

ISSN 1883-1818

No.82

June 2025

AAMT Journal

Asia-Pacific Association for Machine Translation

機械翻訳

機械翻訳



目次

巻頭言		
AAMTセミナーの役割	中澤 敏明	3
事例・研究		
CATER: Enhancing Multidimensional MT Evaluation with LLM-Driven Contextual and Dynamic Equivalence	Kurando IIDA, Kenjiro Mimura	4
生成AIを活用した用語抽出	伊澤 力	9
個人翻訳者による、医薬分野AI翻訳サイトの開発	山本 隆之	12
解説記事		
機械翻訳を利用したツールの持つ可能性	ぬるっぼ	17
自治体による多言語情報発信を支える翻訳資源構築プロジェクト	宮田 玲、藤田 篤、 阪本 章子、香川 璃奈	20
AAMTセミナー		
AI事業者ガイドラインについて	藤本 篤志	29
温故知新		
温故知新5	内山 将夫	35
イベント報告		
第11回アジア翻訳ワークショップ (WAT2024) 開催報告	中澤 敏明	40
AAMT 2024, Tokyo～機械翻訳の進歩と調和～	出内 将夫	42
AMTA 2024 参加報告: 社会学の視点から見るAI翻訳	郭 文静	49
第8回自動翻訳シンポジウム 参加報告	玉咲 知香子	50
編集後記	後藤 功雄	52

C O N T E N T S

Foreword		
The Role of AAMT Seminar	Toshiaki Nakazawa	3
Case Study		
CATER: Enhancing Multidimensional MT Evaluation with LLM-Driven Contextual and Dynamic Equivalence	Kurando IIDA, Kenjiro Mimura	4
Term Extraction using Generative AI	Riki Izawa	9
Development of an AI Translation Site for the Medical and Pharmaceutical Fields by a Freelance Translator	Takayuki Yamamoto	12
Commentary		
Potential of tools using machine translation	Nullpo	17
Project for Building Translation Resources to Support Multilingual Information Dissemination by Local Governments	Rei Miyata, Atsushi Fujita, Akiko Sakamoto, Rina Kagawa	20
AAMT Seminar		
AI Guidelines for Business	Atsushi Fujimoto	29
Learning from the past		
Learning from the past 5	Masao Utiyama	35
Event Report		
Report of the 11th Workshop on Asian Translation (WAT2024)	Toshiaki Nakazawa	40
AAMT 2024, Tokyo --- Progress and Harmony of Machine Translation	Masao Ideuchi	42
AI Translation from a Sociological Perspective	Wenjing GUO	49
Report on the 8th Symposium on Automated Translation	Chikako Tamasaki	50
Editor's Note	Isao Goto	52

AAMT セミナーの役割

中澤 敏明

東京大学

私は AAMT の理事として、セミナー委員を拝命している。AAMT セミナーは機械翻訳の発展・啓蒙の一環として、MT に関するトピックの情報発信を目的としており、2022 年 9 月 28 日に第 1 回が開催された。AAMT セミナーは年に 5~6 回開催されており、2025 年 3 月時点で第 13 回まで開催された。第 14 回は 2025 年 5 月 20 日に開催予定であり、この記事が発行される頃にはすでに終了しているであろう。毎回多くの方に視聴していただき、セミナー委員としても嬉しい限りである。

AAMT セミナーは機械翻訳に関する最新技術や翻訳プロセスに機械翻訳をうまく導入するためのノウハウなど、有益な情報をさまざまな視点から、AAMT 会員の皆様をはじめとして多くの方にお届けすることを目的として運営している。その中で、若手の翻訳・通訳・機械翻訳の研究者を集めて最新の研究成果を発表する場としたり、機械翻訳とは直接は関係しないテーマを選んだり、少しチャレンジングな回も設けている。例えば第 10 回の「新規アルツハイマー病治療薬の進展を見据えた認知症診療の展望～音声解析を用いた早期診断～」や第 12 回の「AI 事業者ガイドラインのご紹介」などである。このようなテーマを選んでいる背景としては、機械翻訳の新たな利用シーンの創出や、これまで見えていなかった共通点の発見など、イノベーションが起こりえると考えてのことである。研究活動においてもそうなのだが、他の分野の研究からインスピレーションを受けたり、アイデアを流用したりしてうまくいくケースは少なくない。機械翻訳とは直接関係のないテーマの回であっても、ぜひ積極的にご参加いただければ幸いである。

さて、この AAMT セミナーは一部の回を除いて有料でご参加いただけるようになっている。当日ライブでの参加が難しい場合であっても、開催後 2 週間は録画の視聴が可能となっているので、後から内容を確認することもできるため、ぜひ積極的にご参加いただきたい。セミナーの参加費用は AAMT 会員でない場合、一般の方が 5000 円、学生が 2000 円となっているが、AAMT 会員だと一般の方が 3000 円、学生が 1000 円となる。AAMT 会員になるための費用は、入会金 1000 円と年間の会費が一般の方が 5000 円、学生が 1000 円なので、AAMT 会員になってセミナーに 2、3 回参加すると元が取れることになる。また AAMT 会員になるとセミナーだけでなく、AAMT の年次大会や機械翻訳の国際会議などにおいても割引価格で参加が可能になったり、他の会員との交流を通じて、より業界内の情報にアクセスしやすくなるなどのメリットもある。ぜひ AAMT へのご入会をご検討いただければ幸いである。

AAMT セミナーは今後も定期的で開催予定であるが、どのような内容のセミナーを開くべきか、どのようにセミナーを運営していくべきかなどについて、皆様からのご意見を頂戴したい。せっかくセミナーを開催するのだから、視聴者のニーズにあったものを提供することが大事であると思っているが、セミナー委員の中だけの意見ではどうしても視野が狭くなってしまふ。ご意見やご要望のある方は、ぜひ aamtseminar@aamt.info までご連絡いただきたい。

我々セミナー委員は今後も定期的の皆様にとって有益な情報提供を行なっていきますので、ぜひご注目いただき、セミナーへの参加や AAMT への入会をしていただければと思う。

The Role of AAMT Seminar

Toshiaki Nakazawa

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/>

CATER: Enhancing Multidimensional MT Evaluation with LLM-Driven Contextual and Dynamic Equivalence

Kurando IIDA, Kenjiro Mimura

ErudAite Inc.

Abstract

Machine translation (MT) evaluation metrics like BLEU (Papineni et al., 2002) prioritize formal equivalence, neglecting contextual and dynamic equivalence—key translation studies principles (Nida, 1964). COMET (Rei et al., 2020) saturates with LLMs (90+/100, WMT 2022; Kocmi et al., 2022), losing discriminative power, while MQM (Lommel et al., 2014) and DQF (TAUS, 2013) lack scalability due to manual processes. CATER (Comprehensive AI-assisted Translation Edit Ratio), refining its arXiv debut, proposes five LLM-driven axes—Grammatical Precision (GP), Semantic Integrity (SI), Factual Consistency (FC), Terminological Consistency (TC), and Contextual & Stylistic Appropriateness (CSA). Integrating MQM/DQF with TER's edit ratios (Snover et al., 2006), CATER ensures MECE evaluation without references, adoptable across MT and human translations. Its novel LLM use targets hallucinations (FC) and cultural fit (CSA), surpassing BLEU's limits and COMET's ceiling. As a proposal, CATER aims to enhance translation processes, with sample prompts on arXiv and future validation planned.

1. Introduction

Machine translation (MT) evaluation often emphasizes formal equivalence, a translation studies concept focusing on fidelity to the source text's grammatical structure, vocabulary, and literal content (Nida, 1964)—e.g., "He is running" as "彼は走っている." BLEU (Papineni et al., 2002) exemplifies this, using n-gram overlap with references, though

critiqued for its surface-level scope (Callison-Burch et al., 2006). This approach neglects contextual coherence and dynamic equivalence—the natural conveyance of intent and audience effect, such as "He's over the moon" as "彼は大喜びだ"—essential dimensions in translation studies.

Modern metrics like COMET (Rei et al., 2020) leverage neural embeddings to align with human judgments (Freitag et al., 2021), yet saturate with LLMs (e.g., 90+/100, WMT 2022; Kocmi et al., 2022), losing discriminative power. METEOR (Banerjee & Lavie, 2005) improves lexical matching but remains reference-dependent. MQM (Lommel et al., 2014) and DQF (TAUS, 2013) provide multidimensionality—MQM via error categories (Accuracy, Fluency), DQF via customizable criteria—but their manual nature limits scalability, especially for MT's hallucination-prone outputs (Guerreiro et al., 2023).

CATER (Comprehensive AI-assisted Translation Edit Ratio), refining its arXiv proposal, introduces five LLM-driven axes—Grammatical Precision (GP), Semantic Integrity (SI), Factual Consistency (FC), Terminological Consistency (TC), and Contextual & Stylistic Appropriateness (CSA). Integrating MQM/DQF frameworks with TER's edit distance (Snover et al., 2006), CATER's novel LLM use ensures reference-free, scalable evaluation, adoptable across MT and human translations. It aims to guide quality improvement beyond human benchmarks, applicable to any language, surpassing BLEU's constraints, COMET's ceiling, and MQM/DQF's manual limits.

2. The CATER Framework

CATER proposes five axes to evaluate translation quality, integrating and advancing prior frameworks with LLM automation.

Table 1: CATER Evaluation Axes

Axis	Definition	Example
GP	Grammatical Precision: Ensures linguistic rule adherence.	"He go" → "He goes"
SI	Semantic Integrity: Assesses meaning fidelity.	"I'm tired" → "私は怒っている"
FC	Factual Consistency: Detects omissions and hallucinations.	"He's a doctor, 40" → "He's a doctor"
TC	Terminological Consistency: Ensures term coherence.	"good" as "良い" then "素晴らしい"
CSA	Contextual & Stylistic Appropriateness: Evaluates context, style, cultural fit.	"Dear Sir" → "やあ、お前"

- **GP**: Ensures formal equivalence's baseline.
- **SI**: Captures meaning fidelity, reflecting METEOR's lexical focus (Banerjee & Lavie, 2005).
- **FC**: Targets MT-specific errors like hallucinations and omissions (Guerreiro et al., 2023).
- **TC**: Broadens consistency to include contextually significant general terms.
- **CSA**: Embeds cultural fit by default.

Table 2: Mapping CATER Axes to MQM and DQF

Axis	MQM Connection	DQF Connection
GP	Fluency (Grammar)	-
SI	Accuracy (Mistranslation)	-
FC	Accuracy (Omission/Addition)	-
TC	Terminology	Consistency
CSA	Style, Coherence	Locale

CATER integrates MQM's detailed error typology and DQF's domain-specific flexibility, advancing them with TER-inspired edit ratios (Snover et al., 2006) and LLM automation.

Table 3: Mapping CATER Axes to Translation Equivalence Types

CATER Axis	Equivalence Type
GP	Formal Equivalence
SI	Formal Equivalence
FC	Formal + Dynamic
TC	Dynamic Equivalence
CSA	Dynamic Equivalence

Caption: Table 3 illustrates CATER's alignment with formal and dynamic equivalence in translation studies (Nida, 1964). FC bridges both types, reflecting its dual role in factual and contextual fidelity.

Framework Evolution:

CATER's five axes clarify MT improvement directions—e.g., FC flags factual errors, CSA adjusts cultural nuance—while evaluating human translations for quality assurance (QA) and skill enhancement, surpassing human benchmarks across languages.

3. Methodology

CATER employs LLMs for reference-free evaluation, processing source and target texts with prompts.

- **Process:**
 - **Input:** Source text, target text, and axis-specific prompts.
 - **LLM Task:** Identify errors, compute Edit Ratio (ER%) = (Words to Edit / Original Word Count) × 100, inspired by TER (Snover et al., 2006). Scores (0-100) derived as: $\text{Score} = \max(0, 100 - \text{ER}\% \times \text{Weight})$, with adjustable weights (e.g., GP=1, SI=3, FC=5, TC=2, CSA=2, tunable by domain—higher FC for technical texts, higher CSA for literary works).
 - **Output:** Axis-specific scores.
- **Prompt Examples:**
 - GP: "Check grammar and spelling errors."
 - FC: "Identify added or omitted content vs. source."
 - CSA: "Assess context coherence and cultural tone."
- **Example:**
 - FC: "He's a doctor" → "He's a Japanese doctor" (ER% = 33.3%, $\text{Score} = \max(0, 100 - 33.3 \times 5) = 0$, Weight=5).
- **Weighting Rationale:**

Weights reflect each axis's practical impact, with a maximum of 5 for usability. FC's default weight of 5 underscores the critical nature of omissions and hallucinations—significant issues in modern MT and generative AI (Ji et al., 2023), where missing key details (e.g., "40" in "He's a doctor, 40") or adding unverified content (e.g., "He's a Japanese doctor") can lead to severe consequences in technical or official documents (Raunak et al., 2021). Existing frameworks like BLEU and COMET often fail to adequately penalize these errors, treating them as equivalent to lexical mismatches (Callison-Burch et al., 2006; Kocmi et al., 2022; Mathur et al., 2020), while MQM/DQF rely on manual weighting without consistent automation (Lommel et

al., 2014; TAUS, 2013). SI (3) prioritizes meaning fidelity, vital yet less critical than factual errors. TC and CSA (2) address coherence and context, significant in specific domains (e.g., literary translations). GP (1) treats grammatical errors as foundational but less severe. These defaults balance sensitivity to critical errors with practical scoring. Users may freely adjust weights based on domain needs (e.g., FC=5, CSA=2 for technical texts; CSA=5, FC=3 for literary works), provided they document the rationale (e.g., prioritizing factual accuracy in contracts) and publicly share the applied weights for transparency and replicability.

- **Advantages:**
 - Reference-free, unlike BLEU or METEOR (Banerjee & Lavie, 2005), leveraging LLM contextual understanding.
 - Easy adoption: No training or resources required, unlike MQM/DQF's manual setup.
 - Language-agnostic, scaling across diverse contexts (Freitag et al., 2021).

CATER's LLM application and TER integration enable broad, immediate usability for MT and human translation evaluation.

4. Comparison with Existing Methods

- **BLEU:**
 - **Approach:** N-gram precision emphasizes formal equivalence (Papineni et al., 2002).
 - **Limits:** Misses context and dynamic equivalence (Callison-Burch et al., 2006), requiring references.
 - **CATER:** CSA and reference-free design address these gaps.
- **COMET:**
 - **Approach:** Neural embeddings align with human judgments (Rei et al., 2020).
 - **Limits:** Saturates with LLMs (90+/100, WMT 2022; Kocmi et al., 2022), lacking

granularity and transparency, especially in low-resource settings (Freitag et al., 2021) and failing to adequately penalize hallucinations (Raunak et al., 2021).

o **CATER**: Retains discriminative power with axis-specific scores (e.g., FC, CSA).

- **MQM & DQF**:

o **Approach**: MQM categorizes errors across Accuracy and Style (Lommel et al., 2014); DQF offers customizable criteria (TAUS, 2013).

o **Limits**: Manual annotation hampers scalability; limited focus on MT-specific errors like hallucinations (Guerreiro et al., 2023).

o **CATER**: Integrates MQM’s typology (e.g., GP=Grammar) and DQF’s flexibility (e.g., CSA), extending them with LLM automation and TER precision.

CATER’s versatility—guiding MT improvement, evaluating human translations, and supporting QA—surpasses prior methods’ constraints.

5. Conclusion and Future Directions

CATER (Comprehensive AI-assisted Translation Edit Ratio) redefines translation evaluation by integrating formal equivalence (GP, SI) with contextual and dynamic equivalence (CSA, TC). Leveraging TER’s edit distance (Snover et al., 2006) and LLMs’ novel application, it subsumes MQM’s error categories (Lommel et al., 2014) and DQF’s flexibility (TAUS, 2013), advancing them with automation and MT-specific focus (e.g., FC for hallucinations, Guerreiro et al., 2023). Unlike BLEU’s reference dependency (Papineni et al., 2002) or COMET’s saturation (Kocmi et al., 2022), CATER’s five axes guide MT improvement, assess human translations for QA and skill enhancement, and enable quality beyond human benchmarks across languages. Its easy adoption—no training or references—ensures broad applicability. As a

proposal refining its arXiv debut, CATER aims to enhance translation processes, with sample prompts to be shared on arXiv.

Challenges:

- LLM judgment variability may affect consistency (Kocmi et al., 2023).
- Empirical validation is pending.

Future Work:

- Benchmarking against WMT datasets for human correlation (Barrault et al., 2020).
- Prompt optimization to reduce variability.
- Open access at [<https://erudaite.ai/cater>] with arXiv updates.

Acknowledgement

The concepts, design, and ideas of CATER are entirely the work of Kurando Iida and Kenjiro Mimura of ErudAite Inc., developed to enhance the ErudAite AI. In preparing this paper, we utilized Grok 3 beta, developed by X.AI Corp., for structuring the manuscript, drafting English text, and surveying prior research. This collaboration enhanced clarity and academic grounding, while the core CATER framework remains our original contribution.

References

- Banerjee, S., & Lavie, A. (2005). METEOR: An Automatic Metric for MT Evaluation. *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures*, Ann Arbor, MI, 65–72.
- Barrault, L., et al. (2020). Findings of the 2020 Conference on Machine Translation. *Proceedings of the Fifth Conference on Machine Translation*, Online, 1–55.
- Callison-Burch, C., et al. (2006). Re-evaluating the Role of BLEU in MT Research. *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Trento, Italy, 249–256.

- Freitag, M., et al. (2021). Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation. *Transactions of the Association for Computational Linguistics*, 9, 1460–1474.
- Guerreiro, N., et al. (2023). Hallucinations in Large Multilingual Translation Models. *arXiv preprint arXiv:2303.16104*.
- Ji, Z., et al. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12), 1–38.
- Kocmi, T., et al. (2022). Findings of the 2022 Conference on Machine Translation. *Proceedings of WMT*, Abu Dhabi, 1–50.
- Kocmi, T., et al. (2023). Large Language Models Are State-of-the-Art Evaluators of Translation Quality. *arXiv preprint arXiv:2302.14520*.
- Lommel, A., et al. (2014). Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Tradumàtica: tecnologies de la traducció*, (12), 455–463.
- Mathur, N., et al. (2020). Tangled Up in BLEU: Reevaluating Metrics with Error Analysis. *Proceedings of ACL*, Online, 4984–4997
- Nida, E. (1964). *Toward a Science of Translating*. Leiden: Brill.
- Papineni, K., et al. (2002). BLEU: A Method for Automatic Evaluation of Machine Translation. *Proceedings of ACL*, Philadelphia, PA, 311–318.
- Raunak, V., et al. (2021). The Curious Case of Hallucinations in Neural Machine Translation. *Proceedings of NAACL*, Online, 1172–1183.
- Rei, R., et al. (2020). COMET: A Neural Framework for MT Evaluation. *Proceedings of EMNLP*, Online, 2685–2702.
- Snover, M., et al. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. *Proceedings of AMTA*, Cambridge, MA, 223–231.
- TAUS. (2013). *Dynamic Quality Framework (DQF): A Collaborative Approach to Translation Quality*. TAUS Reports, Amsterdam.
- Kurando, I., & Kenjiro, M. (2024). CATER: Leveraging LLM to Pioneer a Multidimensional, Reference-Independent Paradigm in Translation Quality Evaluation. *arXiv preprint arXiv:2412.11261*.

