

ISSN 1883-1818

No.78

June 2023

AAMT Journal

Asia-Pacific Association for Machine Translation

機械翻訳

機械翻訳



AAMT Journal

Asia-Pacific Association for Machine Translation 機 械 翻 訳

No.78

目 次

巻頭言		
「二度目の嵐」は吹き荒れるか？	永田 昌明	3
解説記事		
機械翻訳らしさを軽減するための手法の妥当性とMTPEの効率化の条件について	堂本 秋次	5
機械翻訳活用で顧客の期待に応える	Edith Bendermacher, 上田 有佳子	7
Case Study : How NetApp is Utilizing Machine Translation to Meet Customer Needs and Expectations	Edith Bendermacher and Yukako Ueda	11
人間の翻訳と機械の翻訳(7):社会的な受容の一側面	影浦 峯	15
ISO18587 供給者自己適合宣言のススメ	森口 功造	21
MT ユーザーガイドへのお誘い	AAMT	25
AAMTセミナー		
理工系の英語論文執筆における課題と機械翻訳の活用	西山 聖久	26
海外投資家を引き付ける英文IR～第3回AAMTセミナー要旨	西村 賢治	31
温故知新		
温故知新2	内山将夫	35
イベント報告		
AAMT 2022, Tokyo～機械翻訳最前線～	出内 将夫	43
第6回自動翻訳シンポジウム 参加報告	東山 翔平	49
法人会員PR		
ニューラル機械翻訳におけるエヌビディアのハードウェアとSDK	呉先超	55
編集後記	園尾 聡	57

C O N T E N T S

Foreword		
Will the second storm rage?	Masaaki Nagata	3
Commentary		
The validity of the approach to reduce “machine-translationese-ness” and conditions to take advantage of MTPE effectively	Akitsugu Domoto	5
Case Study : How NetApp is Utilizing Machine Translation to Meet Customer Needs and Expectations	Edith Bendermacher and Yukako Ueda	11
On social acceptance of human and machine translation	Kyo Kageura	15
An encouragement of Supplier's declaration of conformity: ISO 18587: 2017	Kozo Moriguchi	21
Invitation to MT User Guide	AAMT	25
AAMT Seminar		
Issues in Writing English Papers for Science and Engineering by Using Machine Translation	Kiyohisa Nishiyama	26
IR in English to Attract Foreign Investors	Kenji Nishimura	31
Learning from the past		
Learning from the past 2	Masao Utiyama	35
Event Report		
AAMT 2022, Tokyo --- Forefront of machine translation ---	Masao Ideuchi	43
Report on the 6th Symposium on Automated Translation	Shohei Higashiyama	49
Corporate PR		
NVIDIA's Hardware and Open-Source SDKs for Neural Machine Translation	Xianchao Wu	55
Editor's Note	Satoshi Sonoh	57

巻頭言

「二度目の嵐」は吹き荒れるか？

永田 昌明

NTT コミュニケーション科学基礎研究所

「この巻頭言は ChatGPT で作ったものではない」と最初に述べる必要がある時代になった。2022 年 11 月に登場した ChatGPT に関する話題は、この原稿の執筆時点(2023 年 4 月上旬)ではまだ勢いを失っていない。AI 業界は数日で状況が激変するので、この巻頭言の内容は公開される頃には陳腐な駄文になっているかもしれないがご容赦願いたい。

私のような機械翻訳の研究者は 2015 年前後にニューラル機械翻訳による破壊的イノベーションを経験した。ルールベース機械翻訳や統計的機械翻訳の頃の知識と経験は使い道がなくなり、「焼け野原に立つ」とはこういうことかと思った。しかしゼロから出発するという条件は皆同じだし、新しい技術を学ぶのは意外に楽しかった。2017 年に Transformer というニューラルネットの新しいアーキテクチャが登場すると、機械翻訳のアルゴリズムはもうこれでよい気がしたので、私は対訳データを集めることに力を入れた。それなりの量の対訳データを集めていい感じになったと思ったら、また嵐が来てしまった。

世間では、日本語の大規模言語モデルを作らないと日本は技術革新から取り残されると騒いでいる。しかし、私個人はこの二度目の嵐に一度目のような焦燥感を感じていない。言語生成モデル GPT は、技術的には Transformer のデコーダを取り出して大量の言語データで訓練したものなので、機械翻訳から生まれた技術が世間を騒がしているのはむしろ心地よい。地元の高校が全国大会で優勝して、別に自分はその高校の出身ではないが、ちょっと誇らしい気分になるようなものである。

機械翻訳の研究者として私が今回の ChatGPT に端を発した技術革新の嵐に対して楽観的な第一の理由は、

Microsoft によれば現時点の ChatGPT/GPT-3 の翻訳精度は国際会議 WMT の汎用機械翻訳タスクの 1 位のシステムと同等と報告されているからである。GPT-3 のような大規模言語モデルは、一般的なニューラル機械翻訳システムの翻訳モデルに比べると、約 1000 倍のパラメタ(1750 億個)を持ち、約 100 倍のテキストデータ(3000 億トークン)で訓練する。大規模言語モデルは、翻訳を含む様々な動作を自然言語で指示できるので、例えば要約してから翻訳するとか、親しみやすい口調で翻訳するというような翻訳とその周辺の作業をまとめて実行する機能を実現するプロトタイプシステムを作るにはとても便利な道具である。しかし、大規模言語モデルは作成するのも運用するのも膨大な費用と労力が必要なので、汎用の大規模言語モデルに比べて 1000 分の 1 の大きさのモデルで小気味よく動く特定目的の翻訳器を作るという技術的な選択肢は依然として残っているように思う。

私が今回の嵐に対して楽観的あるいは好意的なもう一つの理由は、ChatGPT/GPT-3 が教師なし機械翻訳、すなわち対訳データを使わずに機械翻訳を実現するという機械翻訳の研究者の長年の夢を実現したかもしれないからである。OpenAI によれば GPT-3 の訓練データは 92%が英語で、日本語は 0.16%しかない。しかし ChatGPT と日本語で対話してもそれほど違和感はない。GPT-3 の訓練データに対訳データがどれくらい混入しているかはよく分からないが、GPT-3 の内部には言語に依存しない意味表現が存在し、それを介して対話や翻訳が実現されているように見える。

2018 年に登場した言語理解モデル BERT (Transformer のエンコーダを取り出したもの)の頃から、複数の言語のテキストで一つの言語モデルを訓練

Will the second storm rage?

Masaaki Nagata

NTT Communication Science Laboratories, NTT 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/>

すると、訓練データの中に対訳データが存在しなくても、異なる言語で意味的に近い単語は数値的に近いベクトルで表現されると報告されていた。GPT-3 のような大規模言語モデルでは、より大きなモデルをより多くのテキストで訓練することにより、より普遍的な言語の意味表現をモデルの内部に獲得する傾向があると推測される。

GPT-3 による翻訳は、英語から英語以外の言語への翻訳の精度よりも、英語以外の言語から英語への翻訳の精度が高いと報告されているので、対訳データではなく各言語の単言語データの量が翻訳精度を決めることは間違いない。対訳データを必要としない機械翻訳を実現するだけならば、GPT-3 より小さなモデルで実現できるかもしれない。

ここ数年間、対訳データを集めていた私にとってはこれまでの努力を否定される展開ではあるが、破壊的イノベーションとはそういうものである。一方、今回の嵐を経て、SNS のように対訳データが存在しない領域やアジア言語と日本語のように対訳データが存在しない言語対で高精度な機械翻訳を実現できるようになるのならば、これはこれで面白い未来が待っていそうだと感じている。

解説記事

機械翻訳らしさを軽減するための手法の妥当性と MTPE の効率化の条件について

堂本 秋次

個人フリーランス

1. 概要

昨今、機械翻訳の進化が目覚ましいことから、機械翻訳と人間翻訳の差分や棲み分け、あるいは共存について取り沙汰されています。しかし、翻訳する側や発注側の立場からの言及が多く、読み手としての立場から機械翻訳と人間翻訳の差分を比較したり、その有効性について議論をしたりする場面は多くない印象です。

そこで今回は、原文を伏せた状態で機械翻訳かまたは人間翻訳かを読み手に判断してもらい、原文を与えられていない場合でも読み手は機械翻訳と人間翻訳を区別できるのか、区別するとすれば、何を基準に区別しているのかを考察すると共に、従来言われている機械翻訳ツールの活用による MTPE 業務の効率性についても一考しました。

結論として、読み手の機械翻訳か人間の翻訳かの判断は恣意的であり、その正答率はおよそ 50%であること、翻訳者は原文に対する工夫ではなく読み手に対するアプローチとしての工夫を行うことで読み手に狙った効果を与えたり評価を得たりできる可能性があること、また MTPE 業務における効率化には作業内容の限定やプレエディットの活用が有効である可能性があることが示唆されました。

以下、本記事では、機械翻訳に対して人間が行った翻訳を「人間翻訳」と呼称します。

2. 調査手法と主な分析

本調査では、読み手に対して原文を与えず、訳文のみからそれが機械翻訳（今回は DeepL を用いています）であるか、または人間翻訳（今回は本記事執筆者、

つまり堂本が翻訳しています）であるかを判断させています。調査方法はオンラインクラウドソーシングサービスである Lancers のタスク方式での依頼により、1 ユーザーにつき 1 回の回答を行わせることで 450 近い回答を集め、Twitter などを通じてさらに 30 近い回答を集めました。結果として、一般にも機械翻訳と人間翻訳を原文から区別することは非常に難しいらしいことから、すべての訳文を正しく機械翻訳か人間翻訳か判断できた回答者（全問正解者）は出ませんでした。その上で、この回答や同時実施したアンケートなどをもとに、機械翻訳かどうかをどのように判断したか、また読み手が良い訳文を判断する基準および翻訳者のアプローチと読み手の判断基準の一致について検討しました。

分析の中では、読み手は原文と訳文を一般に比較しないため、原文に対する工夫（原文の言葉遊びや表現のユニークさの訳出、原文の語順の維持など）は、一概に読み手に評価されないことが確認されています。一方、例えば読み手に「小難しい印象を与えたい」や「意図的に読みにくさを演出したい」といったような場合に、適切なレトリックやアプローチを用いることは効果があり、読み手の読書体験においてポジティブに受け入れられるものであれば、読み手にも「良い翻訳である」や「人間的な翻訳である」というように受け入れられる傾向があることが理解されました。

このことから、人間翻訳に対しては、「読みやすさ」「理解しやすさ」「直感的な語順」が期待されているとされますが、「人間翻訳らしい工夫」は特に求められていないと考えられます。また、内容自体の難しさによっても、人間翻訳への評価が下がることがありますが、これは「読みにくい」ことや「理解しにくい」ことの

The validity of the approach to reduce “machine-translationese-ness” and conditions to take advantage of MTPE effectively

Akitsugu Domoto

Freelancer (independent)

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/>

原因が翻訳へと帰結されてしまうことによるものです。

こうしたことから、機械翻訳らしさを軽減する（読みやすく、理解しやすく、語順が直感的である文にする）には、「一文を短くする」や「ディスコースマーカ―を用いる」といった方略が一般に有効です。また、意外性のある語彙を用いることが機械翻訳らしくないという印象になることもあるようです。しかし、「機械翻訳らしさ」は最終的に読み手の主観的判断によって決まるため、翻訳時には原文に対してではなく、読み手に対して工夫を行うことが重要であることを強調しておきます。

3. MTPE の限定的有効性

また、今回は機械翻訳の訳例の中に敢えて間違いが含まれているものを紛れ込ませることで、これを文脈などから判断して指摘し、機械翻訳であることの証拠（あるいは人間翻訳であることの証拠）とする回答があるかを確認しました。しかし、こうした指摘は全体を通して見られず、訳文のみから訳出上のエラーを発見し、これを修正することは困難であることが理解されました。これは、機械翻訳ツールによって翻訳された出力結果を与えられ、それに適宜修正を加えるというやり方が一般的である MTPE の効率性に疑問を呈するものです。原文からエラーを発見することは非常に困難であることから、「誤訳判定」までポストエディットで行う場合、原文との比較が必要になるため、効率化が限定的になると言えるでしょう。このため、ポストエディットにおいての修正は限定的なものであるべきであり、その範囲の合意が効率化の鍵となります。ただしこの修正範囲の合意は、翻訳者とエージェント間のみならず、エージェントとエンドクライアントとの間といった場面でも必要になることをつけ加えておきます。

一方、原文との比較が必要であるような場合、プレエディットを行い、出力に対するエラー修正にあたることで、機械翻訳を用いた効率化が可能になる可能性

もあります。もちろんこの場合にも、プレエディット後の機械翻訳の出力に対してどの程度の編集が必要であるかの合意が必要であり、基本的にその程度は最低限であるべきと考えられます。

4. まとめ

今回の調査結果から、翻訳者は読み手の視点に立って翻訳の妥当性を判断し、機械翻訳を用いてプロジェクトに関わる場合には、その出力結果のクオリティを鑑みてどの程度の修正を行うべきかの範囲をエージェントやクライアントとの間に明らかにすることで、効率化を実現していくことが求められると言えます。もちろん、機械翻訳の使用が推奨されない分野を明らかにすること、そうした分野における翻訳のやり方や手法の継承についても、今後の課題として一考の価値があります。これについては翻訳者のみならず、翻訳業務の第一受注先となっているエージェントの協力が不可欠であり、またクライアントの翻訳に対するリテラシーを高めていくことも長期的には必要と考えられます。

解説記事

機械翻訳活用で顧客の期待に応える

Edith Bendermacher*, 上田 有佳子**

*Director, Globalization Strategy and Localization Operations, NetApp Inc.

**ネットアップ合同会社 APAC グローバル化戦略統括

1. 概要

昨今の IT 業界では、情報提供のスピードが、これまで以上に重要な要素となってきました。NetApp 製品を検討される顧客層は、オンラインで提供されている技術的な情報をまずはご自身で検索して入手されることが多くなってきました。

NetApp のグローバリゼーション チームである Globalization Programs Strategy Office (GPSO) では、さまざまなコンテンツのローカリゼーションワークフローに機械翻訳 (MT) エンジンを組み込むようになりましたが、一方、NetApp の技術ドキュメント開発チームでは、ほぼすべての製品マニュアルについて、GitHub プラットフォーム上に移行したうえで一般公開を始めました。GPSO は、技術ドキュメント開発チームと協力し、ニューラル機械翻訳 (NMT) による日本語翻訳を、英語版と同時に提供するという決断をしました。

NMT エンジン日々進化しているとはいえ、日本市場向けには、日本語訳を公開する前には、通常はポストエディット (PE) が不可欠と考えられていますが、私たちは、PE なしの NMT (Raw NMT) で情報提供のスピードを優先することによって、高い顧客満足度を達成することができました。

2. はじめに

ここで、今回の事例の背景をご説明するため、弊社 NetApp について少しご紹介させていただきます。NetApp

は、ハイブリッド マルチクラウド ソリューションでお客様を強力に支援する企業です。IT エクスペリエンスの向上というビジョンの下、統合管理と AI ベースの最適化により、複雑なハイブリッド マルチクラウドの世界において、シンプルさ、セキュリティ、コスト削減、サステナビリティを、データやアプリケーションがオンプレミスにあるかクラウドにあるかに関わらずもたらし続けるソリューションを提供しています。ビジネスに最適なイノベーションをいつでも実行できる自由と将来のどんな状況にも対応できる柔軟性があるという弊社ソリューションの特徴が、弊社情報をどのようにお客様に提供するかにも影響しました。

今のお客様は、自社の要件に合ったソリューションを Web サイトで探し、競合製品各種を比較してどれが優れているかを自ら調べておられます。

一方、弊社製品ポートフォリオは、お客様のニーズに合わせて日々拡大しています。

増え続けるコンテンツに対応し、かつ頻繁な更新にも対応するため、弊社は自社の製品マニュアルを GitHub で、ユーザーか否かに関わらず一般に公開することを決定しました。

外資系の会社が日本で製品を販売するにあたっては、まず製品情報を日本語で届けることを考えなければなりません。日本語の文法的な特徴も相まって、日本は長きにわたって、英語の許容度ランキングで低い順位を維持しています。

Case Study: How NetApp is Utilizing Machine Translation to Meet Customer Needs and Expectations
Edith Bendermacher* and Yukako Ueda**

*Director, Globalization Strategy and Localization Operations, NetApp Inc.

