

# Multilingual Medical Text Development

**Thatsanee Charoenporn, Virach Sornlertlamvanich, Kergrit Robkop,  
Prapass Srichaiwattana<sup>†</sup>, and Hitoshi Isahara<sup>††</sup>**

<sup>†</sup>Thai Computational Linguistics Laboratory, NICT Asia Research Center, Thailand

<sup>††</sup>National Institute for Communications Technology, Japan

{thatsanee, virach, kergrit, prapass}@tcllab.org and isahara@nict.go.jp

## Abstract

This paper describes the collaborative multilingual corpus development. We introduce KUI (Knowledge Unifying Initiator) to be a corpus tool development to facilitate the task of medical text translation of participants of different mother language. KUI provides a chat function and records the communication among participants for the study of the nature of their online communication. The medical text translation is selected as the task for studying the intercultural communication. It contains a list of short questions produced by a doctor in diagnosing a patient. As a result, the translated multilingual medical text is developed together with a chat log of their communication. The chat log is then analyzed to understand the segments of their communication.

## 1 Introduction

Communities cannot deny the importance of the Internet as the bridge of the world and cannot ignore the crucial role of English language in the cyber society either. There are many attempts in machine translation development among languages to/from English with the aim of paving the way for conquering the language barrier and also exterminating the digital divide. But in fact, language barrier problems do not occur only in the Internet society. Language diversity in each community causes the problem in real life as well. Imagine about more than 250 languages spoken in the US, or about 3,764 languages spoken in Africa and Asia. Even though the language diversity can make life exciting, people sometimes need help in communicate across languages. When language diversity is raised into discussion, cultural diversity is always a part. Since the language is a dimension of the culture, one cannot deny establishing the under-

standing and sharing between different cultures when attempting to bridge the gap. According to this phenomenon, we attempt to create a system for multilingual text development as a method to facilitate the empowerment of language communities, although we realize the essential of adopting and implementing measures enhancing equitable multilingualism.

The paper is organized in the following way. Section 2 explains the experiment of intercultural collaboration, namely ICE. Section 3 describes the nature of intercultural and multilingual communication in the scene of doctor - patient dialogue. Section 4 describes KUI as a tool for online collaborative translation. Section 5 explains the experiment and Section 6 analyses the result of the experiment.

## 2 Intercultural Collaboration Experiment

As aware in the current language barrier problem and the accomplishment in cross language communication, many research institutes and organizations spend a lot of efforts to overcome the situation. A joint research project of universities, research institutes, and research societies in Asia, named Intercultural Collaboration Experiment (ICE) is one among many (Nomura et al., 2003).

One objective of ICE is to analyze the interactive translation refinement procedures implemented between humans, and between humans and machines, under the truth that the advancement of technologies increases availability of intercultural collaboration beyond the language barrier. ICE also supports intercultural and multilingual collaborations using machine translation technologies. In short, ICE studies the computer-mediated cross language communications. By this project, many experiments for multilingual inter-communication have been conducted.

One of the experiments is “the experiment for cross-cultural study of expressive avatars survey procedure”, which aims to study whether character representations and facial expressions are

interpreted equally across different cultures (ICE, 2006).

We conducted an experiment on adopting KUI to facilitate the communication between participants from different languages and cultures. It provides chat, chat log and a topic for collaborative work. We selected the task of translating medical text as a task for the intercultural collaboration experiment.

### **3 Intercultural and Multilingual Communication**

As the world was attacked by the recent Tsunami, it was found that during the disaster rescue, relief and recovery time, the communication was often made substantially difficult because of language barriers. Or even for the general medical care, doctors and patients are unable to communicate fluently in their mother language. Trained interpreters are expensive and hardly ready for the emergency cases. Moreover, a study shows that errors of interpretation are often responsible for errors in diagnosis (Flores et al., 2003). To turn crisis, we cannot delay to conduct a system to facilitate cross language communication. We start with creating a knowledge base for helping doctors in diagnosing patients by patients' mother language. In the experiment, limited standard examination questions with only a limited number of one or two responses or gestures are selected to be the source phrases.

