

翻訳者からみた機械翻訳の現況（昨年発表の続編）

- ① はじめに
- ② ニューラル機械翻訳の優れているところ
長文に追従可能、辞書に非記載の新語に強い
- ③ ニューラル機械翻訳で未だに解決しない誤訳
早合点の誤訳、訳抜け、数量と状態の表現など
- ④ 他に指摘したいこと
- ⑤ 機械翻訳の効果的な利用方法
- ⑥ 機械翻訳の発展で、翻訳の担い手が変わる
- ⑦ 機械翻訳の普及のために情報漏洩の対策
- ⑧ まとめ

吉川潔（きっかわ きよし） 2023年11月29日

① はじめに

1-1. 新潟県の自宅で、東京の特許事務所や翻訳会社から原稿を電子メールで受信し、翻訳後に返信という **テレワークで 40 年**働いてきた。

機械翻訳を試訳し、関係者に提示してきた。

防災マニュアルなどの地域資料を、外国人を災害時に支援するために、**機械翻訳**を用いるボランティアで翻訳してきた。

道路や公園用に土地を購入し、他の有志と共に寄付し整備した。これが、19年前に新潟県を襲った水害と中越地震の復旧作業に役立ち、10年前に紺綬褒章を受章した。

1-2. AAMT（アジア太平洋機械翻訳協会）での活動

- ・ 40年前に機械翻訳を試訳したことが縁で、市販の機械翻訳を試訳し、結果と対案をメーカーを訪ねて説明し、関係者に配布してきた。

例えば、

2018年：機械翻訳について翻訳者が語りたくないこと（前編と後編）（A4・38頁）

2022年9月：翻訳者からみた機械翻訳の現況と改善案（A4・13頁）

2023年11月：翻訳者からみた機械翻訳の現況と誤訳の実態（A4・11頁）

－ これに準じて、本日は発表 －

- ・ 特許情報シンポジウムにおける、今までの発表の内容

2010年 第一回 MT試訳の説明と実用化のための提案

2012年 第二回 基本的な誤訳例の対策を提案

2014年 第三回 MTを優れた辞書として活用する方法

2016年 第四回 特許明細書の翻訳の注意

2018年 第五回 ニューラル翻訳の試訳結果の説明

- ・ 2019年 AAMT 2019 Tokyo 「NMTの誤訳と翻訳文の合否」を発表

- ・ 2022年 AAMT 2022 Tokyo 「翻訳者からみたMTの現況と効果的な利用」を発表

- ◆ 2017年 筑波大学で催された言語処理学会の大会で

防災マニュアルの多国語翻訳を発表

1-3. 本論に入る前に

昨年の **AAMT 2022 Tokyo** は、時間不足で説明を省略した項目があった。
今年の **AAMT 2023 Tokyo** は、昨年、省略した項目を主として説明する。

機械翻訳の誤訳を調べる人が少ないので、誤訳の実態が不明瞭である。

誤訳の実態の全体を示し、所々は省略した説明なことを了解してほしい。

文例は、本年11月中旬に機械翻訳の関係者に発信した

「翻訳者からみた機械翻訳の現況と誤訳の実態」に準じている。

該当部分は、参考資料の○頁を参照と記しているので、参照してほしい。

試訳の○×は、発表者が翻訳の仕事を承っている某所の基準に準じている。

4つは常に修正しているので、結果は変わっている場合もあると推測する。

試訳に用いた機械翻訳は、インターネットから利用できる4方式、

グーグル、マイクロソフト、**NICT**（情報通信研究機構）、**DeepL**である。

② ニューラル機械翻訳の優れているところ

2-1. 長文に追従可能（発表者が出会う原稿の90%に追従可能）

下記は今年の AAMT2022と同じだが、説明を省いたので、今回は要点を説明する。

電力会社は、再稼働を目指す原子力発電所の7号機で、**発電タービンを回した後**の蒸気を冷却して水に戻す復水器とその関連部品の点検、即ち、冷却用の海水を循環させるポンプを動かし、復水器内を真空にする機器の動作を調べ、配管に漏れが生じないことを調べ、全てが正常に機能するかどうかを確認する点検を、10月中旬に始めると発表した。

4つの中で3つが×。一つは、赤の部分の英訳が×であるが、△とする。

△ **An electric power company announced that in the middle of October, the company would start an inspection of a condenser that cools steam **after turning a power generation turbine** and returns it to water, and the related parts, that is, an inspection of an operation of an equipment moving a pump that circulates seawater for cooling and vacuums the inside of the condenser, an inspection of whether there is leak in piping, and an inspection of whether everything functions normally, in Unit 7 of a nuclear power plant that is going to be restarted.**

2-1-1. 前頁の補足説明として 日本の英語教育の盲点、分子構文の主語が不鮮明

2016年12月の第四回特許情報シンポジウムで下記を説明
昨年の AAMT 2022 Tokyo では、説明時間の不足で説明できなかった

(北海道を旅行したときに、雪が残っていた) を英訳する際に、

× **When traveling in Hokkaido, snow still remained.**

雪が北海道に旅行したときに、雪が残っていた

分詞構文に主語がないときは、本文の主語（雪）が分詞構文の主語になる

○ **When I traveled in Hokkaido, snow still remained.**

私が北海道に旅行したときに、雪が残っていた。

分詞構文の主語の有無は、日本の学校英語では触れていないようである。
某特許事務所に上記のような英訳文を納入したら、厳重に注意された。
こういう事例を他にも見聞きする。日本の英語教育の盲点！