**APAC Sr. Globalization Strategist, NetApp G.K.

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/>

NetApp が製品マニュアルを PDF で提供していたころ、ローカリゼーションは当初、翻訳者による翻訳 (HT) で行われていました。2013 年からは機械翻訳 (MT) に PE を組み合わせた方法に変更しましたが、それでも日本語版の提供まで、英語コンテンツの公開後数週間がかかりました。

しかし、英語のコンテンツを GitHub プラットフォームで公開するようになり、私たちは、英語圏以外のお客様をローカライズが終わるまで待たせることはできないと考えました。

MT は、国内外を問わずローカリゼーション業界の大きなトピックであり続けていますが、特に日本語に関しては NMT が登場するまで、一般的に PE なしの Raw MT を顧客に見せることが選択肢になることはほとんどなく、PE による品質改善を限られた予算内で実現するという Buyer 側の要望に応えるというのは、日本のローカリゼーション サービス プロバイダ (LSP) にとって、頭痛の種になることが多々ありました。日本の LSP は、これを自社にとっての研究コストと捉え、たとえ PE が HT よりも多くの時間が必要になるケースでも、HT よりも低いレートで受注しているということを、一般的な話として耳にすることもありました。

NMT が業界のゲームチェンジャーとして登場したとき、はたして日本語でもうまくいくのだろうかとか懐疑的だった方も多いのではないでしょうか。NMT のアウトプットはどんどん改善され、日本語でも実用に使われる場面が増えてきました。が、NMT の品質が劇的に向上したとはいえ、ルールベース機械翻訳 (RbMT) や統計的機械翻訳 (SMT) での苦い経験から、今なお Raw NMT のアウトプットをそのまま顧客に提供すると決断するのは、日本人にとって非常に勇気のいることと思われます。

NetApp がローカリゼーションに投資するのは、お客様に、弊社製品がいかに優れているかを理解していただき、最先端の技術を会得していただき、弊社製品を試して購入の意思決定をし実際に使用していただくためです。

私たち GPSO は、単なるローカリゼーション サービス部門ではなく、グローバル化の戦略と方法論を NetApp 全部署に展開するセンター オブ エクセレンス (CoE) です。私たちは、コンテンツのローカリゼーション戦略に基づき、Raw NMT が、日本のお客様に英語圏のお客様と同様のカスタマー エクスペリエンスを届けるソリューションになると考えました。

3. 手法

私たちは日本市場に対応するため、製品マニュアルを掲載している GitHub プラットフォームに NMT エンジンを組み込みました。NMT エンジンの登場で MT の品質が格段に向上したとはいえ、訳文の品質に対する日本人読者の期待は変わりません。そのため、Raw NMT を導入するという NetApp の決断 は、リスクを伴うものでした。しかし、NetApp がクラウド時代にも競争力を維持するには、情報提供のスピードを優先する必要があったのです。

Raw NMT の導入を成功させるために、私たちは、以下のことに取り組みました。

- 公開する英語コンテンツは、ライティング基準に沿って作成されたものであること
- 過去 10 年間に蓄積された HT による膨大な量のコーパスでトレーニングされた NMT エンジンを使用すること
- 法務部門と連携し、適切な免責事項テキストを各ページに記載し、MT を使用していることを読者に通知すること

- お客様と日々接している営業チームやパートナー企業とコミュニケーションをとり、この方法がお客様の期待に沿っているか確認すること
- 読者からのフィードバックを取り込んで生かすためのワークフローを確立すること

4. 結果

製品マニュアル Web サイトに Raw NMT を導入するという NetApp の決断は、日本のパートナー企業やお客様から肯定的に迎え入れられました。

その後の 2022 年 8 月、NetApp は、日経コンピュータ 顧客満足度調査 2022-2023 のストレージ部門で第 1 位にランク インしましたが、この 1 位獲得という結果について、パートナー企業各社からは、NetApp の製品マニュアル提供方法の影響が大きいという声をいただいています。日本市場への情報提供のスピードを重視する姿勢は、米国本社の企業でありながら日本市場を優先していることの現れであり、早いペースで進化するクラウド コンピューティングの世界に欠かせないためです。

GPSO では、これら NetApp 製品マニュアルの日本語ページに対し日本から多数のアクセスがあることも確認しています。

5. 考察

私たちは製品マニュアル Web サイトへの Raw NMT 活用に成功しましたが、このアプローチを適用する場合は、潜在的なリスクや課題に注意する必要があります。企業でお客様向けコンテンツに Raw NMT の導入を検討される場合には、次の 3 つの主要要件の検討をお勧めします。

- 免責事項で MT の使用を明確に告知する
- 十分にトレーニングされた高品質なエンジンを使用する

- 有用なユーザー フィードバックを生かすためのフィードバック ループを用意する

日本では、Raw NMT の結果そのままの公開に積極的な人は多くありませんが、Raw NMT 使用で成果が得られた私たちの事例が、クラウドの時代に競争力を維持する方法を模索している他の企業のモデルケースになると考えています。

6. まとめ

結果として、製品マニュアルに Raw NMT を導入したことで、NetApp は情報提供のスピードを優先し、高いレベルの顧客満足度を達成することができました。このアプローチには潜在的なリスクや課題がありますが、注意深く導入および管理をすれば、よい成果が得られることを示しています。Raw NMT の使用が、最新の情報を入手したいという日本国内の需要の高まりに対応するための実現可能なソリューションになり得るのです。

「高品質」の提供は、日本が誇ってきたことであり、変えるべきものではありません。しかし、ビジネスのスピードは、世界中で速くなっていく一方です。日本経済が成長し続けるためには、最新の技術のキャッチアップを優先しなければなりません。急速に変化する世界においては、スピードが鍵となります。品質の高い翻訳が完成するまで待つことはできません。

今回の成功事例が、顧客のニーズを満たすための手段として Raw NMT を試していただくきっかけとなることを願っています。

7. 謝辞

本稿を書く機会を与えてくださったアジア太平洋機械翻訳協会事務局の皆様、特に、Women in Localization のセッションならびに最新の著書でご自身の研究を共有してくださった、AAMT 会長 隅田英

一郎先生に感謝いたします。

また、Women in Localization Japan の共同設立者である機械翻訳コンサルタント 佐々木優子氏には、10 年前に MT の可能性について上田を開眼させてくれたことに、感謝の意を表します。

8. 参考文献

- NetApp 製品ドキュメント：
<https://www.netapp.com/ja/support-and-training/documentation/>
- EF English Proficiency Index -英語能力指数
ランキング: www.ef.com/epi
- ネットアップ、「日経コンピュータ 顧客満足度調査 2022-2023」のストレージ部門で第 1 位を獲得：
<https://www.netapp.com/ja/newsroom/press-releases/news-rel-20220826-640489/>

解説記事

Case Study: How NetApp is Utilizing Machine Translation to Meet Customer Needs and Expectations

Edith Bendermacher* and Yukako Ueda**

*Director, Globalization Strategy and Localization Operations, NetApp Inc.

**APAC Sr. Globalization Strategist, NetApp G.K.

1. Abstract

In today's IT industry, the speed of information delivery has become more important than ever. Customers who are considering NetApp products are searching for technical information available online on their own.

In response, our Globalization team at NetApp, called Globalization Programs Strategy Office (GPSO), decided to meet the expectations of the Japanese market by incorporating Machine Translation (MT) engine into the localization workflows for a variety of content.

For technical documentation, GPSO collaborated with technical documentation development team who decided to move most of the product documentation to the GitHub platform and make it publicly available. GitHub is providing Japanese translations simultaneously with the English version.

Even though MT engine quality is evolving day by day, post-editing (PE) is usually considered essential for the Japanese market before publishing a Japanese translation. By prioritizing the speed of information delivery by applying MT without PE (Raw MT), we were able to still achieve high customer satisfaction.

2. Introduction

Now, to give you some background on this case, we would like to talk a little bit about NetApp. NetApp empowers customers with hybrid multi-cloud solutions. In a world of hybrid multicloud complexity, under the vision to provide a better IT experience, we build solutions that drive faster innovation — an evolved cloud state where on-premises and cloud environments are united as one, always delivering simplicity, security, savings, and sustainability with unified management and AI-driven optimization.

Customers are now looking for a solution on our website that fits their requirements, investigating the advantages of products by comparing products by themselves.

To keep up with the growing volume of content and keep up with frequent updates, we have decided to publish our product manuals on GitHub and make it publicly available.

When a foreign company sells a product in Japan, it is necessary to think about delivering product information in Japanese. Combined with the grammatical characteristics of the Japanese language, Japan has long maintained a low ranking in English acceptance rankings.

機械翻訳活用で顧客の期待に応える
Edith Bendermacher*, 上田有佳子**

*Director, Globalization Strategy and Localization Operations, NetApp Inc.

**ネットアップ合同会社 APAC グローバル化戦略統括

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/>

When NetApp provided product manuals as PDFs, localization was initially done via Human Translation (HT). In 2017, we implemented MT with PE. The process took a long time and publishing of localized content was done few weeks after the English content was published.

However, as we started publishing English content on the GitHub platform, we realized that we couldn't keep non-English-speaking customers waiting for localization process to finish.

MT has been a major topic in the localization industry both domestically and internationally for many years, but especially for Japanese. Until the introduction of Neural Machine Translation (NMT), in general, presenting Raw MT to customers was rarely an option. It was often a headache for Japanese localization service providers (LSPs). Japanese LSPs see this as a research cost for their own company, and it has been common to hear that they accept requests at a lower rate than HT, even in cases where PE requires more time than HT.

When NMT first appeared as a game changer in the industry, many people were skeptical about whether it would work well in Japanese. The output of NMT has steadily improved, and the use of NMT in Japanese has increased. However, despite the significant improvement of NMT quality, and many years dealing with difficult experiences with Rules-based MT (RbMT) and Statistical MT (SMT), it is still not widely accepted for Japanese people to decide to provide Raw NMT output to their customers.

NetApp invests in localization to help our customers understand how great our products are, to gain access

to cutting-edge technology, to try our products, to make purchasing decisions, and use them.

GPSO is not just a Localization Services organization, but a Center of Excellence (CoE) that applies our globalization strategy and methodology to all NetApp departments. Based on our content localization strategy, we identified Raw NMT as a solution to provide our Japanese customers with the same customer experience as our English-speaking customers.

3. Methodology

To cater to the Japanese market, NetApp integrated NMT engine with GitHub platform for product manuals in 2020. While the MT quality has improved significantly with the introduction of NMT engines, the expectation for high-quality translations among Japanese readers remains high. Therefore, our decision to introduce Raw NMT was a risky one. However, we believed that prioritizing the speed of information delivery was crucial to staying competitive in the cloud age.

To ensure a successful Raw NMT implementation, we worked on the following:

- English content to be published must be prepared in accordance with the writing standards
- Train NMT engine with a huge corpus of HT accumulated over the past decade
- Work with Legal to include an appropriate disclaimer on each page to notify readers of the usage of MT

- Communicate with customer-facing teams such as sales team and partners to ensure that this approach meets customer expectations
- Add feedback loop mechanism and establish a workflow to capture and incorporate feedback

4. Results

NetApp's decision to implement Raw NMT for product documentation translation has been positively received by partners and customers in Japan.

In August 2022, NetApp was ranked #1 in the storage category in the Nikkei Computer Customer Satisfaction Survey 2022-2023. Partners have told us that NetApp's way of delivering product documentation has had a significant impact on their efforts. The focus on the speed of delivery of information to the Japanese market is a reflection of our US-headquartered company prioritizing the Japanese market, which is essential for the fast-paced world of cloud computing.

GPSO has also noticed that the Japanese pages of NetApp's product manuals are highly visited by Japanese audiences.

5. Discussion

We have successfully leveraged Raw NMT for our product documentation website, it is important to note that there are potential risks and challenges associated with this approach. There are three key requirements to consider when an enterprise wants to deploy Raw NMT for customer-facing content:

- Clearly state the use of MT in a disclaimer
- Use well-trained, high-quality engine

- Incorporate feedback loops to take advantage of useful user feedback

While not many people in Japan are willing to publish raw NMT results, we believe that our success with Raw NMT serves as a model case for other companies looking for ways to stay competitive in the cloud era.

6. Conclusion

In conclusion, the implementation of Raw NMT for product manuals has allowed NetApp to prioritize the speed of information delivery and achieve high levels of customer satisfaction. While there are potential risks and challenges associated with this approach, NetApp's success shows that it is possible to achieve positive results through careful implementation and management. Using Raw NMT can be a viable solution for keeping up with the increasing demand for getting the latest information in Japan.

The request of "high quality" is something that Japan has been proud of and that should not change. But the speed of business is getting faster around the world. In order for Japan's economy to continue to grow, we must give top priority to understand and apply latest technologies. In a rapidly changing world, speed is key. We can't wait for a high-quality translation to be completed.

We hope that sharing this success story will inspire you to try out NMT and implement it to solve your customer's needs.

7. Acknowledgments

We would like to thank the AAMT secretariat for giving NetApp the opportunity to write this article, and Dr. Eiichiro Sumita, President of the Asia Pacific Machine Translation Association for sharing his research in the Women in Localization session and in his latest book.

A thank you goes to Ms. Yuko Sasaki, machine translation consultant and co-founder of Women in Localization, for inspiring Yukako Ueda to the possibilities of MT 10 years ago.

8. Reference

- Product Documentation, NetApp:
<https://www.netapp.com/ja/support-and-training/documentation/>
- EF English Proficiency Index - A Ranking of 111 Countries and Regions by English Skills:
www.ef.com/epi
- NetApp ranked #1 in Storage category in the customer's satisfaction survey 2022-2023, provided by Nikkei XTEC:
<https://www.netapp.com/ja/newsroom/press-releases/news-rel-20220826-640489/>

解説記事

人間の翻訳と機械の翻訳（7）：社会的な受容の一側面

影浦 峯

東京大学・大学院教育学研究科

1. これまでのまとめと今回の話題

本連載では、前回までで、翻訳の対象、翻訳の対象である文書の性質、翻訳のプロセス、MT 研究と翻訳論の翻訳に関する認識の違い、翻訳者のコンピテンス、翻訳の中核プロセスについて見てきました。前回の冒頭で述べたように、社会、そして機械翻訳（MT）そのものが、大きな領域として残されています。

本稿では、翻訳・MT と社会の関係を考えます。MT は、これまで人間による翻訳では扱われていなかった範囲をカバーするだけでなく、これまで人間による翻訳がカバーしていた領域でも使われ始めています。また、最近では、MT を使った語学学習、MT で語学学習は不要になるかといった話も論ぜられるようになっており^[1]、MT と社会との関係を考える必要性が高まっています。最近リリースされた ChatGPT が広く社会的な話題となっている中で、ChatGPT よりも先に社会的な話題となり、また、起点言語文書があるため「性能」を考えやすい MT の社会的側面を考えると、他の言語関連プロダクトと社会との関係を考えるためにも有用です。

2. MT にできること・できないこと

MT を適切に利用するためには、MT にできること・できないことを理解しておく必要があります。そして、MT にできること・できないことを把握するためには、MT を使って何がどのようになれば目標が達成されたことになるのかを理解する必要があります。では、翻訳や MT、その応用において、そもそも何が求められるのでしょうか。MT と社会との関

係として論じられるべき話題は多岐にわたりますが、いずれも、この点を踏まえないと評価ができないので、この問いは論理的に中核的な問いです。そしてこの問いは、実際にはもう一步後退した問いを喚起します。すなわち「そもそも何が求められるのか」という問いです。とはいえ、いきなりそこに入るとあまり身の回りの抽象論になってしまいますので、わかっている（と思われる）範囲で、MT にできるあるいはできないと主張されていることを見ることから話を始めます。

2.1 MT の得手不得手はどう理解されているか？

注・参考文献^[1]であげた東大新聞オンラインの記事で、トム・ガリー氏は MT について次のように述べているようです^[2]。

- 専門的文書の翻訳などでは、正確な訳を出す
- 主語を省いたり体言止めを用いたりした文章では精度が落ちる
- 小説や歌詞、落語などの文脈や常識の理解を必要とする文章も苦手である

他に、比較的良好に目にするのは、MT は文化の違いを理解するようなものは苦手、というものです^[3]。大雑把に文芸翻訳は苦手、と言ってもよいでしょう。

ところで、これは本当でしょうか。少なくとも三つのレベルでこれが本当かどうかを考えることができます。第一は、現在の MT 研究の枠組みで文脈や文化、省略の補填などを学習できないのかどうかを理論的・技術的に検証すること、第二は、得手不得手やできないを判定する共通の基準があるものとして、それに照らして MT と人間の翻訳の結果を検証すること、そして第三は、このような「得手不得手」が判定され

る基準が実際に社会で共有されているかどうかも含めてどのように作動しているかをやはり実証的に観察すること、です。第一の点は言語処理の領域で議論されていますし、第二の点は筆者が代表を務める科研プロジェクトを含め様々な検討がなされています。ここでは、第三の点を考えます。

2.2 MT と文芸翻訳：最近の一事例

MT は文化を考慮するのが苦手という点については、例えば 2019 年、「AKB 商法」を Google 翻訳は「AKB commercial law」と訳してしまうとの問題が指摘されていましたが、2023 年の DeepL は「AKB business model」とそれなりにきちんと訳します^[4]。これは第一・第二のレベルで扱われるべき問題で、ここで扱うのは第三のレベルなので少し違う話です。

2023 年 2 月 9 日、Record China のサイトに「韓国で権威ある翻訳賞、韓国語が“できない”日本人が受賞し物議＝韓国ネット「何の問題が？」」というニュース記事が掲載されました^[5]。韓国語を学習中でそれほどできるわけではない日本人が、韓国のウェブ漫画を機械翻訳パパゴも利用しながら翻訳し、「2022 韓国文学翻訳賞」という権威ある賞のウェブtoon部門で新人賞を受賞した、というものです。この賞を主管する韓国文学翻訳院^[6]は、これを受け、「新人賞の公募制度の改善と AI との協業範囲について議論することを決めた」と同記事は報じています。記事では、以下のようなネットユーザの反応も紹介されています。

