Constraint Optimization Approach to Context Based Word Selection

Jun Matsuno **Toru Ishida**
Department of Social Informatics,
Kyoto University, Kyoto 6068501, Japan
matsuno@ai.soc.i.kyoto-u.ac.jp ishida@i.kyoto-u.ac.jp

Abstract

Consistent word selection in machine translation is currently realized by resolving word sense ambiguity through the context of a single sentence or neighboring sentences. However, consistent word selection over the whole article has yet to be achieved. Consistency over the whole article is extremely important when applying machine translation to collectively developed documents like Wikipedia. In this paper, we propose to consider constraints between words in the whole article based on their semantic relatedness and contextual distance. The proposed method is successfully implemented in both statistical and rule-based translators. We evaluate those systems by translating 100 articles in the English Wikipedia into Japanese. The results show that the ratio of appropriate word selection for common nouns increased to around 75% with our method, while it was around 55% without our method.

1 Introduction

Activities are being conducted to improve the accessibility and usability of language services for intercultural collaboration to overcome language and cultural barriers with Language Grid [Ishida, 2006]. We are developing a multilingual environment for the translation of Wikipedia articles in cooperation with the Wikimedia Foundation. However, during this period, we have observed that output words selected by automatic machine translation systems, in both statistical machine translation (SMT) and rule-based machine translation (RBMT), are not consistent. For example, when machine translating the English Wikipedia article "George Washington" into Japanese, 18 nouns appear multiple times and are translated with different meanings. Although 5 of these nouns are context-dependent, the remaining 13 should have consistent Japanese equivalents. Inconsistency in word selection is a major problem since it prevents the user from recovering the meaning of the source text [Yamashita and Ishida, 2006; Tanaka *et al.*, 2009]. Take for example the machine translation of an English document that reads "The paper is excellent. I want to know about the author of the paper." into the Japanese "sono kami ha subarashii. watashiwa, ronbun no chosha wo siri tai. (The sheet of paper is excellent. I want to know about the author of the scientific paper.)" . The word "paper" should be translated into "ronbun (a scientific paper)" in both the first and the second sentences, but "paper" is translated into "kami (a sheet of paper)" in the first sentence. Richer contextual information is needed if we are to resolve inconsistency in word selection. In this example, the machine translation result of a single sentence was inadequate because of the failure to apply global contextual information.

Methods that improve statistical machine translation quality by using word sense disambiguation (WSD) have been proposed in the field of machine translation with contextual information [Carpuat and Wu, 2007; Chan *et al.*, 2007]. These methods, however, consider the contextual information of only neighboring sentences, and the contextual information available in the whole article is not used. Machine learning is the dominant approach in WSD, and huge features have to be treated if sentences other than neighboring sentences are used as the sources of contextual information. Moreover, it is difficult to prepare a sufficiently large training data set to give each feature an appropriate weight.

This paper proposes a word selection method based on constraint optimization. The constraint optimization problem demands that each constraint be weighted according to its degree of importance. A method that applies constraint optimization to word selection has been proposed, but it is unable to use the context of the whole article because constraint is based on single sentences [Canisius and Bosch, 2009]. As a result, consistent word selection can not be performed over the whole article. However, in the constraint optimization approach, it should be possible to use contextual information from the whole article because a variable is assigned to each word appearing in a document and word selection based on constraints between variables is performed. Thus, we propose the use of constraints between words in the whole translated article based on semantic relatedness and contextual distance between words; we resolve word sense ambiguity by using contextual information in the whole translated article. As far as we know, this study is the first to use the context of the whole article for ensuring word consistency.

2 Semantic Relatedness Between Translated Words in a Single Sentence

We formulate the word selection problem based on the weighted constraint satisfaction problem [Bistarelli *et al.*, 1997], one of the constraint optimization problems, to resolve inconsistency in word selection in the machine translation of a document. In this formulation, ambiguity in the sense of a noun in the original document is resolved by using the semantic relatedness between words in each translated sentence. That is, independent word selection is performed for each sentence by using contextual information in a single sentence. We enumerate the requirements for word selection below, and formulate the word selection problem so that it can meet those requirements.

