

ISSN 1883-1818

No.79

December 2023

AAMT Journal

Asia-Pacific Association for Machine Translation

機械翻訳

機械翻訳



目 次

巻頭言		
MT Summit 2023レポート	田中 英輝	3
長尾賞		
医薬品開発業務を迅速化するための機械翻訳の協調的開発と運用	早川 威士、齋藤 敦	5
長尾賞学生奨励賞		
Multimodal Neural Machine Translation based on Image-Text Semantic Correspondence	Yuting Zhao	10
タスクに応じてトークン分割を最適化する試みと現状	平岡 達也	16
解説記事		
人間の翻訳と機械の翻訳(8): 翻訳者がしていること・MTにできること	影浦 峯	22
立命館大学の正課英語授業へのニューラル機械翻訳導入について	坂場 大道、木村 修平、近藤 雪絵	30
AAMTセミナー		
ポストエディットの真実	隅田 英一郎	37
温故知新		
温故知新3	後藤 功雄	41
イベント報告		
MT Summit 2023	本多 純礼	49
AAMT Machine Translation Summit 2023 参加報告	橋本 簡美	50
MT Summit 2023 参加報告書	平岡 裕資	51
Participation Report on MT Summit 2023	山下 優樹	52
第10回特許・技術文書翻訳ワークショップ (PSLT 2023) 開催報告	綱川 隆司、須藤 克仁、後藤 功雄	54
第10回アジア翻訳ワークショップ (WAT2023) 開催報告	中澤 敏明	56
法人会員PR		
ChatGPTを翻訳業務に活かすには	中山 雄貴	59
編集後記		
	出内 将夫	61

C O N T E N T S

Foreword		
Report on MT Summit 2023	Hideki Tanaka	3
AAMT Nagao Award		
Cooperative Development and Introduction of Machine Translation Technology to Accelerate Process of Drug Development	Takeshi Hayakawa / Atsushi Saito	5
AAMT Nagao Student Award		
Multimodal Neural Machine Translation based on Image-Text Semantic Correspondence	Yuting Zhao	10
Task-Oriented Tokenization	Tatsuya Hiraoka	16
Commentary		
What translators do and what MT can do	Kyo Kageura	22
Introduction of Neural Machine Translation into Regular English Classes at Ritsumeikan University	SAKABA Hiromichi, KIMURA Syuhei, KONDO Yukie	30
AAMT Seminar		
Truth of Postedit	Eiichiro SUMITA	37
Learning from the past		
Learning from the past 3	Isao Goto	41
Event Report		
MT Summit 2023	Sumire Honda	49
Participation Report: AAMT Machine Translation Summit 2023	Hiromi Hashimoto	50
MT Summit 2023 Participation Report	Yusuke Hiraoka	51
Participation Report on MT Summit 2023	Yuki Yamashita	52
Report of The 10th Workshop on Patent and Scientific Literature Translation (PSLT 2023)	Takashi Tsunakawa, Katsuhito Sudoh, Isao Goto	54
Report of the 9th Workshop on Asian Translation (WAT2022)	Toshiaki Nakazawa	56
Corporate PR		
How to make use of ChatGPT for translation work	Takeyoshi Nakayama	59
Editor's Note		
	Masao Ideuchi	61

巻頭言

MT Summit 2023 開催報告

田中 英輝

情報通信研究機構

1. はじめに

2023 年 9 月に AAMT（一般社団法人アジア太平洋機械翻訳協会）主催による第 19 回機械翻訳サミット、MT Summit 2023、が開催された。報告者は本サミットの準備から終了まで携わる機会を得たため、本稿では主催者から見た情報を交えて概要を報告する。

会議の概要は以下の通りである。

- 場所：中国マカオ、Studio City ホテル
- 期間：9 月 4 日、5 日、ワークショップ 4 件、9 月 6 日から 8 日、本会議
- 形式：現地と遠隔のハイブリッド
- 参加人数：169 名、うち現地参加者 110 名程度
- 参加国：25 カ国（表 1 参照）
- スポンサー：16 社（表 2 参照）
- 招待講演：基調講演 1 件、招待講演 2 件、招待セッション 2 件（1 件はパネル）
- 研究発表：リサーチトラック 33 件、ユーザーズトラック 21 件

現地開催のみだった 2019 年（ダブリン）の参加者 290 名に比べると、今回はその 6 割程度に留まった。要因はいろいろあると思われるが、表 1 から分かるように開催地の中国からの参加が予想に反して少なかったことが一因である。中国の大学からマカオへの出張手続きの煩雑さ、大学の新学期との重なりなどが原因として推察されるが、だれもが参加しやすい開催地、時期の選択は今後の課題である。

一方、論文数、招待講演の数などは例年通りの規模であり、機械翻訳そのものへの関心は従来以上に大きかった。特に、今年に入って大規模言語モデルの翻訳への応用に高い関心が集まり、招待講演、パネル、研

究発表などで熱心に議論された。なお、今回の IAMT Award of Honor は機械翻訳システムの研究開発に長年にわたり貢献された Huawei Noah's Ark Lab. の Qun Liu 博士に贈られた。

表 1 上位 5 カ国の参加人数

日本	中国*	マカオ	米	独
65	28	14	10	7

*マカオ、香港を除く

表 2 スポンサー国別内訳

日本	中国	米	独	伊
8	3	3	1	1



2. 基調・招待講演概要

以下では興味深かった 2 件の講演の概要を記す。

基調講演：Toward better neural and big model-based machine translation, Min Zhan (Harbin Institute of Tech.)

統計翻訳、ニューラル翻訳と大規模言語モデル(LLM)の翻訳の最大の違いは、LLM の学習には対訳コーパスが

Report on MT Summit 2023

Hideki Tanaka

National Institute of Information and Communications Technology

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License.

License details: <https://creativecommons.org/licenses/by-sa/4.0/>

不要で単言語の巨大なコーパスを使う点にある。LLM は英語の要約など様々な自然言語処理のタスクで高い性能を達成し世間の注目を集めているが、驚くのは単言語データで学習しているにもかかわらず翻訳精度が高い点である。実際には、例えば LLaMA の学習データには多言語テキストからなる Wikipedia が含まれる。しかし全学習データの 4.5%に過ぎず、かつ翻訳の形で学習されたわけではない。なぜ高精度に翻訳できるのか。その謎の解明に向けて世界中で研究が進められている。講演では明らかになった利点として、

1. 文脈の高い理解力
2. 対話的に翻訳可能
3. 翻訳できるスタイルの多様性
4. 文書単位の高い翻訳能力
5. 多言語対応

また、課題として

6. 自動評価が困難
7. ドメイン固有の専門知識不足、転移学習が困難
8. 大きな計算コスト
9. 低資源言語の翻訳性能の低さ

を挙げていた。利点 2 は、翻訳開始時に訳語を指定できることや、出力された翻訳の時制などを再度指示によって変更できることを指す。利点 3 は多様な文体の入力受け付けることを意味する。ただし「ことわざ」などの翻訳では、翻訳に加えてその意味の解説が加わってしまう過剰翻訳の問題を生ずることもある。課題 6. は翻訳精度が高すぎて従来の手法では十分評価できなくなっている問題を指す。課題 7 と 9 は関連しており、ファインチューニングなどである程度所望のタスクに対応できるが、その他のタスクの性能が大きく損なわれる破滅的忘却が発生することを指す。講演では課題に対して研究チームで進めている自動評価、転移学習、計算量削減などの研究紹介があった。翻訳以外にもなす LLM の研究にはいわゆる人工知能一般の課題が山積しており、今後取り組むべきテーマ例として、多モダリティ機械翻訳、信頼できる言語処理、説明可能な言語処理、

人間の意を汲んだ学習などを挙げていた。

招待講演：How to progress beyond human quality?

Ondřej Bojar (Charels Univ.)

機械翻訳が人と同じレベルに達したと語られることがあるが、これは、一部の言語対の文単位の翻訳の場合に過ぎない。例えばチェコ語から英語への翻訳の場合、同じ意味の語順を変えたチェコ語を正しく英訳できない事例などを挙げ、問題は枚挙にいとまがないことを示した。しかし LLM の出現もあり翻訳性能が人間に近づきつつあるのは事実で、ここで Bojar 氏が問題視しているのが基調講演の Zhan 氏と同じく自動評価である。特に、LLM を使った翻訳では学習データが不明な場合が大半で、評価データが未学習であることが保証できない状況になっている。講演ではこの問題を含め、新しい自動評価手法の研究を紹介した。最後に最近の LLM ブームについて、従来型翻訳システムの研究が終わったような論調になりがちだが、LLM の翻訳にも低資源言語翻訳の性能が悪いなど多くの問題がある。また研究の多様性は極めて重要で、従来型の機械翻訳研究を放棄すべきではなく、むしろ、機械翻訳研究者は LLM を批判的に研究し、”Help ChatGPT fall slowly and gracefully”であるべきだと結んだ。

3. おわりに

本稿では紙数の関係で一部の招待講演のみを紹介したが、興味深い研究が多数報告された。これらについては後日 ACL Anthology で公開される予定の予稿集をぜひ参照されたい。会議の最終日には EAMT (ヨーロッパ機会翻訳協会) の関係者から、次回の MT Summit 2025 はスイスのジュネーブで開催されることが発表された。衝撃的に出現し、今回の話題をさらった LLM を使った翻訳システムが今後どのように進化していくか、次回の会議が今から楽しみである。最後に、MT Summit 2023 の開催にご尽力いただいた、各委員会メンバー、査読委員、学生ボランティア、そして多大なご支援を賜ったスポンサー各位に感謝して本稿を終わりたい。

第18回 AAMT 長尾賞受賞記念

医薬品開発業務を迅速化するための機械翻訳の協調的開発と運用

早川 威士

株式会社アスカコーポレーション

齋藤 敦

中外製薬株式会社

信頼性保証企画部 翻訳マネジメントグループ

1. 医薬品開発業務への機械翻訳 (MT) 導入

人工知能 (AI) 技術はすでに様々な形態で実用化され、いまや産業の発展を強力に推進するツールとなっている。ただし、AI は出力に対して確実性や正確性が約束されるものではなく、導入した効果の測定が難しいという課題がある。AI の有効活用には、期待することや目的を明確にし、導入効果を可視化できるような枠組みが必要である。

医薬品業界はグローバル化が進展した産業であり、多言語コミュニケーション手段として翻訳が重要な役割を持つ。翻訳には高い知識水準が要求される一方で時間短縮も課題となっており、機械翻訳 (MT) 技術には、翻訳品質の安定化と翻訳作業の迅速化の双方が求められている。しかしながら、現在までの MT 技術は、医薬品関連文書に求められる高いレベルの翻訳品質に追いついておらず、直ちに人手の翻訳を MT に置換できるわけではない。そのため、人によるポストエディット (PE) を行うことを前提とした最適な MTPE 手法を設計・開発する必要がある、MT の有用性は PE の品質と作業スピードを加味して評価すべきと考えられる。

我々は本研究において、MT を用いた翻訳モデルを最適化することにより、MT 技術が翻訳品質の安定化と迅速化の双方にいかに関与するかを検討した。具体的には、人手翻訳による翻訳プロセスから、MTPE に移行することで、翻訳に要する期間がどれだけ短縮できるかを比較した。翻訳モデルの最適化としては、(1)

翻訳品質を改善するためのドメイン (医薬品開発) に特化した MT の開発、(2) 過不足なく適切な PE を実施するための品質水準の設定、の 2 つの手法を取り入れた。

また本稿では、実証研究後の実業務において翻訳の迅速化が達成されているかを追跡調査した結果についても併せて報告する。

2. ドメイン特化型 MT の開発および評価

本研究の対象となっている医薬品開発業務は極めて専門性が高い領域であることから、非特異的なコーパスで学習させた MT (汎用 MT) は本業務の用途に適さないと考えられる。そこで、転移学習の一手法である domain adaptation を用いて、ドメイン特化型の MT を構築し本研究の MTPE に使用するものとした。

学習データは著者の企業が保有する日本語と英語がペアになった業務用文書約 24 万文を使用し、基盤 MT として採用した「みんなの自動翻訳@TexTra® (国立研究開発法人情報通信研究機構)」に対し domain adaptation を実施した。

上記で構築した MT (カスタマイズ MT) が実際に PE の品質および作業スピードの改善に寄与するかを評価するため、カスタマイズ MT と汎用 MT を用いた英語から日本語の翻訳出力に対してそれぞれ PE を実施し、編集時間と編集率を記録して比較した。編集時

Cooperative Development and Introduction of Machine Translation Technology to Accelerate Process of Drug Development

Takeshi Hayakawa¹, Atsushi Saito²

1. Research & Development, ASCA Corporation 2. Translation Management Group, Business Strategy & Compliance Dept. Chugai Pharmaceutical Co., Ltd.

This work is licensed under a Creative Commons Attribution-Share Alike 4.0 International Public License.

License details: <https://creativecommons.org/licenses/by-sa/4.0/>

間は1文ごとにPEに要した時間(秒)を計測し、平均値を算出した。編集率はPE前訳文とPE後訳文のレーベンシュタイン距離を正規化した値とした。また一般的な評価指標として、BLEUおよびTER(translation error ratio)を算出した。以上の指標の統計学的検定にはpaired Student's t-testを用いた。さらに、医薬品開発業務における専門用語や固有の表現がどの程度正確に再現されているかを評価するため、これらの表現の正答率(固有表現一致率)を比較した。いずれの評価も医薬品開発業務に関連するテキストよりn=220文を用いた。

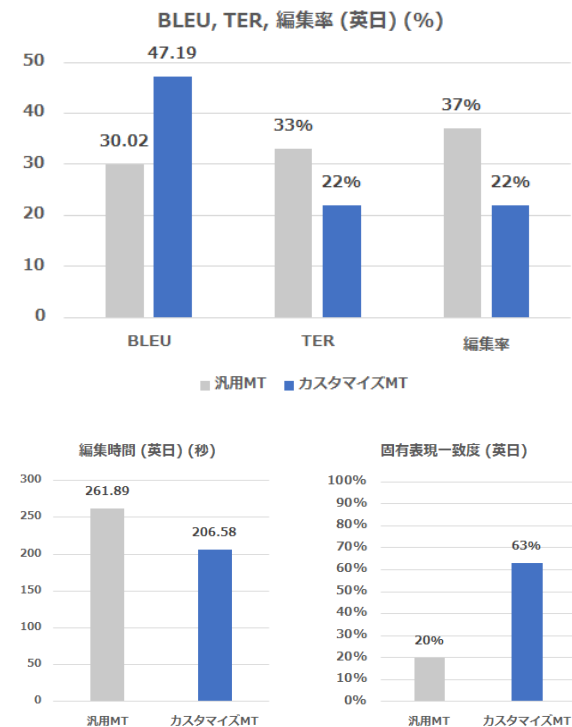
PEの評価の詳細を図1に示す。まずBLEUスコアの平均値はカスタマイズMTが約47を超える高値となり、汎用MTと比べ精度の改善が示された(47.19 vs 30.02; $p<0.01^*$)。TERも同様にカスタマイズMTが汎用MTより優れていた(22% vs 33%; $p<0.01^*$)。

編集率ではカスタマイズMTが汎用MTに対して優れており、統計学的有意差を示した(22% vs 37%, $p<0.01^*$)。編集時間についても、有意にカスタマイズMTが優れていた(206.58秒 vs 261.89秒, $p<0.01^*$)。

固有表現の一致率については、43ポイントの差でカスタマイズMTが汎用MTを上回り、domain adaptationの効果を裏付ける結果となった(63ポイント vs 20ポイント)。

これらの結果から、概ねカスタマイズMTの方が優れた翻訳結果を出力していると考えられ、実際のPE作業においても編集率が減少しており、PEの労力を軽減していることが示された。

図1. カスタマイズMTと汎用MTの比較



3. PE品質水準の設定

PEの作業時間を短縮し、MTPEによる翻訳モデルを最適化するために、PEの品質水準を設定し事前にポストエディターに周知した。これにより、ユーザーが期待する翻訳品質を上回る過剰な編集/重要性の低い編集にかかる作業時間、および編集作業の判断にかかる時間を短縮することを意図した。すなわち、ここで定義した品質水準は、品質に関する要件であるとともにPEの効率性を高める効果を期待したものであるといえる。

品質水準の内容は、(1) 要求事項と(2) 非要求事項をリストアップすることで構成され、PEでやるべきこととやらなくてよいことを明示化している。要求事項は、「正確性」、「ユーザビリティ」、「リーダビリティ」および「申し送り」の4項目から構成され、業務利用のための翻訳文書として比較的重要な項目として絞り

* 統計学的有意差を示す。

込んだ。一方で、非要求事項には表現の好みや一貫性など、一般に翻訳品質の構成要素ではあるものの用途を考慮すると比較的優先度の低い事項を挙げている。図2に品質水準の項目の一例を示す。

4. MTPE と人手翻訳の実証的研究

前章までで記述した翻訳モデルの妥当性を確認するため、実証的研究として人手翻訳によるプロセスとMTPEとで、翻訳に要する時間がどれだけ異なるかを比較した。

本研究では、構築した翻訳モデルを医薬品開発の実務に適用させるため、薬事規制に基づいて実施される臨床試験に実際に使用した「治験実施計画書」を対象文書として選択した。また、異なる2つの治験実施計画書の翻訳を2回に分けて行い、MTPEの時間短縮効果に再現性があるかを確認するものとした。時間短縮効果の目標は、人手翻訳に対し、MTPEで30%以上の時間短縮がみられることとした。同時に品質水準の達成度について、要求事項4項目に対するエラー数をカウントして評価した。1回目の結果で人手翻訳よりMTPEの方がエラー数が多い場合、翻訳およびPE担当者に対してフィードバックを実施し、2回目でのエラー抑制効果を確認した。

1回目の比較では、人手翻訳に対し、MTPEでは48%の時間短縮効果を確認することができた(図3、226時間 vs 106時間)。しかしながら、品質水準の達成度では、4項目いずれもエラーが発生しており、中でも「正確性」においてはMTPEの方がエラー数が多い結果となった(図4、19件 vs 26件)。このため2回目の実験では、エラー数を削減して品質水準をより高度に達成することも考慮することとした。

2回目の比較では、1回目を引き続き、人手翻訳に対してMTPEで43%の時間短縮効果を確認した(図3、191時間 vs 101時間)。なお、翻訳作業の実時間についても、1回目から短縮された。品質水準の達成度については、「正確性」「リーダビリティ」では1回目に比べてエラー数が減少するとともに、人手翻訳に比べてMTPEのエラー数が少ない結果となった。「申し送り」では1回目に比べてエラー数は減少したが、人手翻訳よりMTPEの方が多い結果となった。また「ユーザビリティ」ではエラー数は増加したが、MTPEの方が少なかった(図4)。2回めでエラー数を抑制できた背景としては、1回目の作業を踏まえた改善としてポストエディターへのフィードバックや作業指示書の更新などの対策の効果が大きいと考えられ、PEの品質改善にはコミュニケーションの改善が有効である可能性が示唆された。

- **要求事項**
 - **正確性：原文のプロトコルの情報が、過不足なく翻訳版に反映されている。**
 - 治験を実施する上で問題となる数値や単位の齟齬がない。
 - 否定／肯定の逆転がない。
 - 原文と翻訳版の図表のデータの内容に齟齬がない。
 - 原文中の治験実施に必要な情報について、不足や原文にない情報の追加がない。
- **非要求事項**
 - **リーダビリティ：逸脱を起こす読みにくさがない**
 - 意味や可読性に影響しない表現の好みを適用しない。
 - 意味や可読性に影響しない語順や「てにをは」、文末表現にこだわらない。
 - 文章の内容理解に支障を来さない限り、逐語訳を許容する。

図2. 品質水準の項目の一例

図 3. 人手翻訳と MTPE の治験実施計画書の翻訳に要した時間の比較

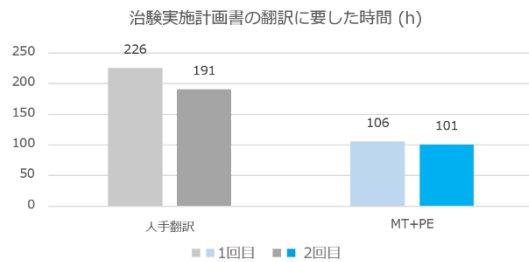
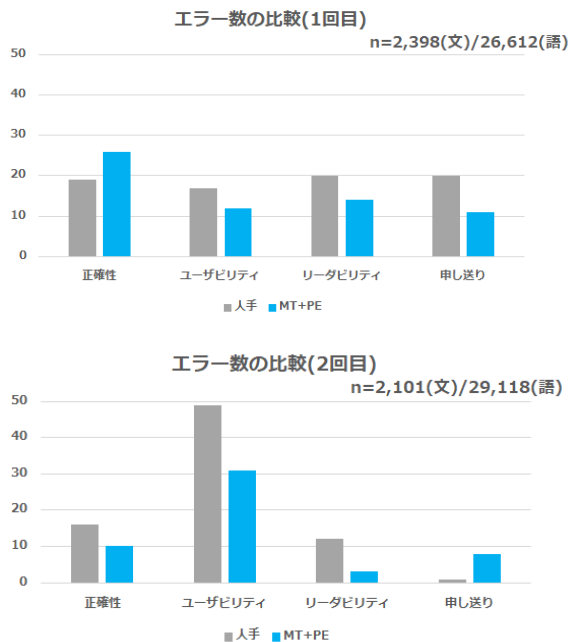


図 4. 人手翻訳と MTPE のエラー数の比較結果 (1 回目及び 2 回目)



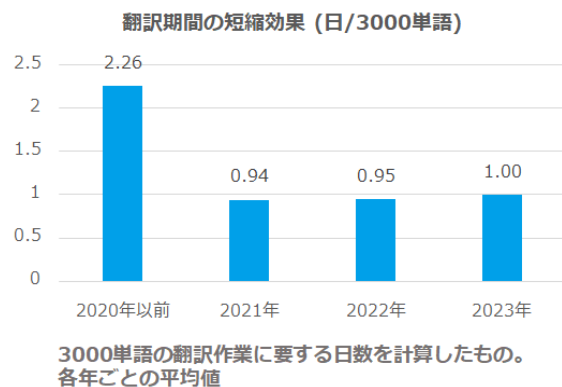
5. MTPE プロセス導入後の効果

これらの実証研究の結果を踏まえて、我々はカスタマイズ MT+PE の翻訳プロセスを治験実施計画書の翻訳において導入することに決定した。実験的環境において MTPE は翻訳プロセスの迅速化に貢献したが、現実の業務においては PE の人的資源の問題やスケジュールの競合など多様な制約が存在する。そこで我々は、MTPE の導入後に実証研究で検討した翻訳プロセスの迅速化が達成できているかを追跡調査した。

MTPE を導入して業務を遂行した 2021 年 1 月～2023 年 5 月と、未導入期である 2020 年以前の翻訳業務について、業務開始から終了までの日数と翻訳した文書のボリューム (単語数) のデータを収集し、単体量 (3,000 単語) あたりの翻訳期間を算出して比較した。

その結果、未導入期 (2020 年以前) では単体量の翻訳に 2.26 日を要していたのが、導入後は 2021 年～2023 年のいずれの年も 1.0 日を下回り、顕著な改善を認めた。このことから、MTPE の導入を機に翻訳処理スピードは大きく改善され、またその効果が維持されており、早期の納品に貢献していることが伺える。また、導入直後の 2021 年から改善効果は明確に現れており、これは実証研究の実施を含め、準備と検証を十分に行ってから導入した効果が現れた結果と考えられる。

図 5. MTPE プロセスの導入による翻訳期間の短縮効果



6. まとめ

本研究では、人手翻訳と比べて品質を落とすことなく、さらなる翻訳スピードを追求することを目的として、科学的手法とエビデンスに基づく MT の導入モデルの構築を試みた。この導入モデルはカスタマイズ MT の開発・評価と品質水準の設定という 2 つの要素により構成された。カスタマイズ MT の開発にあつ

ては、MT の品質を測る指標（BLEU、TER および編集率）と PE 作業時間を計測することで、MT の精度改善を確認した。また品質水準は不必要に高い品質目標を立てるのではなく、水準を明示的に文書化することにより、合理的な品質を効率よく達成することを主眼にデザインした。これらのモデルの効果を検証するために、2 回にわたって実際の翻訳プロセスによる比較を行い、時間短縮効果を確認できた。また、実業務への導入後もこの時間短縮効果は顕著に現れ、一貫して維持できている。

標題からも明らかなように、我々の取り組みは単に新しい技術やシステムを開発したのではなく、その導入と運用を実証研究ベースで行い、定量的評価に基づいて翻訳業務の効率を改善したものである。また、研究体制として MT サービスを提供する企業とユーザー企業が協調的に取り組み、成果を示すことができたという点でも意義は大きく、今後の企業への MT 導入にあたってモデルケースとなることを期待している。

Multimodal Neural Machine Translation based on Image-Text Semantic Correspondence

Yuting Zhao

Kyushu University

1. Introduction

Multimodal neural machine translation (MNMT) extends the conventional text-to-text NMT by exploiting an auxiliary source modality, specifically images, to translate source sentences paired with images into a target language.

