

ISSN 1883-1818

No.84

June 2026

AAMT Journal

Asia-Pacific Association for Machine Translation

機械翻訳

棧載物語



### 目次

<b>巻頭言</b>		
機械翻訳の進化と共に歩む―「AAMTセミナー」のこれまでとこれから	石岡 映子	3
<b>事例・研究</b>		
機械翻訳出力に対するLLMを用いた後編集	早川 威士	5
ポストエディットが翻訳者の仕事満足度に与える影響	Akiko Sakamoto, Darren Van Laar, Joss Moorkens, Félix do Carmo	11
大規模言語モデルは多様な翻訳仕様に追従できるか?	萱野 陽子、菅原 朔	16
<b>温故知新</b>		
温故知新	内山 将夫	21
<b>イベント報告</b>		
第12回アジア翻訳ワークショップ(WAT2025)開催報告	中澤 敏明	31
AAMT 2025, Tokyo～LLM時代における翻訳と人間の共進化～	出内 将夫	33
第9回自動翻訳シンポジウム 参加報告	玉咲 知香子	40
<b>法人会員PR</b>		
AI校正とAI翻訳で進化する翻訳業務	中山 雄貴	43
化合物表記翻訳技術を活かした機械翻訳サービス JAICI AutoTrans	河内 愛	45
特許文書特化の機械翻訳	一般財団法人日本特許情報機構	47
<b>編集後記</b>	新田 順也	49

### C O N T E N T S

<b>Foreword</b>		
Walking with the Evolution of Machine Translation: The Past and Future of the AAMT Seminars	Eiko Ishioka	3
<b>Case Study</b>		
Post-editing of Machine Translation Output Using Large Language Models	Takeshi Hayakawa	5
The impact of post-editing on translators' quality of working lives	Akiko Sakamoto, Darren Van Laar, Joss Moorkens, Félix do Carmo	11
Can Large Language Models Follow Different Translation Specifications?	Yoko Kayano, Saku Sugawara	16
<b>Learning from the past</b>		
Learning from the past	Masao Utiyama	21
<b>Event Report</b>		
Report of the 12th Workshop on Asian Translation (WAT2025)	Toshiaki Nakazawa	31
AAMT 2025, Tokyo --- Coevolution of Translation and Human in the LLM Era ---	Masao Ideuchi	33
Report on the 9th Symposium on Automated Translation	Chikako Tamasaki	40
<b>Corporate PR</b>		
The Evolution of Translation: AI Proofreading and AI Translation	Takeyoshi Nakayama	43
JAICI AutoTrans: A Machine Translation Service Leveraging Chemical Compound	Ai KAWACHI	45
Name Translation Technology		
Machine Translation Specialized for Patent Documents	Japan Patent Information Organization (Japio)	47
<b>Editor's Note</b>	Junya Nitta	49

## 巻頭言

## 機械翻訳の進化と共に歩む―「AAMT セミナー」のこれまでとこれから

石岡 映子

株式会社アスカコーポレーション

## 1. はじめに

会員の皆さま、いつも AAMT の活動にご協力いただきありがとうございます。

私は 2022 年に一般社団法人アジア太平洋機械翻訳協会 (AAMT) の理事に就任し、セミナー委員長を拝命しました。機械翻訳 (MT) 技術の発展・啓発の一環として、MT に関するトピックを継続的に発信する定期セミナーを企画・運営するわけですが、有料開催という責任の重さに、身の引き締まる思いでスタートしたことを覚えています。

## 2. AAMT セミナーのこれまで

セミナーを始める頃、MT を取り巻く環境は大きく変わり、長年の研究の蓄積が一気に応用段階へと移行し、現場で「使える」技術が次々と登場し始めた時期でもありました。この素晴らしい技術を一部の専門家だけのものにせず、翻訳者、翻訳会社、企業、研究者など、できるだけ多くの方に届けることで、MT の社会的価値の底上げに貢献したい—そんな思いでセミナー運営に臨み、研究とビジネス経験を持つ心強い理事 2 名とともに、年間 5 回程度のペースで企画・開催してきました。

2023 年、第 1 回を長尾賞受賞者発表と組み合わせ、第 2 回では MT を活用した論文作成術、第 3 回では英文 IR 作成術と、研究・実務の双方に役立つテーマを取り上げました。2 年目の初回にはポストエディット (PE) に関する翻訳者アンケート結果を共有し、第 2 回では大規模言語モデル (LLM) の登場を受け、アジア・ヨー

ロッパの AI 市場分析や生成 AI 利用の法的留意点など、まさに時代の変化を映す内容となりました。

2024 年度には、MTPE 導入事例、機械学習を用いた音声解析の研究事例、NMT と LLM の活用、政府関係者による AI 事業者ガイドラインの紹介など、多様な講師、視点からの講演が実現しました。

2025 年には、国産 LLM の研究成果報告や、MT・LLM の精度向上が翻訳業界にもたらすインパクトについて、経済的観点からお話いただく機会もありました。PE ガイドラインの実践的な解説に加え、マンガ翻訳への LLM 活用、翻訳に特化した LLM 開発モデルの紹介など、毎回、「今、現場で知りたいテーマ」を意識して企画してきたつもりです。

講師を招いたセミナーだけでなく、次世代の育成にも力を入れています。その一つが、3 月に開催している「AAMT 若手翻訳研究会」です。MT の啓発・普及を促し、技術の発展を目指す取り組みとして、翻訳・通訳に関する様々なトピックの研究事例を募り演題発表を行っています。

2023 年第 1 回目は「サブセット探索を用いた高速な kNN ニューラル機械翻訳」、2024 年は「What language do Japanese-specialized large language models think in?」が最優秀賞を受賞、それぞれの素晴らしい発表内容に大きな勇気をもらいました。2025 年度は、年齢の制限を取り払い、翻訳者、通訳者、翻訳・通訳会社、学生、研究者など、より幅広い層に門戸を開き、「AAMT 翻訳通訳研究会」に改称。3 月 18 日に無事終了し、現在選考作業が進行中です。応募テーマの幅広さと熱量からも、この分野への関心の高まりを強く感じています。

Walking with the Evolution of Machine Translation: The Past and Future of the AAMT Seminars  
Eiko Ishioka

ASCA Corporation

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/>

### 3. AAMT セミナーのこれから

2026年度からは、大学、研究開発、翻訳者育成などに携わる新たな理事3名が加わりました。アカデミアとビジネス双方の視点を持つ AAMT ならではの強みをさらに活かし、より実りあるセミナーをお届けできると期待しています。

AAMT は、学会ではありませんし、業界団体でもありません。研究で生まれた技術を、現場で「使える形」に落とし込み、精度や機能を高めながら社会実装していくアカデミアとビジネスの融合体であり、その橋渡し役を担う存在だと考えています。

LLM や MT を含めた自然言語処理技術は日進月歩で進化しており、そのスピードは今後も加速するでしょう。その行き先は誰にも完全には予測できません。だからこそ、私たちは「技術そのもの」だけでなく、「それをどう使うか」「誰のために使うか」を問い続ける必要があります。

AAMT 創立者の長尾真初代会長は、「MT は世の中を平和にする」とおっしゃっていました。私たちはこの言葉を、単なる理想論ではなく、AAMT の活動の根底にある指針として受け止めています。言葉の壁を少しでも低くし、誤解や対立を減らすこと。その一步一步の積み重ねが、ビジネスの効率化や業務拡大だけでなく、より良い社会づくりにもつながっていくと信じています。

AAMT の理事として、翻訳会社の立場から、そして一人の翻訳・言語サービスに携わる人間として、MT や AI 翻訳を、共に「理解し、活かし、育てていく」立場でありたいと思っています。

これからも会員の皆さまとともに、学び合い、試行錯誤しながら、このダイナミックな変化の時代を前向きに歩んでいければ幸いです。

会員の方はこのセミナーへの参加は無料となりました。AI を活用した業務展開のためにもこのメリットを生かし、これからもご支援、ご参加よろしくお願いたします。

#### <過去のセミナー>

- 第 19 回 第 1 回 AAMT 翻訳通訳研究会
- 第 18 回 PLaMo 翻訳:LLM によって進化する機械翻訳モデル
- 第 17 回 マンガ機械翻訳の現在地と展望～LLM で解けたこと、解けていないこと～
- 第 16 回 AAMT ポストエディットガイドライン解説～成功への実践ガイド～
- 第 15 回 MT をマネタイズできるか～マーケティングフレームワークからヒントを探る～
- 第 14 回 大規模言語モデル Swallow～英語から日本語へ言葉の壁を越える～
- 第 13 回 第 2 回 AAMT 若手翻訳研究会
- 第 12 回 AI 事業者ガイドラインのご紹介
- 第 11 回 翻訳ワークフローにおける NMT と LLM の活用～LLM のインテリジェンスとそれを用いた自動ポストエディット～
- 第 10 回 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～
- 第 9 回 成功事例に学ぶ、機械翻訳および MTPE の導入方法
- 第 8 回 第 1 回 AAMT 若手翻訳研究会
- 第 7 回 生成 AI の隆盛を踏まえた機械翻訳サービスの開発・提供と利用に際しての法的留意点
- 第 6 回 アジア・ヨーロッパにおける「AI 翻訳」ビジネス市場分析及び生成 AI 翻訳の現状から
- 第 5 回 激動の AI 業界～これからのデータ構築にどう取り組む？～
- 第 4 回 ポストエディットの真実～英日 PE に従事した翻訳者のアンケート結果より～
- 第 3 回 海外投資家を引き付ける英文 IR コストや時間をかけない自動翻訳の活用～日本語のみの開示で公平・公正・適時という開示の原則に沿っていると言えるか～
- 第 2 回 理工系の英語論文執筆における課題と AI 自動翻訳ツールの活用
- 第 1 回 第 17 回 AAMT 長尾賞/第 9 回 AAMT 長尾賞学生奨励賞 受賞記念講演

## 機械翻訳出力に対する LLM を用いた後編集

早川 威士

株式会社アスカコーポレーション

### 1. はじめに

機械翻訳 (machine translation: MT) 技術の高度化にともない、ニューラル機械翻訳 (neural MT: NMT) の登場以降の翻訳品質は飛躍的に改善し、産業翻訳の現場では人手翻訳から MT 出力を人間が修正する後編集 (post-editing: PE) ワークフローへの移行が急速に進んでいる。こうした状況のなかで、大規模言語モデル (Large Language Models: LLM) は多様な自然言語処理タスクで顕著な性能を示しており、翻訳品質の面でも初級レベルの人間翻訳者に匹敵するパフォーマンスを発揮することが報告されている。だが、現時点の LLM でも翻訳を単独で完結させるには不十分な場面が存在し、専門的な翻訳品質の要求水準を満たすためには依然として追加的な品質管理が必要とされる。

このギャップを埋める有望なアプローチとして、NMT システムと LLM を組み合わせた自動 PE が注目されている。NMT が生成した翻訳文の誤りを LLM が検出・修正することで、翻訳者が担ってきた PE の一部を代替できる可能性がある。しかしながら、LLM による PE の性能を体系的に評価した研究はいまだ少なく、定量的な知見は限られている。

本研究では、LLM による英日機械翻訳 PE タスクの精度を多角的に評価することを目的とする。誤り訂正の成否を示す正解率 (correct rate) と過剰な書き換えの程度を示す修正率 (edit rate) の 2 つの指標を並行して評価し、それぞれへの影響を定量的に解析する。

評価に用いる要因は、(1) 機械翻訳出力に挿入した誤りのタイプ、(2) LLM への指示文 (プロンプト)、(3) 使用する LLM のモデルの 3 つである。正解率の分析にはロジスティック回帰分析を、修正率の分析に

は三元配置分散分析 (ANOVA) を適用し、各要因の主効果および要因間の交互作用を定量的に評価する。これにより、LLM の PE 性能に最も強く影響する要因とその組み合わせを特定し、実用的な運用指針を提供することを目指した。

### 2. 先行研究

#### 2.1 MT-PE

PE については、人間翻訳者の認知プロセスや作業効率の観点から長年にわたる研究の蓄積がある。Koponen らは、MT 誤りのタイプが PE 難易度に大きく影響することを報告しており、特に意味的な誤りや省略は表層的・局所的な誤りよりも修正が困難であるとされている[1]。NMT 普及後の研究においても、NMT 出力に対する PE は SMT (統計的機械翻訳) 時代と比較して作業効率が改善される一方、流動性が高いため誤りの検出が難しくなる傾向があることが示されている[2]。

#### 2.2 LLM を用いた翻訳・PE

LLM による翻訳タスクの品質評価も活発に行われている。Jiao らは ChatGPT (GPT-3.5/4) を多言語翻訳タスクで評価し、GPT-4 は高リソース言語対において専門翻訳者に匹敵する水準に達するが、低リソース言語や専門ドメインでは依然として限界があると報告した[3]。また Yan らは GPT-4 と複数レベルの人間翻訳者を網羅的に比較し、初級レベルの翻訳者とは同等であるものの、中・上級の翻訳者には及ばない場合があることを示した[4]。

LLM の PE への直接的な適用に関しては、Ki らの報告が重要な先行研究である。同研究では、MQM アノテーションによる誤り情報をプロンプトに組み込むことで LLaMA-2 の PE 精度が改善されることを示しており、外部フィードバックと LLM の能力を統合する方向性を提示している[5]。本研究はこれと問題設定を共有しつつ、誤りの種類・プロンプト・モデルという 3 つの要因を系統的に操作して精度への影響を定量化する点で新たな貢献を目指すものである。

### 2.3 LLM における過剰修正

LLM を校正・訂正タスクに適用する際に顕在化する問題として、過剰修正 (overcorrection / over-editing) がある。これは、モデルが正しい箇所や修正不要な箇所にまで変更を加え、結果として出力品質を低下させる現象である。Katinskaia らは文法誤り訂正 (GEC) タスクにおいて LLM の過剰修正傾向を実証的に示しており、LLM が誤りのない文に対しても書き換えを行う割合が高いことを報告した[6]。翻訳出力への PE においても同様の問題が懸念されており、過剰な書き換えは参照文との乖離を生じさせ、PE の評価指標である Human-targeted Translation Edit Rate (HTER) を悪化させる要因となりうる。本研究ではこの問題を定量化するために修正率 (edit rate) を独立した評価指標として設定し、過剰修正をもたらす要因の組み合わせを明らかにすることを試みる。

### 2.4 プロンプトエンジニアリング

LLM の出力はプロンプトの設計に強く依存することが多くの研究によって示されている。Zero-shot prompting は事前例示なしに指示のみでモデルを動かす基本的アプローチであり、Kojima らは単純なトリガー句を付加するだけで、LLM が多様な推論タスクで性能を発揮すること (Zero-shot-CoT) を示した[7]。Few-shot prompting はプロンプトに少数の入出力例を含めてモデルの挙動を誘導する in-context learning の代表形

式であり、Brown らによる評価研究で LLM におけるその有効性が広く確立された[8]。

Persona prompting は、モデルに専門家としての役割・人格を付与する手法であり、一部のタスクでは応答品質や一貫性の改善が報告されている。しかし、その効果はタスク・ドメインに依存することも知られており、質問応答や推論タスクでは必ずしも有効でない場合もある[9]。Principles-based prompting (原則付与) は守るべき行動原則や制約をプロンプトに明示してモデルの出力を制御する手法であり、Bai らの Constitutional AI (CAI) において、明文化された原則リストによるモデルの自己改訂が有効であることが示された[10]。本研究ではこれら 5 カテゴリーを含む 8 種類のプロンプトのタイプを PE タスクに対して系統的に比較し、正解率・修正率の両面からプロンプト設計の最適化指針を探索する。

## 3. 材料と方法

### 3.1 評価データセット

評価には医薬品開発領域にドメイン特化させた英語→日本語の平行コーパスを用いた。日本語側の参照文に対して、人為的に各タイプの誤りを 1 文あたり 1 箇所挿入したデータセットを構築し、1 条件 (誤りタイプ×プロンプト×モデルの組み合わせ) あたり 200 文を用いた。5 種類の誤りタイプ × 8 種類のプロンプト × 6 種類のモデルの全組み合わせからなる合計 48,000 件のデータを解析対象とした。なお、本データセットは著者のリポジトリにおいて公開している<sup>1</sup>。

### 3.2 要因

#### 3.2.1 誤りタイプ (Error Type)

機械翻訳出力に人為的に挿入する誤りを、NMT システムに典型的に見られる 5 タイプ (表 1) に分類した。

<sup>1</sup> <https://github.com/TakeshiHayakawa>

### 3.2.2 プロンプト (Prompt)

LLM への指示文として、以下の 8 タイプを設定した。比較の基準は最低限の指示のみを含む Vanilla とした (表 2)。なお、プロンプトの具体的な指示文は著者のリポジトリにおいて公開する。

### 3.2.3 モデル

OpenAI 社の API を通じて利用可能な GPT モデル 6 種を比較対象とした。モデルは 2 世代 (GPT-4.1、GPT-5) にまたがり、各世代においてフルサイズ・中規模 (mini)・軽量 (nano) の 3 段階の規模のモデルを採用した (表 3)。なお、GPT-5 系のモデルは低負荷条件での運用を想定し、推論パラメータは reasoning effort = minimal および verbosity = low を指定した。

## 3.3 評価指標

### 3.3.1 正解率 (Correct Rate)

後編集によって機械翻訳出力の誤りが正しく訂正されたか否かを二値 (正解/不正解) で判定し、各条件における正解率 (0~1 の連続値) を算出した。判定は、修正後の出力を参照文と照合することで行った。

### 3.3.2 修正率 (Edit Rate)

後編集によって機械翻訳出力から書き換えられた文字の割合を修正率として定義した (0~1 の連続値)。修正率が高い場合、誤り訂正の範囲を超えた過剰な書き換えが行われていると解釈した。

## 3.4 統計解析手法

### 3.4.1 正解率の分析：ロジスティック回帰分析

正解率 (二値変数) を目的変数、誤りタイプ・プロンプト・モデルを説明変数とするロジスティック回帰分析を実施した。基準カテゴリは誤りタイプ=num、プロンプト=vanilla、モデル=gpt-4.1 とした。有意水準は  $\alpha = 0.05$  に設定した。

表 1

略称	誤りタイプ	説明
num	数値誤り (Numerical Error)	数字・単位・日付等の誤訳
reperr	繰り返し誤り (Repetition Error)	同一単語・フレーズの重複出力
ogen	湧出し (Over Generation)	原文にない内容の過剰生成
ugen	訳抜け (Under Generation)	原文内容の欠落・省略
semerr	意味的誤り (Semantic Error)	意味・語義の誤った翻訳

表 2

略称	タイプ	説明
vanilla	基準指示	最低限の後編集指示のみ
persona	役割付与	翻訳専門家としての役割を指定
context	背景説明	タスクの背景・目的を説明
principles	原則提示	指示の範囲と制約を明示
oneshot	1 例示	後編集の例を 1 件提示
fewshot	複数例示	後編集の例を複数件提示
reason	根拠提示	修正根拠を説明するよう指示
score	自己採点	修正結果を自己採点するよう指示

表 3

モデル名	系統	規模
gpt-4.1	GPT-4.1 世代	フル
gpt-4.1-mini	GPT-4.1 世代	中規模
gpt-4.1-nano	GPT-4.1 世代	軽量
gpt-5	GPT-5 世代	フル
gpt-5-mini	GPT-5 世代	中規模
gpt-5-nano	GPT-5 世代	軽量