「何の問題があるのか分からない」「パパゴの翻訳文をきれいに整えるのは思いのほか難しい。100%信用することもできない。良い訳が思いつかないときに参考にする程度なら可能だけど、パパゴだけで翻訳賞をとるのは無理」「不自然な表現を直せるレベルなら優秀では？ 数学者が計算機を使っても、数学ができないということにはならない」「パパゴの翻訳は正確でない場合も多い。受賞したのは本人の実力が優れていたからだ」「その国の文化について完璧に理解できていない場合は翻訳プログラムを使うべき」

「それなら電子辞書も禁止にするべきでは？」「今の時代、翻訳に賞を与えるのは意味がない」「AI が発達したからこういう賞は意味がない。他の人たちもみんな使っている。賞をなくすか、翻訳機を使ってもよしとするかどちらか選ばないと」

この例で問題になっているのは MT の利用であって、翻訳の品質そのものではありません。実証的な観点からは、この事例は、MT を適切に使うことは文芸翻訳において有効であること（さほど「苦手」でないこと）を示す事例、と考えることもできます。

この記事にはいくつか気になる点があります。受賞作のどこが優れているか書かれていないこと、MT の貢献がどこまでか明確には書かれていないこと、関連して、MT の利用をめぐる検討が翻訳の質と関連付けられていないこと、そして、ネットユーザの声が紹介され逆説的にもネットユーザが当該翻訳の品質に言及していること、そして、AI 時代に翻訳賞は意味がないという主張があること、です。賞を得たということで翻訳の質は高いという前提はあるのですが、これらは、文芸翻訳の質やそれを誰が評価するのか曖昧で、ネットユーザもその曖昧さの中で翻訳を語っていることを示しています。

2.2 「文芸翻訳は苦手」の基準は何か？

では、翻訳の質、とりわけ文芸翻訳の質とはどのようなもので、誰がそれを担保するのでしょうか。そして、「MT は文芸翻訳が苦手」というときに、どのような基準が想定されているのでしょうか。それを考えるために、人間による文芸翻訳をめぐるなされた議論の事例を取り上げます。

2007 年、光文社古典新訳文庫の一つとして野崎歓氏（当時東京大学教授）の訳によるスタンダールの『赤と黒』が刊行されました^[7]。この訳に多数の誤訳や不適切訳が含まれていたことを、下川茂氏（当時立命館大学教授）がスタンダール研究会の会報で詳細に指摘しています^[8]。指摘された問題は、“M. de Rênal”（レナール氏）を “Mme. de Rênal”（レナール夫人）と

読み違えるといった極めて初歩的ながら意味が通じなくなるものや、熟語の取り落とし、訳抜け、原文になり表現、余計な解釈などを含め、数百か所にのぼると下川氏は述べています。挙げられた例を見ると、かなり深刻であることがわかります。

下川氏の批判に対し、光文社文芸編集部編集長の駒井稔氏は、次のように「反論」しています。

・・・読者からの反応はほとんどすべてが好意的ですし、読みやすく瑞々しい新訳でスタンダードの魅力がわかったという喜びの声だけが届いております。当編集部としましては些末な誤訳論争に与する気はまったくありません。もし野崎先生の訳に異論がありなら、ご自分で新訳をなさったらいかがかというのが、正直な気持ちです^[9]。

また、野崎敏氏は、正面から批判に対応せずに、いくつかの箇所を、変更を明示することなく、新たな刷りで修正したようです^[10]。

この事例では、文芸翻訳の質の基準に関して、いくつか重要な点が観察できます。

- 少なくとも一部の文芸翻訳者と編集者は、原文との対応の不在は些末なことであり、「訳」の読みやすさ・瑞々しさが重要であると考えていること
- 少なくとも一部の文芸担当編集者は、読者に受け入れられるかどうか、すなわち人気は訳の評価において決定的に重要であると考えていること
- 少なくとも一部の文芸翻訳者と編集者は、自らが制作した翻訳に対する批判に自らの責任において対応することを放棄し、文句があるなら自分でやれば、という態度を取ることに
- 原文との対応は気にせず訳文が読みやすく瑞々しいことが重要であるとする読者がかなりいること^[11]

多少乱暴にまとめると、この事例は、読解の解像度の低い読者が読みやすい流麗な文体で読み流せること、そしてそのようなものとして人気を博すことが、良い文芸翻訳の基準だと考えている人が、文芸翻訳に従事する人の内部にあり、読者がそれに対応してそのよう

な翻訳を受容していることを示しています。

このような基準で考えると、「MT は文芸翻訳が苦手」だとは言いきれそうにありません。実際、NMT が社会的に大きく注目された大きな要因の一つは、目標言語の流暢度の大幅な改善でした。原文との対応や文芸作品としての性質をあまり考えずに、多少の「超訳」でも読みやすければ良いというのならば、そして読者の側が、レナール氏がレナール夫人であっても気にしないのであれば、この事例で文芸翻訳編集者が主張した基準に従えば、MT はむしろ文芸翻訳に適しているということさえできそうです。2.2 で述べた事例は、恐らく、韓国文芸翻訳院はこのような基準でない基準を前提とし、MT の利用を手続きとして規制しようとしているものと考えられますが、ネットユーザの一部の発言は、翻訳は何を満たしていなくてはならないかをそれほど意識していないことを示していました。

実際、2.1 の冒頭で述べた、「MT は文芸翻訳が苦手」という主張とその理由を人はなんとなく受け入れがちなようです。とはいえ、連載第4回の「おわりに」でも少し言及しましたが、例えば、文芸翻訳をはじめ「やわらかい」翻訳を MT が苦手とすることに関連してしばしば持ち出される「文化」という言葉が（そもそも何かを示しているとして）どのような概念を指し示すのかは極めて曖昧（抽象的なのではなく）であり、その説明が求められるべきものであって、それを何か他のものの説明に用いることができるような内実を有していません。そのため、「文化」をどれだけ考慮できているかといった点の評価は、受容する人間の側で恣意的に調整され吸収されます。つまり、MT が文化を扱えるようになるのは、文化を扱えるかどうかの共通基準を満たすことで達成されるのではなく、ある程度の数の人たちが、それぞれが「文化」と曖昧に考えていることを MT が満たしているかのようなかたちで受け入れるかどうかで決まることになります^[12]。その点では、翻訳の困難に関して一般的なかたちで「文化」を持ち出す議論は、ここで見た文芸翻訳編集者の主張と近い位置にあると言えます。

3. 社会における翻訳の受容

前節では、翻訳においていわば「極端」なところに位置する文芸翻訳を取り上げ、その中で「極端」な事例の検討を通して、文芸翻訳では受け入れられるかどうかの問題でそれは受容者の曖昧な感覚で決まることが前提となると考えている関係者が現実存在すること、可能性としてはそれが文芸翻訳の一般的な基準として作用しうることを指摘しました。そうであるなら、MT が文芸翻訳に対応できるようになるのは MT の技術的改善によるよりも MT 訳の社会的イメージ改善と需要の促進によることになります。では、翻訳一般についてはどうでしょうか。

3.1 翻訳がやっていること

産業翻訳では、翻訳がやっていることはどのようなことか、翻訳において満たされるべき基準はどのようなものかといった点について、共通理解を打ち立てようとしてきました。本連載では、翻訳の対象(第1回)、翻訳の対象である文書の性質(第2回)、翻訳のプロセス(第3回)、翻訳者のコンピテンス(第5回)、翻訳の中核プロセス(第6回)でそれらの一部を、ときに変則的な視点から捉え直して、紹介してきました。

第2節冒頭で示した「何がどのようになれば目標が達成されたことになるのか」という問いについて、産業翻訳ではその基準を外的・手続的に定義し明確化することを試みてきました。例えば日本翻訳連盟の翻訳品質評価ガイドラインは、翻訳が満たすべき品質の指針を一定の解像度で与えています^[13]。そこでは、品質を「翻訳成果物が、関係者間で事前に合意した仕様が満たす程度」と定義しており(p. 5)、仕様の存在が前提とされています。そして仕様は「クライアントや最終読者のニーズや目的などに基づき、翻訳成果物が持つべき要件をまとめたもの」(p. 5)と定義されます。それを満たすことが翻訳の目標が達成されることとなります。その上で、品質を評価する手法として、専門家による主観評価、最終読者による主観評価、評価メ

トリクスを用いた客観評価、そしてその他の客観評価としてユーザビリティ評価、リーダビリティ評価、法的要件などへの準拠が挙がっています。

産業翻訳では「クライアントや最終読者」が具体的に想定されるため、仕様がクライアントや最終読者を想定して定められ、また「最終読者による主観評価」も具体的なかたちで実施され得ます。一方、産業翻訳に埋め込まれた利用の場ではなく、社会一般での受容においては、MT の利用者は「クライアントや最終読者」としての明確なまとまりを持ちません。本連載の第4回で翻訳論と MT 研究を比較した際、人間に関わる概念が MT 研究の対象として欠けていることを指摘しました。これは MT が個別の文書を特定のクライアントに向けて行うものとしては定義されていないことに対応しています^[14]。

立ち戻って 2.3 で取り上げた文芸翻訳の事例を見ると、不特定の曖昧な読者の人気投票が基準として挙げられており具体的なクライアントや最終読者が定まっていない点では MT と似通っていることが見て取れます。そうであれば、繰り返しになりますが、MT に文芸翻訳ができるかどうかの判定は、不特定の曖昧な読者が受け入れるかどうかにかかっており、その基準は「翻訳がやっていること」をどこまでの解像度で理解しているかに依存し、今のところどうやら少なくとも文芸翻訳においてはその解像度は比較的粗いと思われるので、受け入れられる可能性はそれなりにあるのではないか、ということになります。

3.2 「翻訳がやっていること」

文芸翻訳とは対照的に、産業翻訳ではクライアント(と最終読者)が明確なので、例えば JTF の翻訳品質評価ガイドラインに書かれている事項についてクライアントが全く考えないまま個々の翻訳が進むということとはほぼないと考えられます。現在のところ、賛否はあるものの、産業翻訳における MT の利用も、そうした枠組みの中で行われ(ることになってい)ます。その範囲では、関係者が想定する「翻訳がやっているこ

と」は、翻訳がやっていること、やるべきことに、それなりに対応しているはずですし、クライアントとの間で仕様が定められそこで最終読者も具体的に考慮されることを考えるならば、翻訳が曖昧な一般の読者による人気投票で評価される事態にもなりません。

けれども、MT 自体はこのような範囲を超えて広く使われており、使える使えないをめぐる見解も、MT 研究や開発、産業翻訳における明確な利用の範囲を超えて広く表明されています。しばしばそうした場での議論で想定される「翻訳がやっていること」は翻訳がやっていることに対応せず、漠然としているので^[15]、それが標準的なものとして広まると、翻訳の良し悪しが人気投票によってなされることが現実的に予想されます。それがクライアントにも影響を与えるならば、具体的な読者ではなく漠然とした読者の漠然とした人気を想定し、不適切な MT 利用を求めることが一部のクライアントにおいて慣例化する可能性はあります。既に一部ではそうなっているかもしれません。

4. おわりに

翻訳には、人気投票には還元できない基準が存在します。野崎氏訳『赤と黒』に対する下村氏の批判を「些末な誤訳論争」とし、読者の反応が良いから良いとし、批判があるなら自分で訳せばと言うのは、耐震基準を満たしていないことが指摘された家に喜んで住んでいる人が多いから問題はない、批判があるなら自分で建てれば、と言うようなもので、プロとして責任を持って達成すべき基準を放棄しています。

翻訳がそのように扱われる状況が広まると、翻訳がやっていることへの理解も本来翻訳としてなされるべきことをすることができるスキルも失われ、スキルを身につけた人も育たなくなります。そのような事態を避けるためには、翻訳は何をするものかを、漠然としたイメージではなく具体的に共有可能なかたちで把握する必要があります。なお、そのようなかたちで翻訳への理解が進むなら理解の進展に応じて学習に組み込

むことができるので MT はさらに進化しそうです。それで人間の翻訳者がいらなくなるかについてはここで判断しませんが、漠然としたイメージに基づく人気投票に翻訳を委ねると翻訳者も MT も負のサイクルにはまりその中で人間の翻訳者は翻訳の概念とともにいなくなるので、現在の MT にそうした受容を促す状況があるならば、翻訳業界も MT 業界も、その状況を改善する行動を取る必要があるでしょう。

15 年ほど前、長尾真先生は「言語処理は人間を賢くするものでなくてはけません」と言っていました(割と誘導的に筆者が質問したので申し訳ないのですが)。MT が人間を賢くするかたちで使われるためには人間が翻訳の理解に関して賢くなくてはならないので、それは少し辛いのですが、漠然とした翻訳のイメージにもとづく何となくの人気投票で翻訳が評価されることでこの社会から翻訳が失われる事態を避けるために、さらに頭を使って翻訳を理解する必要がありそうです。それはまた、人間が言語において思考することの理解にも寄与するでしょう。

謝辞

本記事の一部は、科研研究費基盤(S)「翻訳規範とコンピテンスの可操作化を通した翻訳プロセス・モデルと統合環境の構築」(19H05660)に関わっています。本項の一部は、東京理科大学主催「知のフロンティア」第 21 回(2022 年 5 月 26 日)講演会の場で、東京理科大学の松崎拓也氏・大石悦子氏、NICT の藤田篤氏と議論した著者の認識を展開したものです。翻訳者高橋さきの氏との対話は後半をかたちにするきっかけとなりました。韓国文学翻訳賞の話は東京大学特任助教の朴恵氏から教わりました。感謝いたします。

注・参考文献

- [1] ワーマン・カエデ(2021)「第 29 回 自動翻訳で英語学習は不要になるのか | カエデの多言語はぐく

- み通信」<https://torja.ca/kaede-trilingual-29/>; 弓矢基貴取材 (2021) 「AI 時代における外国語学習の意義とは? 漫画機械翻訳エンジン開発者と言語学者に聞く機械翻訳と外国語学習のリアリティー」東大新聞オンライン
https://www.todaishimbun.org/machinetranslation_20210724/. 後者は漫画専用 MT エンジン開発を行っている Mantra の石渡祥之佑氏と言語学者のトム・ガリー東京大学教授に取材した記事です。
- [2] 「ようです」と曖昧なのは、これらが地の文に書かれており、取材者の見解なのかトム・ガリー氏の見解の紹介なのか判断しにくいからです。なお、MT の得手不得手については他に例えば、日本翻訳センター (2022) 「機械翻訳にも、向き不向きはあるの?」<https://www.jtc.co.jp/blog/20220801/> のように、翻訳企業の解説もあり、基本的に似たことを言っています。
- [3] 例えば、翻訳会社ジュピター (2023) 「AI・機械翻訳と人間翻訳を比較してみました」<https://honyaku-jupiter.com/ai・機械翻訳と人間翻訳、そのメリットデメリット/> では、MT の問題として、文化と文脈、セキュリティ、専門知識が挙がっています。
- [4] 西澤ロイ (2019) 「自動翻訳で英語学習は不要になるのか—AI による機械翻訳の限界を知っておこう」リクナビ NEXT ジャーナル
https://next.rikunabi.com/journal/20190612_m01/. DeepL は、<https://www.deepl.com/ja/translator/> で 2023 年 3 月 18 日にチェックしました。なお、同日、Google 翻訳は「AKB Commercial Code」と訳しています。
- [5] Record China (2023) 「韓国で権威ある翻訳賞、韓国語が“できない”日本人が受賞し物議＝韓国ネット「何の問題が?」」
<https://www.recordchina.co.jp/b908920-s39-c30-d0191.html>
- [6] <https://ltikorea.or.kr/>
- [7] スタンダール (2007) 『赤と黒』(上・下) 野崎歓訳, 光文社.
- [8] 下川茂 (2008) 「『赤と黒』新訳について」『スタンダール研究会会報』 18 号. pp. 15-20.
<https://infosjes.iza-yoi.net/kaihou18.pdf>
- [9] 産経新聞東京版 2008 年 6 月 8 日「スタンダール『赤と黒』新訳めぐり対立 “誤訳博覧会” “些末な論争”」. なお、書誌情報も含め、佐藤美希 (2009) 「新訳をめぐる翻訳批評比較」メディア・コミュニケーション研究, 57, pp. 3-4 からの再引用。
- [10] 下川, 前掲 「追補」.
- [11] 2008 年 3 月に第 3 刷が出ていることは売れ行きが好調であることを示しています。
- [12] これに対して「文脈」は、定義の揺れや広がりはあるものの、かなり具体的な概念で(おそらくはそれゆえに) MT で対応可能になってきています。文芸翻訳におけるこの曖昧さは翻訳に限らず文芸の評価そのものにも見られるものの反映と考えることもでき、一部はそうだと思いますが、翻訳においてはそれと関連しつつけれども翻訳における評価の曖昧さがやはり関与しています。なお、文学研究でも「豊かな (rich)」といった曖昧な概念の研究上の有効性はしばしば議論の対象となっています。例えば、Love, Heather (2010) "Close but not deep: Literary ethics and the descriptive turn," *New Literary History* 42(2), pp. 371-391.
- [13] 日本翻訳連盟 (2018) 「JTF 翻訳品質評価ガイドライン」
https://www.jtf.jp/pdf/jtf_translation_quality_guidelines_v1.pdf
- [14] この部分がプロンプト・エンジニアリングとして MT を人間がカバーするかたちで対応できるかどうか、どこまで対応できるかは、今のところ明らかではないようです。
- [15] n=30 程度ですが、プロの翻訳者や翻訳研究者、MT 研究者を除き、著者の知る範囲で、MT を使っていて翻訳にそれなりに関心がある人のほとんどが JTF 翻訳品質評価基準を知りませんでした。