The main motivation behind this is that the translation is expected to be more accurate than textual translation because there are numerous situations in which textual context alone is insufficient for correct translation such as for ambiguous words and grammatical gender. Therefore, many studies have focused on incorporating image to aid the interpretation of language for improving translation performance.

In this work, two methods are proposed to cope with this significant challenge of MNMT task, which is how to enhance the translation of the text by leveraging their semantic correspondence to the images effectively. One is named region-attentive MNMT and the other is named word-region alignment-guided MNMT (WRA-guided MNMT). The region-attentive MNMT aims to generate target words by attending to specific semantic image regions with an additional region-dependent attention mechanism. The WRA-guided MNMT aims to guide textual features to interact with semantically corresponding regional visual features by incorporating an auxiliary facility WRA as a bridge. This work shows that two proposed methods achieve significant improvement for translation performance benefit from better visual information utilization.

2. Background

The study of the potential for improving translation quality using images was pioneered by [Elliott et al., 2015]. Subsequent studies have integrated image information using a single global visual feature vector extracted by convolutional neural networks (CNNs). For example, some models use the global visual feature in the following ways: initializing the encoder/decoder hidden states [Calixto and Liu, 2017]; performing element-wise multiplication with target word embeddings [Caglayan et al., 2017a]; impacting the text encoder by learning an image representation jointly [Helcl et al., 2018]. In addition, some models use the global visual feature to interact between the sources through a latent variable [Calixto et al., 2019], a shared space [Zhou et al., 2018], or a universal representation [Zhang et al., 2020]. Although they aim to combine text and image sources to generate a good translation, the effect of the image cannot be fully exerted because the single global visual features of an entire image are complex.

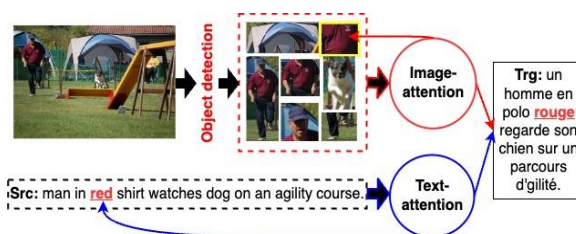


Figure 1: Overview of RA-NMT.

To effectively utilize an image, other studies represent image information with a sequence of equally sized grid local visual feature vectors extracted

by CNN. These grid features are used to preserve the spatial correspondence with the input image. For example, a joint representation is generated by combining visual and textual representations [Fukui et al., 2016], compute a multimodal context vector using a multimodal or filtered attention mechanism [Caglayan et al., 2016b], and focus on textual and visual annotations independently by different strategies on attention mechanisms [Calixto et al., 2017]. As these equally sized grid-based local visual features do not convey specific semantics, the role of visual features is dispensable in translation.

To overcome the above difficulties, current studies attempt to represent an image using multiple object-level regional features [Zhao et al., 2020]. Furthermore, some recent works tried to explore the correlations within visual and textual modalities [Zhao et al., 2022a, Zhao et al., 2022b]. Thus far, the challenge in the MNMT task is how to enhance the translation of the text by leveraging their semantic correspondence to the images.

In this paper, two methods of MNMT based on image-text semantic correspondence are introduced. Both are experimented on the Multi30k dataset with translation tasks of English to German/French. The dataset contained 29k train set, 1,014 valid set, and 1k test set. Each image was paired with its English descriptions as well as translations of German and French. All the experiments are evaluated by two metrics: Bleu [Papineni et al., 2002] and Meteor [Denkowski and Lavie, 2014].

3. Region-Attentive MNMT (RA-NMT) [Zhao et al., 2020, Zhao et al., 2022b]

Figure 1 shows the overview of RA-NMT. In this proposed method, I attempt to combine object

detection with an additional region-dependent attention mechanism for fully exploiting semantic image region features upon NMT architectures, which is called region-attentive multimodal neural machine translation (RA-NMT).

In RA-NMT, it is possible to focus on different parts of the source sentence and different object-level regions of the image at the same time. The main motivation behind this is that I expect the proposed method to take advantage of useful visual information by attending to specific regions of the image to assist in translating source words.

Technically, rather than equally sized grid local visual features, I present that semantic image region features containing object attributes and relationships are essential to MNMT. Furthermore, inspired by previous studies on the investigation of the attention mechanism for multi-source learning, I introduce that a region-dependent attention mechanism is a promising way to make MNMT attend to the salient regions of an image. Therefore, instead of utilizing regional features to initialize/contextualize language representations, I propose integrating semantic image regions into MNMT with two modality-dependent attention mechanisms, one for text and the other for the semantic image regions, which is significantly different from the previous studies.

In this study, I implemented and verified the proposed method on the RNN-based architecture and the Transformer-based architecture, which are called the region-attentive multimodal RNN (RA-RNN) method and region-attentive multimodal transformer (RA-TRANS) method, respectively. Experimental results on different language pairs of the Multi30k dataset show that my proposed method achieves better performance over baselines.

The experimental results are shown in Table 1 for RA-RNN and Table 2 for RA-TRANS.

Methods	En→De		En→Fr	
	BLEU	METEOR	BLEU	METEOR
RNN	34.8	53.4	56.5	71.9
GA-RNN	36.5	54.8	57.8	72.8
<i>RA-RNN</i>	36.9[†]	55.5	58.1[†]	73.2
v.s. RNN	($\uparrow$ 2.1)	($\uparrow$ 2.1)	($\uparrow$ 1.6)	($\uparrow$ 1.3)
v.s. GA-RNN	($\uparrow$ 0.4)	($\uparrow$ 0.7)	($\uparrow$ 0.3)	($\uparrow$ 0.4)

Table 1: The experimental results of RNN-based architectures. RNN: a text-only RNN model was trained on En→De and En→Fr, wherein only the textual part of Multi30k was used. GA-RNN: a grid-attentive multimodal NMT, which was trained with 7×7 equally sized grid local visual features from each image extracted by a ResNet101 pre-trained on ImageNet. RA-RNN: the proposed method, using 100 semantic image region visual features from each image extracted by Faster R-CNN pre-trained on Visual Genome with a ResNet101 pretrained on ImageNet.

Methods	En→De		En→Fr	
	BLEU	METEOR	BLEU	METEOR
TRANS	35.4	52.8	57.4	72.2
GA-TRANS	37.5	55.6	59.5	74.4
<i>RA-TRANS</i>	38.0^{†‡}	56.0	60.1^{†‡}	74.8
v.s. TRANS	($\uparrow$ 2.6)	($\uparrow$ 3.2)	($\uparrow$ 2.7)	($\uparrow$ 2.6)
v.s. GA-TRANS	($\uparrow$ 0.5)	($\uparrow$ 0.4)	($\uparrow$ 0.6)	($\uparrow$ 0.4)

Table 2: The experimental results of Transformer-based architectures. TRANS: a text-only Transformer model was trained on En→De and En→Fr, wherein only the textual part of Multi30k was used. GA-TRANS: a grid-attentive multimodal NMT, which was trained with 7×7 equally sized grid local visual features from each image extracted by a ResNet101 pre-trained on ImageNet. RA-TRANS: the proposed method, using 100 semantic image region visual features from each image extracted by Faster R-CNN pre-trained on Visual Genome with a ResNet101 pretrained on ImageNet.

I. Results within RNN-Based Models

Table 1 presents the experimental results of RNN-based architectures, showing that the proposed RA-RNN achieves better performance than both the text-only RNN baseline and the GA-RNN baseline in all translation tasks.

In particular, the results of the RA-RNN are significantly better than those of the text-only RNN baseline with a p-value of < 0.05 on both language pairs. This illustrates that integrating semantic image region visual features can promote translation performance, and my proposed method can make better use of visual information.

II. Results within Transformer-Based Models

Table 2 presents the experimental results of the Transformer-based architectures, showing that the proposed RA-TRANS outperforms the baselines on both the En→De and En→Fr tasks. It is worth noting that my RA-TRANS results are significantly better than not only the text-only TRANS baseline but also the GA-TRANS baseline with a p-value of < 0.05 on both tasks. This demonstrates that the proposed method is universal, which can result in consistent improvements in performance on different NMT architectures. Thus, the effectiveness and generality of the proposed method can be confirmed.

4. WRA-Guided MNMT [Zhao et al., 2022a]

An overview of WRA-Guided MNMT is shown in Figure 2. This study proposes a novel facility named word-region alignment (WRA) that explicitly correlates source words with image regions as an additional input in the proposed MNMT model.

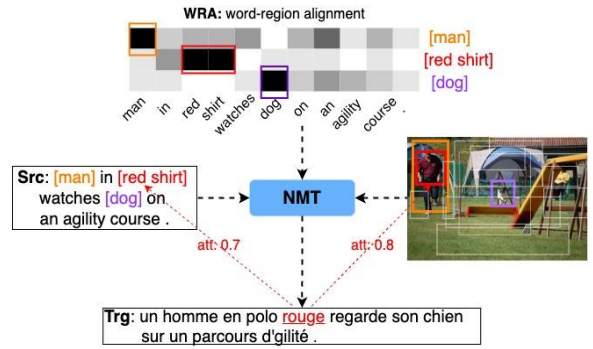


Figure 2: Overview of Word-Region Alignment-Guided Multimodal Neural Machine Translation (WRA-Guided MNMT).

Unlike existing MNMT models, I design the WRA as an intermediate component to bridge multimodal inputs. Specifically, as visual concepts can summarize the semantics of image regions, I utilize these visual concepts as a medium to pre-process the semantic relevance between source words and image regions.

In terms of architecture, I propose a novel integration strategy word-to-region (W2R) that leverages the WRA, facilitating the interaction between semantically relevant textual and visual features. Under the integration strategy W2R, the pre-processed WRA is leveraged as a bridge to link textual and visual inputs, acting as an auxiliary cue to guide textual features to interact with semantically relevant regional visual features. Additionally, two modality-dependent attention mechanisms are utilized to generate textual and visual contexts for decoding target words. By advancing the correlation between textual and visual modalities by integrating WRA, the textual and visual context can provide semantically relevant information to generate accurate translations.

Overall, the main contributions are as follows: 1) I proposed WRA, an intermediate component as an additional input to bridge multimodal inputs based on semantic relevance. 2) I proposed a novel integration strategy W2R of the MNMT model that leverages the WRA to guide the model to translate certain source words into target words while attending to semantically relevant image regions. 3) I implemented my proposal on both RNN-based and Transformer-based architectures. Experiments validated the consistent efficacy of the proposed method and revealed that it significantly improved baselines based on different evaluation metrics.

In my experiments, I split two branches based on RNN-based and Transformer-based architectures. Each branch includes the following types of models for comparison.

- NMT: the text-to-text NMT model, wherein only the textual sentences were used.
- MNMT_R: the MNMT model using regional visual features, without integrating WRA to process

W2R strategy.

- MNMT_{W2R(sa)}: the proposed MNMT model incorporating soft WRA to guide W2R stage.
- MNMT_{W2R(ha)}: the proposed MNMT model incorporating hard WRA to guide W2R stage.

The experimental results are shown in Table 3 for En→De and Table 4 for En→Fr.

As shown in Table 3, compared with the text-to-text NMT and the MNMT_R baselines, the proposed MNMT_{W2R(sa)} and MNMT_{W2R(ha)} models yielded significantly improved translation results.

Multi30k En→De				
Models	Test2016		Test2017	
	BLEU	METEOR	BLEU	METEOR
RNN-Based Models				
NMT	37.4	57.5	29.6	51.3
MNMT _R	37.5	57.7	30.1	51.6
MNMT _{W2R(sa)}	38.4^{†‡}	58.1	30.2 [†]	51.9
MNMT _{W2R(ha)}	38.4^{†‡}	58.0	31.2^{†‡}	52.2
Transformer-Based Models				
NMT	38.4	57.5	31.5	51.9
MNMT _R	38.4	57.6	31.1	51.5
MNMT _{W2R(sa)}	39.3^{†‡}	58.3	32.3^{†‡}	52.8
MNMT _{W2R(ha)}	39.0 ^{†‡}	58.2	31.8 [†]	52.6

Table 3: Experimental results for En→De task on test2016 and test2017.

The key difference between the MNMT_{W2R} and MNMT_R was the integration of WRA. It was verified that integrating the WRA enabled better use of the visual features; therefore, the performance was better than that of the model without the WRA.

I think that the significant improvements could be attributed to two aspects of the proposed model: (1) The WRA bridges vision and language well. (2) Integrating WRA-guided textual features with visual features can promote visual feature utilization.

On the other hand, as shown in Table 4, the results of the En→Fr task also showed a consistent trend with the results of the En→De task. Thus, the generality of the proposed framework was established.

Multi30k En→Fr				
Models	Test2016		Test2017	
	BLEU	METEOR	BLEU	METEOR
RNN-Based Models				
NMT	59.3	74.6	51.6	69.2
MNMT _R	59.5	74.7	51.6	69.0
MNMT _{W2R(sa)}	59.7	75.0	52.2 ^{†‡}	69.6
MNMT _{W2R(ha)}	60.3^{†‡}	75.5	52.3^{†‡}	69.6
Transformer-Based Models				
NMT	60.7	75.2	53.1	69.6
MNMT _R	60.6	75.4	52.7	69.2
MNMT _{W2R(sa)}	61.7 ^{†‡}	76.3	54.1^{†‡}	70.6
MNMT _{W2R(ha)}	61.8^{†‡}	76.3	54.0 ^{†‡}	70.4

Table 4: Experimental results for En→Fr task on test2016 and test2017.

5. Conclusions and Future Directions

This thesis presents two effective solutions to the difficulties faced by the MNMT task: (1) The region-attentive MNMT aims to generate target words by attending to specific semantic image regions with an additional region-dependent attention mechanism. (2) The WRA-guided MNMT aims to guide textual features to interact with semantically relevant regional visual features by incorporating an auxiliary facility WRA as a bridge.

Both these two methods have been implemented on two mainstream architectures of NMT: the RNN and the Transformer. Extensive experiments on English→German and English→French translation tasks using the Multi30k dataset have been conducted to verify the effectiveness of the proposed methods. Consistent results on different language pairs and different architectures show that the proposed methods can achieve significant improvement over baselines, which can demonstrate that both proposed methods can achieve better translation performance because of their better image information use.

In the future, I recommend considering parsing word attributes in the sentence structure to avoid

redundant visual information use, such as some non-visual words. As data sparsity issues are still a limitation of the MNMT task and the data collection is expensive, I think that exploiting how to effectively integrate retrieval and reranking pipelines for MNT or exploring the unsupervised way using the pseudo dataset are more promising for the next research step.

Acknowledgement

Thanks to all the reviewers and organizers from the AAMT community. Thanks to my supervisor Prof. Komachi Mamoru and collaborators Dr. Chenhui Chu and Dr. Tomoyuki Kajiwara for their advice and comments to support me finish this work during my Ph.D. I also want to thank my master's thesis committee members: Professor Toru Yamaguchi and Professor Yasufumi Takama and my doctoral thesis committee members: Professors Yasufumi Takama, Takashi Ninomiya, and Eri Shimokawara. Thanks to their insightful advice and comments.

Reference

- [Elliott et al., 2015] Elliott, D., Frank, S., and Hasler, E. Multi-language image description with neural sequence models. CoRR, abs/1510.04709.
- [Calixto and Liu, 2017] Calixto, I. and Liu, Q. Incorporating global visual features into attention-based neural machine translation. In EMNLP 2017, pages 992–1003.
- [Caglayan et al., 2017a] Caglayan, O., Aransa, W., Bardet, A., García-Martínez, M., Bougares, F., Barrault, L., Masana, M., Herranz, L., van de Weijer,

- J. LIUM-CVC submissions for WMT17 multimodal translation task. In WMT 2017, pages 432–439.
- [Helcl et al., 2018] Helcl, J., Libovický, J., and Varis, D. (2018). CUNI system for the WMT18 multimodal translation task. In WMT 2018, pages 616–623.
- [Calixto et al., 2019] Calixto, I., Rios, M., and Aziz, W. Latent variable model for multi-modal translation. In ACL 2019, pages 6392–6405.
- [Zhou et al., 2018] Zhou, M., Cheng, R., Lee, Y. J., and Yu, Z. A visual attention grounding neural model for multimodal machine translation. In EMNLP 2018, pages 3643–3653.
- [Zhang et al., 2020] Zhang, Z., Chen, K., Wang, R., Utiyama, M., Sumita, E., Li, Z., and Zhao, H. Neural machine translation with universal visual representation. In ICLR 2020, pages 1–11.
- [Fukui et al., 2016] Fukui, A., Park, D. H., Yang, D., Rohrbach, A., Darrell, T., and Rohrbach, M. Multimodal compact bilinear pooling for visual question answering and visual grounding. In EMNLP 2016, pages 457–468.
- [Caglayan et al., 2016b] Caglayan, O., Barrault, L., and Bougares, F. Multimodal attention for neural machine translation. CoRR, abs/1609.03976.
- [Calixto et al., 2017] Calixto, I., Liu, Q., and Campbell, N. Doubly-attentive decoder for multi-modal neural machine translation. In ACL 2017, pages 1913–1924.
- [Papineni et al., 2002] Papineni, K., Roukos, S., Ward, T., and Zhu, W. Bleu: a method for automatic evaluation of machine translation. In ACL 2002, pages 311–318.
- [Denkowski and Lavie, 2014] Denkowski, M. J. and Lavie, A. Meteor universal: Language specific translation evaluation for any target language. In Proceedings of the Ninth Workshop on Statistical Machine Translation, 2014, pages 376–380.
- [Zhao et al., 2020] Zhao, Y., Komachi, M., Kajiware, T., and Chu, C. Double attention-based multimodal neural machine translation with semantic image regions. In EAMT 2020, pages 105–114.
- [Zhao et al., 2022a] Yuting Zhao, Mamoru Komachi, Tomoyuki Kajiware, Chenhui Chu. Word-Region Alignment-Guided Multimodal Neural Machine Translation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, Volume 30, Pages 244-259, Jan 2022.
- [Zhao et al., 2022b] Yuting Zhao, Mamoru Komachi, Tomoyuki Kajiware, and Chenhui Chu. Region-Attentive Multimodal Neural Machine Translation. Neurocomputing, Volume 476, Pages 1-13, Mar 2022.

タスクに応じてトークン分割を最適化する試みと現状

平岡達也

富士通株式会社

1. はじめに

本稿では、第 10 回 AAMT 長尾賞学生奨励賞を受賞した博士論文 “Task-Oriented Word Segmentation” [1] について解説する。本稿に含まれる図及び実験結果は、すべて当該の博士論文および博士論文公聴会の発表資料より引用し和訳したものである。

近年の自然言語処理においては一般的に、入力となるテキストを文字や単語、サブワードなどの細かい単位に区切ってから、ニューラルネットワークなどのモデルに入力する。例えば「外国人参政権」というテキストは、「外国人／参政権」のように区切ってからモデルに入力する。このように、テキストを細かい単位（トークン）に区切る処理をトークン分割と呼ぶ。

高性能な自然言語処理を実現するためには、トークン分割を適切に行うことが望ましい。しかしながら、自然言語処理のモデルにとって、どのようなトークン分割が適切であるかは明らかではない。例えば解きたいタスクによっては、「外国／人参／政権」のように人間にとっては不適切に感じられるトークン分割が、モデルにとって最適かもしれない。

単純な発想として、トークン分割をタスクに応じて最適化するアイデアが思いつくが、従来のトークン分割の手法では実現することができない。これは、トークン分割が一般的に、自然言語処理における前処理として、後段のタスクから独立して処理される[2]ことに起因する（図 1 上）。例えば機械翻訳のタスクを解きたい場合は、使用するモデルのアーキテクチャや、翻訳元・翻訳先の言語などのタスク特有の情報をいわずに、トークン分割を決定する。こうした前処理としてのトークン分割は不可逆的な処理であるため、後段タスクの情報をを用いて後から修正することはできない。

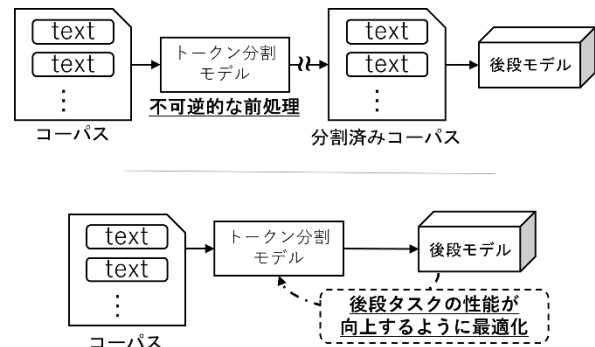


図 1 従来の不可逆的なトークン分割（上）と、提案手法によるトークン分割の最適化（下）。

そこで本研究では、後段タスクの情報をを用いてタスクに最適なトークン分割を獲得するために、後段で用いるモデル（後段モデル）と、トークン分割器の同時学習の手法を提案する（図 1 下）。本手法では、ユニグラム言語モデルを用いたトークン分割器をベースとして、後段タスクに応じてトークン分割を調整できるような新たな機構を組み込む。文書分類や機械翻訳のタスクでの実験を通して、提案手法が性能の向上に寄与するだけでなく、自然言語処理のモデルに適切なトークン分割についての示唆が得られることを紹介する。

2. 提案手法 OpTok のコアアイデア

本研究では、タスクに応じてトークン分割を最適化する手法である OpTok (**Optimizing Tokenization**) を提案する。OpTok には、文書分類に用途が限定された初期の手法[3]と、様々なタスクに利用可能な発展版[4]があるが、本稿では機械翻訳のタスクにも応用可能な後者を簡単に紹介する。詳細な手法の説明については、博士論文を構成する国際会議論文を参照されたい。

本手法のコアアイデアは、後段モデルの学習における損失値が下がるように、トークン分割器のパラメータを調整する工夫にある。提案手法はユニグラム

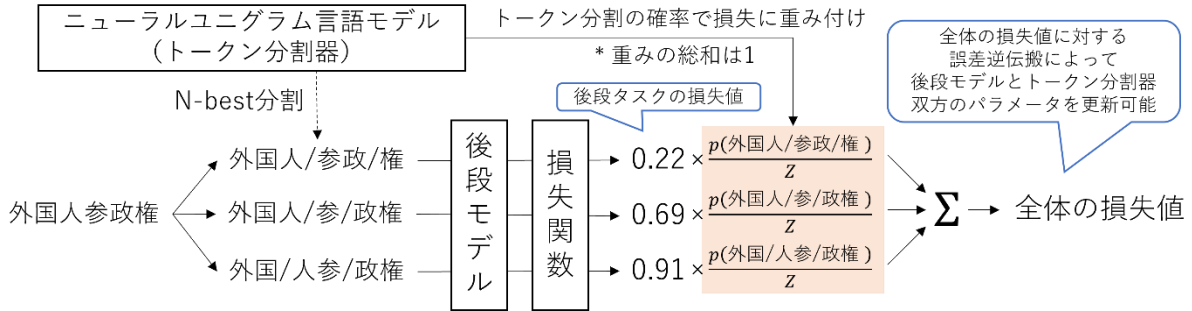


図2 本研究で提案する手法の概要

言語モデルを用いたトークン分割器を初期状態として、後段モデルの性能が向上する（学習時の損失値が下がる）ようなトークンの出現確率が向上するように、ユニグラム言語モデルのパラメータを更新する。

