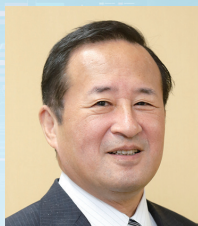


ユーザーから見たNMTの使い勝手と活用の展望

Users review on NMT engines and perspective toward effective use



弁理士。日本弁理士会元会長。久遠特許事務所代表。早稲田大学工学部電気工学科卒、シカゴ大学大学院化学科博士課程修了 Ph.D. AIPPI 本部の副会長。日本知的財産翻訳協会理事長。2013年から2017年まで内閣府知的財産戦略推進本部 有識者本部員。

特定非営利法人日本知的財産翻訳協会 理事長

奥山 尚一

✉ okuyama@quon-ip.jp

TEL 03-3508-9577

1 はじめに

日本知的財産翻訳協会（以下、その英文略称である「NIPTA」という。）は、知的財産翻訳検定を実施する傍ら、これまでも機械翻訳に関係してきた。例えば、国立研究所である情報通信研究機構（NICT）の手伝いで、機械翻訳の評価をしたこともある。

そういった中で2016年11月に事件が起きた。Google 翻訳がニューラル機械翻訳（NMT）になった。その当時の騒ぎは皆さんもご存じのとおりである。

そこで日本翻訳連盟（JTF）の日本翻訳ジャーナル編集委員長の河野弘毅氏の協力を得て、2017年10月に「NIPTA 特許機械翻訳研究会」を立ち上げた。このとき注意しようと考えたのは、次の3点である。

1. 機械翻訳は使えるのかを翻訳者の立場から確認して、もしそうならば最もよい使い方をみんなで考える。
2. 翻訳ソフトの品評会にはしない。NMTの性能はどんどん上がっていくはずなので、現時点での性能評価をしても、半年一年後には研究結果が古くなってしまいうからである。
3. できるだけオープンな集まりにする。NIPTAだけではなく、日本翻訳連盟（JTF）や日本翻訳者協会（JAT）に所属されている翻訳者の方々、ソフトウェア開発をされている方、さらには研究者にもおいでいただいで、いろいろな議論が出来る場にしようと考え

た。

18名のメンバーのほか、ありがたいことに、NICT、日本特許情報機構（Japio）、科学技術振興機構（JST）などの研究者の方々にもゲストとして多数おいでいただき、機械翻訳の最先端の事情をうかがうことが出来た。

この研究会では、まずどういう基準で評価すべきかを検討した。そして、NIPTAは、2004年より知的財産翻訳検定を実施しており、その過去問と標準解答を公開しているが、その中から適当と思われる過去問を選んで3種の代表的なNMTエンジンに入力した。したがって、各エンジンは2018年4月の時点のものだが、本稿では最近の検討も織り込んである。

隔月に開催されたミーティングにおいて、メンバー全員でユーザーの視点から翻訳結果を検討した。分野は、電気、機械、化学、知的財産法務実務の4分野に限定した。また、このミーティングでは、毎回、開発者や研究者の方々との意見交換を実施することができた。以下にその結果の概要を紹介する。なお、上述のように翻訳エンジンの品評会にはしないという趣旨で、どのエンジンを使ったかは触れないが、当時アクセスできた最善と考えられる翻訳エンジンを使用した。

2 評価項目

評価項目は次のとおりである。

- [1] 重要語句の訳抜け (例: クレーム構成部材)
- [2] 軽微な語句の訳抜け (例: 符番抜け)
- [3] 係り受けの間違い (修飾句、入れ子構文等)
- [4] 係り受けの間違い (形容詞、副詞等)
- [5] 「係り受け」以外の誤訳・不適切訳 (例: ~以上・以下の取り違い)
- [6] 専門用語・重要語句の間違い
- [7] 軽微な語句の間違い (例: 符番違い、前置詞の使い方)
- [8] 翻訳 (MT 出力) に起因するもの
- [9] 原文に起因するもの
- [10] 文章になっていない、文章が破綻している
- [11] 主語の取り違い
- [12] 用語のバラつき
- [13] フローティング (原文との脈略がない語句が急に出てくる)

