Development of Translation Memory Database System for Law Translation

Yasuhiro Sekine[ad], Yasuhiro Ogawa[bd], Katsuhiko Toyama[bd], Yoshiharu Matsuura[cd]

[a] *CRESTEC Inc., JAPAN*

[b] *Graduate School of Information Science, Nagoya University, JAPAN*

[c] *Graduate School of Law, Nagoya University, JAPAN*

[d] *Japan Legal Information Institute (JaLII), Nagoya University, JAPAN*

Abstract

There are inconsistency issues with the translations of Japanese laws. A translation memory tool increases consistency by recycling previous translations. However, a translation memory tool is not enough to maintain the consistency among the translations conducted by separate and unspecified translators of the general public. To tackle this problem, we developed a translation memory database system, which is managed in an integrated fashion and available for everyone. The system is open to the general public. Users can refer to the contents of the database directly by using search functions, and download them as a file, which is synchronized with the database. We expect the development and release of this system to help solve the inconsistency issues of law translation.

1. Introduction

In April 2009, the Ministry of Justice released the Japanese Law Translation Database System[1] (JLT), which dramatically improved access to translations of laws. However, some problems with law translation have been pointed out since its release. The problems can be divided into two types: problems of quantity and problems of quality. The problems of quantity refer to the small number of translated laws and delays in translation development plans[2]. The problems of quality refer to translation errors, careless mistakes and inconsistencies within and amongst the translations.

Our goal is to develop a translation support system to solve these problems using

---

1 http://www.japaneselawtranslation.go.jp

2 A list of laws to translate decided by the government every year. It is available at: http://www.japaneselawtranslation.go.jp/rel_info/rel_info_trans?re=02

information technology. There are various technologies applied to the translation process as represented by machine translation or translation memory. We pay more attention to translation memory among these and other technologies, because it seems to be a countermeasure to the problems that we are facing now. A Translation memory is widely used in the translation industry because of its great benefits. A translation memory tool increases productivity and consistency of translation work by allowing translators to recycle previous translations. However, if the translations are conducted by unspecified translators, a translation memory tool is not enough to maintain consistency, since individual translation memory databases used by unspecified and separate translators may be revised or expanded from time to time.

Namely, our purpose in this research is to solve this problem by developing a translation memory database system, which is managed in an integrated fashion and available for everyone. The system is open to the general public. Users can refer to the contents of the database directly by using search functions, and download them as a file, which is synchronized with the database. We expect the development and release of this system to help solve the inconsistency problems amongst the translations as well as reducing translators' burden to increase productivity of translation work and help solve quantity problems of law translations.

This paper is organized as follows. In the next section, we describe the background of this research, what lead us to develop the system. We show the problems we are facing now, and explain how the translation memory has a beneficial effect on the cited problems. In section 3, we introduce the translation memory database system for law translation that we developed. We show the database and some useful functions of the system, focusing on the important technology used in the system. Section 4 lays out the conclusions of this paper and prospects for the future.

2. Background of Development

There are various reasons to translate Japanese laws. The reasons of promoting the translation of Japanese laws are: to facilitate smooth international transactions for Japanese companies, to promote foreign investment in Japan, to support legal system development in developing countries and other reasons such as to enhance international understanding of Japan and to increase the convenience of foreign people living in Japan (Study Council, 2006). To respond increasing demand for law translation, the

Japanese government launched “Japanese Law Translation Database System (JLT)” in April of 2009. The authors of this paper are amongst the developers and we are still in charge of administrating the system and data management.

2.1. Problems We Are Now Facing

While we have been focused on maintaining the JLT, we have noticed some problems with law translations. The problems can be divided into two types: problems of quantity and problems of quality. The problems of quantity relate to the number of translated laws. As of September 1, 2012, the JLT provides translations of 264 laws and regulations, and this number is increasing day by day. However, when compared to the total number of laws currently in effect in Japan—there are over 7,700—the number of laws that have been translated is still low. It seems far from enough to satisfy all of the needs, as is evident in the number of requests we receive from users of the JLT who ask for more translations of laws. Moreover, translation development plans are behind schedule as more than three hundred planned translations from 2007 to 2011 have not been released yet.

The problems of quality refer to the accuracy and consistency of translations. We sometimes notice quality problems when we are checking translations before releasing them. We also sometimes receive feedback from the general public with regards to quality issues. More than one hundred errors in the JLT have been indicated by the general public since its release in 2009. The quality problems can be divided into 3 categories: translation errors, careless mistakes and inconsistent translations. In many cases, the translation errors seem to be caused by a lack of legal knowledge. The example below was an actual translation error, and it seems to have occurred because of a lack of legal knowledge.

> 第十一条　第七条第一項の規定による報告をせず、若しくは虚偽の報告をし、同項の規定による検査若しくは収去を拒み、妨げ、若しくは忌避し、又は同項の規定による質問に対して答弁をせず、若しくは虚偽の答弁をした者は、五万円以下の罰金に処する。
>
> Article 11 A person who has failed to make a report under the provisions of Article 7, paragraph (1) or has made a false report, or has refused, prevented or recused an inspection or removal under the provisions of said paragraph, or has

failed to give an answer or has given a false answer to a question under the provisions of said paragraph shall be punished by a fine of not more than fifty thousand yen.

