3. Statistics-Based Translation : This method uses statistical or probabilistic information extracted from large corpora [Brown 90, Nomiyama 91, Schabes 92],

Still, each has inherent problems and is insufficient for ambiguity resolution. For example, an example-based translation system needs large-scale database of strict pairs of source text and its translation, and it is difficult to collect sufficiently large bilingual corpora.

This paper describes a method for selecting suitable translations using statistical data extracted independently from source and target language texts [Muraki 91, Doi 92c]. The statistical data used here are linguistic statistics representing the dependency degree on the pairs of expressions in each text, especially co-occurrence frequency, i.e., how frequently the expressions co-occur in the same language unit —sentence, paragraph or chapter— of each text. The dependency relation in the source language is reflected in the translated text through a bilingual dictionary by selecting an equivalent translation which maximizes both co-occurrence frequency of the source and target language. We call this method for selecting equivalent translation Double MAXimize Criteria (DMAX Criteria) based on Dual Corpora Statistics.

To prove the effectiveness of our method, we have calculated co-occurrence frequencies between nouns in Japanese and English text and made some experiments on Japanese-to-English and English-to-Japanese translation. The result indicates that, with DMAX Criteria, we can select equivalent translation more accurately than with the method using only the statistics of target language. We also confirmed the effectiveness to enlarge the size of the language unit where the two words co-occur.

In this paper we will comment on the characteristics and the limits of the conventional methods for selecting equivalent translation. Then we will describe the details of DMAX Criteria for selecting equivalent translation. Finally, we will show the procedures and the results of our experiments.

2 Conventional Methods of Ambiguity Resolution

2.1 Rule-Based Translation

In conventional methods, linguistic restrictions described in a dictionary and grammar are used to select a suitable translation. In general, these restrictions are defined logically from the characteristics of another expression which modifies or is modified by the expression being processed. For example, to translate predicates (verbs and predicative adjectives), semantic restrictions are based on essential case arguments in forms of semantic markers to indicate features of words or terms in the thesaurus to show a hierarchy composed of word concepts.

Though these conventional methods have been very useful in realization of natural language processing systems, they have many problems. Restrictions on all dependencies cannot be described in advance. The systems suffer from an inability to select suitable translations if the input expression meets two or more restrictions, and have difficulty in processing any expression that meets no restrictions. Moreover, the description of the restrictions is based on direct structural dependencies, therefore it is difficult to describe restrictions based on sister-dependency or between expressions belonging to different sentences or paragraphs.

2.2 Example-Based Translation

Besides the conventional translation method above, a machine translation system based on translation examples (pairs of source texts and their translations) has been proposed [Nagao 84, Sato 90, Sumita 92]. This type of system, called Analogy-Based or Example-Based Machine Translation, involves storing a large number of bilingual translation examples as a database, and translates input expressions by retrieving an example most similar to the input from the database. There is no failure of output in this method because it selects the most similar example not necessarily an identical one.

However this example-based translation system needs a large-scale database of translation examples, and it is difficult to collect sufficiently large bilingual corpora. Even if possible, there is no way to divide the sentences of such corpora into fragments and link them automatically, and it is expensive to divide and link manually.

To overcome this problem, we have also proposed a new mechanism based on sentential examples in a dictionary, which utilize the merits of both the translation by logical restrictions and the example-based methods, by selecting an equivalent translation which has the most similar to the input expression [Doi 92a, Muraki 92]. This mechanism can guarantee no failure in selecting equivalent translations.

2.3 Structured-Corpus-Based Translation

A method using syntactically and referentially structured bilingual corpora coupled together by cross-coding translation units has been also proposed [Sadler 90]. This bilingual corpora is called Bilingual Knowledge Bank (BKB) and is used as a knowledge-base on various levels for machine translation and other applications. This kind of structured and cross-coded bilingual corpora is useful, but again, it is difficult and expensive to collect, analyze and cross-code this kind of corporus.

2.4 Statistics-Based Translation

Several new methods, especially for machine translation, have been proposed that select suitable translations using statistical or probabilistic information extracted from language text [Brown 90, Nomiyama 91, Schabes 92]. Because many machine readable texts have already been collected, it is not difficult to extract statistical information for expression in the texts semi-automatically. Moreover, the statistical information reflects the context in which each word occurs and implies the logical restrictions based on indirect structural dependencies.

Although all systems are referred to as "statistics-based", the statistical information used in them is diverse, such as translation probability, connectivity of words, (co-)occurrence frequency, etc. We will comment on the characteristics and the limits of these systems.

