

日英ニューラル機械翻訳の定点観測、および、 翻訳者がおこなうプリエディットの有効な手段

新田 順也
エヌ・アイ・ティー株式会社

渡部 孝明
株式会社翻訳センター

1 はじめに

筆者らが所属する日本知的財産翻訳協会 (Nippon Intellectual Property Translation Association: 以下、NIPTAと称する) は、日本における知的財産 (特許) 翻訳力の拡充を目的として2004年に設立された特定非営利活動法人である。

NIPTAの主な活動として、

1. 知的財産翻訳検定試験の実施
2. 月刊「知的財産翻訳ジャーナル」の発行
3. 各種研究活動 (特許機械翻訳研究会)
4. セミナーの企画・実施

などが挙げられる。

ニューラル機械翻訳 (以下、NMTと称する) の出現にともない、特許翻訳業界でもその動向が注目される中、NIPTAは2017年に特許機械翻訳研究会を立ち上げ、NMTの出力結果について様々な研究をおこなってきた。

その研究活動の内容は、日本翻訳連盟 (JTF) が主催する『翻訳祭』など様々なイベントで発表している。

結論を先に述べると、現時点のNMT出力結果では要求される品質基準を満たしていない。それゆえポストエディットを行う必要があるが、ポストエディットを有効化するために翻訳者自身がプリエディットを併用することにより、翻訳作業全体の効率化が期待できる。

本稿では、その具体的な事例の一部を紹介する。

2 NMT出力結果の定点観測

2.1 調査目的

特許翻訳分野における現在の機械翻訳エンジンの性能変化の把握、および、機械翻訳出力結果の利用方法の提案を目的として調査をおこなった。

2.2 調査方法

現在一般に利用可能な3種の機械翻訳エンジンに同一文書を入力し、2018年度の出力結果と、2019年度の出力結果とを比較し、どのような変化が見られたか評価した。

具体的には、NIPTAの知的財産翻訳検定試験の過去問題 (第24回問1、第24回問3、第20回問3) より選択した28個の語句 (特許請求の範囲) および文章 (明細書) を3種の機械翻訳エンジンに入力し、84個の出力結果を評価した。

評価者は4名で、2018年度の出力結果と2019年度の出力結果とを比較して4段階評価 (改善、改悪、ほぼ変化なし、まったく変化なし) をおこなった。

2.3 結果報告 (電気・電子工学分野)

改善:	24%
改悪:	20%
ほぼ変化なし:	48%
まったく変化なし:	8%

2.3.1 改善 (例)

原文

「正極性期間と逆極性期間とを算出し、前記溶接制御御部へ出力する算出部と、」

2018年度出力結果

“A computing unit, comprising: a calculating unit configured to calculate a positive polarity period and a negative polarity period; and a calculating unit configured to output the positive polarity period to the welding controller,”

2019年度出力結果

“A calculation unit configured to calculate a positive polarity period and a reverse polarity period, and to output the positive polarity period and the reverse polarity period to the welding control unit”

ポイント

- “A computing unit, comprising:” の原文にない語句の出現が解消されている
- “and a calculation unit configured to” の不要な繰り返し解消されている
- “to output the positive polarity period and the reverse polarity period” と、原文にない目的語が的確に訳出されている

原文

「ある一定期間あるいは一定計算回数経過した場合に計算を終了する。」

2018年度出力結果

“It is for a certain period of time or to exit calculation on a number of computations”

2019年度出力結果

“When a certain period of time or a certain number of calculations has elapsed, the calculation is terminated.”

ポイント

- 文章が成立している (ただし正確ではない)

2.3.2 改悪 (例)

原文

「過冷却解除が検知されると、過冷却解除した直後の食品温度 Th₂ はその食品の凍結点に相当する。」

2018年度出力結果

“When the subcooling release is detected, the food temperature Th₂ immediately after the subcooling release corresponds to the freezing point of the food.”

2019年度出力結果

“When the supercooling cancel is detected, the food temperature Th₂ immediately after the supercooling release corresponds to the freezing point of the food.”