### **4 Online Collaborative Text Translation**

We develop KUI (Knowledge Unifying Initiator) as a knowledge development tool. We implement KUI to be a Knowledge User Interface for this collaborative translation task. Actually, KUI is a platform to unify the various thoughts following the process of thinking, i.e. initiating the topic of interest, collecting the opinions to the selected topics, localizing the opinions through the translation or customization and finally posting for public hearing to conceptualize the knowledge. The process of thinking is done under the selectional preference simulated by voting mechanism in the case that many alternatives occur.

In this section, we will describe the concept behind KUI, the knowledge development process, various features in KUI, and KUI for collaborative text translation.

#### **4.1 What is KUI?**

KUI is a GUI for knowledge engineering, in other words Knowledge User Interface (KUI) (KUI, 2006). It provides a web interface accessible for pre-registered members. An online registration is offered to manage an account by profiling the login participant in making contribution. A contributor can comfortably move around in the virtual space from desk to desk to participate in a particular task. A working desk can be a meeting place for collaborative work that needs discussion through the 'Chat', or allow a contributor to work individually by using the message slot to record each own comment. The working space can be expanded by closing the unnecessary frames so the contributor can concentrate on the task. All working topics can be statistically viewed through the provided tabs. These tabs help contributors to understand KUI in the aspects of the current status of contribution and the tasks. A knowledge community can be formed and can efficiently create the domain knowledge through the features provided by KUI. These KUI features fulfill the process of human thought to record the knowledge. KUI also provides a 'KUI look up' function for viewing the composed knowledge. It is equipped with a powerful search and statistical browse in many aspects. Moreover, the 'Chat log' is provided to learn about the intention of the knowledge composers. We frequently want to know about the background of the solution for better understanding or to remind us about the decision, but we cannot find one. To avoid the repetition of a mistake, we systematically provide the 'Chat log' to keep the trace of discussion or the comments to show the intention of knowledge composers.

#### **4.2 Knowledge Development Process in KUI**

Adopting the concept of Open Source software development, we will be possibly able to develop a framework for domain specific knowledge development under the open community environment. Sharing and collaboration are the considerable features of the framework. The knowledge will be finally shared among the communities by receiving the consensus from the participants in each step. To facilitate the knowledge development, the process is deliberated into 4 steps (Sornlertlamvanich, 2006).

- Topic of interest

The topic will be posted to draw the intention from the participants. The selected topics will then be further discussed in the appropriate step.

- Opinion

The selected topic is posted to call for opinions from the participants in this step. Opinion poll is conducted to get the population of each opinion. The result of the opinion poll provides the variety of opinions that reflects the current thought of the communities together with the consensus to the opinions.

- Localization

Translation is the straightforward implementation of the localization. Collaborative translation helps producing the knowledge in multiple languages in the most efficient way.

- Public-Hearing

The result of discussion will be revised and confirmed by gathering the opinions to the final draft of proposal

The developed knowledge is started from posting 'Topic of Interest', participants express their supports by casting a vote. Upon a threshold the 'Topic of Interest' is selected for conducting a poll on 'Opinion', or introducing to the community by 'Localization', or posting a draft for 'Public-Hearing' to gather feedbacks from the community. The transition from 'Opinion' to either 'Localization' or 'Public-Hearing' occurs when the 'Opinion' has a concrete view for implementation. The discussion in 'Localization' and 'Public-Hearing' is however interchangeable due to purpose of implementation whether to adopt the knowledge to the local community or to get feedbacks from the community.

The knowledge creating is managed in 4 different categories corresponding to the stage of knowledge. Each individual in the community casts a vote to rank the appropriateness of solutions at each category. The community can then form the community knowledge under the 'Selectional Preference' background.

### 4.3 Features in KUI

#### **Poll-based Opinion or Public-Hearing**

A contributor may choose to work individually by posting an opinion e.g. localization, suggestion etc., or join a discussion desk to conduct 'Public-Hearing' with others on the selected topic. The discussion can be conducted via the provided 'Chat' frame before concluding an opinion. Any opinions or suggestions are committed to

voting. Opinions can be different but majority votes will cast the belief of the community. These features naturally realize the online collaborative works to create the knowledge.

