

Technical Paper

Establishment of Komatsu Technical Word Dictionary and Introduction of NMT Tool

Shinya Tsukahara

The technical terms and sentences are on the increase due to the more and more complicated functions and structure of construction machinery in recent years. Besides, the speeding up of development and globalization require the acceleration of communication beyond language barriers. This paper reports the establishment and operation of a unified technical word dictionary for the purpose of managing the increasing terms and the introduction of a neural machine translation (NMT) tool aimed at speeding up communication.

Key Words: Technical word dictionary, NMT, Globalization

1. Introduction

The recent compliance with exhaust gas regulations, introduction of ICT and electrification has caused the structures of the machines developed by Komatsu to be more complicated. With more complicated machine structures, the volume of the documents explaining the structures increases. Actually, the pages of the operation and maintenance manual of the PC200 hydraulic excavators developed by Komatsu have increased by approximately 25% for the past 5 years.

As the structures become more complicated and the number of pages of documents increases accordingly, the number of terms increases. As a result, “similar but actually different terms” are generated that express the same parts and functions with different terms. The users of such documents misunderstand that the “similar but actually different terms” refer to different parts and functions, which causes a confusion (Fig. 1).

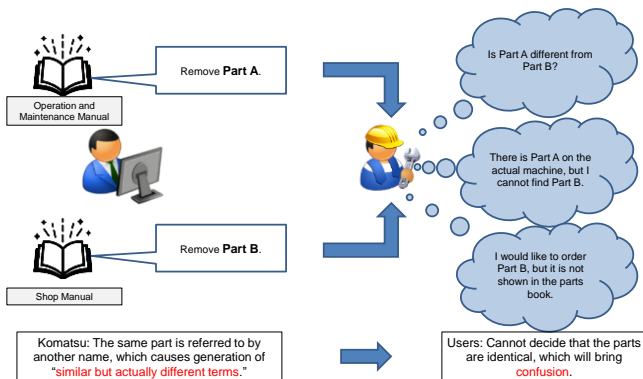


Fig. 1 Confusion caused by similar but actually different terms

To cope with such a situation, we established the “Komatsu Technical Word Dictionary” that defines the terms used uniformly in the company.

Moreover, the global communication using non-native language is continuously increasing due to the globalization in recent years.

In order to meet the needs for global communication in Komatsu (Fig. 2), we introduced a highly accurate and secure NMT tool.

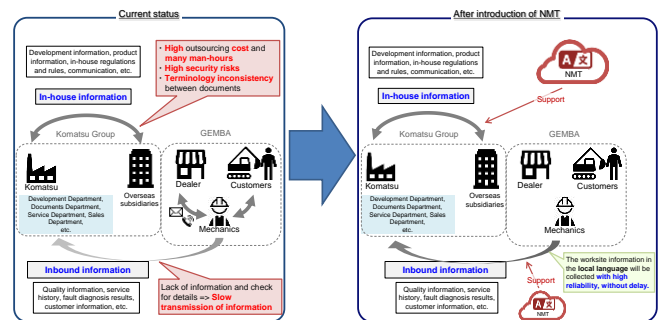


Fig. 2 Concept of using NMT

2. Komatsu Technical Word Dictionary

Komatsu Technical Word Dictionary is an in-house technical word dictionary which has 6,752 words as of January 2022.

This section describes the creation, publication and management of Komatsu Technical Word Dictionary.

2.1 Creation of Komatsu Technical Word Dictionary

Komatsu Technical Word Dictionary was created in the two steps: Extraction and sorting of terms.

2.2.1 Extraction of terms

In the first step, which is extraction of terms, we extracted terms comprehensively, including the terms which seem to be necessary for the communication connecting the people inside and outside the company as well as communication within Komatsu.

Specifically, approximately 120,000 terms in total were extracted from the following three types of sources:

(1) Existing in-house technical word dictionary

We collected the terms shown in the existing in-house technical word dictionaries, which are the “Technical Word Dictionary for Manual Translation” used in translating the service documents such as the operation and maintenance manuals and shop manuals from Japanese to English and the “Monitor UI Term List” which is a collection of the terms for display on the monitors installed on Komatsu’s products (Fig. 3). These technical word dictionaries are already in a multilingual form because they are based on the assumption of use in translation or the data composed of the stored terms as the result of translation. Therefore, we were able to extract the defined terms as they were.