1. The translation candidates of noun w in the original document are all translated nouns of w in the translated document
2. There is semantic relatedness between translated words in the same sentence
3. A solution is the assignment of translated words to the nouns in the original document that maximizes the sum of semantic relatedness between translated words

From requirement 1, one variable x is created for each noun w in the original document, and all translated nouns of w in the translated document are included in a domain D for each variable. From requirement 2, the constraint representing "there is semantic relatedness between translated words" is imposed between x_i and x_j if the original words of x_i and x_j co-occur in the same sentence $(1 \leq i < j \leq n)$. This semantic relatedness is computed quantitatively by function SR.

We use the method of computing semantic relatedness, employed by Wikipedia [Gabrilovich and Markovitch, 2007], to compute function SR. In this method, the relative strengths between x_i and each Wikipedia article are determined by using the tf/idf score based on the number of occurrences of x_i in each article of Wikipedia in the translated language, and a translated word vector weighted for each article $v_{x_i} = (v_{x_i1}, v_{x_i2}, \ldots, v_{x_im})$ is obtained (m is the number of articles in Wikipedia in the translated language). Specifically, x_i appears $tf(i,k)$ times in the k th of the m articles, and appears in l articles. v_{x_ik} is computed as $v_{x_ik} = (1 + \log tf(i,k)) \log \frac{m}{l}$. A translated word vector v_{x_i} is obtained by performing this calculation for all articles. Semantic relatedness between translated words is expressed quantitatively by a value that is not less than 0 and not more than 1 by computing the cosine similarity between v_{x_i} and v_{x_j} which are, respectively, translated word vectors for x_i and x_j. Accordingly, $SR_{ij}(x_i, x_j)$ is determined as:

$$SR(x_i, x_j) = \frac{v_{x_i1}v_{x_j1} + \cdots + v_{x_im}v_{x_jm}}{\sqrt{v_{x_i1}^2 + \cdots + v_{x_im}^2}\sqrt{v_{x_j1}^2 + \cdots + v_{x_jm}^2}}$$

The average of the values of function SR for all pairs of variables in which the constraint is imposed is expressed as:

$$ASR(X) = \frac{\sum_{\{i,j\} \in V} SR(x_i, x_j)}{|V|}$$

(Set V consists of the pairs of indexes that correspond to the pairs of variables in which constraints are imposed.)
The larger the value of function ASR is, the larger the sum of semantic relatedness between translated words in each sentence is. Therefore, context-dependent word selection is performed for each sentence in the original document when the value of function ASR is largest. From requirement 3, the optimal solution for this problem is the tuple of translated words for the variables with maximum value of function ASR.

3 Semantic Relatedness Between Translated Words in a Document

It is thought that semantic relatedness between translated words which appear in the same sentence is really large. However, even if translated words appear in different sentences, there should be semantic relatedness between translated words according to the closeness between the contexts in which translated words appear in a document. It is expected that more accurate word selection will be realized by using the semantic relatedness between words in the translated document. We adopt this approach to formulate the word selection problem based on the weighted constraint satisfaction problem. Word selection using contextual information in the whole article is performed by solving this word selection problem. We enumerate the requirements that the word selection problem should meet below.

1. The translation candidates of noun w in the original document are all translated nouns of w in the translated document
2. There is context-dependent semantic relatedness between translated words in the same document
3. A solution is an assignment of translated words to the nouns in the original document that maximize the sum of context-dependent semantic relatedness between translated words

From requirement 1, one variable x is created for each noun w that appears in the original document, and all translated nouns of w in the translated document are included in domain D for each variable. From requirement 2, constraints representing "there is context-dependent semantic relatedness between translated words" are imposed between x_i and x_j if the original words of x_i and x_j co-occur in the same document $(1 \leq i < j \leq n)$. This context-dependent semantic relatedness is computed quantitatively by function CSR which is based on function SR. Function CSR becomes important when applying machine translation to collectively developed documents like Wikipedia.