The first method uses fertility probabilities, translation probabilities and distortion probabilities [Brown 90, Brown 92]. Fertility means the number of the words in target language that the word of the source language produces, and distortion means the distance between the position of the word of the source language and the one of the target language. The method has been applied to an experimental translation system from French to English. However, since these probabilities are extracted from a large number of text pairs that are translations of each other, this method suffer from the same difficulties as example-based translation in collecting and analyzing an adequately large bilingual corpora, and it's difficult to apply this method to languages whose linguistic structures aren't similar, such as English and Japanese.

The second method uses the occurrence statistics of target language [Nomiyama 91]. It is calculated in advance how frequently the each expression occurs in the target language text, which needs only to belong the same field as the source language text belongs, but not to be a translated text of the source language text. If there are more than one possible equivalent translations, the most frequent translation is selected through this calculated data. Moreover, this method can be applied to make good use of the conventional methods of selecting equivalent translations, for it employs the frequency data exclusively when logical restrictions cannot select one out of candidates.

However this method has one big problem. The high frequency of the expression in the target language text may not originate from the frequency of the expression in the source language text to be translated, because one target language expression does not correspond to only one source language expression in general.

Suppose the following sentence is an example:

ENGLISH:	`The market prices increased because of reductions in the volume of production.`			
	⇓			
JAPANESE:	生産量の減少により、市場価格が上昇した。			
	Seisan- ryou-no	*genshou-niyori,*	*shijou-kakaku-ga*	*joushou-shita.*
	production volume	reduction	market price	increased

Figure 1 indicates four English words and their Japanese equivalent translations in this sentence and the co-occurrence frequencies between the words. These frequencies are calculated from English journals and Japanese newspapers, which we will explain later.

In this case, the co-occurrence frequency of the target language Japanese denotes that the most frequent pair is "市"–"本", because these Japanese words have several meanings. Then using only the statistics of target language text misleads a wrong expression "市(ichi,shi=market,city)" as the equivalent translation of 'market', and "本(hon,moto=book,this)" as the one of 'volume'.

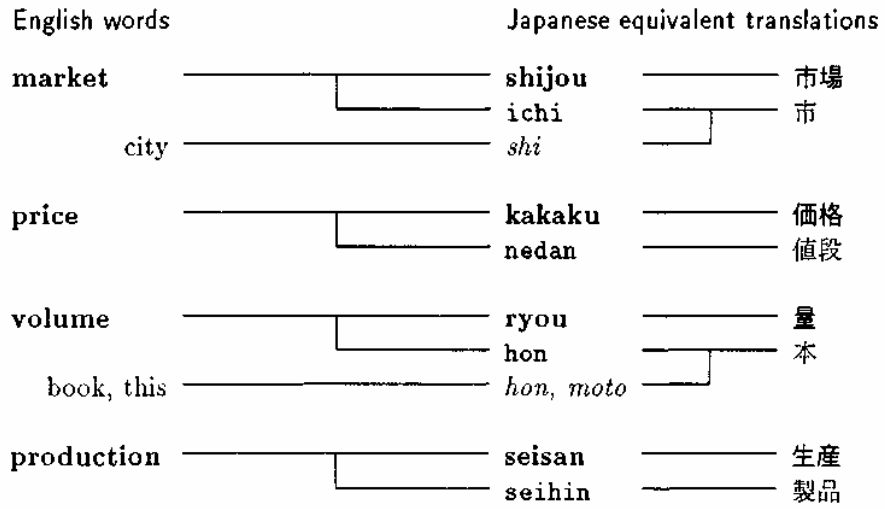
ENGLISH: The market prices increased because of reductions in the volume of production.

⇓

JAPANESE: 生産量の減少により、市場価格が上昇した。

Seisan-	*ryou-no*	*genshou-niyori,*	*shijou-kakaku-ga*	*joushou-shita.*
production volume		reduction	market price	increased

English–Japanese bilingual dictionary



co-occurrence frequencies between the English words

	market	price	volume	production
market	-	1070	117	222
price	-	-	65	251
volume	-	-	-	48
production	-	-	-	-

co-occurrence frequencies between the Japanese equivalent translations

	市場	市	価格	値段	量	本	生産	製品
市場	-	17	228	12	30	7	79	156
市	-	-	10	5	15	437	74	15
価格	-	-	-	37	36	9	174	171
値段	-	-	-	-	3	2	8	15
量	-	-	-	-	-	20	59	9
本	-	-	-	-	-	-	23	6
生産	-	-	-	-	-	-	-	98
製品	-	-	-	-	-	-	-	-

