Context-Based Approach for Pivot Translation Services

Rie Tanaka
C&C Innovation Research Laboratories, NEC Corporation, Nara, 6300101, Japan
r-tanaka@ak.jp.nec.com

Yohei Murakami
National Institute of Information and Communications Technology (NICT), Kyoto, 6190289, Japan
yohei@nict.go.jp

Toru Ishida
Department of Social Informatics, Kyoto University, Kyoto 6068501, Japan
ishida@i.kyoto-u.ac.jp

Abstract

Machine translation services available on the Web are becoming increasingly popular. However, a *pivot translation service* is required to realize translations between non-English languages by cascading different translation services via English. As a result, the meaning of words often drifts due to the *inconsistency*, *asymmetry* and *intransitivity* of word selections among translation services. In this paper, we propose context-based coordination to maintain the consistency of word meanings during pivot translation services. First, we propose a method to automatically generate multilingual equivalent terms based on bilingual dictionaries and use generated terms to propagate context among combined translation services. Second, we show a multiagent architecture as one way of implementation, wherein a coordinator agent gathers and propagates context from/to a translation agent. We generated trilingual equivalent noun terms and implemented a Japanese-to-German-and-back translation, cascading into four translation services. The evaluation results showed that the generated terms can cover over 58% of all nouns. The translation quality was improved by 40% for all sentences, and the quality rating for all sentences increased by an average of 0.47 points on a five-point scale. These results indicate that we can realize consistent pivot translation services through context-based coordination based on existing services.

1 Introduction

Recently, the number of languages used in Web pages has increased rapidly. People using English on the Internet now comprise 30% of all Internet users; users of Asian languages comprise 26%; users of European languages excluding English comprise 25%; and users of all other languages comprise 20%.[1] This trend introduces the requirement for translations between non-English languages in addition to between English and non-English languages. Although the increase in the number of online translation services enables people to access machine translations easily, it is practically impossible to cover all combinations of n languages as the development of (n^2-n) direct translation services would be extremely costly. The pivot translation service generated by combining multiple translation services via a pivot language is a practical solution for such situation.

However, pivot translation often yields drifting for the meanings of words because of inconsistent word selection, making it difficult for users to continue communication. Establishing common ground among users in machine-translation-mediated communication is known to be difficult [Yamashita et al., 2009]; one of the causes of difficulty is inconsistent word selection [Yamashita and Ishida, 2006].

In phrase-based statistical machine translation (SMT), methods for pivot translation with no direct corpora between the source and target languages have been proposed [Utiyama and Isahara, 2007; Wu and Wang, 2007]. In their approach, the phrase-table required for SMT between the source and target languages is generated by combining phrase-tables between the source and pivot languages and the pivot and target languages. The phrase and lexical translation probabilities in the new table are estimated from original corpora, enabling more accurate selection of translated phrases. In the other approach for word selection problems, Kanayama and Watanabe [2003] proposed the linguistic annotation method. They embedded lexical and syntactical information for a source sentence into the intermediated sentence to assure the correctness of the pivot translation. However, the above approaches are not available immediately in practice because it is not easy to prepare the enormous and reliable corpora required to merge phrase tables or to apply the linguistic approach to all translation services. In contrast, we propose a method to realize consistent translation with available dictionaries and translation services.

To coordinate existing translation services, this study used the framework of service computing. In Web service composition, the *WS-Coordination* (Web Services Coordination)[2]

---

[1] The latest estimation of Internet users by language, carried out in May 2008 by Internet World Stats. See:
http://www.internetworldstats.com/stats7.htm

[2] http://www-106.ibm.com/developerworks/library/ws-coor

specification enables the propagation of the service ID or port number as "CoordinationContext" to solve the semantic problems of service composition; it is also used to match input and output data types automatically [Hassine et al., 2006]. Moreover, the method of meta-level control for composite Web services in an open environment, known as "Service Supervision," has been proposed for designers who are not authorized to modify each component Web service [Tanaka et al., 2009]. In terms of improving the performance of composite Web services, a context-aware approach called situated Web service (SiWS) has been proposed to improve the performance of Web services with diverse interfaces and various clients [Matsumura et al., 2006]. We took this type of approach to coordinate word selection of whole component services with context from outside the Web services. In the development of machine translations or language resources, Bramantoro et al. [2008] proposed a method to combine language resources and middleware architecture to integrate deep and shallow natural language processing components. This approach uses both language resources and language processing component as Web services: our context-based coordination approach can contribute towards the improvement of combined services in such areas.

