

# 自動翻訳システム開発に向けた用語調査

新池 昌弘\*

丹後産地機業向けの翻訳システムの構築のために、翻訳ソフトでの翻訳精度を調査したところ、約 4 割の文章に違和感があった。また、収集した文章の中で翻訳結果に違和感を持った用語数は約 2 割であり、その内の約 8 割が専門用語集に掲載されている用語であった。

## 1 はじめに

2020 年 3 月に総務省が作成した「グローバルコミュニケーション計画 2025」によると、日本の多言語翻訳技術は 2030 年までにシビアな交渉にも使える通訳システムの開発を目指している。

丹後産地でも海外バイヤーとのオンライン商談会や個別の商談を通じて販路拡大を行っている中で、言葉の壁がシビアな商談を行う際の課題となっている。

新型コロナウイルス感染症の影響で国内はもちろん海外でも対面での商談の機会が減少した。急速に進むオンライン化の中でも海外のバイヤーと商談をする際には通訳者を雇っており、この通訳者の確保と費用がオンライン商談での課題の一つと考えた。

そのため、多言語同時通訳システムのソフト開発を行う凸版印刷(株)の協力のもと、現在の翻訳システムで丹後産地の機業が商談を行った場合にスムーズに翻訳ソフトを使用することができるのか検証をすることとした。

## 2 調査

### 2.1 商談用語の収集方法

専門用語の使用状況や商談時に使用する用語の収集を目的として、丹後の機業 13 社に協力いただき以下の方法で用語の収集を行った。

- ・オンライン商談会後の聞き取り
- ・百貨店での商談時に同席
- ・テキスタイルの展示会で各社のブースでの聞き取り調査と商談時に同席

### 2.2 翻訳精度の調査方法

凸版印刷(株)の音声翻訳アプリを用いて翻訳精度の確認を行った。このソフトは、ウェブブラウザ上で起動させオンラインで使用するものであり、ボタンを押して音声入力または文字入力を行うことで単文を翻訳することができる。まず発語した言葉を日本語の文章に変換し、変換された日本語を英訳し英文を表示させる。さらに、その英文を和訳した文章を表示する仕組みである。

この仕組みを利用して翻訳結果に違和感を持った例文や用語を翻訳不可と判断することとした。

## 3 調査結果

400 例文を収集し翻訳したところ、約 39 % (156) の例文において違和感がある翻訳結果となった。違和感を持った例文は、①用語を言い換えることで正しく翻訳できるものと、②そもそも翻訳ソフト内に用語の登録がないことで翻訳できないものがあった。

<①の例>

以下に用語の言い換えで翻訳可能となった例を示す。

「このきじのそざいはなんですか。」と発語した場合

日本語の認識:この記事の素材はなんですか。(×)

※例で意図した用語は、記事ではなく正しくは生地

英訳:What is this article made of? (×)

和訳:この品物は何でできていますか。(×)

「このぬののそざいはなんですか。」と言い換えた場合

日本語の認識:この布の素材はなんですか。(○)

英訳結果:What is this cloth made of? (○)

和訳結果:この布は何でできていますか。(○)

\* 技術支援課 副主査

(2021 年度研究課題)

他にも「当社の機(はた)は小幅です。」を変換すると  
 英訳: Our flight is small. (×)  
 和訳: 私たちのフライトは狭いです。 (×)  
 「当社の織機は小幅です。」と言い換えた場合  
 英訳: Our weaving machines are narrow. (○)  
 和訳: うちの織機は幅が狭いです。 (○)

<②の例>

次に翻訳ソフト内に用語の登録がないことで翻訳できない例を示す。

「当社で紋紙(もんがみ)を作成します。」を翻訳ソフトで変換した場合

英訳: We will make Monshi. (×)  
 和訳: もんしを作ります。 (×)

上記では、紋紙(card, Jacquard card)の用語が登録されていないと判断した。

他にも「これは後染めです。」と入力すると

英訳: This is a garment dye. (×)  
 和訳: これは衣類の染料です。 (×)

上記では、後染め(fabric dyeing)の用語が登録されていないと判断した。

翻訳ソフトのデータベース内に用語の登録がないことで翻訳できないと推測された例文があったため、用語に特化して調査を行った結果を図1に示す。

収集した例文から重複分を除いてカウントしたところ用語総数は 474 であり、その中で翻訳文に違和感を持った用語は 24%(115)であった。

本調査を開始した際には、商談時に専門用語は使っていないとのことであったが、翻訳されない用語数 115 の約 8 割がテキスタイル用語辞典などの専門用語集に掲載されている用語であった。

また、用語が登録されていても、会話の中で商品の売れ行きについて「シルクの方が合成繊維よりも動いている」というものは、「Silk is more mobile than synthetic fiber. (和訳: 絹は合成繊維よりも動きやすいです。)」となり、本来伝えなかった「シルクの方が合成繊維よりも売れている(英訳: Silk sells better than synthetic fiber.)」とはならなかった。

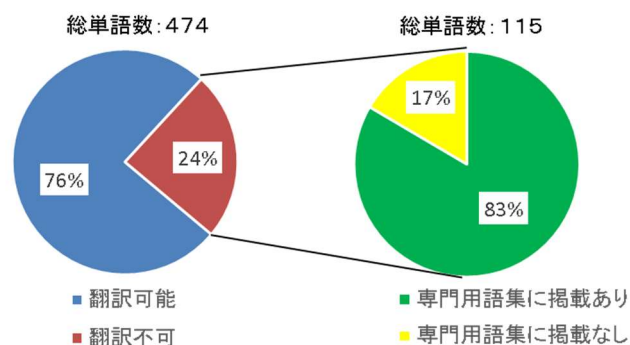


図1 例文に使用された用語の調査結果

翻訳不可と判断した用語の例を以下に記載する。

- 専門用語集に掲載ありとした用語  
紋紙、白生地、後染め、糸染め、先染、シボ、スワッチ
- 専門用語集に掲載なしとした用語  
着分、紋、紋意匠、中、シャリ感、丁(杼の数)、手袖

#### 4 まとめ・考察

本調査を通じて、翻訳ソフトに登録されていない語句の追加で翻訳精度を上げることが可能となることがわかった。

さらに翻訳精度を上げるためには、テキスタイル用語辞典などの専門用語集や日本産業規格に規定されている一般織物用語に掲載がない 2 割の用語の取り扱いについて検討が必要となる。

また、今回は商談時にボイスレコーダーを用いて録音したり、Zoom 等でのオンライン商談時に録画をする許可を得られなかったため、用語の収集を聞き取りによって行った。この方法では手間と時間がかかるため、情報が秘匿され、かつ商談者が商談に集中できるようボタンを押すなどの作業を必要としない商談用語を収集する仕組みが必要となる。

#### 謝辞

末筆ではございますが、本調査にご協力いただきました機業の皆さまに御礼申し上げます。