解説記事

ISO18587 供給者自己適合宣言のススメ

森口 功造

AAMT

1. はじめに

ISO 18587:2017 Translation services-Post-editing of machine translation output-Requirements (翻訳サービス—機械翻訳の出力のポストエディット—要求事項) は、機械翻訳が出力した結果を人間が修正する作業 (後修正=ポストエディット) のプロセスに関する国際標準規格である。機械翻訳の精度が継続的に向上している中、産業界ではポストエディットが普及しており、その業務の標準化を目的として、2017年に同規格が発行された。近年では、グローバル企業のサプライヤー選定や入札にも、認証取得が求められるケースが増えている。

本稿では、ISO 18587: 2017 への適合性を評価する1つの手法として、供給者自己適合宣言を行う背景を説明する。続いて、AAMT内に設置した供給者適合宣言支援WGの活動、および年次大会で実施された本稿と同タイトルのパネルディスカッションの内容を報告する。

2. WG 設立の背景

一般的に、品質マネジメントシステムに関する国際標準規格である ISO 9000 シリーズや、情報セキュリティマネジメントシステム (ISMS) のようなメジャーな国際標準規格の認証取得を支援する第三者機関は国内にも多数存在する。そのため、特定の第三者(機関)が実施した適合性評価が、極端に認証基準の甘いものであったり、認証自体が不適格なものになったりしないように、中立的な立場で第三者機関自体を審査する必要性が生じる。このように、認証機関自体の能力を

審査することを認証とは別に「認定(Accreditation)」と言い、日本国内では、公益財団法人日本適合性認定協会 (JAB) がその役割を担っている。

一方、翻訳業界ではかなり普及しているものの、用途や業界が限定される ISO 17100:2015 「Translation services - Requirements for translation services (翻訳サービス—翻訳サービスの要求事項)」のような国際標準規格は、比較的用户の少ない規格と言え、適合性評価を実施する第三者機関の数も少ない。

この場合、上述した JAB のような役割を担う機関はまだなく、第三者機関が顧客のニーズに応じて適合性を評価する、いわゆる「プライベート認証」と呼ばれる適合性評価が実施されることが多い。ISO 17100:2015 の適合性評価についても、日本国内では日本規格協会ソリューションズ社が実施する適合性評価が有名であるが、これもプライベート認証にあたる。

ところが、ISO18587:2017 については、本稿の執筆時点においてプライベート認証を含め認証 (適合性評価) を支援する第三者機関は日本国内には存在していない。

つまり、国内事業者が同規格について第三者適合性評価を受ける機会が制限されており、せっかく策定された国際標準規格の利用促進の妨げとなっている。

ISO 18587:2017 に限らず、こうした場合には、利害関係者である顧客にプロセスへの適合性確認を直接依頼する第三者適合性評価が実施されることがある。

あるいは、事業者 (供給者) が自ら適合性評価を実施した後、適合を宣言する供給者適合宣言を選択することもできるが、前者は顧客ごとに確認依頼する必要があり、規格の内容によっては依頼が現実的でないこと、後者は自社による評価のため、透明性および信頼

性に疑義が残る可能性があることがデメリットとして挙げられる。

言い換えれば、供給者適合宣言を選択した場合、その適合性評価の仕方が不適切であったり、プロセスが不透明であったりすることで、宣言した各企業の適合性のレベルにバラつきが発生してしまうと、規格の信頼性が低下する恐れもあり、供給する事業者にとっても、利用する顧客にとっても不利益が発生することになる。

そのため、高い透明性と信頼性を保って適合性評価を実施するために、その評価プロセスの定義を行うことが急務であった。

3. 供給者適合宣言支援ワーキンググループ(WG)の設立とガイドライン文書の策定

実は、供給者適合性評価の方法については、以下のような規格が存在し JIS 化もされている。

① ISO/IEC 17050-1:2004:

「Conformity assessment-Supplier's declaration of conformity-Part 1: General requirements (適合性評価-供給者適合宣言 第1部：一般要求事項)」

② ISO/IEC 17050-2:2004:

「Conformity assessment-Supplier's declaration of conformity-Part 2: Supporting documentation (適合性評価-供給者適合宣言 第2部：支援文書)」

AAMT では、①では宣言書の書式や記載すべき事項が、②では透明性を高めるための文書として内部監査報告書が選択肢として提案されていることに着目し、適合性評価に必要な規格の解釈、証憑の例と、内部監査の報告書の書式を含めたガイドラインを提供すれば、供給者による適合宣言を一定のレベル以上に保つことができると判断し、ガイドラインの策定を事業計画の中に含めることとした。

2022 年年次総会での事業計画承認を経て、WG に参加する委員の選定を開始した。委員には、ポストエディットを実施している 3 つの翻訳会社から、小野真

帆様 (株式会社アスカコーポレーション)、菊地大悟様 (株式会社十印)、山本耕様 (株式会社サン・フレア) のほかに、第三者認証機関の視点を取り入れるため浜田良和様 (日本規格協会ソリューションズ株式会社)、および学術研究の立場から意見を伺うため山田優教授 (立教大学) をお招きした。そして、筆者はアジア太平洋機械翻訳協会の理事としての立場から、WG の委員長に就任した。

2022 年 6 月 30 に第 1 回の WG を開催し、それ以降、合計 6 回の WG を実施した。ゴールを明確化するため、以下のように明文化した。

■WG の目的

AAMT の会員が、ISO18587: 2017 の供給者適合宣言をする際に、JIS Q 17050-1 で要求される宣言書の様式、内部監査の判断基準、およびその証憑などの例を示したガイドライン (支援) 文書を策定すること

併せて、ガイドライン文書は、以下の 5 つの原則に沿うことを確認した。

- ガイドライン文書は、AAMT 会員にのみ公開される (無料)。
- AAMT 会員のみが本ガイドライン文書に基づいた自己適合宣言ができる。
- JIS Q 17050-1 で要求される宣言書の様式案を提案する。
- JIS Q 17050-2 で要求される支援文書として内部監査報告書を提案する。
- 証憑の例など、規格の文言の解釈について評価チェックリストを提案する。

ワーキンググループの目的が明確化されたため、第一回～第五回のワーキンググループでは、主に規格の解釈と証憑の例を記載した「適合性評価シート」の内容について議論をし、成果物として以下のガイドライン文書および 4 つの付属書のドラフトを策定することを決定した。

■ ガイドライン文書

本ガイドライン文書の目的、構成内容、および宣言書に対する要求事項を取りまとめた文書。

◆ 付属書 A : 供給者適合宣言書

供給者適合宣言書のテンプレートであり、宣言書に含まれるべき項目などを明記した文書。本文書は、ISO 17050-1:2004 の要求事項をベースに策定した。

◆ 付属書 B : 内部監査結果報告書

内部監査報告書のテンプレートであり、報告書に含まれるべき監査の目的、適合性評価の度合い、および検出事項の対応状況などを報告する書式。

◆ 付属書 C : 内部監査実施要項

内部監査を実施するにあたり、内部監査の目的や実施の基本方針、および監査実施担当の資質、検出事項の評価およびその是正処置などを取りまとめた文書。

◆ 付属書 D : 適合性評価チェックリスト

適合性を判断する際に必要な、規格 (ISO 18587: 2017) の解釈、判断基準、および証憑例を記載したチェックリスト。況などを報告する書式。

2022 年 10 月、会社アスカコーポレーション、株式会社十印、株式会社川村インターナショナルの三社をモデル組織として各社で内部監査を実施し、ガイドライン文書の整合性や改善点を洗い出した。

モデル組織に対する内部監査終了後に実施した WG において、修正点および不明点として挙がったフィードバックを文書に反映し、2022 年 11 月 24 日にガイドライン文書の初版 (Ver.1.0) を発行した。文書名は、「一般社団法人 アジア太平洋機械翻訳協会 ISO 供給者適合宣言支援ガイドライン文書 (AAMTSDC2022-1)」とし、AAMT 会員だけが利用できるようにするため、会員ページに公開した (<https://www.aamt.info/ISO/>)。

4. 年次大会でのパネルディスカッション

2022 年 12 月 1 日に実施された当協会の年次大会「AAMT2022, Tokyo」において、「ISO18587 供給者自己適合宣言のススメ」と題してパネルディスカッションを実施した。下記のまとめは、講演資料と発表内容を参考に、筆者が主観的にまとめたものであることにご注意いただきたい。

◎ 翻訳サービスの適合性評価に関わって: 浜田良和氏

日本規格協会ソリューションズ株式会社の浜田氏から、適合性評価および WG が推進する供給者自己適合宣言に関する基礎的な知識について、およびその種類と特長、認証の種類、供給者適合宣言のポイントについて説明があった。

第三者評価機関としての目線で、適合性評価の一般的な位置づけと、本 WG における議論のポイントについて、JIS Q 17050-1 が要求している一般要求事項に触れながら説明した。また、類似している翻訳プロセスの国際標準規格である ISO 17100: 2015 と、ISO 18587: 2019 の要求事項の違いについても解説した。

◎ 供給者適合宣言支援ガイドライン文書の概要: 森口

森口よりガイドラインの構成と適合性評価チェックリストの例 (抜粋)、本ガイドライン文書の入手方法について説明した。

◎ 供給者自己適合宣言をしてみても良かったこと: 菊地大吾氏

株式会社十印にて、ベンダーマネージャを務める菊地氏は、供給者自己適合宣言に取り組んだ動機を 3 つ列挙した。まず、お客様を含めたステークホルダーとの間で共通の言葉を使用することによるメリットについて述べた。次に、宣言をすることにより、営業活動においてわかりやすく、アピールをしやすくなる点を挙げた。最後に、入札やベンダー選定における加点措置についてビジネスチャンスである旨を説明した。

菊地氏は、同社において ISO 17100: 2015 の認証取得にも携わっていることを明かした上で、このことが内部監査を実施した際に、準備の負担軽減につながったことを補足した。

最後に今後の課題として、社員の教育や内部監査員の確保、顧客への説明の重要性を挙げた。

◎ AAMT ガイドライン作成 WG に参加して：小野眞帆氏

医療領域の翻訳に長年の実績のあるアスカコーポレーションで品質管理に携わる小野氏は、新薬申請文書の翻訳において納期の短縮に対するニーズが高まっており、ますます機械翻訳の重要性が増していると述べた。自社で対訳を収集して、医薬業界に特化した機械翻訳エンジンを活用し、納期を約 3 割短縮できた実績もある。規格に対する理解を深め、自社のポストエディットプロセスがグローバルな基準に沿っているか再確認したいという思いが、本 WG に参加する動機となった。

次に、本ガイドライン文書を活用して同社で実施した内部監査について、同社が確認した具体的な項目や、直面した課題、監査の結果について発表した。

課題として、社員への情報共有不足と教育不足を挙げたが、一方で社内の意識改革につながる可能性を見出したことを述べた。また、自社の現状と規格の要求事項とのギャップを明確化できたことをメリットとして挙げていた。

◎ ディスカッション：モデレーター 森口

ディスカッションの中で挙げたやり取りを抜粋して報告する。

◇ ISO 18587: 2017 ではフルポストエディットに関する要求事項だが、実際の案件はどちらが多いのか。

(菊地)最初はライトポストエディットとして始まったが、最終的にフルポストエディット並みの業務を求められることがある。ISO 18587 のように、明文化され

ているプロセスをお客様にも共有して活用したい。

◇ ISO 18587 では、機械翻訳の精度が悪かった時に、ポストエディットができないと顧客に伝えることを要求している。実際にお客様にお勧めしない時はあるか。
(菊地・小野) 実際にそういうケースは多くはないが、作業員からの意見も踏まえてお断りすることもある。

◇ 作業員に対して、作業レベル、ワークフローなどの作業内容を説明する文書は設けているか

(菊地)プロジェクトごとにガイドラインは設けている。注意点や、ターゲットとなる品質を作業員に提供している。

(小野) 菊地氏と同様。お客様と相談して、項目を追加したり、カスタマイズしたりすることもある。

◇ 多言語の案件の評価について、特に規格では言語の方向は求めている。何か意見があるか。

(浜田) 日本規格協会ソリューションズでは、適合性評価を実施するにあたって、スコアを重視している。顧客に誤解を与えない前提であれば、特に指定する必要はないが、自社で対応できない言語をあたかも対応できるように見えてしまうのは良くない。

◇ ポストエディットに適した人材と翻訳に適した人材の適性は異なるように思うが、依頼時に適した人材を効率よく見つけるためのおすすめの文言は何かありますか(翻訳の場合は、対象言語の母語話者であるなど)。

(浜田) 規格の要求事項に応じたプロセスを構築する必要があるが、自社のルールなどをプロセスに含めても構わないことが規格に明記されている。たとえば、作業員の資格に TOEIC の基準点数を追加しても問題がない。

5. 最後に

WG の委員の皆様には大変お世話になった。この場を借りてお礼を申し上げる。多くの翻訳会社へ AAMTSDC2022-1 が活用されることを期待したい。

解説記事

MT ユーザーガイドへのお誘い

AAMT

1. 機械翻訳で失敗しないための手引き

AAMT では、機械翻訳（MT）で失敗しないための手引きとして、『MT ユーザーガイド』を作成し、

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

より配布していますので、ぜひ、ご一読ください。

機械翻訳は、瞬時に翻訳を提供します。これにより、私たちは外国語を理解したり、外国語で情報発信したりできます。しかし、機械翻訳の精度は 100%ではありませんのでコツつまり、「MT リテラシー」が必要です。

『MT ユーザーガイド』を読んで、MT リテラシーを修得しましょう

2. MT ユーザーガイドの目的と対象読者

MT ユーザーガイドの目的は、MT を正しく使うために必要な知識（MT リテラシー）を提供することです。MT を正しく使うためには、まず「翻訳とは何か」についての知識が基礎となります。そのため、本ガイドでは MT だけでなく、人間が行う「翻訳」についても説明します。

リテラシーが問われる項目は、大きく分けて次の 4 つです。

- ・ 技術の知識：MT の技術的概要を理解する
- ・ 翻訳品質の知識：翻訳品質の評価方法を理解する
- ・ 翻訳制作プロセスの知識：正しい翻訳が作られる工程を理解する
- ・ 法的知識：MT の利用をめぐる秘密保持、責任、著作権を理解する

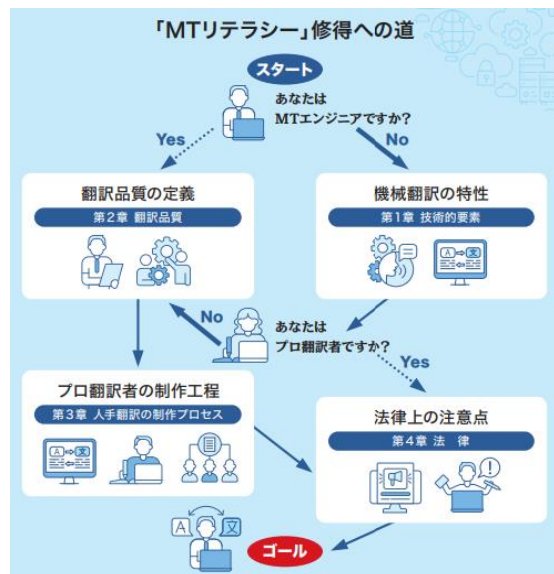
MT ユーザーガイドの対象読者は、MT を利用するすべてのユーザーです。その中でも特に、MT を利用

して情報を発信しようとする人、および（MT を利用した）人手翻訳を翻訳会社に依頼しようとする人には必読です。MT 開発者、翻訳発注者、翻訳会社、プロ翻訳者、教育者、研究者にも役立つ情報を提供しています。

「MT を利用して情報を発信する」ことについて知ることは、「情報収集のために MT を利用する」時に注意すべき点をも包含します。本ガイドでは、最終的にできあがる翻訳文書の品質を担保し、MT に特有のリスクを回避するために留意すべき点をまとめました。

