

技術論文

コマツ用語辞書制定および辞書を用いたAI翻訳の導入

Establishment of Komatsu Technical Word Dictionary and Introduction of NMT Tool

塚原 伸也  
Shinya Tsukahara

近年の建設機械の機能・構造の複雑化に伴い、技術用語や文章は増加傾向にある。また、開発のスピードアップとグローバル化に伴い、言語の障壁を超えたコミュニケーションのスピードアップを図る必要がある。本稿では、増加する用語の管理を目的とした統一用語辞書の制定・運用と、コミュニケーションのスピードアップを目指したAI翻訳ツールの導入について報告する。

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: 用語辞書, AI翻訳, グローバル化

1. はじめに

近年の排出ガス規制への対応や、ICT化・電動化により、コマツの開発する機械の構造は複雑化している。機械の構造の複雑化により、構造を説明する資料のボリュームは増加する。実際に、コマツが開発したPC200系油圧ショベルの取扱説明書のページ数は過去5年間で約25%増加した。

構造の複雑化とそれに伴う資料のページ数の増加は用語の数は増加させ、同一の部品や機能を異なる用語で表現する「似て非なる用語」が生まれる。資料のユーザは「似て非なる用語」が異なる部品や機能を指していると誤解し、混乱を招く。(図1)

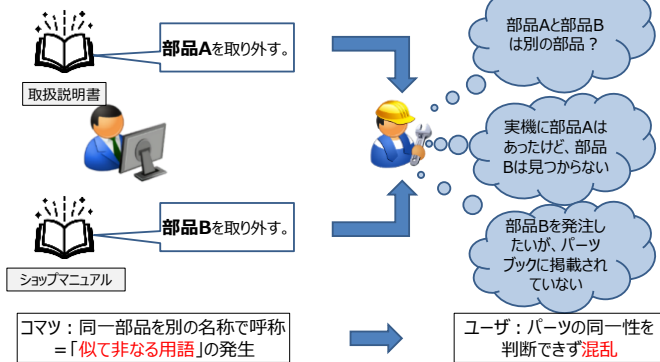


図1 似て非なる用語による混乱

このような状況に対応するために、社内で統一的使用する用語を規定する「コマツ用語辞書」を制定した。

また、近年のグローバル化により、母国語でない言語を使ったグローバルなコミュニケーションは増加し続けている。

コマツ内におけるグローバルなコミュニケーションのニーズ(図2)に対応するために、高精度かつセキュリティ性の高いAI翻訳ツールを導入した。

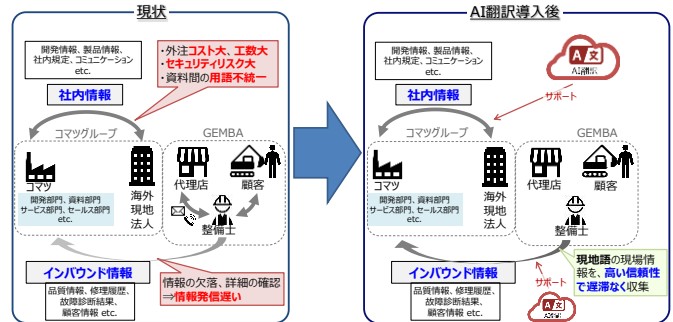


図2 AI翻訳の利用の考え方

2. コマツ用語辞書

コマツ用語辞書は2022年1月時点で6,752語の語数を持つ社内用語辞書である。

本章ではコマツ用語辞書の作成、公開および管理について説明する。

2.1 コマツ用語辞書の作成

コマツ用語辞書は用語の抽出と整理の2ステップで作成した。

### 2.1.1 用語の抽出

用語の抽出ステップではコマツ内のコミュニケーションのみではなく、社内と社外を繋ぐコミュニケーションにおいて必要と思われる用語も含めて網羅的に用語を抽出した。

具体的には、以下の3種類のソースから、合わせて約12万件の用語を抽出した。

#### (1) 既存の社内用語辞書

既存の社内用語辞書としては、取扱説明書、ショップマニュアルといったサービス資料を日本語から英語に翻訳する際に使用していた「マニュアル翻訳用語辞書」や、コマツの製品に搭載するモニタ（図3）に表示する用語を集めた「モニタUI用語リスト」に掲載されている用語を収集した。これらの用語辞書は翻訳時の利用を想定したものや、翻訳結果の用語を蓄積したデータのため、既にマルチリンガルの形態となっている。そのため規定されている用語をそのまま抽出することができた。