Figure 1 English-to-Japanese Translation Example

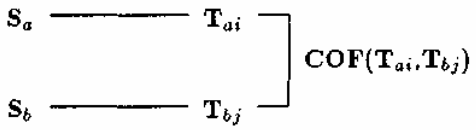
3 Selecting Equivalent Translation Based on Dual Corpora Statistics of Source and Target Languages

Now we propose a method to provide reasonable criteria for selecting suitable translations using the simple statistical data extracted from source language text in addition to the one from target language text. These source and target language texts don't have to be translations of each other. The proposed method gives us a way to select the expression with the highest frequency of the target language that keeps high frequency of the source language at the same time, so it overcomes the difficulty of the method using only the frequency data of target language text, because it does not select the expression with the highest frequency of only the target language text.

3.1 Using statistical data on source language text

The method using only statistical data of target language may mislead a wrong equivalent translation, because in general each target language expression corresponds to more than one source language expression.

When a source language expression $\mathbf{S}_a$ has n target language equivalent translations $\mathbf{T}_{ai}(i = 1 \cdots n)$, equivalent translation selection with co-occurrence frequency is shown as this:



where

$\mathbf{S}_k$	: source language expression
$\mathbf{T}_{ki}(i = 1 \cdots n)$	: n target language equivalent translations of $\mathbf{S}_k$
$\mathbf{COF}(\mathbf{E}_i,\mathbf{E}_j)$	: co-occurrence frequency of two expressions $\mathbf{E}_i,\mathbf{E}_j$

The method using only statistical data of target language selects $\mathbf{T}_{ai}$ which maximizes the co-occurrence frequency of the target language $[\mathbf{T}_{ai}|\max_{b,i,j}\mathbf{COF}(\mathbf{T}_{ai},\mathbf{T}_{bj})]$ as the equivalent translation of $\mathbf{S}_a$, where the partner of the co-occurrence $\mathbf{T}_{bj}$ plays the part of the basis for the equivalent translation selection. The biggest problem of this method is that $\mathbf{T}_{bj}$ which depends both b and j is selected by only statistical data of target language.

Our new method provides reasonable criteria for selecting the basis for the equivalent translation selection using the statistical data of source language text. First the source language expression $\mathbf{S}_b$ which maximizes the co-occurrence frequency of the source language $[\mathbf{S}_b|\max_b\mathbf{COF}(\mathbf{S}_a,\mathbf{S}_b)]$ is selected, then the equivalent translation $\mathbf{T}_{ai}$ which maximizes the co-occurrence frequency of the target language $[\mathbf{T}_{ai}|\max_{i,j}\mathbf{COF}(\mathbf{T}_{ai},\mathbf{T}_{bj})]$ is selected. The dependency relation in the source language is reflected in the translated text through this method. We call this method for selecting equivalent translation DMAX Criteria (Double MAXimize Criteria based on Dual Corpora Statistics).

3.2 Double MAXimize Criteria based on Dual Corpora Statistics

The procedure of this method is summarized as follows:

1. Select domain X — X:politics/medicine/computer science/etc.
2. Prepare the source and target language texts of domain X (the target language text needs not to be a translated text of the source language text)
3. Calculate co-occurrence frequency of every expression in source language text
4. Calculate co-occurrence frequency of every expression in target language text

5. When a source language expression $\mathbf{S}_a$ has n target language equivalent translations $\mathbf{T}_{ai}(i = 1 \cdots n)$

   (a) Select $\mathbf{S}_b$ | $\max_b \mathbf{COF}(\mathbf{S}_a, \mathbf{S}_b)$

   (b) Select $\mathbf{T}_{ai}$ | $\max_{i,j} \mathbf{COF}(\mathbf{T}_{ai}, \mathbf{T}_{bj})$

3.3 Operation Example

We will show operation example of English-to-Japanese translation using the sentence in Figure 1. In this case, with only statistical data of target language, "市(ichi,shi)" may be chosen as an equivalent translation of 'market' and "本(hon,moto)" as the one of 'volume', because the pair of "市"–"本" co-occurs most frequently in the target language. However with DMAX Criteria, the equivalent translations of 'market' and 'volume' are selected accurately.

- The pair of words which co-occurs most frequently in the source language is 'market'–'price'.
- The pair of the equivalent translation of 'market' and the one of 'price' which co-occurs most frequently in the target language is "市場"–"価格".
- As a result, 'market' is translated into "市場".
- Repeating DMAX Criteria, 'volume' is translated into "量".

