

ニューラル機械翻訳の商用利用に関する一考察 ～翻訳会社における特許翻訳での事例紹介～

渡部 孝明 山本 真佑花
株式会社翻訳センター

1 はじめに

ニューラル機械翻訳（以下、NMTと称する）は、産業翻訳業界に大きな影響を与えた。翻訳会社は挙って導入を推し進め、生産性の向上を図ろうとしている。一方、クライアント側も、納期短縮やコスト削減等、NMTに対する期待感が高い。

近年の人工知能（AI）技術の発展が、翻訳業界のみならず、様々な分野に多大なる恩恵をもたらしていることは周知の通りである。

こうした中、NMTに関して、その理解度の差から、クライアント、翻訳会社、翻訳者の三者間にある種の相違が生じてきた。

たとえば、コスト面では、NMTの出力結果を使えば、翻訳にかかる労力が減るため、コスト削減に繋がるという見解がある一方、現状のNMTの出力結果では、一から翻訳した方が効率的で、結果、コスト削減に寄与しないという見解もある。品質面もまた然りである。

2 翻訳業界における認知度

近年、翻訳者の間でも機械翻訳の関心が高まっており、日本翻訳者協会（JAT）が主催する『ijet』や、日本翻訳連盟（JTF）が主催する『翻訳祭』等でも、機械翻訳に関するセッションが増えてきた。

特許分野においても、日本知的財産協会（NIPTA）が、機械翻訳に研究会を立ち上げ、筆者もその一員として参加し、上記の『翻訳祭』等でその研究成果が発表されている。

[1]

3 翻訳作業工程

翻訳会社には、一般に、翻訳者が在勤しており、社内で翻訳物を生産するタイプと、登録されたフリーランスの翻訳者に翻訳物を外注委託するタイプとがある。弊社は後者の翻訳会社に該当する。弊社の場合、クライアントより翻訳業務を受注すると、登録されたフリーランスの翻訳者の中から適任者を選定し、依頼する。

依頼の際、NMTの出力結果を、原稿等とともに翻訳者に送付する場合がある。翻訳者は、NMTの出力結果を下地に翻訳（ポストエディット）をおこなうことがある。

弊社は、『特許』、『医薬』、『工業・ローカライゼーション』、『金融・法務』を主要4分野とする総合翻訳会社で、NMTに対する取組みも各分野様々である。

本稿では、特に、弊社の特許分野におけるNMTの利用例について紹介し、現在抱えている課題、および、その課題を解決するための手段、すなわち、弊社で取り組んでいる作業工程の一例について述べる。

なお、以下の説明は一例に過ぎず、弊社における翻訳作業工程はこれに限定されない。

4 課題

NMTについて、クライアント、翻訳会社、翻訳者の三者で最も関心が高いのが、出力結果の翻訳品質であろう。

以下、国際公開2015/174341号公報（出力日：2018年12月 使用エンジン：A社ニューラルエンジン）を例に挙げ、そのNMT出力結果（日本語→英語）を用いて、NMTによく見られる課題を示す。

なお、弊社では、本稿の共著者である山本による研究[2]において使用されたエラーカテゴリに基づいて独自のNMT用エラーカテゴリを再構築しており、エラーに関して、検出が容易か否か、ポストエディットが容易か否か、NMTに起因するものなのか否か、人手翻訳でも起き得るか否か、等で分類し、その一部を本課題として抽出した。

4.1 用語・表現

特許明細書には「～装置」、「～部」等の要素が数多く登場する。これらには決まった対訳がなく、また、訳出の仕方も複数あるので、不統一を招きやすい。

また、動詞、形容詞等も複数の意味を持つ場合は、不統一を招きやすい。

以下に例を挙げる。なお、カッコ内の数字は、出現回数を表す。

例1)

「パラメータ抽出部」(10)

“parameter extraction unit” (7)

“parameter executing unit” (2)

“parameter extractor” (1)

例2)