The word "忌避(*kihi*)" in the source text is translated as "recuse" in the target text. The Japanese word "*kihi*" can be translated as "recuse" or "evade", and which term is appropriate depends on the context. In other words, the English words "recuse" and "evade" can be translated into one Japanese word, "*kihi,*" although these two English words have totally different meaning from one another. When "*kihi*" means to escape from an obligated inspection like in the example above, it should be translated as "evade" instead of "recuse," and when "*kihi*" means to challenge a judge as incompetent, "recuse" should be used. Usually, the word "*kihi*" is not used in daily life by ordinary people. Moreover, "*kihi*" needs to be translated discriminatively by its legal context. Thus, a legal translator needs to have basic knowledge of laws and be familiar with the legal terminology of both the source language and target language.

Careless mistakes are often found in grammar, spelling, format and especially in numbers: number of days, months, years, dates, amounts of money and so on. In Japanese laws, Japanese calendar is used and conversion to the Christian calendar may result in mistakes. The case of missing translation in the target text is sometimes found, and this is also another careless mistake.

Inconsistent translation is a serious problem in law. Usually, different sentences have different meanings, and for strict interpretation of the law, the same meaning should be written using the same words. Legal texts are intricately linked with each other in intra-document and inter-document ways. To have a right understanding of a certain legal matter, it is not enough to understand a single provision or even a whole law, but it is necessary to understand related provisions among related laws systematically. Inconsistent translations between referrer and referee of law texts may prevent this systematic understanding. Moreover, from a quality control standpoint, consistency is also important factor since it is easier to find errors from among consistent translations in good order. The consistency issue also effects productivity of translation work. It is very difficult and time consuming to choose the best translation from among a lot of variant translations when translator refers to previous translations.

The translation work of Japanese law is conducted under the responsibility of the

competent ministries and agencies. Moreover, they usually outsource this work to translation vendors that are chosen by public bidding. Vendors are chosen only by price since there are no specific criteria to certify their translation skills. In such a situation, keeping the consistency of the translations is very difficult. The standard legal term dictionary contributes to consistency on the word level to some extent, but more effort is still needed to increase the consistency on other levels. On the sentence level, the JLT has 30 variant translations for one sentence as shown in the example below.

この法律は、公布の日から施行する。
This Act shall be enforced from the date of promulgation.
This Act shall come into effect as from the date of promulgation.
This Act shall come into effect as of the date of promulgation;
This Act shall come into effect as of the day of its promulgation.
This Act shall come into effect on the day of promulgation.
This Act shall come into force as from the date of its promulgation.
This Act shall come into force as from the day of its promulgation.
This Act shall come into force as of the date of its promulgation.
This Act shall come into force as of the day of promulgation.
This Act shall come into force from the day of promulgation.
This Act shall enter into force on the day of its promulgation,
…

Inconsistent translations of law titles are more serious problems. If law titles are translated in different ways, they may be recognized as different laws. For example, the JLT has 6 variant translations for one law title as follows.

一般社団法人及び一般財団法人に関する法律
Act Concerning General Corporations and General Foundations
Act on General Associations and Foundations
Act on General Associations and Incorporated Foundations
Act on General Incorporated Association and General Incorporated Foundation
Act on General Incorporated Associations and General Incorporated Foundations
General Incorporated Associations/Foundations Act

To summarize, we have problems concerning quantity and quality. The problems with quantity refer to the small number of translated laws and delays in translation development plans. The problems with quality refer to translation errors, careless mistakes and inconsistencies. Our research group has been working on those issues for a long period of time. In this research, we focus on the problems of quality, especially inconsistency of translations.

A translation memory tool seems to contribute to increasing consistency of translations, since translation memory allows translators to recycle previous translations, and by recycling the previous translations, consistency among translations inevitably increases. In the next section, we give general information on translation memory. If you are familiar with translation memory, skip the next subsection and go to 2.3.

2.2. General Information on Translation Memory

In the industrial world, various technologies are applied to the translation process. The most representative of the technologies are machine translation and translation memory. Sometimes translation memory confused with machine translation, but they are different technologies. Machine translation is an automatic translation by computer. Machine translation is still a developing technology, and leaves much to be improved upon terms of quality. In many cases, texts translated by machine translation need post-editing by human translators, as fully automatic, high quality translation has not been put into practical use yet.

A translation memory is a database consists of source text and target text from previous translations. A translation memory is created by dividing source text and target text into segments (usually sentences). Source segment and its corresponding target segment are paired one-by-one and then stored in a database as shown in Fig. 1, which is usually saved as text file, TMX[3] or other format. When there is a new source segment equal or similar to one already translated, a computer-assisted translation tool retrieves the previous translation form the database. How translation memory works in the translating process as shown in Fig. 2 is as follows: (1) Translator uses a computer-assisted translation tool to search previous translation from the translation

[3] TMX is an open XML standard of translation memory data.

memory that is the same or similar to the sentence now he/she is about to translate. (2) The similar sentence found in the translation memory is often called “fuzzy match”. When there is a fuzzy match, it is retrieved from the translation memory by the computer-assisted translation tool. The parts of a sentence that do not match with the previous translation will be highlighted for the translator to process. (3) Translator copies target segment retrieved from the translation memory and (4) edit the different part of it to use as a new translation.