We now turn to the computational model of function CSR to compute context-dependent semantic relatedness between translated words tw and tw' whose original words are, respectively, w and w' in the same document. First, semantic relatedness SR(tw, tw') between tw, tw' is not less than 0 and not more than 1, and context-dependent semantic relatedness CSR(tw, tw') between tw, tw' does not exceed context-independent semantic relatedness SR(tw, tw'). Namely, the closer the contexts in which tw and tw' appear in a document are, the more the value of CSR approaches that of SR. In addition, we consider that the closeness of the contexts in which tw and tw' appear in the translated document is equivalent to the closeness of the contexts between the sentences in which w and w' appear in the original document. We call this contextual distance. The value of contextual distance is larger than 0, and the smaller the value is, the closer the contexts are. To express the requirements for the computational model of CSR, We describe tw and tw2 as the translations of the same two words, w, that appear in different locations of the original document, and describe tw' as the translated word of word w' in the same original document. Additionally, we describe s as a function that expresses the sentence in which the original word of the translated word appears by accepting a translated word as input, and describe DIS as a function which expresses contextual distance between these sentences upon receiving the two sentences as input. We use the following mathematical expressions to enumerate the requirements for the computational model of CSR.

1. $0 \leq SR(\text{tw,tw'}) \leq 1$
2. $0 \leq DIS(s(\text{tw}), s(\text{tw'}))$
3. $0 \leq CSR(\text{tw,tw'}) \leq SR(\text{tw,tw'})$
4. $DIS(s(\text{tw}), s(\text{tw'})) = 0$
   $\Longrightarrow CSR(\text{tw,tw'}) = SR(\text{tw,tw'})$
5. $DIS(s(\text{tw}), s(\text{tw'})) \leq DIS(s(\text{tw2}), s(\text{tw'}))$
   $\Longrightarrow CSR(\text{tw,tw'}) \geq CSR(\text{tw2,tw'})$

Our computational expression of CSR, shown in Figure 1, meets these requirements.

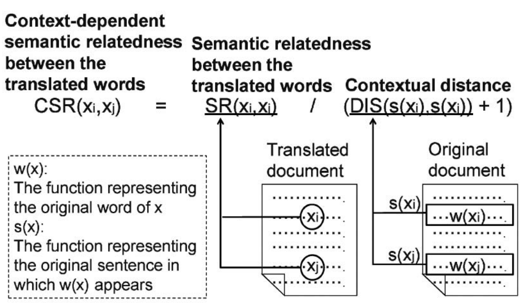

Figure 1: Computation of context-dependent semantic relatedness between translated words

We describe num as a function which expresses the order of the sentence in the article upon receiving an original sentence as input. The order of the sentence is the number of the sentence counting from the beginning of the article. Function DIS is simply based on the physical distance between original sentences as below.

$$DIS(s(x_i), s(x_j)) = num(s(x_j)) - num(s(x_i))$$

The average of the values of function CSR for all pairs of variables is expressed as below.

$$ACSR(X) = \frac{\sum_{j=i+1}^{j=n} \sum_{i=1}^{i=n} CSR(x_i, x_j)}{{}_nC_2}$$

Function $ACSR$ computes the average of the measurement of semantic relatedness between translated words in the whole translated article. The value of function $ACSR$ represents how a translated word which has a context-dependent meaning is selected for each noun in the original document. It also means that the value of function $ACSR$ represents how the same translated word that has the appropriate meaning is selected for the same nouns that have the same meaning in the original document. From requirement 3, the optimal solution for this problem is the tuple of translated words for the variables that maximize the value of function $ACSR$. Figure 2 formulates the word selection problem using semantic relatedness between translated words in a document.