「エラー表示部」(3)

“error display unit” (2)

“error display section” (1)

例3)

「判断する」(28)

“determine” (23)

“judge” (4)

“detect” (1)

例4)

「右方に折れ曲がる線分」(11)

“segment bending to the right” (9)

“segment that is bent to the right” (2)

“segment bending rightward” (1)

4.2 符号

特許明細書では、発明の主題が装置等の場合、符号が頻出するが、数字とアルファベットを組み合わせたものもある。かかる場合、NMTの出力結果では、数字とアルファベットとの間に不要なスペースが介在することがある。

以下に例を挙げる。

例)

原文：「図1 2 A～図1 2 Gの例では、」

NMT：“In the example of FIGS. 12A to 12_G,”

（“12”と“G”の間のスペース）

4.3 未文

特許明細書は、一文が非常に長く、また複雑な係り受けをしている文章が多い。それ故、NMTが上手く翻訳できない場合もある。

以下に例を挙げる。

例1)

As described above, in the information processing apparatus 10, designating the type, position, size, number, interval, and the like of various objects to be generated by manually inputting simple broken lines as commands on the display screen 6 You can do.

例2)

The present invention has been made to solve the above problems, and it is an object of the present invention to easily and quickly designate the type of command instructing generation of an object and the parameters of a command by handwriting without using a command selection menu And to provide an information processing apparatus, a program,

上記例は明らかに文章になっていない。

4.4 翻訳漏れ

機械翻訳であっても翻訳漏れはあるようである。

以下に例を挙げる。

例1)

原文：上記課題を解決し上記目的を達成するために、本発明のうち第1の態様によるものは、情報処理装置であって、コマンド認識部と、パラメータ抽出部と、コマンド実行部と、を備える。

NMT：According to a first aspect of the present invention, there is provided an information processing apparatus including a command recognition unit, a parameter extraction unit, and a command execution unit.

上記例は、英文としては完璧であるが、なぜか、原文の前半部分が翻訳されていない。

例2)

原文：図3A～図3Cは、横書きテキストボックス生成コマンドの一例による表示画面の遷移を示す図である。

NMT：3A to 3C are diagrams showing transition of display screens according to an example of a horizontal writing text box generation command.

上記例では「図」が訳出されていない。本件ではかかる現象が数多く見られた。

4.5 初出／既出

日本語は英語と異なり、冠詞がないため、NMTの出力結果においても、不定冠詞であるべき箇所が定冠詞になっていたたり、その逆であったりすることがある。

特に【特許請求の範囲】においては、権利解釈にも大きく影響するため、非常に重要視する項目である。

4.6 単数／複数

同様に、単数／複数の概念も、日本語の場合は、「複数の」等と記載するか、文脈から判断するか、図面から判断しなければならないので、注意を要する。

上記4.5項同様、【特許請求の範囲】においては、権利解釈にも大きく影響するため、非常に重要視する項目である。

4.7 語順

日本語と英語との言語構造との違いから、NMTの出力結果には、主語と述語とが離れた翻訳文がいくつか見られる。

以下に例を挙げる。

原文：以下において、図3A～図3Cから図12A～図12Gを参照しつつ、手書き入力されるコマンドを構成する折れ線の形状、及びコマンドが指示する処理の内容について例示する。

NMT：In the following, with reference to FIG. 3A to FIG. 3C to FIG. 12A to FIG. 12G, the shape of polygonal line constituting the command to be handwritten input and the contents of processing instructed by the command will be exemplified.

4.8 流暢さ

NMTは、従来の機械翻訳と比べると翻訳文の流暢さが格段に向上していると考えられる。

一方で、出力結果が流暢なあまり、原文との照合だけでは誤訳に気付かない可能性もある。

以下に例を挙げる。

原文：座標入力装置1は、例えば赤外線や静電容量方式等の座標入力機構を有し、ペン、指、レーザーポインタ等により座標の手書き入力を可能にする装置である。

NMT：The coordinate input device 1 has a coordinate input mechanism such as an infrared ray or a capacitance type device and is a device enabling handwriting input of coordinates with a pen, a finger, a laser pointer or the like.