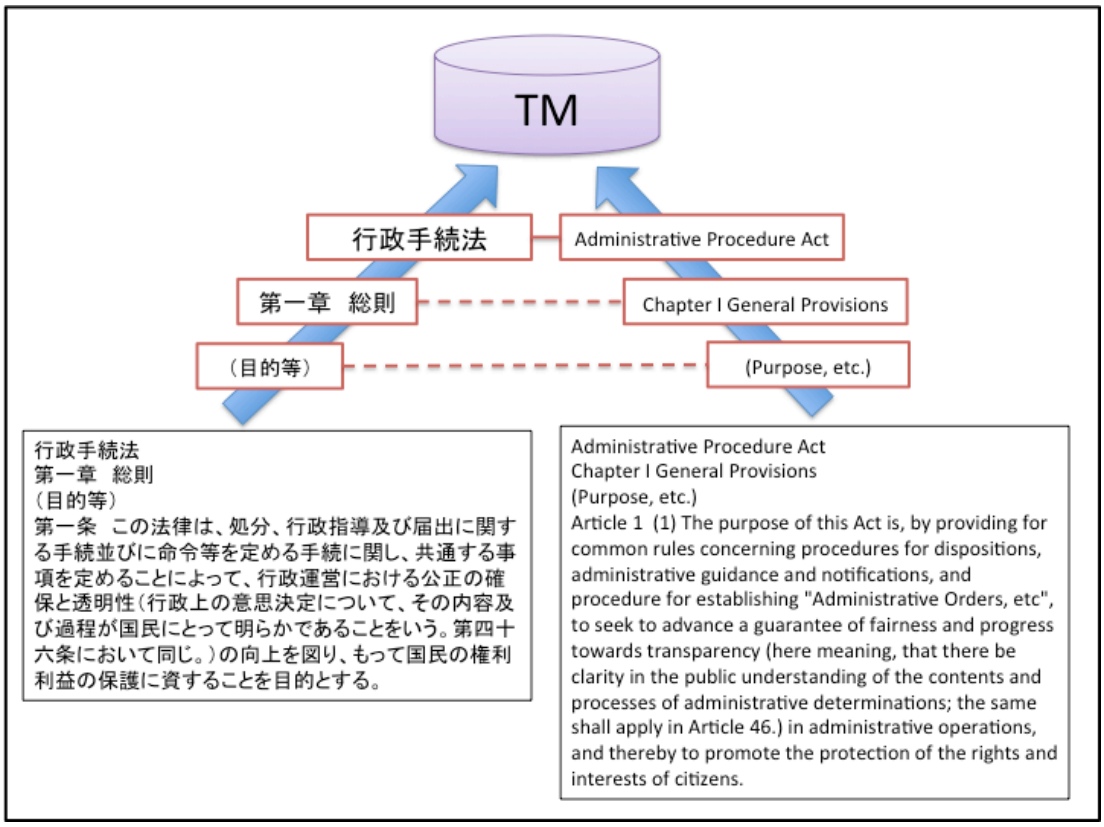

Fig. 1: How to create a translation memory

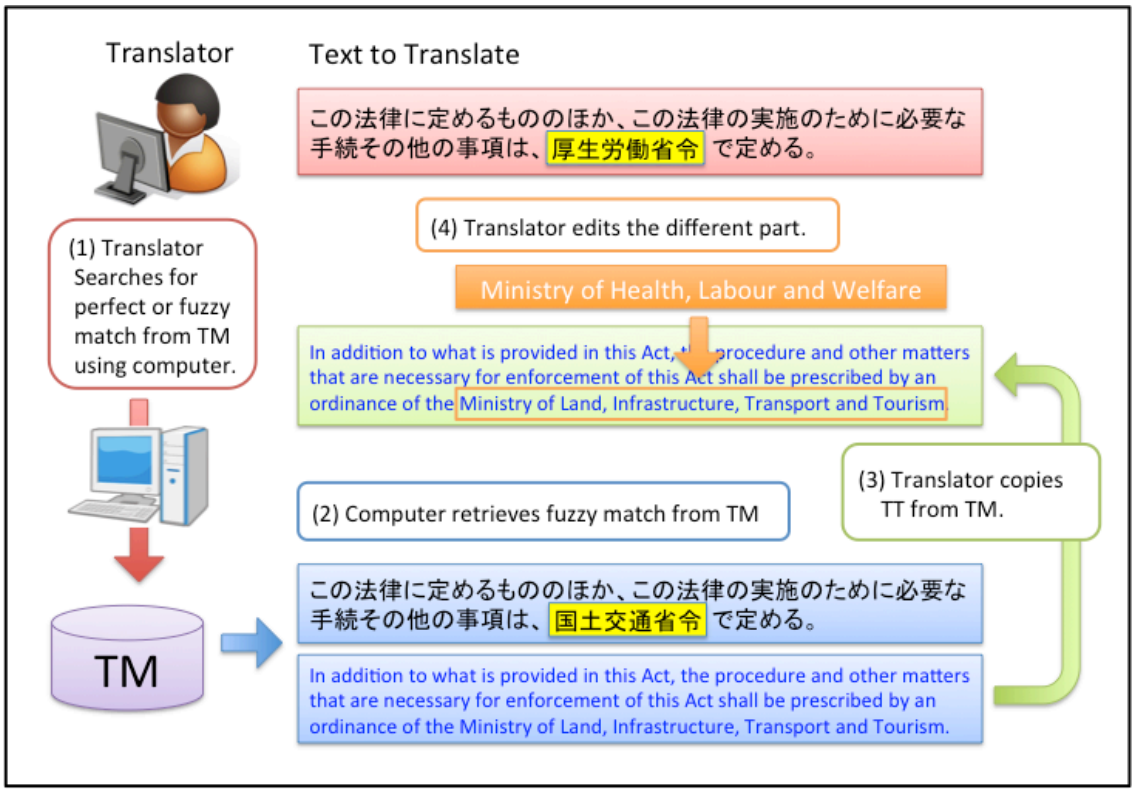

Fig. 2: How translation memory works

Sometimes “translation memory” is a confusing term. A database stores source and target texts is a “translation memory” as mentioned. However, sometimes a computer-assisted translation tool which imports translation memory to utilize and usually includes an editor, terminology, search function and other useful functions for translators also called “translation memory (Fig. 3).” To avoid confusion, hereinafter, we use “translation memory database” for the former, “translation memory tool” for the latter and just “Translation Memory”, if it is not necessary to be distinguished.

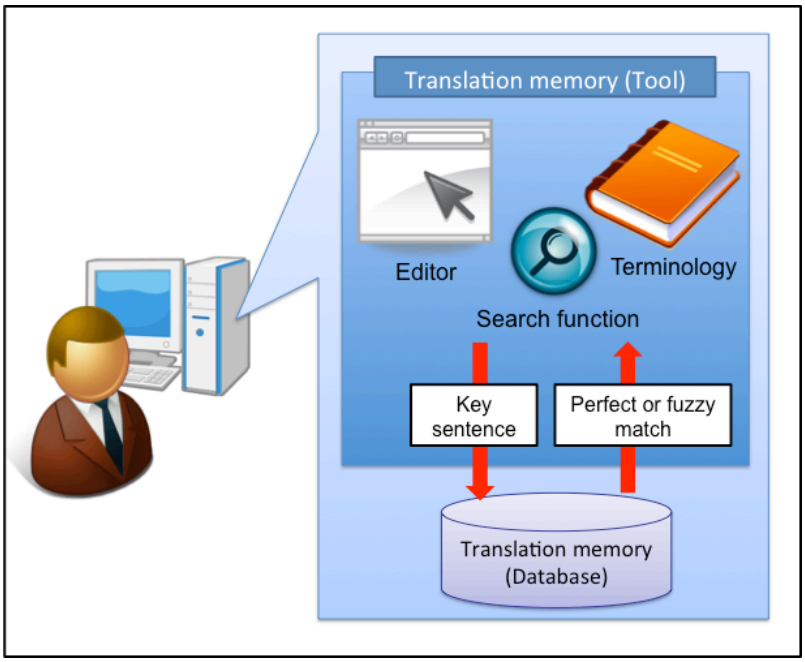

Fig. 3: Translation memory tool

The most often referred benefits of working with a translation memory are to save time and reduce translation efforts, and this productive efficiency shortens the delivery time and lowers the translation cost. The other advantage of working with a translation memory is that it increases consistency, which is needed especially in a document that requires strict understanding such as technical specifications. The aim of a translation memory is to allow translators to re-use previously translated work. It makes sense that the types of texts that are best suited for working with a translation memory are those which are repetitive or which will be updated or revised (Yamada, 2011). That is the reason why a translation memory is used for translating instruction manuals, localization of software, webpages and so on.

A translation memory is no longer an ambitious technology like machine translation. Instead, it is a practical tool already widely used in the translation industry because of its great benefits. Lagoudaki (2006) indicates that many companies producing multilingual documentation are using translation memory. In a survey of language professionals in 2006, 82.5% out of 874 replies confirmed the use of a translation memory (Lagoudaki, 2006). In Japan, over 50% of the translation service providers have adopted a translation memory or some form of computer-assisted translation (Japan Translation Federation, 2009).