2-1-2. 本文の問題の部分を抜き出して英訳してみる。

1つだけ長文に追従し、△と判定したが、問題の部分を再掲する。

(電力会社は、**発電タービンを回した後の蒸気を冷却して水に戻す復水器**の点検を始める)

× The power company starts inspecting **the condenser**, which **cools** the steam back to water **after turning the generating turbine.**

上記の場合、**cools**と**turning**の主語は復水器 (condenser) になるので、復水器が、発電タービンを回した後に、蒸気を冷却して水に戻す意味になる ×

— 復水器は、蒸気を冷却して水に戻すが、**発電タービンを回さない!** —

しかし、日本文の原稿に図面がないので、この機械翻訳は、構造の関係を知らずに直訳と推定。長文にもかかわらず追従しているので、△と判定した。

○ The power company begins inspections of **the condenser**, which **cools** the steam **after the power turbine is turned** and returns it to water.

上記の場合、発電タービンが回された後に、復水器が蒸気を冷却して水に戻す ○

この機械翻訳は、図面が無いにもかかわらず、この部分は正訳したが、長文の全体としての訳文はコマ切れ出力で意味不鮮明なので、×と判定した。

2-2. 辞書に非記載の新語に強い → 参考資料の2頁を参照

2-2-1. (ベロプスカイト太陽電池は折り曲げ可能である) の英訳

A **velovskite** solar cell is bendable

ベロプスカイトが辞書に非記載でも、
機械翻訳から単語や一文として適訳が入手可能

2-2-2. 理工系の資料には (ポリエチル - - ~) などの化学用語が続出し、 今迄なら数時間も要していたが、ニューラル機械翻訳NMTのおかげで 正訳を数分で得ることができるので、大助かりである。

2-2-3. 昨年の AAMT 2022 Tokyo でも説明したが、 **オミクロン株**は最新用語なので書籍辞書や **CD-ROM** 辞書に非記載である。 スマホの翻訳／通訳機能をセットして (オミクロン株) と音声入力すると 「オミクロン株」という文字が、入力確認のためにディスプレイに現れ、 「**omicron strain**」と、英訳文が音声出力と共にディスプレイに出力する。 ※ パソコンも同様である。

2-2-4. 他に、人口無能、連銭形成など、辞書に非記載で困ったときに、 機械翻訳ソフトを用いて適訳を見いだせるので助かる。

2-3. 訳しづらい(紛らわしい) 語句の適訳を入手できる

→ 参考資料の2頁を参照

2-3-1 スピーカーは視聴者に近ければ近いほど良いというわけでない

(近ければ近いほど良い) という部分が訳しづらいので、機械翻訳に入力

The closer the speaker is to a viewer, the better.

上記は面倒な部分が適訳だが、(～良いというわけでない) が訳抜けで×

私案として、下記の英訳文を作成

It is necessarily impossible to say that

the closer the speaker is to a viewer, the better.

この英訳文を機械翻訳に英文和訳として逆方向に入力したら、
和訳文が原文の趣旨に近いので○とした。

2-3-2 昨年の AAMT 2022 Tokyo でも説明したが、

このロボットは人見知りをする → This robot is shy. shyだけでは説明が不十分
(人見知りをする) を国語辞典で調べると

(知らない人を見ると、恥ずかしがったり嫌ったりする) という意味

「このロボットは、知らない人を見ると、恥ずかしがったり嫌ったりする」として
機械翻訳に入力 ○ This robot, when seeing a stranger, becomes shy or disliked.

③ ニューラル機械翻訳 NMT で未だに解決しない誤訳

3-1 早合点の誤訳 NMT で急増

→ 参考資料の3～9頁を参照

製品の品質が、製品のコストダウンよりも重視される - 重要と重視は意味が異なる

- Product quality is emphasized rather than product cost reduction
- × Product quality is more important than product cost reduction

3-2 訳抜け：連語 複数の名詞や形容詞が連なる語句で一部が訳抜け NMT で急増 書類に同様の記載ミスがある

- There is a similar description incompleteness in a document.
- × There is a similar incompleteness in a document 記載が訳抜け