生成 AI を活用した用語抽出

伊澤 力

株式会社川村インターナショナル

1. はじめに

翻訳において、対訳用語集は人手翻訳における用語の統一を図るために不可欠な資料であるが、近年は機械翻訳においても用語を指定するために利用されており、その必要性は益々高まっていると言える。対訳用語集を構築するにあたっては大量のテキストコーパスから対象となる用語を抽出し、適切な訳語を適用する作業が必要だが、そのほとんどにおいて人間の目による確認や人手による作業が不可欠であるため、対訳用語集を気軽に作成できない状況となっている。

具体的には、従来の方法を用いた用語抽出では用語として定義する必要のない一般名詞や、名詞以外の言葉が多く抽出されてしまい、本当に必要な用語の選別には人間による判断が不可欠となってしまう。また、抽出された原文側の用語に対応した訳文側の用語を自動的に抽出することは技術的に困難であるため、適切な訳語の選択にも人間による判断が不可欠であった。

そのため、用語抽出を自動化する方法はこれまで複数提案されているものの、以下の課題によって未だその精度が実用レベルに達していない。

- 特定分野に限定した選択的な抽出
- 対訳での抽出

本研究では、化学分野の特許明細書の対訳コーパスをもとに、生成 AI を活用することで上記課題を解決することができるかを検証し、さらに実用度を高めるためにモデルの学習機能を用いた精度向上の検証も実施した。

2. 対訳コーパス

本研究では以下の対訳コーパスを用いた。

- 対象文書：化学分野の特許明細書
- 原語ペア：日本語、英語
- 文章数：10,000 文

文書の特徴として、化学分野の用語が多く含まれているものの特許明細書の性質上特許用語も頻出しており、従来の統計的手法による用語抽出では不要な用語候補（ノイズ）として抽出されることが予想された。

また、対訳コーパスは対応する日英の特許明細書を自動処理を用いて整合（アライン）した。そのため、日本語と英語が正しく対応していないペアが一定数存在する点も特徴的かつ実践的と言える。

3. 生成 AI

本研究では OpenAI の gpt-3.5-turbo を用いた。当時、GPT 4 が既にリリースされていたが、処理速度と経済性の面から gpt-3.5-turbo がより実用的であると判断した。

プロンプトは System Prompt で役割を定義し、User Prompt には化学用語を抽出する旨を指定し、それらをすべて英語で記述した。

4. 作業工程

本プロジェクトでは当初、特定分野の用語抽出と対訳抽出をひとつのタスクとして行う方法を模索したが、抽出精度が期待したほど高くなかったため、各タスクを独立して行う方法（図 1）を用いることとした。

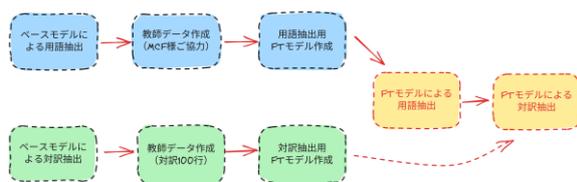


図 1. 用語抽出と対訳抽出を独立して実行するフロー

原文用語抽出

ベースモデルでの抽出

対訳コーパスのうち、原文となる日本語の文章を入力として化学用語を抽出する旨のプロンプトを用いたところ、選択的に化学用語を抽出することができたが、以下の問題が確認された。

- レスポンス形式の一貫性の欠如
- 用語候補の過抽出、漏れ

レスポンスの形式はプロンプトである程度制御することができるものの、用語とは関係のないモデルからの「返事」の内容が含まれていたり、複数の用語が抽出された場合の区切り文字が一定ではないといった結果が見られた。また、モデルが用語を抽出することを優先したためか、用語候補として分野外の用語を抽出するケースも多く見られ、結果としてノイズを含む多数の用語候補が抽出された。

ファインチューニングモデルでの抽出

上記問題を解消するため、ベースとなる gpt-3.5-turbo に対してファインチューニングを行った。教師データは、ベースモデルの出力からランダムに抜き出した 100 文を、専門知識を有する者の目で確認し、修正を行った。また、出力後のデータの加工を効率的に行うために複数候補が抽出された場合の区切り文字を統一した。

ファインチューニングを施したモデルを用いてベースモデルでの出力と比較したところ、ベースモデルの出力における 2 点の問題がいずれも大幅に改善された。

対訳抽出

ベースモデルでの抽出

対訳抽出においても、用語抽出と同様に gpt-3.5-turbo を使用した。上記ファインチューニングモデルによる抽出結果に加え、対応する英文を入力としてモデルに与えた。

結果として、ベースモデルであっても概ね期待通りの結果を得ることはできたが、訳文中に対応する用語が存在しない場合に、誤った訳語を出力する傾向が見られ、この点について改善の余地があることが判った。

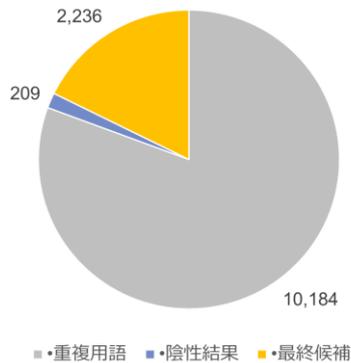
ファインチューニングモデルでの対訳抽出

用語抽出と同様に、ベースモデルの抽出結果からランダムに 100 文を抜き出し、期待する出力となるよう修正を行い、教師データとしてファインチューニングを実施した。このとき、ベースモデルにおける課題を解決するため、訳文中に対応する用語が存在しない場合に期待する出力結果を教師データに含め、プロンプトにも追加した。

5. 結果

上記工程にて作業を実施した結果、12,629 件の用語候補が対訳で抽出された。このうち、図 2 に示すとおり重複した用語 10,184 件および陰性結果 209 件を除いた一意の用語は合計 2,236 件となった。陰性結果は原文と対になる訳文側に、該当する用語が存在しないものを指す。陰性結果の発生原因は、翻訳の結果として訳文に該当する用語が訳出されなかった場合もあるが、主な原因は対訳コーパス作成における自動処理において、誤った訳文が紐づけられたことによるものであった。

図2. 対訳用語抽出結果



6. 考察

本研究の結果から、生成 AI を用いることで従来は難しかった特定分野の用語を大量のテキストコーパスから選択的に抽出することが可能であることが確認できた。ただし、学習を行わなかったベースモデルでは出力の安定性や精度の面で実用レベルに達していなかったため、本研究で行ったファインチューニングを用いた出力制度の向上は必須であると言える。

また、複雑なタスクを処理することが可能な生成 AI においても、可能な場合はタスクをより単純なタスクに分割して実行することが有効であることが確認できた点も本研究の成果と言えるであろう。

もちろん、最終候補として抽出された 2,236 件の用語についても最終的な有用性については専門知識を持った人間、つまり実務において最終利用者となる者による確認を経ることが理想ではあるが、少なくとも大量の文書を前にして用語集を作成必要のある者にとって、生成 AI を用いた抽出は救いの手となることは間違いないであろう。

本研究は、あくまで生成 AI の活用によって従来の用語抽出の課題を解決できるかという点について、実務での活用を前提として検証を行った。そのため、従来の方法による作業との詳細な比較および品質面での評価については今後の課題とする。

7. 今後の展望

本研究では前述の理由から gpt-3.5-turbo を活用したが、本稿執筆時点で OpenAI 社から GPT-4o などのより新しいモデルがリリースされており、今後はこれらの新しいモデルにおいてより精度が向上するかを検証したい。

また、本研究ではファインチューニングにより精度が向上することを確認したが、ファインチューニングを複数回実施した場合に、精度がどのように変化するかといった継続的な改善についても、今後の検証を続けていきたい。

8. 謝辞

本研究を実施するにあたり、言語商会代表の山本和英先生から研究の進め方等について多大なるご支援をいただきました。また、元インターンの高木創志氏には生成 AI を活用した作業に協力いただきました。

個人翻訳者による、医薬分野 AI 翻訳サイトの開発

山本隆之

合同会社クリニカルランゲージ

1. はじめに

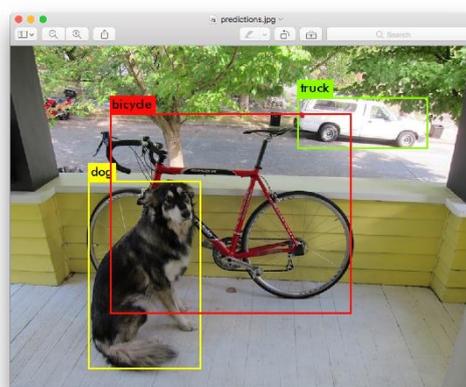
現在医薬翻訳では、特に製薬会社向けの翻訳において MT+ポストエディットの使用が広がっている。筆者も現在は通常の人手による翻訳とポストエディットの仕事をを行っている。多くの個人翻訳者は、今後 AI によって人間による翻訳の仕事が無くなることを心配しており、著者も同様に考えていた。一方で、ポストエディターとして仕事をするうちに、最近のディープラーニングを基礎技術とするニューラル MT の進化を非常に感じるようになり、AI に興味を持った。新しいことにチャレンジするという意味で、AI 専門のプログラミングスクールを受講したところ、ある程度翻訳サイトのウェブアプリを自作できることが分かった。そのため、自分の翻訳専門分野である医薬翻訳で AI 翻訳サイトを開発することを考えた。

2. 方法

医薬分野 AI 翻訳サイトを開発するにあたり、プログラミングスクールの課題として取り組んだディープラーニングによる花の種類判別アプリ（花の萼（がく）と花弁の大きさから花の種類を判定する）と物体認識アプリ開発の書籍（Python Flask による Web アプリ開発入門 物体検知アプリ&機械学習 API の作り方[1]）を基に、Google AutoML Translation の API コードを組み込むことを考えた。Google AutoML Translation は Google 翻訳をベースに対訳で追加学習することで、ドメイン特化のモデルが作れる。今回はこれを使用することで医薬翻訳モデルの構築を目指した。



花の種類を判定する Web アプリ (出典: AI Academy Media サイト[2])



物体検知の見本 (出典: YOLO サイト[3])

その他、翻訳サイト作成に必要なこととして以下のことを並行して行った。

- ① 公表されている対訳を集める（インターネット上のデータを python のコードでスクレイピング、または PDF をエクセル上の対訳にする）
- ② サイト自体の開発（HTML、CSS、JAVA scrip などを学び、本と他社のサイトを参考に Web サイトを作成する）
- ③ 医薬翻訳・薬事関連用語集の作成

2.1 使用した技術スタック（主なもの）

今回の医薬翻訳サイト開発にあたり、主に以下の技術スタックを使用した。それぞれ役割を記載する。

- **Flask** : Python の Web アプリケーションフレームワークであり、ウェブサイト全体を構築する
- **Google AutoML Translation** : Google 翻訳をベースに、対訳を使用してトレーニングすることでドメイン特化の翻訳モデルを作成できる
- **SQLAlchemy** : Python のコードで SQL が操作できるライブラリ
- **Stripe** : API を使用して決済システムをウェブアプリに組み込む
- **Docker** : コンテナを使用してアプリケーションをデプロイできるソフトウェアプラットフォーム
- **Google Cloud Run** : Google によるサーバー環境にコンテナをデプロイすることでウェブアプリを公開する
- **Cloud SQL** : Google によるクラウド型のデータベースサービス
- **ChatGPT API** : ChatGPT を自作ウェブサイト上で使用するための API
- **PubMed API** : PubMed に掲載される医学論文を検索・情報入手できる API
- **ClinicalTrials.gov API** : 米国臨床試験サイト ClinicalTrials.gov に掲載される臨床試験の情報を検索・入手できる API

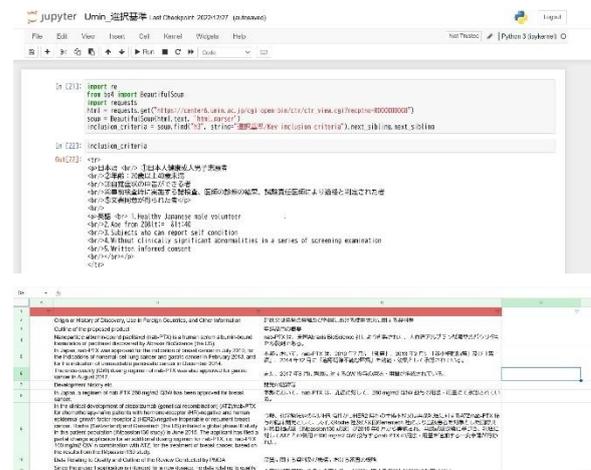
2.2 開発過程

医学や治験翻訳の対訳は、公表されている医薬品開発ガイドラインなどを集め Google AutoML でトレーニングを行った。モデルをトレーニングするための対訳を集めることに時間がかかったため、並行して通常の Google 翻訳 API を使ってウェブアプリの体裁を作った。通常、ウェブアプリを制作する際は、ユーザーが入力する画面を作るフロントエンドと、実際にアプリを動かすサーバー側の機能（翻訳、データベースへの保存、デプロイするインフラの構築など）であるバックエンドに分かれていることを開発を始めた後に知っ

たため、フロントエンド側の HTML、CSS、JavaScript なども、今回ゼロから勉強した。



開発中の画面



インターネット上のデータから対訳を作成

2.3 ChatGPT の登場

AI 翻訳サイトの開発を始めたのは 2022 年 3 月頃であり、当初は書籍やインターネット上の情報を学び、コツコツ制作を進めていた。しかし、2022 年 11 月に大規模言語モデル (LLM) である ChatGPT が登場したことにより、プログラミングの速度が上がった。プログラミングをする際に、本やウェブサイトの情報で理解できない部分や、自作サイトへコードを適用するうえでの不明点を ChatGPT に質問すると、理解できるレベルの内容で返答が得られた。特にデバッグが有用であった。一人でコーディングする中でエラーが繰り返し出現し、その特定と解決に一日中インターネット上を調べていることもあったが、ChatGPT にコードを入力することでこのようなエラーの特定と解決が一瞬でできることが頻繁にあった。

また、用語集作成など簡単なアプリであれば、ChatGPTが瞬時に作成できるため、ChatGPTが使用できるようになってからはこれらを自分のアプリに組み込む作業を行った。

2.4 初公開

フロントエンド、バックエンド両方の勉強と開発を少しずつ進め、集めた対訳を Google AutoML Translation でトレーニングし、医薬分野専用翻訳モデルを作成した。そして、2023年10月頃にインターネット上にデプロイ、初公開した。しかし、公開したものの、この時の翻訳品質は満足いくものではなかった。通常の Google 翻訳はすでに人体の解剖用語などを正しく出力し、流暢に翻訳できるが、AutoMLで対訳をトレーニングすることでかえって文が途切れたりするようになった。さらに、対訳を増やすと BLEU スコアが低下した。この対訳を増やすと BLEU スコアが低下する理由は確認できなかったが、集めたインターネット上の対訳に偏りがあったり（臨床試験のタイトルや組み入れ基準のデータが多い）原文に対し訳文は省略などがあり逐語訳でなかったことが原因ではないかと考える。インターネット上の対訳を集め、誤記のチェックなどの前処理作業は非常に時間がかかった。そのため作成した翻訳モデルの訳文の品質には納得がいかなかったが、とりあえずできる範囲でデプロイすることを目標としていたため、そのまま公開した。



初公開時の画面（2023年10月）

原文	訳文	Bleu	ベース Bleu	Bleuの増加	ベースモデル	対訳数
日本語	英語	37.748	22.134	15.614	Google NMT	14,506
英語	日本語	45.817	25.697	20.119	Google NMT	14,736
日本語	英語	29.797	22.163	7.634	Google NMT	111,654

BLUE スコア

2.5 その後の ChatGPT の進化

2023年10月の初公開時点では、Google AutoML の翻訳モデルに複雑な原文を入力すると訳文が途切れ最後まで訳出することができないことが多かった。そのため、ChatGPT を使用し訳文を修正する校正ボタンを付けていた。しかし、当時 GPT4 の価格は高額であったため GPT3.5 の API を使用していたが、GPT3.5 はである調「常体」で和訳することができず、全てですます調の「敬体」でしか訳文を出力することができなかった。

その後、2024年7月に GPT-4o mini がリリースされ、API 使用料金が大幅に低下した。そして GPT-4o mini を使用することで常体・敬体の修正が正しく反映できるようになった。

3. 開発結果（現在の画面と機能）

ChatGPT を使用することで、PubMed や ClinicalTrials.gov の API コードも自作サイトにある程度簡単に組み込めるようになった。そのため、薬剤や疾患名を用いて、医学論文の抄録や臨床試験の概要などの情報を検索し、そのまま翻訳することまでできるようになった。これらの機能はサイト開発当初は想定していなかったため、生成 AI の出現によって翻訳サイトの機能を増やすことができた。現在のサイトは、画面やインターフェースも一部変更し以下 3.3 項に記載する内容になっている。もともとの翻訳機能以外は、ほぼベースは ChatGPT が出力したプログラムで、サイトにつなぐための調整を筆者が行った。また、翻訳機能も現在は通常の Google 翻訳に用語集を使用し、ChatGPT で修正する方法に変更している。

3.1 ニューラル MT と LLM の役割分担 (ニューラルMTによる下訳+LLMによる ポストエディット)

当初は、Google AutoML Translation を使用した医薬分野専用の翻訳モデルを作成する予定であった。しかし、作成してみるとニューラル MT のみでは思ったほどの精度にならないことが判明した。そのため、ChatGPT を Fine-tuning することで医薬専用モデルを作成することを考えた。しかし、Fine-tuning もある程度大量の対訳が必要であり、また翻訳する際もコストがかかる。さらに、薬剤名や疾患名など大量の用語を含めた対訳で Fine-tuning を行うには膨大な労力がかかることが想定された。個人開発であることからコスト・労力を減らすため、以下の方法を取ることにした。①まず、通常の Google 翻訳 API に、薬剤名や治験用語の用語集をあてる。②そして翻訳の出力結果を ChatGPT を使用して修正する。この方法は、ニューラル MT による下訳作成+LLM によるポストエディットといえる。これにより、膨大な用語は用語集で対応し、ニューラルMTでどこもいかない訳文が出力された場合は LLM でスムーズな訳文にすることができた。この方法を取り入れることで、専門用語をそのまま使用し固い文体で翻訳することと、一般人向けに分かりやすい表現・文体（プレインランゲージ）の2通りで翻訳できるようになった。また、LLM による訳文修正はプロンプトで指示しているため、プロンプトをより詳細に記載することで、さらに細かいスタイルの指定ができるようになると想定される。

3.2 セキュリティ面の検討

今回は医薬分野の AI 翻訳サイトを開発しているため、サイト公開後は翻訳機能面を拡充させる一方、可能な限りセキュリティ上の対応も行った。個人開発のウェブアプリであるため資金・人的資源に限界はあるが、できる範囲の処置を行った。具体的な検討は、情報処理推進機構 (IPA) の発行する「安全なウェブサイトの

作り方」[4]をもとに行った。主な取り組みは以下の内容である。

- ① セキュリティポリシーの作成
- ② 通信の暗号化
- ③ SQL インジェクション、クロスサイト・スクリプティング (XSS)、クロスサイト・リクエスト・フォージェリ (CSRF) などアプリ側の対策、など

3.3 現在の機能 (2025年3月時点)

上記の ChatGPT を使用したサイトの機能改善により、現在の本サイトの機能は以下のようになっている。

URL : <https://clinical-language.com/>

1. AI 翻訳 (和訳・英訳)
2. プレインランゲージで翻訳
3. PubMed の API で医学論文を検索・翻訳
4. ClinicalTrials.gov の API で臨床試験を検索・翻訳
5. 医学・薬事分野用語の検索
6. 英語ページ
7. API でも使用可能



現在の翻訳サイト画面

4. 考察・結論

2021年10月よりプログラミングの経験ゼロから勉強を開始し、2022年3月頃から医薬分野 AI 翻訳サイトの開発を行った。約3年かかったが、途中2022年11

月の ChatGPT の出現などの生成 AI ブームもあり、当初想定していたよりも多機能なサイトが作成できた。ChatGPT を使用することで、翻訳もプログラミングも非常に精度・速度が上がった。一方で、AI に関わるプロジェクトで一般的に言えることのように、大量の翻訳データの前処理など、単純だが労力の必要なデータ作成に関わる作業はマンパワーを必要とし、個人開発では難しい点であると感じた。

誰でも AI 翻訳が作れる時代に

今回、フリーランスの医薬翻訳者である筆者が AI 翻訳サイトを構築したが、このサイトに近いものとして、生成 AI を使用して PubMed 論文を動画にまとめるサイトが個人開発で制作・公開されている[5]。今後は、翻訳もテキストからテキストだけではなく、テキスト→音声翻訳、テキスト→翻訳動画など、モダリティが変わってくるものが出てくる（またはすでに出てきている）。生成 AI で翻訳・プログラミングの難易度が下がったため、翻訳者・翻訳会社（機械翻訳会社も含め）以外でも翻訳エンジンを作ることができる状況になっている。機械翻訳エンジンは、今後 ChatGPT の fine-tuning やプロンプトを組み合わせれば誰でも作成できる。一方で DeepL も医療分野の専用エンジンを出しており、GAFAM 以外の通常の企業では、エンジンに差を付けることはできないのではないと思われる。また、現在の AI の特筆すべき点として、それまでできなかったことが AI のアップデートによって徐々に可能になっていくことが挙げられる。今できないことがあっても、将来的に自然に解決してしまう可能性があるため、現在品質や効率性に不十分な点があっても、あきらめずに少しずつ取り組みを継続していくことが大切であるとする。

