Title	A Cooperative Work Environment for Translation : Integrating MT and TM for Community
Author(s)	Supnithi, Thepchai; Trakultaweekoon, Kanokorn; Klaithin, Supon
Citation	
Issue Date	2007-11
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/4092
Rights	
Description	The original publication is available at JAIST Press http://www.jaist.ac.jp/library/jaist-press/index.html, KICSS 2007 : The Second International Conference on Knowledge, Information and Creativity Support Systems : PROCEEDINGS OF THE CONFERENCE, November 5-7, 2007, [Ishikawa High-Tech Conference Center, Nomi, Ishikawa, JAPAN]

A Cooperative Work Environment for Translation
~ Integrating MT and TM for Community~

Thepchai supnithi† **Kanokorn trakultaweekoon†** **Supon Klaithin†**

†Human Language Technology Laboratory
National Electronics and Computer Technology Center
112 Paholyothin Road, Klong Neung
Klong Luang Pathumthani, Thailand 12120
{Thepchai.supnithi, Kanokorn.trakulthaweekoon, Supon.klaithin}@nectec.or.th

Abstract

Translation becomes an important issue since it enables us to bridge the gap of language divide problem. Because of the unsatisfied accuracy, the automatic machine translation cannot meet users' requirement. We propose a cooperative work environment which is an integration of machine translation and translation memory. It assists users to keep the previous translation results, at the same time, to get a reference from other user's translation result.

Keywords: machine translation, translation memory, cooperative work environment

1 Introduction

Translation is an important task which enables us bridging the gap of language divide problem. There are two main approaches; automatic machine translation [2,3,4] and semi-automatic machine translation or translation memory [1]. Automatic machine translation is a powerful method for rapid translating, but it cannot fulfill user's desire because of the unsatisfied accuracy. Translation memory is a semi-automatic approach which enables translator keep his/her previous translated result. This technique is not only helping translators to reduce the translation time, but also assisting them to translate with higher consistency [5]. However, when we apply this technique into general users, they sometimes cannot design how to correct the results. Cooperative translation work environment has been designed to assist users to share and receive translation idea among users in a community.

In this paper, we develop a cooperative work environment for translation task by integrating MT and TM, and a cooperative tool for sharing and receiving translation idea from other users. The structure of this paper is shown as follows; Section 2 explains a rough idea of machine translation. Section 3 illustrates our cooperative work environment. Section 4 shows an implementation of our cooperative translation work. Finally, section 5 gives a discussion and conclusion.

2 An overview of cooperative work environment for translation

Translation module in cooperative work environment composes of two main components; Parsit, translation memory.

2.1 Parsit

Parsit [7] is English to Thai rule-based machine translation system (MT system). It composes of five main modules: English morphological analysis, English syntactic analysis, English semantics analysis, Thai semantics generation, and Thai syntactic generation. When Parsit receives a sentence, it will be sent to morphological module to stem and retrieve lexicon information. Next, English syntactic analysis, together with semantic analysis will be processed in order to construct an Interlingua representation. Word ambiguity will be resolved in semantic level. Thai semantics generation and syntactic analysis will finally be processed to generate an appropriate Thai sentence. Ordering will be treated in the syntactic generation level. The result of Parsit evaluation by applying BLEU score method is 0.0260 [6]. Parsit is now avail-

able at http://www.suparsit.com with more than 1,500 IP users per day.

2.2 Translation Memory

Automatic machine translation cannot fulfill the needs of users because of the unsatisfied accuracy. It is necessary to provide a tool for users to edit their own results. Translation memory (TM) is an environment for assisting translation by allowing the reuse of previously translated phrases and terms. We combine translation memory to Parsit which enables each user to keep his selected translation. Normally, translation memory database have to be prepared in order to collect sentence pair and update time for tracing the time sequence. Since our system has been designed for supporting cooperative work, we collect additional information, such as sentence owner, sentence original user.

2.3 Process in cooperative work environment

Figure 1 shows two main processes in this system, translation process and translation editing process.