To solve the word selection problem in pivot translation services, we propose the *context-based coordination* method for translation services. We regard the internal translation processes of services as black boxes and realize the coordination outside the services instead of proposing a new machine translation technology. This study addresses the following issues.

Context-Based Coordination with Propagated Context

To ensure consistency in word selection, we propose the propagation of context across cascaded translation services by regarding the context as a set of multilingual equivalent terms. In the research area of bilingual dictionaries, methods to match the meanings of the words of different languages by combining multiple dictionaries are proposed. We refer to those methods and propose a method to generate the multilingual equivalent terms automatically based on commercially available bilingual dictionaries.

Multiagent Architecture for Coordination

This paper proposes a multiagent architecture as one way to implement context-based coordination, wherein the coordinator agent gathers and propagates the context from/to translation agents.

We implemented a coordinated Japanese-to-German-and-back translation service by cascading four translation services and obtained results indicating that the translation quality improved substantially. The advantage of this approach is that high-quality translations can be extracted from existing translation services with existing bilingual dictionaries without modifying their internal coding systems.

<Case 1>
Source sentence (English): Please add that picture in this paper.
□ Translation (Japanese): *douzo, sono shashin wo kono ronbun no naka ni tsuika shinasai.*
(Please add that picture in this thesis.)
<Case 2>
Source sentence (English): Please send me this paper.
□ Translation (Japanese): *douzo, kono kami wo watashi ni okuri nasai.*
(Please send me this paper.)

(a) Inconsistency in word selection

• Japanese user (Japanese): *kinou watashi tachi ha pa-thi wo sita.*
(We had a party yesterday.)
□ Translation (English): There was a party yesterday.

• English user (English): How was the party?
□ Translation (Japanese): *tou ha doudesita ka?*
(How was the political party?)

(b) Asymmetry in word selection

Source sentence (Japanese): *kanojo no ketten ha ookina mondai da.*
(Her fault is a big problem.)
□ Translation (English): Her fault is a big problem.
□ Translation (German): *Ihre Schuld ist ein großes Problem.*
(Her responsibility is a big problem.)

(c) Intransitivity in word selection
Figure 1. Issues in composite translation services

2 Overview of Context-Based Approach

2.1 Issues in Composite Translation Services

We conducted several experiments using the Language Grid [Ishida, 2006] and classified word selection errors into three categories: inconsistency, asymmetry, and intransitivity. *Inconsistency* is when translations of the same source word vary in different sentences. *Asymmetry* is when the back-translated word is different from the source word. The impact of these errors on communication has already been analyzed [Yamashita and Ishida, 2006]. Quantitative results with interview data show that lexical entrainment [Brennan and Clark, 1996] is disrupted by asymmetries in machine translations since they interfere with echoing. *Intransitivity* is when the word sense drifts across the cascaded machine translators.

Figure 1 presents examples of common problems encountered by cascaded translation services. All original Japanese and German sentences in this paper are italicized and their English translations are provided in parentheses. (a) is an example of inconsistency, wherein the English word "paper" is translated to the Japanese word *ronbun* (thesis) in Case 1, while the same word is translated into *kami* (paper) in Case 2. Asymmetry is presented in (b). In the first step of the machine translation-mediated communication, the Japanese word *pa-thi* (party), which means a social gathering, is

translated into English correctly. However, when an English user echoes the word "party," it is translated into the Japanese word *tou* (political party). Intransitivity is presented in (c). The Japanese word *ketten* (fault), which means a weakness of character, is translated into English correctly, but mistranslated to the German word *Schuld* (responsibility). This is because the intermediate English word "fault" has several meanings, and the English-German translator does not have any knowledge of the context for the preceding Japanese-English translation.

2.2 Context-Based Pivot Translation Service with Multiagent Architecture

We propose a multiagent architecture for context-based pivot translation service, as shown in figure 2. The coordinator agent, which plays the role of controlling the whole translation, gathers and propagates context from/to the translation agents in addition to requesting them to translate the sentence. It possesses all possible contexts internally, selects all contexts that suit the context reported by the translation agent, and transfers them to the next translation agent. Translation agents possess the in-built functionality for the original translation service; they perform translations by taking into account the context provided by the coordinator agent, update the context, and transfer the result to the coordinator agent. They have knowledge of the languages and make language-specific processes or decisions. By using the agent framework, more advanced improvements are possible: for instance, adding the ability to interact with users in order to identify the context of the sentence.