ポイント

- 「過冷却解除」の訳揺れが生じている

原文

「【請求項1】」

2018年度出力結果

“[Claim 1]”

2019年度出力結果

“[Wherein 1]”

2.3.3 ほぼ変化なし (例)

原文

「逆極性期間を設定する逆極性期間設定部と、」

2018年度出力結果

“A reverse polarity period setting unit for setting a reverse period polarity period;”

2019年度出力結果

“A reverse polarity period setting section that sets a reverse polarity period;”

ポイント

- ・表現の変化のみで意味上の変化はない

2.3.4 まったく変化なし (例)

原文

「イジングモデルは磁性体の基本的なモデルとして磁性体材料科学において重要な役割を果たしてきた。」

2018年度出力結果

“The aging model has played an important role in magnetic material science as a basic model of magnetic materials.”

2019年度出力結果

“The aging model has played an important role in magnetic material science as a basic model of magnetic materials.”

ポイント

- ・「イジングモデル」はいずれも誤訳

3 ポストエディットの観点から見た出力結果

NIPTAでは、現在のNMT出力結果について、外国出願用特許明細書の翻訳に求められる品質としては不十分であると結論付けている。

その理由として、

- ・初出/既出の概念がない
- ・単数/複数概念がない
- ・用語/表現の統一がなされていない
- ・符号が符号として認識されていない
- ・読み間違いがある
- ・誤った用語/表現が用いられていることがある
- ・原文にない語句が補われていることがある
- ・特許実務は考慮されていない

などを挙げている。

現在のNMT出力結果をポストエディットするに際し、どのような観点でおこなうことが必要なのかについて説明する。

(例1)

原文

「逆極性期間を設定する逆極性期間設定部と、」

2019年度出力結果

“The reverse polarity period setting unit for setting the reverse polarity period;”

ここでは “The reverse polarity period setting unit” “the reverse polarity period” とともに初出の情報であるため、不定

冠詞 “a” にポストエディットする必要がある。

さらに、特許実務の観点から、“for setting” は “configured to set” または “that sets” などとしたほうが好ましい。

(例2)

原文

「正極性期間と逆極性期間とを算出し、前記溶接制御部へ出力する算出部と、」

2019年度出力結果

“A calculation unit that calculates a positive polarity period and a reverse polarity period and outputs the calculated period to the welding control unit;”

“the calculated period” となっているが算出されたのは2種類の期間であるため “the calculated periods” とポストエディットする必要がある。

さらに、特許実務の観点から、“the calculated periods” では不明確であるため “the positive polarity period and the reverse polarity period” としたほうが好ましい。

(例3)

原文

「この温度に基づいて、チルド室下段容器内の目標温度 Tc _ set を、細胞損傷なき程度まで氷結晶を融解できる温度、例えば Th _ 2 + 2 [°C] に設定する (S 9)。」

2019年度出力結果

“Based on this temperature, the target temperature Tc _ set in the chilled chamber lower container is set to a temperature at which the ice crystal can be melted to the extent that cell damage is possible, for example, Th _ 2 + 2 [° C.] (S 9).”

まず、目視で可能なポストエディットとして、符号や記号に含まれる余計なスペースを削除する必要がある。

また、「細胞損傷なき程度まで」に対する “to the extent that cell damage is possible” は誤訳であるため “to the extent of preventing damage to the cells” などと修正する必要がある。

4 結論

2019年度の出力結果について、ポストエディットを必要としない出力結果は84個中6個であった。

すなわち、殆どの出力結果が、程度の差はあれポストエディットを要するものであった。

ポストエディットをおこなう場合、出力結果と原文とを照合する前に、原文に記載の内容を正確に理解しておかなければ誤訳に気付かない可能性がある。

また、翻訳を経験していない限り、ポストエディットをおこなうことは困難であると考えられる。

NMTの出現により、一部の翻訳会社などでは『ポストエディター』と呼ばれる人材の育成・採用をおこなっているが、『ポストエディター』に求められる能力は翻訳者に備わっている能力と同程度である必要があることは明らかである[1]。

5 プリエディットの指針

上記の調査結果を踏まえて、どの程度の原文修正が実務者にとって適当なのか検討した。

NMTを翻訳に活用している現場では、文書全文のNMT出力を後で修正する手法は効率が悪いことが知られてい

る。原文と突き合わせて誤訳や訳抜けなどを発見するのに手間がかかるからだ。

そこで、本稿では、1文単位のNMT出力をプリエディット（原文編集）する方法を前提とし、効果的なプリエディット方法を提案する。文書全体に対するプリエディットとして表記ゆれの修正などが考えられるが、本稿では扱わない。また、NMT出力をより効率的に使う方法を示すが、すべてのNMT出力を無理やり使うことを推奨しているわけではない。プリエディットをすれば訳文の出力がよくなるが、実務ではプリエディットに時間をかけすぎても意味がない。