2.3. Central Translation Memory

A translation memory tool increases consistency by recycling previous translations. Moreover, laws seem to have suitable texts for the translation memory mentioned in the previous subsection. There are boilerplate expressions often used in laws, and laws are often revised by nature. In our previous study (Sekine et al., 2010), we examined how law sentences are recyclable and that results showed that they are highly recyclable, which means a translation memory is effective for law translation.

However, if the translations are conducted by unspecified translators as law translations in Japan are, a translation memory tool is not enough to maintain consistency, since individual translation memory databases used by unspecified and separate translators may be revised or expanded from time to time. The situation calls for a translation memory database managed in an integrated fashion, and available to and shared by everyone. Such an integrated translation memory database is called a "central translation memory" or "central memory", while, in contrast, a translation memory database saved and used in a translator's individual environment is called a "local translation memory" or "local memory."

In the European Union, a central translation memory for law translation entitled the "Euramis" is available for translators of the Directorate General for Translation (DGT) of the European Commission. Euramis is not used directly during the translation process (DGT, 2005). All staff of the DGT have access to Euramis, and they can download the translation memory database as a local translation memory and use it with translation memory tools. Euramis is not open to the general public, but the downloadable translation memory database, "DGT-TM," is available instead to the general public[4] (Steinberger et al., 2012).