Variable Set $X = \{x_1, \ldots, x_n\}$
(x_i:The translated word of the noun which appears in i th order in the original document)
Domain Set $D = \{D_1, \ldots, D_n\}$
(D_i:The set whose elements are all translated nouns of $w(x_i)$ in the translated document
$w(x)$:The function expressing the original word of translated word x)

The function expressing semantic relatedness between translated words
$SR_{ij}(x_i, x_j) = \frac{v_{x_i 1} v_{x_j 1} + \cdots + v_{x_i m} v_{x_j m}}{\sqrt{v_{x_i 1}^2 + \cdots + v_{x_i m}^2}\sqrt{v_{x_j 1}^2 + \cdots + v_{x_j m}^2}}$
($v_{x_k l}$:The weight of x_k for the l th of m articles in Wikipedia in the translated language
m:The number of articles in Wikipedia in the translated language)
The function expressing contextual distance between original sentences
$DIS(s(x_i), s(x_j)) = num(s(x_j)) - num(s(x_i))$
($s(x)$:The function expressing the sentence in which the original word of translated word x appears
$num(s(x))$:The function expressing the order of sentence $s(x)$ which appears in the document)
The function expressing context-dependent semantic relatedness between translated words
$CSR(x_i, x_j) = \frac{SR(x_i, x_j)}{DIS(s(x_i), s(x_j)) + 1}$

The function expressing how inconsistency in word selection is resolved
$ACSR(X) = \frac{\sum_{j=i+1}^{j=n} \sum_{i=1}^{i=n} CSR(x_i, x_j)}{{}_nC_2}$
Optimal Solution
The tuple of translated words for the variables with maximum $ACSR(X)$

Figure 2: Formulation of the word selection problem using semantic relatedness between translated words in a document

4 Example of the Word Selection Problem

We give an example of the word selection problem in Figure 3. Figure 4 and Figure 5 show the constraint networks yielded when this word selection problem is formulated by using the semantic relatedness between translated words in a single sentence and in a document, respectively.

Source document (English): Inuit people have their own peculiar language. However, peoples with different languages do not always have different cultures.
Translated document (Japanese): inuitto no hitobito ha karerajishin no tokuyuuna gengo wo motte imasu.
(Inuit folks have their own peculiar language.)
shikashi, kotonaru gengo wo motu minzoku ha tsuneni kotonaru bunka wo motte inai.
(However, ethnic groups with different languages do not always have different cultures.)

Figure 3: English-Japanese machine translated document in which inconsistency in word selection of "people" occurs

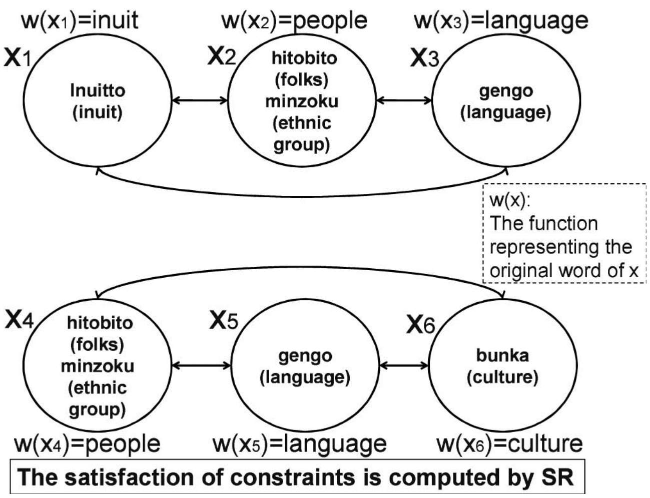

Figure 4: Constraint network representing the word selection problem of Figure 3 which is formulated using semantic relatedness between translated words in a single sentense

In Figure 4, the semantic relatedness between translated words in each sentence is computed, and word selection is independently performed for each sentence. The values of function SR for the pair of translated words are, for example, SR("inuitto(inuit)", "hitobito(folks)") $= 0.0241$ and SR("inuitto(inuit)", "minzoku(ethnic group)") $= 0.0524$. The value of function SR for the pair of "inuitto(inuit)" and "minzoku(ethnic group)" is more than twice that for the pair of "inuitto(inuit)" and "hitobito(folks)".In Figure 5, context-dependent semantic relatedness between words in the translated document is computed, and word selection using contextual information in the whole document is performed. If $x_2 =$ "hitobito(folks)" and $x_4 =$ "minzoku(ethnic group)", the values of function CSR for the pair of x_1 and x_2 and for the pair of x_1 and x_4 are calculated to be, respectively, CSR(("inuitto(inuit)","hitobito(folks)") $=$ 0.0241 and CSR("inuitto(inuit)","minzoku(ethnic group)") $= 0.0262$. The original words of x_1 and x_2 appear in the same sentence,