### 3.4.2 修正率の分析：三元配置分散分析

修正率（連続変数）を目的変数、誤りタイプ・プロンプト・モデルを因子とする三元配置分散分析（ANOVA）を実施した。各要因の主効果に加え、交互作用についても評価した。

## 4. 結果

### 4.1 正解率の解析：ロジスティック回帰分析

ロジスティック回帰分析の結果を図 1 に示す。モデルは MLE 法によって推定され、収束が確認された。擬似決定係数（Pseudo  $R^2$ ）は 0.2460 であり、正解率の変動を相当程度説明していることが示された。

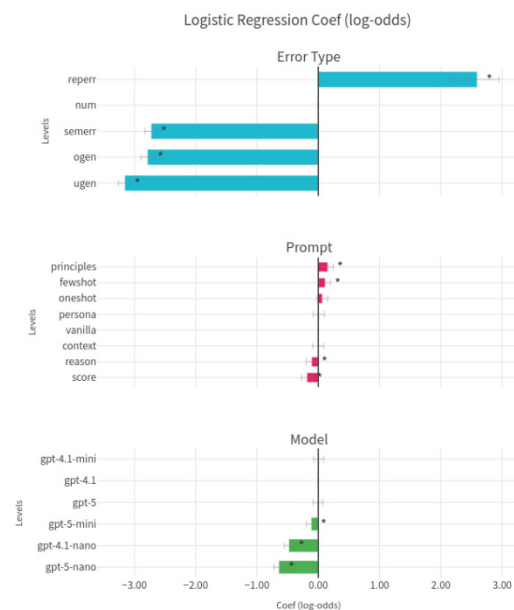


図 1：ロジスティック回帰分析による正解率の解析結果

#### 4.1.1 誤りタイプの効果

誤りタイプは正解率に対して最も強い影響を示した。基準カテゴリ（num）と比較すると、reperr はオッズ比  $\exp(2.591) \approx 13.3$  倍の正解しやすさを示した一方、ugen はオッズ比  $\exp(-3.155) \approx 0.043$  と訂正が最も困難であった。ogen（オッズ比  $\approx 0.062$ ）および semerr（オッズ比  $\approx 0.065$ ）も同様に正解率が大幅に低下した。

### 4.1.2 プロンプトの効果

プロンプトの影響は誤りタイプと比較して小さいものの、一部のプロンプトは統計的に有意な効果を示した。principles（係数 +0.154,  $p=0.001$ ）および fewshot（係数 +0.111,  $p=0.018$ ）は vanilla と比較して正解率を有意に改善した。一方、reason（係数 -0.108,  $p=0.020$ ）および score（係数 -0.183,  $p<0.001$ ）は正解率を有意に低下させた。persona、context、oneshot は vanilla と有意差がなかった。

### 4.1.3 モデルの効果

モデルについては、gpt-4.1 を基準として gpt-4.1-mini ( $p=0.854$ ) と gpt-5 ( $p=0.886$ ) は有意差がなく、実質的に同等の正解率を示した。一方、軽量モデルである gpt-4.1-nano（係数 -0.478,  $p<0.001$ ）および gpt-5-nano（係数 -0.642,  $p<0.001$ ）は有意に低い正解率を示した。gpt-5-mini も有意に低下した（係数 -0.113,  $p=0.006$ ）。世代（GPT-4.1 vs. GPT-5）による差は正解率において認められなかった。

### 4.2 修正率の解析：三元配置分散分析

修正率に対する三元配置 ANOVA の結果を図 2 に示す。

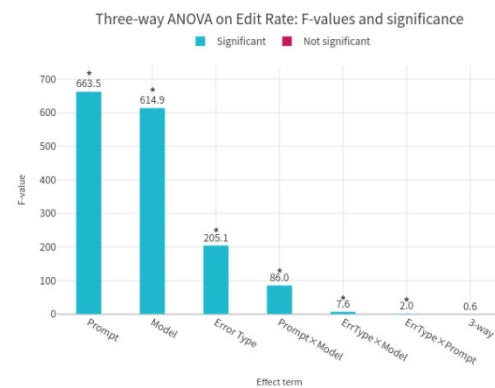


図 2：三元配置分散分析による修正率の解析結果

#### 4.2.1 主効果

プロンプト ( $F=663.46$ ,  $p<0.001$ )・モデル ( $F=614.86$ ,  $p<0.001$ )・誤りタイプ ( $F=205.07$ ,  $p<0.001$ ) の 3 要因

すべての主効果が統計的に極めて有意であった。F 値の大きさからはプロンプトが修正率に最も強い主効果を持ち、次いでモデル、誤りタイプの順であることが示される。

#### 4.2.2 交互作用

2 要因の交互作用についても、プロンプト×モデル ( $F=85.98, p<0.001$ )・誤りタイプ×モデル ( $F=7.63, p<0.001$ )・誤りタイプ×プロンプト ( $F=2.00, p=0.001$ ) のすべてが有意であった。特にプロンプト×モデルの交互作用は  $F=85.98$  と極めて大きく、修正率を規定する上で最も重要な要因の組み合わせであることが示された。一方、3 要因の交互作用 (誤りタイプ×プロンプト×モデル) は有意でなかった ( $F=0.57, p=0.999$ )。

### 5. 考察

本研究において最も顕著な知見は、LLM による後編集の正解率が機械翻訳出力に含まれる誤りのタイプに強く依存するという点である。繰り返し誤り (repperr) は正解率が最も高く、数値誤り (num) も高い正解率を示した。これらはいずれも表層的・局所的なパターンとして識別しやすく、LLM がルールに基づいた対処をしやすと考えられる。

これに対し、訳抜け (ugen)・湧出し (ogen)・意味的誤り (semerr) は正解率が大幅に低下した。訳抜けは原文の一部が日本語出力に反映されておらず、湧出しは原文にない内容が追加されている状態であるが、日本語文として文法的に成立している場合には、LLM がその箇所を誤りとして認識しにくい可能性がある。意味的誤りは文の表層的な形式は正しいが意味内容が誤っており、検出するには意味的整合性を詳細に確認する必要があり、現在の LLM にとって依然として困難なタスクであることが示唆される。

プロンプトエンジニアリングが正解率の改善に寄与し得ることが示されたが、その効果は誤りタイプの影響と比較すると限定的であった。有意に正解率を改善したプロンプトは principles (訂正の範囲と制約を明示)

および fewshot (複数の例示) であり、これらは先行研究における知見とも整合的である。指示の原則を明示することで、LLM が過剰な意識を避け、誤り箇所のみを修正する行動が引き出されたと解釈できる。

一方、reason (修正根拠の提示を求める指示) および score (自己採点の指示) は正解率を有意に低下させた。これは、LLM が根拠を生成する過程や自己評価を行う過程で、不必要な修正を加えたり誤り検出の精度が低下したりした可能性を示唆する。

興味深い知見として、モデルの世代 (GPT-4.1 vs. GPT-5) は正解率においてほとんど差をもたらさなかった。gpt-4.1 と gpt-5 の正解率は統計的に有意差なく ( $p=0.886$ )、gpt-4.1-mini と gpt-5 も同様であった。これは、後編集のような特定の指示追従タスクにおいては、モデルの世代 (能力の全般的な向上) よりも他の要因が支配的であることを示唆する。

一方、モデルの規模 (パラメータ数) が軽量であることは正解率に明確な影響を与えた。gpt-4.1-nano および gpt-5-nano は有意に低いパフォーマンスを示しており (それぞれ係数  $-0.478, -0.642$ )、軽量モデルでは後編集タスクに必要な精緻な指示理解が十分に達成されない可能性がある。実用的な観点からは、後編集の自動化においてコストと性能のトレードオフを考慮した上でモデルを選択することが重要である。

修正率 (edit rate) の分析では、プロンプトとモデルの交互作用が極めて強く ( $F=85.98$ )、過剰な修正を抑制するためにはプロンプトとモデルを個別に最適化するだけでは不十分であり、その組み合わせを考慮する必要があることが明確に示された。

特に顕著だったのは、reason および score プロンプトと nano モデルの組み合わせにおいて修正率が 0.60 を超えるケースが見られた点である。これは、軽量モデルが複雑な指示を適切に処理できず、結果として過剰な書き換えを行ったと考えられる。一方、principles プロンプトは修正率を大幅に抑制する効果があり、修正の範囲を明示的に制約することで過剰修正を防ぐことができると示唆される。

この知見は、実用上の重要な含意を持つ。翻訳において LLM を後編集に適用する際には、正しく修正するだけでなく、必要最小限の修正にとどめることが求められる。本研究の結果は、この目標の達成がモデルとプロンプトの適切な組み合わせ選択によって実現しうることを示している。

本研究にはいくつかの限界がある。第一に、評価データセットは人為的に誤りを挿入したものであり、実際の NMT システムが出力する自然な誤りとは分布が異なる可能性がある。第二に、正解率の判定は参照文との照合によるものであり、参照文と異なる表現であっても許容できる翻訳を不正解と判定された可能性がある。第三に、本研究では OpenAI の GPT シリーズのみを評価対象としており、他の LLM (LLaMA、Claude など) への汎化については検討が必要である。

今後の研究では、実際の NMT 出力に対する評価、人間の後編集者との比較、および多言語設定への拡張が課題となる。また、影響が相対的に低いとはいえ、人為的な介入が容易なプロンプトの最適化についてはさらなる検討の余地があり、プロンプト設計の改善が品質保証タスクにおいて有益な効果をもたらすことが期待される。

## 6. 結論

本研究では、LLM による英日機械翻訳後編集タスクの性能を 3 要因 (誤りタイプ・プロンプト・モデル) の観点から体系的に評価した。これらの結果は、LLM を翻訳後編集に実用的に活用するためには、誤りタイプの特性に応じた対処、適切なプロンプト設計、モデルとプロンプトの組み合わせ最適化が不可欠であることを示している。本研究の知見は、LLM を活用した翻訳ワークフローの設計・最適化に貢献するものである。

## 参考文献

1. Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, (25), 131-148.
2. Zouhar, V., et al. (2021, November). Neural machine translation quality and post-editing performance. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 10204-10214).
3. Jiao, W., et al. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv preprint arXiv:2301.08745*.
4. Yan, J., et al. (2024). Gpt-4 vs. human translators: A comprehensive evaluation of translation quality across languages, domains, and expertise levels. *arXiv preprint arXiv:2407.03658*.
5. Ki, D., & Carpuat, M. (2024, June). Guiding large language models to post-edit machine translation with error annotations. In *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 4253-4273).
6. Katinskaia, A., & Yangarber, R. (2024, May). GPT-3.5 for grammatical error correction. In *Proceedings of the 2024 joint international conference on computational linguistics, language resources and evaluation (LREC-COLING 2024)* (pp. 7831-7843).
7. Kojima, T., et al. (2022). Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35, 22199-22213.
8. Brown, T., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
9. Shanahan, M., et al. (2023). Role play with large language models. *Nature*, 623(7987), 493-498.
10. Bai, Y., et al. (2022). Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*.

## ポストエディットが翻訳者の仕事満足度に与える影響

Akiko Sakamoto<sup>1</sup>, Darren Van Laar<sup>2</sup>, Joss Moorkens<sup>3</sup>, Félix do Carmo<sup>4</sup>

<sup>1</sup> 関西大学, <sup>2</sup> University of Portsmouth, <sup>3</sup> Dublin City University, <sup>4</sup> University of Surrey

### 1. なぜ、今、仕事満足度なのか

効率的な翻訳プロセスを目指して機械翻訳のポストエディット (Machine Translation Post-Editing: MTPE) が普及しつつある一方、効率優先のプロジェクト管理が翻訳者の不満を引き起こしていると言われている。不満の理由としては、ポストエディットの仕事では経済的目標が優先され翻訳者のもつ文化的資本が認識されていないと感じること<sup>[1]</sup>、翻訳会社側でもポストエディットの公平な報酬モデルの確立が難しいこと<sup>[2]</sup>、などが挙げられる。欧州では経験豊かな翻訳者がすでに翻訳業界を去る動向が出ているという報告もある<sup>[3]</sup>。これが本当であれば、優秀な翻訳者の労働市場の長期的な持続は期待できない。このような現象が本当に起きているのか。そうであれば、その理由は何なのか、を探求するのが、本研究の目的である。

昨今では海外で翻訳者の仕事満足度に関する研究が進んできている<sup>[4]</sup>。本研究は翻訳学の研究者 (第1, 3, 4 著者) が産業心理学の研究者 (第2 著者) とタッグを組み、イギリスの翻訳者協会である Institute of Translation and Interpreting (ITI) の協力を得て、イギリスの翻訳者を対象に行った (n=381)。労働者の仕事満足度を測るために第2 著者が開発した Work-Related Quality of Life (WRQoL) 尺度<sup>[5]</sup> をフリーランスの翻訳者用に調整し、アンケートの一部に使用した。残りの部分はポストエディットなど翻訳テクノロジー関係の要因に対する翻訳者の意識・態度を測るためにオリジナルの心理尺度を作成して使用した。そのために2回のパイロット調査を行い、統計的に信頼性と妥当性のある尺度作成につとめた (1 回目のパイロット調査の

結果は[6]参照)。このアンケートを使って翻訳者の労働生活とその満足度に関する広範なデータを集めることができた。本稿ではアンケート結果のうち、ポストエディットに関する一部を紹介する。なお、アンケートの全体の結果は[7]を参照されたい。

### 2. だれがポストエディットを (前向きに) やっているのか

アンケートの回答者は ITI の会員であったため、比較的経験の長い翻訳者が多く、そのため翻訳者の仕事全体に占めるポストエディットの割合は平均で 23.5% (SD = 27.94) と低めだった。高い標準偏差からも分かるように、個人差は大きく、全くポストエディットの仕事をしていない人もいれば (126 人, 33.1%)、翻訳の仕事の 100%がポストエディットだと答えた人も 2 人 (0.5%) いた。ポストエディットの仕事の受注量は、翻訳者個人によってまだ差が大きいことが分かる。

次に、ポストエディットの仕事をどの程度積極的に受けているかを「MTPE への前向き度」というファクターで測定し、属性によって差異があるかを見た。その結果、5つの属性で有意な差があることがわかった (全てのファクターについては次セクション参照)。

- ① ITI の4つの会員レベルのうち、よりエントリーレベルの会員のほうが積極的 (F (3,377) = 6.03,  $p < .001$ )
- ② 大学(院)の翻訳学の学位を持っている人のほうが、持っていない人より積極的 ( $p = .003$ ,  $d = .306$ )
- ③ サイエンス系分野の翻訳者のほうが文系分野の翻訳者より積極的 ( $p < .001$ ,  $d = .354$ )

The impact of post-editing on translators' quality of working lives

Akiko Sakamoto<sup>1</sup>, Darren Van Laar<sup>2</sup>, Joss Moorkens<sup>3</sup>, Félix do Carmo<sup>4</sup>

<sup>1</sup>Kansai University, <sup>2</sup> University of Portsmouth, <sup>3</sup> Dublin City University, <sup>4</sup> University of Surrey

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/>

- ④ 直接の顧客より翻訳会社との取引が多い人ほど積極的 (直接の顧客の割合に対して  $r = -.212, p < .001$ )
- ⑤ 翻訳の経験が浅い翻訳者のほうが積極的 (翻訳の経験の年数に対して  $r = -.141, p = .005$ )

ただし、この分析はポストエディットの仕事量とポストエディットへの態度との因果関係を見ているわけではないので、解釈には注意が必要だ。翻訳の仕事を始めたばかりの人が人手翻訳の仕事をすぐに得るのは難しいのでまずポストエディットの仕事から始めるのであれば、それが必然的にポストエディットへの前向きな態度として数字に出ているのかもしれない。つまり、進んで前向きなのではなく、それしかチョイスがないから前向きなのかもしれない。

ひとつ注目したいのは、②の翻訳学の学位との関係だ。イギリスや欧州の大学院では翻訳学のコースが多く存在し、プロの翻訳者になるには当然求められる資格となっている。また翻訳学のコースではふつう MT やポストエディットの授業が提供されている。したがって、最近大学院の翻訳学コースを卒業した翻訳者はポストエディットの知識と経験をすでに持っているため、翻訳学を学んでいない翻訳者に比べてポストエディットの仕事を受けるチャンスが多いか、または心理的ハードルが低いのかもしれない。欧州のように翻訳学の学位を提供する大学(院)がほぼ存在しない日本では、将来の翻訳者のキャリア形成を高等教育の時点からサポートするのは難しい。しかし今後は若い翻

訳者の育成にどのようにポストエディット教育を組み込むかも考えていかないといけないだろう (ちなみに第1著者の勤務校ではポストエディット教育を翻訳教育に組み込む取り組みを行っている)。

### 3. ポストエディットが翻訳者の仕事満足度に与える影響

本調査でとくに分かったことは、ポストエディットの仕事を多く請け負う翻訳者のほうが仕事への満足度や労働生活の質が低いという点である。その結果を以下に報告する。

翻訳者の労働生活について 12 個の側面を構成概念 (以下ファクターと呼ぶ) として測定し、ポストエディットとの関係性を調べた。12 個のファクターは以下のとおりである。

- (1) ウェルビーイング (2) 仕事のコントロール感 (3) 雇用者 (翻訳クライアント) との関係 (4) ワークライフバランス (5) 仕事のストレス (6) 労働環境 (7) 仕事の安定感 (8) 仕事のフィット感 (9) 同業者とのつながり (10) 翻訳テクノロジーへの肯定感 (11) MT の利便性認識 (12) MTPE への前向き度

これらのファクターと、翻訳者の仕事に占めるポストエディットの割合との相関性を、ピアソンの相関係数 ( $r$  係数) で調べた。すると、図1の左側の3つのファクターが正の ( $r$  係数がプラス)、右側の5つのファクターで負の ( $r$  係数がマイナス) 相関であること

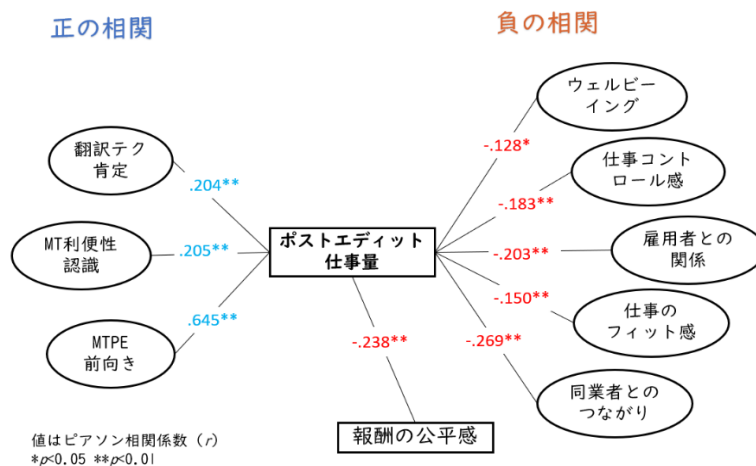


図1 ポストエディットの仕事量と翻訳者の労働生活ファクターとの関係性

が分かった。また仕事の報酬の公平感との相関もマイナスであった。この結果についても、ポストエディットの仕事量と各ファクターの相関関係を見ているのであり、因果関係を見ているのではないことは留意する必要がある。例えば、正の相関の方を見ると、ポストエディットの仕事をたくさんするからポストエディットに前向きになるのか、もともとポストエディットに前向きだからポストエディットの仕事をたくさん受けているのか、断定することはできない。しかし負の相関の方では項目間の時間的ズレを考慮すると、ポストエディットの仕事量が要因となって他のファクターに負の影響を与えていると考えて妥当であろう。例えば、ポストエディットの仕事量と翻訳クライアント（翻訳会社）との関係が  $r = -.203^{**}$  と有意な負の相関であるが、これは翻訳会社になにか不満があるから（関係がマイナス）ポストエディットの仕事の受注量が増える（仕事量がプラス）のではなく、ポストエディットの仕事をたくさん受けるから（仕事量プラス）翻訳会社に不満がある（関係がマイナス）と考えたほうが時系列的に妥当だ。また、6つもの項目で負の相関関係があるということは、ポストエディットが翻訳者の仕事満足度の多くの面にマイナスの影響を及ぼしていることを示唆する。