In 2010, the number of pages to be translated into 22 official languages amounted to 27,000 pages and, as a result, 215,500 pages of translations were produced (DGT, 2012) by the DGT. Moreover, drafted bills and their translations are of equal legal importance, thus they always require high quality accurate translations. Euramis has been contributing to an increase efficiency and ensures consistency of the translation work in the DGT. In Japan, such a central translation memory had not existed. Since the central

[4] http://langtech.jrc.ec.europa.eu/DGT-TM.html

translation memory seems to be the solution to the consistency problem with law translation, we started developing a central translation memory for translating Japanese laws.

Euramis is not open to the general public. Since most of laws are translated within the DGT, there may be less need to open the system to the general public. On the other hand, in Japan, the translation work is conducted under the responsibility of the competent ministries and agencies, and usually this work is outsourced to translation vendors. Therefore, a central translation memory should be open to the general public in this context.

The DGT-TM is a downloadable translation memory database, which is available to the general public. However, it is not simultaneously operated with Euramis. As of September 2012, the latest version of the DGT-TM was Version 2011. It is updated annually (Steinberger et al., 2012). Therefore, improvement or expansion of Euramis is not reflected to the DGT-TM immediately, and general users cannot refer to the latest status of the central translation memory. In Japan, the main users of a central translation memory would be the general public, so a downloadable translation memory should be synchronized with the central translation memory to ensure the availability of the latest translation memory for download.

As mentioned above, Euramis is not used directly during the translation process. The translators of the DGT need to download the translation memory database from Euramis to their local environment to use it, as do public users. This means that there are as many personalized local translation memories as users who use them. A individual's local translation memory can be edited, which may lead to inconsistencies between local translation memories. From the standpoint of maintaining consistency, it is preferable to translate by directly referring to an integrated central translation memory.

Furthermore, a translation memory tool is usually expensive. There are some open source translation memory tools, but they are not popular because of their usability. SDL Trados[5] is a de facto standard in the translation industry, and according to Lagoudaki (2006), SDL Trados is used by a total of 75% of surveyed users in the world. It is also used as a primary translation memory tool at the DGT. As is, on September 29,

[5] http://www.trados.com/en/

2012, the standard price of SDL Trados Studio 2011 Freelance was 845€[6]. Assuming there are users who cannot afford such a translation memory tool, the translation memory database should allow them to utilize the database without any tools, via the Internet, free of charge.

Euramis is a well-designed system and there are many things to learn from it. However, we are in the different situation. We started our development taking our own situation into consideration as well as things learned from the EU's experience with their system.

3. Japanese Law Translation Memory

In September 2011, we released a test version of a translation memory database system named the "Japanese Law Translation Memory[7]" as shown in Fig. 3 and started providing the translation memory via the Internet free of charge. The system is accessible for everyone without any authentication. Users can download the translation memory in CSV or TMX format for use with translation memory tools, and the system also has search functions and reference function for translators. In this section, we show statistics of the database, search functions focusing on important technology used in them, and reference function, which enables users to refer to the standard legal term dictionary.

6 http://www.translationzone.com/en/translator-shop/shop_main.asp
7 http://itrd.crestec.co.jp/transmemoryweb/