翻訳者にとってNMT出力は「訳文を作るための材料」であるため、プリエディットも「訳文のパーツを得るため」に行うものとする。つまり、文法的に正しく誤訳のないNMT出力を目指してプリエディットをするわけではない。

上記の観点から、出力された訳文の構文が崩れるような場合や訳抜けがひどい場合に、短時間で行える必要最小限のプリエディットをする。この時のポイントは以下の3点である。

1. 主語を特定しやすい原文を作る
2. 係り受けがわかりやすい原文を作る
3. 原文が日本語として不自然であっても気にしない

6 プリエディット例

NIPTAの過去問題を例文に中心に説明する。2019年12月現在のウェブの無料版Google翻訳の出力結果を用いて説明をする。

6.1 無生物主語に変換

翻訳者としての知識「無生物主語構文を作ると英語らしく明快な表現になる」をNMTに活用する方法を紹介する。

(例1)

原文

「このような熱による繰り返し荷重により、噴射ノズル10をチャンバの外壁21に結合させるねじ結合部12の部位に緩みやクラックが発生する恐れがある。」

NMT出力

“Due to the repeated load caused by such heat, there is a risk of loosening or cracking occurring at the site of the screw coupling portion 12 that couples the injection nozzle 10 to the outer wall 21 of the chamber.”

上記出力では、和文の構造を正確にとらえた訳になっている。この場合に無生物主語構文の訳文を出力させるための最小限の修正例を示す。以下のように一部分だけを修正する。このように原文の日本語が文法的に正しくなくても、無生物主語の英文が出力されることがある。

(例2)

原文（取り消し線は削除箇所、下線は追記箇所）

「このような熱による繰り返し荷重によりが、噴射ノズル10をチャンバの外壁21に結合させるねじ結合部12の部位に緩みやクラックが発生する恐れがある。」

NMT出力

“Such repeated loads due to heat may cause loosening and cracks at the site of the screw coupling portion 12 that

couples the injection nozzle 10 to the outer wall 21 of the chamber.”

「により」を「が」に変更するだけなので、例1の出力文を無生物主語構文に修正する場合と比較して、小さな手間で修正ができる。

6.2 句点を挿入

プリエディットの手法として文中に句点を挿入してNMT出力を変更する方法が提案されている[2]。原文が長い場合でも、使いやすいNMT出力を得られる場合がある。以下、句点の挿入箇所の例を示す。

(例3)

原文

「チャンバ20と噴射ノズル10は、蒸着工程中は噴射孔11を通過する気化した原料物質と接触するので、気化した原料物質から伝達された熱により高温状態になった後、蒸着工程が完了すると再度低温状態になる。」

NMT出力

“Since the chamber 20 and the injection nozzle 10 come into contact with the vaporized raw material passing through the injection hole 11 during the vapor deposition process, the temperature becomes high due to the heat transferred from the vaporized raw material, and then the vapor deposition process is completed. The temperature becomes low.”

文が長いと、上記下線のように文章がおかしな箇所区切られたり破綻する場合がある。このような場合、この出力をポストエディットをしてもよいが、簡易なプリエディットをして訳文がどのように変化するのか確認するのもよい。

(例4)

原文（下線は追記箇所）

「チャンバ20と噴射ノズル10は、蒸着工程中は噴射孔11を通過する気化した原料物質と接触する。ので、気化した原料物質から伝達された熱により高温状態になった後、蒸着工程が完了すると再度低温状態になる。」

NMT出力

“The chamber 20 and the injection nozzle 10 come into contact with the vaporized raw material passing through the injection hole 11 during the vapor deposition process. Therefore, after the material is brought into a high temperature state by the heat transferred from the vaporized raw material, the temperature becomes low again when the deposition process is completed.”