参考文献

[1] 佐藤昌基（著）、平田哲也（著）、寺田学（監修）. Python Flask による Web アプリ開発入門 物体検知アプリ&機械学習 API の作り方

[2] AI Academy Media :

<https://aiacademy.jp/media/?p=382>

[3] You only look once (YOLO) :

<https://pireddie.com/darknet/yolo/>

[4] 独立行政法人情報処理推進機構「安全なウェブサイト の作り方」:

<https://www.ipa.go.jp/security/vuln/websecurity/about.html>

[5] 医学論文-AI 要約動画オムニサイト MVIDEra :

<https://v.pubmed-ai.com/>

機械翻訳を利用したツールの持つ可能性

ぬるっぽ

フリー

1. はじめに

昨今、機械翻訳は様々なアプリやデバイスで供され、言語の壁はますます低くなっている。しかし、ある特定の分野においては依然として機械翻訳の恩恵を受けづらいものがある。それはゲームだ。

当記事では、拙作の翻訳支援ツール PCOTⁱを利用してゲームを機械翻訳しながら遊ぶというアプローチから、ゲームのみならず、その他様々な用途での使用を通じて、機械翻訳を利用するツールの持つ可能性について言及したい。

2. 筆者について

2008年にIT系企業に入社。2015年にフリーランスとして独立。主に業務システムの開発に従事する。2020年5月に翻訳支援ツール PCOT を開発し、Web サイトにて公開中。ⁱⁱ

PCOTを開発したのはローカライズされていない海外製のゲームをどうにかして遊びたいと思ったのがきっかけである。筆者は外国語に関して会話はおろか、読み書きもままならない。しかし、この事情こそが PCOT を生むきっかけとなった。

理由として、ある程度外国語に精通している場合は、そもそも翻訳ツールを開発するという動機が希薄になりがちであり、また、仮に開発したとしても補助的なツールの枠に収まってしまう可能性が高かったであろう事は想像に難くない。

3. ゲームのローカライズ事情

ゲームのローカライズは昔から取り組まれてきた分野だが、昨今では機械翻訳の精度の向上によって職業翻訳者ではない、有志によるローカライズも盛んに行われるようになった。しかし、ゲームのローカライズは、様々な課題を抱えている。

- パッケージングされたデータから原文テキストを抽出できない
- ゲームの作りとしてマルチバイト文字の表示に対応していない
- 著作者からローカライズの許可を得られない
- テキスト量が膨大過ぎて個人の手には負えない場合がある

上記はほんの一例である。これら様々な要因によってローカライズされる見込みのないゲームを遊びたい場合、採用される手段は限られる。

筆者は限られた手段の中から、ツールを用いてゲームを翻訳しながら遊ぶというアプローチを採用した。

4. PCOT の概要

PCOT は画面に表示された文章を画像データとして OCR (光学文字認識) によってテキストを抽出し、得られたテキストを機械翻訳にかけ、訳文を表示するツールである。インターネット上には、類似したツールが他にもいくつか存在するが、PCOT が他の類似ツールと異なるのは、ゲームに特化しているところにある。

以下に PCOT がゲームに特化している特徴を挙げる。

- 翻訳範囲を固定する「固定翻訳」
- 翻訳範囲を任意に矩形選択できる「フリー選択」
- ホットキーによる操作
- OCR による誤読を翻訳前に置換可能
- 機械翻訳された訳文を置換可能
- 画像を加工して OCR の読取精度を向上させる
- 対象画面を画像ファイルとして保存し、専用画像ビューアから任意のタイミングで翻訳可能

これらの特徴は、可能な限りゲームを快適に遊ぶために工夫した結果である。

ゲームを翻訳しながら遊ぶという観点に立った時、ツールに要求される仕様としては、OCR の読取精度、機械翻訳の翻訳精度、ツールへのアクセシビリティ等が高い水準で求められる。

これらの水準が満たされない場合、最終的に翻訳はできても、翻訳するためにゲームを遊ぶという事態に陥ってしまい、ゲーム体験が著しく損なわれてしまう。

5. ゲーム以外での利用

本来はゲームを翻訳しながら遊ぶ用途で開発した PCOT だが、OCR 技術を利用していることから、理論的には画面に表示されている全ての文章を翻訳可能である。こうした特性により、ゲーム以外での利用を積極的に行っている利用者も存在する。具体的には、電子書籍の洋書や論文、海外製ソフトのインターフェース、仕事に至るまで、様々な用途で利用されている。

特に洋書では、翻訳範囲を固定できる固定翻訳や、左右のページで文章が繋がっている場合に文章を自動的に繋げて翻訳する連続翻訳など、多くの点でゲームから転用可能な機能が存在するため、使い勝手が良いものと思われる。また、訳文をコピーして別のアウトライントール等で、文章を編纂して残す用途でも利用されているようだ。

こうした本来の目的外で利用されている現状を鑑みると、機械翻訳の性能を余すことなく発揮するためには、ツールの完成度は無視できない要素である。

一例として、PDF ファイルの翻訳がある。PDF ファイルはその仕様上、文章が途中で表示のレイアウト通り物理的に改行されてしまうため、コピー&ペーストで原文を翻訳すると、著しく翻訳精度が低下してしまう。かといって、改行を全部繋げて翻訳すると、違った意味で翻訳されてしまう上に可読性を損なってしまう。こうした課題に対して、ツール側である程度文章のブロックや段落を再現し、翻訳精度や可読性を向上させる試みは非常に効果的である。

PCOT には改行に関する様々な機能が搭載されている。

- OCR 読取した文章の座標とサイズから元の文章ブロックと段落を再現（基本動作）
- 箇条書きの文章の可読性を向上させる「対象通りに改行」
- 意図しない位置で改行されてしまうのを防止する「改行を無視」
- 行末にある特定の文字を検知して改行させる「改行詳細設定」
- 原文置換用の機能だが、正規表現により特定の文字の後に改行を付与できる

こうした改行に関する様々な機能は、公開当初の PCOT には搭載されていなかった。しかし、利用者が増えるにつれ、筆者自身の気づきや利用者からのフィードバック等により、多数の機能が追加された。利用者のフィードバックは、専門的な知識に囚われがちな開発者に柔軟な発想を促し、ツールのブラッシュアップに大きく貢献している。

次の項目では、全てのケースで採用できる手段ではないと前置きをした上で、筆者なりに考慮した利用者を増やすための工夫について記す。

6. 利用者を増やす工夫

ツールの利用人口を増やす意図として、利用が増えたとそれだけフィードバックも多くなり、ツールの完成度を高められることが挙げられる。また、利用を通じて多数の翻訳データを収集できるため、機械翻訳そのものの性能向上にも寄与するものと思われる。

以下に、利用者を増やす工夫をいくつか挙げる。繰り返すが、筆者の経験に基づく知見であり、法人向けでなく個人の利用者を想定したものであり、全てそのまま利用できる手段ではないことを予め留意して頂きたい。

● 可能な限り導入の手間を省く

理想は圧縮ファイル展開後に実行ファイルを実行するだけで、基本的な機能を全て利用できるのが望ましい。

導入に手間がかかると、それだけで間口が狭まってしまい、初見の利用者を獲得しづらい。また、利用登録が必須のサービスはオプション機能にすると、利用されやすい傾向にある。

● フィードバックと真摯に向き合う

これは当然のことではあるものの、とても難しく、そしてとても大事だと思っている。

一見無理な要望でも、代替案を示す、或いは根本の仕様を見直すなど、様々な手段を講じる余地がある。どうあっても無理な場合は、どうして無理なのか、何故難しいのかを誠意を尽くして説明し、利用者の理解を得る必要がある。

こうした積み重ねが、利用者との信頼関係を構築し、口コミや紹介記事を通じて利用者を増やす助けになっていると筆者は信じている。

● 宣伝のタイミング

PCOT はゲームを翻訳しながら遊ぶという性質上、利用が増える時期に波がある。利用が増えるタイミングとして、ゲームプラットフォーム

の大型セール時、そして期待の大型タイトルの発売時などが顕著である。このようなタイミングに SNS 等で宣伝すると大変効果的である。

余談だが、筆者は PCOT を宣伝するために広告を出したことは一度もない。また、広告会社やインフルエンサー等に金銭を支払い、宣伝を依頼したこともない。

それは PCOT が実験的なソフトであり、商業的な成功を目的としていなかったためである。利用者の生の声を聞き、改善を繰り返し、どれだけの利用者を満足させることが出来るか、そして、そのアプローチにより口コミや利用者の自発的な紹介記事等で、どれだけ利用者を増やすことが出来るかを挑戦し続けてきた。個人的には、このアプローチは大成功だったと大変満足している。

ただし、これは趣味の範疇で活動してきた経緯からであり、製品として販売する場合には、広告会社やインフルエンサー等はむしろ積極的に活用すべきだということを示し添えておく。

7. まとめ

PCOT は公開してから執筆時点で 4 年が経過した。本誌が刊行される頃には 5 周年を迎えていることだろう。PCOT はこれまでに様々な機能を追加し、沢山の利用者にも恵まれた。PCOT を利用してテキストベースの長編ゲームを快適に遊ぶことが出来たという報告も沢山頂いている。

上記の結果は機械翻訳のみならず、それを利用するツールの持つ可能性を最大限発揮した証左と言えるのではないだろうか。

今後、機械翻訳とそれを利用するツールの双方が発展し、ゲームを含むあらゆる分野の言語の壁が取り払われると信じている。そして、多くの利用者の中で拙作の PCOT がその一助となっていれば幸いである。

ⁱ PCOT は Process Connect OCR Translator の略

ⁱⁱ https://www.gc-net.jp/s_54/

自治体による多言語情報発信を支える翻訳資源構築プロジェクト

宮田玲、藤田篤、阪本章子、香川璃奈

東京大学、情報通信研究機構、関西大学、産業技術総合研究所

1. はじめに

本稿では筆者らが共同で進める「円滑な多言語情報発信を可能にする自治体横断型翻訳資源の構築」プロジェクトについて紹介します。このプロジェクトは、2023～2026年度の4年間にわたり、日本学術振興会・科学研究費助成事業（課題番号：23K28378）の支援を受けて推進しているものです^[1]。簡潔に、自治体翻訳資源（Translation Resources for Local Governments; TR4LG）プロジェクトと覚えていただくと幸いです。

本プロジェクトは、日本の自治体（市区町村・都道府県）による住民向け多言語情報発信を支援するための「翻訳資源」とそれを活用した機械翻訳（machine translation; MT）等のテクノロジーの構築と公開を目的としています。そして、長期的には、自治体において日本語を第一言語としない住民が直面する言語バリアの解消を目指しています。

本プロジェクトが本格的に始まってから2年が経過しました。まだまだ道半ばではあるものの、一定の成果と課題も見えてきました。また、近年は大規模言語モデル（large language model; LLM）の開発と普及が急速に進み、LLMのMTとしての利用も進んでいます。このような著しい技術環境の進展をふまえてプロジェクトのあり方を改めて言語化することが重要だと考えています。

本稿では、自治体の翻訳における課題を概観した上で、我々のプロジェクトの理念と枠組みを説明し、翻訳資源が近年のMT関連技術にどう使えるかを整理します。また、これまでの研究状況の一端と今後の展望

についても簡単にご紹介します。なお、本稿の内容は、学会等で発表した内容^[2]を整理し、広げたものです。

2. 自治体における翻訳の課題

日本における在留外国人数は2024年6月末時点で350万人を超え、過去最高を更新しています^[3]。国籍・地域別の統計から、例えば、ベトナムやインドネシアの国籍の人口が過去10年間で大幅に増えるなど、外国人住民の人口構成の変化も見取れます。このような状況の変化に対応しつつ、外国人住民の地域での暮らしを支える行政情報を多言語で提供することは、自治体にとって重要な課題であり続けています。

各自治体では、多言語サービスの提供に努めているものの、予算的・人的な制約があり、増大・多様化する翻訳ニーズに必ずしも十分対応できていない状況があります。そのような状況下で、住民への多言語情報発信のために、自治体ウェブサイトを中心にMTの導入が進んできました。2019年時点で、1,788ある自治体の内、約73%においてMTが導入されていることが報告されています^[4]。MTを使えば、原理的にはサイト上の全てのテキストは、MTサービスで提供される言語に翻訳できます。しかし、多くの自治体では、MT訳を修正し品質を担保するフローは必ずしも組み込まれていない状況です。

ウェブサイト以外の行政文書の翻訳は、自治体内部で行う場合も外部に委託する場合も、多くは人手でなされており、より多くの文書の翻訳は難しいようです。目標言語によっては、納品結果を自治体内で検査する人的・制度的な体制が整っておらず、納品物をそのま

Project for Building Translation Resources to Support Multilingual Information Dissemination by Local Governments

Rei Miyata, Atsushi Fujita, Akiko Sakamoto, Rina Kagawa

University of Tokyo, National Institute of Information and Communications Technology, Kansai University, National Institute of Advanced Industrial Science and 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/>

ま掲載する他ないという状況です。近年、MT と接続した翻訳支援ツールの導入も進んでいますが、やはり品質管理を含めたベストプラクティスが確立されていないのではないかと思います^[5]。

以上、述べてきた自治体における翻訳の課題は、大きく以下の2点にまとめることができます。

- 多くの文書が住民に必要な言語で翻訳できていない（**規模の問題**）
- 人手翻訳・MT を問わず、品質を担保する工程が十分組み込まれていない（**品質の問題**）

そして、これらの問題の根底には、現実的な制約の下で、どの文書をどの言語でどのような翻訳方法・品質基準によって翻訳すべきかに関する指針が定まっていないという状況があります^[6]。

3. 翻訳資源構築の理念と枠組み

以上の課題認識をふまえ我々が目指しているものは、現状の規模・品質の翻訳ありきで、それをテクノロジーで効率化（すなわちコスト削減）することではなく、できるだけ良い翻訳サービスを、これまでできなかった文書・言語に広げることです。

規模の問題への対処としてMTを利用することは自然ですが、同時に品質の問題にも対処する必要があるため、「MTをよりよく活用すること」が必要です。MTの研究開発が進み、高精度な汎用MTサービスも普及していますが、自治体特有の用語や言い回しへの対応はどうしても必要です。また、各種のMTサービスでは多くの言語への対応が進んでいますが、特に低資源言語における翻訳精度には改善の余地があります。

それでは、MTをよりよく活用するにはどうすればよいでしょうか。ここで、本プロジェクトで焦点を当てる翻訳資源が鍵になります。すなわち、MTを活用した翻訳プロセスに、翻訳資源を適切に組み込むことで、規模の大きさを保ったまま、翻訳の品質を高めることが期待できます。

3.1. 翻訳資源と（機械）翻訳

まず、自治体における文書の特徴として、以下の2点が挙げられます^[7]。

- 過去の文書と似た文書が定期的に作られる
- 異なる自治体で似た文書が作られる

当然これは、原文書だけでなく翻訳文書についても当てはまります。この素朴な観察から、自治体横断的に文書（より細かい単位では文や用語）をその翻訳とともに収集しておくことで、文書作成や翻訳に活用できると期待できます。

ここでMT技術の活用に向けたポイントは2つあります。1つ目は、既訳をそのままの形で参照し、再利用することです。現在のMTの多くは、どんな入力文に対しても、「一から翻訳すること」を前提とした仕組みです^[8]。しかし、実際の翻訳作業においては、「〇〇年〇〇月の翻訳では、〇〇のように訳しており、今回はその一部を修正したものなので、〇〇の部分だけ訳し直せば十分だろう」といった判断がなされることが少なくありません。このような作業は、技術的には翻訳メモリでカバーされてきました。MTをよりよく活用するには、逆説的に、MTを使わない場面（より正確には、一から翻訳しない場面）を適切に見極めることが重要になります。言い換えると、MTありきではなく、その他の翻訳手段も含めた、翻訳プロセス全体にMTを位置付けることがポイントです。

2つ目は、より直接的に、翻訳資源をMTのカスタマイズに利用することです。言語間で対応付けされた対訳文データや対訳用語集は、MT自体の様々なカスタマイズに活用可能です。具体的な方法については、4節で触れます。

3.2. 翻訳資源の種類と構築手順

さて、ここまでで自治体における翻訳の課題を解決するために、MTをよりよく活用すること、そして、その方法として翻訳資源が重要な役割を担うことを説

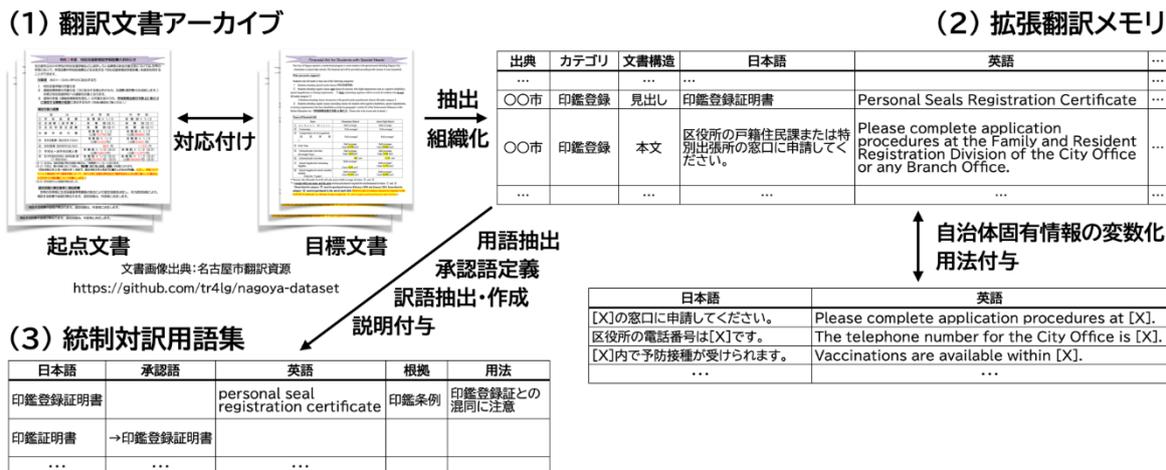


図 1. 翻訳資源構築の流れ

明してきました。ようやく、ここで、我々が扱う翻訳資源とはどのようなものであり、どのように作るのかを紹介したいと思います。

本プロジェクトでは、文書・文・用語のそれぞれの単位で、以下の3種類の翻訳資源を構築します(図1も参照)。

- **翻訳文書アーカイブ:**
適切なメタデータを付与した対訳文書の集合
- **拡張翻訳メモリ:**
文書との対応が保持され、内容・スタイル・用語が統制された対訳文の集合
- **統制対訳用語集:**
標準的な表記を定めた対訳用語集

図1に示すように、まずはこれまで生産された原文書と翻訳文書を対訳で集めることからスタートします(翻訳文書アーカイブ)。ある文書が過去に作られた文書の更新版であるならば、その翻訳も過去に翻訳された文書を適切に踏襲する必要があります。そして、そのような文書参照に基づく翻訳行為を可能とするには、対訳文書を分類し、適切なメタデータを付与して組織化しておくことが重要です。

続いて、文書から文を抽出し、対訳文の集合を作ります(拡張翻訳メモリ)。これは翻訳メモリと同様の考えに基づくもので、文単位で既存の翻訳を再利用することを可能にするものです。このとき、単に多言語で

対応付けされた文を集めるだけでなく、いくつかの拡張を加えます。まず、各文を元の文書と紐付け、抽出位置の情報とともに残します。また、既存の翻訳結果は必ずしも良い品質とは限らないため、品質評価情報を付与することも重要です。さらに、ある自治体のテキストは他の自治体のテキストとして必ずしもそのまま再利用はできないという観察から、自治体固有の要素を変数に置き換えてテンプレート化することも検討しています。このような拡張により、テキスト内外の文脈に応じた訳文の再利用が促進されると期待できます。

最後に、拡張翻訳メモリ内のテキストから、各言語の専門用語を抽出し、用語集として整備します(統制対訳用語集)。ここでも単に対訳用語の一覧のみを作るのではなく、用語の揺れの解消や適切な定義・用法の付与を行います。特に、自治体分野では標準的な用語集が整備されておらず、起点言語・目標言語のいずれにおいても、用語の揺れが広く観察されています。一定の基準のもとで承認語を定めることで、統制された妥当な用語を一貫して使用することが可能となります。