Fig. 3 Words displayed on machine monitors

As one example of the terms extracted from the “Technical Word Dictionary for Manual Translation,” there is a term: “aftertreatment device.” The Japanese term corresponding to this term (“*atoshori sochi*”) can be also translated as “post-processing equipment.” Actually, if the Japanese term “*atoshori sochi*” is translated into English with Google translation as of January 2022, it is translated as “post-processing equipment.” If the registration of “*atoshori sochi*”= “aftertreatment device” is not defined in the technical word dictionary, it may be translated as a wrong term “post-processing equipment.” This “post-

processing equipment” easily misleads the readers of the documents in English into believing that there is another device different from an “aftertreatment device,” which will result in wasted man-hours required to search for the parts that do not actually exist.

Thus, we extracted the terms which can be translated into different words depending upon the translator’s interpretation and may become completely different expressions.

(2) Existing in-house documents

We extracted the terms used in the operation and maintenance manuals, shop manuals and other service documents as the existing in-house documents.

To extract the terms from the service documents (Fig. 4), we used the “morphological analysis” that decomposes the sentences explaining the machine operations and functions into parts of speech and the method of analyzing the numbers of their appearances and the connections between parts of speech and judging the significance of the terms.



Fig. 4 Service documents

The terms extracted from the shop manuals by using the above method include “*nenryo chosei daiyaru*” = “fuel control dial” and “*komon reiru nenryo funsha shisutemu*” = “common rail fuel injection system.” Both of these terms include the word “*nenryo*” = “fuel.” However, this word “*nenryo*” = “fuel” is a general term and therefore does not have to be registered in Komatsu Technical Word Dictionary. On the other hand, the units of expressions: “*nenryo chosei daiyaru*” = “fuel control dial” and “*komon reiru nenryo funsha shisutemu*” = “common rail fuel injection system” are the expressions to be uniformly used as part names and function names and should be registered in Komatsu Technical Word Dictionary.

(3) Technical word dictionaries published by Japanese and overseas public authorities

As for the technical word dictionaries published by Japanese and overseas public authorities, we collected the terms from the Japanese Industrial Standards (JIS) defined by the Japanese Industrial Standards Committee and the ISO standards defined by the International Organization for Standardization. The terms defined by the JIS are consistent with those in the ISO standards. For example, the terms related to hydraulic excavators defined in “ISO 7135: 1993 Earth-moving machinery – Hydraulic excavators – Terminology and commercial specifications” are defined in two languages, Japanese and English, in JIS A8403-1-1996 “Earth-moving machinery – Hydraulic excavators - Part 1: Terminology and commercial specifications.” Therefore, they are advantageous in that the standard terms and definitions in English as well as in Japanese can be collected all at once.

For example, the term “*senkai chusha bureki*” = “swing parking brake” was extracted from this JIS A8403-1-1996 “Earth-moving machinery – Hydraulic excavators - Part 1: Terminology and commercial specifications.”

2.1.2 Sorting of terms

As the next step, we chose the terms to be shown in Komatsu Technical Word Dictionary, from approximately 120,000 terms extracted in the previous step.

In sorting out the extracted terms, the first thing we did was establishing the standards for the terms to be listed in the dictionary.

These standards will become the criteria for making decisions on the needs for registration when adding terms to Komatsu Technical Word Dictionary in the future, together with the conditions for extracting terms which were explained in the term extraction step in the preceding section.

The following terms should be registered in Komatsu Technical Word Dictionary:

- Terms which can be translated into different words and may become completely different expressions (e.g. “*atoshori sochi*”= “aftertreatment device”)
- Technical terms specific to the industry (e.g. “*senkai chusha bureki*” = “swing parking brake”)
- Technical terms specific to Komatsu (e.g. “KomVision”)

The following words/terms will not be registered in Komatsu Technical Word Dictionary:

- Words produced by combining terms
- General words without technicality
- Terms including the words which can be interpreted into multiple meanings
- Terms including the explanatory clauses that contain verbs and/or modifying expressions

According to these standards, we extracted necessary terms from the approximately 120,000 terms collected in the previous extraction step.

As an example of the standards, this paper introduces a case of not registering the “terms including the words which can be interpreted into multiple meanings.”

This standard was adopted from the rules of Komatsu Global Standard Language (KGSL) and Komatsu Manual Standard Language (KMSL) that Komatsu established in 2017 on the basis of the Simplified Technical English Rules, ASD-STE100 ^{*1}.

*1: The descriptive rules for the manuals that even non-native speakers of English can understand without misunderstanding, which were developed by the AeroSpace and Defense Industries Association of Europe (ASD).