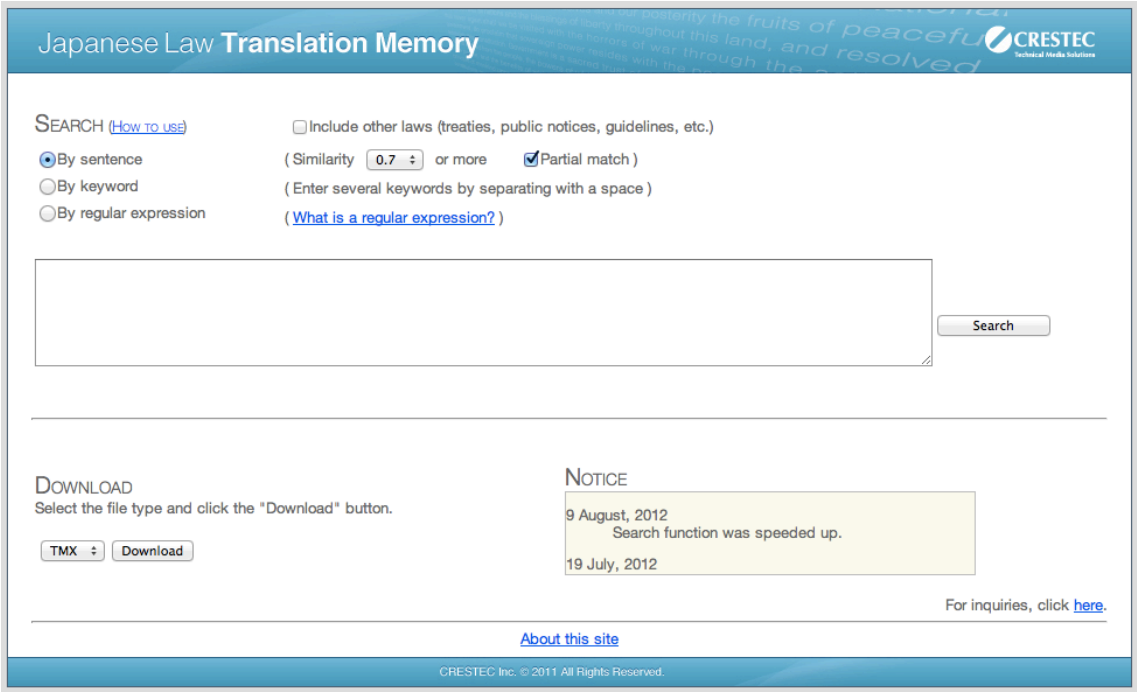

Fig. 3: Top page of Japanese Law Translation Memory

3.1. Size of the Database

As of September 1, 2012, the translation memory database consists of translations of 264 translated laws available in the JLT. In the database, each record consists of a source segment, target segment and other meta-information such as law title, law number, article number and so on. The database consists of 280,797 records including 148,542 different pairs of source-target segments and 130,196 different source segments as shown in Table 1. That the number of different source-target segments pairs is larger than the number of different source segments means that there are some variant translations for the same source segments.

There is also additional data from translations of other law data such as public notices, guidelines, messages, outline of laws and international treaties. The data of these other laws are stored separate of the essential data from the JLT since translations of laws from the JLT are based on the standard legal term dictionary, while these additional translations are not. Including these other laws, the database consists of 306,153 records, 169,028 different pairs of source-target segments and 149,282 different source segments (Table 1). For the search function, there is an option to include this additional data as a target of searching, however this additional data is not included in the translation memory database for downloading.

Table 1: Size of the database

	Essential data	Including additional data
No. of laws	264	597
No. of records	280,797	306,153
No. of different pair of source-target segments	148,542	169,028
No. of different source segments	130,196	149,282

The database contains various kinds of information in addition to the source segment and target segment: information that indicates where the source segment and target segment come from such as the law title, law number, article number, paragraph number, item number, URL to the original data and so on. This meta-information is used to show reference information in the search results. All of the information is generated automatically from the XML document structure downloaded from the JLT. The source segment and target segment contained in the database are also extracted from text nodes of XML files. The XML files of the source language and target language are separate files, so when making pairs of source segment and target segment, we used XPath to do this pairing work so as to avoid a mismatch because sometimes the source segment does not have a corresponding target segment. The text node of the target segment is extracted from the same structural node as the source segment. The reason why we use the XML files to create the database is to get meta-information of referenced laws and to avoid a mismatch of source segments and target segments.

3.2. Updating the Central Translation Memory

When new translation data is added to the JLT or a translation within the JLT is revised, and then the central translation memory is also updated within a few days. The updating is done by manual work, but it is very simple. To edit the existing segment in the central memory, it is only needed to edit the XML file where the segment comes from. To add a new law to the central memory, the XML files of a source and target language are copied to a specified directory. This is everything to do for updating the central memory. From a technical point of view, the central memory can be synchronized with the JLT by means of its simple way to update data.

3.3. Download Functions

Users can download the translation memory in CSV or TMX format for use with translation memory tools. The downloadable translation memory database is operated simultaneously with the central memory. When the contents of the database are revised or expanded, the changes are reflected immediately in the downloadable memory, therefore all users can always obtain the latest translation memory database from the system via download.

3.4. Search Functions

Users can search the content of the central translation memory directly. The system allows the translation memory download for use with translation memory tools such as SDL Trados. However, not all users have translation memory tools nor do they know how to use them. Moreover, to search similar sentences using translation memory tools, there are several steps to take: i.e., create a new project, open files to translate, select a language pair, and import a translation memory database, and it takes some time. Our system offers simple, fast and direct search of the central translation memory for everyone. It means users can utilize the translation memory database without using any tools. In the Japanese Law Translation Memory, there are three ways to search sentences from the central memory: by sentence, by keyword and by regular expression.

3.4.1. Searching by Sentence

When searching by sentence, users can see a list of similar source segments to the specified key sentence, paired with its corresponding target segments as a result of the search (fuzzy match). The purpose of using the translation memory is to recycle the same or similar sentences, so this is a primary function of the system. The search results as shown in Fig. 4 show similarity, key sentence, a pair of source segment and target segment retrieved from the central memory, and reference information including a direct link to the original data in the JLT. The differences between the key sentence and source segment are highlighted in yellow.

The similarity between the key sentence S_1 and source segment in the central memory S_2, $Sim(S_1, S_2)$ is calculated by the equation (a) using the string edit distance, $d(S_1, S_2)$, which is normalized by the maximum length, i.e., number of characters among

S_1 and S_2, $max(|S_1|, |S_2|)$.

$$Sim(S_1, S_2) = 1 - \frac{d(S_1,S_2)}{max(|S_1|,|S_2|)} \text{ (a)},$$

String edit distance is the minimum number of edits (insertion, deletion or substitution) needed to transform one string into the other. The string edit distance is a popular method to calculate similarity between sentences, however, to calculate the similarity sentence-by-sentence takes a lot of time, and it is too slow for practical use.