上記例では、“例えば～可能にする”までが「装置」係るので誤訳である。

4.9 特許翻訳実務

特許権は各国において取得する必要があるため、明細書も特許を取得しようとする国の法律に基づいて翻訳する必要がある。

特に、米国に出願するための明細書の翻訳においては、判例に基づく様々な決まり事がある。

以下に例を挙げる。

原文：図1は、本発明の実施形態に係る情報処理装置のハードウェア構成を示すブロック図である。

NMT：FIG. 1 is a block diagram showing a hardware configuration of an information processing apparatus according to an embodiment of the present invention.

本例は、翻訳上まったく問題はない。しかし、特許実務上は、たとえば、“showing”は“of”のほうが好ましいと言われることがある。また、場合によっては、“according to an embodiment of the present invention”は訳出しないうほうが好ましいと言われることがある。

5 解決手段（作業工程）

係る課題を解決するための手段として、以下に挙げる『プリエディット』、『プリポストエディット』、『ポストエディット』という作業工程を検討し、一部導入している。

そして、上記課題を、ある作業工程単体で解決するのではなく、分担して解決することで、品質および作業効率の向上が図れるのではと考える。

5.1 プリエディット

機械翻訳の出力結果を向上させる手段として、プリエディット[3]が知られている。

特許分野の場合、一般財団法人日本特許情報機構（JAPIO）が発行している『特許ライティングマニュアル』[4]等は、プリエディットをおこなう際の指針となり得る。しかしながら、翻訳会社がクライアントより預かった原

稿に対してプリエディットをおこなうことは困難ではないかと考える。

その最大の理由は、特許翻訳を専門とした翻訳会社は別として、翻訳会社側にプリエディットをおこなう人材が少ないことである。特許明細書は法律文書であり、かつ、技術文書である。記載の技術内容を理解せずに、プリエディットをおこなうと、却って誤訳を招く可能性がある。

また、プリエディットは、その効果の程を、NMTの出力結果を確認するまで把握することができず、出力結果の向上を図るには、繰り返しプリエディットをおこなう必要があり、現状の作業工程に組み込むには相当の試行錯誤を必要とすると考えられる。

一方、優れた翻訳者は、翻訳をおこなう際、頭の中で自然にプリエディットをおこない、長文を短文化したり、主語または目的語を補ったりしている。

したがって、NMTの出力結果が良い翻訳文（プリエディットの必要のない原文の翻訳文）はポストエディットをおこない、NMTの出力結果が悪い翻訳文（プリエディットの必要がある原文の翻訳文）は、ポストエディットをおこなうよりは、翻訳者が一から翻訳したほうが効率的ではないかと考える。

翻訳会社が、プリエディットの一環としてできることは、『特許ライティングマニュアル』等を参考に、プリエディットの必要がある原文に対して、翻訳者に何らかの注意喚起を与えることではないかと考える。

たとえば、原文にハイライトを付し、当該原文に対するNMTの出力結果が悪い可能性がある旨を翻訳者に伝えれば、翻訳者は、ポストエディットするか、一から翻訳するかの判断材料となり得るのではと考える。

また、ハイライトを付す作業は機械的におこなうため、専門的な知識は必要なく、翻訳会社側にかかる負荷は低い。

5.2 プリポストエディット

弊社のように、翻訳を、フリーランスの翻訳者に外注委託する場合、NMT出力結果の品質が、翻訳者の作業負荷に大きな影響を及ぼす。場合によっては、用語・表現の統一、未文の解消等に時間をとられ、本来翻訳者がすべき、技術的観点および特許実務的観点からのポストエディットができなくなる可能性がある。

かかる状況を鑑みて、フリーランスの翻訳者にポストエディットを依頼する前に、NMTの出力結果に応じて、翻訳会社側でプリポストエディットをおこなう必要があると考える。