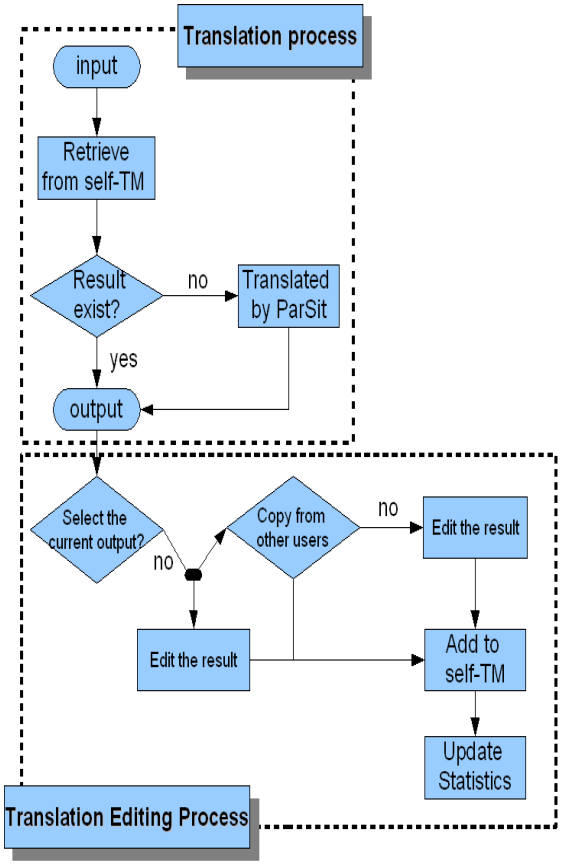

Figure 1. Process on cooperative translation work environment

Translation process: When a user starts sending any sentences, it will be checked whether the sentence is previously translated or not by referring to individual TM database. If the sentence exists in TM database, the translation results will be directly retrieved as an output. Otherwise, Parsit will be applied for automatic translation and return it as an output.

Translation Editing Process: After a user receives the translation result, it is possible to customize based on his/her own desire. If the user is satisfied with the result, the process will be stopped. If a user does not satisfied with the results, it is possible to edit the translation result. However, we cannot guarantee the automatic result because of the uncertainty of input. There are some cases that difficult to get a good reference from machine translation. Our system provides a collaborating environment for applying other translation results within a community. Based on the collaborative tool, it is possible for user to edit the results from others by using a good reference.

3 Implementation

In this section, we explain the implementation of our translation environment. PHP language is used to develop environment and MySQL is used for database manipulation.

3.1 System development

Figure 2 shows an input environment for any users. When a user input any texts and click "translate" button, the input text in English and translated output text in Thai will be displayed as shown in Figure 3. If there are more than one

Figure 2. Input text area

ข้อความภาษาอังกฤษ	ข้อความภาษาไทย
To translate a web page, enter or paste its URL. Then, press Send URL. To change the URL, press Reset.	เพื่อแปลเว็บเพจเข้าหรือวางURLนั้น. ในเวลานั้นกดปุ่มURLส่ง. เพื่อเปลี่ยนURLกดปุ่มลบข้อความ. บันทึกลงไฟล์

Figure 3. Translation results

ประโยคต้นฉบับ	ประโยคผลลัพธ์	
To translate a web page, enter or paste its URL.	เพื่อแปลเว็บเพจเข้าหรือวางURLนั้น.	กดที่นี่เพื่อแก้ไขประโยค (ช่วยกันแก้ไข)
ประโยคต้นฉบับ	ประโยคผลลัพธ์	
Then, press Send URL.	ในเวลานั้นกดปุ่มURLส่ง.	กดที่นี่เพื่อแก้ไขประโยค (ช่วยกันแก้ไข)
ประโยคต้นฉบับ	ประโยคผลลัพธ์	
To change the URL, press Reset.	เพื่อเปลี่ยนURLกดปุ่มลบข้อความ.	กดที่นี่เพื่อแก้ไขประโยค (ช่วยกันแก้ไข)