#### 4. ポストエディット産業の将来 優秀な翻訳者はどこに行く？

最後に、業界の将来に関係する調査結果を提示したい。アンケートは「今後少なくとも5年間翻訳の仕事が続けたいか」という質問項目を含んだ。この項目のスコアと他の質問項目のスコアの相関関係を見て、どのような要因が長期的キャリアの展望につながる可能性があるか、つまりどうすれば翻訳者が長期的に仕事をしたいと思うかを探った。項目間のピアソン相関係数（ $r$ 係数）を比較した結果、アンケートの測定項目のなかでは「仕事の楽しさ」「仕事量」「報酬の公正感」がキャリア継続の3大要因になっていることが分かった。一方で、ポストエディット関係の質問項目との相

関性を見ると、正負双方で有意な相関性が見られる項目は全くなかった。つまり、今たくさんポストエディットの仕事積極的に受けていても、またポストエディットの仕事が楽しいと思っても、それが長期的なキャリア設計に結び付いていないようだ。

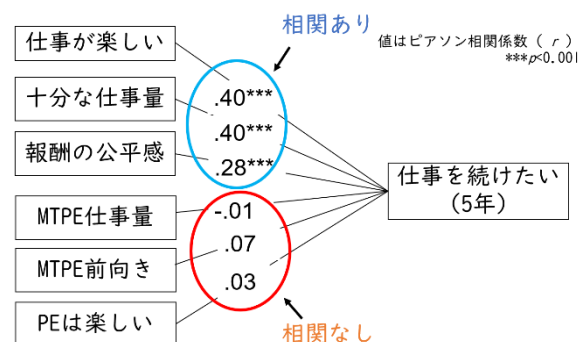


図2 翻訳者の仕事を続けたい気持ちと他項目との関係

翻訳産業におけるプロダクションモデルはこれからどんどん人手翻訳からポストエディットへ移行していくことが予想される。業界全体として優秀なポストエディターが多数必要となるわけだ。しかしこの調査結果を見ると、今のポストエディットの仕事の出し方を続けていては優秀な翻訳者の定着が見込めない可能性が高い。これは翻訳産業の持続可能性の危機につながる重要な問題だ。

#### 5. 持続可能な翻訳産業のために

以上、イギリスの翻訳者を対象に行ったアンケート調査から、ポストエディットが翻訳者の仕事満足度を与える影響について、結果の一部を紹介した。翻訳者の仕事は今後、ますますポストエディットの割合が増えることが予想される。これを受けて、ポストエディットのスキルを伸ばすなど、個人レベルで努力をしている翻訳者はたくさんいるだろう。しかし、本調査が示唆したように、それが仕事の満足度や生活のウェルビーイングにつながらなければ、将来的に優秀

な翻訳者が翻訳産業に定着せず、結局は業界全体のロスにつながる可能性が高い。

翻訳者の仕事満足度を高めるようなポストエディット案件の発注には、何が必要だろうか。ひとつ挙げられるのは、適切なポストエディットのガイドラインを発注時に示すことだ。これについては、AAMTのポストエディット委員会が発行した『機械翻訳ポストエディットガイドライン』<sup>[8]</sup>に見られるように、業界内の努力は行われている。MT訳のどの種類・レベルのエラーを修正するべきかという詳細な指示をポストエディターに提示するべきだと言われている<sup>[9][10]</sup>。またポストエディットにかかる時間について、翻訳会社側は現実的な想定をすることが大切だ。質の高いポストエディットを行うのに十分な納期を翻訳者に与えたほうがよいのは当然である。最近の実証的研究でポストエディットは必ずしも人手翻訳より速いとは限らないことが分かっているにもかかわらず<sup>[11]</sup>、現実には「ポストエディットは人手翻訳より速いはず」という考えが業界やクライアントの間で先行しているのではないだろうか。また、「クライアントがこう要求しているので、翻訳者の仕事もこうでなくてはならない」というナラティブが労働条件の決定に不公平な影響を与えていることはないだろうか。持続可能な翻訳者市場を実現するためのエビデンスとして本調査の結果を少しでも役立てていただければ幸いだ。

なお本アンケート調査は2024年夏にイギリスで行ったものだが、今後定期的に調査を行い、時縦断的な動向を調べる計画である。また2026年度は日本での調査も計画している。

#### 参考文献

- [1] A. Sakamoto, “Why do many translators resist post-editing? A sociological analysis using Bourdieu’s concepts,” *J. Spec. Transl.*, vol. 31, pp. 201–216, 2019.
- [2] A. Sakamoto and S. Bawa Mason, “In search of a fair MTPE pricing model: LSPs’ reflections and the implications for translators,” *Perspectives*,
- vol. 32, no. 3, pp. 460–476, 2024, doi: 10.1080/0907676X.2023.2292572.
- [3] ELIS Research, “European Language Industry Survey 2025,” 2025. [Online]. Available: [https://elis-survey.org/wp-content/uploads/2025/03/ELIS-2025\\_Report.pdf](https://elis-survey.org/wp-content/uploads/2025/03/ELIS-2025_Report.pdf)
- [4] M. Ruokonen, E. Svahn, and A. Heino, Eds., “Translators’ and interpreters’ job satisfaction,” *Spec. issue Transl. Spaces*, vol. 13, no. 1, 2024.
- [5] S. Easton and D. Van Laar, “User manual for the work-related quality of life (WRQoL) scale second edition,” 2018. [Online]. Available: [http://www.qowl.co.uk/researchers/WRQoL\\_User\\_manual\\_2nd\\_Ed\\_ebook\\_Feb\\_2018\\_55.pdf](http://www.qowl.co.uk/researchers/WRQoL_User_manual_2nd_Ed_ebook_Feb_2018_55.pdf)
- [6] A. Sakamoto, D. Van Laar, J. Moorkens, and F. do Carmo, “Measuring translators’ quality of working life and their career motivation: Conceptual and methodological aspects,” *Transl. Spaces*, vol. 13, no. 1, pp. 54–77, 2024, doi: <https://doi.org/10.1075/ts.23026.sak>.
- [7] A. Sakamoto, D. Van Laar, J. Moorkens, and F. do Carmo, “Translators’ Work-Related Quality of Life: Survey Report,” 2025. [Online]. Available: <https://www.iti.org.uk/resource/twrqol-iti-final-report.html>
- [8] AAMT, “機械翻訳ポストエディットガイドライン,” 2025. [Online]. Available: [https://aamt.info/wp-content/uploads/2025/07/AAMT\\_PE\\_Guideline\\_Ver1.0.pdf](https://aamt.info/wp-content/uploads/2025/07/AAMT_PE_Guideline_Ver1.0.pdf)
- [9] A. Sakamoto and M. Yamada, “Managing Clients’ Expectations for MTPE Services Through a Metalanguage of Translation Specifications: MPPQN Method,” in *Metalanguages for Dissecting Translation Processes*, R. Miyata, M. Yamada, and K. Kageura, Eds., London: Routledge, 2022, ch. 14, pp. 191–199. doi: 10.4324/9781003250852-17.

- [10] C. Rico Pérez, “Re-thinking machine translation post-editing guidelines,” *J. Spec. Transl.*, vol. 41, pp. 26–47, 2024, doi: 10.26034/cm.jostrans.2024.4696.
- [11] S. Terribile, “Is post-editing really faster than human translation?,” *Transl. Spaces*, vol. 13, no. 2, pp. 171–199, 2023, doi: 10.1075/ts.22044.ter.

# 大規模言語モデルは多様な翻訳仕様に追従できるか？

萱野陽子<sup>1,2</sup> 菅原朔<sup>1,2</sup><sup>1</sup>総合研究大学院大学 <sup>2</sup>国立情報学研究所

## 1. 背景

翻訳は、その目的や用途、想定読者などの要件に応じて「なにが良い訳か」が変化する。そのため、翻訳品質も固定的な単一基準で測ることが難しい。近年、大規模言語モデル（LLM）を用いた翻訳生成は活発に行われ、翻訳の品質評価にも広く利用されている[1][2][3][4]。しかし、現状では、LLMが多様な翻訳仕様に沿った翻訳がどの程度可能なのか、翻訳仕様に基づいた評価がどの程度妥当なのかは十分に検証されていない。

本稿は、AAMT 2025 年次大会におけるポスター発表「大規模言語モデルは多様な翻訳仕様に追従できるか？」の内容を、ジャーナル向けに再構成したものである。LLMによる翻訳が多様な翻訳仕様へのどの程度追従できるかを評価するため、翻訳目的別（Informative、Persuasive、Legal、Artistic）の翻訳仕様を明示し複数の翻訳モデルによる生成を比較した。また、複数の LLM を用いて、エラー評価と仕様適合評価を行った。その結果、LLM は一定程度仕様に応じた出力を生成できる一方で、その追従の程度は仕様や評価者に依存することが確認された。なお、本研究はアイデア検証段階の実験であり、設計上の制約や改善余地があることをあらかじめ明記しておく。

## 2. 実験設計

### 2.1 原文・言語と評価目的

対象言語ペアは日英および英日とし、原文には 3 ジャンル（政府白書、企業ホームページ、文学作品）から計 18 本を選定し、抜粋して用いた。特定の文体や

分野に偏らないよう、特徴の違うジャンルの文章を含めている。1 原文あたりの平均長は、日本語 896 文字、英語 314 語である。

翻訳生成に際しては、目的別の翻訳仕様として 4 条件（Informative、Persuasive、Legal、Artistic）を設定し、プロンプトを通じて明示的に与えた。これらの 4 条件は、文体や機能の違いが比較的明確で、翻訳出力の差を観察しやすいように選定した。Informative は正確さと明瞭さを、Persuasive は訴求力と読者への働きかけを、Legal は曖昧性の排除と厳密性を、Artistic は文学的な表現をそれぞれ重視する仕様として定義した。加えて、翻訳目的を明示せず、単純な翻訳指示のみを与える仕様なし条件も比較のために設けた。

図 1 は同一の原文に対する翻訳の違いの 1 例である。

<p>原文（トヨタ自動車公式サイト「トヨタ行動指針」より） お客様の笑顔のために、クルマに向き合い、現場で必死に汗をかく。そんな「クルマ屋」としての原点に立ち返った結果、私たちの商品は大きく変わりました。危機の時でも立ち止まらない、企業としての強さも身につけてきたと思います。</p>
<p><b>企業公式英訳</b> By returning to the origins of being a “car maker,” earnestly sweating at the gamba, and focusing on cars for the sake of customer smiles, our products have undergone significant changes. We believe that we have also acquired strength as a company that will not stop moving forward even in times of crisis.</p>
<p><b>DeepL（翻訳仕様なし）</b> We are always working hard to bring smiles to the faces of our customers. As a result of returning to our roots as a “car dealer,” our products have changed dramatically. I believe that we have also gained strength as a company that does not stand still even in times of crisis.</p>
<p><b>ChatGPT Persuasive（説得目的）</b> To earn every customer’s smile, we faced our Cars head-on and worked hands-on at the front lines. By returning to our origins as a true “Car maker,” our products changed significantly—and we gained the resilience to keep moving forward, even in times of uncertainty.</p>
<p><b>ChatGPT Artistic（芸術的目的）</b> For the smiles of our customers, we have faced our cars, perspiring earnestly on the frontlines. Returning to the roots of being a “car-maker,” our products have undergone a profound transformation. Even in crisis, we have cultivated the resilience to keep moving forward as a formidable enterprise.</p>

【図 1】 同一の原文に対する翻訳の違いの 1 例

Can Large Language Models Follow Different Translation Specifications?

Yoko Kayano, Saku Sugawara

The Graduate University for Advanced Studies, SOKENDAI / National Institute of Informatics

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/>

## 2.2 仕様書の自動生成

翻訳仕様は、翻訳プロジェクト仕様書を模した JSON 形式で表現した。仕様書には、目的、対象読者、文体、文化的配慮といった訳文が仕様にとどの程度適合しているかを評価するための項目に加え、用語指定、参照資料、長さ制約、表記方針など、正誤を比較的客観的に判断できるルールも含めた。これらの項目を含む仕様書テンプレートを作成した。

このテンプレートに基づき、各原文に対して4つの目的別仕様を自動生成し、合計72件の仕様書を作成した。生成には GPT-4o を用い、temperature は 0.3 に設定した。仕様書の生成では、多少の表現の多様性がありながらも、項目間の一貫性を重視し、温度を低めに抑えている。生成後には、各仕様書が対応する目的と大きく矛盾していないかを著者が確認した。

## 2.3 翻訳生成モデル

翻訳生成には GPT-4o、GPT-5、Gemini 2.5 Flash、Llama 3 70B Instruct を用い、比較のために DeepL をベースラインとして加えた。各原文について、各 LLM は4つの仕様と仕様なしの5種類の翻訳を生成した。一方、DeepL については仕様を直接入力できないため、仕様なしに相当する翻訳のみを生成した。また、Llama については zero-shot に加え、few-shot (以下) も別途設定した。few-shot では、各目的に対応する具体例を事前に与えることで、出力が改善するかどうかを確認した。したがって、本研究で扱う翻訳の総数は、GPT-4o、GPT-5、Gemini、Llama、LlamaFS の5モデル分の各5翻訳に DeepL の一翻訳を加えた、18 原文×26 翻訳=468 翻訳である。生成時の temperature は、GPT-5 が当時 1.0 のみ利用可能であったため 1.0 とし、それ以外のモデルについては 0.7 に設定した。この値は、語彙選択や文体に一定の創造性を残すことを意図したものである。

## 2.4 LLM による自動評価

評価は、エラー分析と仕様適合度評価の二つの側面から構成した。各原文につき 4 仕様 × 5 モデル (LlamaFS を除く) に加え DeepL を含む計 21 翻訳を一単位として評価モデルに提示し、評価目的ごとにブラインド評価を実施した。翻訳モデル名は匿名化し、提示順はランダム化したうえで、乱数シードを固定した。LlamaFS による翻訳については、別途 Llama (zero shot) と比較する形で評価を行った。

評価者として用いたのは GPT-4o、GPT-5、Gemini 2.5 Flash の3モデルである。各評価者は、全468翻訳を4つの評価目的の観点から判定した。評価時の temperature は、GPT-5 が当時 1.0 のみ利用可能であったため 1.0 とし、GPT-4o と Gemini 2.5 Flash については 0.5 に設定した。評価タスクには、点数付けのみでなく仕様適合の理由説明も含まれるため、ある程度の解釈の柔軟性を残すことを意図した。

評価タスクでは、各翻訳について、エラー情報、仕様適合スコア、およびそれらの判断根拠を JSON 形式で出力させた。エラー情報としては、MQM フレームワーク [5] に基づく Accuracy および Linguistic Conventions のエラーに加え、仕様書に定義した客観ルール違反について、エラー件数と該当箇所のコメントを取得した。また、仕様適合については、目的、読者、文体、文化的配慮の4観点を1~5の尺度で評価させ、その理由をコメントとして出力させた。

## 2.5 スコア定義

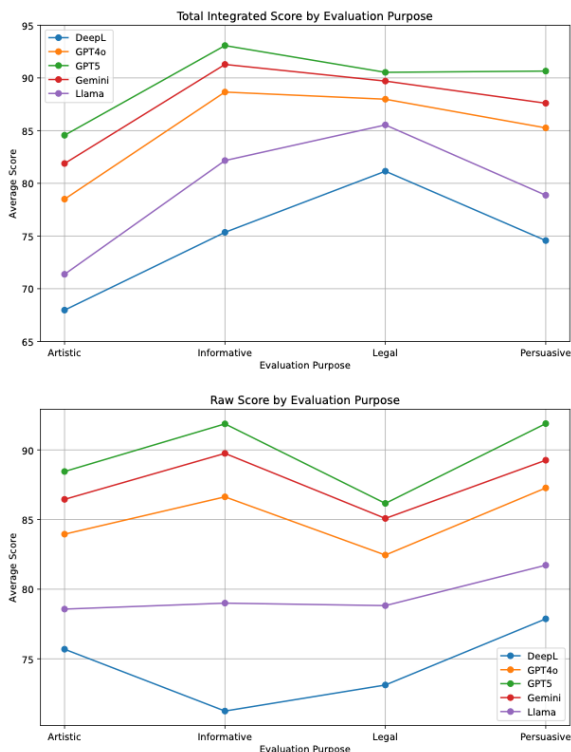
評価結果からは、エラーベーススコア、仕様適合スコア、および両者を目的別に重みづけして統合した全体平均スコアを算出した。エラーベーススコアは、1 エラー1 点減点という単純な方式で算出した。100 点満点でスコアが高いほどエラーが少ないことを表す。

仕様適合スコアは、目的、読者、文体、文化的配慮の4観点をそれぞれ1~5で評価し、その平均を20点満点に換算したうえで、0~100点に再スケールしたも

のである。これにより、仕様適合スコアはエラースコアと同一のスケール上で比較可能となる設計とした。

全体平均スコアは、エラーベーススコアと仕様適合スコアを目的別に重みづけして統合したものである。重み設定（エラーベース/仕様適合）は、Informativeでは0.6 / 0.4、Legalでは0.7 / 0.3、Persuasiveでは0.4 / 0.6、Artisticでは0.3 / 0.7とした。InformativeとLegalではエラーの少ないことを相対的に重視し、PersuasiveとArtisticでは仕様適合をより重視する設計とした。

加えて、エラーベーススコアと仕様適合を0.5ずつの重みで統合した重みなしスコアも算出した。これは、目的別重みづけがモデル順位や全体傾向に与える影響を確認するためである。重みづけの有無によって、全体的な傾向に大きな変化はなかった。



【図2】 評価目的別の全体平均スコア（上）と重みづけなしスコア（下）

### 3. 結果

#### 3.1 全体平均スコアとモデル差

全体平均スコアをみると、GPT-5による翻訳が3評価者すべてで最上位となり、その後にGeminiとGPT-

4oが続いた。重みづけなしのスコアにおいてもその順位は変わらない（図2）。LlamaとDeepLはこれらの上位群より明確に低い位置にあり、とくにDeepLは一貫して最下位であった。仕様を明示的に扱えないDeepLが不利になった点は、予想に沿う結果であるが、同時に、少なくとも本設定では仕様情報が評価に実際に影響していることも示している。

一元配置分散分析（ANOVA）の結果、各評価者において翻訳モデル間の差は有意であった。事後検定でも、DeepLとLlamaが低い群を形成する傾向は比較的一貫していた。一方、GPT-5、Gemini、GPT-4oの上位3モデルについては、評価者によっては有意差が観察されない組み合わせもあり、上位群内の差は下位群との差ほど大きくはないことがうかがえる。