なお、弊社でいうところのプリポストエディットとは、内容には直接関係のない軽微なエラーであるが、看過することができないエラーを解消する作業をいう。

プリポストエディットの作業項目の例としては以下の通りである。

- ・ 頻出用語の統一： 上記4.1項
- ・ 符号の泣き別れの解消： 上記4.2項
- ・ 未文の解消（一部）： 上記4.3項
- ・ 翻訳漏れの解消（一部）： 上記4.4項
- ・ 不要なスペースの削除
- ・ スペルミスの解消
- ・ 全角／半角処理
- ・ 原文に記載のない翻訳文の削除

5.3 ポストエディット

上記を鑑みるに、ポストエディットは、翻訳と同程度もしくはそれ以上の力量が求められる。

弊社には翻訳者以外に校正者も登録されているが、優れた翻訳者が優れた校正者になり得るとは言い切れない。ポストエディットも然りである。

外国出願用特許明細書に対するNMT出力結果をポストエディットする場合、ISO17001、ISO18587:2017[5]等に準じた技量が必要であることは当然である。

特に、上記4.8項および4.9項は熟練の特許翻訳者でなければできない非常に高度な作業ではないかと考える。

一方で、上記項目4.3項および4.4項等は、優れた翻訳者であっても気づかないことはよくある。

また、既に存在する翻訳文に対してさらに良くしようという心構えも必要である。残念ながら、一般的に翻訳単価よりもポストエディット単価のほうが低く、プロ意識の高い翻訳者でない限り、上記のような心構えを持てる動機付けにはなり得ないのが現状であろう。

6 翻訳支援ツールとの併用

基礎となる明細書およびその翻訳文が対となった翻訳メモリの言語資産としての再利用が求められる中、今や、翻訳支援ツールを用いての翻訳は必須となりつつある。

弊社を含む多くの翻訳会社が種々の翻訳支援ツールを導入しており、翻訳品質や作業効率の向上に寄与している。

翻訳支援ツールには様々な種類があり、その機能も様々である。

プリポストエディットおよびポストエディットをおこなうにあたって、翻訳支援ツールの機能を駆使することによって、大幅な作業効率の改善が見込めると考える。

以下、翻訳支援ツールを用いた各作業の手法の一例を説明する。なお、上記NMTの出力結果は、弊社で使用している翻訳支援ツールを用いて出力したものである。

6.1 プリポストエディット

プリポストエディットをおこなうにあたって有効と考えられる翻訳支援ツールの主な機能として、以下が挙げられる。

- ・ 原文および／または翻訳文のフィルタリング機能
- ・ 原文および／または翻訳文の検索機能
- ・ 対訳ファイルのエクスポート・インポート機能
- ・ 用語登録機能
- ・ 翻訳文の一括置換機能
- ・ スペルチェック機能
- ・ 対訳間の差異検出機能

係る機能を駆使することにより、上記5.2項にて列挙した項目を、フリーランスの翻訳者にポストエディットを依頼する前に、ある程度解消しておく。

上述の通り、プリポストエディットは、検出が容易であり、かつ、修正が容易であるエラーのみを解消することを目的としている。

たとえば、頻出用語を一括で統一させたとしても、その妥当性は判断せず、翻訳者に委ねる。

また、未訳、翻訳漏れ等も一見して発見できるものに限る。

それ故、翻訳会社側の負荷もある程度抑えつつ、翻訳者も、技術的観点および特許実務的観点からのポストエディ

ットに傾注することが可能になる。

6.2 ポストエディット

翻訳支援ツールの機能の一つに、翻訳メモリ機能がある。同一文書内で、一度翻訳されたものは都度メモリ化され、以降、同一または同様の原文が出現した際、当該メモリを流用して翻訳する機能である。