KGSL sets a rule of “one meaning for one word,” which means that the same word should be always used with the same meaning to prevent the readers from misunderstanding. In compliance with this rule, we made it a rule not to register the “terms including the words which can be interpreted into multiple meanings” in Komatsu Technical Word Dictionary as much as possible.

2.2 Publication of Komatsu Technical Word Dictionary

Komatsu Technical Word Dictionary in Excel macro format is published to some users on Komatsu's in-house SharePoint site as of January 2022, and they have already started using Komatsu Technical Word Dictionary for translation of manuals (Fig. 5).

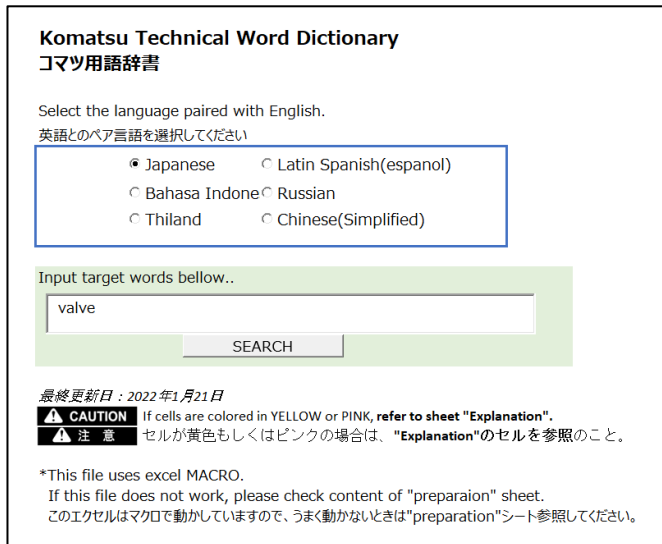


Fig. 5 Screen of Komatsu Technical Word Dictionary

Moreover, if you read Komatsu Technical Word Dictionary into the NMT tool described later in this paper, you can create translation that surely incorporates the contents of Komatsu Technical Word Dictionary.

We will compile Komatsu Technical Word Dictionary into a database in the future for the purposes of publication to the users in the whole Komatsu Group, linkage with other databases and searchability enhancement.

2.3 Management of Komatsu Technical Word Dictionary

New functions and new parts to support them are developed every day for constant improvement of products. The emergence of such new functions and parts leads to generation of new terms. Unless new terms are shown in the technical word dictionary, you cannot make decision on which terms are correct, which will result in generation of “similar but actually different terms.” Therefore, it is extremely important to add new terms to a technical word dictionary without omission at an appropriate timing.

Besides, an approval workflow to deploy the extracted terms in multiple languages is required to share the information globally.

We are considering a term approval workflow as shown in Fig. 6, taking the above into consideration.

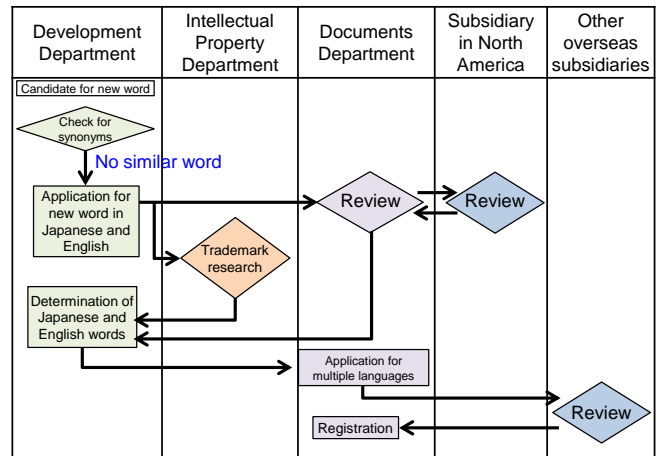


Fig. 6 Term approval workflow

It was decided that Development Department would register new terms at the phase of developing machines prior to starting creating internal and external documents in order to ensure the registration of terms without omission at an appropriate timing.

The registered terms will be translated into the languages as required to share them globally.

3. NMT tool

The precision of AI translation tools has improved significantly in recent years by introducing neural machine translation that uses deep learning through the neural network. Formerly, the machine translation between the languages with different sentence structures such as Japanese-English translation was avoided because its precision was lower than the translation between the languages with the same sentence structures such as English-French translation. However, the precision of translation between the former language pairs has improved in recent years.