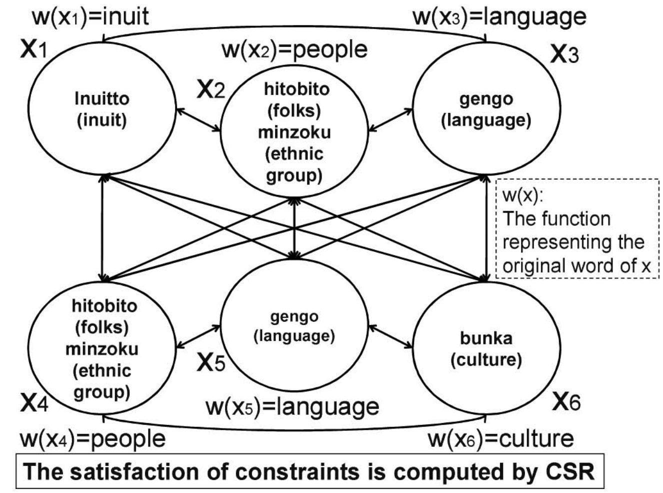

Figure 5: Constraint network representing the word selection problem of Figure 3 which is formulated using semantic relatedness between translated words in a document

but those of x_1 and x_4 appear in different sentences. Accordingly, the value of context-dependent semantic relatedness between "inuitto(inuit)" and "minzoku(ethnic group)" is not much larger than that between "inuitto(inuit)" and "hitobito(folks)".

The translated word that should be selected for $w(x_2)$ and $w(x_4)$ is "minzoku(ethnic group)". Although "minzoku(ethnic group)" and "hitobito(folks)" are selected for $w(x_2)$ and $w(x_4)$, respectively, in the word selection problem represented by the constraint network of Figure 4, "minzoku(ethnic group)" is selected for both $w(x_2)$ and $w(x_4)$ in the word selection problem represented by the constraint network of Figure 5. This is because the semantic relatedness between the translated word of $w(x_4)$ and "inuitto(inuit)" which has strong semantic relatedness with "minzoku(ethnic group)", which is the appropriate translated word for $w(x_4)$, is used in the word selection problem represented by the constraint network of Figure 5.

5 Evaluation

5.1 Evaluation Settings

We implemented the systems of WSD/SR(sentence) and WSD/CSR(article) to formulate the word selection problem using semantic relatedness between translated words in a single sentence and a document, respectively, and resolved the word selection problem by applying the hill climbing approach. Furthermore, we implemented WSD/SR(article). WSD/SR(article) is different from WSD/CSR(article) in that function SR is used instead of CSR to compute the semantic relatedness between translated words. By comparing the evaluation results of WSD/SR(article) and WSD/CSR(article), we can better understand the effectiveness of using function CSR which becomes important when applying machine translation to collectively developed documents like Wikipedia. We used Google Translate[1] and J-Server[2] as examples of SMT and RBMT systems, and used 100 samples which were randomly selected from English Wikipedia articles whose bodies contained more than 500 words as the source documents.

5.2 Evaluation Results

Table 1 shows (a) "the total number of appearances of all common nouns" when translating the 100 samples by Google Translate and J-Server. The common nouns that were included in (a) had different meanings for the translated words selected by machine translation in each document. Table 2 and Table 3 show the number of nouns that were appropriately translated (a) when Google Translate and J-Server were used, respectively.

Table 1: Number of common nouns evaluated

	Google Translate	J-Server
(a)	427	369

(a)"the total number of appearances of all common nouns" (These common nouns had different meanings for the translated words selected by machine translation in each document)

The followings are shown from the evaluation results.