たとえば、特許明細書の場合、【特許請求の範囲】に記載の内容は【要約】、【発明を解決するための手段】、【発明を実施するための形態】に同様の記載があることが多いため、【発明を解決するための手段】欄で確定させた翻訳文を他の箇所に流用するほうが効率的である。

また、複数の実施形態が記載されている場合や、化学関係等の特許明細書で、実験が記載されている場合も、同様に、翻訳メモリの活用が効果的である。

本機能を用いることにより、同一または同様の原文であれば、NMTの出力結果をポストエディットするよりも、翻訳メモリを流用して修正したほうが効率的である。

ポストエディットの際の翻訳メモリの採用基準
NMT出力結果<翻訳メモリ（例：一致率75%以上）

さらに、翻訳文に対してフィルタリングをかけることにより、上記4.5項および4.6項も首尾良く解消することが可能である。

たとえば、特許明細書中、【発明を実施するための形態】欄において、ある要素が最初に出現した時点で、翻訳文に対してフィルタリングをかけ、当該要素の翻訳文の冠詞を適宜修正する。また、複数個ある要素についても、最初に出現した時点でフィルタリングをかけ、複数表現に修正し、個々の説明の箇所のみ適宜単数表現にすれば良い。

また、【特許請求の範囲】においても同様の操作をおこなうことで、上記4.5項および4.6項について効率的に解消することが可能である。

なお、当然のことではあるが、上記6.1項にて列挙した翻訳支援ツールの各機能が、ポストエディットにおいても有効であることは言うまでもない。

7 おわりに

上述の通り、現時点では、NMTを正しく活用するには、プリエディット（クライアント）、プリポストエディット（翻訳会社）、ポストエディット（翻訳者）という作業分担が理想ではないかと考える。

弊社では、現在利用可能なNMT出力結果の定期的な品質評価をおこなっている。評価の際、常に思うのは、NMT技術はまさに日進月歩で、同一のエンジンであっても以前の出力結果より明らかに品質が向上しているものもある。それ故、現時点での取組みも無用になる可能性は十分にある。NMT技術の向上に伴い、近い将来、ポストエディットが主流となり、翻訳に取って代わる作業となるであろう。

筆者を含む翻訳者が抱える現在の不安は、機械翻訳の出現で人間の翻訳者が不要になるのではということではないかと思う。確かに、NMT技術の向上に伴い、生産性が上がり、技量の低い翻訳者は淘汰されると考えられる。

しかし、特許明細書の翻訳は、単なる翻訳の域を超えた外国出願用特許明細書の作成であると考えられる。すなわち、翻訳文として適切である、たとえば、翻訳漏れがない、文法的に正しい、等の最低限の要件のみならず、審査を経て、権利化され、権利行使の際に、出願人の意図した権利範囲

が守られた明細書となっているかが重要なのである。ここに、NMTと、翻訳者との作業分担が見られるように思う。

我々翻訳会社も市場の縮小が予想される中、生き残りを図るには、NMT等の最新の技術を上手く活用し、クライアントの要望を超えるサービスを提供できるようにしていく必要がある。

なお、現時点では、上記手法による作業工程は、弊社の通常の作業工程として確立しておらず、試行錯誤の段階である。特に、プリポストエディットに伴う工数の増加や、人材の確保等の課題を認識している。今後、有識者の方々の助言等を拝借しながら進めていきたい。

参考文献

- [1] 日本翻訳連盟『翻訳ジャーナル』No.298 2018年11/12月号
- [2] 山本真佑花・田辺希久子・藤田篤（2016）。「翻訳学習者の学習過程におけるエラーの傾向の変化」.言語処理学会第22回年次大会発表論文集, pages 865-868.
- [3] 宮田玲 藤田篤（2017）機械翻訳向けプリエディットの有効性と多様性の調査『通訳翻訳研究への招待』No.18
- [4] 一般財団法人日本特許情報機構『特許ライティングマニュアル』（第2版）
- [5] ISO 18587:2017 <https://www.iso.org/standard/62970.html>