Figure 4. Translation Results in sentence by sentence

คำแปลของคุณ	เพื่อเปลี่ยนURLกดปุ่มลบข้อความ.		เลือกใช้งาน บันทึก
คำแปลจากภาษิต	แก้ไข	เพื่อเปลี่ยนURLกดปุ่มลบข้อความ.	เลือกใช้งาน
คำแปลจาก thep	แก้ไข	เพื่อเปลี่ยนURLกดปุ่มลบข้อความ. (แจ้งลบข้อความนี้)	เลือกใช้งาน
คำแปลจาก thep	แก้ไข	เพื่อเปลี่ยนURL, กดปุ่ม Reset. (แจ้งลบข้อความนี้)	เลือกใช้งาน
คำแปลจาก woraphat	แก้ไข	เพื่อเปลี่ยนURL, กดปุ่ม Reset. (แจ้งลบข้อความนี้)	เลือกใช้งาน

Figure 5. Cooperative translation work environment

ข้อความภาษาอังกฤษ	ข้อความภาษาไทย
To translate a web page, enter or paste its URL. Then, press Send URL. To change the URL, press Reset.	เพื่อแปลเว็บเพจเข้าหรือวางURLนั้น. จากนั้นกดปุ่ม ส่ง URL. เพื่อเปลี่ยนURL, กดปุ่ม Reset. บันทึกลงไฟล์

Figure 6. Translation results when applying translation memory

sentence is translated, the input and translated results will be displayed in sentence by sentence as shown in figure 4. At the right most of any sentences in figure4, it displays the status of collaborating. A word "*start editing*" will be displayed, if there is no record of translation from other users. A word "*collaborative editing*" will be displayed, when some users editing this sentence. In the former case, when a user clicks "*start editing*" button, an editing panel will be displayed. In the latter case, an editing panel, together with translated results from other users will be displayed (figure 5). User can select any sentences to get the translated results or to edit the translated results.

The rough idea of using environment is explained as an example in figure 2- figure 5 as follows.

Source sentence in English:
"(1) *To translate a web page, enter or paste its URL.*
(2) *Then, press Send URL.*
(3) *To change the URL, press Reset.* "

Translated result from Parsit:
"(1)*เพื่อแปลเว็บเพจเข้าหรือวาง URL นั้น.*
(2) *ในเวลานั้นกดปุ่ม URL ส่ง.*
(3) *เพื่อเปลี่ยน URL กดปุ่มลบข้อความ.*"

The translated result from Parsit is understandable but it is not grammatically correct. This sentence can be separated into three sentences as displayed in English and Thai. Each sentence has its own cooperative results. User can select other users' result for better translation.

In the third sentence: "*To change the URL, press Reset.*", there are four alternatives (in figure 5). There are two users apply our framework. A user named "woraphat" changed the automatic translation output from "*เพื่อเปลี่ยน URL กดปุ่มลบข้อความ.*" to "*เพื่อเปลี่ยน URL, กดปุ่ม Reset.* " and apply it in his translation.

In parallel, assume that the user changed the second sentence: "*Then, press Send URL*" from "*ในเวลานั้นกดปุ่ม URL ส่ง*" to "*จากนั้นกดปุ่ม ส่ง URL.* ", the system will automatically collect the changed sentence into translation memory. When the user translates the same source sentence again, the translation results will be shown in figure6. As a result, the system applies translation results from MT at the first sentence, and use translation results from TM at the second and third sentences.

Translated result from Parsit+TM:
"(1)*เพื่อแปลเว็บเพจเข้าหรือวาง URL นั้น.*
(2) *จากนั้นกดปุ่ม ส่ง URL.*
(3) *เพื่อเปลี่ยน URL, กดปุ่ม Reset.* "

3.2 Database Design

In order to develop our translation cooperative environment, we design database for translation memory as follows. An example of data is illustrated in Table1.