#### 3.2 評価モデル間の一致

3 評価者間の相関分析では、高い正の相関が確認された。ピアソン相関はおおむね0.77前後、スピアマン順位相関は0.73~0.81の範囲にあり、どの評価者も翻訳の相対順位について類似した判断をしていると解釈できる。その一方で、完全な一致が見られたわけではない。GPT-5およびGeminiの評価では、自身の翻訳を他モデルより高く評価する傾向が観察された。したがって、LLMを評価者として用いる際には、単一の評価者による結論をそのまま採用するのではなく、複数評価者の比較を前提とした設計が望ましい可能性がある。

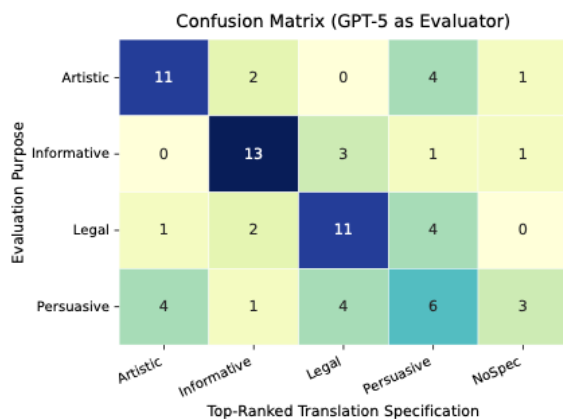
#### 3.3 目的×仕様の対応

各評価目的において、最上位と判断された翻訳仕様を集計すると、評価目的と翻訳仕様の一致率にはモデル差が見られた。GPT-5は一致率が最も高く、与えられた仕様に応じて訳文を正しく区別して評価している傾向が見られた（図3）。

これに対してPersuasiveでは、全モデルで一致率が低下する傾向があった。説得的であることには、語調や流暢さなど複数の要素が含まれ、他の仕様との境界

が曖昧になりやすい可能性がある。また、Legalにおいて一部の評価者が Informative を高く評価する例も見られた。

この結果は、仕様追従の問題が、単に翻訳出力が仕様に応じて変わったかどうかだけでなく、その変化が評価者にどのように解釈されるかにも依存することを示している。すなわち、翻訳モデルの側の追従能力だけでなく、評価者側が仕様をどのように解釈し、それに基づき判断するのも最終的な一致率に影響する。評価者側が、どの程度仕様に沿った評価ができるのかを改めて検証する必要がある。



【図3】 GPT-5による評価の混同行列（評価目的×最上位仕様）。対角線上の値は、評価目的と翻訳仕様が一致した件数を示す。

### 3.4 仕様ありの効果

仕様ありの4種類の翻訳と仕様なし翻訳を比較すると、全体平均スコアは多くのモデルで0.4~0.9程度上昇した。絶対値としては大きな差ではないが、少なくとも一貫したプラス方向の傾向が見られた点は重要である。とくにGPT-5ではこの差が最も大きく、仕様情報を用いて出力を目的別に調整している可能性が高い。

ただし、仕様ありの効果は全目的で一様ではなかった。Persuasiveでは有意な改善が確認される一方で、Artisticでは仕様を与えた方がかえってスコアが下がる場合もあった。

### 3.4.1 LlamaFS (few shot)の結果

LlamaFSでは、各目的に対応する具体例を与えることでfew-shotによる改善を試みた。しかし、その効果は一貫しておらず、評価者によってはzero-shotより低く評価される目的が複数見られた。GPT-4o評価ではArtisticのみ改善があった一方、Informative、Legal、Persuasiveではむしろ低下した。対照的に、Gemini評価では全ての目的において大きく改善が観察された。

### 3.5 評価軸の違い

本実験では、エラーベーススコアは全体として高得点に集中し、モデル間の差は大きくなかった。これは、本実験の翻訳が全体として低品質ではなかったことを示す一方で、単純なエラーカウントでは差が十分に開かないことも意味している。とくに、重大度を区別せず1エラー1点減点とした設計では、軽度なエラーと重大なエラーが同じ重さで数えられるため、結果としてスコア分布が圧縮されやすい傾向があった。

これに対して、仕様適合スコアは翻訳間の違いをより明確に反映していた。本実験では、モデル間の差は主として、エラーの少なさよりも、与えられた目的や文体、想定読者にどの程度合った訳文を生成できていたかによって現れていたと考えられる。

ただし、このことはエラー評価が重要でないことを意味するものではない。誤訳を含む訳文は、どれほど仕様に適合していても高品質とはいえない。エラー評価は翻訳として最低限満たすべき条件を確認する役割を持ち、仕様適合評価はその上で目的に合った出力になっているかをみる役割を持つ。本結果は、両者が異なる側面を担う評価軸であることを示唆している。

## 4. 結論

本研究では、翻訳仕様を明示した条件のもとでLLMの翻訳生成を検討した。その結果、LLMは一定程度仕様に沿った出力を生成できるものの、その追従能力は仕様や評価者に依存することが示された。また、評価においては、エラー中心の評価では差が出にくく、

仕様適合の方がモデル差をより強く反映していた。翻訳評価では、エラー評価と仕様適合評価を区別した多軸的な設計が有用である可能性を示した。

## 5. 限界と今後の課題

本研究にはいくつかの制約がある。

第一に、原文数は18本にとどまり、対象言語およびジャンルも限定されているため、仕様追従の傾向を一般化するには十分ではない。

第二に、評価者はすべて LLM であり、人手評価との直接比較は行っていない。そのため、本研究での仕様適合スコアやモデル順位が、人間の評価者によってどの程度再現されるかは未検証である。

第三に、全体平均スコアに用いた重みは設計上の仮定に基づいている。Informative や Legal でエラーを重く見ること、Persuasive や Artistic で仕様適合を重く見ることには一定の合理性があるが、その配分は唯一の正解ではない。今後、人間評価や実務基準をもとに、重み設定そのものの妥当性を検証する必要がある。

第四に、エラー評価では重大度を区別せず単純件数で扱ったため、深刻なエラーと軽度なエラーが同一視されている。今後は重大度やエラータイプごとの重みづけを行い、実務に近い評価に近づける必要がある。

最後に、本研究で用いた翻訳仕様そのものも、実験のために抽象化されたものである。実際の翻訳実務では、仕様はさらに細かく、複合的である。今後は、こうした実務的な仕様をどの粒度で、どのような形式で表現すれば、翻訳生成と評価の両方において安定して扱えるのかを検討する必要がある。

## 謝辞

本研究は JST 創発的研究支援事業 JPMJFR232R の支援を受けたものである。

## 参考文献

[1] Tom Kocmi and Christian Federmann. 2023. GEMBA-MQM: Detecting translation quality error spans with GPT-4.

In Proceedings of the Eighth Conference on Machine Translation, pages 768–775.

[2] Patrick Fernandes, Daniel Deutsch, Mara Finkelstein, Parker Riley, André Martins, Graham Neubig, Ankush Garg, Jonathan Clark, Markus Freitag, and Orhan Firat. 2023. The devil is in the errors: Leveraging large language models for fine-grained machine translation evaluation. In Proceedings of the Eighth Conference on Machine Translation, pages 1066–1083.

[3] Nuno M. Guerreiro, Ricardo Rei, Daan van Stigt, Luisa Coheur, Pierre Colombo, and André F. T. Martins. 2024. xcomet: Transparent machine translation evaluation through fine-grained error detection. Transactions of the Association for Computational Linguistics, 12:979–995.

[4] Qingyu Lu, Liang Ding, Kanjian Zhang, Jinxia Zhang, and Dacheng Tao. 2025. MQM-APE: Toward high-quality error annotation predictors with automatic post-editing in LLM translation evaluators. In Proceedings of the 31st International Conference on Computational Linguistics, pages 5570–5587.

[5] Arle Richard Lommel, Aljoscha Burchardt, and Hans Uszkoreit. 2013. Multidimensional quality metrics: a flexible system for assessing translation quality. In Proceedings of Translating and the Computer 35.

## 温故知新

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 発展に役立つものが多く見られます。

本号では、1994 年 3 月号と 1994 年 6 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。

【1994 年 3 月 AAMT ジャーナル No.6 より】

### 機械翻訳の今後の技術動向

#### Future Trend of Machine Translation

#### Technology

通産省電子総合研究所 主任研究官 井佐原 均

今回は、機械翻訳システムの新しい技術動向について紹介します。

現在市販されている機械翻訳システムは、(当たり前だといわれるかも知れませんが)文法規則と辞書を使って、入力文を解析し、翻訳出力を生成します。これまで、このシリーズで行なってきた機械翻訳技術の解説も、このようなやり方に沿ったものでした。このようなやり方を、規則(辞書の記述も規則の一つです)を用いる手法という意味で、Rule-based の手法と呼びます。このような手法で作られた機械翻訳システムの

性能を向上するためには、これまでこのシリーズで説明されてきた構文解析や意味解析といった要素技術をさらに深く研究していくことが必要です。

また、実際にシステムが使う文法規則や辞書に書く情報を出来るだけ詳細なものにいくことも必要となります。

これまでは、翻訳システムが用いるさまざまな規則は、人間が自分の言語的直観に基づいて作成してきたわけです。つまり、自分が言葉を理解したり、話したりする時に使う知識を、規則の形に書き下ろしていたのです。しかし、規則が増えてくるにつれて、取り扱う情報が非常に大きなものとなり、全体としての見通しが大変悪くなります。そのため、このような人手に頼る手法だけでは、処理しきれなくなります。規則がどんどん増えていくと、一人の人間が全体を理解する

Learning from the past 6

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/>

ことが難しくなり、その結果、「どうしてこの訳がでてきたのか?」という素朴な質問にシステム開発に直接関わった人ですら答えられないということになります。直感に基づいて規則を作成している以上、どの程度の量の情報を集めれば、翻訳には十分なのかということに対する理論的なきちんとした答えはないのです。

また、規則は単に集めれば良いというものではなく、その規則をその時々に応じて適切に使うことが必要となります。つまり、機械翻訳に必要とされる知識には、規則(言語情報)そのものと、ある時点で適用できる規則が複数あった場合に、それらのうちのどれを使えば良いかを定める適用性の情報(優先順位付け)とがあることとなります。これらを区別して、それぞれを適切に表現できる枠組が確立していないことも、見通しを悪くする原因となっています。

現実のさまざまな文書を翻訳することの困難さはこのような問題点を解決し、きちんと使えるようになった規則化された知識が不足していることから来ているわけです。しかしながら、実際に世の中に存在する膨大な文書を対象に、その翻訳に必要とされる知識を人手で抽出し整理することは、ほとんど不可能でしょう。「それではどうすれば良いのだろうか?」この問いに対する答として、計算機に大規模なコーパスを直接処理させようという考え方が提案されました。

この手法の提案は、10年前に遡る[1, 2]ののですが、当時は大規模なコーパスを解析するのに十分な能力を持った計算機は大変高価なものでした。最近の計算機の高速度・低価格化により、大量のデータを比較的容易に処理することが出来るようになってきたため、事例を利用した(あるいは統計的手法を用いた)自然言語処理と呼ばれるこの手法が脚光を浴び始めたのです。

この手法の機械翻訳への応用としては、まず、事例を用いて翻訳を行なうシステム(Example-based Machine Translation)があげられます。

例を使って説明しましょう。play という英語の単語を日本語に翻訳する場合には、play の目的語に応じて、

「(スポーツを)する」「(楽器を)演奏する」など、適切な訳語を選ばなくてはなりません。これを従来からの規則を用いた手法で翻訳する場合は、

”play スポーツ” → “スポーツをする”

“play 楽器” → “楽器を演奏する”

といった訳し分け規則と

piano	(楽器) ピアノ
violin	(楽器) バイオリン
tennis	(スポーツ) テニス
golf	(スポーツ) ゴルフ

といった、辞書(訳語と意味素性)とを用意することになります。この程度の情報の場合は、どちらも非常に簡単で分かりやすいのですが、一般の雑多な文章をきちんと翻訳するためには、もっと多くの情報を記述することが必要になり、規則を表現する枠組も、非常に複雑なものになります。

一方、事例を用いた翻訳システムの場合には、対訳とシソーラスを準備します。対訳として、

I play tennis.	私はテニスをする。
I play the piano.	私はピアノを演奏する。

シソーラスとして、

楽器
ピアノ
バイオリン
トランペット
スポーツ
球技
テニス
ゴルフ
卓球

というものが与えられていたとしましょう。

ここで、"I play golf."という文が入力された場合、ゴルフが、テニスとピアノのどちらに、より似ているかをソーラスを使って調べます。この場合には、明らかにテニスの方に似ていますから、システムは、"I play golf."は"I play the piano."よりも、"I play tennis."の方に似ていると判断し、「テニス」を「ゴルフ」に置き換えて、「私はゴルフをする。」という訳文を生成します。

このようなシステムにおいては、対訳コーパスを検索し入力文に最もよく似た例を選び出す処理が一番重要な部分です。各単語の「似ている度合」を計算するためには、ソーラスを使う場合が多く、ソーラスの木構造の中での単語間の距離を使って計算します。

ただ、この例のように、単語が1つだけ異なっているような文(対訳例)がコーパスの中に含まれている場合は、現実の文章の翻訳では稀でしょうからほとんどの場合、適当な対訳を組み合わせて、一つの文を作ることが必要になります。このためには、入力と例との間で一致していない部分について、再帰的にコーパスの検索を繰り返す手法[3]等が提案されています。

事例を用いた手法の機械翻訳への応用としては、このようにコーパス中の対訳を使って直接翻訳をするものばかりではなく、事例から知識(あるいは統計的確率)を自動的に獲得しようとする試みも行なわれています。コーパスを使って、翻訳に必要な規則を作り出すわけです。この場合、原文を何らかの形で解析する必要があるわけですが、当然、完璧な解析は出来ません。誤った解析結果が含まれていたり、解析自体に失敗したりすることもあります。

しかし、十分に大きなコーパスを用いていけば、全体としては、間違った解析結果よりも、正しい解析結果がたくさん含まれることになり、意味のある情報を取り出すことが出来ます。たとえば、簡単な構文解析によって、文中の動詞に係る名詞を見つけ、動詞と名詞の共起の頻度を調べることにより、動詞の持つ格フレームを自動的に獲得する研究[4]などが行なわれています

また、自動翻訳を目的とはせずに、翻訳支援システムとしての使用を目的とするものもあります。翻訳したい文に類似した用例を検索し表示することにより、翻訳支援を行なうのです。ユーザが翻訳したい文を入力すると、システムはそれと類似した文を対訳コーパスから選びだし、その対訳と合わせて表示します。ユーザは、それを参考にして、自分が翻訳したい文の訳文を作成するのではありません。もちろん、類似した文をどうやって選び出すかが問題となるわけですが、あらかじめ形態素解析されたコーパスを用いるものや、前処理をされていない対訳コーパスを対象に、文字列の類似度を効率的に計算する手法[5]などが提案されています。

事例を利用したシステムの特徴として、ユーザがコーパスに対訳を追加するだけでシステムの能力を向上することができるということがあげられます。システムの内部構造や自然言語処理に詳しくない人でも、対象とする言語を知っているならば、正しい対訳を追加していくことによって、自分の言語知識に沿って、システムの改良を行なうことが出来るのです。もし、あなたが、play という単語を翻訳する時に、「演じる」と訳した方が良い場合があると知っているならば、上で示した例に、

I play Romeo.          ロミオの役を演じる。

という対訳を付け加えれば良いのです

一方、規則に基づく手法では、システムを改良するために規則を追加する場合には、まず規則の記法を理解してはなりませんし、規則を一つ追加したことによってシステム全体が崩壊してしまったりしないように、注意しなくてはなりません。例えば、"a→b"という文法規則と"b→a"という文法規則を入れたような場合、解析時にループを作ってしまうために、システムが暴走してしまうことがあります。事例を用いる手法では、用いるコーパスが十分に大きければ、一つの対訳を追加することによって、そのような崩壊が

起こることはありません。ゴルフは、テニス、ピアノ、ロミオのうちでは、やはりテニスに一番近いでしょうし、"I play the violin."という入力に対しては「私はバイオリンを演奏する。」と翻訳してくれるでしょう。

勿論これは、システムの性能が簡単に向上するということでは、必ずしもありません。コーパスを大きくしていくことはそれ自体、大変な労力が必要なものです。しかし、このようなコーパスは、個々のシステムとは独立したものであり、一旦作られたコーパスは他のシステムでも用いることか出来ます。あるいは、既に何らかの形で存在する対訳集を使うことも出来るでしょう。

このような、事例を用いる手法は今後一層の発展が期待される分野ですが、それと同時に、既に規則化されている情報が使える場合には、それを利用した方が効率的な場合も当然あるわけです。したがって、今後は、規則を用いた手法と事例を用いた手法とのそれぞれの特長を活かして、両者を融合することが検討されていくでしょう。そのようなシステムでは、言語処理の各時点で利用できる情報のうちで、最も特定化されている情報を用いるという方針で処理が進められることとなります。このような手法は人間の言語処理過程と類似しており、高速な自然言語処理が可能となります。

機械翻訳においては、たとえば、完全に一致する例が見つければ、それを用いて翻訳を行なう。なければ、類似した用例を利用する。それも困難な場合には、規則に基づいて解析・生成を行なう、というようになるでしょう。先ほどの例でいうと、"I play tennis."という文が入力された場合には、完全に一致する文がありますので、そのまま対訳中の訳文が出力されます。挨拶文の翻訳などで、このようなことがよく起こると思われます。また、"I play golf."という文が入力された場合には、すでに述べたように、事例"I play tennis."を使った翻訳が行なわれます。さて、もし、コーパスに疑問文が含まれていなかった場合に、"Do you hate sleep?"という文が入力された場合には、(少なくとも疑問文固有

の部分については)文法規則を用いて解析し、訳文を生成します。もちろん、疑問文を解析する規則をシステムが持っていないとはなりません。

マニュアル等の翻訳では、一旦翻訳した文書の修正版を翻訳することが、しばしば起こります。そのような場合には、以前に翻訳したものをできるだけ使いたいと考えるでしょう。この手法は、そのような状況に対応できることにもなります。

このように考えてきますと、従来からの規則を用いる手法も、やはり重要であるといえます、これまで自然言語処理の研究においては、自然言語の持つさまざまな曖昧性を計算機が正しく理解し、それを解消するためには、意味情報あるいは文脈情報が必要であるということが強調されてきました。そのために、表層の言語現象を詳細に検討するという立場が軽視されていた点があります。その結果、簡単なシソーラス程度の知識と構文解析によって構文構造をかなり正確に把握できるような場合があるにも関わらず、それを意味に関わる困難な問題として放置しておいた面も見られます。詳細な、そして当然膨大な意味情報を用いる意味解析は、もちろん重要な技術ですが、人間が文章を読む場合には、文の内容を完全には理解していなくても、文章の構造(並列句や係り受けなど)をある程度、判断できる場合があることから、人間は言葉の理解過程において、「浅い」意味を有効に用いているものと思われます。このような点に注目した研究の例としては、従来、深い意味処理の対象とされていた名詞句あるいは連用中止による述語の並列構造を、文節同士の類似性を計算することにより、かなりの精度で決定できることを示したもの[6]などがあげられます。