3-3 数量表現

競技者は30数人です（日本語では、30+数人の意味）

- Players are 30-odd persons
- × Players are about 30 persons （27～29人も含む意味なので×）

3-4 状態の表現：単純な状態は正訳するが、比喩的な状態の英訳？

彼は目が染みると嘆いている ○ He laments that his eyes are soaked

3-5 受身（日本原稿が能動形でも、受身で英訳しなければならない場合がある） 前述のスライド 2-1-2を参照

以上は、昨年の **AAMT 2022** の発表の一部だが、未解決の状態なので再掲した 10

④ 他に指摘したいこと

→ 参考資料の (9~11) 頁を参照

4-1 翻訳原稿や原文のミスに起因する誤訳

翻訳原稿の誤字脱字や「てにをは」のミスなど

4-2 日本の英語教育の盲点に起因する誤訳

分子構文の主語、 さきほどの **2-1-1** のスライドを参照

単数と複数の明確な区別、 a n d / o r

4-3 多国語の翻訳と通訳

(日本語 ←→ 英語 ←→ 多国語)の手順で訳しているが、

日英が未だに不十分なので、その先の多国語翻訳も不十分？

4-4 **主語さがし**について： 日本語には、主語の省略が多い。

某所では、原文に主語が省略の場合、

又は、主語を記さずに (これ) (それ) と記してある場合、

「t h i s」や「i t」と逃げずに、具体的な主語を原文から探して意識するように厳命されている。しかし、意識時に青で表示し、自信がないときは「i t」と逃げて青で表示している。¹¹

主語の間違いは**致命傷**になるので、

機械翻訳は、「i t」と逃げて、**青で表示**すべきである。

⑤ 機械翻訳の効果的な利用方法

今回の発表のところどころで、発表者の利用状態を説明してきたが、
昨年の AAMT 2022 Tokyo の発表と同様なので、それを再掲する。

発表者は2台のパソコンを翻訳作業に用いている。

下記の英訳で、1台のパソコンに長文に強い機械翻訳に日本文を入力。

図1は電気的な振動を機械的な振動に変換する回路を示し、
図2は、それらのデータのラプラス変換を示す図である

**FIG. 1 shows a circuit for converting the electrical oscillations
into the mechanical vibrations, and
FIG. 2 is a diagram showing the Laplace transform of these data.**

もう一つの長文に弱いが用語選択が正確な機械翻訳に上記の日本文を入力。

もう1台のパソコン上でwikipediaやCD-ROM辞書を用いて

用語の意味を確かめたり、訳しづらい語句を調べる。

特に、上記の例文では下記の用語の違いに注意！

電気的な振動	oscillation	～を変換する	convert	～
機械的な振動	vibration	ラプラス変換	Laplace transform	

最後に、英文を機械翻訳で逆方向に和訳し、原文の意味と同様なこと訳抜けのないことを確認する。これは1台のパソコンで可能だが、2台のほうが楽で正確である

⑥ 機械翻訳の発展で、翻訳の担い手が変わる！

- 機械翻訳の欠点
誤訳がある。
原稿のミスを訂正しながら、図面を見ながら翻訳できない。
翻訳原稿の丸投げの全面依存の利用条件で、機械翻訳は使えない。
- 大量に翻訳を外注する会社、特に外国出願を主とする特許事務所には、幼少時に米国で育ったという人がいて、職員の大半は技術系である。
発表者が納入した特許明細書の和文英訳の英文を日本語の原稿と一語一語慎重にチェックして米国に出願している。
- そのなかの二つから、吉川流の機械翻訳の利用方法を用いたら効果があるので、以降、自ら和文英訳し、外注は止めるといわれた。
自分で自分の首を絞めた結果になったが、この傾向は増加と推定する。
- 特許明細書は特許事務所の係員が書くので、日本語の原稿のミスを訂正して機械翻訳に投入し、出力した英訳文のチェックも容易である。
外国出願の叩き台として、機械翻訳は人間翻訳者と同程度に役立ち、機械翻訳は無料なので、和文英訳の外注は止めるということである。

⑦ 機械翻訳の普及のために情報漏洩の対策

某社では、情報漏洩を懸念して、インターネットを經由しての
機械翻訳や翻訳支援ツールの使用は厳禁である。

発表者は、10年前、プロバイダーから「吉川のパスワードを使って、吉川に成りすまして悪さをする人がいるので困っている。パスワードを変えて」と要請された。パスワードを変えたら、犯人らしき人？から「パスワード変更の経緯を教えろ」という電話があり困惑した。

どんな悪さをしたのか知らないが、発表者のパスワードと電話番号をどうして知ったのか、いま思い出してもゾツとする。

ものは試しに自社内で機械翻訳を使ってみたいが、情報漏洩が心配で、機械翻訳の使用を厳禁は困る。

機械翻訳の普及のために、情報漏洩の対策も重要である。

⑧ まとめ

機械翻訳が普及しない原因は誤訳にあることは、関係者は認めている

しかし、誤訳の実態を調べて発表する人は少ない

試訳の協力を旧来の翻訳者に呼びかけたいが、
その人の仕事量が減る結果になるので、綺麗ごとを言うことは止めた

翻訳の担い手が変わり、新たな機械翻訳の利用者が現れるので、
そういう人に誤訳調査を依頼し、機械翻訳を支援してほしい

生成A I の根幹を成す大量のデータは英文構成なので、
日本語で利用するには、機械翻訳システムの活用が必須になる
機械翻訳のアプリケーションフィールドが桁違いに拡大するので、
翻訳精度への要求も厳しくなると推測する

発表者は、従来と同様に、翻訳者の立場で機械翻訳を試訳し、
試訳結果を関係者に提示したい