- Both Google Translate and J-Server performed appropriate word selection at the rate of about 55%.
- WSD/SR(sentence) improved word selection quality by 10 points by using contextual information in single sentences. However, the translations still had a word selection rate of about 35%.
- WSD/SR(article) selected the same translated word for the same nouns in the same document by computing semantic relatedness rather than contextual distance although WSD/SR(sentence) selected translated words independently in each sentence. Therefore, WSD/SR(article) consistently selected inappropriate translated words for nouns for which the same translated word should have been selected, and WSD/SR(article) decreased word selection quality more than WSD/SR(sentence) in some cases.
- WSD/CSR(article) yielded better word selection quality than WSD/SR(article) because it uses richer contextual distance to compute semantic relatedness. As a result, WSD/CSR(article) was the best system in terms of word selection quality.

However, we regarded the translation candidates of a word as all translated words which the machine translation system selected for the word in the same document. Therefore, WSD/CSR(article) sometimes failed to select appropriate translated words because appropriate translated words were not included in their translation candidates. Extracting translation candidates from bilingual dictionaries may improve word selection quality.

[1]http://translate.google.co.jp/

[2]http://www3.j-server.com/KODENSHA/contents/entrial/index.htm

Table 2: Comparative evaluation of word selection quality for Google Translate

System	The number of nouns that were appropriately translated
Google Translate	245(57.4%)
+ WSD/SR(sentence)	274(64.2%)
+ WSD/SR(article)	306(71.7%)
+ WSD/CSR(article)	313(73.3%)

Table 3: Comparative evaluation of word selection quality for J-Server

System	The number of nouns that were appropriately translated
J-Server	200(53.9%)
+ WSD/SR(sentence)	241(65.0%)
+ WSD/SR(article)	240(64.5%)
+ WSD/CSR(article)	271(72.9%)

6 Related Work

Existing WSD studies attempt to identify the correct meaning of a polysemous word by using context. Carpuat and Wu [2005] proposed a method that uses words selected by WSD to replace words in a machine translated sentence. They verified whether WSD could improve the translation quality of statistical machine translation (SMT) in the translation of a single sentence or not. The evaluation results using BLEU metric, which is an automatic evaluation method, showed that using WSD decreased the translation quality of SMT. This was because the word replacement degraded the fluency of the sentence. Our method also replaces translated words so we need to manually evaluate the translation quality of the resulting sentences.

In [Carpuat and Wu, 2005], it was shown that the direct use of WSD for SMT could not improve translation quality. Methods that improve the translation quality of SMT by coordinating a WSD model and statistical models of SMT have been proposed [Carpuat and Wu, 2007; Chan *et al.*, 2007]. However, in [Carpuat and Wu, 2007], contextual information from only the original sentence was used for WSD. In [Chan *et al.*, 2007], contextual information in multiple sentences was used for WSD, but sentences that were used as contextual information were limited to the original sentence and the immediately adjoining sentences. This is because a WSD method based on machine learning, such as a support vector machine, needs an impractically large training data set if sentences other than an original sentence and its neighboring sentences are used for WSD. In these methods, consistent word selection is not performed over the whole article because contextual information from the whole article is not used.

SMT methods select translation rules based on context by using the wealth of contextual information available in translation rules and syntax trees have been recently proposed [He *et al.*, 2008; Liu *et al.*, 2008; Shen *et al.*, 2009]D However, using contextual information obtained in the production process of sentences demands the existence of a large training data set. Moreover, these methods select translation rules based on context, while our method uses context to resolve word sense ambiguity.

Our method performs word selection based on the weighted constraint satisfaction problem. Canisius and Bosch [2009] proposed a method that improves the translation quality of SMT based on the weighted constraint satisfaction problem. In this method, constraints on the connections between translated words are initially obtained from a corpus. The line of translated words that maximizes the translation score while satisfying the constraints is produced as the translation output sentence. Therefore, imposing constraints between words in a translated sentence enables the use of contextual information in a translated sentence. In our method, constraints indicating that there is semantic relatedness between words are imposed between words throughout the whole translated article. In addition, constraints are weighted by the degree of importance of the contextual information according to semantic relatedness and contextual distance between words. This realizes word selection based on contextual information from the whole translated article.