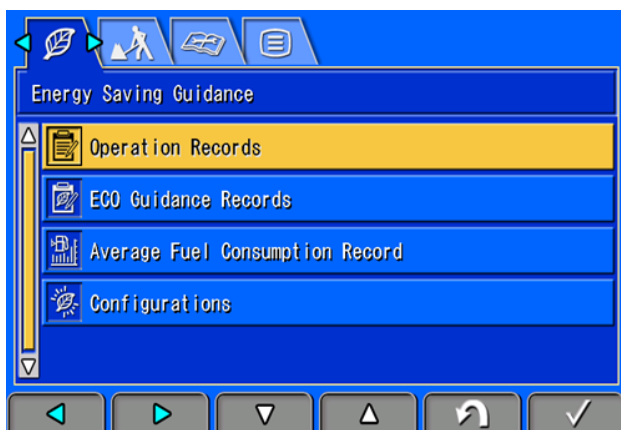


図3 機械モニタに表示される文言

「マニュアル翻訳用語辞書」から抽出した用語の一例として、“後処理装置”=“aftertreatment device”という用語がある。この用語は“post-processing equipment”とも訳出でき、実際に2022年1月時点で“後処理装置”をgoogle翻訳で英訳すると、“post-processing equipment”と訳出される。“後処理装置”=“aftertreatment device”という登録が用語辞書で定義されていない場合、誤った用語“post-processing equipment”と訳出する可能性がある。英語資料の読者はこの“post-processing equipment”を“aftertreatment device”とは異なる別の装置が存在すると誤解し、実際には存在しない部品を探す無駄な工数がかかる。

このように、翻訳者の解釈によって異なる単語を使った訳出が可能で、かつ全く異なる表現になり得る用語は抽出対象とした。

#### (2) 既存の社内資料

既存の社内資料としては、取扱説明書、ショップマニュアルなどのサービス資料で使用している用語を抽出した。

サービス資料（図4）からの用語抽出にあたっては、機械の操作説明や機能を説明する文章を品詞の単位に分解する「形態素解析」や、登場回数や品詞間の接続を分析して用語としての重要度を判断する手法を利用した。



図4 サービス資料

この手法を使ってショップマニュアルから抽出した用語として“燃料調整ダイヤル”=“fuel control dial”と“コモンレール燃料噴射システム”=“common rail fuel injection system”という用語がある。これらの用語にはいずれも“燃料”=“fuel”という表現が含まれているが、この“燃料”=“fuel”という表現は一般的な用語のためコマツ用語辞書には登録する必要はない。一方で、“燃料調整ダイヤル”=“fuel control dial”と“コモンレール燃料噴射システム”=“common rail fuel injection system”という表現の単位は部品名、機能名として統一的に使用するべき表現であり、コマツ用語辞書として登録するべき表現となる。

### (3) 国内外の公官庁が発行する用語辞書

国内外の公官庁が発行する用語辞書については、日本工業規格が規定するJIS規格や国際標準化機構の規定するISO規格から用語を収集した。JISが規定する用語はISO規格の用語と整合性があり、例えばJIS A8403-1-1996「土工機械－油圧ショベル－第1部：用語及び仕様項目」では“ISO 7135: 1993 Earth-moving machinery－Hydraulic excavators－Terminology and commercial specifications”で規定している油圧ショベルに関する用語を日英の2言語で規定している。そのため、日本語と同時に英語の標準的な用語および定義を一度に収集することができるという利点がある。

一例として“旋回駐車ブレーキ”＝“swing parking brake”という用語をこのJIS A8403-1-1996「土工機械－油圧ショベル－第1部：用語及び仕様項目」から抽出した。

#### 2.1.2 用語の整理

次に、抽出した約12万の用語のなかからコマツ用語辞書に掲載すべき用語を取捨選択した。

抽出した用語を取捨選択するにあたって、まず、用語辞書に掲載すべき用語の基準を規定した。

この基準は、前項の用語抽出ステップで説明した用語の抽出条件と合わせて今後コマツ用語辞書に追加する際に登録可否を判断する基準になる。

#### コマツ用語辞書に登録する用語：

- 異なる単語を使った訳出が可能で、かつ全く異なる表現になり得る用語（例：“後処理装置”＝“aftertreatment device”）
- 業界特有の専門用語（例：“旋回駐車ブレーキ”＝“swing parking brake”）
- コマツ固有の専門用語（例：“KomVision”）