Context can be represented in several ways, such as a set of characteristic words in a document, surrounding text, or talk of an expression. Since context in one language can be translated to other languages with multilingual equivalent terms, we represent context by sets of equivalent terms, not sets of terms in one language. In our architecture, we consider a set of terms in the source sentence as context in the source language and use equivalent terms as propagated context.

3 Generating Multilingual Equivalent Terms

The set of equivalent terms can be generated by analyzing generic bilingual dictionaries.[3] However, since it is costly and difficult to manually develop multilingual dictionaries that include all words in all languages, we require an automated method to develop such a dictionary. In previous work on this subject, the concepts for different languages were matched using bilingual dictionaries [Tokunaga and Tanaka, 1990]. We extended this idea to generate a set of trilingual equivalent terms (referred to hereafter as a *triple*). We represent mappings of words belonging to different languages in the form of a graph; a word is represented as a vertex, and a

---

[3] Multilingual equivalent terms can also be developed manually, as in the case of• EuroWordNet [Vossen, 1998].
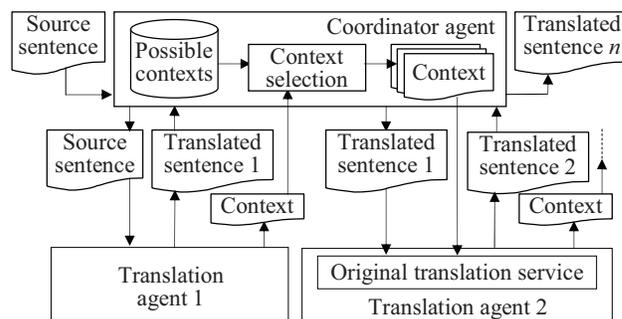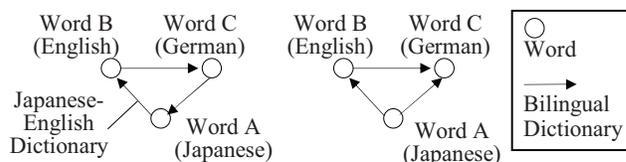

Figure 2. Multiagent architecture for context-based approach

(a) Loop triangle   (b) Transition triangle

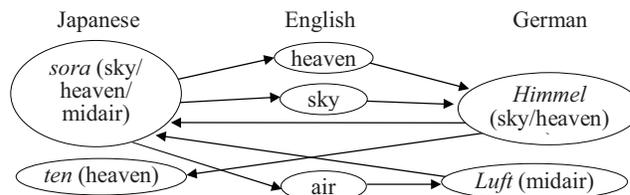Figure 3. Two types of shapes of triangles

Figure 4. A loop triangle representing the sense of "sky•"

mapping in bilingual dictionaries is represented as a directed edge. If the graph contains a triangle, the three words are considered equivalent terms. Figure 3 shows the two types of triangles: *loop* and *transition*. The loop triangle starts from a source language, looks up dictionaries three times, and returning to the source language. The transition triangle starts from a source language and looks up dictionaries to locate transitive and direct routes between the source and target languages. It is easy to generate a triple from such triangles. We call such triples generated from loop triangles loop-type triples hereafter.

Example 1 (A loop triangle representing "sky")
Figure 4 shows an example of a loop triangle, starting with the Japanese word *sora* (sky/heaven/midair). Words such as "sky" are extracted by looking up a Japanese-English dictionary. The German word *Himmel* (sky/heaven) is obtained by looking up the word "sky" in an English-German dictionary. Since the source Japanese word is extracted from a German-Japanese dictionary, {*sora* (sky/heaven/midair), sky, *Himmel* (sky/heaven)} is considered as a triple. Continuing this process further yields other triples.