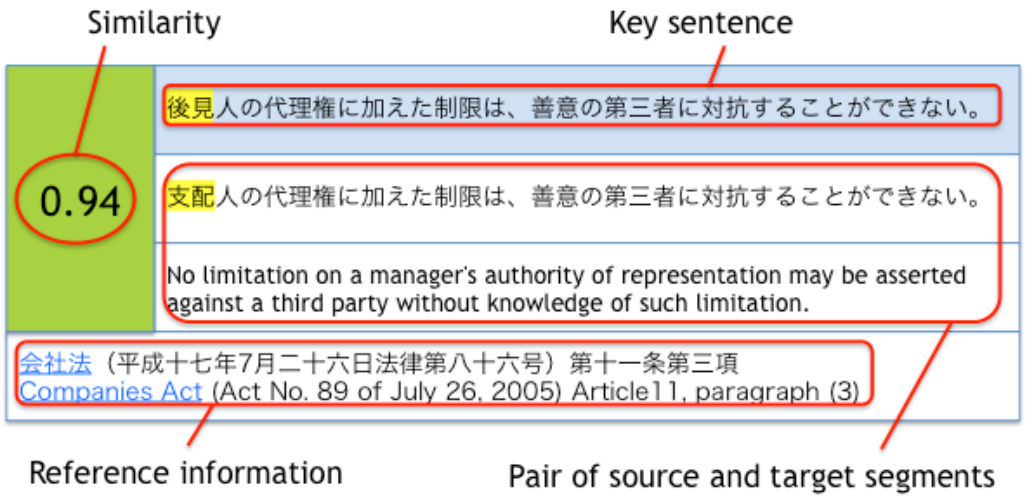

Fig. 4: Search result of "Search by sentence."

3.4.2. Improvement of Search Speed

We carried out a preliminary experiment to measure lookup time using string edit distance for computing similarity. We ran all experiments on Windows 2008 Server with Intel Xeon E5506 CPU (2.13 GHz) and 4 GB main memory. The system is implemented with MySQL and ASP.Net (Visual Basic) as web application. We used sentences from three laws that are not included in the database, containing 592 sentences with an average of 55 characters per sentence. We use each sentence from the law data as a query string and measured lookup time. The average lookup time per sentence is 77.91 seconds, so it is obvious that retrieving similar sentences from a large corpus by using string edit distance is impractical.

To speed up the searching, it is needed to reduce the number of calculation. We applied a novel algorithm proposed by Okazaki and Tsujii (2011) to narrow down the candidates of similar sentences and thereby reduce the number of calculations. Okazaki

and Tsujii (2011) proposed "A Simple and Fast Algorithm for Approximate String Matching with Set Similarity". In their study, strings are represented by unordered sets of arbitrary features (e.g., trigrams). It derives the necessary and sufficient conditions for approximate strings, count occurrence from the index and in the calculation process, strings that do not meet necessary conditions are eliminated, and strings that meet sufficient conditions are picked up from the candidate list. The most important feature of this method is the reduction of the number of the calculations. They measured query performance of approximate string matching by using three large-scaled datasets with English person names, Japanese unigrams, and biomedical entity/concept names and the experimental results demonstrate that their method outperforms state-of-the-art methods including Locality Sensitive Hashing and DivideSkip on all the datasets.

We applied this algorithm to the search function using sets of bigram morphemes for the calculation. After selecting candidate segments by this fast algorithm, then a similarity between the key sentence and source segment of candidates is calculated using the string edit distance. We measured lookup time and compared it to the result of the preliminary experiment as shown in Table 2.

Table 2: Result of the experiment

	String edit distance only	New method
Lookup time (second)	77.91	1.44

Average lookup time per sentence by the preliminary experiment is 77.91 seconds and by this new method is 1.44 seconds. The lookup time is drastically shortened and it is less than 2 percent of the lookup time by the string edit distance only. The calculated similarity is used to rank the multiple pairs retrieved from the database. If the similarities are the same, they are ranked by the number of referenced laws.

3.4.3. Other Search Functions

The system also has other way to search the content of the central memory. When searching by keywords, the list of source segments, which contains the specified keywords paired with their corresponding target segment, will be displayed. Users also can search by regular expression to specify the pattern of the source text to retrieve.

There is an example below to specify a sentence pattern described in a regular expression: “この(法律|政令)において(、)?「[^」]+」とは、”. In the example, “(法律|政令)” means “法律(*Act*)” or “政令(*Cabinet Order*)”, (、)? means that it does not matter if there is a “、(*comma*)” or not, and “[^」]+” means that there are one or more characters other than “」(*closing parenthesis*)”. By this regular expression, sentences fitting the pattern “The term *something* as used in this *Act or Cabinet Order* means…” are retrieved.

3.5. Reference Function

Another useful function to keep consistency of the translations is the reference function. The reference function is an additional function of the search function, by which users can check the translations of the words contained in a key sentence for searching. Since the standard legal term dictionary is an important source for consistent translation, we utilized it for the reference function. If a key sentence, keywords or regular expression for searching contains an entry word of the standard legal term dictionary, then it will be listed, and by clicking, the contents of the standard legal term dictionary will be displayed as shown in Fig. 5. As mentioned above, the JLT provides translations of 264 laws, and in addition to these, translations of over 1,600 titles of laws currently in effect are available. We also utilized these titles for the reference function. As with the standard legal term dictionary reference, if the key sentence, keywords or regular expression for searching contains a law title, then it will be listed, and by clicking, the translation of the title and a direct link to the original data will be displayed. As indicated in section 2.1, inconsistency amongst law titles is a serious problem, so this function can be a countermeasure to that problem.