例3の原文は「蒸着工程中は部材の温度が上昇し、蒸着工程が終了すると部材の温度が低下する」ことの説明文である。この理解のもと、「部材の温度上昇のメカニズム」と「部材の温度変化の説明文」を分けるため、上記のように句点を入れる。日本語としては不自然な「ので、」で始まる第2文は、「Therefore」で始まる訳文として出力された。翻訳者はこの出力文を修正しやすいと思えば、ここからポストエディットを始める。

ただ、上記英文を読むと、下線のように間違った主語が挿入されている。また、温度上昇が蒸着工程中に起こることを第2文に追記してもよさそうである。翻訳者はこの出力文を修正してもよいし、例5のように別のポストエディ

ットし出力される訳文を確認してもよい。

(例5)

原文 (下線は追記箇所)

「チャンバ20と噴射ノズル10は、蒸着工程中は噴射孔11を通過する気化した原料物質と接触するので、気化した原料物質から伝達された熱により高温状態になる。った後、蒸着工程が完了すると再度低温状態になる。」

NMT出力

“Since the chamber 20 and the injection nozzle 10 come into contact with the vaporized raw material passing through the injection holes 11 during the vapor deposition process, the temperature thereof becomes high due to the heat transferred from the vaporized raw material. After that, when the deposition process is completed, the temperature becomes low again.”

今回は、別の個所に句点を挿入し、「蒸着工程中に温度が上昇すること」を明確にしてみた。この場合、第2文の文頭の「った後、」は日本語として不自然な表現であるが、出力された訳文では「After that」として訳出されている。この例からも、不自然な日本語でも自然なNMT出力が出力されることがわかる。

係り受けが複雑で翻訳しづらい原文であってもリエディットをすれば意味を整理して訳文を出力できる場合がある。そのため誤訳を発見しやすくなり、出力文の編集もしやすくなる。結果、ポストエディットの作業負担を下げられることがわかる。

このように、どのようにリエディットをし、どのタイミングでポストエディットをするのかを判断しながらNMT出力を活用する。

6.3 「化合物A」に変換

NMTでは化合物名など長い文字列が続く場合、単純な構造の文章であってもNMT出力では文章構造が壊れる可能性が高まる。

(例6)

原文

「遷移金属触媒として、(1, 3-ビス(2, 6-ジイソプロピルフェニル)イミダゾル-2-イリデン)(3-クロロピリジル)パラジウム(II)クロライドを用いる請求項1~4のいずれか1項に記載の多環式化合物の製造方法。」

(特開2010-120935号 請求項5を改変)

NMT出力

“5 The method according to claim 1, wherein (1,3-bis(2,6-diisopropylphenyl)imidazol-2-ylidene)(3-chloropyridyl)palladium(II)chloride is used as the transition metal catalyst. The manufacturing method of the polycyclic compound of this.”

上記下線の通り、原文にないクレーム番号「5」が挿入され、従属先のクレーム番号が「1」に誤訳されている。また、文章が2文に分割されている。また化合物名の括弧前後に半角スペースが挿入されているため全体として修正箇所が多く使いにくい訳文である。

これを以下のように化合物名を「化合物A」と置き換えることで係り受けがわかりやすくなり、原文の情報を正確に反映した訳文が出力される。

(例7)

原文 (取り消し線は削除箇所、下線は追記箇所)

「遷移金属触媒として、~~(1, 3-ビス(2, 6-ジイソプロピルフェニル)イミダゾル-2-イリデン)(3-クロロピリジル)パラジウム(II)クロライド~~「化合物A」を用いる請求項1~4のいずれか1項に記載の多環式化合物の製造方法。」

NMT出力

“The method for producing a polycyclic compound according to any one of claims 1 to 4, wherein “Compound A” is used as the transition metal catalyst.”

上記では構文が正確に訳出され、数字の誤訳もない。これならば、“Compound A”を正しい訳語に置き換えることでNMT出力を最大限に活用できる。

7 まとめ

本稿では、現在のNMT出力を翻訳者目線で評価した。外国出願用の特許明細書の翻訳としては、現時点でのNMT出力の品質は不十分である。そのため、NMT出力のポストエディットは必ず必要になる。このポストエディットの作業を少なくするために、最小限のリエディットでも効果があることがわかった。この場合、原文の日本語が不自然であっても使いやすいNMT出力が得られることがある。

NMT出力を効果的に活用するためには、リエディットの試行錯誤とポストエディットをバランスよく組み合わせる必要がある。それと同時に、NMT出力が使いづらい場合にはそれを一切使わずに自ら訳文を作成する判断力も必要である。

今後、NIPTAではNMTの活用についてさらなる研究を進めていく予定である。

[参考文献]

- [1] 山田優、大西真奈美 (2018) それでも学生はポストディッターになれるのか 言語処理学会第24回年次大会
- [2] 森口功造、中安裕志、星井智 (2019) .日英機械翻訳前のリエディット~そのポイントとポストエディットによる補完~. テクニカルコミュニケーションシンポジウム2019 論文集, pp. 66-73