Algorithm 1: COORDINATOR-AGENT *CA*

1: s_i /* Source sentence */
2: o_i /* A word in sentence s_i */
3: *MTA* /* An ordered list of translation agents
             ($MTA = \{MTA_1, MTA_2, ..., MTA_n\}$) */
4: $MTA_i = \{(s_i, s_{i+1})\}$ /* A translation agent; a set of pairs of
             sentence s_i and s_{i+1} */
5: T_i /* A set of *n*-tuples $(w_1, w_2, ..., w_n)$, where w_k is included in s_k
             ($k\square i$); All *n*-tuples are *n*-lingual equivalent terms */
6: Q_k /* A set of pairs (o_i, m_{i+1}), where $o_i \in s_i$ and m_{i+1} is the
             modified translated word for o_i */
7: **when** received (ask, s_1) **from** user **do**
8:     $T_1 \leftarrow \{(w_1, w_2, ..., w_n) \mid w_1 \in s_1\}$;
9:     **for each** MTA_i **in** *MTA* **do**
10:        send (request, (s_i, T_i)) to MTA_i;
11:        **when** received (response, (s_{i+1}, Q_i)) **do**;
12:           $T_{i+1} \leftarrow$ SELECT-POSSIBLE-N-TUPLES (T_i, Q_i);
13:        **end do**;
14:     **end loop**;
15:     send (reply, s_{n+1}) to user;
16: **end do**;

Algorithm 2: SELECT-POSSIBLE-N-TUPLES (T_i, Q_i) **return** T_{i+1}

1: $T_{i+1} \leftarrow \square$;
2: **for each** pair (o_i, m_{i+1}) **in** Q_i **do**
3:     $T_{i+1} \leftarrow T_{i+1} \bullet \{(w_1, w_2, ..., w_n) \mid (w_1, w_2, ..., w_n) \in T_i,$
               $w_i = o_i$ and $w_{i+1} = m_{i+1}\}$;
4: **end loop**;
5: **return** T_{i+1};

Figure 5. Algorithms of the coordinator agent *CA*

Algorithm 3: SERVICE-AGENT MTA_i

1: t_i /* Translated sentence */
2: $MT_i = \{(s_i, t_i)\}$ /* A translation service; a set of pairs of s_i and t_i */
3: c_{i+1} /* A word in sentence t_i */
4: P_i /* A set of pairs (o_i, c_{i+1}), where $o_i \in s_i$ and $c_{i+1} \in t_i$ */
5: **when** received (request, (s_i, T_i)) **from** *CA* **do**
6:     $t_i \leftarrow MT_i(s_i)$;
7:     $P_i \leftarrow$ GET-WORD-PAIRS-USED-BY-MT (s_i, t_i);
8:     $Q_i \leftarrow$ CREATE-WORD-PAIRS-TO-BE-USED (P_i, T_i);
9:     **if** $Q_i \bullet P_i$ **then**
10:        $s_{i+1} \leftarrow$ MODIFY-TRANSLATED-SENTENCE (t_i, P_i, Q_i);
11:     **else** $s_{i+1} \leftarrow t_i$;
12:     **end if** ;
13:     send (response, (s_{i+1}, Q_i)) to *CA*;
14: **end do**;

Algorithm 4: CREATE-WORD-PAIRS-TO-BE-USED (P_i, T_i)
               return Q_i

1: $Q_i \leftarrow \square$;
2: **for each** pair (o_i, c_{i+1}) **in** P_i **do**
3:     **for each** *n*-tuple $(w_1, w_2, ..., w_n)$ **in** T_i **do**
4:        **if** $o_i \in (w_1, w_2, ..., w_n)$ and $c_{i+1} \in (w_1, w_2, ..., w_n)$ **then**
5:           $Q_i \leftarrow Q_i \bullet \{(o_i, c_{i+1})\}$;
6:        **end if**;
7:     **end loop**;
8:     **if** $(o_i, c_{i+1}) \notin Q_i$ **then**
9:        $m_{i+1} \leftarrow i+1$th word in *n*-tuple selected from
               $\{(w_1, w_2, ..., w_n) \mid o_i \in (w_1, w_2, ..., w_n)\}$;
10:        $Q_i \leftarrow Q_i \bullet \{(o_i, m_{i+1})\}$;
11:     **end if**;
12: **end loop**;
13: **return** Q_i;

Figure 6. Algorithms of the translation agent *MTA*

This method can easily be extended to four or more languages by combining triples generated in each of the three languages similar to the extension approach proposed by Wu *et al.* [Wu *et al.*, 2008]. For example, for Japanese, English, German, and French words, Japanese-English-German triples are obtained first followed by English-German-French triples. The quadruple is generated by combining two triples with identical English and German words. It is noteworthy that a triangle does not always imply equivalent terms. In the case where word A has word sense C_1 and C_2, word B has C_2 and C_3, and word C has C_3 and C_1, no shared sense exists between the three words. Assume that each word in a triple has *n* senses with uniform distribution, the probability of sharing the same sense is .83 for $n = 2$ and 0.91 for $n = 3$; this probability approaches 1 as *n* increases. In practice, the term frequencies of *n* senses are unequal, and the actual probability is higher than the calculated one. Thus we can obtain reliable equivalent terms by combining triples if the number of languages increases.