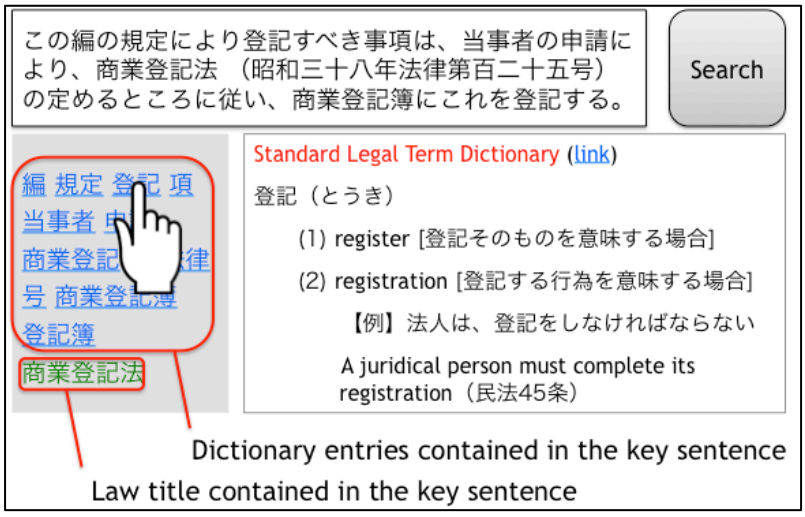

Fig. 5: Reference function

4. Conclusions and Prospects

We developed the translation memory database system for law translation, which is expected to help solve the inconsistency issues of translations of Japanese laws. The main feature of the system is that translation memory database is managed in an integrated fashion and it is open to the general public. With this feature, the central translation memory can be shared among separate and unspecified translators. The other important features of the system are that the central translation memory can be referred directly using search functions, and the downloadable translation memory is synchronized with the central memory. With these features, users can refer to the latest status of the central memory, and that can be utilized regardless of whether they have translation memory tools or not.

Although the translation memory is used to recycle translations, sometimes there is no recyclable translation in the translation memory. Law texts are drafted according to a certain format, but there are less recyclable translations as compared with localizing documents such as instruction manuals of a series of products. Machine translation might be helpful when there is no recyclable translation in the translation memory. In addition, since every translation memory tool, as represented by SDT Trados, has an editor included in the system, adding an editor to the system is another possibility for increasing productivity. The combination of machine translation and an editor is a challenge of the future. We expect that attracting more users by enhancing the usability of the system will contribute to greater consistency within and amongst law translations

in Japan.

Acknowledgement

The authors would like to express their gratitude to Professors Tokuyasu Kakuta, Makoto Nakamura, and Tariho Kimura of Nagoya University for their thoughtful guidance. We would also like to thank Professor Kayoko Takeda and Dr. Masaru Yamada for their advice from the field of translation studies. Our appreciation also goes to the people who kindly supported the development of the system, especially, Mr. Tadashi Ohtani, Mr. Daichi Saito, Mr. Nobuyuki Tsunatori and Mr. Akira Takabayashi of CRESTEC Inc. Our special gratitude is also extended to the administrators of the Japanese Law Translation Database System in the Judicial System Department, Japanese Ministry of Justice, for their supervision and support.

References

Directorate-General for Translation of the European Commission (2012). *Studies on translation and multilingualism: Language and Translation in International Law and EU Law.* Luxembourg: Publications Office of the European Union.

Directorate-General for Translation of the European Commission (2005). *Translation Tools and Workflow*, Brussels, European Commission.

Japan Translation Federation. (2009). *The 2009 Translation White Paper* (in Japanese). Tokyo: Author.

Lagoudaki, E. (2006). *Translation Memories Survey 2006: Users' perceptions around TM use*. In "Translating and the computer 28. Proceedings of the twenty-eighth international conference on translating and the computer, 16-17 November 2006, London", London: Aslib.

Okazaki, N. & Tsujii, J. (2011). A Simple and Fast Algorithm for Approximate String Matching with Set Similarity. *Journal of Natural Language Processing (in Japanese),* 18(2), 89-117.

Sekine, Y., Saito, D., Ogawa, Y., Toyama, K., & Matsuura, Y. (2010). Effectiveness of Translation Memory in Law Translation. *In Proceedings of Annual Conference of Tokai-Section Joint Conference on Electrical and Related Engineering in 2010 (in Japanese)*, F1-1, 1, Nagoya, Japan.

Steinberger, R., Eisele, A., Klocek, S., Pilos, S. & Schlüter, P. (2012). DGT-TM: A

freely Available Translation Memory in 22 Languages. *In Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'2012)*, Istanbul, 21-27 May 2012.

Study Council for Promoting Translation of Japanese Laws and Regulations into Foreign Languages. (2006). Final Report. Tokyo: Author. Retrieved from www.cas.go.jp/jp/seisaku/hourei/report.pdf

Yamada, M. (2011) *Revising Text: An empirical investigation of revision and the effects of integrating a TM and MT system into the translation process*, Thesis for the degree of Doctor of Philosophy, Graduate School of Intercultural Communication, Rikkyo University.