Sentence ID: An identification number for each sentence
User ID: An identification number for user who own this sentence
English sentence: A source sentence in English

Table 1. An example of data in translation memory database

Feature	Example
Sentence ID	328
User ID	2776
English Sentence	To change the URL, press Reset.
Thai Sentence	เพื่อเปลี่ยนURL, กดปุ่ม Reset.
Freq_sentence	5
Source_ID	361558
Update_Date	2007-05-27
Update_Time	17.08

Thai sentence: A target sentence in Thai which is defined manually
Freq_sentence: A frequency of using this sentence
Source_ID: An identification number of the user who construct source of this sentence
Update_date: Date for sentence registration
Update_time: Time for sentence registration

3.3 Database manipulation

In translation editing process there are two main functions for manipulating data.

Translation memory management: This function supports updating edited sentence. It will add the translation pair into database iff

1) Translated results are copied from other users
2) Translated results are edited. It might be edited from his/her own results or from other user's results.

Comparing to the result from Parsit, it will not be added in the database, if there is no change between Parsit and user edited results. Criteria are assigned to avoid careless mistakes in using our cooperative work, such as clicking mouse errors.

Statistics update function: This function is designed to update statistics. It enables us understand behavior of using the cooperative tool. When any translation pairs are added into database, the system will update the following information

1) Frequency used of sentence
2) Sentence owner
3) Source sentence owner

4 Discussion

We developed and launched a web based environment in http://www.suparsit.com There are 23,721 sentence pairs from users in the environment during May 2006- August 2007. The statistics of frequency used in each sentence and user's participation are shown in table2 and table3.

Table 2. Sentence usage statistics

Sentence_used	frequency
>5	17
5	5
4	7
3	40
2	311
1	23341

Table 3. User behavior statistics

User behavior	frequency
>3	1159
3	488
2	1346
1	5658

We request users to send a parallel sentence to the system. 3,419 sentence pairs are sent to the system during Oct 2004-August 2007. Since users get a benefit on keeping previous translation results, the collection rate between requesting translation pairs and providing the cooperative translation work is about 6:1.

5 Conclusion

In this paper, we proposed a cooperative environment work which enables users share idea of translation in communities. Translation pairs are collect with more accelerated than requesting the translation pairs from users. In the future work, we plan to continue the usage on cooperative work in the community. Furthermore, we plan to extend our work by applying the system to classroom, we aims to assist students to learn by observing and diagnosing.

References

[1] Christian Boitet. *Machine-aided Human Translation*. in COLE, R.A.- MARIANI, J.- USZKOREIT, H.- ZAENEN, A.- ZUE, V. (Eds.) *Survey of the State of the Art in Human Language Technology*. Cambridge: Cambridge University Press. (1997)

[2] Doug Arnold and Louis Des Tombe. *Basic Theory and Methodology in Eurotra*. In S. Nirenburg, editor, *Machine Translation: Theoretical and Methodological Issues*, pages 114-135. Cambridge University Press, Cambridge, England, (1987)

[3]Franch Josef Och. Statistical Machine Translation: Foundations and Recent Advances, MT Summit X Tutorial, Phuket Thailand. (2005)

[4] Makoto Nagao. Framework of a Mechanical

Translation between Japanese and English by Analogy Principle, in: Artificial and Human Intelligence, thorn, A., and Banerji, R.(eds.), North Holland, Amsterdam,. pp173-180 (1984)

[5] Lynn E. Webb. *Advantages and Disadvantages of Translation Memory: A cost/Benefit Analysis.* Submitted in partial satisfaction of the requirements for the Degree of MASTER OF ARTS (1992)

[6]Nattapol Kritsuthikul and Thepchai Supnithi "English-Thai Example-Based Machine Translation using n-gram model", IEEE-SMC 2006 (2006)

[7] Virach Sornlertlamvanich, Paisarn Charoenpornsawat, Mothika Boriboon and Lalida Boonmana. *ParSit: English-Thai Machine Translation Services on Internet.* 12th Annual Conference, ECTI and New Economy, National Electronics and Computer Technology Center, Bangkok, June (2000). (in Thai)