表計算ソフトに原文と翻訳、標準解答、そしてこれらの項目についての評価の一覧を作成していった、弱みと強みを検討していった。これを英和、和英の両方について行った。

3 化学分野¹

[気になる点1: クレームの翻訳が苦手] (特に和英)

下記は、クレームの英和 NMT 翻訳の一例である。

物品の第 1 の領域が実質的に第 1 のポリマー性の水膨潤性材料を含み、

第 1 の領域に隣接する第 2 の領域が第 2 のポリマー性水膨潤性材料を含むマルチポリマーヒドロゲル物品であって、

第 2 のポリマー材料を実質的に含み、

第 2 のポリマー材料は、第 1 の領域から第 2 のポリマー材料を通して移動する漸増する濃度勾配を示す。

第 2 の領域から第 3 の領域に移動する。

これは一番できの悪いものであるが、クレームの形式にはなっていない。原文を示すまでもなく、相当荒れた翻訳であることはお判りいただけよう。原文によれば、「マルチポリマーヒドロゲル物品」は第 1 のポリマー性水膨潤材料と、第 2 のポリマー性材料とを含んでいて、異なる領域での特性が異なっているものである。

1 メンバー (敬称略): 平林千春、梶木正紀、上野哲也、柴田知恵、岡本光弘

[気になる点2: ピリオドを含む略語が苦手]

いずれのエンジンも、3 ml of conc. HCl はうまく処理できない。sat. や Pat. も同様である。

[気になる点3: 単複の見分けが苦手]

例えば、次の文章をエンジンにかけても、温度条件が3つあるので、サンプルが複数必要であることは正しく反映されない。

長さ 100mm の直線をそれぞれ MD 方向および TD 方向に対して平行に試験片 (フィルム: 15mm × 120mm) 上に描いた。この試験片を標準状態 (温度 23℃ × 湿度 50%) に 2 時間放置し、その後試験前の直線の長さを測定した。続いて、150℃ または 180℃ または 200℃ の雰囲気 に設定した熱風循環式オープン内で一角を支持した状態にて 30 分間放置した後、取り出して標準状態に 2 時間放置冷却した。その後各方向の直線の長さを測定し、...

[気になる点4: 用語の統一が苦手]

さほど長くない翻訳課題の中で、試験片が test piece や specimen になり、熱収縮率が thermal shrinkage や heat shrinkage ratio になる。

[気になる点5: 原文との脈略がない語句が唐突に出てくる]

いわゆるフローティングといわれる現象であるが、原文に全くない語句が出てくる NMT に特有の現象である。

[気になる点6: 化合物名の誤訳]

例えば、polyvinyl chloride を塩化ポリビニルにしてしまうといった例があった。

[まとめ]

訳文自体は読みやすい場合が多い。仕事でそのまま使えるレベルではないが、全体的には上手く訳せている。ときには驚くほど上手である。その分、十分な専門的知識があって、翻訳スキルが高くないと、NMT のエラーを見落としがちかもしれない。

4 機械分野²

まず、NMT が得意な点としては、概して、訳文の流暢さがある。例えば次の英文は、” Having two stems

2 メンバー (敬称略): 湯浅豊裕、浜口宗武、新田順也、鈴木章浩、奥山尚一

11 enables easier access if the component 1 is located in a hard-to-reach place or is placed close to a wall.” は、「2つのステム 11 を有することにより、コンポーネント 1 が届きにくい場所に配置されている場合、または壁の近くに配置されている場合にアクセスが容易になる。」と訳出される。これで完璧というわけではなからうが、マニュアル翻訳の際のタイピングが省略でき、ポストエディットの手間は相当軽減されよう。

不得意なのは次のような点である。まず、訳語の揺れ。噴射ノズルが spray nozzle と injection nozzle になってしまっている。“In the process of repeating the vapor deposition process, the chamber 20 and the spray nozzle 10 are exposed to repeated loads in which expansion and contraction are repeated by repeated heat, and (中略) there is a possibility that a minute gap is formed between the injection nozzle 10 and the outer wall 21 of the chamber.”