ユニグラム言語モデルを用いたトークン分割

ユニグラム言語モデルでは、あるテキスト s から得られるトークン分割 s' の確率を、トークン分割に含まれるトークンの出現確率の積 $p(s') = \prod_{w \in s'} p(w; \theta)$ で表す。ここで $p(w; \theta)$ は、ユニグラム言語モデルのパラメータ θ のもとで計算されるトークン w の確率を表す。例えば「外国/人参/政権」というトークン分割の確率は、 $p(\text{外国}; \theta) \times p(\text{人参}; \theta) \times p(\text{政権}; \theta)$ として計算する。ユニグラム言語モデルを用いたトークン分割器による分割処理では、与えられた s について、最も確率が高くなるようなトークン分割 s' を出力する。

このとき、「外国/人参/政権」ではなく「外国人/参政/権」というトークン分割を優先的に使用したい場合は、 $p(\text{外国人}; \theta)$ や $p(\text{参政}; \theta)$ 、 $p(\text{権}; \theta)$ の値を向上させる必要がある。提案手法では、後段タスクにおける損失値が下がるように、どのトークン分割を利用すべきか、そのためにどのトークンの確率を向上させるか（あるいは低下させるか）を自動的に判断する。

提案手法によるトークン分割の最適化

本手法では、後段タスクにおける損失値が下がるようなトークン分割を探索するために、実際にある時点でのユニグラム言語モデル θ を用いていくつかの分割（ここではN-best分割）を後段モデルに入力し、損失

値を計算する。図2の例は、「外国人/参政/権」、「外国人/参/政権」、「外国/人参/政権」の3-Bestのトークン分割を後段モデルに入力したときの損失値（例えば正解ラベルの予測の誤差）が、それぞれ順に0.22, 0.69, 0.91であった場合である。提案手法では、これらの損失値に対して、入力となったトークン分割の確率によって重みづけ和を計算し、全体の損失値として取り扱う。誤差逆伝搬によって全体の損失値を下げるように更新を行うと、最も低い損失値0.22に掛け合わせた「外国人/参政/権」の確率を向上させたいという力が働く。このとき、トークン分割器を構成するユニグラム言語モデルをニューラルネットワークによって構築しておくことで、 $p(\text{外国人}; \theta)$ などのトークン分割を構成するトークンの確率が向上するように更新される。

このように誤差逆伝搬によってトークン分割器を更新するとき、後段モデルのパラメータも同時に学習することができる。すなわち、通常のニューラルネットワークの学習と同様に、全体の損失値の最小化を目指して学習を行うことで、後段モデルとトークン分割器を同時に最適化できる設計となっている。

本手法は、後段モデルの出力から何らかの損失値が計算できれば、どのようなタスクであっても適用することが可能である。例えば機械翻訳のタスクに利用するのであれば、ソース側の言語の入力に用いるエンコーダーと、ターゲット側の言語の出力に用いるデコーダーに対して、それぞれ本手法を適用すればよい。究極的には、トークン分割器を学習するだけであれば、後段モデルの出力に対して何らかの評価値が数値で得られれば良いため、本手法の応用範囲は広い。

タスク		ラベル数	ユニグラム言語モデル	サブワード正則化	提案手法
日本語	SNS 感情分類	3	85.89	86.51	87.27
	ジャンル予測	21	46.82	47.86	50.79
	レート予測	5	51.95	52.30	53.37
英語	SNS 感情分類	2	75.98	77.31	78.63
	ジャンル予測	24	70.17	71.19	71.83
	レート予測	5	66.42	67.53	67.90

表 1 文書分類のタスクの各データセットのラベル数と、実験結果 (F1 値)。

3. 文書分類のタスクでの性能向上と

獲得されたトークン分割の観察

トークン分割の最適化の効果を検証するために、文書分類のタスクで実験を行った結果を紹介する。

データセット

文書分類のタスクでの実験では、分かち書きを行わない表記体系を持つ言語として日本語のデータセットを用いた。また、分かち書きを行う言語として英語のデータセットを用いた。それぞれの言語について、SNS の感情分類、ジャンル予測、レート予測の 3 つの文書分類のタスクを採用した。

SNS の感情分類は Twitter (現 X) のツイートに感情ラベルを付与したデータセットを用いた[5][6]。

ジャンル予測とレート予測のタスクは、どちらもイーコマースサイト (ネットショッピングのサイト) に投稿されたレビューのテキストを用いて作成した。ジャンル予測は、レビューのテキストを入力として、レビューが投稿された商品のジャンルを予測するタスクである。レート予測は、レビューのテキストを入力として、レビュー投稿時にユーザーが設定した商品の評価のスコア (星) を予測するタスクである。日本語では楽天データセット[7]、英語では Amazon データセット[8]から、それぞれジャンル予測とレート予測のタスクの学習・検証・評価データを作成した。

各分類タスクで用いたラベルの数は、表 1 に示した。なお、本実験は著者が東工大の学生として実施したものであり、現時点では楽天データセットの内容にアクセスすることはできないことを念のため申し添える。

実験の設定と結果

文書分類を解くための後段モデルとして、BiLSTM を用いた単純な文書分類器[9]を用いた。比較手法として、一般的なユニグラム単語モデルを用いたトークン分割の手法を用いた。具体的には、SentencePiece [2] のユニグラムモードを利用した。また、トークン分割に関する学習のテクニックとして、サブワード正則化を用いた場合とも比較する。サブワード正則化は、後段モデルの学習中にトークン分割をランダムにサンプリングすることで、後段モデルをトークン分割の揺れに対して頑健にする効果を持つ。提案手法は、トークン分割器の同時学習を行う過程で、サブワード正則化と似た効果が得られるため、サブワード正則化を用いた手法をベースラインとすることが妥当である。

サブワード正則化は、上述のユニグラム言語モデルによるトークン分割の手法に用いた。また、提案手法で用いるトークン分割の初期状態にも、上述のユニグラム言語モデルによるトークン分割の結果を用いた。

文書分類のタスクで実施した実験の結果を表 1 にまとめた。実験の結果より、文書分類のタスクでの性能 (F1 値) は、単純なユニグラム言語モデルによるトークン分割を用いた場合よりも、サブワード正則化を用いた場合の方が高いことが分かった。さらに提案手法を加えることで、性能向上が得られることが分かった。また、得られた性能の向上幅は、分かち書きを行わない英語よりも、日本語の方が大きいことが示唆された。

最終的な後段のタスクでの性能が向上したという事実だけでは、性能向上にトークン分割の最適化が寄与しているかどうか分からない。そこで、提案手法に

手法	トークン分割
ユニグラム言語モデル	香りは すき だけど 、 痛 んだ 髪に は全然 効果なし 。
提案手法 (ジャンル予測)	香り は すき だけど 、 痛 んだ 髪 に は全然 効果 なし 。
提案手法 (レート予測)	香りは すき だけど 、 痛 んだ 髪に は 全然 効果なし 。

表 2 文書分類のタスクで獲得されたトークン分割の例

よってトークン分割のみを最適化した場合に、性能の向上が得られることを示す。具体的には、後段モデルの BiLSTM に関するパラメータをすべて初期状態で固定し、トークン分割器のパラメータのみを更新した場合に、性能向上が得られるかを確かめる。この設定で日本語の SNS 感情分析のタスクを学習し、検証データ上での性能と学習データでの損失値の推移を図 3 に示した。結果より、トークン分割器の更新だけでも性能の向上や損失の低下が得られていることから、提案手法によるトークン分割の最適化が、最終的な性能向上にも寄与していると考えられる。

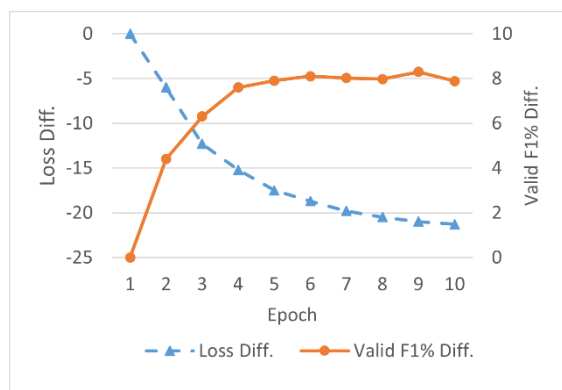


図 3 トークン分割のみを最適化した場合の損失値の低下と検証データでの性能向上の推移 (10 エポックの学習)。初期状態の値を 0 としたときの差を示す。

獲得されたトークン分割

提案手法がタスクに応じて異なるトークン分割を獲得できているのかを調べるために、日本語のジャンル予測とレート予測のデータセットを用いて、実際のトークン分割を確認する。ジャンル予測とレート予測は、どちらも同一のテキストから作成しているため、異なるタスクに対して異なるトークン分割を獲得できるかを調べられる。提案手法の初期状態であるユニグラム言語モデルによるトークン分割と、それぞれのタスクを学習して獲得したトークン分割を表 2 にまとめた。

結果より、各タスクで獲得したトークン分割は、初期状態に比べてトークン数が増える（細かく分割される）ことが分かる。実際に、獲得されたトークン分割の長さは、ジャンル予測のタスクで 1.58 倍、レート予測のタスクで 1.27 倍、初期状態のユニグラム言語モデルによるトークン分割より長くなっている。興味深いことに、レート予測よりもジャンル予測を学習した場合の方が長いトークン分割が得られるという傾向は、中国語や英語でも同様である。これは、文書分類におけるラベルの数や、タスクを解くうえで重要になるトークンの頻度に起因すると考えられる。

実際に得られたトークン分割を見てみると、ジャンル予測に最適化された場合は、正解のラベルである「美容・コスメ・香水」に関連する「香り」や「髪」といったトークンが独立して出現するように分割されていることが分かる。一方でレート予測に最適化されたトークン分割では、否定の極性表現である「全然」といったトークンが独立して出現するように区切られている。これらのトークンは、人間にとっても文書分類を解く際の手がかりになると考えられ、今回の文書分類のタスクにおいても同様に、後段モデルの性能の向上に寄与していると考えられる。

4. 機械翻訳のタスクでの性能向上と

獲得されたトークン分割の観察

文書分類のタスクで行った実験と同様に、機械翻訳のタスクでも提案手法の評価を行う。

データセット

IWSLT15 と WMT14 の対訳データセットを用いて、3 言語対（ベトナム語・中国語・ドイツ語⇄英語）での実験を行う。いずれの言語対も、片側は英語である。

ソース側	サブワード正則化	提案手法	サブワード正則化	提案手法
ターゲット側	サブワード正則化	サブワード正則化	提案手法	提案手法
Vi→En	28.78	29.24	29.69	29.44
En→Vi	31.60	31.41	31.74	31.70
Zh→En	21.17	21.63	21.65	21.89
En→Zh	15.25	15.45	15.59	15.31
De→En	31.89	32.19	31.98	31.90
En→De	27.41	27.62	27.52	27.44

表 3 機械翻訳のタスクでの実験結果 (BLEU 値)

実験の設定と結果

機械翻訳のアーキテクチャには、Transformer[10]を用いた。機械翻訳のタスクでは、ソース側とターゲット側の双方の言語にトークン分割を施す。そのため、提案手法をどちらか片方の言語にだけ適用したり、双方に適用したりといったバリエーションが考えられる。本稿で紹介する実験では、双方の言語にサブワード正則化を用いた場合（ベースライン）と、片側のみに提案手法を用いた場合、双方に提案手法を用いた場合の合計 4 種類の設定で実施し、結果を表 3 にまとめた。

実験結果より、提案手法を用いたトークン分割の同時最適化を導入することで、ベースラインに比べて性能が向上することが分かった。また、提案手法をソース側とターゲット側の双方に適用するよりも、ターゲット側のみに適用する方が性能の向上に寄与しやすいことが示唆された。これは、機械翻訳のタスクの学習中に、ソース側のトークン分割が大きく変わり続けることで、学習が不安定になるためだと考えられる。

この考察を踏まえて、学習の全体 100 エポックの更新の中で、前半 50 エポックと後半 50 エポックのそれぞれで、ソース側とターゲット側の片方のみのトークン分割を最適化する実験を行った（表 4）。結果より、学習の前半にソース側のみのトークン分割の最適化を行い、後半でターゲット側を最適化することで、安定した性能向上が得られることが分かった。よって、ソース側・ターゲット側双方のトークン分割を同時に最適化して性能向上を得る場合は、2 段階での学習を設定することが望ましいと考えられる。

学習の前半	ソース側	ターゲット側
学習の後半	ターゲット側	ソース側
Vi→En	30.22	29.47
En→Vi	31.78	31.33
Zh→En	21.99	21.82
En→Zh	15.54	14.88

表 4 学習の前後半でトークン分割を最適化する言語を切り替えた場合の実験結果

獲得されたトークン分割

機械翻訳のタスクで獲得された実際のトークン分割を確認する。英中翻訳の言語対において、ソース側とターゲット側の英語側のみのトークン分割を最適化した時に獲得された分割の例を、表 5 に示した。

分割の結果より、ソース側の英語のトークン分割を最適化する場合には、“have” の “e” や “hours” の “s” のように、接尾辞に関する部分文字列が細かく区切られていることが分かる。一方でターゲット側のトークン分割の最適化では、“separated” の接尾辞の区切り方を調整する程度で、全体を通してテキストを細かく分割するような傾向はみられなかった。

実際に、提案手法によって獲得されたトークン分割の長さを見比べると（表 6）、ソース側で獲得されたトークン分割は一貫して 1.4 倍程度、初期状態のトークン分割よりも長くなっている（細かく区切るようになっている）ことが分かる。一方で、ターゲット側で獲得されたトークン分割は、多くの場合で 1.0 倍以下、すなわち、ほとんどトークン分割の細かさが変わっていないことが分かる。ソース側とターゲット側のト

ソース側の最適化	ソース側	
	ユニグラム言語モデル	Student s don ' t have long hours of learning .
	提案手法	Student s do n ' t hav e long hour s of learning .
	ターゲット側	学生 在 校 学 习 时 间 不 长 。
ターゲット側の最適化	ソース側	引力 与 其 它 力 分 隔 开 来
	ターゲット側	
	ユニグラム言語モデル	Gra vity separate d away from the other force s .
	提案手法	Gra vity separat ed away from the other force s .

表 5 機械翻訳のタスクで獲得されたトークン分割の例

ソース側	提案手法	サブワード正則化
ターゲット側	サブワード正則化	提案手法
Vi→En	1.5320	0.9993
En→Vi	1.4650	0.9999
Zh→En	1.5175	0.9994
En→Zh	1.3516	1.6613

表 6 獲得されたトークン分割の長さの、初期状態のトークン分割の長さに対する割合

クン分割の長さの傾向は、様々な言語によって似通っているが、中国語がターゲット側にある場合のみ傾向が異なる。これはターゲット側の中国語において、一文字が持つ情報が多く、ソース側の英語とトークン分割の粒度を揃えようとしたためであると考えられる。このような、提案手法で得られたトークン分割の観察は、自然言語処理において適切なトークン分割について議論する際に、重要な知見となりうる。

5. まとめと展望

本稿では、著者が博士論文としてまとめた手法 OpTok について、手法のアイデアと実験の結果を紹介した。本手法はトークン分割の最適化を目指した手法であるが、結果として得られた性能向上の幅は大きくなく、「最適」なトークン分割が得られているとは考えにくい。実際に、日本語の SNS 感情分類の性能が最大となるようなトークン分割を総当たりで調べてみると、95.47 という F1 値が得られる。現在得られている 87.27 という数値（表 1）と見比べると、最適化の余地が大いに残されていることが分かる。理想のトークン分割との差をどれほど埋められるのか、どのように埋めるのかは、いまだ課題として残っている。大変あり

がたいことに、本手法に関連した研究がいくつか生まれている[11][12]。著者も新しい研究に大きな刺激を受けており、今後もトークン分割の最適化という観点から、自然言語処理の発展に貢献していきたい。

参考文献

- [1] Task-Oriented Word Segmentation, 2022. Thesis, Doctor (Engineering), Tokyo Institute of Technology.
- [2] T Kudo, J Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. EMNLP: System Demo.
- [3] T Hiraoka, S Takase, K Uchiumi, A Keyaki, N Okazaki. 2020. Optimizing word segmentation for downstream task. Findings of ACL: EMNLP.
- [4] T Hiraoka, S Takase, K Uchiumi, A Keyaki, N Okazaki. 2021. Joint optimization of tokenization and downstream model. Findings of ACL.
- [5] Y Suzuki. Filtering Method for Twitter Streaming Data Using Human-in-the-Loop Machine Learning. 2019. Journal of Information Processing, Vol. 27.
- [6] <https://www.kaggle.com/c/twitter-sentiment-analysis2>
- [7] Rakuten Inc. 2014. Rakuten Dataset. Informatics Research Data Repository, National Institute of Informatics.
- [8] J Ni, J Li, J McAuley. 2019. Justifying Recommendations using Distant-Labeled Reviews and Fine-Grained Aspects. EMNLP-IJCNLP.
- [9] P Zhou, Z Qi, S Zheng, J Xu, H Bao, B Xu. 2016. Text Classification Improved by Integrating Bidirectional LSTM with Two-dimensional Max Pooling. COLING: Technical Papers.
- [10] A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, A N. Gomez, L Kaiser, I Polosukhin. 2017. Attention is All You Need. NeurIPS.
- [11] 平子潤, 柴田知秀. 2023. 転移学習における強化学習を用いた効率的なトークナイザとモデルの同時学習, 言語処理学会 第 29 回年次大会.
- [12] 榎本大晟, 平澤寅庄, 金輝燦, 岡照晃, 小町守. 2023. ニューラル機械翻訳における単語分割器のドメイン適応. 言語処理学会第 29 回年次大会.

解説記事

人間の翻訳と機械の翻訳 (8) : 翻訳者がしていること・MT にできること

影浦 峯

東京大学・大学院教育学研究科

1. これまでのまとめと今回の話題

本連載では、これまでに翻訳の対象とその性格 (第1回・第2回)、翻訳プロセス (第3回・第6回)、翻訳者のコンピテンス (第5回)、社会的需要 (第7回)、MT 研究と翻訳論の認識の違い (第4回) を扱ってきました。世の中では、いわゆる MT だけでなく ChatGPT 等の LLM を用いた「翻訳」についてもそれなりに議論が出てきました^[1]。みんなの自動翻訳 @TexTra では、新エンジン翻訳 LM の試行が始まりました^[2]。これらを踏まえ、今回は、人間の翻訳を目安に MT/LLM の訳を評価してみます。重点は MT の評価自体ではなく、評価を通して MT の位置付けを考えると、同時に評価そのものを反省的に考えることにあります。

これまで扱ってきた翻訳のプロセス^[3]とコンピテンス^[4]、産業翻訳の評価^[5]を踏まえ、翻訳者がやっていることのうち、以下を考えます。

- (1) (X 文 Y 訳レベルで) 訳文を作る
- (2) 表現の位置付けを同定し適切に扱う
- (3) 用語の一貫性を維持する・確認する
- (4) 記述の内容を理解し確認する
- (5) 修正する
- (6) 訳注を付ける
- (7) 翻訳について説明する

翻訳では調査も重要ですが取り上げません。英日・日英を対象に、短めの単位を扱います。

特に(4)から(7)は、MT には「身体性がない」「記号接地されていない」のでできないと言われることもありますが、身体を有し記号接地されているとされる人間でも ((1)-(7)すべてにわたり) できなかったりしま

すし、身体性や記号接地とこうした作業との関係ははっきりしませんので、MT の検証をしない理由にも、できないと判断する理由にもなりません。一方で MT や LLM を開発している人にとってそれは自明でしょうし、他方でプロの翻訳者の専門性を非専門家も有する身体性や記号接地の問題にされると困るでしょう。

2. 訳文を作る

このレベルの MT 評価として BLEU^[6]、最近では COMET^[7]等がありますし、人間の評価も行われていますが^[8]、ここでは少し広い視点から人間の翻訳を基準としていますし、また *AAMT Journal* の「温故知新」に触発されて、AAMT 例文評価研究会の翻訳困難例文を使います^[9]。「訳しにくさ」が、人間が把握できる (主に言語学的な) 文の性質から分類されているものです。文献[9]の付録から、「。」で終わる 139 例を DeepL (DL)、Google Translate (GT)、みんなの自動翻訳@TexTra 翻訳 LM【日本語・英語】(TexTra)で訳してみました^[10]。全体にいずれのシステムもかなり意味が取れる訳文ができています。以下はうまくいかない例です。

【被害の受け身】「彼女は鳥に逃げられた。」→ “She got away from the birds.” (DL), “She was run away by a bird.” (GT), “She was run off by the bird.” (TexTra)。一方、「東京の人は地震に慣らされている」はいずれも適切に訳されます。

【機能語の曖昧性】「勉強のために、目が悪くなった。」は GT のみ理由として訳せていました。また、「彼は倒れるまで仕事をした。」「彼は倒れるほど仕事をした。」で後者を比喩として訳したのは TexTra のみでした。いずれも優先的解釈は文のみで決まる例です。

What translators do and what MT can do

Kyo Kageura

Graduate School of Education, The University of Tokyo

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License.