3.3. 学術的な特徴

ここまでの説明で、本プロジェクトで扱う翻訳資源は、単に文書・文・用語を幅広く集めて言語間で対応付けただけではないことをお分かりいただけたかと思

います。ここからは、もう一步踏み込んで、それが学術的にどのように位置付けられるのかについて簡単に触れておきたいと思います。

本プロジェクトにおいて我々は、大きく2種類の「翻訳に関する知識」を扱います。1つは、翻訳資源そのものです。適切に組織化されて公開された翻訳資源は、それ自体、翻訳時に参照される社会的知識としての役割を果たします。ここで重要な点は、資源が組織化されるときの基本認識単位が、一定の内容的なまとまりをもって社会において固有の役割を持つ「文書 (document)」^[9]であるということです。翻訳者は言語学的な単位ではなく、文書論的な単位を扱います^[10]。個々の語句を扱う場合でも、それらが、当該分野の文書・文書群においてどのような役割を果たしてきたかが常に考慮されます。例えば、「～します」という文末表現は、英語において、その文の機能（例えば、納税手続き文書における「義務」）や文書中の位置（例えば、手続きステップの「箇条書き項目」）に応じて、“You must...”や“You can...”などの構文になったり、命令文になったりと訳し分けられます。このような翻訳時の判断を支えるためには、資源の側も、文書論的な単位で組織化されている必要があります。

これに近いことは既に様々な場面で指摘されています。例えば、文単位の翻訳メモリでは、翻訳者は文脈情報が捨象された細切れの訳文を扱うこととなります。このような問題は、センテンスサラダ^[11]やコラージュ翻訳^[12]という表現で批判されてきましたが、翻訳メモリのメンテナンス自体が翻訳業務として必ずしも重要視されていない状況も相まって、結局解決されないままで見受けられます。MT研究の分野でも、長らく文単位の翻訳の問題点と文書単位の翻訳の必要性について議論され、研究も進められてきました^[13]。しかし、そもそも文書の属性が保持され、ある程度文書構造が復元可能な形で構築された研究データセットが不足していることもあり、文書レベルMTの研究は十分に進んでいるとは言いがたい状況です。本プロジェクトにおいて、自然言語処理 (NLP) 分野で普及

している「言語資源」という表現をあえて使わずに、「翻訳資源」という表現を使っているのは、言語学的な単位で解体された要素ではなく、翻訳論的な文書の単位を扱うことを表明するものです。

もう1つの翻訳に関する知識は、翻訳資源構築時の意思決定に関わるメタ知識です。本プロジェクトでは、現状の翻訳結果を出発点としつつも、それを無批判に受け入れるのではなく、「よりよい」品質を追求することも重要な課題です。例えば、用語レベルの翻訳資源を作る際に、ある日本語の専門用語に対して複数の翻訳事例が得られたとします。それらを全て列挙するだけでなく、その中でも妥当な訳（1つとは限らない）を選び、その判断理由を記録することまで行っています。朴ら^[14]による専門用語の日中翻訳の研究では、規範性、正確性、形式的妥当性、理解容易性、参照可能性、用語性、通用性、一貫性、識別性の規準を定義し、訳語候補を一つ一つ精査しています。これにより例えば、「再生手続」の中国語訳として、複数の候補を検討した上で最終的に、日本の慣習にある程度慣れている人にとっての参照可能性を重視した“再生手続”という訳語と、そうでない人にとっての理解可能性を考慮した“重整手続”という訳語の両方を妥当な訳語と判断しています。我々は、このように精査された用語を、規準に基づく判断の結果とともに用語集に掲載することを目指しています。

翻訳におけるこのような規範に関わる意思決定は、多くの場合わざわざ明示されませんし、そもそも全てを明示し尽くすことは困難です。しかしながら、翻訳資源の有効範囲と妥当性を第三者が検証可能にするには、どのような手順や基準でその訳を選んだ・作ったのかについて可能な限り言語化し、公開することが重要だと考えています。

筆者らの一部は、これまで翻訳プロセスについて説明するための「メタ言語」の構築に関する研究プロジェクト^[15]に関わっており、今回のプロジェクトはこれを学術的に継承・発展させるものでもあります。

4. 翻訳資源の機械翻訳への活用

3.1 節で触れたように、我々が構築している翻訳資源は MT のカスタマイズに使えます。すなわち、翻訳資源を使うことで、自治体分野の翻訳に特化した MT の開発が可能となります。ここで MT はニューラルネットワークを用いた手法であることを前提として、大まかに、エンコーダ・デコーダ型のアーキテクチャで対訳文を使って翻訳専用のモデルを訓練する従来型のニューラル MT (従来型 NMT) と、デコーダ部を多様な形式のテキストで事前訓練した LLM を翻訳器として利用する手法 (LLM ベース翻訳) に分けて説明します。

従来型 NMT では、各種の分野適応手法^[16]が使えます。多様な分野の大規模な対訳文で訓練したベースモデルを、対象分野の比較的小規模な対訳文で追加訓練 (微調整) する手法がよく使われます。また、用語レベルの翻訳資源は、語彙制約付きデコーディング^[17]、プレースホルダ翻訳^[18]、用語の事前置き換え^[19]などの枠組みで活用できます。また、翻訳メモリ中の類似する既訳を参考訳として使うことも考えられます。

LLM ベース翻訳でも、従来型 NMT と同様に、対訳文を教師あり微調整に用いることができます^[20]。また、LLM ベース翻訳はそもそもプロンプトそのものからの文脈内学習 (in-context learning; ICL) によって実現されるので、このプロンプトに翻訳資源の知識を直接入力することができます。例えば、原文に含まれる用語の対訳を検索し、訳語の指定を明示的にプロンプトに与えることで、用語の正確な訳出が期待できます。用語だけでなく、入力文に関連した既存の訳文や文のテンプレートを与え、訳文の構造や語彙をゆるやかに指定する方法も考えられます。これらはいわゆる、検索拡張生成 (retrieval-augmented generation; RAG) と呼ばれます。

しかしながら、従来型 NMT も LLM ベース翻訳もニューラルネットワークに基づく手法であり、翻訳に関する所定の仕様を完全に満たすような訳出は容易で

はありません。ルールベースの品質保証 (quality assurance; QA) の仕組みを併用することも必要になるでしょう^[21]。

5. これまでの状況

本プロジェクトのこれまでの取り組みの状況について、紙幅の都合上、名古屋市と連携して構築してきた名古屋市翻訳資源とそれを起点とした学術および社会における展開に絞って紹介します。

名古屋市翻訳資源とは、名古屋市が 2020 年に開始した行政文書翻訳向け MT 活用プロジェクトにおいて構築が進められてきた対訳文書集・対訳文集・対訳用語集のことです。自治体翻訳資源プロジェクトの開始以前から本稿第一著者が構築に協力してきました。名古屋市から提供を受けたこれらの資源を、クリエイティブコモンズライセンス CC-BY を付与し、2023 年に第一弾を GitHub で公開しました^[22]。その後も、我々のプロジェクトで管理が引き継がれ、翻訳資源の追加がなされています。対訳文書は日本語文書とその英語訳・中国語訳を中心に約 90 組公開されています。対訳文と対訳用語は、それぞれ約 800 文、約 3000 語の日本語に対して、英語、中国語、ベトナム語、フィリピン語、ネパール語、ポルトガル語、朝鮮語、スペイン語の 8 言語の翻訳が付与されています。

これらは基本的には生の対訳データであり、我々が目指す翻訳資源の前段階にありますが、CC-BY という使い勝手のよいライセンスで公開できたことで、学術・社会の両面において、様々な活用が広がっています。

5.1. 学術的展開

名古屋市翻訳資源を出発点として、様々な MT/NLP 研究用データセットを構築してきました。図 2 にこれらの派生関係を示します。いずれのデータセットも、文書レベル MT の発展に寄与すべく、文書のまとまりを保持している点が特徴です。そして、MT モデル自

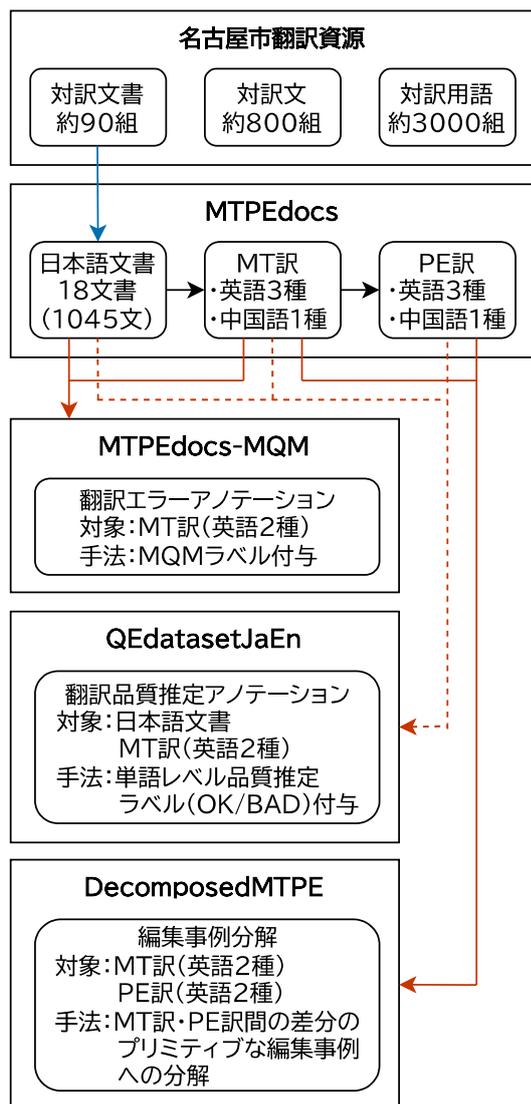


図 2. MT 関連研究用データセット

体の研究だけでなく、後編集 (post-edit; PE)、分析的評価、翻訳品質推定といった各種関連タスクで使えるように、データの加工やアノテーションを行っています。また、MT 関連の研究分野で不足している日本語を起点言語としたデータセットを補うものでもあります。

まず、MTPEdocs^[23]は、名古屋市翻訳資源から選定した日本語文書 18 文書を対象に、複数の MT 訳と人手 PE 訳を作成・付加したデータです。MT 訳と PE 訳の比較により MT 訳の改善点を可視化できますし、また MT 訳と PE 訳の対を学習することで、自動 PE の研究にも利用可能です。そして、この MTPEdocs を起点に、続く 3 つのデータセットを開発しました。

1 つ目は、MT 訳に対して人手で翻訳イシューラベルを付与したデータセット MTPEdocs-MQM^[24]です。翻訳イシュー体系の MQM (Multidimensional Quality Metrics) をベースに、日英翻訳の特徴を考慮して作成したイシュー分類体系を使用して、ラベルを付与しています。これにより MT の出力結果に対する分析的な評価が可能となりますし、またそのような分析的な評価の自動化に関する研究にも利用できます。

2 つ目は、原文と MT 訳の両方に、単語レベル翻訳品質推定ラベルを付与したデータセット QEdatasetJaEn^[25]です。MTPEdocs 中の PE 訳を正解とみなして、MT 訳中の修正すべきテキストスパンに BAD のラベルを、それ以外のスパンに OK のラベルを付与しています。また、MT 訳の OK/BAD ラベルに対応する日本語原文箇所にも OK/BAD のラベルを付与しています。このデータにより、翻訳文中の誤り箇所やその原因となりうる原文箇所を推定する単語レベル品質推定の研究が可能となります。

3 つ目は、MT 訳と PE 訳の対をプリミティブな編集操作の系列に分解したデータセット DecomposedMTPE^[26]です。具体的には以下のように、MT 訳と PE 訳の対を最小単位の書き換えでつなぐような中間文を作成しています。

- **MT 訳** : We are **making** an operation plan ...
- **中間文** : We are releasing an operation plan ...
- **PE 訳** : We have released an operation plan ...

このように分解することで、MT 訳に観察される問題を、訳語の選択 (make→release) とアスペクトの選択 (現在進行形→現在完了形) の問題に細分化して捉えることが可能になります。このデータセットは、MT 訳の分析的評価の定式化や自動化の研究に使えます。

以上、それぞれのデータセットによって可能になる技術や研究課題のいくつかに言及してきましたが、これに留まらない様々な用途で利用されることを期待しています。また、ここで示したデータセットは MT 訳と PE 訳の利用が中心です。名古屋市翻訳資源中の人

手翻訳を利用したデータセットの開発とそれを活用した研究も今後の重要な課題です。

5.2. 社会的展開

名古屋市翻訳資源の社会応用的な展開について取り上げます。まず、最も直接的な利用として、名古屋市役所における行政文書の翻訳業務に名古屋市翻訳資源が活用されています。特に対訳文と対訳用語でカスタマイズした MT を、対訳文と対訳用語を登録した翻訳支援ツールと連携させて使うことで、翻訳効率が向上したことが報告されています^[27]。

また、名古屋市翻訳資源の対訳文と対訳用語を活用した資源検索プラットフォームが企業により開発・公開されています^[28]。これも翻訳資源を特定の自治体の中だけで使うのではなく、外部に公開したからこそ可能になった試みだと言えます。このプラットフォームは誰でも無料で使用できます。

また、能登半島地震被災者支援サイト^[29]に搭載されている MT のカスタマイズに、名古屋市翻訳資源が使われています。2024 年 1 月に震災が起きた後、急ピッチでサイトの構築が進み、多言語翻訳の話が持ち上がりました。この時、企業による多大な協力のもとで、10 日程度の期間で、名古屋市翻訳資源でカスタマイズした LLM ベースの MT システムを導入することができました。とはいえ、現在の翻訳資源にはカバレッジの課題があり、例えば「再生手続」や「半壊」といった被災者支援に関する用語が含まれておらず、それが MT 出力における訳語の揺れを生んでしまいます。そこで、さらなる用語集の拡張にも取り組んでいます^[30]。

6. おわりに

本稿では、我々が進めている自治体翻訳資源プロジェクトについて、MT と関わる側面を中心にその背景・理念・枠組みを説明した上で、現在の状況の一部を紹介しました。

冒頭に述べたような大きな目標を掲げていますが、比較的少数のアカデミックなメンバーのみでできることはどうしても限られています。カスタマイズされた MT システムを無料のサービスとして全自治体に提供するようなことは当初から本プロジェクトの範囲外です。あくまで、ベースとなる翻訳資源の構築と、利用環境の概念実証が短期的なゴールです。しかし、数年間の研究に終わらせず、この試みを全国に広げ、持続可能にしていくための方法も考えなければなりません。そのためにも、これまで行ってきたように、自治体や企業の方との連携が非常に重要だと感じています。

必要な情報を、希望する言語で、しかも、良い品質で得られることが当たり前の社会になることを目指して、プロジェクトを進めていきます。この理念を共有し、ご協力いただけるという方は、ぜひご連絡いただけますと幸いです。

謝辞

本研究は、JSPS 科研費（課題番号：23K28378）の支援を受けました。

注・参考文献

- [1] <https://tr4lg.p.u-tokyo.ac.jp/>
- [2] 宮田玲 (2024). 「自治体情報の多言語化支援の取り組み」言語処理学会第 30 回年次大会 (NLP2024) 併設ワークショップ「自治体における生成 AI (ChatGPT) の利活用と問題点」.
- 宮田玲, 阪本章子, 藤田篤, 香川璃奈 (2024). 「多言語情報発信を支援する自治体横断型翻訳資源の構築：プロジェクトの理念、枠組み、現状」日本通訳翻訳学会第 25 回年次大会.
- 宮田玲 (2024). 「自治体情報の多言語化を支える翻訳資源とテクノロジー」AAMT 2024, Tokyo.

[3] 出入国在留管理庁 (2024).「令和6年6月末現在における在留外国人数について」https://www.moj.go.jp/isa/publications/press/13_00047.html

国籍・地域別の在留外国人数の推移は、以下を参照。
<https://www.moj.go.jp/isa/content/001425981.pdf>

[4] 宮田玲 (2020). 「日本における自治体ウェブサイトの多言語化の現況と課題」『通訳翻訳研究』20, 1-24.

[5] この情報は筆者らがこれまでの活動で見聞きした範囲の内容に基づいています。自治体における翻訳技術の活用状況については、日本だけでなく世界的にも調査・研究が不足しており、問題点の全体像が明らかになっていないという課題があります。

Miyata, R. (2025, forthcoming). Machine translation in local government. In S. Baumgarten & M. Tieber (Eds.), *The Routledge Handbook of Translation Technology and Society*. Routledge.

我々のプロジェクトでは、自治体における翻訳の問題点やニーズを明らかにすべく、複数の自治体を対象に質問紙やインタビューによる調査も進めています。

Sakamoto, A., & Miyata, R. (2025). Machine translation in Japanese local governments: Human resource strategies and challenges. The 11th Asia-Pacific Translation and Interpreting Forum (APTIF).

[6] このような翻訳に関する指針策定の問題は、翻訳政策 (translation policy) の研究分野で扱われており、日本でも近年注目を集めています。

島津美和子 (2022). 「translation policy をめぐる先行研究の検討と今後の研究課題」『通訳翻訳研究』22, 53-74.

武田珂代子, 辛島デイヴィッド, 宮田玲, 島津美和子, 吉田理加 (2023). 「『日本におけるトランスレーション・ポリシー』プロジェクトの活動報告」日本通訳翻訳学会第24回年次大会.

[7] これらの特徴は、自治体文書に限らず、マニュアルやレシピといった情報伝達型の文書について、多かれ少なかれ当てはまります。

[8] 「一から翻訳する」のではないパラダイムとしては、伝統的には用例ベース MT の研究がなされてきました。近年でも、用例を活用した NMT の研究があります。これらの多くは、言語的・テキスト的・ベクトル表現的な類似性に基づく用例の検索に基づいていますが、これを文書論的な参照妥当性に基づく用例の検索に展開することが理論的な課題です。

[9] Miyata, R. (2023). Formulating a terminology for source document profiling through a literature review: From functionalist to documentational approaches. *Perspectives*, 31(5), 864–881.

[10] 影浦峯 (2019). 「人間の翻訳と機械の翻訳 (1): 翻訳者は何を翻訳しているか?」*AAMT Journal*, 71, 14-19.

[11] Bedard, C. (2000). Memoire de traduction cherche traducteur de phrases... *Traduire*, 186, 41–49.

Bowker, L. (2005). Productivity vs quality: A pilot study on the impact of translation memory systems. *Localisation Focus*, 4(1), 13–20.

[12] Mossop, B. (2006). Has computerization changed translation? *Meta*, 51(4), 787–805.

[13] Maruf, S., Saleh, F., & Haffari, G. (2022). A survey on document-level neural machine translation: Methods and evaluation. *ACM Computing Surveys*, 54(2), 1–36.

[14] 朴恵, 山浦育子, 宮田玲 (2024). 「日中対訳用語集構築に向けた翻訳の規準と手続きの明確化: 被災者支援分野を対象に」日本通訳翻訳学会第25回年次大会.

[15] 科研費・基盤研究(S)プロジェクト「翻訳規範とコンピテンスの可操作化を通じた翻訳プロセス・モデルと統合環境の構築」(課題番号: 19H05660) .

これまで開発されたメタ言語については以下を参照。
<https://tntc-project.github.io/>

[16] Chu, C., & Wang, R. (2020). A survey of domain adaptation for machine translation. *Journal of Information Processing*, 28, 413–426.