In related research on dictionary formulation, a method to construct a bilingual dictionary using a third language as an intermediate is proposed [Tanaka and Umemura, 1994]. This study takes the example of generating a Japanese-French dictionary by connecting Japanese-English and English-French dictionaries. It addresses the problem that a French word with a meaning different from that of the original Japanese word is obtained due to ambiguity in the intermediate English word; this problem is solved through inverse consultation with French-English and English-Japanese dictionaries. We focus on obtaining more reliable equivalent terms when dictionaries exist between each pair of languages and differ from the above research in terms of our assumptions and objectives. In order to realize coordination even when sufficient dictionaries are not available, methods such as inverse consultation are required to obtain equivalent terms.

4 Context-based Coordination Algorithms

Algorithms of the multiagent architecture for the context-based coordination are shown in figure 5 and 6. These algorithms are simple implementations of our multiagent model. Let machine translator MT_i input source sentence s_i and output translated sentence t_i. Let the translation agent MTA_i receives source sentence s_i, generate and modify t_i, and output s_{i+1}, which is a source sentence of MTA_{i+1}. Let the coordinator agent *CA* repeat the coordination process from MTA_1 to MTA_n and receive s_{n+1} as the final result in the target language. Multilingual equivalent terms in *n* languages are grouped into *n*-tuples. The context T_i is a set of *n*-tuples and the *i*-th word in each *n*-tuple in T_i is included in s_i. In a *n*-tuple $(w_1, ..., w_n)$, the words $w_2, ..., w_n$ have the same meaning as w_1 i.e. the same meaning as original sentence s_1, and their use assures the correct translation.

First, *CA* prepares the initial context T_1 from s_1 received from the user and starts translation. After MTA_i returns the translated sentence s_{i+1} and Q_i—representing word pairs of the source word in s_i and translated word in s_{i+1}—*CA*

Coordinator agent *CA*

T₁ SELECT-POSSIBLE-N-TUPLES

s₂
Her fault is a big problem.

Q₁
ketten (fault)
□ fault,
mondai (problem)
□ problem

s₂

T₂
{{*ketten* (fault), fault, *Fehler* (fault)},
{*ketten* (fault), fault, *Mangel* (fault)},
{*mondai* (problem), problem, *Problem* (problem)}}

Japanese-English translation agent *MTA₁*

English-German translation service *MT₂*

English-German translation agent *MTA₂*

t₂
Ihre Schuld ist ein großes Problem. (Her responsibility is a big problem.)

CREATE-WORD-PAIRS-TO-BE- USED

GET-WORD-PAIRS -USED-BY-MT

P₂
fault□ *Schuld* (responsibility), problem□ *Problem* (problem)

MODIFY-TRANSLATED-SENTENCE

fault□ *Fehler* (fault), problem□ *Problem* (problem)

Q₂

s₃
Ihre Fehler ist ein großes Problem. (Her fault is a big problem.)
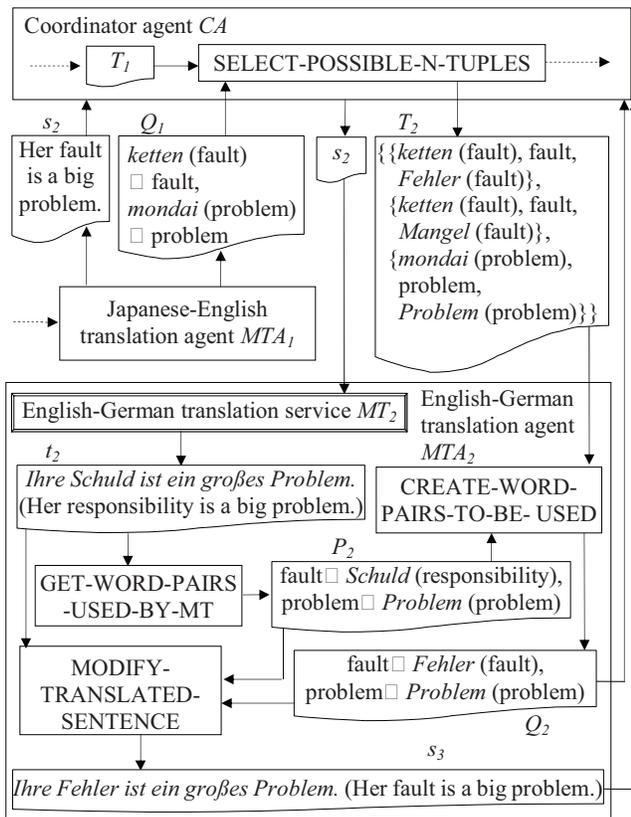
Figure 7. Example of Coordinated Translation Services