3. MT リテラシー修得への道

MT リテラシー修得への道は険しくありません。下の図に沿って、必要な章だけを読めば OK です。



4. 終わりに

『MT ユーザーガイド』を気に入っていただけたら、お近くの方にご紹介・再配布いただけたら幸いです。

Invitation to MT User Guide

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/>

第2回 AAMT セミナー

理工系の英語論文執筆における課題と機械翻訳の活用

西山 聖久

タシケント工科大学

1. はじめに

本稿は、2023年1月、筆者が登壇したセミナー「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」についてまとめたものである。筆者は工学部の教員として、長年にわたり理工系のための英文ライティングの啓蒙活動を行ってきた（西山ほか2015a, 2015b, 西山2019）。その一環として、多数の工学部学生の英語論文執筆指導を行うなか、数年前から機械翻訳を活用することにより学生が書く英文のレベルが目に見えて上がってきた。筆者自身も機械翻訳を活用しており、大変便利であると感じている。一方で、機械翻訳の訳出する英文には誤訳などのトラブルが起きることもあり、現時点では機械翻訳を完全に信頼することはできないだろう。これを踏まえ、本セミナーでは、主に、英語論文執筆における課題と対策、理工系の英文に求められること、3Cな英文と機械翻訳、機械翻訳に頼ってはいけないことについて講演した。

2. 英語論文執筆における課題と対策

英語で論文を書くことは、理工系の学生や研究者による英語学習の目標の一つである。英語論文執筆で苦労する多くの学生を指導する中、彼らが英語論文を執筆できないのは、そもそも論文の概要が自分の頭の中で整理されていないことにより、内容の方向性が教員や共同研究者とうまく共有できていないことが原因であると考えられた。英語以前に、内容が整理されていない状態で執筆された論文は理解することは不可能であり、たとえネイティブや業者に添削してもらったとしても、そもそも内容が理解できないというのでは、論文を投稿できるレベルにまで仕上げることは難しい。

そこで筆者は、研究が人類が共同して問題解決に取り組むための報告書であるとの考えのもと、企業などで活用される問題解決手法（産能大学VE研究グループ・土屋裕, 1998; ダレル・マン, 2014）を活用して、論文の概要のための以下のようなテンプレートを考案した。

【1】(する)ためには、【2】(する)が必要である。しかし、この際、【3】(する)ことが問題となっている。そこで、【4】(する)ことがその解決策として検討されている。その実現のためには、【5】を明らかにすることが必要である。これに関連して、昨今は、【6】が着目されて多数の研究が行われている。しかし、【7】に関する知見は未だ十分に得られていない。そこで、本研究では、【8】と【9】の關係に着目し、【10】による【11】への影響を調査する。【10】の調整は【12】により行い、【11】の測定は【13】により行う。

テンプレートの空欄となっている13項目について考えることにより、研究の内容を明示するために最低限必要な情報を吟味することができる。これらに、研究の結果、考察、結論を加えることで、一定の品質に達した論文の概要の叩き台を完成させることができる。テンプレートをを用いて作成した概要に合うように本論の構成を行うことにより、英語論文執筆が行き詰まることを防ぐことが可能となることを強調した。

3. 理工系の英語に必要なこと

ここでは、理工系のための英文ライティングの前提として3Cの考え方の概要を説明する。3Cについての詳細は『技術系英文ライティング教本』『英語論文ライティング教本』中山, 2009, 2018を参照されたい。

理工系の英語とは、例えば、論文に限らず、技術報告書、機械や装置の使用マニュアルといった理工系の産業や研究に関連した資料などにおいて使用される英語を意図している。

特に英文ライティングを中心に、理工系の英語は大学入試や TOEIC などの英語検定試験で必要とされる英語とは大きく異なる。試験においては、学習を頑張った人だけが理解し、解答できるように意図的に難解な英文が含まれる一方、論文や技術報告書、マニュアルなどの英文は、英語が苦手な人でも理解できる内容や表現を意識する。これは、例えば、難しい英語で書いてある機械のマニュアルを英語が苦手な人が読み、間違った理解に基づく操作をしたことで怪我をした、などという事態は許されないからである。

このような理由から、理工系の英文ライティングでは、英文を誰もが理解できるよう、明確、正確、簡潔であることが求められる。これは、3C (Clear: 明確、Correct: 正しい、Concise: 簡潔) として知られている。

3C の考え方を理解することのメリットは、自分が作成した英文に根拠を与えることができるようになることである。私の周囲の学生や研究者の中には、自力で英文を書き上げた後に、ネイティブや添削業者にチェックしてもらったことをもって完成としてしまう人が少なくない。しかし、たとえ英語に精通しているネイティブや添削業者であっても、誤りを含む英文の意図を正しく理解し、分かりやすく書き換えることはできない。そこで、3C の考え方を理解することにより、根拠のある英文を作成し、ネイティブや添削業者、同僚との深い議論を通じた英文のブラッシュアップをすることの重要性を強調したい。

3C な英文を書くためのポイントは、(1) 誤りが少ない英文を書く、(2) できるだけ少ない語数にする、(3) 名詞を意識して扱う、の三つである。

以下、それぞれの項目について説明する。

(1) 誤りが少ない英文を書く

英文における代表的な誤りには、文法が間違っている、単語やつづりが間違っている、表現が間違っている、

主語と述語が一致していない、名詞の扱い方が間違っている、カタカナ語を英語であると勘違いしている、などが挙げられる。これらについては、大学入試や英語検定試験においても気を付けるべきことなので、特に目新しいことはないであろう。

(2) できるだけ少ない語数にする

日本語に引きずられた英文は複雑で長くなりがちである。そこで、極力シンプルな文型(SVO、SVC、SV、S: 主語、V: 動詞、O: 目的語、C: 補語)を使うように意識する。これは、限られた語数の中に、より多くの情報を格納した方が読みやすいからである。特に能動態(SVO の文型)を用いることを意識することによって、受動態の多用や、冗長になりがちな「It...to...」「It...that...」「There is...」「there are...」「If ...」「when...」などの構文の使用を避けることができる。これらの表現は、日本語との相性が良いため多用しがちであるが、文が長くなる上に誰が何をしたのかといった重要な情報(主語と動詞)が文の後半に出てくるため、読みにくい英文になってしまうので注意が必要である。

原則的に能動態にすることにより、英文はかなり読みやすくなる一方、動作主が重要でない、動作主が分からない、動作主を隠したい、主語をそろえたい場合には、受動態の使用を検討する。これにより、英文の語数をさらに削減したり、読みやすくすることが可能となる。

(3) 名詞を意識して扱う

名詞の扱い方とは、ある名詞を単数にするのか複数にするのか、冠詞は「a」にするのか、「the」にするのか、無冠詞にするのかと言った判断をさす。これらの選択を正確に扱うことは、3C の英文を書く上で極めて重要である。ある程度の長文で名詞の扱い方が一貫していない場合、内容自体の正確さにも疑問が持たれてしまうかもしれないので注意が必要である。

4. 機械翻訳と理工系の英作文

筆者は特に理工系に関連する英文作成においては、機械翻訳を活用すべきであると考えている。その理由の一つは、機械翻訳は理工系の英文に求められる 3C (Clear, Collect, Concise) の考え方に沿った英文を出力してくるためである。ここでは、機械翻訳が訳出する英文と 3C の関係について説明する。

(1) 誤りが無い文を書く

日本人が英作文で起しやすいミスの一つに、英語ではないカタカナ語(サラリーマン、バックミラー、ノートパソコンなど)を英語と勘違いして使うことが挙げられる。これまで筆者は学生に、カタカナ英語を英訳する時は必ず正確な英語を確認するよう指導してきた。ところが、機械翻訳は、カタカナ英語を正確な英語に瞬時に訳す上、人間のようにカタカナ語を英語と勘違いするようなことはしない。その他、先に示した、他の「英文における代表的な誤り」についても、機械翻訳が人間のような間違いをすることはないだろう。

(2) できるだけ少ない語数にする

例えば、機械翻訳は「There is...」「There are」構文を使わず、SVO の簡潔な英文を出力する。例えば、「酸化ナノワイヤによりその薄膜が形成された可能性がある。」と言う日本語を英訳する場合、日本語を見て、「There ...」の構文を使い、「There is possibility that the oxide nanowire formed the thin film.」としてしまう。先の説明に従えば、書き手は、SVO の構文を意識し、「The oxide nanowire 【S】 may have formed 【V】 the thin film 【O】.」とすべきとした。しかし、機械翻訳を活用すれば、SVO の英文が自動的に出力される。その他、機械翻訳は、3C の英文においては避けるべきとした、「It...to...」「It...that...」「If ...」「when...」の乱用も避けるようである。

5. 機械翻訳に頼ってはいけないこと

上述の通り、理工系の英文においては機械翻訳を活用することによりライティングの効率を高めることができ

る。しかし、現時点では機械翻訳により対応できないこともあり、完全に頼ることはできないことには注意が必要であろう。これに関しては、機械翻訳が

3C の英文に訳しやすい日本語を入力し、書き手が機械翻訳の弱点を理解しつつ、訳出された英文を修正することが重要になるであろう。筆者は、機械翻訳が対応できない英作文の作業として、主に(1)名詞の扱い方、用語、時制の一貫性、(2)能動態、受動態のどちらにするかの判断、(3)長い文章の英訳(パラグラフの構成)、(4)専門家としてのこだわり、が挙げられると考えている。以下、各項目について説明を行う。

(1) 名詞の扱い方、用語、時制の一貫性

先に説明した通り、単数系や複数形、冠詞の選択といった名詞の扱いは、英文を書く上で極めて重要なポイントである。適切な名詞の選択は書き手にしかできない作業であり、機械翻訳は、書き手の意図通りの名詞の扱い方をすることは限らない。例えば、ある英文において、orange (オレンジ) という単語を使う場合、そのオレンジが、一つであるのか、複数であるのか、液体のような数えられない状態なのか、また、特定のオレンジを指しているのかといったことは、本来は、本人にしか分からない。機械翻訳を使用して英作文を行うとしても、作成した英文の中のすべての名詞について書き手は自らの意図に合った選択をしなければならない。筆者はこのために名詞の扱い方に関するフローチャートを考案した。

その他、機械翻訳が苦手な分野として、用語や時制の一貫性も挙げられる。例えば日本語の「車」は、「car」「vehicle」「automobile」など、異なる英語に訳されるため、どの名詞を使うかは書き手が判断する必要がある。フローチャートや、用語、時制の一貫性についての詳細は拙著『理工系の AI 英作文術』を参照されたい。

(2) 能動態、受動態の判断

理工系の英文においては日本語に引きずられた受動態を避けることが重要であるとした。基本的に、機械

翻訳は受動態の日本語を入力すると受動態の英文を作る。意識していないと受動態が多くなりがちなのが日本語である。そこで、読みやすい英文を作成するために、基本的には能動態を意識して日本語を入力することで能動態の英文を作成することを薦めたい。ただし、先に英文は基本的に能動態の方が読みやすいが、受動態にする方が良い場合もあることを述べた。この、ある英文を受動態にする方が良いかどうかの判断は、機械翻訳にはできないので書き手が行うべき作業の一つであろう。

(3) 長い文章の英訳

長い日本語の文章を機械翻訳に英訳すると、誤訳が発生する可能性が高くなる。これに関しては、あらかじめ短く分割した日本語を入力し、出力された英文をパラグラフの構成に合わせてつなげることで対処することができる。

(4) パラグラフの構成

正確な英文を訳出できるよう機械翻訳には、短く分割した日本語を入力すると良いとした。ただし、さらに読みやすい英文とするためには、パラグラフの構成を意識する必要がある。パラグラフとは、英語特有の概念で複数のセンテンスのまとまりを指し、一行目のトピックセンテンスで話題を提供し、続くサポーティングセンテンスで話の内容を補足する。

機械翻訳で作った短い英文をうまくつなげて、パラグラフに仕上げるのは人間の役目であり、このためには、文と文のつなぎ方を学んで理解しておく必要があるだろう。文と文のつなぎ方に関しては、参考文献『英語論文ライティング教本—正確・明確・簡潔に書く技術—』が詳しい。

(5) 専門家としてのこだわり

そもそも、英作文の目的は、書き手自身が伝えたい情報を伝えることであり。翻訳された英文に見られない表現や単語があったら、自分で考え、調べ、適切な表現を選ばなければならない。判断に迷う時は、最後は専門家としての自分の知識や経験を信じて決めるべきであろう。

6. 機械翻訳の活用手順

ここまでで説明を踏まえると、機械翻訳の利点を最大限に引き出しつつ、その弱点を補うためには、(1) 日本語の文章を短く分割する、(2) 「主語」「目的語」「動詞」の構文にする(婉曲な表現、慣用・比喩表現・ことわざなどに注意)、(3) 機械翻訳に入力して英訳する、(4) 文の構成、単語や表現、受動態の使用、名詞の扱い方について検討する、(5) 修正した英文を機械翻訳で日本語にして確認する、の順序で機械翻訳を活用して英作文を行うのが良いだろう。この手順を経てもうまくいかなない部分は(1)に戻り再検討する。この手順に従えば、誰もがかなりの高品質の英文を作成することができるであろう。

7. おわりに

本セミナーでは、まず、機械翻訳の活用以前に、英語論文執筆などだが、そもそも論理的かつわかりやすい内容を練ることが重要であることを強調した。そして、理工系の英語に必要な3C(Clear: 明確、Correct: 正しい、Concise: 簡潔)について紹介し、3Cの英文を書くために機械翻訳は極めて有効なツールであることを述べ、その弱点も踏まえた上で機械翻訳活用の手順を提案した。

機械翻訳の登場以降、英語学習の意義についての議論がなされるようになっていく。筆者は、少なくとも、理工系の分野では、現段階では、機械翻訳の登場により、英語学習が不要になることはないが、英語学習のあり方には大きな変革が迫られると考えている。機械翻訳は英作文の作業の負担を大幅に軽減する一方で、依然、英文の書き手の判断が求められることは存在しており、今後の英語学習においては、これらのポイントが、より重視されるべきであろう。いずれにせよ、今や、誰もが機械翻訳を活用して作成した3Cな英文により効率良く情報発信を行い、専門分野の知識や経験を深めることにより、仕事や研究の成果を上げることが可能となったことを指摘し本稿を締めくくる。

なお、英語論文執筆や機械翻訳に関する詳細は、拙著『最短ルートで迷子にならない！理工系の英語論文執筆講座』や『理工系の AI 英作文術』に詳しいので、興味のある方はご一読いただきたい。

参考文献

- 西山聖久 (2022). 『理工系の AI 英作文術』化学同人.
- 中山裕木子 (2009). 『技術系英文ライティング教本: 基本・英文法・応用 単行本』日本能率協会マネジメントセンター.
- 中山裕木子 (2018). 『英語論文ライティング教本—正確・明確・簡潔に書く技術—』KS 語学専門書.
- 京都大学 HP「令和3年度 試験問題および出題意図等」
<https://www.kyoto-u.ac.jp/ja/admissions/undergrad/past-eq/r03-eq/>(2023 年 3 月 21 日閲覧)
- 西山聖久・レレイト・エマニュエル・曾剛・古谷礼子 (2015). 「工学分野における科学技術英語ライティング教育の実施と課題」『工学教育』, Vol.63-4, 10-15.
- 西山聖久・レレイト エマニュエル・曾剛・古谷礼子 (2015). 「工学分野英語 教育の為の語彙に関する考察—大学入試, TOEIC と工学分野の比較から—」『名古屋高等教育研究』, 15 号.
- 西山聖久 (2019). 『最短ルートで迷子にならない！理工系の英語論文執筆講座』化学同人.
- 土屋裕・産能大学 VE 研究グループ (1998). 『新・VE の基本 価値分析の考え方と実践プロセス』産能大学出版部.
- ダレル・マン著・中川徹 監訳・知的創造研究グループ訳 (2014). 『体系的技術革新—新版矛盾マトリックス matrix 2010 採用 (TRIZ 実践と効用)』クレプス研究所.