#### コマツ用語辞書に登録しない語：

- 用語同士を合成した語
- 技術的な専門性のない一般的な語
- 複数の意味に解釈できる単語を含む語
- 動詞や修飾表現を含む説明的な節を含む語

これらの基準によって、先の抽出ステップで収集した約12万語のなかから必要な用語を抽出した。

基準の一例として、「複数の意味に解釈できる単語を含む語」に登録しない点について紹介する。

この基準は、ASD-STE100 Simplified Technical Englishルール（※1）をベースにコマツが2017年に規定したKomatsu Global Standard Language (KGSL) およびKomatsu Manual Standard Language (KMSL)のルールから採用した。

（※1）：欧州航空宇宙防衛産業協会（ASD）が開発した英語を母国語としない話者でも誤解なく理解できるマニュアルの記述規則

KGSLでは読者の誤解をふせぐために、同一の単語は常に同じ意味で使用する「一語一義」のルールを定めている。このルールに合わせて、コマツ用語辞書でも「複数の意味に解釈できる単語を含む語」は極力登録しないこととした。

#### 2.2 コマツ用語辞書の公開

コマツ用語辞書は2022年1月現在、コマツ社内のSharePointサイト上にエクセルマクロ形式で一部のユーザに公開し、既にマニュアルの翻訳にはコマツ用語辞書の利用を始めている。

（図5）



図5 コマツ用語辞書画面

また、後述するAI翻訳ツールにコマツ用語辞書を読み込むことで、コマツ用語辞書の内容を確実に織り込んだ翻訳ができるようになっている。

今後は、全コマツグループユーザへの公開、他のデータベースとの連携、検索性の向上のために、コマツ用語辞書のデータベース化を進めていく。

#### 2.3 コマツ用語辞書の管理

製品に対する絶え間ない改善のために、新しい機能やそれを支える新しい部品は日々開発されている。このような新しい機能や部品の登場は、新しい用語を生む。新しい用語が用語辞書に掲載されていないと、どの用語が正しい用語であるかを判断することができず、「似て非なる用語」が生まれていく。そのため、用語辞書においては、新しい用語を漏れなく、適切なタイミングで追加していくことが非常に重要になる。

また、情報をグローバルに共有するために、抽出した用語を多言語に展開するための承認ワークフローが必要である。

これらを考慮して、**図6**のような用語承認ワークフローを検討している。

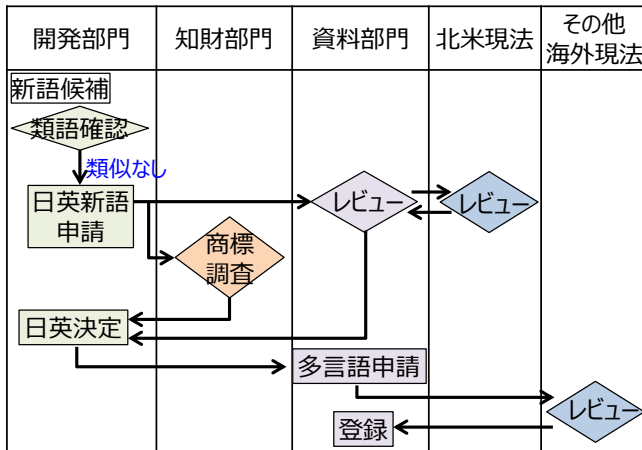


図6 用語承認ワークフロー

用語を漏れなく、適切なタイミングで登録するために、社内外の資料の作成を開始する前の機械を開発する段階で、開発部門が新しい用語を登録することとした。

登録した用語はグローバルに共有するために、必要な言語に翻訳する。

### 3. AI翻訳ツール

ニューラルネットワークによるディープラーニングを使ったニューラル機械翻訳の導入により、AI翻訳ツールの精度は近年大きく改善している。かつては英仏翻訳のような文章構造が一致する言語間の翻訳と比較して、日英翻訳のような文章構造が異なる言語間の機械翻訳は精度が低く、敬遠されていたが、近年ではこのような言語間でも翻訳精度が向上している。

世界148か国に210社の販売・サービス代理店を展開するグローバル企業であるコマツでは、母国語でない言語でのコミュニケーションの機会が日常的に発生する。そのため、タイムリーな情報の共有をサポートすることのできるAI翻訳ツールを日々のさまざまな業務のなかで効率よく利用する必要がある。

本章では、コマツが導入しているAI翻訳ツールの概要、導入事例および改善事例について報告する。

#### 3.1 AI翻訳ツールの概要

コマツではAI翻訳ツールとしてロゼッタ社のT-400（ティー・フォー・オー・オー）を導入している。2022年1月時点でコマツグループ内の51部門585人が利用している。過去1年間の翻訳文字数は約1180万ワードだった。