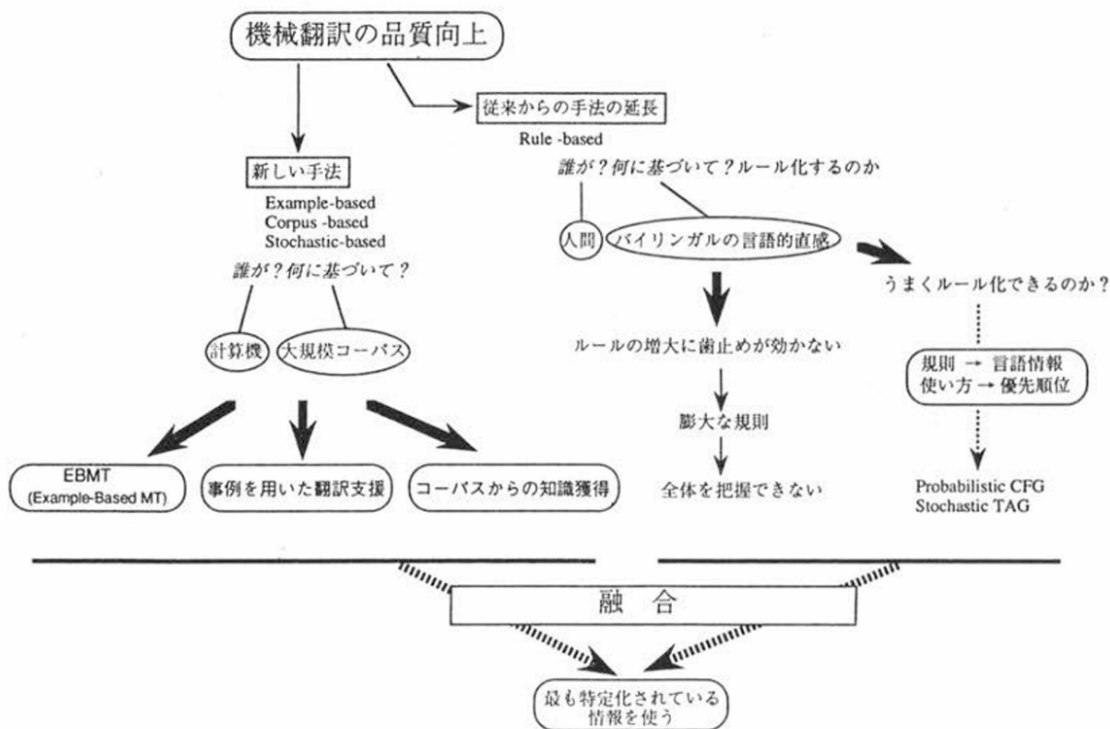
なお、自然言語処理における困難な課題を避ける手段の一つとして、取り扱う領域を限定しようという考え方もあります。この場合、実際には、領域を限定することによって、取り扱う知識の範囲だけではなく、取り扱う言語表現(あるいは言語能力)をも制限していることとなります。特定の場面で発話される文は、そ

の言語で可能なすべての言語表現の中では、ほんの一部となります。そのような制約があることによって、辞書中に表層の情報に対応した知識を書き込んで行くことが可能となります。自然言語処理において、辞書の記述が主要な知識源であることは、どのような手法を用いる場合にもいえることであり、現実の文章から着実に辞書記述を改良して行くことは、最も重要な作業の一つです。

《参考文献》

[1] M. Nagao. Some Rationales and Methodologies for Example-based Approach. Proc. of International Workshop on Fundamental Research for the Future Generation of Natural Language Processing, 1992.  
 [2] M. Nagao. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. In A. Elithorn, ed. Artificial and Human Intelligence, Elsevier, 1984.  
 [3] 佐藤理史, MBT2: 実例に基づく翻訳における複数翻訳例の組み合わせ利用, 人工知能学会誌, Vol. 6, No. 6, 1991.

[4] R. Grishman, et al. Combining Rationalist and Empiricist Approaches to Machine Translation. Proc. of Fourth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92), 1992.  
 [5] S. Sato. Example-based Translation Approach. Proc. of International Workshop on Fundamental Research for the Future Generation of Natural Language Processing, 1991.  
 [6] M. Nagao, et al. Dynamic Programming Method for Analyzing Conjunctive Structures in Japanese. Proc. of COLING'92, 1992.



【1994年4月 AAMT ジャーナル No.7 より】

## ヨーロッパの現在の MT 活動

UMIST・KIT ラルフ・スタインバーガー

1992年末をもって、10年間にわたりヨーロッパ連合(EU)が膨大な努力を注いだ MT 研究(EUROTRA)終わったが、現在さえ、TRADE と CAT2-EDS などのヨーロッパにおける MT 活動のかなりの部分は EU 主導であり、資金供与を受けている。これらの EU 関連の活動の他に、ヨーロッパのいくつかの企業が商業システムの開発に投資している。主だった企業プロジェクトとは、METAL(ドイツ・シーメンス社開発)、EUROLANG(フランス・ソフトウェアハウスの SITE 開発)などである。どちらのシステムにも、従来の第2世代技術が多かれ少なかれ使用され、非常に精巧なツールとの併用により、翻訳者の仕事をより効率化する。新しく従来とは異なるアプローチを採用するのは VERBMOBIL プロジェクトで、これはドイツ連邦政府の開発技術庁 BMFT(Bundesministerium fuer Forschung und Technik)のスポンサーで開発されたものだ。

VERBMOBIL は、音声言語翻訳を必要とするバイリンガルユーザの支援ツールを目指す。そのほかの MT 活動には、PaTrans (デンマーク EUROTRA ベースのインハウスシステム)、欧州シャープが開発する多言語商業ソフトウェア(英国・英語、ドイツ語、フランス語、日本語)ならびにマンチェスター大学で進められる MEG(英国・UMIST)、これはエグザンプルベース技法を用い、雪崩情報を英語、フランス語、ドイツ語およびイタリア語に翻訳する。

### 有名なヨーロッパの MT プロジェクト

EUROTRA は、10年以上かかってヨーロッパ共同体(EC)の12カ国すべての参加を得た共同作業を1992年に終えた。のべ200名の研究者が平行して研究を行い、37,500万 ECU(ヨーロッパ通貨単位)をかけたこのプロ

ジェクトの成果発表は、1992年にフランス(ナント)で開かれたコンピュータ言語学の国際会議 COLING で提示された。

EUROTRA は、多言語のルールベース MT の将来性とリミットについて、統計的手法とトランスファベース手法を使用して調査するという目標があり、非常に野心的な研究プロジェクトだった。

EUROTRA が負った政治上の制約とは、平等性を理由に、EC加盟の12カ国すべてがかかわらなければならなかったということである。そのため、9つの公式ヨーロッパ共同体言語すべて(英語、ドイツ語、フランス語、オランダ語、デンマーク語、スペイン語、イタリア語、ポルトガル語、ギリシア語)という、およそ管理可能な範囲を越えた数の言語を同時に扱う必要があったのである。これら9つの言語を各々ソース言語とターゲット言語として用いると、言語対は72、翻訳方向が144にのぼる。これに対して、商業翻訳ソフトウェアの大部分が扱うのは2つの言語であり、ソフトウェア開発者は単一の翻訳方向だけを提供する場合が多い。

最終的研究成果評価委員会(英国人ブライアン・オークレ座長)が設定した EUROTRA の主要な成果の1つとして、ヨーロッパ全域に数百人の専門 NLP 学者を組織化し訓練するプランがあった。具体的な目標とは、それまで NLP に関する経験が浅い、もしくは全く持たないポルトガル、スペイン、アイルランド及びギリシアなどの国へ、NLP 知識を移植するというもので、これは首尾よく達成された。

前出の評価委員会によりこれまでに EUROTRA が蓄積した資料と経験について、さらに多くの実用化を図るため、より小規模のフォローアッププロジェクトの開始が示唆された。ECでも、EUROTRA その他の研究開発プロジェクトにおける学界と産業界の両方にかかわる多国籍プロジェクトを共催した経緯から、この提

案を認めている。プロジェクト参加企業に各社の研究開発費用の 50%以上を分担させることで、EU は、アプリケーション指向の研究を促し、かつ、企業側がアカデミックなノウハウから利益を得させようとした。

多言語アクションプラン(MLAP)の予算枠によって、EU は 2 つの MT プロジェクト、TRADE と CAT2-EDS に共同出資した。これらはいずれも 1994 年に 2 年計画で開始されたもので、研究課題は、EUROTRA タイプのシステムからプロトタイプを開発し、より少数の言語対について、さらに制限された言語領域(サブランゲージ)を用いる場合も、この方式でリーズナブルな翻訳結果が得られると示すことである。

トランスレーションデモンストレータ(TRADE)は、専門化した社会保障ドメインの英語とスペイン語とイタリア語のテキストを、他の言語(6 つの翻訳方向)のいずれにも翻訳する。TRADE はスペインのソフトウェアハウス Centro de Calculo de Sabadel が、コーディネーションを行い他のパートナーとして UMIST(英国・マンチェスター大学)と Fundacio Bosch Gimpera(スペイン・バルセロナ)と Gruppo DIMA(イタリア・トリノ)と INPS(イタリア社会保障機関)である。E スターと呼ばれる Gruppo DIMA 開発のもので、EUROTRA から派生した形式を採用している。

TRADE は、3 つの言語に対し、EUROTRA 文法と辞書の最適化されたバージョンを利用する。既存の資源の再利用の他に、辞書ツールの利用で、専門辞書を各言語あたり 5000 単語のデモンストレーションサイズに拡大できる。その他の機能として、長文を分割するテキスト・セグメンテータなど、ユーザインタフェースが開発される予定だ。さらには、形式は EUROTRA と異なり、不当な入力や、未知語を含む文をも扱えるように強健になる。

MLAP 枠で資金供与を受けた第 2 のプロジェクトは、CAT2-EDS と呼ばれ、ドイツの会社エレクトロニック・データ・システムズ(EDS)がコーディネーションし

た。このうち CAT2 は、EUROTRA に、参加した「応用情報科学研究所」(IAI=ドイツ・ザールブリュッケン)によって開発されたもので、やはり EUROTRA ファミリーの一つである。その他、UMIST(イギリス)とフランスパリ第七大学のタラナ研究所が開発に参加している。TRADE 同様の機能を備えるものとなるだろうが、CAT2-EDS チームは 1995 年中には、ドイツ語・英語・フランス語相互の MT として、オペレーティングシステムメッセージとソフトウェアマニュアルの翻訳専門の、プロトタイプを提示するだろう。

より少ない言語対を対象として使用する場合、また専門化された分野のサブランゲージに適用するとき、EUROTRA のような形式が成功するという点は、最近、デンマーク語・英語の MT システム PaTrans によって、示された。PaTrans は実際に EUROTRA 形式を基礎に開発され、デンマーク語と英語の文法を用いた。TRADE や CAT2-EDS と異なり、PaTrans は、デンマーク言語技術センター(デンマーク・コペンハーゲン)が単独で開発した。1 言語対だけを翻訳対象として、原言語は英語に制限している。

PaTrans の研究資金は EU 供与ではなく、デンマークの特許翻訳企業 Lingtech および開発者である CST の自己資金である。2 年の開発期間の後、1993 年の終わりにインハウスの MT システムを配布した。開発者とクライアントの両方が翻訳結果に満足したことから、サブランゲージへと研究対象を拡大する計画である。

これらのアプリケーション指向プロジェクト開始の他に、EU は、中・長期の目標として、ヨーロッパ圏内の MT ソフトウェアその他の NLP 応用技術の開発を上げている。これらの目標とは、共同開発を通じて、9 つの EU 公用語すべての標準化を図り、言語資源を作成すること、および、それらの取得・生成・適用およびそれらを多くの目的のために応用するためのソフトウェアツールの開発を含む。この言語学的インフラは、EU 圏内のすべての企業と研究機関の利用に供することで、研究者と開発者のために非常に貴重な資源

となるだろう。現在の EU における MT 関連の諸活動の成果が、時間とお金の節約を助けるだけにとどまらず、さらに将来の文法の開発と、辞書及びツールの再利用を促すよう、望まれる。EU におけるこの方向の貢献の例として、ALEP と LS-GRAM の 2 つを取り上げたい。

ALEP(高度言語エンジニアリング・プラットフォーム)は言語対象をトレイシングしデバッグし見るためのツールを持つ形式で、グラフィックのユーザ・機械インタフェースとテキスト処理システム他、すべての NLP 開発に必要なツールがある。

1991 年以来開発を支持する EC 委員会(CEC)が ALEP に期待するのは、堅固かつ携帯可能で広く実用可能なソフトウェア・プラットフォームの確立であり、専門化されたアプリケーションのプロトタイプの開発にこれを提供することだ。CEC では、産学一致してこのプラットフォームを使用し、多くのさまざまな NLP アプリケーションを開発すれば、それら研究開発の結果は、ヨーロッパ全域でコンパチブルかつ再使用可能な状態となるという理想像を描いていた。すなわち、この共通のプラットフォーム利用を通じて、学界および企業の研究機関の間にシナジーを助成するわけである。ALEP の最終バージョンは今年中に実用化されるだろう。

ヨーロッパでは、完全な NLP 研究のインフラを確立するため、もう一つの主要なプロジェクト LS-GRAM(EC 言語のための大規模文法)が進められてきた。研究期間 2 年のこのプロジェクトは、9 つすべての公用語の大規模なドキュメンテーションを含む、コア文法をヨーロッパ NLP 学界へ提供することを目指した。文法は ALEP を用いて書かれ、それらは「基本セット」として ALEP を使用するすべての研究開発機関に分配される予定だ。ALEP とコア文法の利用が、ヨーロッパ全域の産学協同をさらに助長し、またそれがアプリケーション開発に対しても、トレーニングや研究、実用化を意識したものを生む基盤として魅力あ

るものとなるよう、望まれている。1994 年開始の同プロジェクトの中心となった「応用情報科学研究所」(IAI = ドイツ・ザールブリュッケン)はじめ、エセックス大学(イギリス)、Fundacio Bosch Gimpera(スペイン・バルセロナ)スツットガルト大学(IMS-CL・ドイツ)が参加した。1993 年までに EC が言語学関連のプロジェクトに対して行った資金供与は総額 1,5 億 ECU に上る。内訳は、MLAP(1977 年～)、EUROTRA(1982～92 年)、LRE(1991 年～)、そして ESPRIT(1984 年～)関連である。これらの諸活動は、1994 年から 98 年にかけて「第 4 次基幹プログラム」に編入され、予算規模 4 億 ECU を投入して継続される。

しかしながら、EC の一般的原則として、加盟諸国諸地域が独力では行えない研究活動、すなわち標準化の開発と各の研究努力の連携を図ることに助成対象が決められている。商業的指向の強いプロジェクトは、企業自身が資金供給しなければならない。ここで取り上げる主要な MT プロジェクトは、既存の METAL システム(ドイツ企業シーメンス開発)EUROLANG(フランス中心の商用プロジェクト)およびドイツの国家研究プロジェクト VERBMOBIL である。

METAL の開発は 1960 年代半ばにアメリカ・テキサス州のオースティン大学で始まった。1978 年より、同プロジェクトはドイツ・ミュンヘンに本拠を置くシーメンス社の支援を受け、80 年には同プロジェクトは同社に移籍された。同社はドイツ語一英語の市販版をリリースしたのは 89 年になってからである。また 92 年まで、同システムは稼働システムに限られており、記号リスマシンもしくはシーメンス社のワークステーションをソフトウェアと同時に購入するほかなかった。そのため、価格的にも 25 万 DM(ドイツマルク=約 1,500 万円)となり、販売実績は 20 台以下であった。

その後、シーメンス社は翻訳対象言語対を拡大、英語、ドイツ語、フランス語、オランダ語、スペイン語、

デンマーク語を加えた。また、SUN ワークステーションに移植、現在では約 5.5 万 DM(約 330 万円)で販売している。

METAL の特徴は、フルオートマチックな翻訳の提供に加え、ユーザに便利なツールを多く用意した点だ。大型の専門辞書、セミオートマチックな辞書アップデート機能、また後処理のための機能などが含まれている。ユーザにとっては、図表を含む入力文の書式が保存されるため、翻訳結果が入力文と全く同じ書式で得られる点が、時間短縮の大きな味方である。

シーメンス社は、フランスの SITE 社が率いる EUROLANG の開発にも参加しているが、このプロジェクトには 10 以上のヨーロッパの企業および大学が名前を連ねる。1992 年～1994 年の見込み予算は 7 億 3500 万 ECU である。1994 年末までに SITE は、ドイツ語、英語、フランス語、イタリア語、スペイン語(その他多くの言語が計画)の全自動 MT のみならず、完璧な「翻訳者のためのワークベンチ」の完成と提供を計画している。全自動翻訳は、フランスの Ariane (GETA)システムと METAL システムの両方で行われる。

METAL がもたらすツールやその他の利点以外にも、EUROLANG は、人間翻訳の支援として、理想的なワークベンチの役割を担う。一般辞書、専門辞書ともオンラインでいくつも提供する上、用語管理ツール、あるいは前回翻訳したテキストを検索するツールを提供できるからだ。特に後者は仕事のむだな重複を防ぎ、例えば修正やデータ更新を行った文書の翻訳には有効である。用語管理ツールはデータベースとして同システムを利用する複数の翻訳者に利用され用語統一に役立つ。EUROLANG は、SUN ワークステーション上を走るだろう。

もう一つ、より研究指向の高いプロジェクトとして VERBMOBIL がある。これはドイツ人工知能研究センター(DFKI=Deutsches Forschungsinstitut fuer Kuenstliche

Intelligenz)とって、ドイツ政府(BMFT=Bundesministerium fuer Forchung und Technik)が資金供与するものだ。VERBMOBIL の目標は大胆で、対面型の会話のために、携帯用のドイツ語英語日本語翻訳支援のプロトタイプを作成することである。2 年の準備研究の後、プロジェクトは 1993 年に公式に始まり、2001 年終了を計画している。

VERBMOBIL はさしあたり国家プロジェクトであるが、アメリカと日本人のパートナーとの共同研究が計画されている。日本では、同様のプロジェクトを行っている京都の ATR 音声翻訳研究所との協力が検討されている。アメリカでは、カーネギーメロン大学、スタンフォード大学の CSLI 及びバークレイの国際コンピュータ科学研究所(ICSI)との共同研究が準備された。ドイツ国内での研究提携先は、シーメンス社、ダイムラー社、テレフンケン社及び 9 大学が含まれる。

VERBMOBIL は、電話の会話の自動逐次翻訳システムとは規定されていない。むしろ、面と向かってともに英語を話す日本人とドイツ人に対して、会話の細かい表現に困ったときに支援するものである。この二人が何か英語表現で問題を感じたらすぐその場で VERBMOBIL を動かすと、日本語かドイツ語の単語の英訳を即座に与えられるわけだ。これらは整った文の形である必要はない。翻訳プロセスは増加つまりシステムは直ちに荒翻訳を提供することができるし、もっと時間があるなら、さらに洗練された翻訳を代わりに提供することもできるのだ。

最初のデモンストレーション版は 1995 年リリースの予定だが、その会話のシチュエーションは、2 人のビジネスパートナーの会合という設定で、目の前にカレンダーを広げて次のミーティングをいつ行うか話し合っている。このバージョンから 2001 年の完成版までに、アプリケーションドメインは徐々に広げられるだろう。

このプロジェクトの主要な目標は非常に野心的で音声の認識と合成、ダイアログと知識処理及び機械翻訳など、高度な情報テクノロジーと自然言語処理の各分野を、統合化することである。最初の4年で、BMFTの資金は6,000万DMに達した。また VERBMOBIL へのドイツの総合的な投資は、たぶんプロジェクト終了までに12,000万ECUに届くだろう。