第3回 AAMT セミナー

海外投資家を引き付ける英文 IR～第3回 AAMT セミナー要旨

西村 賢治

ジャパンエレベーターサービスホールディングス株式会社 常務執行役員 IR 室 室長

1. IR という仕事

筆者は現在、ジャパンエレベーターサービスホールディングス株式会社に IR を担当している。IR というのは Investor Relations の略であり、全米 IR 協会の定義によると、IR とは、「企業の証券が公正な価値評価を受けることを最終目標とするものであり、企業と金融コミュニティやその他のステークホルダーとの間に最も効果的な双方向のコミュニケーションを実現するため、財務活動やコミュニケーション、マーケティング、そして証券関係法の下でのコンプライアンス活動を統合した、戦略的な経営責務である」とされているⁱ。当社は東京証券取引所（以下、東証）プライム市場に上場しているが、上場企業として市場から適正な評価を受ける、すなわち適正な株価を形成するために市場とコミュニケーションをとることが主要業務である。

適正な株価とは何か。様々な議論があるが、株主資本利益率（ROE）、すなわち「株主が提供した資本を会社がどれだけ増やすことができるか」に集約されるという意見に反論の余地は少ないのではないかと。実際、東証上場企業の ROE と株価純資産倍率（PBR、会計上の純資産額と投資家間の売買によって時価として決まる株価の関係）の関係をみると、一部の例外を除き、概ね正の相関があることが指摘されている。ROE は親会社株主に帰属する当期利益を純資産で除することで算出されるが、実際には財務・非財務の様々な要素が影響している。最近では ESG の要素も関係しているという見方も広がっており、そうした要素に関する投資家の問い合わせも増えている。こうした財務・非財務の諸要素を分かりやすく整理し、株主をはじめとするステークホルダーに説明することで、ROE の期待値

が株価に反映される。当社の場合、ROE は 2023 年 3 月期の実績で 25.1% であり、これは東証プライム上場企業の中でも高い方である。このため、当社の PBR も東証プライム上場企業の中で上位 20 番以内に入る。

また、ROE の水準が同じ企業でも、投資家層の違いによって PBR の水準が異なるケースもよくみられる。例えば外国人保有比率の高い企業である。当社もその代表企業であり、外国人保有比率は 40.0% と市場平均（約 30%）を大きく上回っている。外国人投資家に本邦投資家と同質の情報を提供するため、今年 5 月にホームページの改修を行った。主な変更点は、上場来日本語のみで開示していた有価証券報告書をすべて英訳し、ホームページに掲載したことだ。

2. 英文開示の必要性和実勢

日本の上場企業が、関わりを持つすべてのステークホルダーの権利を保護し、共に価値を高めていくためにどのような行動をとるべきかについては、2015 年に導入されたコーポレートガバナンス・コードに記載されているⁱⁱ。東証は従来の「東証 1 部」「東証 2 部」「マザーズ」という市場区分を 2022 年 4 月より再定義し、現在は「グロース」「スタンダード」「プライム」という 3 つの市場区分が設けられている。「グロース」は高い成長可能性を有する企業向けの市場、「スタンダード」は公開された市場における投資対象として十分な流動性とガバナンス水準を備えた企業向けの市場、「プライム」はグローバルな投資家との建設的な対話を中心に据えた企業向けの市場と定義されている。この変更と併せてコーポレートガバナンス・コードには「上場会社は、自社の株主における海外投資家等の比率も

踏まえ、合理的な範囲において、英語での情報の開示・提供を進めるべきである。特に、プライム市場上場会社は、開示書類のうち必要とされる情報について、英語での開示・提供を行うべきである」という文章を追加している（下線は筆者による）。

実際、上場企業がどの程度の情報を英文で開示しているのだろう。東証が毎年、上場企業に対してアンケート調査を実施しており、直近の2022年12月に公表された資料ⁱⁱⁱによると、主要開示情報の英文開示状況は下図の通りである。

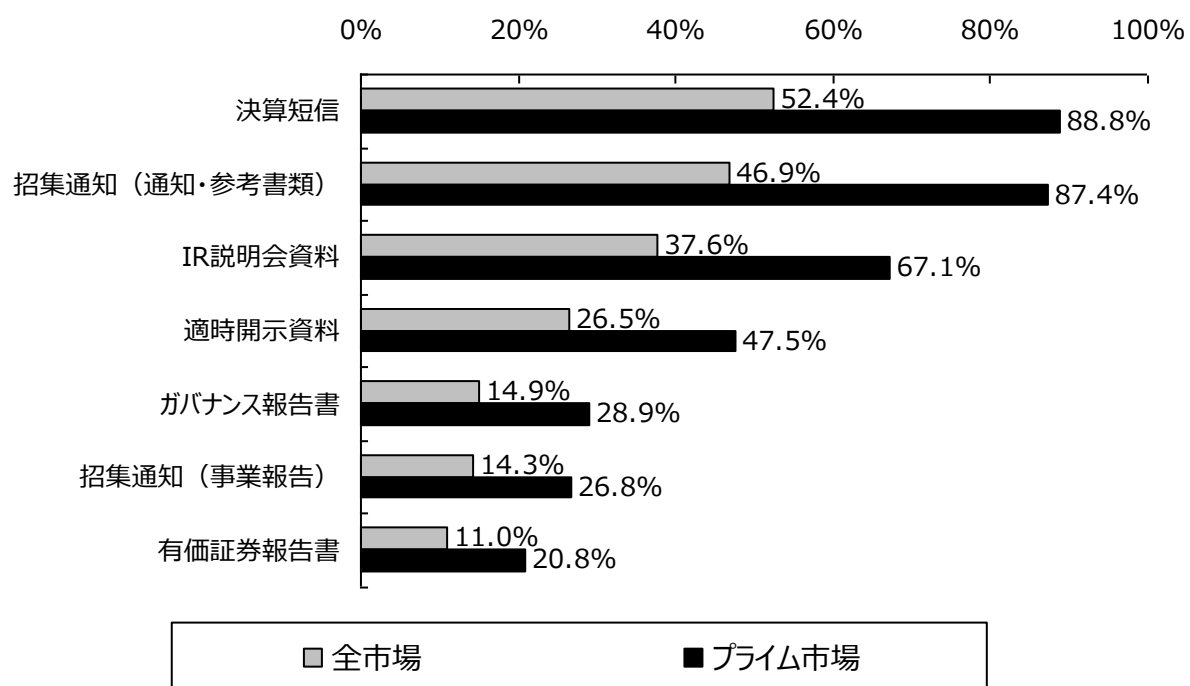
プライム上場企業については約9割が決算短信や株主総会招集通知（通知）を英訳して開示していることが確認される。これで充分だろうか？筆者が問題だと考えるのは、図の下の方で、株主総会招集通知（事業報告）については3割、有価証券報告書についてはわずか2割しか英文を開示している企業がないことである。

ここで、決算短信とは何か、有価証券報告書とは何か、について補足しておこう。有価証券報告書は会社の成り立ちから現在までの経緯、ビジネスモデルや財

務諸表、財務諸表の注記など100ページを超える文章で説明しているもので、米国でいえば10-K、欧州ではアニュアル・レポートと呼ばれているものにあたる。欧米ではサステナビリティに関する記述も充実させているため数百ページにわたる文書である。決算短信は四半期毎の決算発表時に開示される、財務諸表と重要な補足のみ説明しているもので、10ページ前後の短い文書である。投資家にとって重要なのは将来にわたってのROEの期待値を計算するために必要な情報であり、そのためには毎四半期の財務諸表のみならず、その注記を読み込んで様々な会計方針を確認するほか、ビジネスモデルの妥当性、リスクについて検討しなければならない。そのような重要な情報を含んでいる有価証券報告書について、プライム上場企業の2割にあたる企業しか英文開示していないのである。

株主総会招集通知も同様である。株式会社の最高意思決定機関である株主総会において、剰余金の処分や取締役の選任など、株主の決議を促すため、参考となる財務状況やガバナンス体制、報酬決定方針などを事

図：東証上場企業の資料別英文開示実施率



出所：東京証券取引所「英文開示実施状況調査集計レポート（2022年12月末時点）」

業報告で説明し、その上で招集通知において議案を説明するものである。これについても通知部分はプライム上場企業の8割の企業が英文で開示しているが、その前提である事業報告については3割の企業しか英訳していない。

3. なぜ英文開示が普及しないのか

なぜ日本の上場企業のほとんどが投資家に必要な情報を英文で開示しないのか。いくつかの課題が挙げられよう。例えば、決算数値確定から決算発表までの時間が流動的かつ限定的であること。決算数字は社内で確認するだけでなく、監査法人と会話しながら正確性・適正性を担保しなければならない。このため、決算数字が確定し、決算短信や有価証券報告書としてまとめられるのが決算発表直前、ということもありうる。これを英訳し、ネイティブ・チェックを受けている時間は限られている。また社内にリソースがなく外部に翻訳を依頼する場合にはさらなる時間とコストが発生する。英訳にかかる関係者が増えるほど、記述や数字を誤訳する可能性は高まり、適時開示文章の正確性・適正性は失われるというジレンマがある。さらに、日本語は主述関係を曖昧にしたまま文章を完結することができ、略語を多用することが多いが、これを英語に訳すと、どうしても冗長な文章になる。テキストのみの文章であれば問題はないが、図表やフローチャートが入っていると、日本語から英語に訳した際にレイアウトを再構築する必要がある。これにも相当の時間を食われることになる。

同業者からこうした話を聞いている中で、筆者が感じているのは、日本の上場企業の開示担当者は少し硬直的に考えすぎているかということだ。まず、正確性・適正性については、適時開示文書の正本はあくまでも日本語の文書であり、英語は参考文書である。どの英文資料を見ても、表紙には「この文書は原文である日本語を翻訳（妙訳）したものであり、両者に相違がある場合、日本語が優先する」と、ディスクレーマーが

掲載されている。もう少し、肩を張らずに取り組んではいかがだろうか。

また、機械翻訳の精度はここ数年、著しく改善している。数年前までの機械翻訳は確かに、短い文書までしか実用には耐えず、しかも非常に古臭い、固い文書しか提示されなかった。しかし、最近の機械翻訳では100ページを超える文章でも1時間もかからず、こなれた英文が提示されるようになってきている。しかもAIで学習するので、会社特有の文書の癖にも対応するようになってきている。決算短信などは5分とかからないのである。

そもそも決算短信や有価証券報告書は会社法や金融証券取引上といった法規制に基づき、会社の財務状況の説明をするものであり、主述の分かりやすい文章が多い。投資家は事実関係が知りたいのであり、これを“きれいな”表現に直す必要はなからう。世界の資本市場参加者は様々な国籍の方がおり、共通言語として英語を使用しているケースが多い。米国の証券市場で投資家保護のための法規制に責任を持つ米国証券取引委員会（SEC）は、簡素な表現を推奨している^{iv}。美しく見せるためのネイティブ・チェックはどこまで必要なのだろう。また決算短信や有価証券報告書の中の文章は、一度使った文章構造が一部表現を変えながら翌年も使用されることが多い。100ページを超える文章で前回の文章と同じ部分と変わった部分を瞬時に判断して変わった部分だけを翻訳でアップデートするのは機械翻訳の方が適しているのではないか。

4. 機械翻訳の精度を上げるための工夫

機械翻訳の精度は上がってきているが、完全ではない。誤訳は発生する。しかし、なぜ誤訳が発生するのかという背景を理解しておけば、英語に対する知識がそれほど高くなくても対応は可能になる。

最初に認識すべきことは、誤訳は翻訳エンジンの問題というよりも、日本語の構造や使い方に因って発生することが多いということだ。先にも触れたが、日本

語は主述を曖昧にしたまま文章を完結することが可能な言語だ。説明している主体は経営者なのか、会社なのか。「誰が」「何をする」ことが明確でない文章は、翻訳者によりあらかじめ原文を補足することも必要である。

複雑な事象の説明にあたっては後ろに説明が続き、「誰が」「何をする」ことが分かり難くなるのはどの言語にもあることだが、日本語はさらに様々な括弧を使って説明することもあり、誤訳に繋がりやすい。特に会計方針の説明においては法令の引用が多くなる。こうした表現については、枝葉の部分を取り除き、「誰が」「何をする」かの根幹を先に訳しておいて、枝葉の訳を補足していくことで解決できる。

よく日本語を英語に翻訳すると原文に比べて文字数が増えると言われているが、簡潔な文章は日本語も英語もそれほど文字数に相違はない。むしろ、英語訳の文字数が日本語訳の倍以上になっている場合には、略語が多用されているか、日本語が分かり難いので誤訳が生じているケースが多い。そうしたところを重点的に検証していけば、機械翻訳の検証作業も効率化されるのではないかな。

筆者は前職で約5年、米国の上場企業について調査する立場にあったため、米国企業の年次報告書(10-K)を読み込んでいた経緯がある。このため、英語ネイティブではないものの、決算短信や有価証券報告書の英訳にあたり、どのような単語を使ってどのような表現をするのが適当か、一定程度の判断基準を持っている。一方、そのような経験のない、ネイティブでもないIRオフィサーが機械翻訳の結果を評価することは難しいのも理解できる。しかし、だからと言って上場企業の株主の3割、売買代金の6割を占める外国人投資家に日本語と同程度の情報を公開しなくてもよいという結論にはならないと思う。もっとフレキシブルな翻訳スタイルが出てきてもおかしくないのではないかな。日本語が正本であるという前提の上で、機械翻訳により時間を有効に使い、英文エディター（ネイティブである

必要はない）が限定的な正確性・適正性を担保するようなサービスが出てくれば、もう少しこの比率は上がるのではないかな。

5. セミナーを終えて

IRオフィサーの英訳の実態という、かなりニッチな話題であり、どの程度話を聞いていただけるのか心配していたが、予想以上に多くの参加者があり、30分にはわたって様々な質問をいただいた。機械翻訳の可能性を見極めたいというニーズはかなり強いのであろうと感じた。利用頻度が高まればそれに比例して翻訳エンジンの精度も高まるのだろうと考えているので、是非一緒に普及を後押ししたい。筆者のような立場の人間がどれだけ見ているかは分からないが、株式市場を動かすひとつの要素として外国人投資家の存在は無視できないはずである。冒頭のROEの話が続けると、IRオフィサーはROEを変えることはできないが、外国人投資家に認められ、企業価値が上がれば、数十億円、数百億円レベルの経済効果を生み出すこともできる。翻訳に係る費用がハードルとして挙げられることが多いが、英文開示を充実することで企業価値が上がるのであれば、著しい費用対効果を得られるとは言えないかな。

ⁱ Investor Relations の定義は National Investor Relations Institute のホームページ「About NIRI」の中に掲載されている。

<https://www.niri.org/about-niri>

ⁱⁱ コーポレートガバナンス・コードは金融庁もしくは東京証券取引所に掲載されている。

<https://www.jpix.co.jp/equities/listing/cg/index.html>

ⁱⁱⁱ 報告書は東京証券取引所ホームページに掲載されている。

<https://www.jpix.co.jp/equities/listed-co/disclosure-gate/survey-reports/index.html>

^{iv} 米国証券取引委員会のホームページには明確な開示文書を作成するための英語ガイドブックが提供されている。

<https://www.sec.gov/reports-and-publications/investor-publications/newsextrahandbook>

温故知新

温故知新 2

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いです。

本号では、1991 年 11 月と 1992 年 1 月号の JAMT ジャーナルから 2 つの記事を転載して紹介します。これらは、PDF から OCR で読み取ったテキストを修正したものです。画像等を含むオリジナルの原稿については、上記サイトをぜひご覧ください。

【1991 年 11 月 JAMT ジャーナル No.3 より】

F1 マシンと機械翻訳

牧野武則

F1 グランプリも最終戦、セナとマンセルのチャンピオン争いも最後まで楽しませてくれます。体力のぎりぎりまで使い果たしマシンを走らすドライバー達、必死の作業で彼らを支えるメカニック、マシンの改良に明け暮れるコントラクター。そうしたスペシャリスト達が一体となってサーキットを盛り立てます。

そんなサーキットの賑わいを見ながら、80 年半ば、機械翻訳システムの開発が日本で一気に立ち上がった頃を思い出します。それぞれ異なった翻訳方式を唱える各社の開発競争、ユーザとの折衝、ユーザ側でのシステムや辞書の気が遠くなるようなチューンナップ作業。そして、機械翻訳と言うサーキットの中でシステ

ム開発側とドライバーに当たるユーザが一緒になってレースを始めたわけです。

機械翻訳システムの開発は、それ自体がビッグプロジェクトです。各社とも、おそらくシステムの開発に 10 億円をはるかに上回る投資をしたことでしょう。F1 マシンの開発費にも匹敵するかも知れません。それだけ企業にとっても機械翻訳システムの開発は魅力的だったのです。社内で大量に発生する製品のマニュアルは、製品のライフサイクルが短くなっていく中で、翻訳の迅速化、翻訳コストの低減が叫ばれていましたし、機械翻訳システムは爆発的な需要を作りだしたワードプロセッサに続くものと期待されもしました。

日本語ワードプロセッサは、入力する人が出力するものを知っており、誤りは簡単に発見でき、その場で修正します。ところが機械翻訳の場合、余程その機械翻訳システムに熟知していないかぎり、入力に対する出力は予測できません。また出力が正しいのかを判断するにはかなりの言語に対する知識が必要です。さら

Learning from the past 2

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/>

に始末が悪いのは、出力に対する好みの問題まであります。翻訳はそれ自体が創作活動の側面を持っているわけですから、機械的翻訳に対して批判や失望が示されたのも当然かも知れません。

その後の機械翻訳システムの開発改良は、コンピュータの高速処理能力と一度覚えれば忘れることのない大容量の記憶能力の向上とあいまって、人間には不可能な翻訳の高速化と多分野にわたる大量の用語の利用を実現し、実用場で用いられるまでになってきました。けれどもまだ機械翻訳システムの本格的な利用は翻訳のエキスパートに限られています。機械翻訳システムは、その能力を引き出すには F1 マシンのようにエキスパートの力が必要と言えます。

したがって、マシンを限界まで早く走らせる F1 ドライバーと機械翻訳システムの限界の中でより質の高い翻訳を行なおうとする機械翻訳者とは同じような状況にあるのではないかと思います。

そこまで考えると、機械翻訳でのメカニック達はどこにいるんだろうかと思えます。多様性に富む翻訳の状況に機械翻訳システムを調整し、開発側と翻訳者のインターフェイスをとる人達です。システム開発側だろうか。もしそうだとすると個々の翻訳者の要求に答えることは難しい。それとも翻訳者側だろうか。そうだとすると、個々の翻訳者がシステム開発者と強いパイプを持つ必要があります。しかしそんなケースバイケースの要求に開発側が応じるとは考えられません。