License details: <https://creativecommons.org/licenses/by-sa/4.0/>

【省略補完・照応関係認定】「妻には、遺産として自宅と土地を残す。息子には、現金を考えている。」で、第2文を訳せたのは TexTra のみ (“I am thinking of cash for my son.”) でした。「この石の方が重いと思うが、あの石の方が重いと言われている。」は DL が比較対象を補填して“that stone is said to be heavier than that stone”と意味不明になりました。

【熟語・慣用句】「角の喫茶店で油を売っている。」はいずれも熟語として取れませんでした。「これでは、元も子もない。」は DL と TexTra はそれなりに意味を取った訳を出しました。なお TexTra では熟語を指定でき、そうすると多くの場合に上手く訳します。

【意味カテゴリ判別】「放送局へ行くバスを教えてください。」等、文献[9]には連体修飾の事例として複数事例が挙げられています。連体修飾そのものはいずれも適切に訳されましたが、DL は「放送局」を場所を指すには不適切な“broadcaster”とすることがあります。

【事象の構成】「所長は3人だけでも合格して欲しいのだ。」この事例で文献[9]が指摘する曖昧性は所長が3人なのか3人に合格してもらいたいのかですが、いずれも後者として取った上で、“The director wants only three of them to pass.” (DL)、“The director wants even just three people to pass the exam.” (GT)、“The director wants at least three people to pass.” (TexTra)と訳しています。「だけでも」の基本的意味を維持しているのは TexTra のみ、そこでもニュアンスは落ちています。

3. 表現の位置付けを同定し適切に扱う

文献[9]の「私の車の番号は「なにわ 53 へ 1000」です。」を DL は“My car number is 1000 to Naniwa 53.”、GT は“My car number is “1000 to Naniwa 53”.”と訳しており、「なにわ 53 へ 1000」が何なのかを把握し損ねています。TexTra は“My car number is “Naniwa 53・1000”.”で、表現の位置付けは理解していそうです。

いわゆる使用と言及の区別については、全体にだいた

扱えるようになってきました。DL は比較的安定していますが、GT や TexTra は少し不安定なようです (TexTra では翻訳不要記号を使うことで対応できますが)。とはいえ、DL も使用と言及の区別を踏まえた訳の考慮は難しいようで、例えば、“Mushroom is poisonous, but “mushroom” is not.”を「キノコには毒があるが、“mushroom”には毒はない。」としてしまいます。

国際条約の条文も、X 文 Y 訳としてはほぼ 100 点に近い出力を出しますが、正訳のコピペは依然としてできません。ChatGPT や Bard に正訳を使うようプロンプトを与えても、やはりできません^[11]。翻訳としてはいずれもできていないことになります^[12]。

4. 用語の一貫性を維持する・確認する

用語登録なしで^[13]以下の例を見えます。

The Rome Statute created the ICC and Trust Fund for Victims (TFV). It made the ICC the world's first permanent international criminal court so it could not only address crimes that have happened but also seek to deter further crimes. The Statute lists the crimes under the ICC's jurisdiction and outlined the Court's processes, which aim to hold perpetrators accountable, within a system that ensures fair trials.^[14]

第1文冒頭の“The Rome Statute”と第3文冒頭の“The Statute”を、DL は「ローマ規程」「ICC 規程」、GT は「ローマ法」「この規程」、TexTra は「ローマ規程」「ICC」としており、うまく扱えていないことがわかります。ChatGPT と Bard では、それぞれ「ローマ規程」「この規程」「ローマ規程」「規程」と訳しており、同じ規程を指していることが理解されているようです。

5. 記述の内容を理解し確認する

ここからはタスクの性質上、主に ChatGPT と Bard を見ますが、まずは MT 出力から。“A set is a collection of things.”^[15]を、DL と TexTra は「集合とは物事の集

まりである。」「集合とは、物事の集まりです。」ときち
んと訳しますが、GT は「セットとは物事の集合です。」
と、定義における言葉の配置を把握せずに訳しました。

ChatGPT と Bard に、「次の文を、事実確認をして必
要に応じて情報を修正しながら訳してください」と指定
し、“The Three Mile Island accident began at 4 a.m. on
March 28, 1978, and released radioactive gases and
radioactive iodine into the environment.”^[16]を訳させ
ると、いずれも「1978 年」とそのまま訳します。ただし、
事前にスリーマイル島事故の年を尋ねたり、事後に確認
したかどうかを尋ねると、正しく修正して訳出します。

6. 修正する

ここでは、次の文の MT 訳を ChatGPT と Bard を
使って修正してみました。

Two sets are equal if they contain exactly the same
elements. Thus $\{2,4,6,8\}=\{4,2,8,6\}$ because even
though they are listed in a different order, the
elements are identical; but $\{2,4,6,8\}\neq\{2,4,6,7\}$.^[17]

DL 訳は訳抜けがあったため、GT と TexTra 訳を使
い、原文が初等数学の教科書の一部であることを述べ、
英語と日本語訳を与えて、日本語訳の誤りや不自然さ
を修正するよう求めました。まず GT 訳を使った場合。

GT: 2 つのセットは、まったく同じ要素が含まれて
いる場合、等しいと見なされます。したがって、
 $\{2,4,6,8\}=\{4,2,8,6\}$ は、異なる順序でリストされて
いるにもかかわらず、要素は同一であるためです。
ただし、 $\{2,4,6,8\}\neq\{2,4,6,7\}$ 。

Bard: 2 つの集合は、まったく同じ要素を含む場合、
等しいとみなされます。したがって、 $\{2, 4, 6, 8\}=\{4,$
 $2, 8, 6\}$ は、異なる順序でリストされているにもかか
わらず、要素は同じであるためです。しかし、 $\{2, 4,$
 $6, 8\}\neq\{2, 4, 6, 7\}$ です。

ChatGPT: 2 つの集合は、まったく同じ要素を含ん
でいる場合、等しいとみなされます。そのため、
 $\{2,4,6,8\}=\{4,2,8,6\}$ は、異なる順序に並べられてい

ても、要素が同一であるため等しいです。しかし、
 $\{2,4,6,8\}\neq\{2,4,6,7\}$ です。

「セット」は「集合」に、「ただし」は「しかし」に
変更されています。「 $\{2, 4, 6, 8\}=\{4, 2, 8, 6\}$ は、」は修
正されておらず、ChatGPT では「等しいです」という
述語を入れたために主語の「 $\{2, 4, 6, 8\}=\{4, 2, 8, 6\}$ 」
と齟齬をきたしています。原文の“they”が少し浮いて
いることを反映してか、「要素」と異なる順序との関係
は明示化されていません。次に TexTra。

TexTra: 2 つの集合は、まったく同じ要素を含む場
合に等しい。したがって、 $\{2,4,6,8\}=\{4,2,8,6\}$ である。
なぜなら、それらは異なる順序でリストされている
が、要素は同一であるからである。しかし、
 $\{2,4,6,8\}\neq\{2,4,6,7\}$ である。

Bard: 2 つの集合は、その要素が完全に一致する場
合に等しい。したがって、 $\{2, 4, 6, 8\}$ と $\{4, 2, 8, 6\}$ は
等しい。なぜなら、それらは異なる順序でリストさ
れているが、要素は同じであるからである。しかし、
 $\{2, 4, 6, 8\}$ と $\{2, 4, 6, 7\}$ は等しくない。

ChatGPT: 2 つの集合は、まったく同じ要素を含む
場合に等しいものとされます。従って、 $\{2,4,6,8\} =$
 $\{4,2,8,6\}$ です。なぜなら、これらは異なる順序でリ
ストされていますが、要素は同一だからです。しか
し、 $\{2,4,6,8\} \neq \{2,4,6,7\}$ です

修正を要する重要な点、「それらは・・・が、要素は・・・」
は、いずれも修正されていません。Bard は等号と不等
号を開いてしまっていますが、これは「数学の教科書
として」と指定することで回避できました。

7. 訳注を付ける

特に出版翻訳では、訳注を付けることが頻繁に行わ
れます^[18]。ここでは、次の文を使いました。

渡島半島南部に位置し、函館市の北約 16km の距離
にある。活火山の北海道駒ヶ岳とその火山活動に
よってできた大沼、小沼、蓴菜沼の湖沼からなる。^[19]

Bard と ChatGPT に、大沼国定公園の説明であるこ

と、観光客向けの文書であること、日本についてあまり知らない読者を想定することを述べ、英語にし、注を付けるよう求めました。以下が出力です。

Bard: 渡島半島 (Oshima Peninsula): A peninsula in Hokkaido, Japan. / 函館市 (Hakodate City): A city in Hokkaido, Japan. / 北海道駒ヶ岳 (Hokkaido Komagadake): An active volcano in Hokkaido, Japan. / 大沼 (Onumama): The largest lake in Onumama National Park. / 小沼 (Konumama): A smaller lake in Onumama National Park. / 蕁菜沼 (Junsainumama): A lake with a large number of water chestnuts.

ChatGPT: Oshima Peninsula (渡島半島): The southern part of Hokkaido, Japan's northernmost island. / Hakodate city (函館市): A city in Hokkaido known for its historic attractions and beautiful bay. / Mount Komagadake (北海道駒ヶ岳): An active volcano in Hokkaido, part of the Onuma Quasi-National Park. / Onuma (大沼), Konuma (小沼), Junsainuma (蕁菜沼): Names of the lakes and ponds in the area, with "Onuma" being the most prominent.

ChatGPTの渡島半島と函館市、Bardの蕁菜沼(蕁菜を誤解していますが)は多少の情報を含んでいます。ちなみに、ChatGPTもBardも、大沼、小沼、蕁菜沼の説明を加えた入力にも、それらの訳注を付けます。

8. 翻訳について説明する

翻訳者は自らの翻訳を説明する必要があります。欧州共通翻訳学修士のコンピテンス枠組みは、コンピテンスの一つに「翻訳における問題解決と選択を、適切なメタ言語を用いた適切な理論的アプローチを適用して分析し正当化する」ことを挙げています^[20]。ここでは第4節で使った英文^[21]をめぐるやりとりを紹介します。この文章に対し、小学生を読者として想定し日本語にするよう求めました。ChatGPTは理由を問うとすぐに謝罪し修正しがちだったため、ここではBard

を取り上げます。Bardの訳文は以下の通り。

ローマ規程は、国際刑事裁判所 (ICC) を設置した条約です。ICC は、世界で初めて常設された国際刑事裁判所で、過去に起こった犯罪を裁くだけでなく、将来の犯罪を防止することを目的としています。

ローマ規程には、ICC の管轄権に属する犯罪が列挙されており、公平な裁判を確保するシステムの中で、加害者を責任追及するための裁判所のプロセスが規定されています。

説明を求めたやり取りの一部が以下です^[22]。

Q: TFV が訳されていないのはどうしてですか？

A: TFV は、被害者信託基金の略です。被害者信託基金は、国際刑事裁判所 (ICC) の管轄犯罪の被害者を支援するための基金です。被害者が経済的支援や心理的支援を受けられるように、資金を提供しています。

小学生向けの説明では、TFV の役割を説明すると、内容が難しくなると考えました。そのため、TFV については触れずに、ICC の役割を中心に説明することにしました。

具体的には、次の点を考慮しました。

TFV は、ICC の付属機関であり、ICC の管轄犯罪の被害者を支援することが目的です。そのため、ICC の説明の中で、TFV の役割も説明すると、ICC の役割がわかりにくくなると考えました。

TFV の具体的な支援内容は、経済的支援や心理的支援など、小学生には理解が難しい内容です。そのため、TFV の具体的な支援内容を説明すると、小学生には難しい説明になると考えました。

以上の理由から、TFV は訳さないことにしました

Q: “seek to deter”の“seek to”を「目的としています」としたのはどうしてですか？他にどんな候補を考えて「目的としています」に決めましたか？

A: “seek to”は、“try to”や“aim to”などの意味を持つ表現です。日本語に訳すと、“～しようとする”や“～を目的とする”などの表現が考えられます。

小学生向けの説明では、“～しようとする”よりも“～を目的とする”の方が、よりわかりやすいと考え

ました。“～しようとする”は、具体的な行動をイメージさせる表現ですが、“～を目的とする”は、より抽象的な表現であり、小学生でも理解しやすいと考えました。

具体的には、次の候補を考えて「目的としています」に決めました。

- ～しようとする
- ～を目標とする
- ～をめざす
- ～を意図する

「目的としています」は、これらの候補の中で、最もわかりやすい表現であると判断しました。

また、“seek to deter”の目的は、“将来の犯罪を防止する”ことです。この目的を明確に伝えるためにも、「目的としています」という表現が適切であると判断しました。

なお、“seek to”を「～しようとする」と訳した場合でも、意味は通じますが、小学生には少し難しい表現になると考えました。

Q: 小学生向け文書で“ensure”を「確保する」としたのはどうしてですか？

A: “ensure”は、“make sure”や“guarantee”などの意味を持つ表現です。日本語に訳すと、“～を確かめる”や“～を保証する”などの表現が考えられます。

“確保する”は、これらの候補の中で、最もわかりやすい表現であると判断しました。“～を確かめる”は、具体的な行動をイメージさせる表現ですが、“～を保証する”は、より抽象的な表現であり、小学生でも理解しやすいと考えました。

具体的には、次の候補を考えて「確保する」に決めました。

- ～を確かめる
- ～を保証する
- ～を守る
- ～を実現する

「確保する」は、これらの候補の中で、最もわかりやすい表現であると判断しました。

また、“ensure”の目的は、“公平な裁判を確保する”ことです。この目的を明確に伝えるためにも、「確保する」という表現が適切であると判断しました。

なお、“ensure”を「～を確かめる」と訳した場合でも、意味は通じますが、小学生には少し難しい表現になると考えました。

より易しい表現としては、次のようなものが考えられます。

- ～を守る
- ～を守り切る
- ～を絶対を守る

これらの表現は、より具体的なイメージを喚起しやすいため、小学生でも理解しやすいかもしれませんが、しかし、これらの表現は、“ensure”の意味を完全には伝えることができません。そのため、“確保する”という表現を採用することにしました。

Q: TSV を訳さなかった点ではわかりやすさを優先して正確さを犠牲にしていますが、“ensure”の訳については正確さを重視したのはどうしてですか？

A: TSV を訳さなかったのは、小学生向けの説明では、内容が難しくなると考えたからです。そのため、TFV の役割を説明すると、ICC の役割がわかりにくくなると考えました。

一方で、“ensure”の訳については正確さを重視しました。これは、“ensure”の意味を誤って理解すると、ICC の役割を誤解する可能性があると考えたからです。

具体的には、次の点を考慮しました。

TSV は、ICC の付属機関であり、ICC の管轄犯罪の被害者を支援することが目的です。そのため、ICC の説明の中で、TSV の役割も説明すると、ICC の役割がわかりにくくなると考えました。

TSV の具体的な支援内容は、経済的支援や心理的支援など、小学生には理解が難しい内容です。そのため、TSV の具体的な支援内容を説明すると、小学生には難しい説明になると考えました。

“ensure”の意味を誤って理解すると、ICC の役割を誤解する可能性があります。そのため、“ensure”の意味

を正確に伝えることが重要であると判断しました。

以上の理由から、TSV は訳さないことにしました。一方で、“ensure”の訳については、正確さを重視することにしました。

長くなりましたが、このやりとりには、「説明すること」をめくり、いくつか興味深い点が観察されます。第一に、候補を列挙して性質を考慮し基準に照らして決定するという行為を作用因、目的因^[23]、その両方から説明することが形式的にはなされていること（例えば“ensure”の訳語説明例）、第二に、でも説明になっていないことがあること（「最もわかりやすい」の一部など）、第三に性質や評価基準が奇妙なことがあること（抽象的だから小学生にわかりやすい、「確かめる」が具体的、等）、第四に説明が一貫していない場合があること、第五に説明に誤りが含まれること（TSV は ICC の付属機関ではない）、第六に質問の問題は確認しないこと（“ensure”を「確保する」としたのは「最もわかりやすい」と判断したためでもあるはずだが、末尾近くでは正確さのみが強調される）、などです。

個人的には、Bard も ChatGPT も試す分には面白いのですが、このような応答をする人と一緒に働くのは少しストレスになるのではないかと思います。

9. MT にできること・できないこと

以上の観察は体系的なものではないため（以下で「体系的」に関わる話もします）、MT や LLM の評価を厳密にできるわけではありませんが、第 2 節から第 5 節で扱った事例のいくつかは、現在の MT や LLM が苦手とすることを示唆しています。例えば 2 節の熟語・慣用句、3 節の表現の位置づけ、4 節の用語の一貫性は苦手と判断できそうですし、逆に、文献[9]「翻訳困難事例」のいくつかのカテゴリについては（例は挙げませんが）ほぼできるようです。

ここで、それなりにはっきりと苦手判定（やできている判定）が可能な事象のカテゴリは、翻訳とは何をどのレベルで維持しなくてはいけない行為かをめぐっ

て人間が比較的明確に認識しているものであると同時に、MT や LLM の側でどうしてできていないか（あるいはできているか：興味深いことに上で挙げたできていない事象は学習の内的なメカニズムよりも学習の枠組みの設定に関係していそうです）がそれなりの妥当性をもって把握できると思われるものです。

これに対して、2 節の一部、5 節、6 節、7 節の事例は、個別の例を前にして修正が必要なところを診断し指摘できるけれどもそこで修正が必要であることを一般化されたかたちで述べにくい事象であること、事象に対する MT や LLM の側の挙動に関する一般化された特徴を認識しにくいことの、いずれかあるいは両方に関わっているため、「MT や LLM にはこれができない」というかたちで明示しにくいものとなっています。このことは、第 8 節の事例でやりとりに見られる極めてあからさまな気持ち悪さとしても露呈しており、同時に、この気持ちの悪さがまさに分析的な把握を（今のところ）拒絶するものであるが故にもたらされるということから、いっそう明らかになっていると思われます（もちろん、これらについて、自らが知らないことを認識していないまま説明をしようとするとか、問に対して何を言えば答えが与えられることになるかを把握していないといった大きなところの指摘はできますが、それを指摘することで問題の解決に向かうアプローチを確定できるわけでもありません）。ここまでで、MT や LLM の挙動に対してしばしば擬人的な述語を用いていることに気付いた読者もいると思います。それは、こうした点と関係しています。

プロンプトでこれらの事象への対応を一貫して促すことは、個々の例に対する一般的な言明を人間の側で構成できないこと（その手前でそもそも個々の例の範囲が把握できているかどうかかわからないのですが）^[24]、仮に一般的な言明を構成できたとしても 8 節のやりとり事例を見るとシステムの挙動がそれに対応したものになる保証がないことから、難しそうです。

こうした状況を鑑みると、以下のような課題に取り組む必要があることがわかります。

- (1) 人間が翻訳を行い翻訳について説明をする際に、どのような手続きで何を行いどのような結果を出すともそもも翻訳を行ったことになるのか、どのような手続きで何をするか説明を行ったことになるのかを、知識として対象化し把握すること。これは、人間の側が機械に歩み寄るためではなく、そもそも自動化において目標とされる翻訳（とその説明）を科学的に認識するためです。
- (2) 明確にされた人間の認識・評価のカテゴリに機械の挙動を対応づけること。特に深層学習の登場にともない説明可能な AI (XAI) は重要な課題になっていますが^[25]、これまでの話を踏まえると、私たちは、翻訳をめぐる事象について何をすればそれをしたことになるのか、そしてそれについて何をすればそれを説明したことになるのか(より一般的にそもそもある行為について何をすれば説明をしたことになるのか)がよくわかっていない状況にあるので、入出力の関係をつなぐ機序の説明を超えて結果を人間と結ぶ部分で評価しその説明を扱う場合、問題は現在定義されたタスクを前提とした技術的なアプローチが扱う範囲を超えることになります^[26]。

これらの課題は、本連載第一回で最初に述べたことでした。本連載では、第一回の冒頭で紹介した研究プロジェクト「翻訳規範とコンピテンスの可操作化を通した翻訳プロセス・モデルと統合環境の構築」(科学研究費基盤 S)で扱ってきた主に(1)に関する知見の一部を簡単に紹介し、それを踏まえて MT や LLM の挙動を検討した今回で、振り出しに戻ったことになります。MT は文芸翻訳が苦手といった印象論に過ぎない言明や MT や LLM は身体性に欠き記号接地できていないからできないといった人間同士のうなずきあいの確認以上の意味をもたない言明よりも多少は先に進むことができたのではないかとはいいますが、本来の到達点をはるか先のようです。まだ扱っていないことは色々ありますが、これでひとまず連載を終わります。

謝辞

本記事の一部は、科研研究費基盤(S)「翻訳規範とコンピテンスの可操作化を通した翻訳プロセス・モデルと統合環境の構築」(19H05660)に関わっています。

注・参考文献

- [1] Robinson, N. R., Ogayo, P., Mortensen, D. R., and Neubig, G. (2023) “ChatGPT MT: Competitive for high- (but not low-) resource languages,” <https://doi.org/10.48550/arXiv.2309.07423>
- [2] <https://mt-auto-minhonn-mlt.ucr.jgn-x.jp/>
- [3] 影浦峽 (2020) 「人間の翻訳と機械の翻訳 (3): 翻訳プロセスは何から構成されるか?」 *AAMT Journal* 73: 34-39.; ISO (2015) *ISO17100:2015 Translation Services — Requirements for Translation Services*. Geneva: ISO.
- [4] 影浦峽 (2021) 「人間の翻訳と機械の翻訳(5): 翻訳者のコンピテンスとは何か?」 *AAMT Journal* 75: 34-38.; ISO (2015) 前掲; European Master’s in Translation (2022) *European Master’s in Translation Competence Framework*.
- [5] 日本翻訳連盟 (2018) 「JTF 翻訳品質評価ガイドライン」 https://www.jtf.jp/pdf/jtf_translation_quality_guidelines_v1.pdf
- [6] Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J. (2002) “Bleu: A method for automatic evaluation of machine translation.” *ACL 2002*, 311–318.
- [7] Rei, R., Stewart, C., Farinha, A. C. and Lavie, A. (2020) COMET: A neural framework for MT evaluation. *EMNLP 2020*, 2685–2702.
- [8] Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q. and Macherey, W. (2021) “Experts, errors, and context: A large-scale study of human evaluation for machine translation,” *TACL* 9: 1460-1474.; Koeurwc, A. and Ryabchikov, A. (2022) “Picking out the best MT model: On the

- methodology of human evaluation,” *AMTA 2022*, 12-23.
- [9] 横山晶一・熊野明・望月泰行・植村昌俊・川越睦・時岡洋一・安藤進・田中穂積 (1994) 「翻訳困難例文の収集と分類」 情報処理学会 自然言語処理研究会報告 94 (47): 33-40.
- [10] DeepL (<https://www.deepl.com/translator>) と Google Translate (<https://translate.google.com/>) は 2023 年 9 月 10 日、みんなの自動翻訳@TexTra ([2]を参照) は 2023 年 9 月 23 日に実施。
- [11] ChatGPT: <https://chat.openai.com/>; Bard: <https://bard.google.com/>。文書としては人種差別撤廃条約の第 6 条を不定期観測に使っています。「猿真似」(コピペ) が人間にしかできないことについては、「山極壽一さん 総合地球環境学研究所所長〈インタビュー〉 コロナ禍は人間らしさの源泉である「共鳴」を妨げる Part 1」(2022 年 5 月 10 日: <https://note.aktio.co.jp/education/20220510-1133.html>) を参照。
- [12] 「中学生向け」といった制限がある場合に正訳をそのまま使うかどうかといった決断は別の話です。
- [13] 用語の同定と訳語の確定も翻訳者(ターミノロジストも含む広義で捉えて)が行なっているタスクですが、ここでは扱いません。
- [14] <https://www.icc-cpi.int/get-involved/sdg16>。一部略。
- [15] Hammack, R. (2018) *Book of Proof*, 3rd ed. <https://www.people.vcu.edu/~rhammack/BookOfProof/Main.pdf>, p. 3.
- [16] Wikipedia “Three Mile Island accident”の項第 2 文の主語“It”を“The Three Mile Island accident”に、“1979”を“1978”に置き換えたもの。
https://en.wikipedia.org/wiki/Three_Mile_Island_accident
- [17] Hammack, 前掲[15], p. 3.
- [18] Pascal, B. *Pensées*. [塩川徹也訳 (2015)『パンセ(中)』岩波文庫] の凡例には「訳注には大別して二つの種類がある。一つは、当該断章の本文批判に関する情報であり、もう一つは断章の意味の理解を助ける情報と説明、つまり解釈に関わる注である。」(p. 11) とありますが、後者は一般には必ずしも「解釈に関わる」わけではないでしょう。
- [19] Wikipedia 「大沼国定公園」概要の第 1 文からよみを削除したもの。<https://ja.wikipedia.org/wiki/大沼国定公園>
- [20] European Master’s in Translation, 前掲[4].
- [21] 前掲[14].
- [22] Bard では回答を選べますがいずれも第一案です。
- [23] 日本翻訳連盟 (2018) 「JTF 翻訳品質評価ガイドライン」 https://www.jtf.jp/pdf/jtf_translation_quality_guidelines_v1.pdf は「翻訳成果物が、関係者間で事前に合意した仕様を満たす程度」(p. 5) と、目標側から定義しており、「そのためにどうしたか」という説明が重要であることが伺えます。人間の行動については目的因的説明が受け入れられているだけでなく、そうでないと説明にならないことがあります。
- [24] Miyata, R. and Fujita, A. (2021) “Understanding pre-editing for black-box neural machine translation,” *EACL 2021*, 1539-1550.で報告されているような作業を大規模にやることは有効かもしれせん。
- [25] Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B. and Sen, P. (2020) “A survey of the state of explainable AI for natural language processing,” *AACL-IJCNLP 2020*, 447-459.; Gurrapu, S., Kulkarni, A., Huang, L., Lourentzou, I., Freeman, L. and Batarseh, F. A. (2023) “Rationalization for explainable NLP: A survey,” <https://doi.org/10.48550/arXiv.2301.08912>
- [26] 機械の挙動を外部から理解するための説明ではなく、人間と機械が関与する作業の中で機械が人間に説明することを考えると、説得と説明の区別 (Dummett, M. (1978). *Truth and Other Enigmas*. Harvard University Press.)、注[23]でも述べた目的論的説明の位置づけと構成方法なども考える必要が出てきます。ところでこれら、特に前者は、機械にとってもだけでなく人間にとっても難しいことで、それは大元では何をすれば説明したことになるのかが認識されていないことを反映しているようです。

解説記事

立命館大学の正課英語授業へのニューラル機械翻訳導入について

坂場 大道、木村 修平、近藤 雪絵

立命館大学

1. はじめに

本稿では、立命館大学で展開しているプロジェクト型英語教育にニューラル機械翻訳サービス「Mirai Translator」を導入した背景および授業実践の変化について報告するとともに、学生アンケート調査に基づいた現段階での評価を報告する。