7 Conclusion

Inconsistency in word selection is a problem that occurs when the instances of one source word are given different translations. Consistent word selection can be realized for the translation of documents like Wikipedia by resolving this problem. Contextual information taken from the whole article must be used to resolve this problem. We proposed a word selection method based on constraint optimization. Our method can suppress inconsistency in word selection by using contextual information from the whole article, not just single sentences.

Evaluations on Wikipedia articles showed that our method was effective for both statistical and rule-based translators. The ratio of appropriate word selection for common nouns was around 55% with previous approaches. However, it was around 75% with our method. Using contextual information from the whole document improves the word selection quality of machine translations. We will evaluate the translation quality in terms of fluency to highlight the benefits of our method.

Acknowledgments

This research was supported by Strategic Information and Communications R&D Promotion Programme (SCOPE) from Ministry of Internal Affairs and Communications of Japan and a Grant-in-Aid for Scientific Research (A) (21240014, 2009-2011) from Japan Society for the Promotion of Science (JSPS).

References

[Bistarelli *et al.*, 1997] Stefano Bistarelli, Ugo Montanari and Francesca Rossi. Semiring-Based Constraint Satisfaction and Optimization. *Journal of the Association of Computing Machinery(JACM), vol.44 no.2*, pages 201-236, 1997.

[Canisius and Bosch, 2009] Sander Canisius and Antal van den Bosch. A Constraint Satisfaction Approach to Machine Translation. *In Proceedings of the 13th Annual Conference of the European Association for Machine Translation(EAMT-09)*, pages 182-189, 2009.

[Carpuat and Wu, 2005] Marine Carpuat and Dekai Wu. Word Sense Disambiguation vs. Statistical Machine Translation. *In Proceedings of the 43th Annual Meeting of the Association of Computational Linguistics(ACL-05)*, pages 387-394, 2005.

[Carpuat and Wu, 2007] Marine Carpuat and Dekai Wu. Context-Dependent Phrasal Translation Lexicons for Statistical Machine Translation. *In Proceedings of Machine Translation Summit XI*, pages 73-80, 2007.

[Chan *et al.*, 2007] Yee Seng Chan, Hwee Tou Ng and David Chiang. Word Sense Disambiguation Improves Statistical Machine Translation. *In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics(ACL-07)*, pages 33-40, 2007.

[Gabrilovich and Markovitch, 2007] Evgeniy Gabrilovich and Shaul Markovitch. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. *In Proceedings of the 20th International Joint Conference on Artificial Intelligence(IJCAI-07)*, pages 1606-1611, 2007.

[He *et al.*, 2008] Zhongjun He, Qun Liu and Shouxun Lin. Improving Statistical Machine Translation using Lexicalized Rule Selection. *In Proceedings of the 22nd International Conference on Computational Linguistics(COLING-08)*, pages 321-328, 2008.

[Ishida, 2006] Toru Ishida. Language Grid: An Infrastructure for Intercultural Collaboration. *IEEE/IPSJ Symposium on Applications and the Internet(SAINT-06)*, pages 96-100, 2006.

[Liu *et al.*, 2008] Qun Liu, Zhongjun He, Yang Liu and Shouxun Lin. Maximum Entropy based Rule Selection Model for Syntax-based Statistical Machine Translation. *In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing(EMNLP-08)*, pages 89-97, 2008.

[Shen *et al.*, 2009] Libin Shen, Jinxi Xu, Bing Zhang, Spyros Matsoukas and Ralph Weischedel. Effective Use of Linguistic and Contextual Information for Statistical Machine Translation. *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing(EMNLP-09)*, pages 72-80, 2009.

[Tanaka *et al.*, 2009] Rie Tanaka, Yohei Murakami and Toru Ishida. Context-Based Approach for Pivot Translation Services. *In Proceedings of the 21st International Joint Conference on Artificial Intelligence(IJCAI-09)*, pages 1555-1561, 2009.

[Yamashita and Ishida, 2006] Naomi Yamashita and Toru Ishida. Effects of Machine Translation on Collaborative Work. *In Proceedings of International Conference on Computer Supported Cooperative Work(CSCW-06)*, pages 515-523, 2006.