機械翻訳をすでに確立された自動車の世界と重ね合わせて考えましたが、広がりを持った技術の世界ならばどの世界でも作る側と使う側の協力、時には闘争があり、その中で技術が育まれています。

機械翻訳で今まで欠けていたのは、F1 マシンでメカニック達が行なっているような機能かもしれません。システム開発者と翻訳者によって設立されたこの機械翻訳協会が、今後の活動の中で、その機能の一部を果たすことが期待されます。

近い将来、簡単な教育だけで誰もが使え、十分な質の翻訳が得られる、一般の乗用車のような機械翻訳シ

ステムが登場することでしょう。そのときでさえ、まさに翻訳エキスパートのための F1 マシンのような機械翻訳システムも存在していると思います。そこで翻訳のエキスパートとシステム開発者によって研究され開発された技術は、自動車が現在そうであるように、一般ユーザの機械翻訳システムに移転されていく。そんな近未来の機械翻訳システムの姿が見えてくる気がします。

【1992 年 11 月 JAMT ジャーナル No.4 より】

機械翻訳ビジネスは儲かるか

パネリスト：柏原武利氏、鎌田紳二氏、富永勲氏
司会：鳴海武史氏

司会：本日はお忙しいところ、お集まりいただきまして有難うございます。はじめに会社の概要を述べていただき、それから本日のテーマに沿ったお話をお聞かせ願えればと思います。まずは、鎌田さんからお願いします。

鎌田：私どもは富士通 100%出資の子会社で従業員が約 140 名、昨年度の売上げが 26 億円です。主な業務は翻訳、マニュアル執筆(テクニカルライティング)、それと編集(版下・レイアウト)です。

我々は機械翻訳センター課に属しており、翻訳部の中の一つの課です。機械翻訳関係の昨年度売上は、1 億円に達しない程度でした。機械翻訳については数年前、富士通が ATLAS という機械翻訳システムをはじめて世に出した時、その辞書の作成を富士通より委託されていました。かなり積極的に事業部から請け負う形でやってきました。しかしプログラム会社ではないので、ソフトウェアには手を出さずに、辞書作りを英語の専門家と日本語の専門家を派遣する形でやっていました。

大型コンピュータ M シリーズの端末を会社にひいて、川崎工場の大型汎用コンピュータと接続して使っていました。ここにきて大型コンピュータのプロジェクトは、ほぼ終わっています。

今、世の中はダウンサイジングといって、ワークステーションとかパソコンとかの小さなコンピュータに移りつつありまして、その流れに乗って ATLAS そのものもワークステーションに 2 年前にのりました。

それで、そのプロジェクトでは、我々はお客さんに売る時に、どういう風に使えばよいかということ、富士通から請け負って導入コンサルティングもしております。今もそれはつづいております。

昨年「ニフティサーブ」と「アスキーネット」というパソコン通信を利用して、この中に機械翻訳メニューをのせて、お客様から依頼があった文章を直ちに機械で翻訳して返すというサービスを行っています。

これは、やっと一年経ったんですけど、スタートしてすぐに収益がでてビジネスとして大成功でした。今もどんどん需要がのびています。これはしばらくいけるだろうと思います。ただ、こういうビジネスは経営資源としてソフトウェアを持っていることが必要です。

今回、儲かるかどうかというテーマに沿うと「かなり儲かるよ」と言えます。ただ、機械翻訳システムの開発は数十億かかると言われていますから、これができるところは数が限られることになります。

これ以外に、機械翻訳を使って儲ける仕事もしています。新聞のデータベースの翻訳をある会社から請け負って毎日毎日やっています。土、日は休みですが。朝受けて夕方には返すので、機械翻訳なしではこのビジネスは無理です。なぜかというと、普通の文章だと機械翻訳はほとんど使いものにならない(*プロの翻訳者は原文をみたときに英語を生成できる)のですが、新聞のデータベースになりますと社名の正式英文名称などが必要になります。数千社の正式名称を翻訳者が全部間違いなく覚えるのは無理です。機械でないとできないことしか、機械にやらせていません。

日本に会社が数千数百ありますし、世界中でも会社があります。ペンツといったって、ペンツの正式英文名称は何なんだということになります。そういうところは全部正しく翻訳できるし、その時できなかったら、お金をかけて調べて、一回入れてしまえば、後々使えますから。そういう使い方、ある会社からデータベースを請け負っているのです。

あとは富士通の社内ユースですけど、それも我々が絡まないといけません。今の機械翻訳システムはズブの素人が使うと、とんでもないことになってしまうんです。辞書登録の問題一つとっても、使いなれずに勝手に登録すると、とんでもないことになってしまいます。例えば、A という大文字のアルファベットにどこかの「アメリカ云々」という社名を登録しますと、ABC ときたら、アメリカ云々 BC と翻訳が出てしまいます。

ですから今、機械翻訳はそういう意味で、まだまだ、使い慣れて、問題を知らないと、全部駄目になることがあります。機械翻訳システムはまだまだ使い勝手が悪いし、ユーザインタフェースが問題だと言う人もいますが、システムそのものも問題なんです。様々な問題を抱えています。ですから非常にナイーブな使いにくいシステムを今後はどうにかして、うまく使いこなすことが我々の課題となってきます。恐らくそういう問題をクリアしたらもっと機械翻訳の売上げが伸びると思います。

司会：今のお話をうかがっていると、どちらかというと日英翻訳が多いということですか？新聞データベースも。

鎌田：そういう使い方はほとんど日英です。ニフティサーブだけが日英/英日は 1 対 1.5 の売上げになっています。これは一般のお客さんに機械翻訳システムを裸で使ってもらっています。それ以外の、翻訳者が絡むものはほとんど、今のところは日英になってしまいます。

柏原：ニフティーサーブでは不特定多数の人のドキュメントを対象としているようですが、膨大な辞書ができると思いますが、ニフティでの辞書登録はどうなっているのですか？

鎌田：ニフティではすべてお客様の責任でやっています。法律上(有線通信法)でお客様の文書等は一切見てはいけないことになっています。電話の盗聴と同じことです。文字化けしているなどとクレームや問い合わせなどがきてはじめて何が起きているかわかるわけです。それから辞書に関しては、名詞レベルですがユーザ預かり辞書という形でお客様に辞書登録をしてもらい、うちのシステムに登録しておいてキープするサービスも始めます。そうなりますと今、抱えている通信の問題は全て一応は解決ということになりますが、究極的にはやはり品質の問題ということになります。

柏原：その場合、各方面にコントリビュータをたくさん作ることにはなりますが、その辞書は、当然お客様の物だと思いますが、富士通ドキュメントサービスさんが預かり責任を持つんですね？

鎌田：そうです。お客様の資産なので勝手にお客様の許可なくコピーしたりできません。

柏原：将来のあるべき辞書データベース化の一つの大きな方向性を富士通ドキュメントサービスさんは模索されていると思いますね。

司会：それでは、情報センターの富永さんに業務概要の説明をお願いします。特に Mu 計画と現在の連続性についてもお話下さればと思います。

富永：私どもは科学技術庁傘下の特殊法人で、32 年に開設されました。情報処理の中核機関という位置付けで従来は主として海外の情報を国内に提供してしま

た。84,85 年頃アメリカ等から海外の情報ばかり取り入れないで、日本国内の情報を提供してくれという要望があり、いわゆる情断(情報断絶)というやつで、これ以上、海外の情報だけ持っていった国内の情報を出してくれなければ海外の情報も入れさせないと強く言われました。そういうことで何とか日本の情報を海外に提供しなければならないということになった訳です。ただ、山ほどある情報を一体どうやって提供するのかという点から考えると、簡単に言えば全部翻訳して出せばいいということになります。それでは人手や予算の問題で駄目だということになります。そこで科学技術庁が中心となって開発に着手したのが、例の Mu 計画です。

Mu システムというのは科学技術庁の中では、オフィシャルな名前ではなくて、いつの間にか、Mu という名前があって、今は Mu2 だと色々な所で書かれていますけれども、私どもは科学技術庁を含めて Mu という言葉は使っていませんが、ここでは便宜上、Mu としておきます。

経過としては 82 年～86 年(5 年間)に、長尾先生を中心として Mu システム(Mu1)の開発がされて、そこで、まあ、プロトタイプができあがったわけです。その後、それをどう実践の場に移していくかという段階で情報センターが Mu1 をベースにしてさらに実用化という目的で開発することになりました。それから 87 年～90 年春(4 年間)にかなりのお金をかけて実用システムなるものの第一歩が出来上がったという次第です。

昭和 32 年ぐらいから作り続けている科学技術速報誌(文献速報誌)のうち、国内で発表された論文だけを対象としまして 1984 年くらいから、少しずつ英文データベース化していました。JOIS オンラインシステム(海外につながっている)で、細々と情報を海外に提供していました。しかし手翻訳中心だったので、本当に少しのデータしか提供できませんでした。

お話にならない程度の情報量だったので、冒頭述べたような海外の非難があったわけです。これではしょうがないので、機械翻訳システムが何とか使えるとい

うことで、90年8月から手翻訳をなくしてしまいました。年間220万タイトルの論文抄録を英文データベースとして提供しているわけですが、その40%強を機械翻訳にかけています。

残るところは、最近日本語論文を書く時はきちんと英文抄録や英文タイトルをつけなさいという指導がなされているので、58.9%ぐらいは少なくとも英語が著者によって書かれているという状況です。

辞書のお話をしますと、開発費の半分ぐらいは辞書に当てました。現在は54万語の辞書をもっています。特に官公庁関係の資料を扱うことが多いため官公庁の部局まで含めた英語名称が重要になり、多くの機関名が辞書にはいっています。勿論民間会社も非常に重要です。約91,000タイトルの抄録の英文化を、前・後編集付きで月に7,600タイトルのペースで行なっています。それを前・後編集をしながらやっていくわけですが、8人の校閲者(ネイティブ2人)が最終のフィルターとして控えています。

前・後編集は翻訳会社さんへ(5社)お願いしています。5社に対してはそれぞれの専門分野のところへ振り分けて行なっています。一番最後の後編集段階で残る専門用語の英訳の不備は専門家が一番最後に控えていますから、ここで修正します。会社名に関しても全部フォローすることになっていますので、そこからどんどん言葉が抽出されます。

科学技術文献の抄録という特性から、膨大な情報を300字程度に納めようとコンパクトにしますので、ひどいのになると40~50字全部漢字が並ぶようなこともあり、従って前編集はとても苦労します。そういうものを機械翻訳にかけると、ツルツル切れる所で切ってしまうと、いかにも英語ですみたいな形で出てくるので、ツルツル英語と言っていますが、それでは役に立ちません。

従って、そういうものを全部前編集の形で補いながら、また、長い漢字列がでてきても対応できるような形で辞書作りをやってきました。その結果、医学を含めて現在54万語までできました。

手翻訳は90年8月でやめたとはいいましたが、その前は手翻訳と機械翻訳を併用した試験期間が約半年~1年弱ありました。

今日のテーマに沿った話をしますと、何が儲かるかという点では、私どもは一般の翻訳業務もやっていますが、それにはまだ機械翻訳は使っていません。今、申したデータベース作りだけなんですけれど、人手翻訳に比べて翻訳発注費用が3割程度ダウンできました。それは私どもから見たら儲かったという話になります。それとともに計算機使用時間が減りました。機械翻訳をやるから、人手翻訳の時代と比べて、前・後編集などもあるのもっとトータルタイムは増えるだろうと言われてましたが、計算機の作業は下がっています。一件あたり30分くらいのが25分ですんでいます。それからサービスタイムやターンアラウンドタイムも小さくなりました。工夫すればもっと小さくなると思います。

現在は大型コンピュータの上に登載して、自社内のデータベースだけをやっていますが、将来は一般の論文も受け付けて、素翻訳や後編集付きの翻訳サービスもしたいと思っています。あるいは今は完全なバッチ処理でやっていますが、日本語抄録を作っている段階でTSSで翻訳ができたりとか、全国ネットワークを張って、一般論文の翻訳のお手伝いするというのを、近い将来に実現する方向で計画中です。

司会：どうもありがとうございました。それでは柏原さんをお願いします。会社の概要とか、機械翻訳を導入したキッカケとか、現在どういう利用のしかたをしているのかとか含めた形でお話していただければと思います。

柏原：私どもは従来は手翻訳専業でして、昭和52年からスタートしまして、法人になりましたのが、昭和60年です。今のよう、本格的に機械翻訳に取り組むようになったのは、正確には昭和63年なんです。ただ、60年の法人設立の目的には機械翻訳をベースにした業務展開を将来、必ず担いたいというプランがあっ

たものですから、その前提として法人を設立したということがございました。機械翻訳導入のキッカケについては、後ほど申し上げます。

現在は、機械翻訳と翻訳の後処理として DTP による版下作成、またいくつかの電子編集システムを使って、EP といいますか、新聞、雑誌を発行できるところまでの陣容を整えつつあります。それからこれも機械翻訳の後処理を追求する中で、でてきたのですが、DTP のスタッフの中で、マッキントッシュの特性を生かした、グラフィックアプリケーションの開発とかも始めています。これは最近でいう、バーチャルリアリティという言葉を使わせていただいています、実はこれも機械翻訳というペースがあって技術情報を得てできたものです。

現在は大きく分けて機械翻訳、DTP、機械翻訳ではできない、従来の手翻訳を行なっています。比率としては、現在は、機械翻訳と機械翻訳に全く関わらない翻訳が半々ぐらいですが、2 年以内には機械翻訳の比率を 8 割まで高めたいと考えています。それは、手翻訳の仕事を減らすというよりも、機械翻訳の処理を増やして、相対的に比率を高めたいということです。この他に通訳とかイベントに関わる業務も行なっています。売上げは 4 億円弱位です。

それで機械翻訳を導入したキッカケは翻訳業者としてやってきて、一番泣かされてきたというか、これはいずれも同じかも知れませんが、やはり翻訳のクオリティそのもの、それを支える翻訳者の不足、マネジメントが非常に難しいということです。で、一度いい翻訳結果を出しますと、お客さんは当然次も同じことを期待して発注されますが、同じ品質のものを常に提供し続けることは至難の技でございます。

とりわけ私も翻訳者のはしくれとして一番残念に思い、かつ疑問に思ってきたのは、データを共有するとか、蓄積したノウハウを分かち合うとか、それを組み立てて新たなものを生み出すという感覚をもった翻訳者は本当に稀だということです。全くスタンドアローンで仕事をしています。ですからその人に何かがあればそれまで蓄積

されたノウハウは消えてしまいますし、その時点でお客さんの信頼も失うということになります。

なかには例外的な人もいますが、大体、一定レベルに達した人は蓄積した資産で食っていこうとする人が多いのです。ということは、技術革新がどんどん進んでいくと、対応できなくなるということになります。現実には、コンピュータのマニュアルなどは新しいバージョンのソフトウェアができていくのに、前のバージョンの翻訳がまだ終わっていないことが珍しくないのです。

こういうことに翻訳者はほとんど危機感をもっていないどころか、いかに自分たちの存在が有難いかわかるだろうという対応しかしませんので、実際のところは改善を求めるといっても難しいところです。お客さんの利益をどう勝ちとって、いかに自分たちの事業を展開していくかという点では、それは全てのスタッフの利益にかかわっていることなのですが、それが根底から崩れさります。

まだ翻訳業界そのもののステータスが非常に低いということに、色々な場面でぶつかります。だいたい金融業界では、我々は産業分類にはいっていなかったですね。最近やっとサービス業の中にできるようになりましたが。何か売れない文学作品を書いている作家と同様で本当に水商売と言われてました。

今の翻訳者の質をいかに高め、維持するかという問題が一つ。それと同時に海外での日本の技術に対する情報の公開とか、日本は非常にクローズされた世界だと言われているんですが、これは翻訳の問題がネックとなっているだけなんですね。私も、JETRO とか通産省のお仕事をずっとさせてもらってそういう場面にたくさんぶつかりました。

これはもうお話してもさしつかえないと思いますが、JETRO で昭和 61~63 年にかけて、技術的なバックグラウンドを持った企業の海外での進出先アンケートをやったことがあるんです。正式には『海外向け技術進出アンケート調査』ということで、アットランダムに選んだ日本の技術系企業からアンケートをとって、どういうプロダクトを持って、どこの国に進出したいか、

どういう目的で海外に進出したいのか、という目的の調査でした。

これはマシーンが、N5200 というのを使うことが条件で、他の業者さんは全て降りて、うちだけが残ったんですが、一つの画面の中にその会社のプロフィールを掲載し、当然その進出先、例えばベルギーとかオランダに進出したいのはどの企業か、とかのアウトプットも可能なので、技術提携とか、海外での工場進出とか目的別区分けもできるということで、非常に好評だったようですね。アメリカに進出したい企業が一番多いのですが、アメリカ以外の国からも、このデータは貴重だと非常に喜ばれたようです。

ところが喜ばれたことが新たな問題を作ってしまったんです。つまり、フォローアップの体制ができていなかったし、予算もそれしかありませんので、当然、海外からもっとディーテールを知りたいとか、この会社とコンタクトとを取りたいとか言われても、そこまで面倒を見る体制がなかったので、イギリスから「日本の体制はけしからん」と正式なクレームがついたり、トラブルがあって「フォローができないのならやめるべきだ」との上の判断で、2 年間でそのプロジェクトは終わってしまったんですね。ただ、N5200 から PC98 におとしてということだけはされましたので、海外での利用という点では多少はつながったのではないかと思います。