In Komatsu, a global company with 210 distributors and service agents deployed in 148 countries over the world, the opportunities for communication in non-native languages occur on a daily basis. Accordingly, the NMT tool which can support sharing of timely information needs to be efficiently used in various daily services.

This section reports the overview, introduction case and improvement case of the NMT tool that Komatsu has introduced.

3.1 Overview of NMT tool

Komatsu has introduced T-400, the NMT tool supplied by Rozetta Corp. As of January 2022, 585 people at 51 departments in Komatsu Group are using the tool. Approximately 11,800,000 words were translated for the past one year.

Mainly the following texts are translated with the tool:

- In-house technical documents such as product information
- Communication information such as e-mail messages
- Procedure manual for using in-house systems
- In-house regulations and rules

On the other hand, T-400 is not used for translation of external sales documents and service documents such as product catalogs, operation and maintenance manuals, and shop manuals.

The reasons for the above are that it cannot be used together with the computer aided translation tool that allows reuse of the sentences translated in the past and that they are the documents intended for external users and therefore require high translation precision to translate more subtle nuances accurately.

Specifically, T-400 is not used for translation of the following documents:

- Operation and maintenance manuals and other manuals for external users
- External sales documents such as product catalogs
- Documents using many abbreviations such as character strings on in-vehicle monitor screens

3.2 Introduction of NMT tool

This section introduces the introduction process of the NMT tool.

The NMT tool was introduced in three steps: selection, trial and publication.

3.2.1 Selection of NMT tool

Prior to the selection of an NMT tool, we analyzed the in-house needs for translation. As a result, we have found the following four needs:

(1) Needs on translation language

Komatsu has been developing critical components such as engines and hydraulic components in Japan. Therefore, the communication through the Japanese language is required both for the inbound information from the outside to the inside and the outbound information from the inside to the outside. Thus, we have found that there are high needs for a tool with high precision in translation of Japanese which is a pivot language.

(2) Needs on data for translation

We have found that many data in Excel format are used to manage the schedule for managing the development progress situations of products and the raised issues in a list format.

Some of these files in Excel format exceed 10,000 words per file because text boxes are used for the purpose of supplementing the data while information is separately input in multiple sheets.

Therefore, we set the following conditions: The comments and the characters in text boxes contained in the Excel files should be extracted without omission; the layout should not be deformed; and the number of processable characters and the file capacities should be large.

(3) Needs on use of translation and operational cost

An NMT tool is a tool which may be used by a great many employees. Therefore, we also have taken into consideration long-term stability in use, that is, low maintenance costs.

We explored an NMT tool available on the market, instead of the In-house original tool that may require additional development with the OS version upgrade.

Additionally, we also considered a low cost of usage to be an important element so that we can use the tool without hesitation in our various daily services.

(4) Needs for security of data for translation

There were needs for a highly secure tool without information leakage risk due to many opportunities of translating highly confidential texts for internal use only for products under development.

We explored an NMT tool that meet the above four needs and compared and studied the tools available on the market, and as a result, decided to introduce T-400, the NMT tool supplied by Rozetta Corp.

3.2.2 Customization of NMT tool

Prior to using the NMT tool on a trial basis, we customized the translation engines to reflect the technical terms specific to the company and construction machinery field in the translation results, and registered Komatsu Technical Word Dictionary.

(1) Customization of translation engines

T-400 has 2,000 types of translation engines tailored to the users' data for translation. We customized the translation engines in the "machinery" field, which is the closest to Komatsu's business domain among these existing translation engines by letting them read the translation data of Komatsu's operation and maintenance manuals and shop manuals. In order to evaluate the performance of the customized translation engines, we translated 100 sentences extracted from those contained in the operation and maintenance manual for the PC200 series, Komatsu's major products, from Japanese into English with T-400 and evaluated the results.

For evaluation, we used the human evaluation method for evaluating the accuracy of the contents of the 100 sentences translated in English, using the following grading scale from 0 to 2:

0 point: Cannot be understood.

1 point: The intention can be understood though it has a sense of linguistic discomfort.

2 points: The meaning is understood without linguistic discomfort.

The average points of the 100 sentences were 1.76 as shown in Fig. 7.