- [17] Post, M., & Vilar, D. (2018). Fast lexically constrained decoding with dynamic beam allocation for neural machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* (pp. 1314–1324).
- [18] Post, M., Ding, S., Martindale, M., & Wu, W. (2019). An exploration of placeholder in neural machine translation. In *Proceedings of Machine Translation Summit XVII* (pp. 182–192).
- [19] Song, K., Zhang, Y., Yu, H., Luo, W., Wang, K., Zhang, M. (2019). Code-switching for enhancing NMT with pre-specified translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* (pp. 449–459).
- [20] 継続事前学習に使うことも可能ですが、学習コストやデータ規模を考慮する必要があります。
- [21] Nitta, J. (2022). A customisable automated quality assurance tool. In R. Miyata, M. Yamada, & K. Kageura (Eds.), *Metalanguages for Dissecting Translation Processes: Theoretical Development and Practical Applications* (pp. 200–215). Routledge.
- [22] 名古屋市, 宮田玲 (2023). 「名古屋市翻訳資源」 <https://github.com/tr4lg/nagoya-dataset/>
名古屋市 (2023). 「AI 翻訳に活用可能な用語集データの公開について」 <https://www.city.nagoya.jp/kankobunkakoryu/page/0000162160.html>
- [23] Miyata, R. (2024). MTPEdocs. <https://github.com/tntc-project/MTPEdocs>
- [24] Fujita, A. (2024). MTPEdocs-MQM. <https://github.com/tr4lg/MTPEdocs-MQM>
- [25] Shimada, S. (2024). QEdatasetJaEn. <https://github.com/tntc-project/QEdatasetJaEn>
- 島田紗裕華, 山口大地, 宮田玲, 藤田篤, 梶原智之, 佐藤理史 (2024). 「機械翻訳向け原文編集の支援に向けた日英翻訳品質推定データセットの設計と構築」 言語処理学会第 30 回年次大会.
- [26] Yamaguchi, D. (2024). <https://github.com/tntc-project/DecomposedMTPE>
Yamaguchi, D., Miyata, R., Fujita, A., Kajiwara, T., & Sato, S. (2024). Automatic decomposition of text editing examples into primitive edit operations: Toward analytic evaluation of editing systems. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)* (pp. 1899–1914).
- [27] 伊藤毅 (2023). パネルディスカッション「アフターコロナを見据えた今後の AI を活用した多言語翻訳技術の役割について」 第 6 回自動翻訳シンポジウム. https://gcp.nict.go.jp/event/article_20230216.html
- Miyata, R. (2023). Managing translation technologies in local governments through collaborations with internal and external actors: A case of Nagoya City. The 8th AAS-in-Asia.
- [28] 株式会社システートソリューションズ. 自治体言語資産データベース. <https://jichitai.translation-db.com/view>
- [29] 株式会社アスコエパートナーズ. 令和 6 年能登半島地震 支援情報ナビ. <https://aidfor.noto-peninsula-earthquake2024.supportnavi.jp/>
- [30] 注[14]参照。用語集は以下で公開。
<https://github.com/tr4lg/controlled-multilingual-terminology>

AI 事業者ガイドラインについて

藤本 篤志

総務省 情報流通行政局 参事官室

1. はじめに

総務省及び経済産業省は、令和6年4月19日、「AI事業者ガイドライン（第1.0版）」（以下「本ガイドライン」という。）を策定・公表した。本ガイドラインはAI活用に取り組むすべての事業者が、AIのリスクを正しく認識し、必要となる対策を自主的に実行できるよう、AIの開発・提供・利用の観点から基本的な考え方を示したものである。

本稿では、AI事業者ガイドラインの概要について解説すると共に、令和6年度中に行った更新について示すこととしたい。なお、本稿中意見にわたる部分は筆者らの個人的見解であり、所属する組織を代表するものではないことを予めお断りしておきたい。

2. AI事業者ガイドラインの概要

本ガイドラインは、【図1】のとおり、①事業者の自主的な取組の支援、②国際的な議論との協調、及び③読み手にとってのわかりやすさという点を重視している。



【図1】

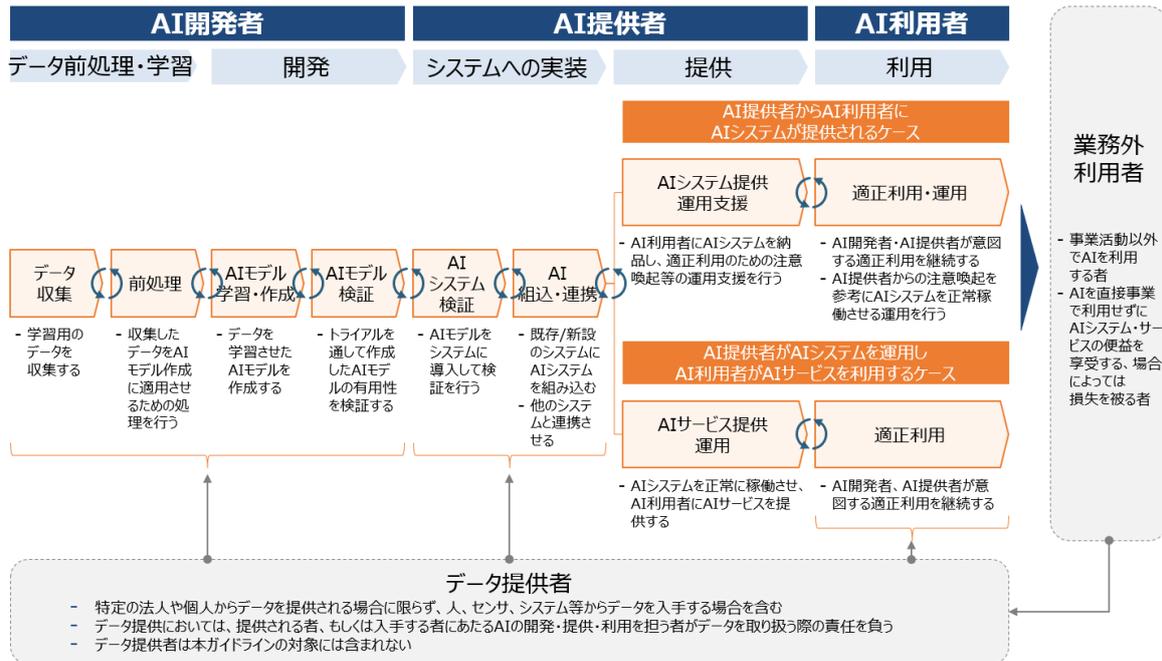
また、政府が一方的にその内容を定めたものではなく、教育・研究機関、一般消費者を含む市民社会、民間企業等で構成されるマルチステークホルダーで検討を重ねた結果として策定されたものであること、及び

Living Document として継続的改善・更新等が予定されていることに特徴がある。

本ガイドラインでは、AIの事業活動を担う一般的な主体を「AI開発者」、「AI提供者」、「AI利用者」の3つに大別しており、それぞれ以下のとおり定義している（【表1】）。

主体	定義
AI開発者	AIシステムを開発する事業者（AIを研究開発する事業者を含む） ⇒AIモデル・アルゴリズムの開発、データ収集（購入を含む）、前処理、AIモデル学習及び検証を通してAIモデル、AIモデルのシステム基盤、入出力機能等を含むAIシステムを構築する役割を担う。
AI提供者	AIシステムをアプリケーション、製品、既存のシステム、ビジネスプロセス等に組み込んだサービスとしてAI利用者（AI Business User）、場合によっては業務外利用者に提供する事業者 ⇒AIシステム検証、AIシステムの他システムとの連携の実装、AIシステム・サービスの提供、正常稼働のためのAIシステムにおけるAI利用者（AI Business User）側の運用サポート又はAIサービスの運用自体を担う。AIサービスの提供に伴い、様々なステークホルダーとのコミュニケーションが求められることもある。
AI利用者	事業活動において、AIシステム又はAIサービスを利用する事業者 ⇒AI提供者が意図している適正な利用を行い、環境変化等の情報をAI提供者と共有し正常稼働を継続すること又は必要に応じて提供されたAIシステムを運用する役割を担う。また、AIの活用において業務外利用者に何らかの影響が考えられる場合は、当該者に対するAIによる意図しない不利益の回避、AIによる便益最大化の実現に努める役割を担う。

【表1】



【図2】

AIの開発から利用までの一般的なバリューチェーンにおいては、AI開発者が収集データを用いてAIモデルの作成を行い、AI提供者はそのAIモデルをシステムに組み込むことでAIシステムを構築する。構築されたAIシステムまたはそのシステムによるAIサービスがAI利用者に提供され、利用に至る（【図2】）。これらの主体は事業者（または各者内の部門）を想定しており、AIの活用方法によっては同一の事業者がAI開発者、AI提供者またはAI利用者の複数を兼ねる場合も想定される。

「AI」と言っても、その種類は多岐にわたり、今後のAI技術の在り方については有識者であっても予測することは困難である。このような状況を踏まえつつ、本ガイドラインにおける「AI」の概念については、「AIシステム」自体または機械学習をするソフトウェア若しくはプログラムを含む抽象的な概念」と定義している。また、「AIシステム」については、「自律性をもって動作し学習する機能を有するソフトウェアを要素として含むシステム」と定義し、その例として、機械、ロボット、クラウドシステム等を挙げている。本ガイドラインにおける関連する用語の定義は右の【表2】のとおり。

用語	定義
AI	現時点で確立された定義はなく、広義の人工知能の外延を厳密に定義することは困難である。本ガイドラインにおけるAIは「AIシステム（以下に定義）」自体又は機械学習をするソフトウェア若しくはプログラムを含む抽象的な概念とする。
AIシステム	活用の過程を通じて様々なレベルの自律性をもって動作し学習する機能を有するソフトウェアを要素として含むシステムとする（機械、ロボット、クラウドシステム等）。
AIモデル/MLモデル	AIシステムに含まれ、学習データを用いた機械学習によって得られるモデルで、入力データに応じた予測結果を生成する。
AIサービス	AIシステムを用いた役務を指す。AI利用者への価値提供の全般を指しており、AIサービスの提供・運営は、AIシステムの構成技術に限らず、人間によるモニタリング、ステークホルダーとの適切なコミュニケーション等の非技術的アプローチも連携した形で実施される。
生成AI	文章、画像、プログラム等を生成できるAIモデルにもとづくAIの総称を指す。
AIガバナンス	AIの利活用によって生じるリスクをステークホルダーにとって受容可能な水準で管理しつつ、そこからもたらされる正のインパクト（便益）を最大化することを目的とする、ステークホルダーによる技術的、組織的、及び社会的システムの設計並びに運用。

【表2】

わが国では、平成31年3月に「人間中心のAI社会原則」を策定した。これは、国内外における倫理的な検討の成果を踏まえた上で、より上位の理念として定めているものであり、著しい技術の発展によっても不変かつ目指すべき理念であり続けていることから、本ガイドラインにおいてもこれに基づき、「人間の尊厳が尊重される社会 (Dignity)」、「多様な背景を持つ人々が多様な幸せを追求できる社会 (Diversity and Inclusion)」、「持続可能な社会 (Sustainability)」の3つの価値を基本理念として掲げている。

本ガイドラインでは、「基本理念」を実現するために各主体が念頭におくべき「原則」に紐付けられる「共通の指針」として、次の10項目を掲げている(【表3】)。なお、この「原則」及び「共通の指針」は、「人間中心のAI社会原則」を土台としつつ、諸外国における議

論状況や新技術の台頭に伴い生じるリスクへの対応を踏まえ、再構成を行ったものである。

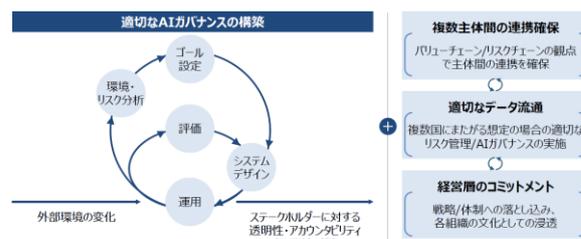
各主体は、原則・共通の指針に掲げる取組にあたり、「1) 人間中心」に照らし、法の支配、人権、民主主義、多様性、公平公正な社会を尊重するようAIシステム・サービスを開発・提供・利用すべきであり、加えて、憲法、知的財産関連法令および個人情報保護法をはじめとする関連法令、AIに係る個別分野の既存法令等を遵守するとともに、国際的な指針等の検討状況についても留意することが重要である。なお、これらの取組は、各主体が開発・提供・利用するAIシステム・サービスの特性、用途、目的及び社会的文脈を踏まえ、各主体の資源制約を考慮しながら自主的に進めることが期待される。

指針	内容(主な項目の抜粋)
各主体が 取り組む事項	1) 人間中心 <ul style="list-style-type: none"> ✓ AIが人々の能力を拡張し、多様な人々の多様な幸せ(well-being)の追求が可能となるよう行動する ✓ AIが生成した偽情報・誤情報・偏向情報が社会を不安定化・混乱させるリスクが高まっていることを認識した上で必要な対策を講じる ✓ より多くの人々がAIの恩恵を享受できるよう社会的弱者によるAIの活用を容易にするよう注意を払う
	2) 安全性 <ul style="list-style-type: none"> ✓ 適切なリスク分析を実施し、リスクへの対策を講じる ✓ 主体のコントロールが及ぶ範囲で本来の利用目的を逸脱した提供・利用により危害が発生することを避ける ✓ AIシステム・サービスの特性及び用途を踏まえ、学習等に用いるデータの正確性等を検討するとともに、データの透明性の支援、法的枠組みの遵守、AIモデルの更新等を合理的な範囲で適切に実施する
	3) 公平性 <ul style="list-style-type: none"> ✓ 特定の個人ないし集団へのその人種、性別、国籍、年齢、政治的信念、宗教等の多様な背景を理由とした不当で有害な偏見及び差別をなくすよう努める ✓ AIの出力結果が公平性を欠くことがないよう、AIに単独で判断させるだけでなく、適切なタイミングで人間の判断を介在させる利用を検討した上で、潜在的なバイアスに留意し、AIの開発・提供・利用を行う
	4) プライバシー保護 <ul style="list-style-type: none"> ✓ 個人情報保護法等の関連法令の遵守、各主体のプライバシーポリシーの策定・公表により、社会的文脈及び人々の合理的な期待を踏まえ、ステークホルダーのプライバシーが尊重され、保護されるよう、その重要性に応じた対応を取る
	5) セキュリティ確保 <ul style="list-style-type: none"> ✓ AIシステム・サービスの機密性・完全性・可用性を維持し、常時、AIの安全な活用を確保するため、その時点での技術水準に照らして合理的な対策を講じる ✓ AIシステム・サービスに対する外部からの攻撃は日々新たな手法が生まれ、これらのリスクに対応するための留意事項を確認する
	6) 透明性 <ul style="list-style-type: none"> ✓ AIを活用する際の社会的文脈を踏まえ、AIシステム・サービスの検証可能性を確保しながら、必要かつ技術的に可能な範囲で、ステークホルダーに対し合理的な範囲で適切な情報を提供する(AIを利用しているという事実、活用している範囲、データ収集及びアノテーションの手法、AIシステム・サービスの能力、限界、提供先における適切/不適切な利用方法、等)
	7) アカウンタビリティ <ul style="list-style-type: none"> ✓ トレーサビリティの確保や共通の指針の対応状況等について、ステークホルダーに対して情報の提供と説明を行う ✓ 各主体のAIガバナンスに関するポリシー、プライバシーポリシー等の方針を策定し、公表する ✓ 関係する情報を文書化して一定期間保管し、必要なときに、必要なところで、入手可能かつ利用に適した形で参照可能な状態とする
社会と 連携した 取組が 期待される 事項	8) 教育・リテラシー <ul style="list-style-type: none"> ✓ AIに関わる者が、その関わりにおいて十分なレベルのAIリテラシーを確保するために必要な措置を講じる ✓ AIの複雑性、誤情報といった特性及び意図的な悪用の可能性もあることを勘案して、ステークホルダーに対しても教育を行うことが期待される。
	9) 公正競争確保 <ul style="list-style-type: none"> ✓ AIを活用した新たなビジネス・サービスが創出され、持続的な経済成長の維持及び社会課題の解決策の提示がなされるよう、AIをめぐる公正な競争環境が維持に努めることが期待される
	10) イノベーション <ul style="list-style-type: none"> ✓ 国際化・多様化、産学官連携及びオープンイノベーションを推進する ✓ 自らのAIシステム・サービスと他のAIシステム・サービスとの相互接続性及び相互運用性を確保する ✓ 標準仕様がある場合には、それに準拠する

【表3】

【表2】に記載のとおり、AI ガバナンスは、本ガイドラインにおいて、「AI の利活用によって生じるリスクをステークホルダーにとって受容可能な水準で管理しつつ、そこからもたらされる正のインパクト(便益)を最大化することを目的とする、ステークホルダーによる技術的、組織的、及び社会的システムの設計並びに運用」と定義される。AI を安全安心に活用していくために、事業者としては、経営層のリーダーシップのもと、AI ガバナンスを適切に構築することで、リスクをマネジメントしていくことが重要となる。AI ガバナンスの構築に当たっては、【図3】のとおり、事前にルール又は手続が固定された AI ガバナンスではなく、「環境・リスク分析」「ゴール設定」「システムデザイン」「運用」「評価」といったサイクルを、継続的かつ高速に回転させていく、「アジャイル・ガバナンス」の実践が重要となる。そして、具体的な検討にあたっては、AI のもたらすリスクの程度及び蓋然性、並びに各主体の資源制約に配慮することが重要となる。

さらに、AI ガバナンスの構築に当たっては、複数主体に跨る論点について、バリューチェーン上における主体間の連携を確保すること、バリューチェーンが



【図3】

複数国にわたる場合、データの自由な越境移転の確保のための適切な AI ガバナンスを検討すること、経営層のコミットメントの下、各組織の戦略や企業体制への落とし込みや文化としての浸透を図ることなどにも留意することが求められる。

3. 令和6年度の更新点

令和6年4月19日に策定後、同年11月に時点更新を行った後、AI ネットワーク社会推進会議・AI ガバナンス検討会 合同会議及びAI 事業者ガイドライン検討会における議論や事業者へのアンケート及びヒアリング、最新技術等の動向調査を踏まえ、【表4】のとおり令和6年度中に更新すべきポイントを整理の上、令和7年3月28日に更新版を公表した。

令和6年度の更新論点及び更新方針 一覧

#	更新論点	主な更新方針	主な修正箇所
1	AIによるリスクの洗い出し・分類	リスク例の追加・分類案の記載・「共通の指針」とのマッピング (過度な依存、金銭的損失などを追加、技術的リスク・社会的リスクなどに分類)	別添
2	AIの契約に関する留意事項	AIの開発、提供、利用における契約に関して留意すべき事項の記載 (契約モデルの多様化や責任分界に関する記載の充実)	別添
3	生成AIに関する記載の追加	生成AIに関する新たなリスクや留意点の記載 (マルチモーダルな生成AI、RAG導入、プログラムコード生成など記載を追加)	本編、別添
4	AIガバナンスに関する事例の充実	AIガバナンスの取組事例の充実化 (「リスクベース・アプローチ」を実施する上で考慮すべき点、様々な企業の事例追加、「人材不足」の課題を追加)	別添
5	AIガバナンスの動向等の反映	AIガバナンスに関する国内外の最新動向を追記 (AI制度研究会、広島AIプロセスの動向等、最新状況を追記)	本編、別添
6	特定単語の整理・見直し	AIガバナンスにおいて重要な単語の定義や表現の見直し (「バイアス」、「透明性」、「多様性」、「包摂性」などの定義や表現の揺れを修正)	本編
7	その他	UIの改善 (目次から該当ページへのリンクを追加)	本編、別添

【表4】

(1) AIによるリスクの洗い出し・分類

事業者が AI リスクを把握・対応しやすくなるように、AI によるリスクの記載を整理・拡充した。具体的には、「過度な依存」、「労働者の失業」及び「データや利益の集中」などリスク例を追加した上で、これらリスク例を「技術的リスク」と「社会的リスク」に大別し、技術的リスクは「学習及び入力段階」「出力段階」「事後対応段階」に、社会的リスクは「倫理・法」「経済」「情報空間」「環境」の4つに分類する等して【表5】のとおり分類案を示し、これらの具体的な対応策につなげるべく、本ガイドラインの「共通の指針」とのマッピングを示した。

(2) AIの契約に関する留意事項

AI 技術を用いたサービスが普及し、契約類型が多様化する中で、典型的な契約類型を整理するとともに、経済産業省にて令和7年2月に公開した「AIの利用・

開発に関する契約チェックリスト」を参考資料として追記した。また、責任分界の明確化が求められる背景や、事故発生リスクとの関係で契約上の留意が有益な事項等を追記した。

(3) 生成 AI に関する記載の追加

生成 AI に関する技術の進歩や事業者における導入の進展を踏まえ、マルチモーダルな生成 AI、RAG、プログラムコードの生成、AI エージェントに関する便益や留意する事項等の記載を拡充した。

(4) AI ガバナンスに関する事例の拡充

グローバルな AI ガバナンスを構築している企業や、中小・スタートアップ企業、地方自治体による実際の取組事例の追加や、「リスクベースアプローチ」の実践例、「人材不足」に関する対応方法等の記載を拡充した。

【表5】

・下表はAIのリスクを網羅したものではなく、想定に基づく事例も含んでおり、あくまで一例として認識することが期待される
 ・下表には政府等の公的機関も含めた社会全体での対応・議論が必要となるリスクも含まれる