次に、脈絡のない言葉が出てくること。” As will be appreciated from the above, turning one of the stems 11 will cause the ball 7 to rotate between a closed position, where the bore 9 is perpendicular to the main bore 4, and an open position, where the bore communicates with the main bore.” を「上記から理解されるように、ステム 11 の 1 つを回転させることにより、ボア 9 が主ボア 4 に対して垂直である閉位置と、ボアが主ボア 4 に連通する開位置との間でボール 7 が回転する。ボア。」となる。下線をした「ボア」はどこから湧いたのか。もっと長いフレーズが現れることもある。

さらに、訳抜け。” The invention will now be described, by way of example, with reference to the accompanying drawings, in which: Figure 2a is a side-sectional view of a valve component in accordance with a first embodiment of the invention; Figure 2b is a side perspective view of the valve component shown in Figure 2a.” が、「本発明を、添付の図面を参照して例として以下に説明する。本発明の第 1 の実施形態による弁構成要素の側断面図である。図 2b は、図 2a に示される弁構成要素の

側面斜視図である。」Figure 2a に対応する訳語が消えている。これも唐突感があり、注意しないとポストエディットの際に見落としそうになる。

そして、意味は掴めない。文章を理解しているわけではないので、字面を追うだけになる。「ゴルフクラブ（以下クラブ）には飛んで曲がらない（曲がりにくい）ことが最も要求されている。」は、“It is most demanded that a golf club (hereinafter referred to as a club) does not bend by bending (hardly bending).”

これらの弱点にかんがみて一つ提案できるのは、CAT（コンピューター支援翻訳）ツールの利用である。CAT ツールを使って長い文章を読点でぶつぶつ適当な長さに切り、機械翻訳にかけ、それをパズルのように組み立てるのである。

5 電気分野³

まず、実用性評価と分析結果のまとめからはじめる。

[A 社のエンジン]

英日：同じ翻訳文が繰り返し出現する現象が起こった。

日英：未訳（日本語が残る）があった。

[B 社のエンジン]

英日：「です・ます」調の翻訳文があった。

日英：並列関係が不適切な翻訳文があった。

[C 社のエンジン]

英日：単語の意味としては正しいが、翻訳としては不適切な箇所があった。

日英：文章として成立していない翻訳文があった。

比較の結果、それぞれ傾向というか癖はあるものの、実用性の点からみて各社の出力結果に大きな差は見られなかった。原文の複雑さに依拠するにせよ、最終成果物としてはポストエディットは必須であり、その作業は項目化してチェックリスト方式で進めるのがいいのではないかと考える。

翻訳ツールの開発者への提言としては、次の 2 点がある。

(1) 湧き出し・フローティング・脱訳について、注記・警告らしき表示が出るとポストエディット時に便利であ

3 メンバー（敬称略）：菊地公一、田村和彦、宮本伸也、渡部孝明

ろう。

(2) 用語登録オプションが必要である。登録する用語に関して品詞指定が出来れば便利であろうが、かえってNMTの品質低下を招くことは防がなければならない。例文として、「送風機は、ファンの回転数を増減させるインバータを備える。」を英訳するとする。用語登録がなければ、“The blower includes an inverter for increasing or decreasing the number of revolutions of the fan.” となろう。これが、「増減」を“variation”と訳語登録できれば、“The blower includes inverters for variation of the rotational speed of the fan.” に一発でなる。

ユーザ側で対応可能な改善策としては、次のような点がある。

(1) ネスト構造文を避ける。つまり、機械翻訳に不向きな日本語をリエディットすることがあろう。例えば、修飾句の挿入は1つに限定するといったことである。

(2) 文理解釈の容易な日本語にリエディットする。Japioが発行している「特許ライティングマニュアル」の第2版は大変よくできており、それを参考にリエディットを試みたり、「色 de チェック」といった翻訳支援ソフトを利用することも考えられる。