#### **Individual or Group works**

Thought may be formed individually or through a concentrated discussion. KUI facilitates a window for submitting an opinion and another window for submitting a chat message. Each suggestion can be cast through the 'Opinion' window marked with a degree of its confidence. By working individually, comments to a suggestion can be posted to mark its background to make it better understanding. On the other hand, when working as a group, discussions among the group participants will be recorded. The discussion can be resumed at any points to avoid the iterating words.

#### **Record of Intention**

The intention of each opinion can be reminded by the recorded comments or the trace of discussions. Frequently, we have to discuss again and again on the result that we have already agreed. Misinterpretation of the previous decision is also frequently faced when we do not record the background of decision. Record of intention is therefore necessary in the process of knowledge creation. The knowledge interpretation also refers to the record of intention to obtain a better understanding.

#### **Selectional Preference**

Opinions can be differed from person to person depending on the aspects of the problem. It is not always necessary to say what is right and what is wrong. Each opinion should be treated as a result of intelligent activity. However, the majority accepted opinions are preferred at the moment. Experiences could tell the preference via vote casting. The dynamically vote ranking will tell the selectional preference of the community at the moment.

### 4.4 KUI for Collaborative Text Translation

In this collaborative text translation, participants of different mother language work online as a virtual group by using KUI. After registering the system, KUI automatically provides a group of discussion for each task. The group consists of participant from different languages. Multi groups operate in parallel. Before translating, they are encouraged to discuss by the provided

chat function about the topic in question, system, personal information and so on.

## 5 Experiment

As the initiative purpose of developing multi-lingual medical text, we set an experiment for online translation process. The source phrases used in the experiment are the English emergency diagnosis phrases from MedSLT, an Open Source project for developing a medical speech translation system (MedSLT, 2006). Translation process is done online via KUI. The volunteer translators join the discussion group to translate the source text (English) in to their mother language sentence by sentence. Each group participated by translators from more than one language. During the translation process, they all are encouraged by the group communication to build their own community.

As a result, we obtain both the translated medical text and “chat log” which is considered to be the background intention of translation. This chat log will be a resource for further analysis on cross language communication.

### 5.1 Medical Text

As mentioned in the previous section, source text for the translation is the English emergency diagnosis phrases from MedSLT which doctors suspect that patients may be suffering from. These phrases are all range of utterances of standard examination question about chest pain and factors that increase or decrease such pain, which can be accomplishedly communicated by one or two word responses or gestures (Starlander, et al., 2005). The patterns of the question utterances are grammatically enough to ask about most domain concept in a natural way.

In term of content, questions are usually limited to be of the basis form as followings.

- (Do you, have you)
- (How long, how)
- (usually, often, ever, ...)
- (experience symptom, do something)
- (how long, when)
- or
- (does something cause symptom)

For example:

Type of utterance	Samples
Sentence	have you had pain for weeks how long do your headaches usually last do you ever have chest pain in the morning do your chest pains appear at night

	is the pain gradual
Phrase	muscle aches after meals high blood pressure heat cheese

Currently, there are 915 utterances in the experiment, including noun phrases, verb phrases and simple sentences.

### 5.2 Medical Text

The volunteer participants will be translated the English medical text into their mother languages, by using KUI. They act as a virtual group and participate in the translation via this web interface. With different backgrounds and degrees of translation abilities, they, therefore, are encouraged to chat, discuss or exchange their opinion while translating each utterance. Not only to better know each other, but the discussion is also aimed to deepen the understanding of utterance before translation. Figure 1 shows the participation work flow.