大分類	中分類	リスク例
技術的リスク (=主にAIシステム特有のもの)	学習及び入力段階のリスク	データ汚染攻撃等のAIシステムへの攻撃
	出力段階のリスク	バイアスのある出力、差別的出力、一貫性のない出力等 ハルシネーション等による誤った出力
	事後対応段階のリスク	ブラックボックス化、判断に関する説明の不足
社会的リスク (=既存のリスクがAIにおいても発生又はAIによって増幅するもの)	倫理・法に関するリスク	個人情報への不適切な取扱い
		生命等に関わる事故の発生
		トリアージにおける差別
		過度な依存
	経済活動に関するリスク	悪用
		知的財産権等の侵害
		金銭的損失
		機密情報の流出
		労働者の失業
	情報空間に関するリスク	データや利益の集中
資格等の侵害		
偽・誤情報等の流通・拡散		
民主主義への悪影響		
フィルターバブル及びエコーチェンバー現象		
環境に関するリスク	多様性・包摂性の喪失	
	バイアス等の再生成	
		エネルギー使用量及び環境の負荷

(5) AI ガバナンスの動向等の反映

AI 制度研究会による中間とりまとめや文化庁による著作権に関する検討等の国内動向や、広島 AI プロセス等の国際動向において、注視すべき最新状況を追記した。また、経営層のガバナンスの管理・監督責任が問われる可能性があることにも留意するため、経営層の取組を補足した。

(6) 特定単語の整理・見直し

「バイアス」「多様性・包摂性」「透明性」など、AI ガバナンスにおいて重要な単語の明確化や表現の揺れの修正を行い、定義の参考となる文献を追記等した。

(7) その他

UI の改善として、目次から該当ページへのリンク等を追加した。

4. おわりに

急速に進展しつつある AI の高度化や国際的な議論等も踏まえ、本ガイドラインは Living Document として継続した見直しを行うことが必要である。AI によるリスク事例やその分類案については、最新の技術やサービスの発展等を踏まえ柔軟に追加、削除及び修正等がなされるべきである。

今後、AI をめぐる課題はより一層多岐にわたり、継続的かつ多面的な検討が必要であることは間違いない。本ガイドラインを通じて AI ガバナンス等に関与する一人一人が、好事例や失敗事例等の実事例の共有や議論等を通じて、本ガイドラインの普及・活用・アップデートに主体的に関与していくことを期待したい。

温故知新 5

内山 将夫

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、1993 年 9 月と 1993 年 12 月号の AAMT 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。

【1993 年 9 月 AAMT ジャーナル No.4 より】

翻訳の現場から

条件付期待感

今井 孝昌

翻訳をやりながらつねづね思うことは、息をぬけるところで思う存分息をぬいて、もっと知的な作業やゲーム感覚を楽しめるようにしたいということである。これはきっと翻訳者であれば誰でも考えていることだろう。もちろん、そこへ至るまでには、いろいろなプロセスを自分なりに経ていって、ぜいたくを言えるだけの高度な資質を身に付けなければならないわけだが、ある段階を超えたときにそういった喜びが得られな

かったら、仕事としての言語生活は実に味けないものになってしまう。こういう思いは、最近ますますつのるばかりである。もっともこんなことを言えるのは、頭の片隅に、いつも電子的なツールの存在が想定されているからである。

他の翻訳者に負けないように、電子的なツールは、すでに何種類かのものを利用している。しかし、何とんでも、データベースがまだ不十分である。私は、あまり深いことは考えずに、とにかく、翻訳に活かせる基礎的なデータベースを自分で作り続けることにしている。

ところで、半年ほど前に、市販の翻訳用ソフトを使っているところを、友人に見せてもらったことがある。このソフトは、自動翻訳という機能が組み込まれていることになっていた。しかし、はっきり言って使いものにならないのではないかと思った。開発した人間の言語感覚がすかさずかであることが露呈しているのであ

Learning from the past 5

Masao Utiyama

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/>

る。優れたソフトは、パワープログラマーの手によって生まれる。しかし、機械翻訳の分野となると、それだけではすまされないのではないか。

さて私の方は、もっぱらデータベース作りだが、しょせんこちらは一人だから、ソフトハウスの真似をするわけにもいかない。そこで、まずは作家あるいはノンフィクション・ライターを数人に絞りこんで、その作品を文章データベースの下地作りの参考にしている。この作業では、単なるテキストの通読では得られない思わぬ発見に行き当たったりする。

たとえば、漱石の作品には、テクニカル・ライティングに通じる要素が、ふんだんに盛り込まれているような気がする。これは、試しに「明暗」を手にとって、ほんの数十行を読んでみれば分かることではないだろうか。

これからは検索の設定を慎重に検討していかなければならない。しかし、うまく行けばそこでもまた、思わぬ収穫が得られるのではないかと秘かに考えている。何とかして、作品に投入されている膨大なエネルギーを自分なりに取り込んでいきたいものだ。

矛盾するようだが、ツールとしての今後の機械翻訳ソフトには条件付きの期待感を抱いている。多少値ははっても、納得の行くものであれば、すすんで購入することになるだろう。ただし、使ってみて、期待以上の効果が得られたとしても、コンピュータはあくまでツールであり、決定権は自分にあるのだということを自分に言い聞かせるつもりでいる。「人工知能」に頼りすぎて、無味乾燥な文章に甘んじてしまつては、元も子もない。それに、手元の二次資料から察すれば、人間の知性原理、ファジー論理、情動原理をふまえた開発が実現して、Aランクの翻訳者を震えさせるようなソフトが登場するまでには、まだすこし時間がかかりそうである(しかし、それがいつになるかは分からないが、そういったソフトは必ず現実のものとなるはずである)。

今のうちに、自分なりのデータベースや文章力をさらに充実させておき、優れたソフトが登場しても十分

にそれを統括していけるようにしたいものである。煩雑な作業から解放されて、大きな知的喜びを獲得できることを願って。

著者紹介:(いまたかまさ)

1953年 東京生まれ。上智大学法学部、出版社勤務を経てフリーランス翻訳者となる。翻訳の主な分野は、国際情勢、特許、コンピュータ、通信など。「エース・オブ・エース」(日本ソフトバンク)他、3冊の訳書がある。

【1993年12月 JAMT ジャーナル No.5 より】

利用技術ヒアリング

果てしなく広がる翻訳の世界

株式会社サイマル・インターナショナル
翻訳室長 早良哲夫

アジア太平洋機械翻訳協会(AAMT)には、研究者、システムメーカー、ソフトメーカー、ユーザーなどが加盟しています。ユーザーの中には機械翻訳(MT)システムを積極的に使っている企業や個人もいますが、サイマル・インターナショナルは、まだMTシステムを導入していない"ノン・ユーザー"です。この「まだ」という表現に込めた謙虚な気持ちを感じ取っていただきたいわけで、できることなら一日も早く導入したいと考えていることは事実です。

では、なぜ導入しないのか。まず、取り扱う翻訳の分野がMT向きではありません。翻訳の分野を大きく文科系と理科系に分けるとすれば、サイマルは主として文科系、つまり政治、経済、社会、文芸といった分野の翻訳を扱っています。当然のことながら、その中に理科系の要素が入っていることもありまた医学や薬学といった分野の翻訳を行うこともあります。いずれにせよ、MTが得意とするコンピューター関係とか技術マニュアルといった種類の文書は、限りなくゼロに近い状態です。

次に翻訳の方向が MT 向きではありません。サイマルの場合、外国語から日本語への翻訳よりも日本語から外国語への翻訳のほうが多く、そのうえ、外国語といっても必ずしも英語だけではなく、欧州系の諸言語や中国語、韓国(朝鮮)語なども取り扱っています。これに対して、MT が英語から日本語という方向性に優れていることは周知のことです。

第 3 に、翻訳の手順が MT 向きではありません。MT の場合、一般的には、例えば英語の原文を機械が読み取りやすいように「前編集」してから機械にかけ、翻訳されて出てきた日本語を「後編集」という手順になると思います。これに対して、サイマルの場合は、その分野を得意とする翻訳者——翻訳者を指名する際、サイマルでは「外国語から母国語へ」を原則としています——に翻訳を依頼し翻訳者から提出された訳文を社内で「エディット」します、原則として「前編集」がない代わりに、MT の「後編集」に相当する作業を重視し、殆ど全文を書き直してしまうことも稀ではありません。またサイマルの翻訳者は殆どが在宅で、社内には最少限度の要員しか置いていませんが、MT の場合は「前編集」や「後編集」の担当者が機械の至近距離——つまり社内——で待機していなければなりませんし、機械は 24 時間稼働しますから、それだけ社内の要員を増やす必要も生じてきます。

こういった諸々の要素が絡み合い、MT 導入に踏み切れないのが現状です。しかし、最初にも述べたように、現段階では「まだ」導入できないというだけのことで、サイマルが AAMT に加盟しているのも、この「まだ」という接続詞を無意味なものに終わらせたくないからです。メーカー筋からも何度となくアプローチがありました、その都度、真剣に検討していますが、もし現段階で MT を導入するとすると、今の仕事の分野を捨てて MT 向きの分野に乗り換えるか、今の仕事を続けながら、それとは別に MT 担当部を発足させて独自の業務と取り組むか…といった難しい選択を迫られ、いつも時期尚早という結論に至るわけです。

では、サイマルの翻訳部はメカに弱いのか。いえそうではありません。コンピューターやワープロは使いこなしていますし、もちろん電子メールのシステムなども日常的に駆使しています。ただ、我々のプロフェッショナルなニーズに対して、MT が「まだ」充分に応えてくれないのが悩みの種なのです。

参考にしていただくために、実際の翻訳例を 2 つ 3 つ持ってきました。最初は、訪米した日本の経済人が日米経済摩擦について講演した時の原稿の一部で、日本語から英語への翻訳例です。クライアント側の最初の原稿では、講演の締めくくりの部分が「(経済摩擦解消のため)皆様も米国サイドで同様のご努力をお願い申し上げます」で終わっていましたが、その後、クライアント側から

「その夜の大リーグのオールスター戦に引っかけてこの部分に『ご質問に対しては全力投球でお答えします』という文を付け加えたい」

という連絡がありました。これに対してサイマルは「質問に対して"全力投球"するというのは変だ。まず質問側が投球し、回答者が打ち返すのが自然の順序ではないか」と反対提案し、

「ご質問には全力で打ち返すので、思い切って投げ込んでほしい。ただし、カーブは困る」

という趣旨の英訳案を出しました。すると、クライアント側から

「投打の順序についてはサイマル案に従うが、もう少し謙虚に『こちらはマイナーリーガーだが、皆さんはメジャーリーガーだから・・・』という表現を加えたい」

という申し入れがありました、これに対しては、

「日米関係について熱弁を振ったあとで『私はマイナーリーガーだ』では、逆に『私の言っていることは無視してくれ』と訴えているように——少なくとも米国人には——聞こえる」

と反論しました。それでもクライアント側は

「オールスター戦の直前のことでもあり、何とかして"メジャーリーガー"という言葉を入れたい」と固執。結局、これならばと作ったのが

"I hope I am able to hit pitches off major leaguers like yourselves. But, please, no curves"

という英文でした。

この程度のやり取りは毎日のようにやっていますが、これは翻訳というよりも、むしろ如何にして異文化間のギャップを埋めるとかいう非常に人間的な側面が強く、いわゆる「後編集」の範囲を超えているのではないかと思います。

次は普通の「後編集」の例です。これは英語から日本語への翻訳で、知日の英国人が英語で原稿を書き、日本語で行なったスピーチです。翻訳家が訳したものに社内で手を入れた事例ですが、誤訳があるかどうかは別にして、殆ど全面的に書き直しました。最初の訳文が「である」調だったのを、スピーチに適した「ですます」調に改め、漢字を少なくすると同時に、視覚的に読みやすくするために動詞の語幹に当たる部分は、仮名書きではなく漢字を増やしました。また、耳で聞いて判別しやすい単語に書き改めた個所の少なくありません。

最後の例は、日本画の解説文で、日本語からフランス語への翻訳です。絵には梅の枝が描かれており、解説文を仏訳したのは日本語に堪能なフランス人ですが、さすがに菅原道真と天神様と梅の相関関係は知らなかったようで、訳文には正直に「この部分は原文の意味が分からない」と添え書きしてありました。そこで、あの「東風吹かば・・・」の故事を理解するのに必要と思われる最少限度の言葉を補ってみました。

さて、一連の翻訳作業を進めていく上で軸になるのが「コーディネーター」です。コーディネーターは、翻訳の受注から完成品の納入、そしてアフターサービスに至るまで、一貫して"プロジェクト・マネジャー"の役割を果たします。翻訳者やエディターを手配した

り、クライアントからの注文を聞いて翻訳者やエディターに伝え、翻訳者からの問い合わせをクライアントに伝えるなど、すべて同一のコーディネーターが窓口になって、全体の動きを調整しながらプロジェクトを進めていくわけです。もちろん、どの翻訳エージェンシーでも大なり小なりやっていることですが、コーディネーターを軸として関係者の間で密度の濃い連絡が繰り返されるうちに、少しずつ翻訳が磨き上げられ、固まっていきます。サイマルは、別表に示されているように、この流れを非常に重要視しています。

この段階になると、先にも述べたように、もはや「後編集」ではなく、人間的対話を積み重ねて異文化間に橋を架けていくようなもので、この中でMTが占める場所があるか、MTを導入した場合に効率が上がるか、正直なところ現時点では疑問に思います。

質疑応答

問：MTが得意とするのはマニュアルなど大量翻訳の分野だと思うが、その他の分野にもMTが入り込める余地はあると思うか。

答：MTは、今の段階では主として技術分野の翻訳に生かされていますが、メーカーの展示会などでは文化系の文書の翻訳にも相当な才能を発揮しています。これが展示会のような特別の機会だけではなく、日常的になればよいわけです。日進月歩の時代ですから、そういう日が意外に早く到来するかもしれません。ただ、行間を読んで異文化間に橋を架けるようなことは、機械には無理でしょう。いずれにせよ、MT"システム"といった大袈裟な装置ではなく、例えば普通のラップトップ型ワープロにフロッピーを挿入するだけで、一定程度の翻訳作業が自動的に行えるようになれば、個人翻訳者の間にもMTが普及するのではないのでしょうか。

問：翻訳業界における「翻訳」の定義について。

答：実際問題として定義するのは不可能に近いし翻訳団体の会則などを見ても「翻訳とは何か」という定義はないようです。文芸作品の翻訳こそ本当の意味

での翻訳だと主張する人もいますし、超訳とか翻案といった方式が翻訳といえるかどうかについても賛否両論があります。ついでながら、日本には翻訳者の団体も幾つかあるうえ、ローマ字の略称が“JTF”, “JTA”, “JAT”, “JST”というように互いに似ているため、先ごろ開かれた翻訳者の国際会議でも何とかならないかという要望が出たそうです。

問い：放送や映画のメテリア翻訳も、文書の翻訳と相当異なるものがあるが・・・

答え：映画やテレビは画面に出せる字幕の字数に、時間的にも空間的にも制限があり、テキストに沿っての翻訳ではなく、要点をピックアップしてストーリーの流れを伝えることを重点を置かざるを得ません。

テレビやラジオのニュースも同様で、例えば英語のニュースと日本語のニュースでは構成が違うため、そのまま訳しても結果は極めて不自然です。むしろ「英語で送られてきたニュースの内容を咀嚼して日本語のニュースを書き下ろす」といった感じです。

問い：前編集や後編集の比重について。

答え：先にも触れたように、前編集は原則として行なっていません。後編集はどうか、単に「エディット」と呼んでいますが、この過程は誤訳チェックと品質管理の両面で大いに重視しています。翻訳者が英語のネイティブ・スピーカーの場合は、日本語のネイティブ・スピーカーがチェックする、その逆の場合は逆の組み合わせにする・・・というのが原則です。翻訳者といえば、とかく「この単語は知っている」と思いがちです。事実、翻訳者の語彙は豊富ですが、知っている単語でも、知らなかったつもりで辞書を引くと、思い違いに気づくこともあれば、新しい意味を発見することもあり、決して無駄ではありません。

翻訳や編集をする時は、いろいろな辞書や参考文献を最大限に利用すべきです。その点 社内にいると各種の辞書や参考書を使えるので便利です。こういった辞書や参考文献の電子化が一段と進み、一般の翻訳者も気楽に使えるようになれば、それに越したことはないと思います。

第 11 回アジア翻訳ワークショップ (WAT2024) 開催報告

中澤 敏明
東京大学

1. はじめに

本稿では WAT2024[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、2024 年の第 11 回 (WAT2024) はアメリカのマiami で開催された EMNLP2024 の併設ワークショップとして、2024 年 11 月 16 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 20 名程度で、現地参加が 1/3 ほどであった。

2. 研究論文

WAT2024 では 5 件の研究論文を採択した。採択した研究論文のタイトルを以下に示す。

- Machine Translation Of Marathi Dialects: A Case Study Of Kadodi
- AI-Tutor: Interactive Learning of Ancient Knowledge from Low-Resource Languages
- An Empirical Study of Multilingual Vocabulary for Neural Machine Translation Models
- Are Large Language Models State-of-the-art Quality Estimators for Machine Translation of User-generated Content?

- Creative and Context-Aware Translation of East Asian Idioms with GPT-4

この 5 件の中で、査読の評価が最も高かった “Machine Translation Of Marathi Dialects: A Case Study Of Kadodi” を best paper として選出した。

3. パネルディスカッション

WAT2024 では招待講演のかわりにパネルディスカッションを行なった。パネリストとして蘇州大学の Min Zhang 氏 (ビザの関係で残念ながら当日出席できず)、National Electronics and Computer Technology Center (NECTEC) の Thepchai Supnithi 氏、川村インターナショナルの森口功造氏、TransPerfect の Fred Bane 氏をお迎えし、以下のようなテーマで討論を行なった。

- Current status and future of MT in your country
- How is MT adopted by human translators in your country
- What should the MT researchers do for practical MT
- NMT vs. LLM MT for human translators
- Future of MT
- etc.

パネルディスカッションの中では「NMT/LLM の技術の利用はどんどん広がっているが、適切に利用しないと良い結果は得られない」「NMT の長所は一貫した出力が出せること、ドメイン適応に優れていること」「LLM の長所はプロンプトによる高いカスタマイズ性、長い文脈の理解、より流暢な翻訳」「LLM の短所は速度が遅い、費用が高い、英語以外の言語における精度が低い」「今後必要なことは多言語化、マルチモー

ダル、データ不足の解消、セキュリティーなど」といった意見が出た。また、タイではそもそも自然言語処理の研究者が少ないといった問題点も挙げられた。

また、今後も定期的にこのような情報交換の場を設け、アジア全体での機械翻訳コミュニティの活性化を目指し、WATがそのプラットフォームとしての役割を果たすべきであるという結論となった。

ワークショップ全体の録画は EMNLP に参加登録を行った方であれば、以下の URL から視聴できる。

<https://underline.io/events/470/sessions?eventSessionId=18628&searchGroup=lecture&searchGroup=lecture>

4. 翻訳タスク

WAT2024 は世界最大の機械翻訳のワークショップである WMT2024 と同じ会場で開催していた。そのため参加者の利便性を図るため、今回に限り、翻訳タスクは WMT の翻訳タスクとして実施した。実施したタスクは以下の 4 つである。

- MultiIndic22MT [2]
- English-to-Lowres Multi-Modal Translation [3]
- Non-Repetitive Translation [4]
- Patent Translation [5]

どのタスクにも数チームの参加があった。それぞれのタスクの結果は、それぞれの Findings 論文にまとめられているので、興味のある方は参考にいただきたい。

5. まとめ

本稿では WAT2024 全体の概要を報告した。残念ながら WAT2024 では当日の参加者が少なかったが、論文の投稿数や翻訳タスクの参加者は一定数あり、またパネルディスカッションも有意義であったことから、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。2025 年度は 12 月にインドのムンバイで開催される AACL-IJCNLP2025 のワークショップとしての開催を計画しているが、未定である。WAT2025 では新たに日英の特許翻訳、特に請求項に特化した翻訳のタスクを実施予定である。翻訳精度の向上もちろんだが、評価の部分に力を入れる予定である。WAT では評価を行うための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。

参考文献

- [1] Toshiaki Nakazawa and Isao Goto. 2024. Proceedings of the Eleventh Workshop on Asian Translation (WAT 2024). Association for Computational Linguistics, Miami, Florida, USA.
- [2] Raj Dabre and Anoop Kunchukuttan. 2024. Findings of WMT 2024's MultiIndic22MT Shared Task for Machine Translation of 22 Indian Languages. In Proceedings of the Ninth Conference on Machine Translation, pages 669–676, Miami, Florida, USA. Association for Computational Linguistics.
- [3] Shantipriya Parida, Ondřej Bojar, Idris Abdulmumin, Shamsuddeen Hassan Muhammad, and Ibrahim Said Ahmad. 2024. Findings of WMT2024 English-to-Low Resource Multimodal Translation Task. In Proceedings of the Ninth Conference on Machine Translation, pages 677–683, Miami, Florida, USA. Association for Computational Linguistics.
- [4] Kazutaka Kinugawa, Hideya Mino, Isao Goto, and Naoto Shirai. 2024. Findings of the WMT 2024 Shared Task on Non-Repetitive Translation. In Proceedings of the Ninth Conference on Machine Translation, pages 715–727, Miami, Florida, USA. Association for Computational Linguistics.
- [5] Shohei Higashiyama. 2024. Results of the WAT/WMT 2024 Shared Task on Patent Translation. In Proceedings of the Ninth Conference on Machine Translation, pages 118–123, Miami, Florida, USA. Association for Computational Linguistics.