(3) 用語のバラつき等を抑える。そのためには、Memsourcesといった翻訳支援ツールの利用が考えられる。

6 知財法務実務分野⁴

裁判例等の日本語テキストを素材として翻訳結果の分析を行ったが、このようなテキストは個々の文章が長く、主述の関係が入り組んでいるなど複雑な構造を持っていて、実用的な結果は得られないことが明らかになった。そこで、機械翻訳処理後のチェックの労力を避けるためのリエディットの効果を確認した。

リエディットを行う際の指針としては、上述の「特許ライティングマニュアル」第2版が提唱する、「短文にする」、「省略しない」、「読点を工夫する」などの7つのポイントを採用した。

試行対象の文例は次のようなものである。

4 メンバー（敬称略）：宮城三次、河目志津加、河野弘毅、柴田純一郎

「原告は、各基礎出願明細書の図2から、螺旋溝27の傾斜角度が『約20度』であることが読み取れ、当該『約20度』を中央値として『±10度』とすることで、『10度から30度』の範囲が設定できることは、図面および明細書の記載から当業者にとって自明な事項である旨主張する。」

これを先の指針に従って次のように処理した。

まず、入れ子構成を下記A～Cのように解消する。

A－原告は、(B, C) (である) 旨主張する。

B－(C) (である) ことは、図面および明細書の記載から当業者にとって自明な事項である。

C－各基礎出願明細書の図2から、螺旋溝27の傾斜角度が「約20度」であることが読み取れ、当該「約20度」を中央値として「±10度」とすることで、「10度から30度」の範囲が設定できる。

そして、Cを分割するとともに、「中央値として『±10度』とする」という表現を言い換える。

このようなリエディットにより以下のようなテキストが得られる。

「原告は、下記を主張する。

以下のことは、図面および明細書の記載から当業者にとって自明な事項である。各基礎出願明細書の図2から、螺旋溝27の傾斜角度が『約20度』であることが読み取れる。したがって、当該『約20度』を中央値として『±10度』とすることで、『10度から30度』の範囲が設定できる。」

ちなみに、機械翻訳にかけながら、どのようなリエディットが有効であるのか試行錯誤することも容易である。

7 まとめ

以上みてきたように、一定の品質を維持するためにポストエディットは必須であって、機械翻訳に任せきることとはできない。一方、NMTは使っていかなければいけないツールになっているのが現実であろう。そして、今回の研究で、リエディットの重要性も確認された。

本研究会においてもう一点明らかになったのは、機械翻訳エンジン自体は一気にコモディティー化してしまったということである。今回、当然トップを走る翻訳エンジンを選んだ結果ではあるが、若干の優劣はあってもい

すれの翻訳エンジンも「使える」という結果になった。例えば、ハーバード大の OpenNMT というオープンソースのプラットフォームも提供されるようになっていく。ディープラーニングに必要な GPU などの高度な計算能力へのアクセスもお小遣いレベルで得られるようになっていく。そうすると、コーパスさえあれば、誰でもというわけではないが、統計的翻訳エンジンの複雑さからは考えられないような容易さで翻訳エンジンが完成してしまう。現在、関心は翻訳エンジン自体の進化に加えて、プリエディットとポストエディットを支えるプラットフォームの構築に向かっている。

このような急速な進歩は、翻訳者にとって果たして脅威なのであろうか。まっとうな疑問である一方、これに答えることは意味がなかろう。「脅威」ととらえたところで機械翻訳の進歩は止まらない。むしろ、テクノロジーの進歩は好機と捉えるべきである。タイピングという労働から解放され、生産性をより向上させることができると考えるべきである。

「日経エレクトロニクス」誌の9月号の特集は、「AI 翻訳が人間を超え 言葉の壁崩壊へ」というセンセーショナルなものであるが、日本語という地球上で比肩するものがない優れた言語圏において、翻訳文化を営々と築いてきた日本人にとって、言語のバリアーは日常レベルで低くなりこそすれ、AI がそのまま人間を置き換えるとは到底考えられない。

これからの展開を注視していきたい。

(本稿は上記の4分科会の報告をもとにまとめた。文責は奥山にある。)



特集

特許情報分野におけるAI活用のススメ