generates a new context T_{i+1} for the $i+1$-th translation by narrowing down T_i such that the $i+1$-th word in each *n*-tuple appears in s_{i+1} by the SELECT-POSSIBLE-N-TUPLES procedure. T_{i+1} may contain ambiguity in word selection for the $i+2$-th word, as more than two *n*-tuples containing the same j-th word ($1\Box j\Box i+1$) can exist with different $i+2$-th words. If there are several candidates for the $i+2$-th word, the $i+1$-th translation agent MTA_{i+1} determines the most appropriate one. The choice is noted to *CA* by Q_{i+1}, and *CA* reflects it to the next translation.

MTA_i generates a translated sentence t_i using MT_i to create P_i—a set of word pairs of source word o_i and translated word c_{i+1}—using the GET-WORD-PAIRS-USED-BY-MT procedure. One way to implement this function is to divide s_i and t_i into morphemes and map between them using bilingual dictionaries. Then, MTA_i modifies words in P_i based on the context T_i using the procedure CRE-ATE-WORD-PAIRS-TO-BE-USED and Q_i. Since T_i preserves the words used in the preceding i translations, the translated words excluded from T_i may have different meanings. Such words are replaced by words included in T_i, selected from among a few candidates if T_i contains ambiguity. Finally, t_i is modified by the procedure MOD-IFY-TRANSLATED-SENTENCE, wherein the words are replaced using P_i and Q_i. The word selection process can be improved through several methods: for instance, by referring

to term frequency or priority of words, in case the translation agent possesses this information. If the entire document or conversation logs are available, this information can be utilized by *CA* to create an initial context T_1.

Example 2 (Context-based translation)
We show the translation process for the sentence shown in figure 1(c). In this example, the replacement of target words is limited to nouns. Figure 7 shows the process of the English-German translation agent MTA_2 after the Japanese-English translation agent MTA_1 completes its translation process. In the first step, the coordinator agent *CA* receives the Japanese source sentence s_1 = "*kanojo no ketten ha ookina mondai da* (Her fault is a big problem)," sets all possible *n*-tuples including the words in s_1 and transfers s_1 and T_1 to MTA_1. MTA_1 then translates s_1 into the English sentence t_1 = "Her fault is a big problem" using the Japanese-English translation service MT_1. MTA_1 obtains pairs P_1 of words in s_1 and t_1: P_1 = {{*ketten* (fault), fault}, {*mondai* (problem), problem}}. MTA_1 then examines the translated words. For example, if T_1 contains triples including both *ketten* (fault) and "fault," MTA_1 realizes that they share the same meaning. If that is not the case, the triples may remain incomplete, and MTA_1 has to abandon efforts to maintain context. If the triples are complete, then triples including both *ketten* (fault) and "fault" as well as those including both *mondai* (problem) and "problem" should be contained in T_1. Therefore, translated words are not modified: $Q_1 = P_1$ and $s_2 = t_1$. MTA_1 then sends s_2 and Q_1 to *CA* and *CA* generates the new context T_2. For example, both triples of T_1 including both *ketten* (fault) and "fault" are to be included in T_2, as shown in figure 7.