主な翻訳対象文章は以下のとおりである。

- ・ 製品情報等の社内の技術資料
- ・ メールなどのコミュニケーション情報
- ・ 社内システムの利用手順書
- ・ 社内の規則・ルール

一方で、製品カタログ、取扱説明書、ショップマニュアルなどの社外向けのセールス資料およびサービス資料の翻訳には、T-400を利用していない。

その理由として、過去に翻訳した文章を再利用することができる翻訳支援ツール（Computer Aided Translation tool）と併用することができない点、および社外ユーザ向けの資料のため、より細かなニュアンスを正確に翻訳する高い翻訳精度が求められる点がある。

具体的には以下のような文書の翻訳にはT-400を使用していない。

- ・ 取扱説明書等の社外向けマニュアル
- ・ 製品カタログなどの社外向けセールス文書
- ・ 車載モニタ文字列などの略語を多用する文書

#### 3.2 AI翻訳ツールの導入

AI翻訳ツールの導入プロセスについて紹介する。

AI翻訳ツールは選定、試行、公開の3ステップで導入した。

##### 3.2.1 AI翻訳ツールの選定

AI翻訳ツールの選定に先立ち、社内の翻訳ニーズを分析した。その結果、以下の4点のニーズがあることがわかった。

##### (1) 翻訳言語に関するニーズ

コマツではエンジンや油圧機器などの重要なコンポーネントの開発を国内で行っている。そのため、社外から社内への内向き（インバウンド）の情報および、社内から社外への外向き（アウトバウンド）の情報の双方において、日本語を介したコミュニケーションを必要としている。基軸言語となる日本語の翻訳精度が高いツールへのニーズが高いことがわかった。

##### (2) 翻訳対象データに関するニーズ

製品の開発進捗状況を管理するスケジュールや発生した課題をリスト形式で管理するために、エクセル形式のデータを多く利用していることがわかった。

これらのエクセル形式のファイルには、データを補足することを目的としてテキストボックスを利用したり、複数のシートに分けて情報を入力しているために、1ファイルの文字数が10,000ワードを超えるファイルもある。

そのため、エクセル形式のファイルに含まれるコメントやテキストボックスの文字を漏れなく抽出し、かつレイアウトを崩さないこと、および処理可能な文字数やファイル容量が大きいことを条件とした。

### (3) 翻訳の利用および運用コストに関するニーズ

AI翻訳ツールは多数の社員が利用する可能性のあるツールである。そのため、長期的に安定して使用できる、すなわちメンテナンスコストが低いことも考慮した。

OSのバージョンアップに伴って追加開発が必要になる可能性のある社内独自のツールではなく、市場で購入可能なAI翻訳ツールを検討した。

また、日々のさまざまな業務において躊躇なく利用できるように、利用コストが低いことも重要な要素と考えた。

### (4) 翻訳対象データのセキュリティに対するニーズ

開発中の製品に関する秘匿性の高い社外秘文章を翻訳する機会が多いため、情報漏洩のリスクがない、セキュリティの高いツールへのニーズがあった。

これら4つのニーズを満たすAI翻訳ツールを求めて、市場で利用可能なツールを比較検討した結果、ロゼッタ社の提供するAI翻訳ツールT-400を導入することを決定した。

#### 3.2.2 AI翻訳ツールのカスタマイズ

AI翻訳ツールを試行するに当たり、社内および建設機械分野に特有の専門用語を翻訳結果に反映させる、翻訳エンジンのカスタマイズおよびコマツ用語辞書の登録を実施した。

#### (1) 翻訳エンジンのカスタマイズ

T-400はユーザの翻訳対象データに合わせた2,000種類の翻訳エンジンを持っている。この既存の翻訳エンジンのなかでコマツの事業領域に最も近い「機械」分野の翻訳エンジンをベースに、コマツの取扱説明書およびショップマニュアルの翻訳データを読み込ませるカスタマイズを実施した。カスタマイズ後の翻訳エンジンの性能を評価するために、コマツの主要な製品であるPC200系の取扱説明書の文章から抽出した100の文章をT-400で日本語から英語に翻訳した結果を評価した。