2. PEP について

本節では、立命館大学およびニューラル機械翻訳サービスを導入した英語プログラムについての概要を述べる。

立命館大学 4 学部で展開する PEP

立命館大学は関西の大規模私立大学の 1 つである。本稿執筆時点で 16 学部、21 研究科を擁し、約 3 万 7000 人の学部生・大学院生が所属する。このうち、生命科学部・薬学部・スポーツ健康科学部・総合心理学部の 4 学部の正課必修英語授業となっているのが「プロジェクト発信型英語プログラム」(Project-based English Program。以下、PEP と略す)であり、4 学部の 1・2 年生あわせて 2000 名が履修している (図 1)。

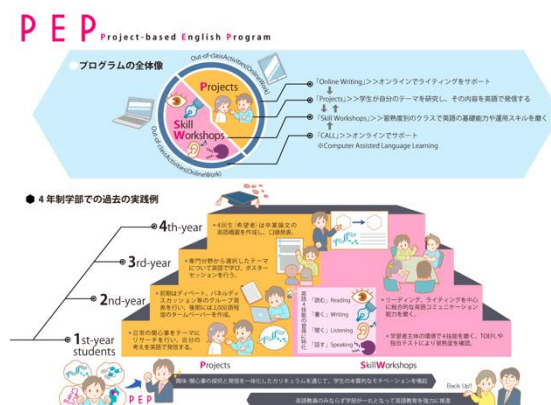


図 1. PEP の全体像 (生命科学部、薬学部の例)

PEP は週 2 コマの授業で構成されている。1 つは読む・聞く・話す・書く・文法などの基本的な英語スキルを磨く Skill Workshops で、こちらは外部教育機関との連携で行われている。もう 1 つが Project で、こちらは Skill Workshops で培ったスキルを活かす表現・発信の場として位置づけられ、著者らを含む立命館大学の教員が担当する。Project はプロジェクト型学習 (Project-Based Learning) に基づく教授法を採用しており、近年では探究型学習とも呼ばれる¹⁾。

Project 授業は、学習者自身の興味・関心からリサーチ対象を選択してプロジェクトを起ち上げ、授業を通じて発展させることで、進捗・成果を英語で発表することを基本スタイルとする。学年の進行とともに求められる発表形態のレベルやクオリティは高くなる。たとえばインフォーマルなエッセイからアカデミックな書式のペーパーへ、あるいはカジュアルな口頭発表から学会のようなフォーマルなプレゼンテーションへ、のように変化する。

ICT をフル活用する英語プログラム

PEP は 2008 年度のプログラム開設当初から ICT 活用をコアに据えてきたプログラムであり、特に Project 授業は事実上の BYOD (Bring Your Own Device) 体制を実施してきた。学生がほぼ全員ノート PC を開き教室 Wi-Fi に接続しながらプロジェクトに取り組み発表する PEP の授業風景 (図 2) は、従来の大学英語教育の一般的なイメージとはかなり異なると思われる²。



図 2. PEP の授業風景 (2018 年度)

15 年以上にわたり教育 ICT 化を推進してきた PEP にとってニューラル機械翻訳の導入は必然であったと言える。そもそも PEP ではこれまで機械翻訳の使用を明示的に禁止しておらず、むしろ著者らを含む一部の PEP 教員はニューラル機械翻訳や AI (人工知能) テクノロジーの合理的な教育利用の可能性を模索していた (山中, 2019)。

3. Mirai Translator 導入の経緯

本節では、PEP4 学部に導入されたニューラル機械翻訳サービス「Mirai Translator」について、その導入の経緯と機能について述べる。

教育開発 DX ピッチの開催

立命館大学は 2030 年にあるべき姿を学園ビジョンとして掲げ、そのビジョンの実現を推進する一連の取り組みを「立命館大学チャレンジ・デザイン」(通称

R2030) と呼んでいる³。R2030 の中には教育や研究活動の高度化および DX (デジタル・トランスフォーメーション) が含まれており、ICT を積極的に活用した教育手法の開発および実践を支援する学内競争予算施策「教育開発 DX ピッチ」が全学に向けて通知された⁴。選考の結果、3 組の採択が決定し、その中の 1 つに PEP が選ばれた。PEP は獲得した予算で 2022 年度秋学期から複数の先進的なサービスと契約し、その中の 1 つがニューラル機械翻訳サービスの「Mirai Translator」である。

Mirai Translator の選定

「Mirai Translator」は、株式会社みらい翻訳が提供するニューラル機械翻訳サービスである⁵。同社は NTT ドコモの子会社として 2014 年に設立され、三菱商事やダイキン工業など主に B2B 分野において多数の導入実績を持つ。Mirai Translator は、Google 翻訳や DeepL と同様にニューラルネットワーク技術に基づいており、その翻訳精度は、同社によると、和文英訳で「TOEIC960 点の日本人ビジネスマンを超える」とされ、英文和訳では「TOEIC960 点の日本人ビジネスマンと同等レベル」である⁶。

企業への大規模導入実績があること、契約やサポート、共同研究を日本語ベースで行えること⁷に加え、逆翻訳を確認しやすいインターフェイスになっていることが PEP による Mirai Translator 選定の理由となった。

2022 年度秋学期から Mirai Translator を PEP の授業に導入するにあたり、PEP 授業を担当する教員間でアジア太平洋機械翻訳協会 (AAMT) の発行する『MT ユーザーガイド』を Slack 上で共有した⁸。また、AAMT 理事の山田優・立教大学教授の監修のもと、ニューラル機械翻訳の仕組みと適正な使い方を解説した教員および学生向けの動画教材を作成し、YouTube 上で一般公開した⁹。

一連の PEP の取り組みは、2022 年 10 月に大学から Mirai Translator 導入についてのプレスリリースが発表

され¹⁰、次いで 11 月に取材記事が公表された¹¹ことで、全国的に知られることとなった。かくして、おそらく日本の大学では史上初の試みとなるニューラル機械翻訳の正課英語授業への導入が開始されたのである。

4. NMT の導入で授業はどう変わったか？

本節では、ニューラル機械翻訳 (Neural Machine Translation: NMT) サービスである Mirai Translator を導入したことによる変化について PEP 授業を担当する著者らの目線から報告する。

NMT 導入で授業はどう変わったか？：坂場の場合

最も顕著な変化の一つは、英語の得意ではない学生の発表に表れていた。NMT の積極的導入前は、学生が発表テーマについて熱心に調べていても、文法エラーや語彙の選択ミスによって、プレゼンテーションの内容がうまく伝わらないこともあった。時間をかけて準備した内容が伝わらないことが発表者にとって残念であるのは言うまでもないが、聴衆としても要領を得ない発表を聞き続けるのには辛抱を要する。NMT 導入後は、こういう機会に出くわすことはほとんどなくなった。NMT の助けを借りることで、プレゼンテーションに用いる英語の質を高めるなど、学生が自分の英語力では追いつかないことを成し遂げることができていた。こうした発表の成功経験を積むことは、英語に対する苦手意識の軽減にも寄与すると考えられる。

他方、NMT 導入後から頻繁に観察されるようになった誤りもある。たとえば、発表日の振り返りとして、“I made an announcement today.”のように日本語の「発表」に相当する部分に、announcement という英語を用いる学生が急増した。日本語の「発表」という語は presentation だけでなく、公式に何かを通知する announcement のようなものも表せる。こうした日英間の語彙のズレによって、NMT の産出した英文が文脈に適さない訳になったと考えられる。この事例が示しているのは NMT の問題ではなく、機械翻訳され

た文を人間が修正するポストエディットの重要性である。もちろん、機械翻訳しやすいように原文を修正するプレエディットによって、そもそも「発表」ではなく「プレゼンテーション」という日本語を用いる方法もある。これらのプロセスは、より文脈に適した英語の産出だけでなく、学生が自分の英語として発信する自信にも寄与する (近藤 et al., 2023)。ここで取り上げたような日英間で生じやすい翻訳のズレを意識することは、学生が自律して NMT を活用する意識の醸成にも役立つと思われる。

NMT 導入で授業はどう変わったか？：木村の場合

機械翻訳の活用と言うと和文英訳のような産出タスクに注目が集まりがちだ。確かに基本的な文法ミスやスペルミスを指摘する機会は劇的に減ったので、教師は学生のプロジェクトの内容や方向性に関する助言やフィードバックにより多くのエフォートを振り分けることができるようになったと実感している。

その一方で、受容タスクにも NMT は大きな効果があると思われる。このことを示す顕著な変化の 1 つとして、学生がプロジェクトの論拠や参考資料として収集する英語の論文や新聞記事などのクオリティが急激に上がったことが挙げられる。NMT を導入する以前は、日本語のまとめサイトやニュースポータルサイトなど信頼性やトラッキングに問題のあるソースを使わないよう何度も繰り返し注意していた。しかしながら、2022 年度秋学期に担当した 3 つの授業 (1~3 年生) では、学生たちが Mirai Translator を用いて信頼性の高い論文や公的機関の英文ドキュメントを参考文献に挙げてくる事例が急増した。

こうした NMT の活用は PEP のようなプロジェクト型・探究型の授業だからこそ可能なのだという声もあるだろう。そのことは否定しないが、英語は多くの大学生が中学・高校と最低 6 年の歳月を費やして学んできた既修言語であること、2022 年度から高校の学習指導要領に探究活動が加わったこと、そして全国の SELHi (Super English Language High Schools) や

SSH (Super Science High Schools) 指定校において PEP のような授業形態は決して珍しいものではない¹² ことを考えると、大学英語教育が NMT を導入して新たな教育モデルの可能性を模索しない理由はもはや無いのではないだろうか。

5. アンケート調査について

2022 年度秋学期の英語授業を受講した 1 年生および 2 年生を対象に、NMT の活用状況と NMT を活用して作成した英文への意識に関するアンケート調査を実施した。実施時期は秋学期の終了時で、回答者数は 1 年生 115 名、2 年生 126 名の合計 241 名であった。本アンケートにおいて、回答者はあるメディア（グラフ）から自分が最も重要だと思う情報を抜き出し、自分の意見を添えて 2 文の英文でまとめるという課題を行った後に、質問に回答した。本節では、今回のアンケート結果を報告し、秋学期の開始時に行った近藤 et al.(2023)の調査結果と比較する。

NMT を何に活用したか

学期を通じて学生がどの活動で NMT を利用したかを確認するため、4 つの項目（「プレゼンテーション」「タームペーパー」「英語情報収集」「質疑応答やディスカッション」）における利用度を 5 段階で尋ねた。結果を図 3 に示す。

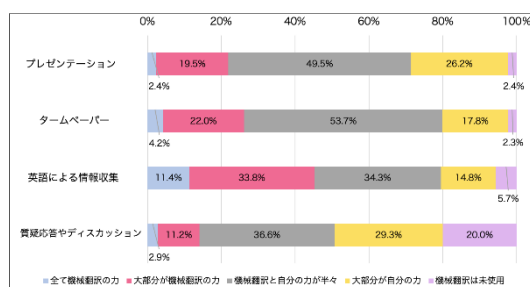


図 3. どの活動にどれくらい NMT を活用しているか

「プレゼンテーション」と「タームペーパー」は似た傾向を示し、約半数が NMT と自分の力を同等に併用

したと回答した。情報収集活動においては、NMT の利用が顕著に増え、NMT を主に使用する学生が 33.8% に上り、全て NMT の力で情報収集したと回答した学生も 11.4% 存在した。一方、質疑応答やディスカッションは反対の傾向を示し、全てまたは主に NMT を活用したと回答した学生を合わせても 14.1% と割合が低く、全てまたは主に自分の力を使ったと回答した学生を合わせると 49.3% であった。これらの結果から、学生は情報を受信する活動においては NMT を頼りにする傾向があるが、自分のプロジェクトに関するプレゼンテーションやタームペーパーのように自らの意見や情報を発信する活動では、自分の力を併用する傾向が高まった。また、事前に準備することが難しい質疑応答やディスカッションのような動的な活動でも、NMT は一定の割合で使用されていることがわかった。

NMT を活用して作成した英文への意識

NMT を活用した英文作成に対する意識を調査するため「最終的にできた英文を気に入っていますか？」および「最終的にできた英文を自信をもって発信できますか？」の 2 問と回答の理由を尋ねた。理由の回答は任意で、自由記述形式であった。

英文を気に入っているか

NMT を活用して作成した英文に対して、「気に入っている」の回答は 32.4%、「やや気に入っている」は 39.4% であり、ポジティブな意識をもった学生は合わせて 71.8% であった。ネガティブに回答した学生は、「気に入っていない」0.4%、「あまり気に入っていない」4.1% であり、合わせて 4.6% であった（表 1）。秋学期開始時の調査結果（近藤 et al., 2023）では、ポジティブな回答が 66.9%、ネガティブな回答が 7.2% であり、秋学期終了時には開始時に比べて NMT と活用した作成した英文を気に入っている学生の割合がやや高まった。

表 1. 作成した英文を気に入っているか

項目	選択数	選択割合
気に入っている	78	32.4%
やや気に入っている	95	39.4%
どちらでもない	57	23.7%
あまり気に入っていない	10	4.1%
気に入っていない	1	0.4%

気に入った理由として最も多かったものは、「わかりやすい／簡潔／簡単」、「(翻訳が) うまくできた／意図が反映された」など英文や翻訳の質の高さであり、「正確／間違いがない」など英文の正確さがそれに続いた。一方、気に入っていない理由として最も多かったものは、「あっているかわからない」など翻訳結果の不確かさであり、「自分で作った英文ではない/自分の手を加えていない」など自分ではなく機械が英文を作成した点がそれに続いた。

自信をもって発信できるか

NMT を活用して作成した英文に対して、「自信をもって発信できる」の回答は 21.6%、「やや自信をもって発信できる」の回答は 44.4%であり、ポジティブな意識をもった学生は合わせて 66.0%であった。ネガティブに回答した学生は、「自信を持って発信できない」0.4%、「あまり自信を持って発信できない」10.8%であり、合わせて 11.2%であった(表 2)。概ね学生は MT を活用して作成した英文を、自信をもって発信できると捉えていた。秋学期開始時の調査結果(近藤 et al., 2023) では、ポジティブな回答が 64.7%、ネガティブな回答が 7.2%であり、秋学期全体での学生の英文への自信に大きな変化はなかった。

表 2. 作成した英文を自信をもって発信できるか

項目	選択数	選択割合
自信をもって発信できる	52	21.6%
やや自信をもって発信できる	107	44.4%
どちらでもない	55	22.8%

あまり自信を持って

発信できない	26	10.8%
自信を持って発信できない	1	0.4%

自信を持って発信できる主な理由として、学生たちは英文の正確さを挙げた。さらに、自分で英文に手を加えたり、確認・修正を行ったことも主な理由として挙げられた。その次には、「わかりやすい／簡単／簡潔／伝わる」など英文の質の高さが続いた。一方、自信をもって発信できない理由として最も多かったのは「あっているかわからない」など英語の不確かさであり、自分で英文に手を加えたり、確認・修正を行わなかった点がそれに続いた。

秋学期開始時と終了時の比較

秋学期に NMT を授業で取り入れた結果、NMT の助けを得て作成した英文を気に入っている割合は 71.8%に上り、自信をもって発信できる割合も 66.0%という高い結果であった。特に、英文を気に入っている学生の割合は、秋学期開始時と比べ約 5%上昇した。高く評価する学生たちは、NMT によって得られる英文の質や正確性を重視していた。一方で、評価が低い学生は、自ら英文に手を加えなかったり、英文の正確性の確認が不十分であると感じていた。

学生の発信力を向上させるためには、学生自身がその英文を気に入り、自信を持って発信することが鍵となる。その実現のために、NMT の助けを単に利用するだけでなく、真に活用すること、すなわちその英文を自ら手直して自分のものにすることが求められる。自分自身で英文を確認するための方法として、逆翻訳や Grammarly を活用したと回答した学生もいた一方で、NMT による自分の習熟度を越えた単語の選択や自身の英語力不足を理由に自信が持てないと感じる学生もいた。発信力向上のためには、授業で各自の習熟度に応じた英文の修正や確認の方法を学び、独自の修正や確認の経験を重ねることが重要であると考えられる。

6. おわりに

本稿では、立命館大学のプロジェクト型英語教育においてニューラル機械翻訳サービス「Mirai Translator」を導入した背景と、授業実践の変化について論じた。

第一に、NMT の力を借りてプレゼンテーションに用いる英語の質を高めることで、英語が得意ではない学生も、自身のプロジェクトの内容を一定水準以上で発信できるようになった例が観察された。また、そうした影響は和文英訳のような産出タスクだけでなく、プロジェクトに用いる英語の論文や新聞記事の質の向上など、受容タスクにも見られた。英語の授業を受講した学生に対するアンケート調査によると、学生は情報を受信する活動においては NMT を頼りにする傾向があるが、自らの意見や情報を発信する活動では、自分の力を併用する傾向が高まる。多くの学生は、NMT を活用して作成された英文を気に入っており、それらを自信を持って発信できていた。

学生の発信力のさらなる向上には、習熟度に応じた英文の修正や確認方法を学び、それらの経験を積み重ねることが重要であると考えられる。

7. 参考文献

木村修平(2018)「BYOD 型大学英語プログラムが IT スキルの自己評価に及ぼす影響—英語と IT をインフラとして捉える教育実践—」『コンピュータ&エデュケーション』45, pp.127-132.

¹ PEP に関する詳細は山中 et al.(2021)を参照。

² PEP 授業における ICT の具体的な活用について、詳細は山中 et al.(2021, pp.148-178)および木村(2018;2020;2022)を参照。

³ R2030 : <https://www.ritsumei.ac.jp/features/r2030/>

⁴ 教育開発 DX ピッチ : <https://www.ritsumei.ac.jp/itl/2022DXpitch/>

⁵ Mirai Translator : <https://miraitranslator.com/>

⁶ 機械翻訳サービスの和文英訳がプロ翻訳者レベルに、英文和訳は TOEIC960 点レベルを達成 : https://miraitranslate.com/uploads/2019/04/MiraiTranslate_JaEn_pressrelease_20190417.pdf

木村修平(2020)「BYOD 環境下でのプロジェクト型英語授業の要件に関する考察」『CIEC 春季カンファレンス論文集』11, pp.75-80.

木村修平(2022)「プロジェクト発信型英語プログラムが見出す新たな 4 技能一次世代研究大学の基盤としての英語教育—」『大学教育と情報』176, pp.29-32.

近藤雪絵, 木村修平, 坂場大道, 豊島知穂, 中南美穂, 山下美朋, 山中司(2023)「AI 機械翻訳の英語正課授業への大規模導入とその課題—英語発信力向上のための機械翻訳活用に向けて—」『CIEC 春季カンファレンス論文集』14, pp.41-44.

田中博, 武田菜々子, 廣松光一郎, 家長大輔(2021)「オンラインによる国際科学教育の取組とその成果—Online Science Fair の開催—」『科学教育研究』45(2), pp. 256-264. <https://doi.org/10.14935/jssej.45.256>

山中司(2019)「大学にもう英語教育はいらない：自身の『否定』と『乗り越え』が求められる英語教育者へのささやかなる警鐘」『立命館大学人間科学研究所』38, pp.73-89. <http://hdl.handle.net/10367/11541>

山中司・木村修平・山下美朋・近藤雪絵(2021)『プロジェクト発信型英語プログラム：自分軸を鍛える「教えない」教育』北大路書房.

⁷ 2023 年 7 月 13 日、みらい翻訳社は PEP との共同研究の最初の成果を発表した。:

<https://miraitranslate.com/news/1870/>

⁸ MT ユーザーガイド :

<https://aamt.info/act/MTuserguide>

⁹ 英語学習のパートナーとしての機械翻訳 :

<https://www.youtube.com/watch?v=YQDm8qLpIzc>

¹⁰ 大学の英語授業に AI 自動翻訳サービスを試験導入 学生・院生約 5,000 人を対象に、翻訳ツールを用いて新しい英語教育の可能性を検証 (2022 年 10 月 3 日) :

https://www.ritsumei.ac.jp/profile/pressrelease_detail/?id=719

¹¹ 大学の英語授業で AI 自動翻訳サービスを試験導入-
新しい英語教育の可能性とは (2022 年 11 月 28 日) :
<https://www.ritsumei.ac.jp/news/detail/?id=2908>

¹² 立命館大学の附属校である立命館高等学校は 2002

年度から SSH 指定を受け続け、国内外の高校生を集めてサイエンスフェアを毎年開催し、英語と ICT を活用した探究型教育を実施している (田中 et al., 2021)。

解説記事

ポストエディットの真実

隅田 英一郎

アジア太平洋機械翻訳協会会長(情報通信研究機構フェロー)

1. AAMT セミナーの紹介

アジア太平洋機械翻訳協会（AAMT）では、機械翻訳を活用するために有用な情報を広く発信する手段として、2022 年度にオンライン・セミナーをはじめた。

同年度実施した 3 回の講演が好評だったことを受け、翌年度、企画・実行を強化するために人員を増強し実施回数も増やす目的で、AAMT の理事会は委員会を新設した。セミナーの第一回から内容が決まっている回までについての情報がサイト¹に公開されている。

通算 4 回目のセミナーは、ポストエディット（機械翻訳の出力を検査して必要な修正を施し翻訳としての完成度を上げることを目的に翻訳者が行う作業であり、MTPE と略記されることも多い）に関する関心が高まっている状況をとらえ、「ポストエディットの真実」というタイトルで小職が登壇した。

2. 「ポストエディットの真実」の背景

セミナー当日、すなわち 2023 年 5 月 31 日の時点で、ポストエディットを敵視する意見とポストエディットを支持する意見が拮抗しているようにみえた。より正確には、強い反対意見が深く浸透している一方で、ポストエディットのメリットを認識している翻訳会社や翻訳者は（目立つ形で発信することせずに粛々と）ポストエディットを実践していた。

ポストエディットは、その名のとおりに必然的に機械翻訳の精度に強く依存することから、機械翻訳が AI によって高精度化しその活用が広がってきた²ことに

じて、ポストエディットの特長も変化している。にもかかわらず、多数の人々は、かつて低精度であった時代の機械翻訳とそれを対象にしたポストエディットの経験に基づいて形成された負のイメージを引きずって、現在のポストエディットの実像を正確に認識していなかった。

小職はこの状況を打開すべきと考え、そのための一つの手段として 2022 年度に情報通信研究機構が実施した 2 件の大規模なポストエディット案件を活用した。同案件に携わった翻訳者にアンケート調査への協力をお願いし、その集計結果をセミナーで発表した。本稿で要点を記し所見を述べる。

3. 質問「今後もポストエディットの仕事を受注したいと思いますか？」への回答

衝撃的なことに、この直接的な質問への回答は、93%が「受注したい」であった。強い反対意見が深く浸透しているにもかかわらず、実際に体験した後は次も受注するということだった。

次節から、アンケートの詳細を説明する。

4. 大規模なポストエディット案件①の概要

本件は論文要約の英日 MTPE (図 1) で総語数は 7000 万強と大きく期間も 1 年弱と長期に亘るものである。

回答いただいた人数は 178 で、アンケートの回収率としては約 64%であった。回答者の年齢の分布は、40 代未満 9%、40 代 29%、50 代 29%、60 代 21%、70

¹<https://www.aamt.info/event/seminar>

²翻訳会社の機械翻訳活用率は 2022 年度には、2017 年度の 15.6%の約 3 倍に(日本翻訳連盟の翻訳白書)。

代以上 12%であり、大きな偏りはない。

【語数】約 7,200 万
 【分野】医療系 (PubMed) と IT 系 (arXiv)
 【文数】約 350 万文
 【翻訳エンジン】汎用
 【翻訳支援ツール】不使用
 【基準】フル PE。原文は論文要約を文分割したテキストであり、同一要約と分かる範囲で、文脈参照し、訳語統一、論文スタイルの訳文を依頼。
 【期間】2022 年 3 月～2023 年 2 月。
 【翻訳者数】277