4 Experiments

To prove the effectiveness of our method, we have calculated co-occurrence frequencies between nouns in Japanese and English text and made some experiments on Japanese-to-English and English-to-Japanese translation. We compared the outputs of translation with DMAX Criteria and translation with the method using only the statistics of target language. The result indicates that, with DMAX Criteria, we can select equivalent translation more accurately than with the latter method.

4.1 Calculation of Co-occurrence Frequencies

We prepared Japanese newspapers for one year and English journals published in the same year, and calculated inner-sentence co-occurrence frequencies —how frequently the two words co-occur in the same sentence— between nouns appeared in the texts. For example, English words 'government' and 'policy' co-occur 643 times in the same sentence of the English text, and equivalent Japanese words 'seifu' and 'houshin' co-occur 539 times in the same sentence of the Japanese text.

4.2 Algorithm of DMAX Criteria

The algorithm of DMAX Criteria is as follows:

1. Prepare the source language text to translate and the source language-target language bilingual dictionary of the nouns appeared in the text.
   NOTATIONS
   $\mathbf{S}_a(a = 1 \cdots m)$: source language noun
   $\mathbf{T}_{ai}(i = 1 \cdots n_a)$: n_a target language equivalent nouns of $\mathbf{S}_a$
   $\mathbf{COF}(\mathbf{S}_a, \mathbf{S}_b)$: co-occurrence frequency between source language nouns $\mathbf{S}_a, \mathbf{S}_b$
   $\mathbf{COF}(\mathbf{T}_{ai}, \mathbf{T}_{bj})$: co-occurrence frequency between target language nouns $\mathbf{T}_{ai}, \mathbf{T}_{bj}$
2. Select $\mathbf{S}_p, \mathbf{S}_q$ | $\max_{p,q} \mathbf{COF}(\mathbf{S}_p, \mathbf{S}_q)$
   where target language equivalent noun of $\mathbf{S}_p$ or $\mathbf{S}_q$ has not been fixed
3. Select $\mathbf{T}_{pr}, \mathbf{T}_{qs}$ | $\max_{r,s} \mathbf{COF}(\mathbf{T}_{pr}, \mathbf{T}_{qs})$
4. Fix the target language equivalent noun of $\mathbf{S}_p$ as $\mathbf{T}_{pr}$

5. Fix the target language equivalent noun of $\mathbf{S}_q$ as $\mathbf{T}_{qs}$
6. Repeat 2-5 until every target language equivalent noun of $\mathbf{S}_a(a = 1 \cdots m)$ is fixed

And the algorithm of the method using only the statistics of target language is as follows:

1. Prepare the source language text to translate and the source language-target language bilingual dictionary of the nouns appeared in the text.
2. Select $\mathbf{T}_{pr}, \mathbf{T}_{qs} \mid \max_{r,s}\mathbf{COF}(\mathbf{T}_{pr}, \mathbf{T}_{qs})$
   where target language equivalent noun of $\mathbf{S}_p$ or $\mathbf{S}_q$ has not been fixed
3. Fix the target language equivalent noun of $\mathbf{S}_p$ as $\mathbf{T}_{pr}$
4. Fix the target language equivalent noun of $\mathbf{S}_q$ as $\mathbf{T}_{qs}$
5. Repeat 2-4 until every target language equivalent noun of $\mathbf{S}_a(a = 1 \cdots m)$ is fixed

4.3 Experiment 1

According to the algorithm above, we made an experiment on Japanese-to-English and English-to-Japanese translation. We choose 70 Japanese-to-English bilingual sentences from *"Obunsha Hyougen Jiten"*(dictionary of expressions) and extract pairs of corresponding Japanese noun and English noun. Each sentence contains 3 to 10 corresponding noun pairs. We made JtoE and EtoJ translation experiments on the pairs and compared the outputs of translation with DMAX Criteria and translation with the only target language statistics.

Figure 2 shows an EtoJ translation example of this experiment. We extracted 7 corresponding noun pairs from this sentence and translated the 7 English words into Japanese with two methods; one is DMAX Criteria using co-occurrence frequencies both of English and Japanese and the other is the method using only statistics of target language Japanese. We evaluate the out of the translation by three levels, *agree*, *correct* and *incorrect*. *agree* means that the selected Japanese equivalent translation agrees strictly with the word appeared in the original Japanese sentence. *correct* means that the equivalent translation doesn't agree with the original word but its meaning agrees with the one of the original word. *incorrect* means that the meaning of the equivalent translation doesn't agree with the one of the original word. In this example, with DMAX Criteria, we can obtain 5 correct(*agree* or *correct*) translations out of 6, where we can get only 3 correct translations with only target language statistics.