これらのプロジェクトの他に、ヨーロッパではNLPの分野で多くの他の活動が行われている。多くのアプリケーションは、研究成果を即座に実用化する方向であり、例えば文法と文型のチェッカー、自動テキスト分類機能及び知的情報検索などが上げられる。しかしながら主にEU主導だが、多くは中・長期的な目標達成を目指している。将来的な投資として、大きいコーポラや他の言語学のデータ収集、評価テクノロジーの開発、既存の資源の再利用性の検討、コーポラからの知識取得の自動化、その他を含むアクティビティが行われている。ヨーロッパには、言語障壁に打ち勝つことにおいて経済的、政治上の大きな関心がある。いくつかの不可避な退行があったにもかかわらず、健全なNLPインフラの構築を目指し、協調して努力することを通して、ヨーロッパ諸国はこの問題を解決しようとする強い姿勢を示している。

## イベント報告

## 第 12 回アジア翻訳ワークショップ (WAT2025) 開催報告

中澤 敏明  
東京大学

## 1. はじめに

本稿では WAT2025[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、2025 年の第 12 回 (WAT2025) はインドのムンバイで開催された IJCNLP-AAACL2025 の併設ワークショップとして、2025 年 12 月 24 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。参加者は現地参加が 25 名、オンライン参加が 11 名で合わせて 36 名程度であり、盛況であった。

## 2. 研究論文

WAT2025 では 4 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。

- Segmentation Beyond Defaults: Asymmetrical Byte Pair Encoding for Optimal Machine Translation Performance
- Speech-to-Speech Machine Translation for Dialectal Variations of Hindi
- A Systematic Review on Machine Translation and Transliteration Techniques for Code-Mixed Indo-Aryan Languages
- CycleDistill: Bootstrapping Machine Translation using LLMs with Cyclical Distillation

この 4 件の中で、査読の評価が最も高かった “Speech-to-Speech Machine Translation for Dialectal Variations of Hindi” を best paper として選出した。

## 3. 招待講演

招待講演は University of Surrey の Senior Lecturer である Diptesh Kanojia 氏より “Optimizing Large Language Models for Low-resource Quality Estimation” というタイトルで行われた。大規模言語モデル (LLM) を用いた機械翻訳の品質評価 (QE) に関する課題と、その解決策に関する講演であった。以下にその概要を示す。

LLM は多くの自然言語処理タスクで高性能だが、機械翻訳の QE、特にデータが少ない「低資源言語」の評価は苦手としている。この原因の一つは、複雑な単語構造を持つ言語におけるトークン化(テキストの分割)の不一致である。この課題を解決するために、2 種の手法を提案した。一つ目は評価のガイドラインに基づいた推論ルールを、直接プロンプト(コンテキスト内)に組み込む手法である。もう一つは LLM の最終層が持つ生成特有のバイアスを避けるため、Transformer の「中間層」に LoRA (低ランクアダプター) を接続する方法である。結果として、既存の最高性能モデル (COMETKiwi など) や通常のファインチューニングを上回る精度を達成した。

また応用として、この高精度な QE データを LLM のガイドとして活用することで、LLM がより正確に翻訳の修正を行えることを WMT のタスクで実証した。これにより、人間の正解データから逸脱することなく、より自然で流暢な出力へと改善できるようになる。

#### 4. 翻訳タスク

WAT2025 では日英特許請求項翻訳タスク[2]、日英文書レベルニュース翻訳タスク[3]、英語からインド諸語へのマルチモーダル翻訳タスク[4]の3つのタスクに参加者があった。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させていただきたい。

この中で日英特許請求項翻訳タスクは今年度から新たに実施されたタスクである。ニューラル機械翻訳(NMT)やLLMの性能は劇的に向上しているが、ドメインが異なると精度が不安定になるなど、万能な自動評価手法は未だ存在しない。特に特許文書(特許請求項)の翻訳は、独特の文体や長さゆえに用語や一貫性の評価が困難である。

そこで、正確な自動評価手法の開発を最終目的とし、日英の特許請求項翻訳の翻訳タスクを実施した。第1回となる今回は、多様な翻訳出力を収集して人手で誤りをアノテーションし、将来の自動評価モデル開発に向けた学習データを作成することを主目的とした。

本タスクの実施結果は NLP2026 において日本語でも発表しているので、興味のある方はぜひご覧いただきたい[5]。

#### 5. まとめ

本稿では WAT2025 全体の概要を報告した。論文の投稿数や翻訳タスクの参加者は一定数あり、また招待講演に関しても多くの質疑があり、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。2026 年度は 11 月に中国の珠海市、横琴島で開催される AACL2026 のワークショップとして採択されている。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. 2025. Proceedings of the Twelfth Workshop on Asian Translation (WAT 2025). Association for Computational Linguistics, Mumbai, India.
- [2] Toshiaki Nakazawa, Takashi Tsunakawa, Isao Goto, Kazuhiro Kasada, Katsuhito Sudoh, Shoichi Okuyama, Takashi Ieda, and Masaaki Nagata. 2025. Findings of the First Patent Claims Translation Task at WAT2025. In Proceedings of the Twelfth Workshop on Asian Translation (WAT 2025), pages 1–15, Mumbai, India. Association for Computational Linguistics.
- [3] Naoto Shirai, Kazutaka Kinugawa, Hitoshi Ito, Hideya Mino, and Yoshihiko Kawai. 2025. Findings of the WAT 2025 Shared Task on Japanese-English Article-level News Translation. In Proceedings of the Twelfth Workshop on Asian Translation (WAT 2025), pages 93–97, Mumbai, India. Association for Computational Linguistics.
- [4] Shantipriya Parida and Ondřej Bojar. 2025. Findings of WAT2025 English-to-Indic Multimodal Translation Task. In Proceedings of the Twelfth Workshop on Asian Translation (WAT 2025), pages 103–108, Mumbai, India. Association for Computational Linguistics.
- [5] 中澤敏明, 綱川隆司, 後藤功雄, 笠田和宏, 須藤克仁, 奥山尚一, 家田堯, 永田昌明. 2025. 特許請求項機械翻訳およびその人手評価の課題と挑戦 ~ WAT2025 特許請求項タスクからの知見~, pages 414–4149, 宇都宮, 栃木. 言語処理学会第 32 回年次大会 (NLP2026).

## イベント報告

## AAMT 2025, Tokyo

### ～LLM 時代における翻訳と人間の共進化～

出内 将夫

国立研究開発法人 情報通信研究機構

#### 1. はじめに

2025年12月2日に開催されたAAMT 2025, Tokyoのイベント報告をします。AAMT 2025, Tokyoはオンラインと会場のハイブリッド形式で開催され、パネルディスカッション(図1)、公募ポスター7件の発表、招待講演3件のほか、ランチョンセミナーや企業展示も行われました。講演資料は<https://aamt.info/aamttokyo2025/handout-20251202/>からアクセスできますので、併せてご覧ください。



図1：パネルディスカッションの様子

#### 2. パネルディスカッション

パネルディスカッションでは、「機械翻訳とどうつきあう?—PEガイドラインと実務の声から見えるこれから—」をテーマとして、AAMTのポストエディット(PE)ガイドライン策定に携わった方々を招いて、ディスカッションを行いました。モデレーターはAAMT理事 室田 洋子氏で、パネリストは以下の3名の方です。

◎荒木 慎太郎 氏(株式会社カルテモ 取締役 事業推進部 部長 翻訳品質管理責任者)

◎上田 有佳子 氏(ネットアップ合同会社 APAC グローバル化戦略統括)

◎大村 雅之 氏(MSD株式会社 メディカルアフェアーズ シニアスペシャリスト)

#### ◎ディスカッション&質疑

室田氏：

AAMTでは、機械翻訳(MT)の品質管理に役立つPEガイドラインを策定した。このディスカッションでは策定に関わった3名の方々をお招きした。まず、それぞれの自己紹介とMTとの関わりをお話してください。

荒木氏：

PEに取り組む会社で品質管理責任者を務めている。自分自身がPEネイティブの翻訳者であり、プロジェクトの様々な工程、PEの教育にも携わっている。

上田氏：

30年くらい前からMTに関わっており、MT結果の修正が必須だった時期から、MTのまま公開することも多くなった現在まで、海外ベンダーのローカリゼーションに関する業務に携わっている。

大村氏：

製薬企業において、薬の医学的情報を取り扱う部門で、ローカリゼーションを含めた、社内の効率化・デジタル化に携わっている。

室田氏：

3名それぞれが、PE ガイドライン策定に携わったモチベーションは何ですか？

大村氏：

利用者の立場から PE ガイドラインが分かりやすくなるように、と思い参加した。参加しているうちに知識が付き、本当に利用者にとって分かやすいかどうかは不明だが、ご意見をいただきながら改善していければ良いと考えている。

荒木氏：

自分自身が PE の受託者や PE の講師という立場で活動する中で、PE や MT に対するリテラシーに差があると感じた。PE ガイドラインをまとめることで、業界や発注者が PE や MT を正しく理解できると良い、と思い参加した。

上田氏：

PE ガイドライン策定の話をお願いした時は、各社独自に作るものではないか、とも感じた。しかし、これから始める企業もあることや、LLM の導入などを考慮すると、新たな視点での PE ガイドラインが今必要だと思い、業界に貢献するモチベーションで参加した。

室田氏：

PE ガイドラインに期待することは何ですか？

荒木氏：

PE に対するリテラシーの底上げ、PE 案件の顧客とのコミュニケーションに有効なツール、PE の教育の3つの観点で、非常に役立つドキュメントだと考えている。特に教育では、PE 実務のスキルだけでなく、全体のプロセスや自らの役割を理解することも重要である。

大村氏：

発注側が PE を理解せずに翻訳会社に丸投げでは、思うような成果物が得られないと考えている。発注側が PE を理解し、スピード向上やセキュリティ面など、業務で必要な点を考慮する必要がある。

荒木氏：

製薬業の文書において、PE が向く・向かない文書はありますか？

大村氏：

社内文書はライト PE で良いと思う。規制当局への提出資料や医学情報は正確性重視なので、ライト PE は向かない。

上田氏：

荒木さんに尋ねたい。PE ガイドライン冒頭で、PE ガイドラインはスタイルガイドではないと断っているが、スタイルガイドの実態は顧客ごとにどの程度違うのか。

荒木氏：

基本は顧客ごとに大きく異なるが、そもそもスタイルを細かく指定するかどうか顧客ごとに異なる。

上田氏：

PE はある種の哲学だと思う。これから PE を始めよう、という企業は、スタイルガイドを作る前に、是非 PE ガイドラインを一読してほしい。日本語の細部表現にこだわった結果、リリースが遅れるのは、本末転倒である。品質に対する考え方自体も変遷しているので、見直しが必要な可能性もある。

大村氏：

製薬企業も、外向けに出す文書に対してはまだ厳しいが、社内向けの文書だと、MT で作成した文書も散見されるようになってきた。

荒木氏：

スピード重視のスタイルガイドを用意しても良いし、使い分けの視点も重要だと思う。そのような点も、PE ガイドラインを見ていただければ分かると思う。

上田氏：

発注側とのコミュニケーションは今後も大事だと思うし、そのための関係づくりが重要だと思う。

室田氏：

お使いの MT エンジンや選定理由は？

荒木氏：

海外ベンダーの案件では、MT エンジンを指定される

ことが多い。国内の案件では、MT エンジン選定を行うことがある。顧客が好む翻訳に近いかどうか、翻訳支援ツールと相性が良いか、といった観点で案件ごとに最適な MT エンジンを選ぶ。

室田氏：

海外ベンダー案件は MT エンジンを選べない、とのことだが、何種類くらい MT エンジンを使っているか？国内顧客の好みはどのようなものか？

荒木氏：

使っている MT エンジンが明かされないことも多く、カスタマイズなどを含めると数は分からない。

好みは、例えば、文章が長めと短めのどちらが良いか、顧客の業態に応じた文体など。特定の MT エンジンを推奨ということはなく、最近はどれも実用レベル。

大村氏：

2016 年くらいに Google 翻訳がすごく良くなり、その頃出てきた医薬専門翻訳エンジンが素晴らしく、導入したことがきっかけ。更に用途ごとに用語や言い回しが異なる点にも対応した MT エンジンが欲しい、となり、2018 年から 2019 年に PharmaTra という特化型 MT エンジンを使用している。ほかに、薬事文書で膨大な情報が頻繁に更新されるようになり、翻訳会社の MCL が作成した MT エンジンと PE を組み合わせている。

上田氏：

弊社では 10 年くらい同じ MT エンジンを使い続けてきたが、現在は言語ごとに MT エンジンを使い分けている。切り替える動機となったのは、通常のテキストの翻訳品質ではなく、タグなどの扱い。自動化を考慮すると、テキストでない部分、例えば画像や翻訳禁止部分、の扱いが重要となる。MT エンジンを再検討する場合は、弊社のコンテンツでテストするフェーズを経て、別の MT エンジンに切り替える。時間とともにコンテンツも変わるので、フレキシブルさは強み。

荒木氏：

タグ関連は翻訳会社として悩まされる部分。そういった部分に柔軟に対応できるソリューションも、今後多様化していくのだと思う。

室田氏：

翻訳のプロジェクトを成功させるために翻訳会社とクライアントとの間のコミュニケーションで大切だと思うことは何ですか？

大村氏：

クライアントにとって、MT ガイドラインは重要なコミュニケーションツールになり得る。クライアントは翻訳会社がどういうプロセスやツールを使って翻訳しているか、知らないことも多い。クライアントとして知っておくべき内容、過去の翻訳資産の何がどのツールでどう活用できるか、を MT ガイドラインに盛り込んだ。MT ガイドラインを通じて、コミュニケーションが充実できると良いと思う。

荒木氏：

翻訳会社はクライアントに対して、翻訳支援ツールや、翻訳メモリ、マッチ率など、翻訳作業に関わる用語を丁寧に説明した方が良いと思う。また、クライアント側も、専門的な知識で見て誤訳と判断でき、それが翻訳者視点では誤訳と判断できない場合に、翻訳者を育てるような長期的視点で、付き合うのが良いと思う。丁寧なフィードバックを出せば、良い翻訳会社であれば改善が見られると思う。

上田氏：

専門的な知識によって誤訳と判断できるが、その知識が無いと誤訳と判断できない場合、原文を見直すきっかけだと思う。ポストエディターが頑張って直して、原文が間違っているのに日本語だけきれいになっている、というのは避けるべきで、本来クライアント側が自分ごととして、対応すべき問題だと思う。

四半期ごとに、翻訳会社とクライアントが資料を持ち寄り、問題点やエラーの傾向を見つつ、MT エンジン切り替えや、プロセス改善を検討できると良いと思う。

大村氏：

公開している WEB サイトに対して読者から誤訳報告をいただくことがあるが、誤訳なのか原文誤りなのか、翻訳会社と一つ一つチェックしている。最近は誤訳が

ほとんどなく、原文がおかしいケースもある。日々の業務の中でフィードバックできるかは課題も感じる。

上田氏・荒木氏：

差分ヤリスト形式で、翻訳会社からクライアントに見やすい形で報告できると良い。

上田氏：

今までは翻訳メモリ (TM) が神、のような考えがあったが、最近はクリーンアップする必要がある、ということも感じている。古い TM で訓練した MT エンジンの翻訳が汚くなることがあった。また、TM を活用したプロセスやワークフローも見直す必要があると思う。今までは TM マッチ率 N% 以下の場合に MT というのが通例だが、スピードを求められる場合に何% だったら確認省略、などを考える状況になってきている。100% 以外をすべて MT にする選択肢もある。コミュニケーションの中で決めていく必要がある。

荒木氏：

フル PE とライト PE と言われているが、コストとスピードに応じて、いろいろな PE のやり方が生まれる可能性がある。例えば、訳文側で整合性が取れば良い、TM 適用部分は差分確認のみで可、など。最近の懸念事項として、PE のプロジェクトで、修正率に応じて単価が変わることがあると聞いた。PE は訳文を修正する仕事ではなく、7 割くらいは正しい訳文であることを確認し品質を保証することが本質だと思う。

室田氏：

今後の展望についてどう思うか？

荒木氏：

PE によって翻訳市場が縮小するのでは、翻訳者の取り分が減るのでは、という話をよく聞くが、私は PE によって今まで翻訳されなかった部分が翻訳されると考えている。例えば、エンタメ系や動画、専門分野の小規模メディアなどである。PE のスキルを身につけている方にとっては、良い時代になってくると思う。

大村氏：

PE ガイドラインがクライアント側にとってより良い

ものになっていくと良いと思う。製薬業界に限らず、スピードが求められつつ、正確性を担保しなければならないシーンは数多くあると考える。

上田氏：

今後、PE において、原文の問題や機械的な修正は、LLM で対処できるようになってくると思う。ローカリゼーション業界を目指す人が減っており、業界が高齢化してきている中、少ない編集量で PE の処理量が増やせるのは良いことだと思う。日本にいろいろな海外の情報を取り入れ、世の中を良い方向に持っていければ良いと思う。

会場参加者：

荒木さんの話で、品質保証では PE の 7 割は結果が正しいことの確認となる話があるが必要な能力は何か。

荒木氏：

PE を実施すると翻訳能力が落ちるという話を聞いたことがあるが、正しいことの確認には翻訳能力が必要。また、多くのワードを処理する中で、スタイルガイドに定義されたスタイルを保てるか、ツールを正しく使えるか、といった能力も重要だと思う。

会場参加者：

必要な能力として、語義、文法、事実の把握、論理の正当性について、翻訳者は 100% を目指すべきと考えているがどうか。

荒木氏：

お客様が必ずしも 100% を求めるわけではないため、顧客の品質要求への適応力も必要と思われる。

### 3. 公募ポスター

公募ポスター発表は以下の 7 件でした。

◎XTM を用いた機械翻訳・PE の実務報告：古河 師武氏（翻訳テクノロジー株式会社）

◎翻訳実務での機械翻訳の活用と成果を翻訳者の体験から紹介：吉川 潔氏（翻訳者）

◎LDX hub を活用したハイブリッド PE の事例：伊澤力氏（株式会社川村インターナショナル）

◎大規模言語モデルは多様な翻訳仕様に追従できるか？：萱野 陽子 氏（総合研究大学院大学）／菅原 朔 氏（国立情報学研究所）

◎機械翻訳出力に対する LLM を用いた後編集：早川 威士 氏（株式会社アスカコーポレーション）

◎生成 AI の時代に翻訳者が主導する新しい翻訳システム：三浦 由起子 氏（医学翻訳者・翻訳エージェント）／津山 逸 氏（医学翻訳者）

◎ポストエディットが翻訳者の仕事満足度に与える影響：阪本 章子 氏（関西大学）

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。

◎XTM を用いた機械翻訳・PE の実務報告：古河 師武 氏

MTPE で作成した訳文と人手翻訳で作成した訳文について、3 種類のブログ記事翻訳を対象に、作業時間・翻訳品質（主観評価）・作業付加（主観評価）を実施した。MTPE は人手翻訳に比べ作業時間を 50%～60%削減でき、翻訳品質は人手翻訳と遜色なく、作業付加は軽減できた、という結果だった。

◎翻訳実務での機械翻訳の活用と成果を翻訳者の体験から紹介：吉川 潔 氏

最新の MT は、長文に追従可能であり、新語の正訳や訳しづらい語句の適訳が入手できるので実務でも有用である。また、翻訳後の訳文を逆方向に MT して、原文と照合する用途で MT を活用している。MT に残存する課題も継続してまとめている。