図 1 論文要約の英日 MTPE

5. 質問「ポストエディットと通常の人手翻訳（ゼロから翻訳）で作業効率は変わりますか？」への回答

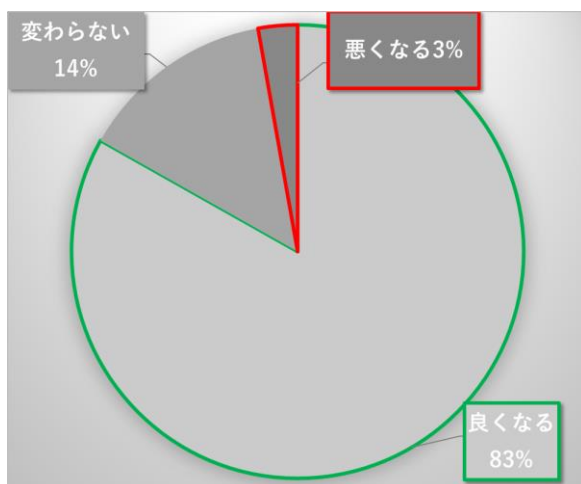


図 2 MTPE による作業効率の改善

指導的な立場にある翻訳者の多くは「ポストエディットは翻訳を完成させるのに反って余計に時間がかかる。ゼロから翻訳の方が早く済む。」という趣旨で頻繁に発信してきた。「何処に誤りがあるかどうか分からないので全数 CHECK が必要である。」「機械翻訳

の訳文が下手なので直す手間が大きい」等いくつかの理由が挙げられることが多かった。

効率について実際を見てみよう。図 2 に示した円グラフから分かるように、実際に大規模なポストエディットに参加しアンケートに答えた翻訳者の中で、「効率が悪くなった」としたのは少なかった(3%)。一方「効率が良くなった」としたのは8割を越えた(83%)³。

流布している認識と真逆の認識を多数の翻訳者がもったことになる。今回は『汎用』の翻訳エンジンを産業翻訳の異なる分野である医療や IT に適用したケースであることを考慮すると、他分野のポストエディット案件にも通底することが期待される（そして多分野に通底することが期待される）。

集計からポストエディットの単位時間当たりの翻訳語数が、ゼロから翻訳する場合の単位時間当たりの翻訳語数に対して平均 1.5 倍⁴になったことがわかった。

一方で、100 人のオーダーで、多様なバックグラウンドの翻訳者が作業して平均 1.5 倍だから、好ましい結果である。機械翻訳システムは、医療や IT などの分野に特化させると高精度化出来ることが分かっているので、『特化した』翻訳エンジンの訳文を対象にポストエディットに使える 1.5 倍を超えるだろう。

長期に亘る案件であったことから MTPE は続けることによってどんどん効率が上がるということも分かった。翻訳者は「機械翻訳のエラー発見になれた」「原文の理解が進み内容把握に時間がかからなくなった」「定訳等の調査方法に習熟し調査の時間が減った」等の理由を上位にあげていた。少し試しただけでは分からないポストエディットの【深層】にある真実であろう。

³ 「今後もポストエディットの仕事を受注したいと思いますか？」の 93%が「受注したい」で「効率が良くなった」の 83%を越えた。「効率が良くなかった」場合に「受注したい」という回答が過半であるので効率化以外のメリットの存在が示唆される（6 節参照）

⁴ この倍率 1.5 は様々な要素で変動しうる。対象となる分野が翻訳者の得意分野であるか否か、翻訳支援ツールを利用するか否か、機械翻訳の訳文の質が高いかどうか等で変わるので、絶対値についての議論は控える。

6. 【ポストエディットを実務として体験した】 翻訳者の positive な感想

翻訳者の方々に自由書式で感想をいただいたので以下に引用する（下線は著者）。

「複数の翻訳会社からの案件で経験を積むにつれて、ポストエディットの可能性を感じるようになりました。例えば、今回の案件では医薬分野の文献とのお話でしたが、実際には昆虫学や社会学など、私には門外漢のものも多く、これを一から翻訳するのは到底無理でした。機械翻訳の土台があったからこそ、専門用語を調べる手掛かりが得られたのだと思います。」

「機械翻訳がまさかと思うような用語を使用している場合でも、調べてみると実際に使われているものだったりすることもあり、正直なところ大変勉強になりました。」

「レビュー作業と同じくらい勉強になります。他の翻訳者の翻訳から勉強しているのに近いものと感じておりました。翻訳の仕方はもちろん、スタイルガイドに沿った表現を選ぶトレーニングにもなります。本案件に従事している間、仕事をする日のうち、ほぼ毎朝 1～2 時間程度で本案件に取り組むと決めておりました。筋トレのように、毎日続けることで翻訳の「筋肉(力)」がつくような気がしておりました。」

「最初は自分に向いているか、知らない分野の訳文をチェックできるのか等、不安はありましたが、今ではポストエディット作業に大変やり甲斐を感じています。というのも、沢山の訳文をチェックする中で、これまでに身につけた英文法の知識を大いに生かせ、リサーチに時間はかかりますが未知の分野について知る機会を多く得られるからです。これから是非続けていきたいです。」

「ゼロから翻訳するよりも、機械翻訳を通して PE をする方が、圧倒的に効率が良いと感じました。人の翻訳スピードには限界があるので、データを蓄積して機械翻訳の精度を高めていくのが良いと思います。」

「毎日やっても苦にならなかった。常に一定量の作業が確保されるのであれば、毎日でもポストエ

ディットをやりたいとさえ思った。自分はゼロからの翻訳よりも、ポストエディットや校正の方が向いているのかもしれない。

・苦手（知識があまりない）な分野は時間・労力ともにかかってしまい大変だったが、それはそれで楽しみながら作業できた。」

「翻訳ツールの優秀さやその有用性を理解した。今後、最初のドキュメント翻訳を、このような翻訳ツールを用いることによって、大幅に負担を下げる事が可能となり、その分、科学的な見地からの訳文に仕上げることに注力できます。加えて、ケアレスミスも削減できると考えます。」

「長文になると係り受け訳の誤訳が頻繁にみられ、その解説が逆に人手翻訳より時間がかかってしまった印象がありました。その他では、タイピングの手間も省略できるので、PE 作業は迅速に大量の翻訳が可能であると実感できました。」

「PE では、タイピングの時間を削減でき、知識が乏しい分野の専門用語の調査も大幅に削減できることが多いと思います。その分、日本語の精査に時間がかけられるメリットがあると思いました。」

MTPE が効率を上げるという有用性は想定内だが、翻訳者が専門以外の分野に進出することを支援するツールとしての機械翻訳の有用性は新しい発見である。

7. 【ポストエディットを実務として体験した】 翻訳者の注文等

6 節の感想の他に、伝統的な見解もあったし、作業方法の改善を求める注文や、ポストエディット普及策の提案もあった。

「機械翻訳の能力の高さに感嘆させられる箇所と、何でこんな単純なミスが発生するかと感じさせられる箇所が混在していることに面白さを感じています。」

「ポストエディットでは、機械翻訳が直訳としては正しく、訳抜け等を除けば基本的に直さなくても意味が通じる場合が多いため、どこまで手を加えるべきなの

か？という点で迷うことが少なくありません。自分であればこう訳すけれども、この機械翻訳でも問題ない
という場合のさじ加減が難しいと感じています。」

「ポストエディットを多くやるにつれ、作業を行ううえで押さえておくべきポイント、機械翻訳の得意な点と弱点が少し分かったような気がしました。指示書に基づき自己流で行っているため、一度体系的にポストエディットについて学んでみたいと思いました。」

「今回のように「あくまでも原文に忠実な、過不足のない直訳」を目指すのであれば、MTPEの方が断然早いと思います。ただし「読みやすさ」「ナチュラルさ」を求められると、MTでも意味はとれるものの文章を書き直さなければならないことが多く、MTのメリットといえば自分で訳文を一からタイプしなくてよいことぐらいかなと感じます。」

「今回の案件のように非常に多くの人がかかわる場合は難しいと思いますが、他の翻訳会社の案件での経験から、翻訳支援ソフトのプロジェクトを使用したほうがポストエディットによる作業効率化をさらに実感できるとも感じております。」

【語数】(約 9,000 万語)

【分野】法令、医療、教育、経済、情報・通信・ネットワーク等の多様なトピック。

【文数】 約 380 万文

【基準】フル PE。原文に前後関係が無く文脈は利用できないが、法令や医療等の専門知識を要するため可能な範囲で用語統一や分野に合わせた訳出を依頼。

【期間】 2022 年 8 月～2023 年 3 月。

【翻訳者数】 268

図 3 多分野の英日 MTPE

8. 大規模なポストエディット案件②の概要

案件①と並行して実施された案件②の概要は図 3 の通りである。単語数が若干多いが従事した翻訳者はほぼ同規模。主たる違いとしては、②の方が分野の分散が大きい点がある。アンケートで得られた結果は①と大差なく、**85%が「作業効率が良くなる」と回答し、98%が「今後もポストエディットの仕事を受注したい」と回答した。**

9. おわりに⁵

今回は、小職が知りうる限り初めての試みとして、2 件の大量（7～9 千万語）のポストエディット実務に従事していただいた「延べ 500 人を越えるプロの翻訳者」にアンケートへの回答をお願いし一定の回収率を得た。

今まで実施されることがなかった大規模アンケートによって、ポストエディットに対する「今日まで喧伝されてきた negative な見解」と異なる「positive な意見」が多数あったことは大きな収穫であった。

実案件体験後のアンケートは、一人一人の思考に内在しがちなバイアスを越える手段となるものであり、また、ポストエディットに関する従前の意見・論考と 180 度違う意見が多数あったということから、ポストエディットの有用性が否定しがたく確認された。これは多くのステークホルダーにとって新規な情報である。

小職は、ステークホルダーが機械翻訳の訳文の質を踏まえて、ポストエディットのゴールについて意識を合わせておくことがポストエディット案件の成功の要であることも、今回のセミナーで言及した。また、単価の適切な設定も重要な課題であり、サステナブルな設定をステークホルダーが連携して追及することが求められるのだろう。

最後に、アンケートにお答えいただいた翻訳者様、アンケートを作成・配布・回収していただいた翻訳会社様に深く感謝する。

⁵ 通訳・翻訳ジャーナル AUTUMN 2023 号、pp.36-37 に同誌記者による報告が掲載されている。著者が異なるため本記事と補完的といえる。

温故知新 3

後藤 功雄

AAMT

AAMT では、AAMT 創立 30 周年記念事業として、過去の AAMT ジャーナルおよび JAMT ジャーナル (AAMT の前身である JAMT (日本機械翻訳協会) の会誌) を PDF 化して公開しています。

<https://www.aamt.info/act/journal/>

公開されているジャーナルは次の通りです。

- JAMT ジャーナル No.01,1991 年 7 月～No.07,1992 年 8 月 (今回 PDF 化)
- AAMT ジャーナル No.01,1992 年 11 月～No.70,2019 年 6 月 (今回 PDF 化)
- AAMT ジャーナル「機械翻訳」No.71,2019 年 12 月～ (当初より PDF 版を公開)

「温故知新」シリーズでは、過去の AAMT ジャーナルおよび JAMT ジャーナルの記事を紹介します。

過去のジャーナル記事を読むと、当時の MT の事情が分かるとともに、現在も解決していない課題が残っていることが分かるなど、これからの MT 発展に役立つものも多く見られます。

本号では、1992 年 3 月と 1992 年 5 月号の JAMT ジャーナルから 2 つの記事を転載して紹介します。これらは、PDF から OCR で読み取ったテキストを修正したものです。画像等を含むオリジナルの原稿については、上記サイトをぜひご覧ください。

【1992 年 3 月 JAMT ジャーナル No.5 より】

第 2 回 「機械翻訳ビジネスは儲かるか」

前号からのつづき.....

ご出席： 柏原武利氏 (株式会社スバルインターナショナル 代表取締役)

鎌田紳二氏 (株式会社富士通ドキュメントサービス 機械翻訳センター課長)

富永勲氏 (特殊法人日本科学技術情報センター 主任情報員)

司会： 鳴海武史 (ニューズレター編集委員 株式会社十印 言語研究所) (以下敬称略)

司会： どうもありがとうございました。お立場がそれぞれ違う方から貴重なお話を聞けたと思います。そこで、少し話を変えます。前から気にしていることなんです、それは機械翻訳の料金

の問題です。翻訳市場としては買い手はいいものをなるべく安く、売手はなるべく高く売ろうとする訳ですが、こと翻訳については機械翻訳だからという理由で安くてよいということになるのか、逆に機械翻訳には諸々のメリット、例えば均質性や用語の統一があるから高くしていいという点ではどうですか。素翻訳とポストエディット付きでは異なるとは思いますが。

鎌田： 価格は非常に単純だと思います。この 2 年、機械翻訳を売る商売をやってきたんですが、機械翻訳だから安くするのは一理あります。パソコン通信だと人手を全然かけないでやれば、かかる原価というのは設備投資と営業経費だけです。だから、本当に安いんです。今までは、私たちはお客さんから文書フロッピーをもらってきて人手をかけて機械にかけていましたが、間接経費がかかりすぎてダメでした。スタンドア

Learning from the past 3

Isao Goto

AAMT

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License.

License details: <https://creativecommons.org/licenses/by-sa/4.0/>

ローンは高すぎるし、パソコンに載る機械翻訳ソフトはせいぜい 20 万円ぐらいであるが、パソコンのマニアクラスでないと自分では使いこなせない。そういう色々な問題があって、マーケットリサーチもやって、結局、パソコン通信だったらいけるということで、リリースしてその通りだったんです。機械翻訳がナマのままですと、勿論、プロの翻訳者と比べればひどい品質ですよ。それでも皆さん、ガンガン使っています。ずっと一年間見ていると、一度に 3 千円、5 千円使っているんです。最高記録は 1 日で 16 万円使った人がいます。必要だから使っているんです。お客さんから見れば機械でやっているんだから電気代だけじゃないのと言われてもしょうがない訳で、もっとトラフィックが上がれば値段は下げられます。

柏原：それはあくまでニフティサーブを利用した MT ということですね。それ以外は例えば後編集をしたようなサービスというのはないのですか。

鎌田：もちろんあります。そちらの方も売り上げはかなりあります。

柏原：私どもは MT だけのサービスはほとんどないですね。あることは多少はあるんですが。その時点で言えば手放れもいいし、コストもかからないので、お客さんにメリットを還元すべきで、安くなると思います。ですから、とにかくデータでいただければあとは流すだけですから、まあ OCR だけの料金は別途ということにはなりますが。それでも大量にかつ安く、ただ品質はそれだけおちますよ、というところをご納得いただけるお客様に対してだけです。それ以外のお客様に対しては、別に機械翻訳を使っていることがあっても単価を安く下げたりは今はしていません。

鎌田：それはうちも同じです。下げたくてもできないんですよ。現に、機械を使ったからといっても

単価は下がらないんですよ。下手したら上がるんですよ。

柏原：下手なやり方をしたら上がっちゃうんですよ。

司会：それは手翻訳と同時の品質を保証するレベルということですか？

柏原：結局、お客さまは「プロセスはどうでもいい。とにかくいい成果を出してくれ」ということですから。手翻訳と同じレベルになるプロセスを説明しますが、お客様のメリットとしてまずユーザ辞書ができてくるので、お客様としても色々な翻訳会社に出していて、それこそバラバラな品質なので、自社のリライトの負担が増えたのが緩和されるかなということですので、むしろ上げてもいいという所もあります。ですから MT の使い方として、エディットがきちんと必要なドキュメントと全く MT レベルのみの手放れのいい翻訳と分けて、その趣旨もお客さんに理解いただいてやれば、それぞれでマーケットがひろがると思います。やはりマーケットも使い方・目的も違うとおもいますね。

富永：コスト設定に対する辞書のメリットというのは大きいですね。私どもも特殊法人は会計検査を受ける法人ですが、指摘されることはアコギな商売をしてはならない、民間企業を圧迫してはならないということです。先ほどコストを 30% 程度ダウンさせたと申しましたが、その 30% のダウンを根拠もなく計算したかというところではなく、きちんと説明できなければなりません。30% の根拠は、人手でこれだけの仕事をかけて、これだけの品質を得ました。同じ品質を得るために MT ではこれだけの時間がかかりましたということで、時間が大きなファクターになります。時間短縮の説明には、50 万語の翻訳辞書を使ってこの言葉は他の分野ではこういう言い方をしてますよとザーッと見せる訳です。そうしますと文章上では、ある特定の分野の専門語をポンともってきて、その他

の専門語も同時に全部、下に見せる訳です。その後は後編集を頼んでいる先で、最適の用語を選択して下さいとって交渉します。

司会：訳語候補としてですか。

富永：多訳語という形で出力します。従来ですとその方の頭にあるものや手近にあるブック形式の辞書から取り出したのが、その手間がはぶけると説明します。ということで数千抄録の実験を致しまして、どのくらい作業効率が向上するか、時間が減っていくかということで、だいたいこれくらいじゃないかという当たりの値段で割り出していくと、30%ダウンぐらいではないかと計算し、「これでお引き受けいただけますか」ということで、お引き受けいただいていますので(笑い) 従来からの手翻訳会社さんが全部引受けていただいている上に、更に2,3社加わったという実績でありますので、まあ、それほどアコギなことはしていないと思います。

司会：ちょっと確認したいのですが、富永さんのところは、JOISのデータベースを英文としてサービスしているだけで、飛び込みの仕事はしていないんですね。

富永：現在はおしていません。飛び込みの仕事は手翻訳です。

司会：今は抄録ということですが、これは何かアクセスがあってから機械翻訳が始まるのですか。

富永：日本語の抄録がまず作られます。年間60万タイトルで、これはJOISで提供されていますが、その中の22~23万が日本で発行されている論文なのです。この部分を海外に提供するというので、60万件の中から取り出して、JOIS用データベース作りと同時並行で翻訳にかけています。ですから一方では日本語でサービスをし、一方では英語でサービスするためにMTを利用した作業をしているわけです。

司会：辞書を確か、売り切りの形でサービスされてますよね。

富永：はい、販売しています。30万語収録となっています。形式としては単純な日英対訳で、あと専門分野コードがはいっています。翻訳辞書とは全く違った形になっています。

鎌田：うちにも売り込みがありました。(笑い) 1,200万円で買わないですかときました。実は買いたいですよ。効果がありますから。

柏原：だいぶ売れているんですか。

富永：まあ、そんなにたくさんはありませんが、北は北海道、南は九州まで売ってはおります。

柏原：それは開発のベンダーさんばかりでなくてですか。

鎌田：額が高いですからね。一般の人にはとても手が出せません。

富永：翻訳辞書として汎用的なデータを収録しながら、ご提供申し上げることも考えなければならぬのでなかろうかということで、今は、30万語から離れまして翻訳辞書の中から必要なデータだけを抽出して、全く新しいフォーマットで提供する計画があります。

鎌田：おそらくMTのビジネスの中で、辞書が最終的なターゲットになると思います。翻訳アルゴリズムは研究者が考えればよい訳ですが、研究者がいくら頭をしばっても、辞書はつくれませんでしょう。ですからJICSTさんの辞書も今後生き残るための武器になると思いますし、私も回線経由でやっており、不特定多数のお客さんの指摘などから、辞書がどんどん良くなっていきます。

柏原：すごいですね。PC VANと比べると優位性をもっていますね。

鎌田：多少変でも何でもいいから10分で送り返してもらうのをお客さんは選んでいるということだと思います。ますね。利用している人はみんな、品質が悪いのは知っているんです。でも、レスポンスが早いから使っているのだと思います。辞書を引くより早いんですよ。私は事業

部を回って執筆指導もしますが、かなり優秀な人も辞書引きながら英語を書いているんですよ。英語を書くのは結構苦しいです。ですからうちの機械を使うと、何か変な英語だけど、スベルには間違いがないし、それに文節単位で見れば、自分で考えるよりもずっといい英語ができることもあるのです。機械だからと言って、バカにできないですよ。そしてうまく、組み直して文章を作っているんです。

柏原： できる人ほど自分の時間をセーブしたいと要求を持っていますし、MT は彼らの希望をかなえていると思います。話は変わりますが、MT では絶対成功事例を出さないといけないと思いますよ。うちのような規模の会社が MT に手を出しているとなると、間違い扱いされますよ。だいたい機械翻訳に触って潰れる方が多いといわれていますよね。だけど MT で潰れたというのはだいたい語弊があると思います。魔物だとか道楽と言われるのが一般的ですよ。でも何とか MT で、どこでもいいから成功事例が、つまり儲けるということですが、欠かせないと思います。MT が拓くべきマーケットというのはこれまでの翻訳世界と全然違うとおもいます。

鎌田： そうです。翻訳の業界と機械翻訳の業界は全く別物ですね。これはあきらかにデータベース産業ですよ。ですから辞書に力を入れてあたりまえです。

柏原： それで、仕事上の MT の利用法にかかわりますけど、マルチメディア関係のマニュアル翻訳を実際、MT でやっているんですよ。今の翻訳業界の状況から言いますと、MT をツールとして使う業者以外にはできないんですよ。というのは、言葉がこれまで存在しない用語が多いからなんです。今までの翻訳者にやらせるとダメなんですね。言葉も新しいし、お客さまの苦労も大変です。新しい辞書作りが同時にできます

し、逆に言うと、手翻訳者がダメなんで機械翻訳にすんなりとはいれます。お客さま自体が、まだ言葉を決めていない状態です。これを日本語にかえるかとか、カタカナのままでいくとか、例えばバーチャルリアリティにしても「仮想現実」とか「人工現実」とか言っていますが、人によっては「バーチャルリアリティ」と言っていますから、それはユーザさんによって使い分けはするんですが、その手の話はいっぱいある訳です。そういう問題を翻訳者に預けたら、いつまでもラチの明かないことになってしまいます。ですからそういう意味ではこれから、先端的な技術を開発していく人達には、MT は欠かせないし、武器になると思います。どこか1社でも MT を武器に成功してもらいたい。そういうところがあれば僕は協力してでもプッシュしなければと思いますね。

富永： 成功させるためにもということですが、私なんか毎日毎日、前編集のテスト、後編集のテストを眺めている訳ですが、言葉を辞書に追加するのは第一義的に大切ですが、機械翻訳がどうしようもないそと言う前に日本語がどうしようもないと言う必要がありますね。ですから、ぜひこういう機械翻訳協会さんの設立を機に、そういう機関から、いかにテクニカルライティングが大切か、日本語の書き方が大切かということ PR してほしいということをつくづく思います。辞書を増やすこと、文法の構文解析などのソフトも大切ですが、それよりも正しい日本語の書き方がどれだけ不足しているかということがために、今機械翻訳が中々上手くないんじゃないんじゃないかと思っています。しかし、元をただせば、言葉の不足もありましたが、なにしろ機械翻訳は自動翻訳とのイメージが強すぎましたね。

柏原： そうですね、あれでかなり悪いイメージとか制約を受けていますね。

富永： それででてくる英語は単語を並べているだけで、とんでもないとか。一つでも二つでも来知語でもあろうものなら、とんでもないということになる。やっぱり、その根底に大きく日本語という問題があります。

柏原： それは重要な問題で、根元的に言葉の問題に帰っていくと思います。私どもは今、英日しかやっていないのですが日英のニーズは高いと思うし、将来的にはターゲットとしたいのですが、私どものスタッフでは、日英からやる場合、日本語のプリエディットができないんです。それから、英日においても最終的には日本語のポストエディットがありますが、テクニカルライティングの分野は翻訳と同じで日本では軽視されてきました。だいたいその分野に予算というのをさいたというのはきいたことがないし、テクニカルライターというのもほとんど育っていません。そういった意味で言うと言葉というのは、非常に不当に軽く扱われている。英語圏の分野ですら、特定のコンピュータメーカーのマニュアルを見てますと同じシリーズでもライターが違ふと変わってしまうんです。向こうはライティングが早いし開発も早いので、調整なしで書いてしまうケースがままあるようなんですね。機械翻訳で対応する場合、前のシリーズのデータが使えないケースがあります。それを避けるため YHP さん、IBM さんは先進的に取り組んでいますね。ドキュメンテーションのスタートの時点から文章の情報をきちんと考えているわけです。機械翻訳を成功させる為には、きちんとしたきれいな標準的な文章を書くことから始めることが必要だと思います。特に日本はね。ですからシステムが読み易いような書き方をするということです。

鎌田： それは非常に明るいと思います。なぜかというところ、私のところでは延べにして 3 万人近い利用者がいて、色々なクレームもあります。これ

だけきれいな日本語を短く書いているのに何故上手く行かないのだとのクレームもありますが、これはもう少し機械翻訳システムの開発者に頑張ってもらうしかありません。しかしよく使っている人ほど本当に上手く文章を書いています。