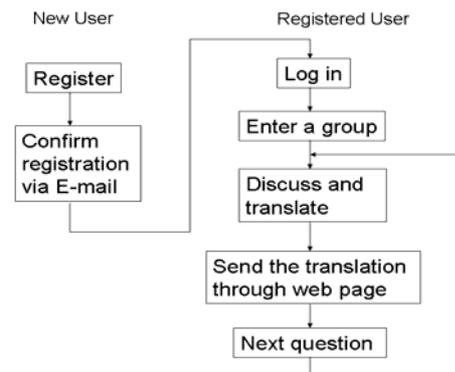


Figure 1: Participation Work Flow of Collaborative Translation Task

There are 3 tracks in the experiment. Each consists of participants from two languages as shown in Table 1. Each group was assigned automatically to translate different questions into their language.

Table 1: Participants Joining in the Collaborative Translation Task Phase 1.

Countries	Languages	No. of Group	No. of Participants
Thailand - China	Thai - Chinese	2	4
Thailand - Japan	Thai - Japanese	5	10

Thailand - Korea	Thai - Korean	3	6
Total		10	20

And after two days of experiment, 539 utterances were translated from these 20 participants.

## 6 Chat Log Analysis

As we believe that human has a smart way to communicate cross language, it is one of our purposes to find out how they communicate in the specific situation.

From the chat log, we can divide topics that the participants talk during the experiment into two groups. One is the topics related to the experiment and another is related to personal data.

### Topics related to the experiment

The topics related to the experiment can be classified into 4 sub-topics as followings.

- Requiring more information about the system, KUI

There are some questions asking about how to work with KUI, its conditions, the components and features of KUI. For example:

<b>English text:</b> Does it hurt in the left chest?
<b>Thai1:</b> what does desk 1, desk 4 mean?
<b>Thai2:</b> It means you do different topics
<b>Thai1:</b> but I still in medicine na
<b>Thai2:</b> yeah, medicine has a lot of topics
<b>Thai1:</b> but i still in the medicine na
<b>Thai1:</b> all right
<b>Thai2:</b> if you stay in the same group so you do in same item in that topic
<b>Thai2:</b> sorry same 'Desk' not same group
<b>Thai1:</b> can i use Thai here?
<b>Thai2:</b> of course, yes
<b>Thai1:</b> why I don't see my message?
<b>Thai2:</b> every things in here show in Unicode, you can type every languages
<b>Thai2:</b> that because it take time to take you messages from server

- Discussion on the question for better understanding and word meaning

The participants are all encouraged to discuss about the identified question for the clear and correct understanding before translating. For example:

<b>English text:</b> can exertion sometimes cause chest pains
<b>Japanese1:</b> ok

<b>Japanese2:</b> if you have same words like 'sometimes' do you use the translation you did before?
<b>Japanese1:</b> i am trying to mention the possibility
<b>English text:</b> can exertion sometimes give you chest pains
<b>Japanese1:</b> because sometimes means there are occasion that symptom doesn't occur
<b>Japanese2:</b> that's right.. and also I am not sure the frequency of 'sometimes'
<b>English text:</b> can your headache be caused by exertion
<b>Japanese2:</b> for example two or three times par day... or something like that.. but 2 or 3 par day is a lot ! :)
<b>English text:</b> can exertion cause abdominal pains
<b>Japanese1:</b> even in the Japanese language difinition of 'sometimes' in the Japanese language difinition of 'sometimes' = tokidoki difficult
<b>Japanese1:</b> excuse me. my last sentense was difficult to read
<b>Japanese2:</b> and also we should take care not to translate into particular symptoms..
<b>Japanese2:</b> abdominal pain is general I think
<b>Japanese1:</b> yes I think it's correct
<b>Thai:</b> maybe
<b>Japanese1:</b> it doesn't immediately mean 'cancer' or some particular deseases..