AAMT 2024, Tokyo

～機械翻訳の進歩と調和～

出内 将夫
情報通信研究機構

1. はじめに

2024年12月3日に開催されたAAMT 2024, Tokyoのイベント報告をします。AAMT 2024, Tokyoはオンラインと会場のハイブリッド形式で開催され、招待講演3件、公募ポスター3件、パネルディスカッションのほか、ランチョンセミナーや企業展示も行われました。今回は、公募ポスターと企業展示が同じ会場で開催され、説明者と参加者の間で密にコミュニケーションが取られ、非常に盛況でした。(図1) 講演/発表資料は、<https://aamt.info/aamttokyo2024/handout-20241203/>からアクセスできますので、ぜひご参照ください。

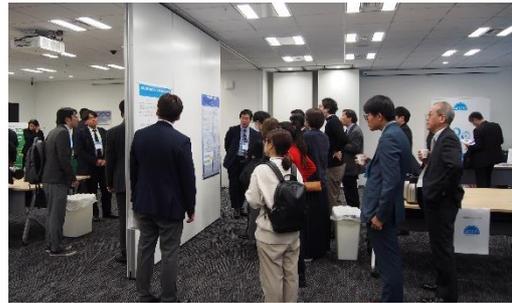


図1: 公募ポスター・企業展示会場の様子

2. 招待講演

招待講演として、以下の3つの講演がありました。

◎機械翻訳はどこまでも可能か？

隅田 英一郎 氏 (国立研究開発法人情報通信研究機構フェロー)

◎国産 LLM の現在と未来

小田 悠介 氏 (国立情報学研究所 特任准教授)

◎自治体情報の多言語化を支える翻訳資源とテクノロジー

宮田 玲 氏 (東京大学 講師)

下記のまとめは、講演資料と発表内容より、報告者が主観的にまとめたものであることにご注意ください。

◎「機械翻訳はどこまでも可能か?」: 隅田 英一郎 氏

40年ほど前に機械翻訳(MT)の可能性と限界をテーマとした「機械翻訳はどこまで可能か」という本が出版された。この本を参照しつつ、今までのMTを振り返り、MTの今後について述べる。

冒頭で紹介した書籍に、MTの必要性について「翻訳に対する需要は急速に増大し、翻訳者の数が全く不足する」とある。翻訳需要が指数関数的に増えると仮定し5年で2倍になるとすると、現在は1986年の256倍の需要があり、MTは依然必要な状況である。

文書翻訳において、MTはコーパスベースの技術と大量データにより大きく進歩した。英文ニュースはMTのみでも自然な日本語で読め、金融庁や法務省においても本格的にMTが利用されている。法務省では日本の法律をタイムリーに公開するため、WebサイトでMT結果を暫定公開した後、人手で翻訳して正式公開する方式を採用した。またNICTでは、日次でarXivやPubMedに掲載された論文のMT結果を公開¹し、日本の研究者が最新の科学技術論文や医療系論文を日本語で読めるようにした。

MTによる同時通訳も商用可能なレベルになった。AI音声認識技術は人間の精度を超え、音声合成も機械

¹ <https://mt-auto-minhon-mlt.ucri.jgn-x.jp/content/tool/search/>

と人間の違いを区別できないレベルになった。ここま
で説明した技術を組み合わせれば文単位の逐次通訳が
できる。同時通訳の MT では、同時通訳者による分割
を模倣した翻訳単位であるチャンクごとに MT を行う
ことで、話者が話している途中で翻訳が開始できる。
これにより低遅延と高精度を両立した MT を実現した。

MT は語学にも影響を及ぼしている。小学校の語学
教育で、先進的な先生はすでに MT を導入している。
生徒が外国人との会話の楽しさを実感することで、語
学学習の動機付けとなっている。また、プロ翻訳者の
MT ポストエディット (MTPE) の大規模実験では、
MTPE は専門外領域の勉強になった、という趣旨のコ
メントが得られ、プロ翻訳者の語学学習にも役立つこ
とが分かった。

冒頭で述べた書籍に、MT の限界として、MT が文
単位の処理であるため文間の関係を考慮できないとい
う課題が挙げられていた。翻訳では、数量、視覚、背
景知識などのコンテキストが無いと、適切に翻訳でき
ない事例が存在する。このコンテキストの課題は、生
成 AI の登場により解決されつつある。生成 AI を活用
することにより、複数文や語の関係性を考慮した MT
や、画像情報を活用した漫画の MT が可能になった。
実際に生成 AI を用いて漫画や小説の試訳を行ったと
ころ、従来の MT に比べ改善が見られた。多言語 MT
においては、少量のデータしかない言語の MT も高精
度化しつつあり、経済的成長が著しいグローバルサウ
スとの多言語コミュニケーション活性化が期待できる。

生成 AI のメリットを活用してハイコンテキストな
翻訳も可能となりつつあるため「機械翻訳はどこまで
も可能」と言えるのではないか。

◎「国産 LLM の現在と未来」：小田 悠介 氏

オープンかつ日本語に強い大規模言語モデル (LLM)
の開発について、NII (国立情報学研究所) と LLM-jp
(LLM 勉強会) の活動を中心に紹介する。LLM では
MT の研究で発展した Attention や Transformer の技
術が使われており、MT と関係が深い。LLM への懸
念として、一部組織が寡占していることや、中身が不

明であること、英語中心のモデルは日本語性能が低い
こと、海外 LLM 利用時の法的懸念などがあり、国産
LLM の開発を進めている。2023 年春頃に、自然言語
処理の勉強会として始まった LLM-jp が規模を拡大し、
現在は国立情報学研究所に LLM 研究開発センターが
設置されるに至った。LLM の原理解明を目的とし、
LLM に関する議論過程や失敗を含め、すべて公開する
方針で活動が続いている。LLM の研究開発はビッグサ
イエンスであるため、組織横断的な参加者の貢献や計
算資源を活用しながら活動している。

現時点で LLM-jp から公開している最新モデルは、
LLM-jp-3 であり、そのうち最も高性能な LLM-jp3-
172B-beta2-instruct を llm-jp-eval 1.4.1 で評価する
と、平均スコアは GPT-3.5 を超え GPT-4 に近づいて
いる。LLM-jp3-172B の最終版は年末の公開を目指し
ており、研究用に小さめのモデルも公開している。

事前学習に用いる LLM-jp コーパス構築では、公開
レベルに応じて分類を行い、現時点では日本語 6,000
億トークンを収集済である。今後は日本語 3 兆トー
クンを目標にコーパスを開拓する予定である。

モデルの事前学習では、実際にモデル開発を経験し
ないと分からない知見が得られている。小規模モデル
では問題にならない最適化関連のパラメタ設定や、ロ
ススパイクの発生、低精度計算の影響、平均化モデル
の効果など、大規模モデルに特有の課題・現象に関す
る知見が得られた。そういった知見は、LLM 事業者か
らは基本的に情報公開されないため、公共の活動とし
て行う重要性を感じている。

LLM の分析では、LLM の学習経過であるチェック
ポイントごと、評価セットの事例ごとの評価結果を
DB に残すことで、きめ細やかな分析を可能としてい
る。事例レベルの学習曲線の傾向分析や、訓練データ
を暗記する現象の分析、LLM の回答の根拠と思わしき
文書を特定する事前学習コーパス検索を可能としてお
り、LLM が知識を獲得する過程の分析が行える。

他にも、インストラクションチューニング用のデー
タセット構築や、LLM の安全性向上、画像理解などの

取り組み、音声や映像などへの対応も進行中である。興味がある方は、是非 LLJ-jp に参加されたい。

◎「自治体情報の多言語化を支える翻訳資源とテクノロジー」：宮田 玲 氏

日本の自治体では、在留外国人の増加により、翻訳への需要が高まっている。2019年に行った調査によると、自治体 Web サイトの多言語化に MT のみを導入している自治体は約 2/3 で、人手翻訳まで行えていない状況が明らかになった。MT の誤訳が引き起こし得る問題は度々指摘されており、改善が求められている。その問題を解決する一つの手段として、自治体の翻訳プロセスを技術的に支援する研究を進めている。

「円滑な多言語情報発信を可能にする自治体横断型翻訳資源の構築」という研究プロジェクトで、自治体文書の翻訳資源を構築している。自治体文書は法令に基づくため、過去の文書や他の自治体の文書と似ている、という特徴を持つが、自治体ごとに表現や用語の揺れがある。自治体間の共通資産として翻訳資源をまとめ、共有・活用することで、翻訳に関する自治体の課題を解決できると考える。翻訳資源は、対訳文書（文書単位）、対訳コーパス（文単位）、対訳用語集（語単位）という 3 つのレベルで管理している。まず、翻訳文書アーカイブから文書単位の対訳関係を特定し、対訳文書から対訳コーパスと対訳用語集を獲得する。対訳コーパス構築では、連絡先など自治体に依存した表現を変数化し、対訳用語集構築では、用語の省略形と正式名称などの情報を付与する。翻訳資源は MT だけでなく生成 AI の質問回答などにも役立つと考える。活動理念として、公共利益の追求、必要な翻訳は実施するというサービス改善・拡充の視点、持続可能なエコシステムの構築を重視している。

研究活動として、基礎調査、翻訳資源構築、翻訳資源利用の 3 つを進めている。基礎調査では、自治体における翻訳業務の課題と翻訳ニーズ洗い出しのアンケート調査を進めている。翻訳資源構築では、名古屋市を始めとする複数自治体から翻訳資源を受領し、整備後に github で公開している。公開前に、翻訳時に参

照されうる情報を文書属性として付与している。他に、ユニバーサルメニュー普及協会が公開している自治体文書テンプレートの翻訳も行っている。翻訳資源利用では、MTPE 研究や翻訳品質推定に役立てたほか、企業による翻訳資源活用を推進している。民間企業が令和 6 年能登半島地震支援情報ナビを MT で 10 言語に翻訳した事例からは、公開済の翻訳資源活用による早期 MT カスタマイズを実現した一方で、翻訳資源が未カバーの分野における翻訳品質が課題となった。

今後も、自治体横断で利用可能な翻訳資源の構築を続け、翻訳テクノロジーへの応用を検討する予定である。また、学术界、政府・自治体、産業界の間で、翻訳資源構築・活用のエコシステム構築を目指す。

3. 公募ポスター

公募ポスター発表は以下の 3 件でした。

◎機械翻訳の品質を迅速かつ包括的に評価する新手法 CATER (Comprehensive AI-assisted Translation Error Rating) の提案：飯田 蔵土 氏

◎医薬分野個人翻訳者による、医薬分野 AI 翻訳サイトの開発：山本 隆之 氏

◎生成 AI を活用した用語抽出：伊澤 力 氏

下記のまとめは、資料と発表内容より、報告者が主観的にまとめたものであることにご注意ください。

◎「機械翻訳の品質を迅速かつ包括的に評価する新手法 CATER (Comprehensive AI-assisted Translation Error Rating) の提案」：飯田 蔵土 氏

LLM を活用し、参照訳不要で原文と訳文を包括的に評価する手法である CATER(Comprehensive AI-base Translation Error Rating)を提案する。包括的な評価軸として、言語的正確性、意味の正確性、文脈的整合性、スタイルの適切性、情報の完全性の 5 つを用いた。LLM に与えるプロンプトには、評価軸に沿って評価させる指示と、評価対象の原文と訳文を与え、出力内容の指定として、評価軸ごとのスコアと編集率、全体スコア、エラー箇所、エラー理由を与えた。川端康成の

「雪国」を、実際に MT で翻訳した結果を、GPT4o で評価した。

LLM のモデルや設定によるスコア変動が課題である。評価の安定性向上に向けて Temperature 等のパラメータ調整や、複数 LLM の利用、複数回評価の平均取得などが有効だと考えている。

◎「医薬分野個人翻訳者による、医薬分野 AI 翻訳サイトの開発」：山本 隆之 氏

医薬分野の AI 翻訳サイトを個人で自作する試みを行った。プログラミング経験無しの状態から 3 年間かけて AI 翻訳サイトを構築した。ChatGPT は、翻訳結果の後編集、プログラミング支援の双方で有用だった。医学や治験に関する対訳を、スクレイピングで収集し、14,000 超の対訳文を用いて Google AutoML Translation を追加訓練した。また、MT 結果に対して、ChatGPT による文体変換や平易化の後編集を行った。現在は、PubMed や ClinicalTrials.gov に API 経由でアクセスし、それらの記事の MT 結果を参照できる。

一連の開発を行った結果の所感として、AI 翻訳応用サービスの構築が容易になり、分野特化型 MT エンジンがコモディティ化し、音声や動画などの異なるモダリティでの MT 利用が増加していると感じた。

◎「生成 AI を活用した用語抽出」：伊澤 力 氏

gpt-3.5-turbo を利用して、日英対訳の化学分野特許明細書から、化学用語を抽出した。対訳コーパスとして、複数の特許明細書から自動でアライメントを取得した対訳文を利用した。用語抽出と対訳抽出を同時に行うと精度が低かったため、日本語原文からの用語抽出、抽出した用語と英語訳文からの対訳抽出、という 2 つのタスクに分けた。また、gpt-3.5-turbo のベースモデルでは、ノイズや誤った語が出力されたため、タスクごとに 100 文を用いて gpt-3.5-turbo をファインチューニングした。その結果、10,000 文の対訳文から、重複を除いて 2,236 件の用語が抽出できた。

一連の実験から、過去の対訳文書からの用語抽出に、生成 AI が有効であることが確認できた。また、単純

なタスクへの分割やモデルのファインチューニングが精度向上に有効だった。

4. パネルディスカッション

パネルディスカッションでは、「企業における MT 活用と生成 AI の可能性」をテーマに、事例紹介を含むパネル発表の後、議論が行われました。パネル発表とパネリストは以下のとおりです。

◎生成 AI 時代における、エンタメ翻訳領域での活用事例と課題

関野 遼平 氏 (Mantra 株式会社 ToB 事業責任者、
米国公認会計士)

◎メーカーにおける機械翻訳と翻訳支援ツールの活用事例と AI 活用の可能性

染谷 隆一 氏 (オリエンタルモーター株式会社 オリ
ム・エンジニアリング・カンパニー 販促支援部 プ
ランナー)

◎「生成 AI 時代における、エンタメ翻訳領域での活用事例と課題」：関野 遼平 氏

Mantra 株式会社は、マンガ翻訳 (Manga Translation) を縮めた社名であり、マンガに特化した MT 技術の研究開発を事業としている。マンガ用 AI 翻訳ツールとマンガ・コミック・小説の翻訳サービスを紹介する。

マンガの市場は世界で急成長しており、最近 4 年間に北米で売上 5 倍、フランスで売上 2 倍以上となった。一方で、翻訳出版されるマンガは 10~20%前後と全体のごく一部にとどまっており、出版時期が数か月から数年遅れとなっている。出版時期に遅れがあると、ファンが翻訳して海外海賊版が出回るため、海外同時出版が望ましい。一般的なマンガ翻訳では、翻訳者、チェッカー、デザイナーの 3 者が、それぞれ異なる形式でデータをやり取りし、人手作業が多く煩雑なデータやり取りが発生していた。Mantra が提供するマンガ AI 翻訳ツールでは、翻訳・文字組み・校正を一元的に管理することで、作業を効率化し、マンガの海外版制作まで

の時間をおよそ半分に圧縮し、海外海賊版の取下に貢献した。また viviON が運営しているファン自身がマンガを翻訳できるサービスにもマンガ AI 翻訳ツールを提供している。

マンガ処理 AI 技術は、テキスト抽出、翻訳、画像生成の3段階に分かれる。このうち翻訳について説明する。翻訳では、意味が伝わるだけでなく「面白い」翻訳が必要で、翻訳もその評価も主観が入るため難しい。翻訳の評価では、人手評価と機械評価を併用している。機械評価手法の COMET を用いた実験では、人手翻訳と MT による翻訳の違いを区別できなかった。そのため、まず「良い翻訳」の要件の言語化から始めた。MQM (Multidimensional Quality Metrics) に基づいてマンガの人手評価を設計し、実際に人手評価を行った。人手評価のうち、Critical と判定されたのは正確性に関するエラーで、最も多いエラーは不自然な用語や文構造だった。この人手評価の結果を基に LLM で自動評価したところ、人手翻訳と MT を区別できたが、「面白い」翻訳の判定にはまだ至っていない。また、マンガ翻訳では多種の文脈情報が重要なため、長い文脈、翻訳ガイドライン、画像情報などを扱える、LLM を活用した MT (LLM-MT) を開発した。

LLM によりマンガ MT は大きく進展したが、信頼できる自動評価や「面白い」翻訳など、課題も多い。今後もマンガ MT でマンガの世界展開を支えていきたい。

◎「メーカーにおける機械翻訳と翻訳支援ツールの活用事例と AI 活用の可能性」: 染谷 隆一 氏

メーカーにおける MT や翻訳支援 (CAT) ツールを活用した多言語化業務を紹介する。発表者が所属する販促支援部では、技術部門と営業部門双方のドキュメント多言語化を行っている。技術部門では取扱説明書や外形図・特性図といった技術情報、Web サイトが、営業部門では製品カタログが主な多言語化対象となる。新商品や設計変更への対応、翻訳言語数増加などにより、品質を保ったままの作業効率化が求められている。メーカーの場合、安全性に関する規格に関する文は決

まった文言が使われるほか、文章や項目の流用性が高いため、その特徴を活かした効率化を行っている。

多言語化工程である、前処理、翻訳工程、MT での逆翻訳、原文との整合性確認の各工程について説明する。前処理では、原文編集やデータクリーニング、例えば、テキスト分割や統合を行うことで、翻訳メモリ (TM) にマッチしやすくする。翻訳工程では、CAT ツールで TM 一致率の調整や MT の適用条件を設定し、TM 不一致部分や MT 結果を対象に、用語集に基づいた修正を行う。MT での逆翻訳と原文との整合性確認では、複数の MT エンジンを用いた翻訳結果の逆翻訳結果を原文と比較し、余分な単語や訳抜けなどを修正する。工程終了時に TM や用語集を更新し、次工程に備える。

生成 AI の活用例を紹介する。技術部門からの原文の一部では、長文、複数内容を含む文、読点が無い等の理由で、理解しにくいことがある。前処理の工程で、生成 AI を活用して、このような原文の改善ができることが分かった。また、翻訳のサポートとして、2 つの MT 結果が大きく違う場合に、生成 AI を活用してどちらが適切か判定理由とともに出力することが有用だと分かった。現状における生成 AI 活用方針として、生成 AI の出力には誤りが含まれることから、生成 AI は前処理や翻訳の支援に使用し、過去に実績がある文の翻訳には、TM の継続利用が良いと考えている。

◎質疑とディスカッション

会場参加者：

マンガ翻訳の評価に用いた LLM や COMET はマンガ用にチューニングしたものか。

関野氏：

チューニングはしていない。マンガとは別に小説の翻訳を行っており、そちらはチューニングを行う予定。

会場参加者：

マンガ MT で紹介されていた誤訳は、ニューラル MT (NMT) によるものか、LLM-MT によるものか。

関野氏：

資料は LLM-MT の誤訳を集計したもので、NMT では主語誤りがさらに多く見られた。

会場参加者：

メーカーの翻訳で、外注を検討したことはあるか。

染谷氏：

営業部門で広告宣伝文書の翻訳を担当した際は、クリエイティブな要素が求められるため外注した。その後、技術部門の翻訳を担当した時に、文書の構造などに着目して、社内で翻訳対応する文書と翻訳工程を決めた。技術者と直接やり取りできるのが、社内で翻訳対応するメリットだと考えている。