In the second step, s_2 and T_2 are sent to the second English-German translation agent MTA_2. MTA_2 translates s_2 to the German sentence t_2 = "*Ihre Schuld ist ein großes Problem* (Her responsibility is a big problem)." Pairs P_2 are then obtained: P_2 = {{fault, *Schuld* (responsibility)}, {problem, *Problem* (problem)}}. It appears that the word *Schuld* (responsibility) has semantically drifted, as there is no triple in T_2 that includes both "fault" and *Schuld* (responsibility). Thus it is replaced by a word that is included in a triple in T_2, which also includes "fault." If the first triple in figure 7 is selected, Q_2 would be {{fault, *Fehler* (fault)}, {problem, *Problem* (problem)}}. MTA_2 modifies t_2 to s_3: s_3 = "*Ihre Fehler ist ein großes Problem* (Her fault is a big problem)." s_3 is finally returned to the user.

5 Evaluation

We constructed Japanese-English-German triples limiting their parts-of-speech to nouns. Table 1 lists the dictionaries used and the number of triples obtained from them. Transition-type triples start with Japanese words. A total number of 21,914 triples were obtained. We first analyzed the effectiveness of the 21,914 triples in covering arbitrary Japanese documents. We used the term frequency of nouns in a Web corpus storing 470 million sentences containing 5000 million Japanese words [Kawahara and Kurohashi, 2006]. The triples

Table 1: Dictionary and generated triples

(a) Bilingual dictionaries used to obtain triples

Dictionary	Number of headwords
Genius Japanese-English dictionary	31,944 (noun)
Concise Japanese-German dictionary	38,487(all words)
Oxford English-German dictionary	31,180 (noun)
Crown German-Japanese dictionary	34,255 (noun)

(b) Number of triples of each type

Type	Number of triples
Loop	15,627
Transition (starting from Japanese)	13,757
Total (no overlaps)	21,914

Source sentence (Japanese; A):
  torakku ga michi wo husaide ita.
  (A truck was blocking the road.)
B: *torakku ha houhou wo samatageta.*
  (A truck was blocking the method.)
C: *torakku ha michi wo samatageta.*
  (A truck was blocking the road.)

Figure 8. Example of an improvement from 4 (Most) to 5 (All)

appeared to cover 58% of all nouns in the corpus and 40% of all parts-of-speech words. If the triples are used in descending order of term frequency, 6,000 triples can cover 50% of nouns and 38% of all parts-of-speech words. This implies that a relatively small number of triples can cover the majority of frequently used nouns.

We then conducted a preliminary evaluation of the quality of Japanese-German back translation using the cascade of Japanese-English, English-German, German-English, and English-Japanese translations. We compared the source Japanese sentence (A), back-translated Japanese sentence generated without context (B), and that generated based on context (C). For purposes for accuracy, we took the subjective evaluation by three Japanese subjects who were native speakers of Japanese. The subjects were asked to evaluate the translation quality on a five-point scale, how much of the original meaning of sentence A was conveyed through sentences B and C (5-All, 4-Most, 3-Much, 2-Little, 1-None). Source sentences were selected from the Machine Translation Test Set provided by the NTT Natural Language Research Group[4]. We randomly selected 100 samples in which B and C were different. The results of Welch's test show that there is a difference in quality between B and C with a confidence level greater than 98%.

On average, the translation quality improved for 41 sentences and the score increased by an average of 0.47 points using context-based coordination. For example, in figure 8, without context the Japanese word *michi* (road) is mistranslated to *houhou* (method). This error occurs because the intermediate English word "way" has several meanings. The quality improved in the case of 34% for the sentences that were previously assigned a rating of 4 when translated

---

4 http://www.kecl.ntt.co.jp/mtg/resources/index.php

without coordination. Similarly, sentences with ratings of 3, 2, and 1 showed improvements for 32%, 49%, and 60% respectively with the context-based approach.

6 Conclusion

This study proposes a method for context-based coordination to overcome mistranslations during pivot translation, which occurs because of inconsistent word selection. The major aspects are summarized below.

Context-based Coordination with Propagated Context
  We took an approach to propagate context across combined translation services. Treating context as a set of multilingual equivalent terms used in translation, we propose to obtain all possible terms based on triangle forms formed by the relationships between words and translated words extracted from bilingual dictionaries. Our triangle method can be easily extended to four or more languages, and it is efficient in obtaining a sufficient amount of terms; the evaluation results show that the generated equivalent noun terms cover 58% of nouns and 40% of all parts-of-speech appearing in arbitrary sentences.