Result of this experiment is shown in Table 1. The frequencies are total of the 70 sentences. The frequency of the word which has only one equivalent translation doesn't included in this result, but we used such words in other words' equivalent translation selection. The result indicates that, with DMAX Criteria, we can select equivalent translation more accurately than with the method using only the statistics of target language.

Table 1 Result of The Experiment 1

translation direction	translation method	evaluation			correct percentage
		agree	*correct*	*incorrect*	
Jap.⇒Eng.	DMAX Criteria	193	56	42	85.6%
Jap.⇒Eng.	only with Eng. COF	186	49	56	80.8%
Eng.⇒Jap.	DMAX Criteria	167	80	82	75.1%
Eng.⇒Jap.	only with Jap. COF	157	71	101	69.3%

Original Sentence:

The reason that prices stabilize at a high level in a seller's market is that there are reductions in the volume of production based on a shortage of raw materials.
売手市場による価格の高値安定化は，原料不足による生産量の減少にその原因がある。

Extracted Corresponding Noun Pairs:

ENGLISH:	market	price	production	reason	reduction	shortage	volume
JAPANESE:	市場	価格	生産	原因	減少	不足	量

English-to-Japanese Word Dictionary:

Eng. Words	*Japanese Equivalent Translations*
market	市 市場 マーケット 食料品店 業界 需要 市況 市価 相場 商機
price	価格 値段 代償 代価 犠牲 歩合 比率 懸賞金
production	製作 生産 産出 生産額 製品 著作 著作物 作品 結果 提供 提出 提示 演出 上演
reason	理由 原因 わけ 動機 道理 理屈 理性
reduction	縮小 削減 割引 減少 低下 下落 格下げ 変形 整理 分類
shortage	不足
volume	本 巻 号 体積 量 容積 容量 統計

Translation with DMAX Criteria:

Eng. Pair	*COF*		*Jap. Pair*	*COF*	*Eng.*	*J.equiv.*	*evaluation*
market–price	[1070]	⇒	市場－価格	[228]	market	市場	*agree*
price–production	[251]	⇒	価格－生産	[174]	price	価格	*agree*
price–reduction	[166]	⇒	価格－下落	[97]	production	生産	*agree*
market–volume	[117]	⇒	市場－量	[30]	reason	わけ	*correct*
market–reason	[108]	⇒	市場－わけ	[80]	reduction	下落	*incorrect*
price–shortage	[33]	⇒	価格－不足	[13]	shortage	不足	–*
					volume	量	*agree*

Translation with only Target Language Statistics:

Eng. Pair		*Jap. Pair*	*COF*	*Eng.*	*J.equiv.*	*evaluation*
market–volume	⇒	市－本	[437]	market	市	*incorrect*
price–production	⇒	価格－生産	[174]	price	価格	*agree*
price–reduction	⇒	価格－下落	[97]	production	生産	*agree*
price–reason	⇒	価格－理由	[57]	reason	理由	*correct*
reason–shortage	⇒	理由－不足	[23]	reduction	下落	*incorrect*
				shortage	不足	–*
				volume	本	*incorrect*

*'shortage' has only one translation equivalent.

Figure 2 English-to-Japanese Translation Example with Two Methods

4.4 Experiment 2

We prepared 5 Japanese newspaper articles and made an experiment of Japanese-to-English translation of the nouns in the articles. The articles consist of 6 to 14 sentences, and the average length of them is 679 Japanese characters. We translated the nouns in four ways:

a) translate all the nouns in the article together with DMAX Criteria

b) translate all the nouns in the article together with only English COF

c) translate the nouns in the same sentence at one time with DMAX Criteria

d) translate the nouns in the same sentence at one time with only English COF

In a) & b), we can utilize not only inner-sentence co-occurrence but also inner-article co-occurrence relations, and can translate the noun of the sentence even if the noun is the only one noun of the sentence. To clear the effect to enlarge the size of the language unit where the two words co-occur, we made comparison between a)&b), a)&c) and b)&d).