- Problems found during the translation process both related to the system and linguistic knowledge or the question itself. For example:

<b>English text:</b> Does it hurt in the lower abdomen?
<b>Thai:</b> it's hard to use the tools. cursor jumps all the time :(
<b>Thai:</b> yes
<b>Thai:</b> when i write chat box, it often jumps to translation box.
<b>Thai:</b> i have to wait until start icon acivates, right?
<b>Thai:</b> Anybody there? can u read me?
<b>Chinese:</b> yes ,i ok now ,go ahead
<b>Chinese:</b> yes ,i ok now ,go ahead

- The feeling during the translation process

During the translation, some participants express their feeling concerning on the question such as how difficult or how long of the question. For example:

<b>English text:</b> a stabbing pain
<b>Thai1:</b> quite difficut...
<b>Thai1:</b> Are U OK Ou?
<b>English text:</b> after meals
<b>Thai2:</b> Do you think it' s difficult?
<b>Chinese:</b> has login

### Topic related to the personal data

This kind of topic includes all general conversation since the participants are introduced to the group which is not concerning about the experiment. They include greeting phrases and persuading utterance to join the experiment next time. Background of the participants: nationality, country, address, age and so on. For example:

<b>English text:</b> is there family history of heart disease
<b>Thai1:</b> hello guys!
<b>Japanese2:</b> hello!
<b>Thai1:</b> you are so ontime!
<b>Thai1:</b> it's just 3.00 p.m. in Thai
<b>Japanese1:</b> hello!
<b>Japanese2:</b> yes, here in France, it's 10 in the morning
<b>Thai1:</b> oh i c
<b>Thai2:</b> how are u?
<b>Japanese2:</b> so we are all three or more?
<b>Thai1:</b> did your guys type anything?
<b>Japanese2:</b> type? where? here? or translation box?
<b>Thai1:</b> i see nothing in the translation tab
<b>Japanese2:</b> ah, translated already?
<b>Thai1:</b> yes, i just done it
<b>Thai1:</b> and yes for the upper question, we have 3 person in our group
<b>Japanese2:</b> hello, yes, I came here to know about the situation

<b>English text:</b> is your headache caused by bright light
<b>Thai:</b> have you ever been out of your country?
<b>Japanese:</b> what do you like?
<b>Thai:</b> i like movies, travelling, reading, swimming, hanging out with friends
<b>Japanese:</b> I never go to out of japan

<b>Thai:</b> i also like the beaches
<b>Japanese:</b> it is nice
<b>Thai:</b> yeap here has many nice beaches

## 7 Conclusion

We proposed an efficient online collaborative work in producing and maintaining the multilingual corpus.

KUI was efficiently introduced to encourage the communication among participants from different language background by providing a task of translating a list of medical text. We collected and analysed their communication resulting in a set of common segment of the online communication. This will lead to an efficient retrieval system of the response to either the request of knowledge about the system or the topic in question. KUI was also proved to support the collaborative work in producing the multilingual medical text. The translated text will be voluntarily maintained by the online participants under the selectional preference based on the voting function.

## References

- G. Flores, B. Laws, S. Mayo, B. Zuckerman, M. Abreu, L. Medina, and EJ. Hardt. 2003. *Errors in medical interpretation and their potential clinical consequences in pediatric encounters*. Pediatrics (111): 6-14.
- ICE. <http://character.kuis.kyoto-u.ac.jp/ICE/>. As of 9 March 2006.
- KUI. <http://www.tcllab.org/KUI/>. As of 9 March 2006.
- Marianne Starlander, Pierrette Bouillon, Manny Rayner, Nikos Chatzichrisafis, Beth A. Hockey, Hitoshi Isahara, Kyoko Kanzaki, Yukie Nakao, and Marianne Santaholma. 2005. *Breaking the Language Barrier: Machine Assisted Diagnosis using the Medical Speech Translator*. In: Proceedings of the XIX International Congress of the European Federation for Medical Informatics MIE, Geneva, Switzerland.
- MedSLT. <http://sourceforge.net/projects/medslt/>. As of 9 March 2006.
- Saeko Nomura, Toru Ishida, Mika Yasuoka. 2003. *Open Source Software Development with Your Mother Language: Intercultural Collaboration Experiment 2002*. HCI International.
- Virach Sornlertlamvanich. 2006. *KUI: The OSS-Styled Knowledge Development System*. In: Handbook of The 7th AOSS Symposium, Kuala Lumpur, Malaysia.