会場参加者：

用語集は1つとのことだが TM も1つなのか。

染谷氏：

マニュアルやカタログで文章の構造や特徴が異なるため、TM は分けて管理している。

森口氏：

新たに対応する言語を、どのように決めているか。

染谷氏：

マニュアルは、日本語を原文として英語、中国語の簡体字と繁体字、韓国語への翻訳を基本としている。営業の販促系資料は、現地法人の有無に合わせ、タイ語やベトナム語、欧州系の言語を扱う。欧州系の言語は英語を中心に MT を活用してローカライズしている。

関野氏：

お客様からのリクエストに応じて言語を決める。マンガ翻訳は基本的に MTPE を行うため、ポストエディターを確保できた時点で対応、という形を採っている。

森口氏：

どこまでの品質を求めるのか。誤訳の許容水準や、誤訳が許されない部分はあるか。また、市場に出てから翻訳を修正することはあるか。

関野氏：

マンガ翻訳に対する要求品質は、翻訳を行う顧客とその先の読者に大きく依存する。作家個人が顧客となる場合、作家個人が許容すれば良く、MT の品質のままで受け入れられることが多い。出版社が顧客となる場合、要求レベルは高くなり、MTPE が必須。翻訳の修正について、作家自ら翻訳版を用意した場合、読者からの指摘で作家が修正することがある。出版社が翻訳版を用意する場合、後で修正することは無い。

染谷氏：

メーカーで一番間違えていけないのは、仕様や寸法などの数値である。数値は TM の翻訳対象外であり、1 バイトの場合は、TM 適用時に自動で置換される。しかし、数値が 2 バイトで記述されることもあるので、人手や自動のチェックを組み合わせで確認している。操作手順の翻訳誤りを指摘され、修正したことがある。

森口氏：

NMT と LLM-MT の使い分けはあるか。

染谷氏：

商品のマニュアルでは、TM の適用率が高く、NMT も LLM-MT も大きくは変わらない。どちらを使う場合も、正答率が 100% でなければ、全部人手チェックが必要。現状では、生成 AI を前編集と MT 結果の選別に用いる方向で考えている。

関野氏：

マンガ翻訳では、NMT は速度重視の場合に使用し、LLM-MT は時間がかかっても良い翻訳が欲しい、という場合に使用している。

森口氏：

LLM を評価に使う話があったが、翻訳や修正をさせるより評価をした方が良い、という判断なのか。

染谷氏：

発表で紹介した LLM による前編集でもポストエディットが必要だった。NMT から LLM-MT に置き換えるのはもう少し先と考えている。NMT は時期に

よって訳が変わるので、LLM で 2 つの NMT を評価した。

関野氏：

MT の品質は、LLM-MT で相当良くなった。修正は人が行う、という点は NMT も LLM-MT も変わらない。評価については、人手評価が数量的にボトルネックになるので、LLM による評価を積極的に活用している。

森口氏：

メーカーの文書では TM や用語集が有効とのことだが、マンガ翻訳にそういう要素はあるのか。

関野氏：

キャラ名や技名などは、1 巻と 2 巻で訳語が変わってはいけなくて、用語集を整備しており、過去の訳を参照する仕組みがある。

会場参加者：

エンターテインメント翻訳は、いろいろな会社が生成 AI を活用して参入しているが、どう差別化するのか。

関野氏：

生成 AI の登場以降、他社も似たツールが作れる状況である。今までの翻訳実績を活用して良い翻訳ができる、という点を大事にしている。マンガ翻訳業者の声をもとに、ツールの使い勝手も改善していきたい。

オンライン参加者：

染谷氏の発表で、LLM を用いた場合でも一文ずつ処理していたが、それでも効率は上がるのか。

染谷氏：

逆翻訳の処理は一括で行えるようにツール化している。産業翻訳では流用率が高く、すべての文に MT を使っておらず、新商品の説明などが MT の対象になる。MT の翻訳量は多くなく、効率の問題は起きていない。

会場参加者：

翻訳結果を使うユーザーの想定は明確か。

染谷氏：

マニュアルは対象者が明確で、ペルソナが設定しやすい。例えば、安全性に関する記事は全員、配線であれば電気的な知識が前提、など。MT を使いやすい部分、使ってはいけない部分を分けて対応している。

関野氏：

翻訳先言語のネイティブを想定して翻訳を行っている。マンガ翻訳の場合、翻訳が合っているかより、ネイティブにとって面白いか、という点を重視している。

会場参加者：

小説の翻訳で、1 冊の情報を生成 AI に入力する、とあったが、マンガ翻訳では実施していないのか。

関野氏：

まだマンガ翻訳では実施しておらず、小説の翻訳から実施している。小説はテキスト量が膨大なので、text-to-text の生成 AI の研究材料として良いと考えている。マンガ翻訳では、まず画像から表情の情報を取得する研究を進めている。マンガ翻訳で 1 冊の情報を生成 AI に入力するのはまだ先だと考えている。

会場参加者：

TM の出力を LLM-MT の入力とする実験はしたか。

染谷氏：

TM から翻訳対象に一番マッチする文が引き出せるよう TM を整理できれば、翻訳もチェックもやりやすい。まずは生成 AI で TM を整理したいと考えている。

会場参加者：

TM の出力がきれいでもなくとも、それを受け付けられる LLM を用いる方法もあるのではないか。

5. おわりに

生成 AI が翻訳や翻訳前後で実用される事例や、他サービスに MT が組み込まれる事例の増加、MT の価値や評価基準の多様化が伺えた、大変興味深い年次大会となりました。今後の年次大会にもご期待ください。

AMTA 2024 参加報告：社会学の視点から見る AI 翻訳

郭 文静

同志社大学

1. 生成 AI と機械翻訳

近年、生成 AI (Generative Artificial Intelligence) と機械翻訳 (Machine Translation: MT) の組み合わせは社会の視野に入ってきた。AMTA 2024 会議の中に、AI が効率的なツールとして、機械翻訳の内容を評価し、品質を保証するという役割を期待され、生産性を向上させるように使われている内容が多く発表されてきた。たとえば、Aneta Sapeta が「Is AI the new “Human evaluator” ?」報告の中で、AI は補助ツールとして機械翻訳した内容の質、データの良さ、ポストエディット (Post Edit, 以下 PE) の状況を判断すると期待されている。その結果、AI を利用すると、MT の内容を評価する時間が短縮になるメリットがある。しかし、Aneta Sapeta が AI の限界はいくつかあると指摘し、AI は 100% 指示通り評価しなく、同じの文書に対して異なる評価結果を出した場合もある。また、AI は機械が翻訳されていない部分を認識できず、特に、1 つの問題に焦点を当て、他の問題点を見逃しており、全体的に高い評価を付ける傾向がある。以上の結果から見ると、AI の判断の信頼性が低く、実際の問題を見逃す潜在的なリスクが存在するとわかった。AI の開発はまだ時間とコストが必要である。特に AI が出した結果や判断の根拠を説明することが難しく、ブラックボックス化の問題が存在する。また、AI の活用に伴う著作権侵害のリスクやセキュリティ上の倫理的問題が顕在化しているため、AI 技術の適正な運用方針と法制度の強化が喫緊の課題であろう。したがって、現在、AI は完全に人間を代替する力がなく、補助として人間の作業負担の軽減と業務効率の向上のために非常に期待されている。

2. 翻訳者としての職業の変動期

日本の雇用制度の下で、AI の導入により、専門職においても正規雇用と非正規雇用の二極化が大きくなる可能性がある (藤本 2024)。正規雇用者は雇用が保障されているが、フリーランスの翻訳者は AI に仕事を代替されるという社会問題をもたらす懸念も少なからずある。この問題に対して、AMTA 会長パネルのセッションでは、翻訳内容の 1% でも人力が必要であれば、人的翻訳はなくなるという。その上、プロ翻訳者は翻訳の精度や専門性の確保が求められる。また、人的翻訳や機械翻訳以外に、PE の業務が発生している。Grinbaum & Adomaitis (2022) は人間と機械の区分を曖昧しないと指摘している。それに関して、Michel Simard は「Should Machine Translation be Labelled as AI-Generated Content?」の報告の中で、機械が訳した内容を人間でチェックして修正したものは人間の成果物と認めるべきだと述べている。しかし、翻訳業界では、ポストエディターの役割をはっきり取り扱われなくなり、PE の単価が低く、やりがいを感じにくい問題がある。加えて、職人気質を持つプロ翻訳者が PE 業務に抵抗感があり、専門性の向上や仕事のモチベーションが下がる傾向が見られる (郭 2025)。このように、業務効率化とコスト削減を目的とした AI 翻訳の普及は翻訳業界の不可避な趨勢である一方、AI と人間翻訳者の共存にはジレンマが存在しているといえる。

AI 翻訳の普及は翻訳業界に大きく影響をもたらし、特に翻訳者の職業は過渡期として徐々に変化しつつあり、その変化の記録を正確に取るべきである。これから AI は社会に実装する際に、利便性や効率化、コストの削減を求める同時に、翻訳者の働き方、仕事の質や働きがいについてさらに深く理解すべきであろう。

第8回自動翻訳シンポジウム 参加報告

玉咲 知香子

国立研究開発法人情報通信研究機構

1. はじめに

2025年2月19日、第8回自動翻訳シンポジウム「生成 AI と AI 翻訳～自治体での活用～」が品川インターシティホールにて開催されました。

講演会では、岡崎 直観氏（東京科学大学 情報理工学院 教授）による基調講演「大規模言語モデルはどのように『ことばの壁』を超えるのか」に続いて、自治体における多言語通訳・翻訳の活用に関連して、鶴田 知佳子氏（東京外国語大学 名誉教授）の「自治体における多言語翻訳技術の活用」、小笠原 堂裕氏（ユニカミノルタ株式会社 DW-DX 事業統括部 国内 ICW 推進部）の「自治体での多言語通訳サービスの活用事例～東京都板橋区での実績～」、永野 量平氏（TOPPAN 株式会社 多言語ビジネス開発部チームリーダー）の「自治体でのインバウンド対応における自動翻訳の活用」の3つの講演が行われました。

講演会場の隣では、22企業・団体による自動翻訳製品・サービス等に関する展示があり、出展者と参加者が活発に議論を交わす姿が見られ、会場は熱気に満ちていました。

本稿では、展示に焦点をあて、実際に出展者から話を聞くことができた中から4点の展示について報告します。なお、報告内容は、展示会場にて配布された資料や出展者から伺った話の内容に基づき、報告者が主観的にまとめたもので、五十音順で紹介します。

2. 展示会場の報告

1)

株式会社 IP DREAM の「VoiceOn® LLM」が組み込まれた「多言語対応 AI コンシェルジュ」では、AI アバターである観光ガイド“葵”が、国内・インバウンド観光客向けに、自然な会話で秋葉原を案内します。21言語の多言語対応で、QR コードを自身のスマートフォンで読み取るだけで、すぐにサービスが利用可能となり、話者の言語を自動認識して21言語の中からいずれかの言語で AI アバター“葵”と会話することができます。

一般的なウェブ検索を使用すると、不確実な情報が出力されることもありますが、こちらの AI アバターは地域関係者等から収集、学習したデータを用いているため、提供される観光情報は信頼性の高いものとなっています。また、テレビ番組のデータも学習されているため、具体的な番組名を出して、「○○の番組で映っていた△△△のカフェに行きたい。」という質問にも答えられます。本シンポジウム開催時点では実証実験中でしたが、今後は他の観光地へのサービス拡大も視野に入れているそうです。

2)

株式会社エミューズインターナショナルとマインドワード株式会社による「∞L i V E」は AI による同時通訳・同時字幕サービスです。話し終わってから通訳を行う逐次通訳とは異なり、リアルタイムの通訳が可能で、イベント会場のような一方向（登壇者→観客）の通訳やオンライン会議のような双方向（参加者間）で利用することができます。スタンドアロンで動作す

る同時通訳端末には、イベント・会議の前に、業界特有の用語、社内用語や名称（商品名・作品名・キャラクター名）を事前に登録することができ、この端末を使うと登壇者が発した言語の内容を日本語の音声で聞くことが可能となります。さらに観客は自身のスマートフォンを用いて翻訳された日本語音声を聞いたり、日本語の字幕を読んだりすることも可能です。

展示担当者の方のお話では、「昨今の K-POP 等の人気の高まりにより海外アーティストが日本でコンサートやファンミーティング等の催しをする機会が増えており、今後ますます本サービスの需要が高まることが期待される。」、ということです。

3)

株式会社 CMC エクスマニコムの「ATOM KNOWLEDGE」は AI 自動翻訳システムです。AI 翻訳には、同じ文書を複数回翻訳すると必ずしも同じ翻訳にならない問題があります。これに対して本システムでは一度翻訳した内容を記録する翻訳メモリを組み込むことでこの問題を解決しています。記録は文単位に加えて用語単位で可能となっており、用語集を併用することでより精度の高い翻訳が可能となります。作業指示書等の場合、本システムの翻訳結果をそのまま使用できるケースもあるそうです。

製造業においては近年、既存の日本語資料の現地法人での活用のための翻訳、国内の非日本語話者労働者への情報提供のための翻訳の需要が増加しています。特に東南アジア言語の翻訳需要が高まっているとのことで、本システムの活用の方はさらに広がるだろうとのことでした。

4)

株式会社 聴覚研究所が開発する「TeamLog」は複数人の会話を同時にリアルタイムで文字起こしをするアプリです。コロナが大流行した時に人々がマスクをつけるようになり、聴覚に障害のある人は話者の口元の動きを読むことができず、コミュニケーションに大

きな問題を生じました。そこで、マスクをつけたままでもコミュニケーションが可能になる「TeamLog」が開発されました。

聴覚に障害のある人は一度に複数の方の口の動きを確認することが難しく、複数人と同時に会話をすることが困難です。このアプリでは複数人の音声認識結果と話者名を同時に表示することで、会話参加者全員の発言を理解することが可能になっています。

またマイクを複数使用した独自の仕組みにより発言が重なっても声が混ざりません。iOS 版のアプリでは最大 8 人まで同時に会話が可能です。

さらに、マイク毎に言語を設定することで日→英、英→日などの同時翻訳（現在は日本語・英語・中国語・韓国語に対応）ができます。文字起こしされた文書はリアルタイムで修正可能で、議事録として利用可能です。今後は、テキストの出力だけでなく、音声での出力や対応可能な言語数を増やすことを検討中だそうです。

3. 最後に

今回、動画の自動翻訳やマルチモーダル同時通訳技術などの展示が多くみられ、新たな技術が日進月歩で進化していると感じました。一方、翻訳システムの精度は今でも言語によるバラつきがあり、特に希少言語の精度が低いという課題が残っています。希少言語を話す人は観光で来日するだけでなく、日本に定住することも増えてきており翻訳精度の向上が望まれます。

また、機密性が高い情報、プライバシーの秘匿が必要な翻訳をいかに実現するか、いわゆる翻訳時のセキュリティへの配慮は大きな関心事となっており、今後の研究・開発に大きな期待が寄せられていると感じました。

編集後記

後藤功雄

AAMT 編集委員会

LLM の活用が浸透してきました。LLM はアイデア出しの壁打ちのほか、翻訳、要約、フォーマット変換、プログラムのコード生成など多岐にわたる用途に対応でき、これらをプロンプトの指示だけで柔軟に組み合わせ活用することも可能です。最近の商用 LLM では RAG (Retrieval-Augmented Generation) の仕組みを取り入れ、新しい情報に関する質問にも対応できるものもあります。東京大学の入試問題の画像を LLM に入力して生成された回答が、理 III の合格最低点を上回ったという報道もあり、LLM という優秀なアシスタントを手軽に活用できる時代になったと感じます。LLM ができることと人間 (自分) でなければできないことをうまく見極めて LLM を活用していきたいものです。ただし、使いこなすためにはノウハウも必要です。本号でも LLM の活用に関する報告があり、評価手法の提案、対訳用語の抽出、翻訳サービスの個人による構築についての報告が参考になります。

巻頭言では、中澤敏明様より AAMT セミナーの役割についてご寄稿いただきました。AAMT セミナーは、機械翻訳の発展・啓蒙の一環として、MT に関するトピックの情報発信を目的としています。機械翻訳とは直接は関係しないテーマが取り上げられることもあります。その理由は機械翻訳の新たな利用シーンの創出や、これまで見えていなかった共通点の発見など、イノベーションを考えてのことだと述べられています。

AAMT Tokyo 2024 年次大会のポスター発表からは、以下の 3 件が事例・研究として掲載されています。1 件目は Kurando IIDA 様、Kenjiro Mimura 様による、LLM を用いた新しい自動評価尺度「CATER」の提案です。5 つの評価軸に合わせたプロンプトを用い、それぞれの評価軸ごとに LLM で翻訳を評価する方法が

解説されています。2 件目は伊澤力様による、生成 AI を活用した用語抽出の研究です。従来は自動構築が困難だった対訳用語集を、プロンプトを定義して生成 AI で対訳コーパスから対訳用語を抽出させることで自動構築し、検証しています。OpenAI の gpt-3.5-turbo を用いた実験では、ファインチューニングによって大幅な改善があったと報告されています。3 件目は山本隆之様からの個人翻訳者による医療分野の AI 翻訳サイトの開発の報告です。翻訳サービスや LLM の API を活用し、医療分野の用語集を構築することで、個人による AI 翻訳サイトのシステム開発を実現しています。ChatGPT がデバッグやコード開発に有用だったことも紹介されています。

解説記事としては、以下の 2 件が掲載されています。1 件目はぬるっば様より外国語のゲームの画面表示を画像データとして OCR によってテキストに変換して翻訳するツールを解説いただいています。このツールは画面上のテキストを翻訳することができるため、ゲーム以外での用途にも応用可能であることが述べられています。2 件目は、宮田玲様、藤田篤様、阪本章子様、香川璃奈様より科学研究費助成事業の自治体翻訳資源プロジェクトについて解説いただきました。自治体文書の多言語化における課題とその解決に向けた翻訳資源の構築の取り組みが紹介されています。

第 12 回 AAMT セミナーの記事として、総務省の小倉知洋様より AI 事業者ガイドラインの概要について解説いただきました。総務省及び経済産業省が令和 6 年 4 月に策定・公表した AI 事業者ガイドライン (第 1.0 版) について、その概要と令和 6 年に行った更新情報がまとめられています。

Editor's note

Isao Goto

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/>

温故知新では、1993年9月と1993年12月のAAMTジャーナルから、2件の記事を転載しています。1件目の「翻訳の現場から」の記事では、「Aランクの翻訳者を震えさせるようなソフトが登場するまでには、まだすこし時間がかかりそうである（しかし、それがいつになるのか分らないが、そういったソフトは必ず現実のものとなるはずである）」と書かれており、約32年前に今日の状況が予想できていたかのような記載が興味深いと思いました。2件目は「利用技術ヒアリング」の記事で、機械翻訳を活用したいが導入できない理由として、当時の課題が挙げられています。

イベント報告では、以下の4件が掲載されています。中澤敏明様よりアメリカ・マイアミで開催された第11回アジア翻訳ワークショップ(WAT2024)の報告をいただきました。5件の研究発表と、パネルディスカッションが行われ、翻訳タスクは参加者の利便性を図り、WMT2024の翻訳タスクとして実施したことが説明されています。出内将夫様よりAAMTの年次大会であるAAMT 2024, Tokyoへの参加報告をいただきました。招待講演、公募ポスター、パネルディスカッション、企業展示、ランチョンセミナーが行われ、特に公募ポスターと企業展示の同会場開催により、参加者との密なコミュニケーションが取られて非常に盛況だったとのこと。郭文静様よりAMTA2024の参加報告をいただきました。生成AIを機械翻訳の評価に使う発表が多かったということですが、その課題についても指摘しています。また、翻訳者の役割の変化についても議論しています。玉咲知香子様より第8回自動翻訳シンポジウムについて報告をいただきました。観光ガイドのAIアバター、同時通訳・同時字幕サービス、翻訳メモリと組み合わせた自動翻訳、文字起こしアプリの4つの展示について、詳しくご紹介いただいています。

なお、今年6月にはMT Summit 2025がスイス・ジュネーブで開催されます。国際的で活発な交流と情報交換が行われることを楽しみにしております。

AAMTジャーナル「機械翻訳」No. 82

- 【発行日】2025年6月15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】〒160-0004
東京都新宿区四谷4-7新宿ヒロセビル5F
一般社団法人アジア太平洋機械翻訳協会 (AAMT) 事務局
【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】奥麻里
【印刷所】株式会社 プリントパック