Comparison between a)&b)

number of the words to translate	166	
the translations are the same	142	85.5%
DMAX's translation is better	11	6.6%
both correct / both incorrect	10	6.0%
with only Eng. COF is better	3	1.8%

Comparison between a)&c)

number of the words to translate	257	
the translations are the same	226	87.9%
a)'s translation is better	17	6.6%
both correct / both incorrect	9	3.5%
c)'s translation is better	5	1.9%

Comparison between b)&d)

number of the words to translate	257	
the translations are the same	231	89.9%
b)'s translation is better	5	1.9%
both correct / both incorrect	11	4.3%
d)'s translation is better	10	3.9%

The result of this experiment indicates following two features:

- We can select equivalent translation more accurately with DMAX Criteria than with the method using only the statistics of target language.
- Translation with DMAX Criteria can be improved enlarging the size of the language unit where the two words co-occur, while the method using only target language statistics may be degraded with enlargement of the size.

5 Conclusion

We proposed and evaluated a new method Double MAXimize Criteria (DMAX Criteria) based on Dual Corpora Statistics in this paper. It can select suitable translations using statistical data extracted from both source and target language corpora even when linguistic restrictions described in the dictionary or grammar cannot. The dependency relation in the source language is reflected in the translated text through a bilingual dictionary. Moreover, the method has the following features:

1. It utilizes linguistic statistics as context information in addition to logical restrictions effective for ambiguity resolution.
2. The resource of the linguistic statistics is a dual corpora of source and target languages, not a bilingual corpora, i.e. the target language text doesn't have to be the translation of the source language text.
3. Translation from target language text to source language text can be achieved using the same statistics, because the statistics of two texts are independent.
4. The linguistic statistics can be computed semi-automatically in advance.
5. When the equivalent translation list includes a high-frequency word in addition to the suitable translation, the method with only target language statistics tends to choose the high-frequency word as a translation. DMAX Criteria can avoid this problem.
6. When the (co-)occurrence frequency is very low, both DMAX Criteria and the method with only target language statistics cannot select suitable translation. However, with DMAX Criteria, we can check this before applying the method.

We are preparing further experiments on a large-scale hybrid machine translation system with rule-based method and DMAX Criteria for selecting equivalent translation.

References

[Brown 90] P.F.Brown, J.Cocke et al. "A Statistical Approach to Machine Translation", Computational Linguistics, Vol.16, No.2, 1990

[Brown 92] P.F.Brown, S.A.DELLA PIETRA et al. "Analysis, Statistical Transfer, and Synthesis in Machine Translation", TMI-92, pp.83–100, 1992

[Doi 92a] S.DOI, K.MURAKI "Robust Translation and Meaning Interpretation Mechanism based on Examples in Dictionary", 44th Annual Convention of IPSJ, 1P–2, 1992 (in Japanese)

[Doi 92b] S.DOI, K.MURAKI "Hybrid Translation and Meaning Interpretation Mechanism based on Examples and Text Corpora", 6th Annual Conference of JSAI, 15–5, 1992 (in Japanese)

[Doi 92c] S.DOI, K.MURAKI "Translation Ambiguity Resolution based on Text Corpora of Source and Target Languages", COLING-92, pp.525–531, 1992

[Muraki 91] K.MURAKI, S.DOI "Translation Ambiguity Resolution by using Text Corpora of Source and Target Languages", 5th Annual Conference of JSAI, 11–7, 1991 (in Japanese)

[Muraki 92] K.MURAKI, S.DOI "Robust Translation and Meaning Interpretation Mechanism based on Examples in Dictionary", PRICAI92, 1992

[Nagao 84] M.NAGAO "A Framework of a Mechanical Translation between Japanese and English by Analogy Principle", Artificial and Human Intelligence, ed. A.Elithorn and R.Banerji, North-Holland, 1984

[Nomiyama 91] H.NOMIYAMA "Lexical Selection Mechanism Using Target Language Knowledge and Its Learning Ability", IPSJ-WG, NL86-8, 1991 (in Japanese)

[Sadler 90] V.SADLER, R.Vendelmans "Pilot Implementation of a Bilingual Knowledge Bank", COLING-90, Vol.3, pp.449–451 1990

[Sato 90] S.SATO, M.NAGAO "Toward Memory-based Translation", COLING-90, Vol.3, pp.247–252, 1990

[Schabes 92] Y.SCHABES "Stochastic Lexicalized Tree-Adjoining Grammars", COLING-92, pp.425–432, 1992

[Sumita 92] E.SUMITA, H.IIDA "Example-Based NLP Techniques -A Case Study of Machine Translation–", AAAI-92 Workshop "Statistically-Based NLP Techniques", AAAI Technical Report W-92-01, 1992