◎LDX hub を活用したハイブリッド PE の事例：伊澤 力 氏

AI と人手の二段階 PE を行うハイブリッド PE を提案し、AI による PE を反復させ人手の PE で検出したエラーをどの程度改善できたかを調べた。AI による PE は 1 回で人手の PE で対処したエラーのうち 75%を対処でき、3 回以上反復することで 95%を対処できた。

◎大規模言語モデルは多様な翻訳仕様に追従できるか？：萱野 陽子 氏／菅原 朔 氏

企業 IR 資料の MT において、LLM に説得目的や芸術的目的などの翻訳仕様を与えた場合に、LLM の翻訳仕様への追従能力や評価基準の妥当性を、複数の原文・翻訳仕様・モデルに対して検証した。翻訳仕様に基づいた翻訳結果の評価において LLM は人手評価と高い相関を示し、LLM に翻訳仕様を与えて生成した訳文は評価が高い傾向を示した。

◎機械翻訳出力に対する LLM を用いた後編集：早川 威士 氏

LLM を用いて英日翻訳の PE を行い、精度を評価した。MT の誤りを 5 タイプに分類し、LLM への指示を 8 タイプに類型化した上で、6 種類のモデルを用いて評価した。PE の正解率が高かった要因は誤りタイプで、繰り返しや数値誤りは修正が容易、意味的誤り、湧出し、訳抜けは修正が困難、という傾向が見られた。

◎生成 AI の時代に翻訳者が主導する新しい翻訳システム：三浦 由起子 氏／津山 逸 氏

医療系の翻訳では、文書ごとに文の適切な長さや語調が存在し、それらに応じた PE が必要である。発表者自身が作成した医療特化型 AI 翻訳システムが、汎用 MT エンジンである Google 翻訳より上記 PE に適していることを、複数の翻訳例を比較することで示した。

◎ポストエディットが翻訳者の仕事満足度に与える影響：阪本 章子 氏

統計的手法を用い、MTPE と翻訳者の労働生活の関係を分析した。分析の結果、MTPE に前向きな翻訳者は翻訳の仕事始めて間もないことが伺え、MTPE の仕事量増加は翻訳者の労働生活の質と負の相関があり、MTPE に前向きでも翻訳の仕事が続けたいわけではないことが分かった。将来の人材不足が懸念されるため、業界全体で施策を講じた方が良いと考える。

#### 4. 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。

◎翻訳からコミュニケーションへ——生成 AI が変える現場のローカライズ

戸田 広美 氏 (パナソニック コネクト株式会社 デザイン&マーケティング本部 コミュニケーション統括部 特命担当 兼 コーポレート PR マネージャー)

◎AIの進化と人の役割

吉川 健一 氏 (株式会社 BRIDGE MULTILINGUAL SOLUTIONS 代表取締役社長)

◎WEB サービスをローカライズする技術～Localize the Internet を目指して～

上森 久之 氏 (Wovn Technologies 株式会社 取締役副社長 COO/CPO)

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。

◎翻訳からコミュニケーションへ——生成 AI が変える現場のローカライズ：戸田 広美 氏

外資や国内の企業で、マーケティング部門を歴任した MT ユーザーとしての立場から、MT と生成 AI の特長に応じた使い分けについて述べる。パナソニックコネクトでは、2023年2月という早い時期から法人契約した複数種の生成 AI を社員が自由に使える状態にした。生成 AI の使われ方は「聞く」から「頼む」に変化し、社内に浸透してきた状況である。

自身の経験から、ローカライズ業務での MT と生成 AI の活用方法を紹介する。外資系企業では、急ぎ案件で MT と生成 AI の翻訳結果を比較して自身で最終調整することが多かった。日本企業では、MT を活用し業務が高速化しておりネイティブの人でも一次翻訳に MT を使用している。さらには、生成 AI によって人の作業はレビューと調整にシフトしている。残る課題は、セキュリティや機密性の制約で生成 AI が使えない状況や、ニュアンスや文化など表現の調整を要する場合である。

今後は、MT が持つ大量処理や用語統一などの良い特長を基盤として、生成 AI による要約・言い換え・補足説明などを組み合わせ、人間が最終判断して品質保証する形に変わっていくと思う。AI を使いこなすためには人の知識も進化させていく必要があると感じる。翻訳は単なる文章の変換作業から、情報を最適な形で届ける技術、つまり誰にどう伝えるかというコミュニケーションに進化していくと考えられる。

◎AIの進化と人の役割：吉川 健一 氏

通訳サービスを主軸とする多言語ソリューション企業の立場から、通訳業務における MT・AI の役割と活用可能性について述べる。10年前にも AAMT で講演する機会があり、その際に通訳需要を、利用者が多く提携的な対応が中心となるマス領域、一定量の即時対応や双方向のやり取りが発生する領域、数は少ないが高度な専門性を要する領域の3段階に分けた。現在、MT は圧倒的にマスの領域に有効であり、双方向のやり取りもかなりの部分が任せられるようになってきた。しかし、また AI で代替できない部分もある。例えば、感情・衝突のある場面や、医療・法律・危機管理など、機微な翻訳が必要な場合、文化的背景や倫理的判断が必要となるため、AI ではなく人が対応した方が良いと思われる。AI の進化により人の仕事は「専門性」「倫理判断」「共感」「状況判断」が求められる領域に高度化すると考える。

新型コロナウイルスの初動に関係した、日本政府観光局が設置した訪日外国人向けの通訳センターの対応例を紹介する。このセンターでは、オンラインで感染が疑われる問合せがあり、官邸等に連絡した経緯がある。このような判断は人間がやらなければいけないことだと考える。

最近では、滞在期間が短期の旅行者、中期の留学者、長期の滞在者のそれぞれに対して、通訳会社としてどのような事業展開ができるか、という検討も実施している。関係する省庁向けに、インバウンド対策や日本

語教育など、それぞれの所管で必要となる対応を提案している。

#### ◎WEB サービスをローカライズする技術～ Localize the Internet を目指して ～：上森 久之 氏

Wovn (ウォーブン) という WEB サービスのローカライズソフトウェア開発に携わってきた立場から、MT を多言語サービスに組み込む際の考え方や位置づけを整理する。Wovn は、既存の WEB サービスに後付けで多言語対応できる仕組みを提供している。多言語対応を自動化することで、WEB ページの一貫性や新しいデザインへの変更が容易となる。

システム開発分野では、グローバル化と国際化、ローカリゼーションはそれぞれ異なり、G11n、i18n、l10n という区別がある。G11n はソフトウェアを世界対応するための様々な要素を含み、i18n や l10n を包含する。i18n は様々な地域・言語に対応するための基盤で、翻訳はl10n に関係する。翻訳は基本的に AI の翻訳を使っているが、人手翻訳を行うこともある。人手翻訳は AI 翻訳と比べ 1 万倍くらい価格差があるので、範囲や言語ペアを限定して導入し、AI 翻訳は範囲や言語ペアを気にせず導入、となる。また翻訳だけではないカルチャライゼーションには力を入れており、LLM を使うことが増えている。

Wovn の製品思想として、シンプルに多言語対応できることを目指している。バックグラウンドでは 3 人の AI エージェントが稼働しており、1 人目が元の WEB サイトを収集する役目、2 人目がテキストの翻訳や違う国向けのデザインに切り替える役目、3 人目が外国語にしたコンテンツを埋め込むページを作成しドメインなどを国際化対応して公開する役目を担う。翻訳では、基本的に MT の API を呼び出すが、数字などの翻訳不要部分指定や、対訳集自動生成が可能である。MT エンジン開発しない方針としており、MT や LLM について翻訳品質や価格などの評価軸を設けて選定し、ラッピングした上でオリジナルのプラットフォームとしている。世界中で使われる製品にしていきたい。

## 5. おわりに

様々な翻訳目的に応じて、MT のみならず、LLM を組み合わせて翻訳品質を改善し、必要に応じて人が最終的な品質担保を行う、という形が一般化しつつあると感じました。また、その過程を経る中で人はさらに自らの専門性を高めていく、という「LLM 時代における翻訳と人間の共進化」というテーマにふさわしい講演や発表が多く、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。

## イベント報告

## 第9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

2026年2月20日、第9回自動翻訳シンポジウム「AIによる翻訳でジャパンを世界へ」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、石渡 祥之佑氏 (Mantra 株式会社 代表取締役)による基調講演「マンガ自動翻訳の現在地」を皮切りに、菅谷 史昭氏 (マインドワード株式会社 代表取締役 CEO)による「自動通訳の実装と応用の最新状況と可能性」、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)による「生成 AI のメリットを取り込んだ自動翻訳」の講演が行われました。

続いて行われたパネルディスカッションでは、瀬戸 優樹氏 (ヤマハ株式会社 新規事業開発部 SoundUD 室 室長)をファシリテーターに、上述の石渡氏、菅谷氏、隅田氏をパネリストとして迎え、『日本の価値 (コンテンツ・文化・自然)』を伝える自動翻訳・通訳に対する期待」をテーマに議論がされました。

また、会場ホワイエでは企業・団体による最新の自動翻訳製品・サービス等の展示があり、参加者が実際に最新のサービスを体験できるなど、参加者と出展者が熱心に議論を交わす様子が見られました。

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展示から3件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あり、紹介順は五十音順としています。

## 2. 展示会場の報告

## 1) 株式会社時事通信社

**概要:** 同社は、日本語記事の英訳を従来人手で行っていましたが、最近では機械翻訳と LLM による校正を導入して英文ニュースを作成する割合が増加しているそうです。また、自社で作成する記事の他に、提携先である AFP 通信が配信する英文記事を生成 AI による高レベルな日本語訳と共に配信するという新たなニュースプラットフォーム「FASTLOOK」というサービスを開始しました。

**「FASTLOOK」の特徴:** パリに本社を置く AFP 通信から送られてくる1日300超の記事を時間差なく、日本語訳と共に読者へ届けるというサービスです。AFP 通信社は欧州、アジア、北米に加え、中東、グローバルサウスのニュースを幅広くカバーしており、特にアフリカには10か所を超える支局網を有しているため、FASTLOOK では世界各地で起こっている様々なニュースをリアルタイムに日本語で読むことが可能になっています。

**所感・示唆:** 海外メディアが発信する情報に触れると、国内メディアでは詳しく報じられていないニュースや、同じ出来事であっても、国内メディアとは異なる視点で報道されていることに気付くことがあります。昨今の不安定な国際情勢の中で、海外メディアの情報に日本語翻訳付きで触れられることは、情報の偏りを防ぎ、多角的な視点で物事を捉える上で非常に有意義であると感じました。

Report on the 9<sup>th</sup> Symposium on Automated Translation  
Chikako Tamasaki

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/>

## 2) ジョルダン株式会社

**概要:** 電車の経路探索サービスで知られているジョルダン株式会社は、鉄道会社等の公共交通機関を運営する企業向けに、遅延や運休などの運行情報を瞬時に翻訳して配信できる「MovEasy お知らせ情報配信システム」というサービスを提供しています。株式会社みらい翻訳が提供する AI 翻訳機能と連携しており、登録された日本語の情報は英語・中国語・韓国語等の多言語へ自動翻訳可能とのことでした。

**特徴:** 本システムの利用方法は非常にシンプルで、ブラウザ上から日本語で情報を登録することができ、雪や台風、事故などの突発的なトラブルが原因となる遅延、行き先変更、運休などの情報が先述の AI 自動翻訳で翻訳され、利用者へ正確かつ迅速に届けられます。また登録された情報は API による連携が可能で、自社ホームページに加え、SNS 等のソーシャルメディアにも自動反映されるので、複数の媒体を個別に更新する手間を省くことができます。

**導入状況:** 現在は、IGR いわて銀河鉄道株式会社が本システムを導入しており、今後は鉄道会社にとどまらず、バス会社や航空会社等への導入が期待されています。

**所感・示唆:** 近年、インバウンド旅行者や日本に住む外国人は増加の一途をたどっていますが、鉄道会社が配信する遅延・運休情報は意外にも日本語の案内のみが多く、非日本語話者には十分に情報が行き届いていないのが現状です。また、これらの情報は即時性が求められる重要な情報であるにもかかわらず、翻訳作業のために鉄道会社で新たにリソースを割くということは容易ではありません。このような状況で、今後は鉄道会社にとどまらず、他の公共交通機関への導入が進められれば、訪日外国人を含む多様な利用者に向けた情報発信の実現が可能となり、利用者の利便性・満足度の向上につながると考えられます。

## 3) マインドワード株式会社

**概要:** 同社が開発した「スタンドアローン同時通訳(可搬型)」は、インターネット接続を必要としないスタンドアローン型のリアルタイム翻訳により多言語コミュニケーションを支援するシステムです。

**特徴:** 展示会場では、「令和7年度 屋久島国立公園における同時音声翻訳技術モデル実証業務(環境省)」で使用されたガイド向け同時通訳システムのデモンストラーションを体験することができました。ガイドがヘッドフォンのマイクを通して話した内容が、低遅延でリアルタイムに同時通訳され、ツアー参加者はイヤホンに接続された小型端末のボタン操作によって、英語・中国語・韓国語の3言語を簡単に切り替えながらガイドの説明を聞くことができます。また、ガイドが説明で使用することが予想される地域固有名詞や専門用語等は、事前に辞書登録によるカスタマイズが可能であるため、より正確な翻訳を届けることができます。

**所感・示唆:** 近年の訪日観光客、とりわけリピーターは、定番の観光地は一通り訪れていることが多く、いわゆる「秘境」と呼ばれる地域へ足を延ばし、観光客が少ない場所でしか得られない唯一無二の体験に価値を見出す傾向があると言われていています。しかし、こうした地域では、定番の観光地とは異なり、多言語対応の準備が十分に整っていないケースや、安定した通信環境が常時確保できないといった課題が想定されます。そのため、本システムのように通信環境に依存せずに多言語翻訳を行える仕組みが求められます。秘境と呼ばれる地域においても本システムを活用することで、訪日観光客がその土地や体験内容について十分に理解を深めることが可能となり、結果として観光体験の満足度向上につながることを期待されます。

### 3. 最後に

最近 LLM による翻訳が大きな注目を浴びていますが、今回の講演・展示を通じて、翻訳の用途・対象・利用環境によって、最適な翻訳システムは異なる、ということ具体的な事例を通して確認することができました。

スタンドアロン型のシステムでは、インターネットが不安定な環境下で自動翻訳をすることが求められるため、LLM を搭載・運用することは困難です。また、定型文の多いマニュアル等のドキュメントを大量に翻訳する場合には、高速かつ軽量で、LLM と比べて低コストである従来のニューラル機械翻訳の方が適している場合があります。逆に、文脈を考慮した翻訳という点では、LLM による翻訳の方が高い精度を示すこともあります。特に、これまで翻訳が難しいとされてきた小説分野においては、LLM による翻訳精度の向上が著しく、その進展には目を見張るものがあります。英国における日本小説のブームや、海外で高い人気を誇る漫画・ゲームなど、まだ十分に海外展開されていない日本のコンテンツを、今後さらに世界へ発信していくことに対して、大きな期待が寄せられていると感じました。

## AI 校正と AI 翻訳で進化する翻訳業務

中山 雄貴

株式会社ヒューマンサイエンス

### 1. はじめに

機械翻訳 (MT) は、ニューラル翻訳から大規模言語モデル (LLM) 基盤の生成 AI 翻訳へと進化し、多くの現場でインフラとして定着しつつあります。

MT エンジンの基盤がニューラル翻訳か LLM かに関わらず、その翻訳精度を左右するのは、入力される原文の質です。曖昧で不明瞭な原文を翻訳すると、誤訳となる可能性が高くなります。

この課題を解決する方法として、翻訳前工程で AI を活用し、原文を検証・整備する「AI 校正」が注目されています。本稿では、テクニカルライティングにおける AI 校正の実践的な活用方法と、それが翻訳工程全体に与える効果について解説します。

### 2. テクニカルライティングにおける AI 校正の役割

テクニカルライティングには、意味が一意に定まる正確さと、読者の理解を妨げない簡潔さが求められます。従来の校正ツールは、誤字脱字や送り仮名の誤りなど、あらかじめ定義されたルールに合致するかを判定する「パターンマッチング型」の形式チェックが中心であり、文章の意味や論理の整合性までは検証できませんでした。

生成 AI による校正では、LLM の文脈理解能力を活用することで、従来の校正ツールでは検出不可能だった以下のような問題を指摘できます。

- ・ 主語と述語のねじれ：長文化に伴い対応関係が崩れている
- ・ 不明瞭な指示語：「これ」「それ」等が何を指すか特定できない

- ・ 前後の記述における矛盾や論理の飛躍
- ・ 一文への過度な情報集約

AI の指摘内容を人間の執筆者が確認・修正することによって、MT が解釈に迷わず処理できる「曖昧さのない文章」へと原文を上げることができま

さらに、自社のスタイルガイドや TC 協会の「日本語スタイルガイド」を AI 校正ルールとして登録すれば、執筆者ごとの品質のばらつきを平準化できます。国際規格 IEC 82079-1 への準拠を確認することもでき、「情報は簡潔、一貫性、理解可能であること」といった規格要件を文脈に即して検証できるため、法規制や国際品質基準への適合を執筆段階から担保できます。

### 3. 実務への組み込み方と MT への波及効果

原稿執筆後に AI 校正を実行します。原稿に含まれるわかりにくい表現や論理的な不整合を特定します。人間が AI の指摘事項や修正案を参照しながら改善することで、校正の属人性を排除し、再現性のある品質改善が可能になります。

原文の曖昧さが解消されていれば、MT の出力精度は向上します。具体的には、以下の改善が期待できます。

- ・ 誤訳の低減：主語の明示や指示語の具体化により、MT が文意を取り違えるリスクが減少
- ・ ポストエディット工数の削減：訳文の修正箇所が減り、翻訳者が内容確認に集中できる
- ・ 品質のばらつき抑制：原文品質が安定することで、エンジンや翻訳者が変わっても訳文品質を維持しやすくなる

AI 校正による原文品質の向上は、翻訳工程全体のコストと品質に対して大きな効果があります。

#### 4. 運用上の留意点

AI 校正の業務運用には、情報セキュリティの確保が前提となります。未公開の技術情報を扱う現場では、入力データが AI モデルの再学習に利用されないことが保証された法人向け環境を選定する必要があります。

あわせて、AI はあくまで検証と提案を行う補助ツールであるという認識が不可欠です。特定業界の技術的慣習や製品固有の仕様に関して AI の判断が必ずしも正確とは限りません。AI が提示する修正案の採否は最終的に人間が判断する「ヒューマン・イン・ザ・ループ」の原則を運用プロセスに組み込むことが必要です。

#### 5. 「MTrans for Office」による統合環境

AI 校正と MT の双方の機能を、使い慣れた環境上でシームレスに利用できるのが「MTrans for Office」（エムトランス・フォー・オフィス）です。Microsoft Office 上での作業を前提に設計されており、既存の業務フローを大きく変更することなく導入できます。

MTrans for Office は Word のアドインとして提供されます。原稿の執筆後、Word のサイドパネルを表示して、その場で AI 校正を実行できるため、校正作業が執筆プロセスの一部として定着します。

校正ルールはオンラインで一元管理され、メンバー間で共有できるため、各執筆者が同一のルールセットを参照することで、担当者の経験差に左右されない安定した品質を組織的に実現できます。ルールの改訂や追加も即座に反映されます。