評価方法は、100文の英訳の内容の正確性を以下の0～2点で人的に評価する手法を採用した。

0点：理解できない

1点：言語的な違和感はあるが、意図は分かる

2点：言語的な違和感が無く、意味が通じる

100文の平均は図7に示すとおり、1.76点となった。

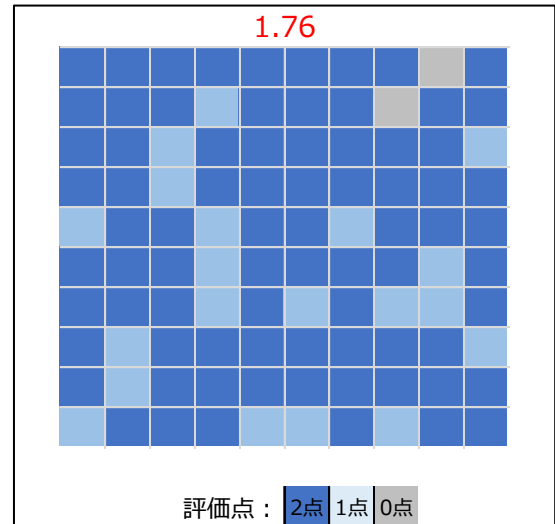


図7 100文評価のヒートマップ

#### (2) コマツ用語辞書の登録

T-400ではユーザが作成した用語辞書を登録し、翻訳結果に反映することができる。コマツでは、前項で紹介したコマツ用語辞書をT-400に登録することで、社内の専門用語を翻訳結果に確実に織り込めるようにした。

#### 3.2.3 AI翻訳ツールの公開

AI翻訳ツールは開発部門および要望があった部署に対して、先行して導入している。

導入部門における、工数削減効果、外注費削減効果、および業務効率改善効果は、現在検証中である。

効果が実証出来次第、コマツグループ全体へ導入を拡大していく。

## 4. 今後の課題

今後の課題として、以下の3点がある。

- ・ 用語辞書の効率的な運用ワークフローの構築
- ・ サービス資料の翻訳におけるAI翻訳ツールの導入
- ・ AI翻訳ツールの全社への展開

#### 4.1 用語辞書の効率的な運用ワークフローの構築

用語辞書に完成はない。新しい機能や部品の増加は、新しい用語を必要とする。正しい用語を迅速に追加するための運用ワークフローを構築しなければ、必要な用語が登録されていない信頼性の低い用語辞書になってしまう。

正しい用語を迅速に追加するためには、2.3のワークフローのように、開発部門、知的財産情報の管理部門、資料制作部門、各海外現地法人を巻き込んだ、コマツグループ全体で協力して用語辞書をメンテナンスできる運用ワークフローを構築する必要がある。

#### 4.2 サービス資料の翻訳におけるAI翻訳ツールの導入検討

3.1に記載したとおり、サービス資料の翻訳には現在T-400を使用していない。

その理由は繰り返しの表現が多く、文章の再利用が多いサービス資料の翻訳で高い効果を出している翻訳支援ツール(Computer Aided Translation tool)とT-400を同時に使うことができない点がある。

しかし、本稿で繰り返し述べているとおり、大量の文章を短期間で翻訳する社内のニーズは日々増え続けており、サービス資料の翻訳もまたスピードアップが必要である。その対策として、サービス資料の翻訳に利用できるAI翻訳ツールの導入を検討する必要がある。

具体的には翻訳支援ツール上で、人手翻訳の代わりにAI翻訳を使用し、AI翻訳の結果を人的に手直し(ポストエディット)することのできるツールの導入を検討している。

#### 4.3 AI翻訳ツールの全社への展開

3.2で述べたとおり、AI翻訳ツールは対象を限定して、導入効果を検証している。導入効果が確認出来次第、コマツ全社への展開を進めていきたいと考えている。

### 5. おわりに

コマツ用語辞書の制定とAI翻訳ツールの導入はまだ始まったばかりで、4章に挙げたような解決すべき課題がある。今後も国内海外各部門の協力を仰ぎながら、引き続き改善を重ねていきたい。

#### 謝辞

コマツ用語辞書の制定にご協力いただいたAI翻訳プロジェクトチームおよび国内外の用語レビュー担当者、またT-400に関する情報の提供および本レポートのレビューをしていただいた株式会社ロゼッタの皆さんに、この場をお借りして深く御礼を申しあげます。

#### 参考文献

- 【1】 “土工機械-油圧ショベル- 第1部：用語及び仕様項目”，JIS A8403-1-1996
- 【2】 松村幸治，“プロフェッショナルのマニュアル翻訳”，2014年，p.60-86

#### 筆者紹介



Shinya Tsukahara

塚原伸也 2007年、コマツゼネラルサービス入社。  
開発本部 資料技術開発センタ所属

#### 【筆者からひと言】

コロナ禍でface to faceのコミュニケーションが難しくなっている。このような状況で、共通の理解のもとで情報交換をするために、統一された用語の利用と言語の壁を超えた資料の制作はこれまで以上に重要になってきている。本稿で紹介したコマツ用語辞書の制定とAI翻訳ツールの導入がこのようなコミュニケーションの障害を解決する一助になるように、これからも改善を続けていきたい。