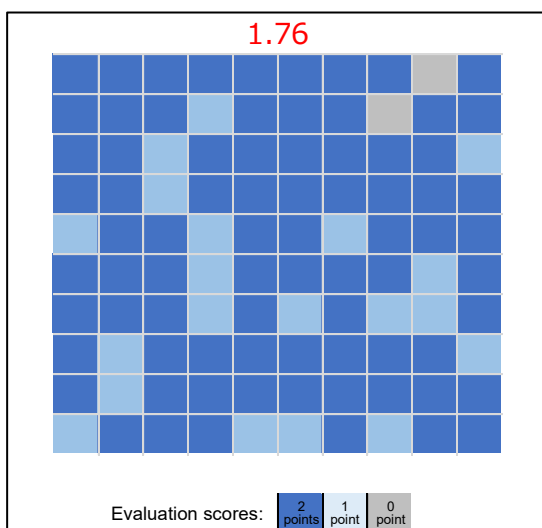


Fig. 7 Heat map for evaluation of 100 sentences

(2) Registration of Komatsu Technical Word Dictionary

With T-400, the technical word dictionary created by the users can be registered and reflected in the translation results. At Komatsu, we registered in T-400 Komatsu Technical Word Dictionary introduced in the preceding section to ensure that the In-house technical terms can be incorporated into the translation results.

3.2.3 Publication of NMT tool

The NMT tool has been introduced in advance to Development Department and to other departments as requested.

The man-hour reduction, outsourcing cost reduction and work efficiency improvement effects at the departments that installed the tool are currently under verification.

We will expand the introduction of the tool throughout the whole Komatsu Group as soon as the effects can be demonstrated.

4. Future tasks

We have the following three tasks for the future:

- Construction of an efficient operational workflow for the technical word dictionaries
- Introduction of the NMT tool in translation of service documents
- Deployment of the NMT tool throughout the whole company

4.1 Construction of efficient operational workflow for technical word dictionaries

Technical word dictionaries are never completed. Increase of new functions and parts requires new terms. Unless an operational workflow for adding correct terms quickly is built, the technical word dictionaries will be unreliable, without registration of necessary terms.

In order to add correct terms quickly, an operational workflow as shown by the workflow in Section 2.3 needs to be built that allows the maintenance of the technical word dictionaries in collaboration throughout the whole Komatsu Group, including Development Department, Intellectual Property Information Management Department, Document Production Department and overseas subsidiaries.

4.2 Consideration of introducing NMT tool in translation of service documents

Currently T-400 is not used for translation of service documents as stated in **Section 3.1**.

The reason for the above is that the computer aided translation tool, highly effective in translation of service documents that have many repeated expressions and reused sentences, cannot be used together with T-400.

However, the In-house needs for translating a large volume of texts for a short time are increasing day by day as repeatedly stated in this paper. The translation of service documents also needs to be expedited. As countermeasures against that, it is necessary to consider the introduction of an NMT tool which can be used for translation of service documents.

Specifically, we are considering the introduction of a tool which enables us to use an NMT tool, instead of human translation, on the computer aided translation tool and human translators to revise (post-edit) the texts translated with the NMT tool.

4.3 Deployment of NMT tool throughout the whole company

The introduction effect of the NMT tool is under verification by limiting the tested objects as stated in **Section 3.2**. We intend to deploy the tool throughout the whole company of Komatsu upon the confirmation of the introduction effect.

5. Conclusion

The establishment of Komatsu Technical Word Dictionary and the introduction of an NMT tool have just started and have the issues to be resolved as listed in **Chapter 4**. I would like to continue improvements, asking the Japanese and overseas departments for their cooperation, in the future.

Acknowledgments

I would like to take this opportunity to express my deepest gratitude to the NMT project team and the local and overseas term reviewers in charge for their collaboration in the establishment of Komatsu Technical Word Dictionary, and to the members of Rozetta Corp. for their delivery of the information on T-400 and their reviews of this paper.

References

- [1] Japanese Industrial Standards Committee, "Earth-moving machinery – Hydraulic excavators - Part 1: Terminology and commercial specifications," A8403-1-1996, JIS.
- [2] Matsumura, Y., "Professionals' manual translation," pp. 60-86 (in Japanese), 2014.

Introduction of the author



Shinya Tsukahara

Joined Komatsu General Service Co., Ltd. in 2007.

Document Technology Development Center, Development Division

[A comment from the author]

Face-to-face communication is currently difficult due to the COVID-19 crisis. Under the circumstances, the use of unified terms and the production of documents beyond language barriers are becoming more important than ever, for the purpose of exchanging information under a common understanding. I would like to continue improvements in the future so that the establishment of Komatsu Technical Word Dictionary and the introduction of an NMT tool as introduced in this paper may help resolve such communication obstacles.