司会： 日本では国語の時間に作文というのがありますが、作文を書く技術についてはきちんと教えられていませんね。こう書きなさい、こう書くという明確に伝わりますよというのはありませんね。その点、多民族国家のアメリカでは大学にテクニカルライティングコースというものがあって、一つの学問になっているそうです。万人に伝わるような書き方の探求です。翻訳は元原稿が悪いと、翻訳されたものはもっと悪いということになります。

柏原： コミュニケーションというのは日本というか日本人は大事にしていらないと思いますね。単一民族と多民族国家の違いがあっても、やはり、言葉をつくしていうとか、きちんとすることをいやがる風土があって、その結果、コミュニケーションが悪くなっていると思います。で、これを埋めることは容易でないにしても、機械翻訳事業にたずさわる方々は同時にプロジェクトを進める中でそういう問題も解決しなければと思うんですね。

鎌田： 富士通の事業部で、機械翻訳を使わなければいけない。使わなければ翻訳しないというプロジェクトがありました。80 人のプロジェクトだったんですが、その人たちが作ったドキュメントをすべて機械にかけて、その結果をその人たちに全部送り返したんです。複文や重文ならばなるべく分解しろというふうに機械翻訳の為のライティングを教えたんです。その時のコストダウンは 36% でした。ものを作る時から執筆を含めた QC をやったことは効果がありました。少なくとも社内ですらやった経験では、10 人中 8、9 人は正しい日本

語を書けます。で、今ではその事業部では機械で翻訳し、その英語でアメリカ人に通じると思うまで日本語を書きかえて、その日本語をベースにやるんだとしています。

富永： 私どもも同じようなことをやっておりまして、日本語がいかに大切かという大論争の後、機械翻訳をもう少しおもちゃ的に使ってもらって、あなたの書いた日本語がどう訳されるかということが、試せるように端末を自由に使ってもらっています。さかんに面白半分に使っていますが、その結果がだいぶ貯まってまして、「こう書けば、こう出る。」というのが山ほどあります。ひょっとしたら一冊の本にできるぐらいだと思います。正しく、わかりやすく、短くていい文章というものは、それぞれシステムによってはちがうと思いますが、随分似たようなところがあると思います。ですから機械翻訳協会ができたのを機に、そういう日本語の書き方、機械翻訳向けの日本語の書き方というものを、データをとって集めて、なんとかしていただきたいものですね。

司会： 国際化を円滑に進めるという中に、日本人にしかわからない文章を残しておいてそれが翻訳されないものだから外国に「かくしている。かくしている。」と言われていきます。富士通さんなどは、どういう局面であってもこれを英語に直せといわれたら、直さなければならない状況にあると思います。将来必ず英語に直すというのだったら機械翻訳にかけるためには、ある程度の執筆基準があって、それに従って文章を書いていくことが必要だと思います。その場合、その執筆基準というのは自然でかつ十分記述力があるというのが条件となると思います。いかにも機械に合わせたものというのはまずいでしょうから。

鎌田： 今まで、いろんな企業がライティングスタンダードというものに取り組んで、ほとんど失敗

しています。それは人間の感性を無視していると思うんですよ。例えば使える動詞はこれだけです、といってもダメなんですね。アメリカとは異なり日本語のライティングスタンダードは難しいですよ。国語の先生だって、我々の言っているようなことを聞いたら反発しますよ。

柏原： その点は文化のバックグラウンド抜きでは考えられませんよ。でもそういうことは文部省の管轄だったら駄目だと思いますよ。やはり実践というか現実とかけ離れたところで議論していても駄目ですね。

鎌田： だから MT は一つの道具、これを使ったらあなたに絶対メリットがありますよというような言い方で、社内の若い人に使ってもらえないんじゃないですか。機械翻訳を通して、あるレベルの、それなりの英語がでるまで日本語を書き換えてもらう。ある事業部の数百人がやっていますが、これが成功したら、億円単位のコストダウンになります。これは設計なんです。今時、設計といっても、海外と分担して設計することも多いのです。ある部分は日本で、ある部分はアメリカでということですから、向こうは英語で書いて、こちらは日本語でといったら、設計変更があった時には会議ばかりになってしまいます。だから、日本語と英語が完全にシンクロできるようにしなければならないのです。そのメリットは、翻訳コストどころか、翻訳コストはメじゃないですね。最大のメリットは設計の失敗がなくなることです。今まで日本語で書いても、日本人同士お互いわからない設計書になっていたのが、設計者が機械翻訳を通すために日本語を分かりやすく書くことから、理解が早く正確になるわけです。

柏原： そういう意味では鎌田さんのところは自社内にそういうニーズを抱えているのは優位性がありますね。

鎌田： そうです。今、私は外に向いていないで内に向いています。内側で使いこなせないで、何でお客さんのところへ提案に行けるんだ、ということとです。

(つづく)

【1992年5月 JAMT ジャーナル No.6 より】

サグラダ・ファミリアにも似て...

MT 誕生は 83 年と予言

電子翻訳機（ただし、「学力は大学程度」との断り書きあり）—その試作時期は 1973 年（昭和 48 年）ころ、実用と普及の時期は 1983 年（同 58 年）ころ、と予測した記録がある。これは、日本電子工業振興協会が 1968 年（同 43 年）9 月に発表した「電子工業の長期展望」というタイトルの調査書に掲載されているデータである。

わかりやすく説明すると、いまからおよそ四半世紀前に、いわゆる電子翻訳（現在の一般呼称では、マシントランスレーション/機械翻訳/MT）というテクノロジが、その時点から 3 年後に試作が始まり、15 年後にはすでにシステムの実用化が具現されて普及への一途をたどるだろう、と想定した事実を物語っている。

「学力は大学程度」とわざわざ断っているが、これはなかなか意味深長だ。学力という表現の自体がおかしな響きをもっているが、それはそれとして、「大学程度」とは中学、高校とだんだんレベルが上がって最高級の能力を有しているということなのか、それとも逆にせいぜい現今の大学生くらいの能力だから、とても満足すべきシロモノではない...と遠慮しているのか、そのあたりが判断に苦しむ。

ところで、この調査は電子翻訳のみを扱ったわけではなく、未来技術や未来システム全般について行われたものだ。当然のことながら、エレクトロニクス 1 コンピュータを駆使する＜スマート＞な技術/システ

ムが目白押しに取り上げられ、その数は全部で 22 項目もある。その中で、電子翻訳という技術は実現時期順で 17 番目の位置として予測され、前後には「キャッシュレス/チェックレス社会—普及時期 1981 年」、「家庭用ファクシミリの普及—同 1985 年」などが並んでいる。脚注に＜デルフィ・メソッドによる調査＞と記してあるから、かのデルファイ法 (Delphi technique) を適用した専門家たちの反復アンケート調査であって、消して無責任なト占のたぐいではない。

予言は当たった

さて、時は流れる。全国紙の A 新聞がその一面トップに「自動翻訳いよいよ開始」という意味のばかでない見出しをかかげて関係者の度肝を抜いたのは、1984 年 5 月 18 日付けの朝刊である。この日の朝は、とりわけ天気がよく、文字どりのさわやかな五月晴れ。一見したところ業界紙と見紛うばかりの紙面構成と記事内容は、当時生身のトランスレータであり翻訳会社経営者であったわたしにとって、まさに青天の霹靂だった。読めば読むほど、私の頭の中には、失業・倒産・離散・物乞いなどといった絶望的なことが声高く飛び交う。そして、その渦の真ん中で、何ともやりきれないひとつの決断が生まれた—これはいま、妻に見せてはならない！

こうして、新聞はかばんの奥深くにかくされることになる。その結果、幸か不幸か、彼女はこの記事の存在はもちろん、主人のおぞましい狼狽ぶりをいまだに知らない。

同年、夏から秋にかけては、業界紙（誌）や週刊誌がさらに追い打ちをかけ、翻訳における人間敗退説ないし不要説を強調する。いわく、「自動翻訳機の登場で語学力の革命」「早くも英語屋の失業が始まった」「自動翻訳機をついに完成させた Y 社長の言葉の壁への挑戦」（いずれも特集記事の大見出し）など、など。

ともあれ、こうして電子（機械）翻訳機の実現予測は、数字上はわずか 1 年の遅れで＜現実＞と符号する

ことになった。1984 年、その時点からもどって 16 年前の読みの正しさが立証され、コンピュータ応用技術としては日本における最初の発明品<機械翻訳システム>（厳密な意味での翻訳機の開発史から見れば、改良発明品）がここに誕生したのである。

生まれた子供たちはいとしいが...

第一子に続き、遅れてはならじとばかりに第二子、第三子...が雨後の竹の子のごとく誕生する。この現象は、MT 開発戦争において単純に先を越されたあせりもあったかもしれないが、第二子以下しばらくの間、その親たちがこの道では正統派(?)の電子機器メーカーであったことを考えると、その相次ぐ出産はやむをえぬ事情もしくは圧力、つまり国の指導や資金援助のからみがあったのころだろう。

さて、これで、めでたしめでたしの大団円となればよかったのだが—。残念ながら、このストーリーにはく実は.....>というおまけがつく。生まれた子供はすべて、能力的に見てあちこちに欠陥があるばかりでなく、生まれつきの疾病をたくさんもった病弱児であったり、また全体的にバランスが取れていないので、一筋縄ではとても面倒を見きれない性悪っ子だったのだ。

このように書くと、機械翻訳開発におけるせつかくの結晶に対して、たいへん不遜かつ不埒な言を弄しているように見えるが、これはあくまでも起承転結の<転>を述べている便法のひとつにすぎない。開発側はもちろんのこと、わたしを含む使用者側も生まれでたものにいとしさを感じないはずはない。かわいければ、どうしたらよいかというのが現実の思惑であろう。

最後のきざしははるかなる彼方

視点が誤っていると、結論も当然まちがう。機械翻訳の場合、いちばん重要な視点は、人間の翻訳と機械による翻訳とは次元が異なる、ということである。夢とか永遠の目標として描くならば、このふたつの一致

を課題にするのはよいが、少なくとも今の段階では無意味である。どのように次元が異なるかについては措くとして、かりに機械翻訳技術が次のターゲット年度である 2020 年に格段の進展をとげても、それは機械ができることを成し得た、あるいは機械にしかできないことを達成した結果であって、人間の翻訳とは明かに一線を画しているであろう。

このような認識が一般の人たちにも植え付けられないと、機械翻訳に関する批判はいつまでたっても両極端にとどまってしまう。1984 年に私自身が体験したような錯覚（すなわち、機械による翻訳力は人間の言語活動能力と同じかそれ以上でなければ藪に馬鉄のたとえに等しい、と誤解していること）は、この世界から迫放しなければならない。

現行システムのいわゆる<欠陥>については、これを永久欠陥とは考えずに、<未完成>とみたらどうだろう。いま、わたしたちは終りの見えないきざしを登っており、その途中々々の位置を何度か有意の時期に発表して世に問うことこそ義務と権利である、と自覚することだ。

シューベルトの未完成交響楽は、かれの死によってあたかも完成したかの姿で終りを告げたが、機械翻訳はおそらくガウディーのサグラダ・ファミリアのように、今後 100 年も 200 年もかけて継承構築すべき性質のものにちがいない。

((株) CSK 技術開発本部
MT 開発部顧問 水上達郎)

MT Summit 2023

Sumire Honda

University of Potsdam

1. Overview

This report summarises my online attendance at MTSummit2023, which was held in Macau SAR, China, from 4 to 8 September. The conference provided a valuable platform for experts, researchers, and industry professionals to share insights, exchange ideas and explore emerging trends in machine translation. This report outlines my online participation, networking opportunities and overall takeaways from the conference.

2. Online Participation

Participating online was a great opportunity as it allowed people from various locations to join. What made it even more special for me was the student scholarship I received for the online attendance fee, which eased the financial concerns that students often encounter at such events. Thanks to this support, I presented our study *Context-aware Neural Machine Translation for English-Japanese Business Scene Dialogues* at Research Track on 8 September.

The event used a combination of Whova and Zoom for the online platform. Whova was easy to use, enabling the networking system and simplifying the participation process, while the integration with Zoom ensured a smooth conference experience for all participants.

3. Networking Opportunities

The conference featured both a research track and a user track, providing valuable networking opportunities for attendees, which was very beneficial. This set-up encouraged links between industry professionals and researchers, fostering collaboration and knowledge sharing.

In addition, the conference provided a platform for

several job postings, enhancing career development opportunities.

Even in the online format, participants were able to use chat rooms to connect with other people and engage in meaningful discussions. For those attending the conference onsite, there were also further social events to enhance the networking experience.

4. Overall Takeaways from the Conference

Attending the conference gave me valuable insights into the field of machine translation, particularly in areas that are less familiar to me. For example, evaluating the performance of industries can be challenging due to the absence of standardized evaluation criteria in machine translation, making direct comparisons difficult. Therefore, research focused on establishing evaluation methodologies currently holds significant importance.

Additionally, it's important to acknowledge that the datasets used in real-world industrial settings may differ substantially from those typically employed in academic research. For example, when working with website data, there is a need to deal with HTML tags. This underlines the necessity for tailored and customized approaches to effectively handle such diverse datasets.

Furthermore, the constructive feedback I received during the presentations is very helpful for my future work. The conference allowed me to identify aspects of my research that I may have overlooked from a different perspective.

MT Summit 2023

本多純礼

ポツダム大学

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License.
License details: <https://creativecommons.org/licenses/by-sa/4.0/>

イベント報告

AAMT Machine Translation Summit 2023 参加報告

橋本 簡美

立教大学異文化コミュニケーション研究科修士課程

1. AAMT Machine Translation Summit 2023 の概要

AAMT Machine Translation Summit 2023 は、9 月 4 日から 9 月 8 日の 5 日間にわたり、中国のマカオで開催された。基調講演、ワークショップ、ポスター発表などが行われ、オンラインに加えてオフラインでの配信も行われた。本会議は、機械翻訳に関わりのある人たちの意見交換や知識共有を目指したものであった。本稿では、報告者が参加したセッションの発表内容について報告する。

2. User Track Session 3: New workflows with MT/LLMs

本セッションでは、機械翻訳や大規模言語モデルを用いた新しいワークフローについての発表が行われた。中でも興味深かったのは、Qian & Zhang (2023) による “The impact of machine translation on the translation quality of undergraduate translation students” である。本研究では、翻訳学習者が 1 から人手翻訳をした場合と、機械翻訳のアウトプットをポストエディットした場合の訳文のクオリティを比較している。結果として、ポストエディットの場合は人手翻訳と比べて、正確性エラーが多く出現していることが分かった。またポストエディットで発生したエラーは人手翻訳の場合と比較して、エラー内容が均質的であることも分かった。ポストエディットに正確性エラーが多く、エラー内容がより均一であるということは、ポストエディットの際に出現するエラーの多くは、機械翻訳で出現したエラーがそのまま残ったものであ

ると考えられる。つまり、翻訳学習者は機械翻訳のエラーを修正するより前の段階である、機械翻訳のエラーを見つけるという段階で躓いている可能性が示唆されている。

さらに興味深い点は、翻訳学習者はポストエディットを行った場合の方が人手翻訳を行った場合よりも、正確性エラーが多かったにもかかわらず、より自信があったということである。先述したように、翻訳学習者はポストエディットをした際、機械翻訳のエラーを見逃している可能性がある。つまり翻訳学習者は、機械翻訳のアウトプットに発生したエラーに気付かず、その品質を過大評価している可能性が考えられる。

機械翻訳の普及が広まるに伴い、ポストエディットの需要も増えており、翻訳学習者へのポストエディットの教育も進めていく必要があるだろう。Qian & Zhang (2023) の研究は、翻訳学習者がポストエディットを行う際には、そもそも機械翻訳のエラーをスポットすることが難しいのではないかということを示唆している。また、学習者はポストエディットをしている際に自身がエラーを見逃している可能性に気付いていないということも考えられる。今後は、こうした問題点に対する具体的な解決策を考える必要があるだろう。

3. おわりに

以上が、報告者が参加した AAMT Machine Translation Summit 2023 とその中で特に興味深かった発表についての報告である。機械翻訳が急速に発展していく中で、翻訳コミュニティに多様な情報が共有されることは、さらなる機械翻訳の成長と翻訳研究の発展に繋がっていくだろう。

イベント報告

MT Summit 2023 参加報告書

平岡 裕資

立教大学異文化コミュニケーション研究科博士後期課程在籍

1. はじめに

本稿では、2023 年 9 月 4 日から 9 月 8 日まで開催された MT Summit 2023 について報告する。MT Summit は、民間および教育分野における機械翻訳 (MT) の開発者や研究者、機械翻訳のユーザー (翻訳者) など、多様な背景を持った人々が集う国際会議である。今回で第 19 回目となる本会議は中国・マカオで開催され、世界各国からおよそ 200 名の参加者が集まった。最初の二日間はワークショップのみの開催となり、その後の三日間は招待講演、パネルディスカッション、個々の研究発表が行われた。日程や登壇者の情報はすべてイベントアプリ (Whova) 上で参照でき、録画やスライドなども閲覧できたため非常に便利な仕組みであった。

2. Invited panel (山田優氏 & Felix do Carmo 氏主催)

以降では印象的であったセッションについて述べる。1 つ目は 9 月 6 日に行われた山田優氏と Felix do Carmo 氏によるパネルディスカッションである。テーマは『Artificial Intelligence and Translation: effects on theories, professional practices and education』であった。Chan Sin-wai 氏、阪本章子氏、Kirti Vashee 氏の 3 名による発表が行われた後、発表者 3 名と司会 2 名でディスカッションがなされた。トピックは機械翻訳の翻訳品質から訳文の著作権など多種多様であったが、共通した主張は他分野間および異なるステークホルダー間の協働が必要であるということであった。この考え自体は目新しいものではないものの、その重要性が再認識されたことには意味があるだろう。

3. Dorothy Kenny 氏による Keynote speech

9 月 7 日には Dorothy Kenny 氏による基調講演が行われた。Kenny 教授は Dublin City University で翻訳研究に従事しており、機械翻訳関連では『Machine Translation for Everyone: empowering users in the age of artificial intelligence (2022 年)』という書籍で言語教育や翻訳業界など機械翻訳が影響を与える分野において重要な指針を示している。基調講演のタイトルは『Translating machines: implications for translation theory and practice』で、その内容は翻訳学の理論を用いて機械翻訳の可能性と限界を説明するものだった。例えば、機械翻訳はその仕組み上、context よりも co-textual な情報を捉えることに長けていることや、機能主義的理論 (Nord, 1995/2018) の観点から Yamada (2023) によるプロンプト例を用いて ChatGPT の可能性を説いた。このような翻訳学の理論を援用した説明は機械翻訳による翻訳の性質を適切に解釈することができるため、非常に有効な手段であると感じた。

4. おわりに

今回の MT Summit では他分野間の協働に前向きな雰囲気を感じられた。特にユーザートラックでは開発者からの質問が積極的になされていたのが印象的であった。自分自身も機械翻訳の知識が深まり、翻訳を多様な視点で考えられる良い機会となった。

Participation Report on MT Summit 2023

Yuki Yamashita

Graduate School of Information Science and Technology, The University of Tokyo

1. Introduction

To stay in touch with the latest developments in the machine translation field, I attended MT Summit 2023 online with student financial support from AAMT. In this report, I take up some important and interesting conference presentations.

2. Overview of MT Summit 2023

The 19th Machine Translation Summit (MT Summit 2023) was held in Macau on Sep 4-8, 2023, in a hybrid style. This was a biennial international conference, organized by AAMT this year, which brought together researchers, developers, and users of machine translation. The conference overviewed state-of-the-art technologies, major contemporary trends, and practical applications. There were 3 keynotes, 7 invited talks, 4 workshops, 19 research track oral presentations, 22 user track oral presentations, and 14 poster presentations.

3. Keynote

There was a keynote talk “Toward Better Neural and Big Model-based Machine Translation” by Prof. Min Zhang from Harbin Institute of Technology. He listed outstanding contextual understanding, interactive translation capabilities, versatility in text handling, superior document-level translation, and multilingual capability, as advantages of big model-based machine translation. Also, he pointed out challenges in automated evaluation, lack of domain expertise and difficulty in knowledge

transfer, computational expense, and suboptimal performance in low-resource languages, as remaining issues. To overcome these limitations, he introduced recent advances such as candidate- and entity-aware neural evaluation metrics [1], test-time adaptation by uncertainty [2], curriculum consistency learning, and complementarity between prompting and tuning. He concluded that potential future directions of machine translation with big models are multimodal translation over three or more modalities and a core downstream task for big model research.

4. Invited Talk

Prof. Tong Xiao from Northeastern University and NiuTrans gave a talk “Translation between Chinese and Over 300 Other Languages: Methods and Systems” in the invited session on machine translation for Asian languages. He introduced the minor languages in China and referred to the differences in the amount of bilingual data resources. He explained knowledge transfer from high-resource languages to low-resource languages derived from similarities among Asian languages. Also, he described the model architecture, specific text processing, and the improvement in translation quality. He also mentioned that large language models are good few-shot multilingual translators for high-resource languages. On the other hand, they are not well-trained for low-resource languages, and the improvement through few-shot demonstrations or bilingual dictionaries is modest.

MT Summit 2023 参加報告

山下 優樹

東京大学大学院情報理工学系研究科

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License.
License details: <https://creativecommons.org/licenses/by-sa/4.0/>

5. Workshop

The 10th Workshop on Asian Translation (WAT 2023) is an open machine translation evaluation campaign focusing on Asian languages. There was a talk on task descriptions and participation reports. Team BITS-P introduced their work for multimodal translation with image augmentation using diffusion models [3]. They used Stable Diffusion [4] to generate a synthetic image for each caption in the training set and doubled the size of the training set.

6. Oral Presentation

In the user track session on new workflows with machine translation and large language models, human-in-the-loop pipeline that guides LLMs to produce customized outputs with revision instructions [5] is proposed. The pipeline initiates by prompting the LLM to produce a draft translation, followed by the utilization of automatic retrieval or human feedback as supervision signals to enhance the LLM's translation through in-context learning. Revision instruction is automatically synthesized based on the optimal edit path between the reference and generated translation.

7. Conclusion

There were various kinds of topics including case studies and practical applications in industry as well as research. For the recent trend, large language models have gained strong interest from many researchers or engineers. The conference focused mainly on Asian languages, and there were contributions to the corpus of low-resource languages and to translations with cultural backgrounds. Since machine translation technology is currently studied mainly in European languages,

this kind of work on Asian languages is important for global universality and equality.

References

1. Liu, X., Wang, Y., Wong, D. F., Zhan, R., Yu, L., & Zhang, M. (2023, July). Revisiting Commonsense Reasoning in Machine Translation: Training, Evaluation and Challenge. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 15536-15550).
2. Zhan, R., Liu, X., Wong, D. F., Zhang, C., Chao, L. S., & Zhang, M. (2023, July). Test-time Adaptation for Machine Translation Evaluation by Uncertainty Minimization. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 807-820).
3. Dash, A. R., Gupta, H. R., Sharma, Y., & IQVIA, B. (2023). BITS-P at WAT 2023: Improving Indic Language Multimodal Translation by Image Augmentation using Diffusion Models. In Proceedings of the 10th Workshop on Asian Translation, (pp. 41-45).
4. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10684-10695).
5. Yang, X., Zhan, R., Wong, D. F., Wu, J., & Chao, L. S. (2023). Human-in-the-loop Machine Translation with Large Language Model. In Proceedings of Machine Translation Summit XIX, Vol. 2: Users Track, (pp. 88-98).