Multiagent Architecture for Coordination
  We proposed a multiagent architecture as one way to implement coordination with propagated context, wherein the coordinator agent gathers and propagates context from/to translation agents. Evaluation results of the translation quality of the indicated improvements in 41% of the total 100 sentences used and that the quality rating increased by an average of 0.47 points on a five-point scale. This architecture offers the flexibility of extension and the possibility of constructing a more complex composition of translation services and other types of language resources.

By considering the translation services as black boxes, a substantial improvement in translation quality was realized. The advantage of our approach is that we can improve the translation quality without any corpora, training of translation services with training sentences, or changing the inner components of systems; we only use available language resources and add some components outside existing translation services. This improvement is not trivial in the intercultural collaboration domain [Ishida *et al*., 2007]. Context-based coordination approach will play an important role in the quality improvement of the component service itself making up the composite service, which is frequently considered an issue of the component technologies.

References

[Bramantoro et al., 2008] Arif Bramantoro, Masahiro Tanaka, Yohei Murakami, Ulrich Schäfer and Toru Ishida. A Hybrid Integrated Architecture for Language Service Composition. *ICWS-08*, pages 345–352, 2008.

[Brennan and Clerk, 1996] Susan E. Brennan and Herbert H. Clark. Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(6):1482–1493, 1996.

[Hassine *et al*., 2006] Ahlem Ben Hassine, Shigeo Matsubara and Toru Ishida. A Constraint-Based Approach to Horizontal Web Service Composition. *ISWC-06*, pages 130–143, 2006.

[Ishida, 2006] Toru Ishida. Language Grid: An Infrastructure for Intercultural Collaboration. *SAINT-06*, pages 96–100, keynote address, 2006.

[Ishida *et al*., 2007] Toru Ishida, Susan R. Fussell and Piek Vossen. (Eds.): *Intercultural Collaboration*. Lecture Notes in Computer Science, 4568, Springer-Verlag, 2007.

[Kanayama and Watanabe, 2003] Hiroshi Kanayama and Hideo Watanabe. Multilingual Translation via Annotated Hub Language. *MT-Summit IX*, pages 202–207, 2003.

[Kawahara and Kurohashi, 2006] Daisuke Kawahara and Sadao Kurohashi. Case Frame Compilation from the Web using High-Performance Computing. *LREC-06*, 2006.

[Matsumura et al., 2006] Ikuo Matsumura, Toru Ishida, Yohei Murakami and Yoshiyuki Fujishiro. Situated Web Service: Context-Aware Approach to High Speed Web Service Communication. *ICWS-06*, pages 673–680, 2006.

[Tanaka and Umemura, 1994] Kumiko Tanaka and Kyoji Umemura. Construction of a Bilingual Dictionary Inter-mediated by a Third Language. *COLING-94*, pages 293–303, 1994.

[Tanaka et al., 2009] Masahiro Tanaka, Toru Ishida, Yohei Murakami, and Satoshi Morimoto. Service Supervision: Coordinating Web Services in Open Environment. *ICWS-09*, to be published, 2009.

[Tokunaga and Tanaka, 1990] Takenobu Tokunaga and Hozumi Tanaka. The Automatic Extraction of Conceptual Items from Bilingual Dictionaries. *PRICAI-90*, pages 304–309, 1990.

[Utiyama and Isahara, 2007] Masao Utiyama and Hitoshi Isahara. A Comparison of Pivot Methods for Phrase-based Statistical Machine Translation. *HLT-NAACL*, pages 484–491, 2007

[Vossen, 1998] Piek Vossen. (Eds.) EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Dordrecht, Netherlands: Kluwer, 1998. See: http://www.hum.uva.nl/ ewn/.

[Wu and Wang, 2007] Hua Wu and Haifeng Wang. Pivot Language Approach for Phrase-Based Statistical Machine Translation. *ACL'07*, pages 856–863, 2007.

[Wu *et al*., 2008] Yanchen Wu, Fang Li, Rie Tanaka and Toru Ishida. Automatic Creation of N-lingual Synonymous Word Sets. *SKG-08*, pages 141–148, 2008.

[Yamashita *et al*., 2009] Naomi Yamashita, Rieko Inaba, Hideaki Kuzuoka and Toru Ishida. Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation. *CHI'09*, pages 679–688, 2009.

[Yamashita and Ishida, 2006] Naomi Yamashita and Toru Ishida. Effects of Machine Translation on Collaborative Work. *CSCW-06*, pages 515–523, 2006.