翻訳が必要な場合には、Google 翻訳・DeepL・OpenAI 等を MT エンジンとして利用して機械翻訳を実行できます。用語集機能や自動ポストエディット機能を利用できます。

MTrans for Office を利用すれば、原文の品質向上から翻訳、訳文確認まで同一環境内で完結するため、ツール間のデータ移動に伴う手間やセキュリティリス

クを最小化しながら、品質向上と効率改善を同時に実現できます。



MTrans for Office について詳しくは、

[https://www.science.co.jp/nmt/service/mtrans\\_office.html](https://www.science.co.jp/nmt/service/mtrans_office.html)

をご覧ください。14 日間の無料トライアルも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。

#### 6. おわりに

多言語ドキュメントの品質向上に取り組む際、翻訳エンジンの選定やポストエディットの精度改善に目が向きがちですが、翻訳の「前工程」を改善することが重要です。AI 校正を執筆業務の基盤として導入し、原文品質を組織的に底上げすることによって、翻訳業務全体の質と効率が向上します。

## 化合物表記翻訳技術を活かした機械翻訳サービス JAICI AutoTrans

河内 愛

一般社団法人化学情報協会 情報技術部

### 1. はじめに

化学情報協会 (以下 JAICI) は「化学を軸とした情報ソリューションにより、科学技術の発展と豊かな社会の実現に貢献する」ことをミッション<sup>1</sup>としています。国際的な情報流通において言語の違いは依然として大きな障壁となっていますが、JAICI では長年の業務を通して蓄積してきた化合物名の対訳辞書や機械翻訳に関する技術を基盤として、その解決に継続して取り組んでいます。

現在 JAICI は 2 種類の機械翻訳サービス<sup>2</sup>を提供していますが、本稿ではそのうち JAICI AutoTrans<sup>3</sup>に焦点を当て、機能的特徴を概説します。

### 2. JAICI AutoTrans とは

JAICI AutoTrans は、科学文献・特許・中間処理文書などの外国語文書の内容を日本語で効率よく把握したい方に特に有用な機械翻訳サービスです。手軽な操作で迅速に高精度な翻訳結果を得ることができます。原文の種類や用途に合わせて最適な翻訳メニューを選択できるのも大きな特長です。例えば、原文ファイル (PDF/DOCX/PPTX/XLSX) をアップロードするだけで同じレイアウトで翻訳結果を入手できるメニューでは、両者を比較しながらの内容確認が容易です。また、原文が格納された Excel ファイルに和訳結果の列を自動挿入できるメニューでは、大量のスクリーニング作業を効率化できます。その他にも多様な機能を備えています。特に注目すべきは化合物表記翻訳技術です。

### 3. 化合物表記翻訳とは

科学文献には、IUPAC 命名法・CAS 命名法といった国際的な標準規則に基づく体系名から、歴史的に定着した慣用名や一般名のような非体系的な化合物名まで、さまざまな表記が混在します。

化合物表記の例：

*Acetanilide*

*1H-imidazole-4-carbaldehyde*

*2,4-di-tert-butyl-6-(1-methylpropyl)phenol*

これらの化合物表記は、いわゆる”普通の単語”とは違い、名称の内部に化学構造や置換基の位置を表す規則や情報が組み込まれている特殊な用語です。そのため、その翻訳には化学的知識に基づく対訳辞書や命名規則の理解といった専門的な技術基盤が不可欠となります。JAICI では、このような化合物表記の特性を踏まえ、正確な翻訳を実現するために化合物表記翻訳技術を独自に開発しました。

JAICI 化合物表記翻訳は、2 ステップの独自技術の組み合わせで成り立っています。まずステップ1では、原文に含まれるテキストの中から化合物表記のみを自動的に認識します。続くステップ2では、認識した化合物名に対して適切な対訳を付与します。対応言語方向は、英日・日英・中日 (簡体字・繁体字) です。

### 4. 化合物表記翻訳の効果

科学文献において、化合物名は重要度が極めて高い専門用語と位置付けられます。正確な化合物名を把握することは、研究開発・特許調査・安全性評価といっ

たあらゆる化学関連業務の出発点と言え、以降の業務上の判断を左右します。本技術を活用することで、化合物名に関する誤訳・未訳・訳抜けを低減できます。

化合物名の誤訳は、場合によっては技術理解そのものを誤らせることがあります。例えば *ethyl acetate* は日本語で「酢酸エチル」ですが、これを「エチルエステル」のように一般化した表現で訳出してしまうと、対象物質の特定が曖昧になり、反応経路や用途の理解を誤るおそれがあります。さらに、誤った化合物名に基づいて実験計画や調達を進めた場合、無関係な試薬を選定してしまい、評価試験のやり直しや追加検証が発生するなど、研究費や工数の損失につながり得ます。したがって、化合物名の翻訳精度を担保することは、翻訳品質の向上にとどまらず、業務効率の観点からも重要です。

## 5. 化合物表記翻訳の実例

化合物表記翻訳の有効性を示すため、実例を紹介します。ある英語原文を ①JAICI AutoTrans ②一般的な Web 翻訳 ③一般的な生成 AI の 3 種でそれぞれ日本語訳し、化合物名 *acetyldialuric acid* の翻訳結果に着目しました (表 1)。まず、①JAICI AutoTrans では「アセチルジアルル酸」と訳出されました。これは正しい訳であり、*dialuric acid* が慣用名として「ジアルル酸」に対応することに基づきます。一方、②一般的な Web 翻訳では「アセチルジアルギン酸」と誤訳されました。これは *aluric* を、綴りや音が部分的に近く相対的に頻出度の高い *alginic* に誤推定した可能性があります。また、③一般的な生成 AI では「アセチルジア尿酸」と誤訳されました。*dialuric* を “*di + uric*” のように分解して解釈する誤りが生じた可能性があります。以上の結果は、化合物名の翻訳が、一般的な処理だけでは安定にくく、専用辞書と化学的知識に基づく処理が必要であることを示しています。なお、表 2 には中国語の化合物名についても同様の比較結果 (①~③) を示しています。

表 1

原文	acetyldialuric acid	
① JAICI	アセチルジアルル酸	○
② Web 翻訳	アセチルジアルギン酸	×
③ 生成 AI	アセチルジア尿酸	×

(2026 年 3 月現在)

表 2

原文	萊莫維韋	
① JAICI	レテルモビル	○
② Web 翻訳	レモビウエ	×
③ 生成 AI	レモビビル	×

(2026 年 3 月現在)

## 6. おわりに

本稿では JAICI が提供する機械翻訳サービス JAICI AutoTrans の概要について述べるとともに、特に化合物表記翻訳技術の有用性を示しました。科学文献において化合物名は、技術内容の理解の正確性を左右する中核的要素である一方、一般的な機械翻訳においては、医薬品や農薬などの慣用名の訳出において、なお改善の余地が残る場面があります。JAICI では、こうした課題に対応するため、対訳資源の整備・拡充の強化に継続的に取り組んでいます。本サービスにご興味をお持ちの方は、ぜひ JAICI ホームページをご覧ください。

### 参照

- 1: <https://www.jaici.or.jp/about-us/mission/>
- 2: <https://www.jaici.or.jp/translation/>
- 3: <https://www.jaici.or.jp/translation/autotrans/>

## 特許文書特化の機械翻訳

一般財団法人日本特許情報機構

### 1. はじめに

一般財団法人日本特許情報機構 (Japio) は、1985 年 8 月 1 日に総合特許情報サービス機関として設立し、昨年で創立 40 周年を迎えた。

この間、特許情報の正確かつ容易な提供を目指し、従来の紙媒体から電子媒体への移行、外国公報の日本語翻訳公報の提供など、特許情報の利用拡大を進め、我が国の技術開発さらには産業の発展に寄与してきた。

また、特許庁等が進める知財施策にも多年にわたり協力し、特許電子図書館 (IPDL)、特許情報プラットフォーム (J-PlatPat) の運用、抄録の翻訳事業などにも携わり、それらで得られたノウハウ・スキルを活かし当財団独自のサービスとして、「世界特許情報全文検索サービス (Japio-GPG/FX)」、「Japio-AI 翻訳」を提供している。今回はその中でも特許文書特化の翻訳サービスである「Japio-AI 翻訳」を紹介する。

### 2. Japio-AI 翻訳の開発

Japio では、以前より独自の大規模な言語資源を収集し、その対訳コーパスを活用して特許公報に特化した機械翻訳エンジンの研究開発を行ってきた。2017 年には統計的機械翻訳 (SMT) 手法により、翻訳需要の高い中日機械翻訳のサービス化を実現、その後、英日機械翻訳の実現に成功した。統計的機械翻訳の特徴は、大量のコーパスから辞書や文法などの言語的特徴を自動的に学習させるもので、日本語としてはやや不自然ながらも原文を忠実に翻訳できる特徴を有する。

一方で、対訳コーパスの生成手段や収集規模が重要となる上、文法間で発生する語の並び替えなどや言い回し等の点では課題が多く、特に自然言語としての流

暢さについては機械翻訳の研究開発において長年の課題であった。

その状況を一変させたのが深層学習 (Deep Learning) と呼ばれるニューラル機械翻訳 (NMT: Neural Machine Translation) の登場である。当初は RNN ベースの sequence to sequence モデルがベースのニューラル機械翻訳が主流であったが、Google が開発した Transformer と呼ばれるアーキテクチャの登場により、係り受けの正確性や流暢さが飛躍的に向上した。また、ニューラル機械翻訳の研究開発において必須条件となるのが膨大な演算処理を持つ GPU サーバであるが、Japio は、ニューラル機械翻訳が主流となる前より精力的に GPU 環境を確保し、潤沢な計算資源を保有することでスムーズに研究開発を進めてきた。そして、Japio では、統計翻訳 (SMT) の学習用に独自に収集した大規模な特許対訳コーパスを利用して、NMT についても独自に学習を行い、2019 年 11 月には Japio-GPG/FX に、特許文献をリアルタイムで翻訳して表示する AI 翻訳機能を導入、自然で読みやすい訳文を提供できるようにした。

当初 AI 翻訳を導入した言語は、英語、中国語、ドイツ語であったが、特に、中日翻訳での品質向上が顕著であった。SMT では原文での preordering(事前並び替え)を行っていたため、この段階での間違いがその後の処理に大きく影響していたが、AI 翻訳では preordering をしなくとも高精度な構文的変換ができる学習がなされるため、大きな精度向上につながったと考えられる。

現在では、英語、中国語 (簡体字・繁体字)、韓国語、ドイツ語、フランス語、スペイン語、ロシア語、ポルトガル語まで翻訳言語を広げている。

以下に、Japio の AI 翻訳の主な特徴を挙げる。

### 3. Japio-AI 翻訳の特徴

#### 大規模・高品質な特許データを学習した AI が、複雑なテキストも正確な構文で読みやすく翻訳

Japio で独自に収集した大量の特許対訳テキストで AI 翻訳の学習を行うことで、難解な特許文が正確で読みやすい形の翻訳文となっている。

特に中日の特許対訳テキストは、Japio 独自の言語資源である対応特許由来の大規模な対訳テキストから、厳選したスクリーニングをかけることにより、量とともに質も確保している。そして、量・質の両者を満たした中日特許データのみで学習をしているので、翻訳結果も特許文章に特徴的な出力結果となる。

例えば、日本語訳文のスタイルは「である調」に統一される点、あらゆる分野の技術用語を網羅した翻訳結果となる点、などの特徴を有した翻訳が可能となる。この特徴は、読み易さの向上はもちろんのこと、内外明細書の作成の際に、ポストエディットの作業負担を大幅に軽減させる効果もある。

#### 独自開発の処理技術 (X-STEP®) により、特許翻訳者のノウハウを AI に移植

一般的な技術文献と比べ特許文献は長文が多いことが特徴の一つである。このような長文をそのまま翻訳しようと思っても、翻訳エンジンの文字数制限などにより適切に翻訳できない場合も少なくない。この問題を解決するため、Japio では、X-STEP®(クロスステップ: XML Translation Framework with State-of-the-art Translation Engines and Automatic Claim Pre-editor)と呼ぶ機械翻訳用フレームワークを独自開発した。

例えば、前処理として、請求項の原文を翻訳しやすい形に自動で書き換えてから翻訳を行うことで、一度に翻訳できないような長い請求項に関しても、なるべく読みやすく理解しやすい訳文が得られるようにしている。請求項は、項番の記載や、箇条書き、文の途中での改行などといった通常の文章にはない記載や、特許文章に独特な文構造が採用される傾向にあり、そのまま機械翻訳処理を行うと品質が低下するおそれがある。この前処理は翻訳品質の向上に大いに貢献している。

また、特許文献に頻出する上付、下付文字の訳文での高い再現性を有しており、上述と同様に、ポストエディットの作業負担を大幅に軽減させる効果がある。**AI 翻訳と SMT の併用により、訳抜け、訳語の繰り返し等のエラー文が少ない翻訳を実現**

AI 翻訳は、原文に書かれているフレーズが訳出されない現象(訳抜け)や、原文に書かれていない語が訳出される現象(湧き出し)が一定の確率で発生する。そこで、AI 翻訳の訳文に対する不具合チェックを行い、顕著な訳抜けや湧き出しが発生している場合には SMT でバックアップするハイブリッド方式を採用している。

#### セキュアな通信規格 (TLS 暗号化通信) と翻訳原文・訳文を保管しないシステムにより機密性を確保

ユーザーの環境から翻訳サーバまでの通信を TLS 暗号化通信で行っている。また、テキスト翻訳機能を使用する場合は、原文も翻訳結果も翻訳サーバには記録しておらず、セキュリティには万全の対策をとっている。

特に、海外特許出願原稿の翻訳などに機械翻訳を用いる場合、出願前の特許文書は未公開であることが多い。Japio-AI 翻訳では、ユーザーが出願前の特許文書を安心して翻訳いただけるよう対策を講じている。

### 4. まとめ

「Japio-AI 翻訳」は、高品質な特許対訳データで学習した AI と独自技術により、専門用語や長文が多い特許文献を構造分析して最適に分割し、AI 翻訳と SMT を組み合わせることで訳抜けや不要な補足を防ぎつつ、主要言語に対応した正確で読みやすい翻訳を提供している。さらに文献番号を入力するだけで各国特許公報の原文と翻訳文を取得できるため、研究者の検索効率向上と負担軽減を実現できる。

興味のある方はサービス窓口: [service@japio.or.jp](mailto:service@japio.or.jp) まで。

Japio-AI 翻訳紹介ホームページ

URL: <<https://japio.or.jp/service/service06.html>>

## 編集後記

## 編集後記

新田 順也

AAMT 編集委員会

本号では、大規模言語モデル（LLM）の急速な普及と、それが自動翻訳ツール開発や翻訳者の心理にどのような影響をもたらしているかという点について論じています。私が専門とする特許分野でも LLM の活用事例が増えています。最近では、知財業務全般において LLM を活用する企業の取り組みが話題になっています。特許翻訳も LLM で行い、その出力された訳文の修正を事務員が担い、最終的には出願先の現地の弁護士が修正して出願するという事例が紹介されています。これにより翻訳費用は削減されたようです。公表されている運用方法だけで特許を翻訳して外国に出願できるという点が、私には信じがたいものです。このように LLM を運用している企業において、知的財産権が適切に保護されるのか、また各社の LLM の運用方法についても注視していきたいと思えます。

今後も、翻訳分野における LLM の活用は広がっていくと考えられます。また、分野ごとの差はあるものの、翻訳者に求められる役割も変化していくでしょう。ますます、LLM と人間との健全な協働が求められる時代になってきたと感じます。LLM との関わりにおいては、人間が主導権を持ち、人間が幸せになるような活用方法が求められていると言えるでしょう。本号での多面的な議論が翻訳における LLM 活用の在り方を考える一助になると思えます。

## 1. 巻頭言

石岡映子様から「機械翻訳の進化と共に歩む」を寄稿いただきました。石岡様は、長年の研究蓄積をビジネスの現場へとつなぐ「橋渡し」の役割を担ってきた

AAMT セミナーの歴史を振り返りました。技術が専門家の手を離れ、社会実装の段階に入った今、単なる効率化だけでなく「誰のために使うか」を問い続ける重要性が強調されています。ここで紹介されている「MT は世の中を平和にする」という AAMT 創立者の長尾真初代会長の言葉が、今の時代に胸に響きます。

## 2. 事例・研究

早川威士様から「機械翻訳出力に対する LLM を用いた後編集」を寄稿いただきました。NMT（ニューラル機械翻訳）の出力に対し、LLM を用いて自動ポストエディット（PE）を行う際の精度を、「正解率」と「修正率（過剰修正の程度）」の観点から定量的に解析した研究報告です。複数の LLM モデルを用い、誤りタイプやプロンプトによる挙動の違いを統計的に明らかにしています。

阪本章子様から「ポストエディットが翻訳者の仕事満足度に与える影響」を寄稿いただきました。イギリスの翻訳者を対象とした 2024 年の調査を通じ、MTPE がフリーランス翻訳者の心理的な満足度やキャリアにどのような影響を与えているかを分析しています。また、持続的な翻訳産業に関する考えも紹介されています。2026 年は日本での調査を計画しているとのことでした。

萱野陽子様から「大規模言語モデルは多様な翻訳仕様に追従できるか？」を寄稿いただきました。本記事は、LLM が文体や用語の指定、不適切な表現の回避といった「翻訳仕様」にどの程度応えられるかを検証しています。この評価自体を LLM が担っています。

Editor's note  
Junya Nitta

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/>

### 3. 温故知新

AAMT ジャーナルの過去の記事として、1994年3月号の「機械翻訳の今後の技術動向」と同年6月号の「ヨーロッパの現在のMT活動」を取り上げます。これらを手がかりに、30年以上前の技術的課題と、研究開発を取り巻く財政的・政治的課題を振り返ります。

### 4. イベント報告

中澤敏明様から、第12回アジア翻訳ワークショップ(WAT2025)を報告いただきました。招待講演では、LLMを用いた機械翻訳の品質評価に関する課題と、その解決策が紹介されました。

出内将夫様から、2025年12月に開催されたAAMTの年次大会を報告いただきました。パネルディスカッションや公募ポスター発表、招待講演について詳細に報告をいただきました。

玉咲知香子様から、2026年2月に開催された第9回自動翻訳シンポジウムを報告いただきました。会場にて展示した企業から3社を取り上げ、それぞれの課題解決のためのサービスについて説明をいただきました。

### 5. AAMT 創立 35 周年

今年で、AAMTは創立35周年を迎えます。これを記念して、ロゴデザインを全面的に刷新しました。新しいロゴは、AAMT ジャーナルの2026年6月号の冊子版から使用されます。また、今年9月には創立35周年の記念イベントとして国際会議を奈良で開催する予定です。

自動翻訳技術がこれまでになく盛り上がっている昨今、AAMTはさらに積極的に情報を発信し、みなさまから信頼される組織に成長していきたいと思っています。今後ともどうぞよろしくお願いいたします。

---

**AAMTジャーナル「機械翻訳」No. 84**

- 【発行日】2026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 プリントバック

Asia-Pacific Association for Machine Translation (AAMT)  
Shinjuku Hirose Bldg. 5F, 4-7 Yotsuya, Shinjuku-ku, Tokyo 160-0004 Japan

---

**AAMT**  
*since 1991*  
アジア太平洋機械翻訳協会