イベント報告

第 10 回特許・技術文書翻訳ワークショップ (PSLT 2023) 開催報告

綱川 隆司¹須藤 克仁²後藤 功雄³¹ 静岡大学² 奈良先端科学技術大学院大学³ NHK

1. 開催概要

2023 年 9 月 5 日、中国・マカオにて開催された機械翻訳サミット (Machine Translation Summit) 2023 の併催ワークショップの一つとして、特許・技術文書翻訳ワークショップ (PSLT 2023; The 10th Workshop on Patent and Scientific Literature Translation) が開催された。本ワークショップは、アジア太平洋機械翻訳協会(AAMT)、および一般財団法人日本特許情報機構(Japio)による AAMT/Japio 特許翻訳研究会が中心となり 2005 年から隔年で開催しており今回で 10 回目を数える。前回 2021 年の第 9 回ワークショップは新型コロナウイルス拡大の影響で完全オンライン形式での開催となったが、今回は機械翻訳サミット本会議と同じく現地とオンラインのハイブリッド形式での開催となった。

本ワークショップでは特許庁、機械翻訳ユーザー、および機械翻訳研究者のそれぞれの立場から 3 件の招待講演をいただいた。現地にて 22~23 名程度、オンラインで 15 名程度の参加があり、質疑も活発であり盛況であったと言える。

2. 招待講演

招待講演 1 件目は日本特許庁の名和大輔氏から、特許情報ユーザーのための特許庁の情報サービスに関する講演をいただいた。特許庁では即時に正確な情報を提供するため多様な情報提示を行っている。特許情報プラットフォーム (J-PlatPat) は特許庁が運営する日本で最大の知的財産データ検索サービスであり、その中で海外特許および日本特許のニューラル機械翻

訳等による翻訳結果を提供しており、企業等に向けた API を介したサービス提供も行っていることが紹介された。

招待講演 2 件目は中国の知的財産権出版社 (IPPH; 知识产权出版社) から YAN Jiang 氏による、IPPH 及びその特許機械翻訳サービスである“中知慧译” (ZhongZhi HuiYi) のご紹介があった。IPPH は 2010 年に特許向け AI・機械翻訳サービスの提供を開始し、2023 年には中国の知的財産情報を扱う最大のメディアグループとなっている。特許機械翻訳の課題である省略、長い文章、特許に特有のパターン、専門用語の一貫性等に対応する、ニューラル機械翻訳ベースの 10 言語に対応した機械翻訳サービス“中知慧译”は、1 日に 10 万件の特許を翻訳しており、中国語と英語、日本語、ドイツ語等の特許翻訳において高い性能を示している。

3 件目の招待講演は澳門大学の Derek W. Wong 氏による、彼らの研究グループにおけるニューラル機械翻訳の研究に関する下記の研究内容についてのご発表であった。

- 不確実性を考慮したカリキュラム学習
- 文ベクトルのノルムによるカリキュラム学習
- 自己適応学習
- 分野適応のためのメタカリキュラム学習
- 難易度に基づく機械翻訳評価
- 分散を考慮した機械翻訳評価テストセット

ニューラル機械翻訳モデルの機械学習においては、人間に対する効果的な学習方法を模したカリキュラム学習の有効性が知られている。効果的なカリキュラムとして、訓練データの難易度や不確実性を考慮した手法や、モデルが自己の学習を調整するアプローチにつ

いて、翻訳性能やその収束速度が評価、分析されている。また特定の分野に適応させるため、カリキュラム学習の方法を学ぶメタカリキュラム学習を取り入れ、過学習を軽減し、長文に対してより優れた性能を示している。また、機械翻訳システムはテストセットにおいて正解訳との類似性を求めることで性能評価されているが、テストセット中の文の翻訳の難しさに応じた評価や、テストセットを分散指標により選別することでよりバラツキの多いものにし、翻訳システムの弁別力を測る評価指標を提案している。

3. 所感

本会議前のワークショップとしては比較的多くの方にご参加いただけたこと、また特に現地会場での質疑が活発であり、プログラム編成時に想定していた総合討論の時間をも活用して招待講演に対する質疑が続いたことはワークショップ提案者としては喜ばしい。招待講演が1件オンライン、1件が録画放映であった点はワークショップの盛り上がりの面で懸念していたが、ハイブリッド開催に対する参加者の慣れもあってか、大きな障害にはならなかったと言える。

一方で、一般講演については論文投稿がなく招待講演のみの短時間での開催になったことについては残念であり、今後の企画に向けた大きな課題である。ニューラル機械翻訳の進展を上回る勢いで急速に活用が広がる大規模言語モデルは、特許や技術文書の翻訳やその活用の上でも重要な役割を果たすことがほぼ確実であり、応用事例も含めて幅広く技術的課題の解決に向けた議論の場として本ワークショップをご活用いただくべく、募集分野やワークショップ名の再検討を行った上で引き続き実施していきたいと考えている。



イベント報告

第 10 回アジア翻訳ワークショップ (WAT2023) 開催報告

中澤 敏明

東京大学

1. はじめに

本稿では WAT2023[1]の開催報告を行う。アジア翻訳ワークショップ (Workshop on Asian Translation, WAT) はアジア言語を中心とした評価型機械翻訳ワークショップであり、2014 年に第 1 回 (WAT2014) を開催して以降、毎年開催している。本稿の著者は初回からオーガナイザーの一人としてワークショップの運営を行っている。2016 年の第 3 回 (WAT2016) 以降は自然言語処理の国際会議との併設ワークショップとして開催しており、2023 年の第 10 回 (WAT2023) は中国のマカオで開催された MT-Summit 2023 の併設ワークショップとして、2023 年 9 月 4 日にオンラインと現地でのハイブリッド形式で開催された。参加者は現地参加とオンラインと合わせて約 50 名程度で、現地参加が 2/3 ほどであった。

2. 研究論文

WAT2023 では 1 件の研究論文を採択した。採択した研究論文のタイトルは “Mitigating Domain Mismatch in Machine Translation via Paraphrasing” である。

本論文は入力文の言い換えと翻訳結果のリランキンクにより、訓練データとテストデータでのドメインミスマッチの影響を低減する方法を提案している。言い換えは単語レベルの言い換えと文レベルの言い換えを用いている。提案法は機械翻訳モデルのファインチューニングが不要であり、オンラインの翻訳サービスなどのブラックボックスな機械翻訳に対しても適応可能であるという特徴がある。

実験により提案手法の有効性が示され、特に言い換えの手法に関しては文単位での言い換えよりも単語単位の言い換えの方が精度が高くなる傾向にあることがわかった。これに関して会場にいた参加者から「単語単位の言い換えの方が精度がいいという結果は LSP (Language Service Provider) にとっては朗報で、他のドメインでの効果も検証してほしい」とのコメントが出た。これはおそらく、単語単位での言い換えは LSP が独自に持っている辞書などを用いることで容易に行うことができるため、提案手法が実用的にも期待できるからではないかと思われる。

3. 招待講演

招待講演は Wikimedia Foundation の Principal Software Engineer である Santhosh Thottingal 氏より “Machine Translation at Wikipedia” というタイトルで行われた。Wikipedia は 320 以上の言語で提供されており、記事の翻訳に機械翻訳を使用している。さまざまな機械翻訳サービスを利用することができる統合ツールにより最初の翻訳を提供し、編集者による修正を経てから翻訳された記事が公開される。この方法で現在までに約 160 万件の記事が翻訳されている。

講演ではテキストのみの機械翻訳と人手による情報のキュレーションを組み合わせることで高品質なリッチテキスト翻訳を提供することに焦点を当てて、ヒューマン・イン・ザ・ループの製品設計が紹介された。また翻訳品質と翻訳者に関する洞察と分析が共有された。フリーでオープンソースのシステムから、セルフホスト型のサービス、外部の有料 API まで、Wikipedia で採用している機械翻訳エンジンが紹介さ

れ、これらを用いることで 198 の言語を翻訳していると述べられた。

Thottingal 氏の講演で用いられたスライドは WAT2023 のホームページ⁶でダウンロードすることができるので、興味のある方は是非ご覧いただきたい。

4. 翻訳タスク

WAT2023 では以下の Shared Task を設定した。

- ・ 文書単位翻訳タスク：
 - ASPEC+ParaNatCom: 英 → 日 科学技術論文
 - BSD Corpus: 英 ⇄ 日 ビジネスシーン対話
 - NICT-SAP: ヒンディー/タイ/インドネシア/マレー/ベトナム ⇄ 英 非構造的文書
 - 日中韓 ⇄ 英 構造的文書
- ・ マルチモーダル翻訳タスク：
 - Hindi Visual Genome: 英語 → ヒンディー語
 - Malayalam Visual Genome: 英語 → マラヤーラム語
 - Bengali Visual Genome: 英語 → ベンガル語
 - Ambiguous MS COCO: 英 ⇄ 日
 - Video Guided Ambiguous Subtitling: 日 → 英
 - Khmer Speech Translation: クメール → 英/仏
- ・ インド諸語翻訳タスク：
 - アッサム/ベンガル/グジャラート/ヒンディー/カンナダ/マラヤーラム/マラティ/ネパール/オリヤー/パンジャブ/シンド/シンハラ/タミル/テルグ/サンタル/カシミール/マイティリー/サンスクリット/ウルドゥ ⇄ 英
- ・ 特許翻訳タスク：
 - JPC3: 英中韓 ⇄ 日
- ・ 制限翻訳タスク：
 - ASPEC: 英/中 ⇄ 日
- ・ 対訳コーパスフィルタリングタスク：
 - JParaCrawl: 日 ⇄ 英
- ・ 非反復翻訳タスク：
 - JLIJ Corpus: 日 → 英

残念ながら今年度は 2 チームの参加にとどまった。またどちらのチームもインドの言語のマルチモーダル翻訳タスクに参加した。表 1 に参加者のリストを示す。

Team ID	Organization	Country
ODIAGEN	Odia Generative AI	India
BITS-P	Birla Institute of Technology and Science, Pilani	India

表 1 : WAT2023 Shared Task 参加者リスト

評価の結果、3 つのタスク全てにおいて WAT2023 の精度は WAT2022 での精度よりも良くなっていること、テキストのみを用いる手法とマルチモーダルな情報を用いる手法との精度の差はほとんどないことなどがわかった。

5. まとめ

本稿では WAT2023 全体の概要を報告した。残念ながら WAT2023 では論文の投稿数や翻訳タスクの参加者は非常に少なかったが、ワークショップ当日の参加者はそれなりにあった。また招待講演、研究発表、翻訳タスクの結果それぞれで議論も活発に行われたことから、ワークショップとしては成功だったと言える。

WAT は今後も継続して開催予定であるが、来年度の開催予定は未定である。来年度はより多くの参加者が集まるよう、工夫していきたいと思う。また WAT では翻訳評価にかかる費用等のためのスポンサーを募集しているため、興味のある方はご連絡いただければ幸いである。

⁶ <http://lotus.kuee.kyoto-u.ac.jp/WAT/WAT2023/index.html#invited-talk.html>

参考文献

[1] Toshiaki Nakazawa, Kazutaka Kinugawa, Hideya Mino, Isao Goto, Raj Dabre, Shohei Higashiyama, Shantipriya Parida, Makoto Morishita, Ondrej Bojar, Akiko Eriguchi, Yusuke Oda, Chenhui Chu and Sadao Kurohashi. 2023. Overview of the 10th workshop on Asian translation. In Proceedings of the 10th Workshop on Asian Translation (WAT2023), Macau SAR, China. MT-Summit 2023.

ChatGPT を翻訳業務に活かすには

中山 雄貴

株式会社ヒューマンサイエンス

1. はじめに

ChatGPT の自然言語処理能力は高く、翻訳業務のさまざまな場面で活用できます。本記事では、ChatGPT のセキュリティについての留意事項と、プロジェクトマネージャー (PM)、翻訳者、レビューアの担当業務での活用方法を説明し、弊社の機械翻訳製品である MTrans for Office (エムトランス・フォー・オフィス) と MTrans for Trados (エムトランス・フォー・トラドス) について紹介します。

2. セキュリティについて

ChatGPT をウェブブラウザから使う場合、デフォルトでは入力したデータが ChatGPT のモデルの改善に使用されるため、機密が保持されません。データの二次利用を回避するには、左下の「...」→「設定」→「データ制御」→「チャット履歴とトレーニング」のスイッチをオフにしてください。詳細については、<https://openai.com/blog/new-ways-to-manage-your-data-in-chatgpt> (英語) をご覧ください。

ChatGPT を API 経由で使う場合、データは二次利用されず、機密は保持されます。詳細については、<https://openai.com/enterprise-privacy> (英語) をご覧ください。

3. PM としての活用法

翻訳スタイルガイド作成：翻訳者やレビューアが使用するスタイルガイドを生成できます。既にスタイルガイドがある場合は、記載内容に不足がないかどうか

確認するためにも利用できます。プロンプト例：「翻訳用のスタイルガイドを作成して」

用語集作成：文章内の重要な語句を抽出して用語集を作成できます。従来の用語集作成ツールでは出現頻度や品詞などの情報から用語を抽出していましたが、 unnecessary 用語が多数抽出されることがよくありました。ChatGPT を使うと、重要と思われる用語のみを抽出できるため、より効率的に用語集を作成できます。プロンプト例：「文章の中にある重要な語句を抽出して」と入力し、改行してから対象の文章を入力します。

メール作成：例えばグローバルに販売する製品の取扱説明書を英語から多言語に翻訳する際に、日本語を使用しない翻訳者とのやりとりが必要になる場合があります。一から英文メールを書く時間がかかりますが、ChatGPT を使えば、日本語で要点だけを書き、それをもとに英文メールを書き起こすことができます。プロンプト例：「次の内容の英語のメールを書いて」と入力し、改行してから日本語で要点を入力します。

4. 翻訳者およびレビューアとしての活用法

機械翻訳：プロンプトに翻訳分野を指定し、既存の原文と訳文がある場合はそれを渡すことで、適切な用語や表現を使って翻訳できます。従来の機械翻訳エンジンでは訳文をカスタマイズするためにモデルの事前トレーニングが必要で、時間も費用もかかっていましたが、ChatGPT ではプロンプトを変えるだけで済みます。プロンプト例：「医療機器の取扱説明書用の文章を日本語に翻訳して」と入力し、改行してから翻訳対象の原文を入力します。

自動ポストエディット：ChatGPT や DeepL、Google

How to make use of ChatGPT for translation work

Takeyoshi Nakayama

Human Science Co., Ltd.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Public License. License details: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

などの機械翻訳の訳文を書き換えることができます。スタイルガイドに準拠するように表記を修正でき、文書の用途に応じて表現をブラッシュアップすることもできます。自動ポストエディットによって、人手のポストエディット作業を軽減できます。プロンプト例：「次のルールに従って文章を書き換えて。主語に You を使用しない。」と入力し、改行してから書き換え対象の英文を入力します。

校正：英文の誤りを検出して、単語の重複やスペルミスなどを修正できます。プロンプト例：「文法上の誤りを修正して」と入力し、改行してから校正対象の英文を入力します。

辞書：英和和英辞書、類語辞書、表現辞書として使用できます。英和和英辞書では、単語の意味だけでなく使用例を生成させることで一般的な辞書と比べてより理解が深まります。また、類語や別の表現を提案できるため、訳文をブラッシュアップしたいときにも役立ちます。プロンプト例：「＜英単語＞の類語と例文、日本語訳を教えてください」「＜日本語の語句＞の別の表現と例文を教えてください」

5. 機械翻訳ソフト MTrans for Office と MTrans for Trados が ChatGPT API と連携

ヒューマンサイエンスでは ChatGPT を利用できる機械翻訳ソフト MTrans for Office（エムトランス・フォー・オフィス）と MTrans for Trados（エムトランス・フォー・トラドス）を提供しています。

MTrans for Office は Windows/Mac/Web 版 Outlook、Word、Excel、PowerPoint 用のアドインです。インストールすると各 Office 製品からワンクリックで機械翻訳を実行できるようになります。

Windows 版では、複数の機械翻訳エンジンで同時に翻訳して、それらの訳文を比較できます。

ChatGPT のプロンプトが多数用意されており、独自のプロンプトを追加することもできるため、翻訳業務のさまざまな場面で ChatGPT を活用できます。複数

のプロンプトを同時に実行して、その出力を比較することもできます。Office ファイルの他に、スキャン PDF や画像の OCR 翻訳にも対応しています。

MTrans for Trados は Trados Studio 用のプラグインです。インストール後にプロジェクトの設定で「翻訳メモリと自動翻訳」に追加して使用します。機械翻訳と自動ポストエディットを、セグメント単位で実行することも、複数ファイルを対象に一括で実行することもできます。

両製品共通の機能として、機械翻訳エンジンでは ChatGPT、DeepL、Google、Microsoft の各 API と連携しています。すべての機械翻訳エンジンにおいてデータは二次利用されないため、機密情報の翻訳に利用できます。自動ポストエディット機能が搭載されており、文字列置換、正規表現置換、ChatGPT を使って訳文を自動修正できます。また、用語集機能を使って機械翻訳エンジンに対して製品名やサービス名などの名詞の訳語を指定できるため、手作業による用語の統一作業が不要になります。自動ポストエディットと用語集により、人手のポストエディット作業を大幅に軽減できます。（Mac/Web 版 MTrans for Office は DeepL と Google のみに対応しており、一部機能が利用できません。）

MTrans for Office について詳しくは、

https://www.science.co.jp/nmt/service/mtrans_office.html をご覧ください。

MTrans for Trados について詳しくは、

<https://www.science.co.jp/nmt/service/nmt.html> をご覧ください。

両製品ともに 14 日間の無料トライアルをご利用いただけます。

<https://www.science.co.jp/localization/contact/index.html> までお気軽にお問い合わせください。

編集後記

編集後記

出内 将夫

AAMT 編集委員会

本稿執筆時点で、ChatGPT や Bard 等の LLM（大規模言語モデル）は、その有用性やインパクトの大きさから、LLM 利活用を拡大する動きだけでなく、LLM 利活用におけるリスクも取りざたされています。LLM を構成する技術が MT で提唱され、MT 精度向上により MT 利用拡大が先行する形で MT ユーザーガイドがまとめられたように、LLM 利用拡大が先行する形で LLM 利用のルールが決められていくのでしょうか。

巻頭言では、田中英輝様から「MT Summit 2023 開催報告」をご寄稿いただきました。MT Summit 2023 の基調講演および招待講演では、LLM を用いた MT システムの評価、分析、課題などについて最新の研究が紹介されました。なぜ LLM に高精度な翻訳が可能なかの解明や、自動評価が難しくなっている点、低資源言語翻訳の性能改善などについて、さらなる研究が進められているとのことでした。

早川威士様と齋藤敦様からは、「医薬品開発業務を迅速化するための機械翻訳の協調的開発と運用」という題で、医薬品開発業務においてドメイン特化型 MT の開発と MTPE の導入の実証研究結果についてご寄稿いただきました。ドメイン特化型 MT の開発と MTPE の導入により合理的な品質を迅速に実現できた、との報告をいただきました。

Yuting Zhao 様からは、「Multimodal Neural Machine Translation based on Image-Text Semantic Correspondence」という題で、画像情報付きの原文を翻訳する際に、画像の情報を加味し、より正確に翻訳する MT 技術についてご寄稿いただきました。画像の特定領域を翻訳結果に反映する技術、画像の特定領域と原文テキストのアライメントを取る技術の提案により、既存手法を上回る翻訳精度を達成したことを報告

いただきました。

平岡達也様からは、「タスクに応じてトークン分割を最適化する試みと現状」という題で、従来前処理として行われていたトークン分割を、文書分類や MT のような後段タスクに応じて最適化する技術についてご寄稿いただきました。また、特に MT におけるトークン分割最適化について、ソース言語とターゲット言語の 2 段階に分けてトークン分割の最適化を行うことで精度が向上したことや、言語による分割長の特徴について興味深いご報告をいただきました。

影浦峯様からは、「人間の翻訳と機械の翻訳 (8) : 翻訳者がしていること・MT にできること」をご寄稿いただきました。連載の締めくくりとして、訳文の生成に関連した様々な作業が MT や LLM にできるかについて評価いただき、現在の MT や LLM が苦手とする領域について示唆をいただきました。

坂場大道様、木村修平様、近藤雪絵様からは、「立命館大学の正課英語授業へのニューラル機械翻訳導入について」をご寄稿いただきました。大学のプロジェクト型英語教育において MT を導入した結果、学生のプロジェクト発表における英語の質の品質向上や、論拠とする情報源の質の向上が見られた、との報告をいただきました。

隅田英一郎様からは、「ポストエディットの実実」をご寄稿いただきました。2022 年度に実施した大規模なポストエディット案件において、翻訳者に対して実施したアンケート結果の分析から、作業効率を上げるだけに留まらない有用性が示唆された、との興味深い報告をいただきました。

第 3 回目となる温故知新では、1992 年 3 月号と 1992 年 5 月号からの記事を紹介しております。日英翻訳に

Editor's note

Masao Ideuchi

AAMT Editorial Board

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International Public License.

License details: <https://creativecommons.org/licenses/by-sa/4.0/>

において、MT 以前に、入力となる日本語の品質改善が必要、という話は、今後 LLM を利用する上でも同様に重要となるように思えます。また、MT の誕生に関する記事では、MT をサグラダ・ファミリアになぞらえ「100 年 200 年かけて継承構築するもの」としておりました。奇しくもサグラダ・ファミリア同様に、MT も急速に完成に近づいているのかもしれません。

Sumire Honda 様からは、「MT Summit 2023」をご寄稿いただきました。MT Summit 2023 における、ご自身の発表のほか、オンライン参加のしやすさ、ネットワーキングの機会、産業分野での MT 研究を聴講することで得られた知見を紹介いただきました。

橋本簡美様からは、「AAMT Machine Translation Summit 2023 参加報告」をご寄稿いただきました。翻訳学習者が人手翻訳した場合とポストエディットした場合のエラーの比較から、翻訳学習者が MT の誤りを見抜けず、品質を過大評価してしまう可能性を指摘した論文を紹介いただきました。

平岡裕資様からは、「MT Summit 2023 参加報告書」をご寄稿いただきました。Invited panel や Keynote speech の内容を中心に、議題や発表の要旨を紹介いただきました。

Yuki Yamashita 様からは、「Participation Report on MT Summit 2023」をご寄稿いただきました。Keynote、Invited Talk のほか、Workshop (WAT2023) や Oral Presentation の発表について、要旨を紹介いただきました。

綱川隆司様、須藤克仁様、後藤功雄様からは、「第 10 回特許・技術文書翻訳ワークショップ (PSLT 2023) 開催報告」をご寄稿いただきました。MT Summit 2023 のワークショップの一つとして開催した PSLT 2023 について、招待講演の内容や、活発な質疑があったことを報告いただきました。

中澤敏明様からは、「第 10 回アジア翻訳ワークショップ (WAT2023) 開催報告」をご寄稿いただきました。入力文の言い換えと翻訳結果のリランキングによるドメインミスマッチの低減をテーマとした採択論

文や、Wikipedia における MT 活用をテーマとした招待講演の内容を紹介いただきました。また、設定した Shared Task と評価結果について報告いただきました。

法人会員の PR 記事として、株式会社ヒューマンサイエンス様から、「ChatGPT を翻訳業務に活かすには」という題で、寄稿をいただきました。引き続き法人会員の PR 記事を募集しておりますので、ご希望の会員様は、AAMT からの募集メールに記載されている連絡先までご連絡ください。

11 月 29 日には、年次大会の「AAMT 2023, Tokyo ～機械翻訳の今と未来を探る～」が、現地とオンラインのハイブリッドで開催されます。みなさまと活発な情報交換ができることを楽しみにしています。

AAMTジャーナル「機械翻訳」No. 79

- 【 発 行 日 】 2023年11月29日
【 発 行 】 アジア太平洋機械翻訳協会 (AAMT)
ホームページ: <https://aamt.info/>
【 住 所 】 〒619-0289
京都府相楽郡精華町光台3-5
国立研究開発法人情報通信研究機構 先進的翻訳技術研究室内
【 編 集 委 員 会 】 内山将夫 後藤功雄 中澤敏明 新田順也 園尾聡 森口功造 隅田英一郎
石川弘美 早川威士 出内将夫
【表紙デザイン】 泉谷東十郎
【 題 字 】 長尾真
【 事 務 局 】 石川弘美
【 印 刷 所 】 株式会社 プリントバック

Asia-Pacific Association for Machine Translation (AAMT)
c/o National Institute of Information and Communications Technology
3-5 Hikaridai Seika-cho Soraku-gun Kyoto 619-0289, Japan