これはフォローアップとかユーザインターフェースの問題ですが、私はマッキントッシュとか IBM とか、海外のユーザが使いやすいマシーンでのデータ提供をすべきだと思っていました。その時いくつかの海外の企業とか政府の方々とやりとりを通じて、いかに日本の技術情報に対するニーズが高いのかを痛感しまして、それでこの JETRO さんでの仕事が私にとって、機械翻訳に踏み切らせる最終的な決断になった訳です。

その前に、実は、もうひとつありまして、私どもはあるコンピュータメーカーさんのマニュアル翻訳をずっとやっておりまして、4 年程前に合併された会社なんですが、そこの方々と日本の新聞の記事情報を翻

訳して、それをその社のマシーンで、つまりパソコンを持っているお客様にネットワークで提供する(新聞発売と同時に)という話がありました。

まあ、元々私も一部提案しておりましたが、ジョイントでやろうということになったわけです。日本の新聞記事を翻訳して出すというのは量としては多いと思うんですよ。で、しかもあれほど面倒な事もないのですよ。固有名詞とか色々な問題があります。新聞というのは限られた時間で書いていましすし、翻訳しやすい文章はほとんどありません。

翻訳される記事というのは、いくつかの所で必要とされるケースが多いですね。コンピュータとか、通信とか、半導体とかはそうです。いくつかの分野に限定して、ヘッドラインだけではだめでしょうが、少なくともアブストラクトには必ずお客がいると思ひまして、当時、資本金 180 万で法人成りしたばかりでしたが、500 万円掛けました。うちとしては死ぬような金を掛けました。日本の外資系企業と海外のお客様を相手に 1 ヶ月ためしてみました。非常に好評だったんですが、ネックはやはり翻訳でした。

翻訳者も最初は元気のいい者だけ選んだんです。朝 9 時に出勤した時はその時、手にする記事のアブストラクトがファックスで入っていることにしようとしたんです。ところが、第一版が出たとき、記事をピックアップして始めたんですが、10 日ぐらいたらだんだん息ぎれがしてしまったんです。考えた以上に翻訳のコストがかかったんです。それに、ジョイント先の事情もあって、1 ヶ月で切れてしまいました。

ただその時、本格的にそういう事をやるところがあつていいんじゃないかと思いました。日本にある外資系企業のマネージャーの任務の一つは日本における情報収集です。本国にタイムリーに最も肝心な情報を送る事です。そういうことによってその人の能力が問われるわけです。そういう方々に対するサービスだけでも大変なニーズがあるなと思っています。ただ人手の翻訳者を前提とする限り、これは絶対にできない話だと思っています。

その後の JETRO さんの件とかがあって、とにかく機械翻訳を利用しない限り、いずれは廃業に追い込まれるだろうと思いました。その時点では、既にブラビスさんのマイクロパックなどが出て、出た当座はこれで翻訳者が廃業になるとかの言われ方がされましたが、全然そういう甘いものではなくて、業界では、優秀な翻訳者ほど敵対的な意識を持ちましたし、今でもまだ、そうです。「こんなことをして、俺の仕事の邪魔をするな」という言い方をする人が多いですね。

私どものサービスはコンピュータ支援翻訳サービス、つまり CATS (Computer Aided Translation Service) といいまして、勝手に付けたんですが。そうではなくて、人間の、優秀な翻訳者の優秀な武器として使えばいいわけですよ。感覚的には昔の鉄砲が生まれる前の、刀とか槍の名人・達人に鉄砲を持たせると、「こんな卑怯なものが使えるか」という感覚があったと思うんですが、それに近いと思います。ですから、敵対的な存在として認定するのではなく、自分の武器として使いこなす、そういうインターフェースをもてる人材を育てない限り、機械翻訳の世界は広がらないと本当に思いました。

(つづく)

イベント報告

AAMT 2022, Tokyo

～機械翻訳最前線～

出内 将夫

情報通信研究機構

1. はじめに

2022 年 12 月 1 日に開催された AAMT 2022, Tokyo のイベント報告をします。2022 年はオンライン開催に加え、3 年ぶりの会場開催もあり、ハイブリッド形式での開催となりました。招待講演 2 件、公募セッション 7 件の発表のほか、MT Summit 2023 紹介、パネルディスカッション、企業展示コーナー、ランチョンセミナーもあり、非常に盛況でした。講演/発表資料は <https://www.aamt.info/aamttokyo2022/handout-20221201/> からアクセスできますので、ぜひこれらの資料をご参照ください。

2. 招待講演

招待講演として、2 件の講演に加え MT Summit 2023 の紹介がありました。

◎日本語を含む言語対における制約付き翻訳タスクの近況：

阿部 香央莉氏／東北大学大学院 情報科学研究科 博士後期課程

◎NICT における自動『同時通訳技術』のご紹介：

内山 将夫氏／情報通信研究機構 先進的音声翻訳研究開発推進センター 上席研究員

◎Back to Asia: MT Summit 2023 Macau：

田中 英輝氏、内山 将夫氏、山田 優氏、森口 功造氏／AAMT 理事

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。



招待講演時の会場の様子

◎日本語を含む言語対における制約付き翻訳タスクの近況：阿部 香央莉氏

専門用語を含む文書を翻訳する際、特定の用語を一貫して翻訳したい、というニーズがある。現在、一部の機械翻訳（MT）サービスには用語集機能があり、指定した語の訳語を指定できる。本講演では、訳語を指定する MT の仕組みや、それを評価する制約付き翻訳共有タスクについて説明する。

制約付き翻訳共有タスクは、与えられた制約語彙に従って文を翻訳するタスクである。ASPEC 科学論文ドメインデータを元に再設計した。共有タスク設立のため、ASPEC 対訳文から制約語彙リストを作成し、制約語彙の出力を考慮した自動評価指標を提案した。2021 年に日英・英日、2022 年に日中・中日を対象に制約語彙のアノテーション付きコーパスを作成し、共有タスクに参加した MT システムを評価した。

上述のタスクに対応した MT システムがいくつか提案されている。ニューラル機械翻訳（NMT）以前の統計的機械翻訳（SMT）では、制約語彙を入れることが容易であったが、NMT では容易に制約語彙を入れら

AAMT 2022, Tokyo --- Forefront of machine translation ---

Masao Ideuchi

NICT

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/>

れない。そこで、NMT の翻訳結果出力時に制約を追加する試みが行われている。追加する制約の種類は「できる限り制約語彙を含める」という緩い制約と「必ず制約語彙を含める」という厳しい制約がある。一方の制約もしくは双方の制約の組み合わせや、緩い制約と後編集を組み合わせるなど、様々な MT システムが提案されている。

複数の MT システムの評価結果から、考案した自動評価指標が人手評価と相関が高いことが示された。しかし制約語彙以外の部分で、欠落や誤訳などにより人手評価との差が大きい事例もあり、改善点も見られた。

◎NICT における自動『同時通訳技術』のご紹介：内山 将夫氏

NICT で進めている、総務省委託研究開発「多言語翻訳技術の高度化に関する研究開発」について説明する。「グローバルコミュニケーション計画 2025」で提唱している 2025 年の大阪・関西万博をショーケースとした同時通訳の実現を目指している。複数の企業と連携し、様々な利用シーンでの社会実装を進めている。

同時通訳のための技術について紹介する。同時通訳では、話者が発話を停止しないので、入力分割が必要となる。区切りの無い音声認識結果のテキストを、機械学習によりチャンクという文以下の短い単位に区切り、その単位で翻訳することで、低遅延で翻訳できる。短い単位に分割すると翻訳精度が悪化するので、付加情報として文脈やタグを与え、翻訳精度の悪化を抑える技術を開発している。

チャンクに区切る技術について詳しく説明する。チャンクへの分割は、分割点より N 語先まで読み込み、分割点の位置を予測するモデルを構築して実現する。訓練データとしては、人手で文やチャンクに分割したコーパスを用いる。複数の会話コーパスを用いて、F1 と latency を組み合わせたスコアを用いて評価している。評価指標では、正解であることと遅延を少なくすることを重視している。最初に文分割モデルを作成し、そのモデルをチャンク用にファインチューニングして

チャンク分割用モデルを作成している。チャンク翻訳は遅延が少ないが、文翻訳より精度が劣るため、文末で同期し、訳し直し可能としている。

また同時通訳用のコーパスとして、通訳経験者が原文をチャンクに分割し、チャンクごとに順送りの対訳としたコーパスを作成している。原文内容の 60～80% 程度の一次訳、原文同等内容の二次訳の双方を作成している。このコーパスを用いて、チャンク単位の翻訳を評価したところ、人手による分割と遜色ない翻訳精度を実現しつつ、低遅延を達成できた。チャンク単位の翻訳は、文単位の翻訳に比べて、日本語で平均約 5 単語、約 40% の短縮ができた。また、短文化エンジンをみんなの自動翻訳@TexTra から公開した。

今後、2025 年の大阪・関西万博に向けて、開発した技術の実証を進めていく予定である。

◎Back to Asia: MT Summit 2023 Macau : 田中 英輝氏、内山 将夫氏、山田 優氏、森口 功造氏

MT Summit 2023 を紹介する。MT Summit は、IAMT（国際機械翻訳協会）が隔年で開催している国際会議で、AAMT の他、アメリカの AMTA とヨーロッパの EAMT が持ち回りで開催している会議である。MT Summit は AAMT 初代会長の長尾眞先生が創設し、幅広い層の参加者が特徴の会議である。MT Summit 2023 は AAMT が担当し、2023 年 9 月 4 日から 9 月 8 日まで、マカオの現地開催とオンライン開催のハイブリッドで開催する予定である。詳細は 12 月中旬に AAMT ホームページで公開する。

MT Summit 2023 のリサーチ・トラックでは、招待講演、オーラル発表、ポスター発表を予定している。招待講演は、1 時間枠の招待講演 2 件と、講演者 3 名による 2 時間枠の招待講演セッション 1 件を予定している。研究発表で募集するテーマとして、MT に関する各種タスクや、MT の応用に関わる広範なテーマを扱う予定である。

MT Summit 2023 のユーザー・プロバイダトラックでも、招待講演、オーラル発表、ポスター発表を予定

している。ユーザー・プロバイダトラックでは、翻訳研究の要素も考慮に入れ、MT が社会に与える影響なども扱う予定である。研究発表で募集するテーマは、今までと同様に翻訳会社による発表を想定したテーマのほか、社会一般で MT が使われることを想定したテーマを取り扱う予定である。

MT Summit 2023 では、スポンサーシップを募集している。4 ランクに分けた募集となっており、ランクによって学生招待権や求職中の会議参加者情報閲覧が可能になる。詳細は MT Summit 2023 の公式サイトから公開予定である。

3. 公募セッション

公募セッション発表は次の 7 件でした。

◎コネクテッドカー業界における表現の機械翻訳の対応状況について：

太田 圭佑氏／株式会社ファイブアイズ・イングリッシュ

◎機械翻訳活用で顧客の期待に応える：

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

◎翻訳者からみた機械翻訳の現況と効果的な利用方法：

吉川 潔氏／特許翻訳者

◎翻訳チューリングテストの結果に見る、“機械翻訳らしさ”の正体とその軽減方法、および人間翻訳に期待されている価値について：

堂本 秋次氏／翻訳家、英語学習アドバイザー、講演家、作家、YouTuber

◎meta 翻訳 ～個人開発から最高精度を目指す～：

藤井 隆太郎氏／株式会社 ClassIII

◎EU から世界へ、日本から世界へ：

岩永 時春氏／pangeanic

◎クラウドベースの翻訳管理システムで機械翻訳を活用する：

目次 由美子氏／XTM International Ltd.

紙幅の関係から一部の発表について次にまとめます。

◎コネクテッドカー業界における表現の機械翻訳の対応状況について：太田 圭佑氏

コネクテッドカーにおけるテレマティクス技術に関する翻訳事例をもとに「新ビジネスが日本でどの程度浸透したかを計測する方法」について考える。

世界最先端のビジネスは海外で発祥し、数年遅れで日本に上陸することが多い。当初は、ほとんどの文書が英語で書かれており、新しい単語の日本語訳が定まっておらず、ある程度日本市場に浸透した段階で訳語が定まる。訳語が定まれば、MT が訓練データとして訳語を学習できるため、MT が新ビジネスの英文を正しく訳せるかどうかで、日本での新ビジネス浸透度合いが測れる、という仮説を立てた。

コネクテッドカーに関する MT の翻訳結果を 2020 年と 2022 年で比較した。“incident”の訳語が「出来事」から「インシデント」に、“connected car”の訳語が「接続された車」から「コネクテッドカー」に変化しており、仮説が成り立っていることが確認できた。

このことから、日本市場浸透度のような数値化が、MT 結果の滑らか度を元に算出可能と思われる。翻訳者の主観に左右されない指標としては、ポストエディット (PE) の作業時間や日本語訳の Google 検索ヒット数、用語集の Google 検索ヒット数などの経時変化を利用できる可能性がある。

◎翻訳者からみた機械翻訳の現況と効果的な利用方法：吉川 潔氏

MT は、翻訳業のライバルではなく、サポータであることを伝えたい。NMT は近年、長文に追従可能となり、特許翻訳でも参考にできるようになった。新語の訳語を調べたり、訳しづらい語の適訳を調べたりする際に重宝する。しかし、訳抜けや構文エラー、語の訳し分け、数量表現など、問題も残っている。これらは翻訳分野に詳しい人によるチェックが必要である。

NMT が発表されてから 6 年以上となり、MT の出力が大きく改善したことは事実である。翻訳者が MT を実務で利用し、気づいたことを開発側に提示し、開

発側が改善してフィードバックすることで、良い関係が築けると考える。

◎meta 翻訳 ～個人開発から最高精度を目指す～：

藤井 隆太郎氏

meta 翻訳の開発について説明する。meta 翻訳は、発表者が個人で開発を始め、最高レベルの精度を達成した英日英翻訳サービスである。英日翻訳では世界 2 位、日英翻訳では世界 1 位相当の精度を達成した(藤井氏調査による)。

meta 翻訳は、「日本語くらい日本のサービスで」という思いから開発を開始した。また一般人が有名翻訳サービスより優れた翻訳機を開発できれば、個人開発にチャレンジする人を増やせる、という狙いもあった。

精度が良い理由として、OSS の恩恵を受けていること、多くの論文から必要な知識を網羅したこと、他と異なる尖った技術を採用したこと、が挙げられる。尖った技術は、同じ AI 同士が戦い続けて強くなる特徴を持つ将棋 AI から着想を得た。将棋 AI は個人でも成功しており、計算量が現実的、勝手に成長するという特徴を持つ。後者の特徴は meta 翻訳にも取り入れた。

meta 翻訳は、これまで有料化していなかったが、これを機に有料プランや API 公開などを検討する。今後も社会に良い影響を与えていこうと考えている。

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「AAMT ガイドラインに基づく ISO18587 供給者自己適合宣言のススメ」をテーマに発表と議論が行われました。冒頭にモデレーターの森口功造氏から、PE 案件で ISO18587 準拠ニーズはあるものの、国内に認証組織が無く ISO18587 認証取得が難しいことから、AAMT が自己適合宣言の支援を行うことにしたと説明がありました。

パネリストとして 3 名から発表がありました。浜田氏の発表後に、モデレーターの森口功造氏から AAMT が公開したガイドライン、様式、チェックリストの説

明があり、続けて事例の発表がありました。

◎翻訳サービスの適合性評価に関わって：

浜田 良和氏／日本規格協会ソリューションズ株式会社 審査登録事業部

◎供給者自己適合宣言をしてみてわかったこと：

菊地 大悟氏／株式会社十印 翻訳事業部

◎AAMT ガイドライン作成 WG に参加して：

小野 眞帆氏／株式会社アスカコーポレーション 事業戦略室

◎翻訳サービスの適合性評価に関わって：浜田 良和氏

日本規格協会は、ISO17100 の審査業務を行ってきた。AAMT が行う ISO18587 の供給者適合宣言支援について、日本規格協会も協力したので、翻訳サービスの適合性評価について説明する。

適合性評価には、第三者による認証、関係者による二者監査、供給者自身による供給者適合宣言がある。供給者適合宣言を行う際は、JIS Q 17050-1 で定義された供給者適合宣言の手続きに従う必要がある。

供給者適合宣言では、対象となるサービスを明確化する必要がある。そして基準となる規格や要求事項の明確化、責任者の明確化が必要となる。さらに力量・資格を持つ実務者の管理・識別や業務プロセスを確立し、必要な業務文書の整備、記録を保持することが必要となる。さらに、上述の事項が保たれていること確認する内部監査を定期的に行う必要がある。

ISO18587 を ISO17100 と比較すると、PE 特有の項目もあるが、ISO17100 と共通する要素も多い。ポイントを表にまとめたので、必要に応じてそれぞれ参照されたい。

◎供給者自己適合宣言をしてみてわかったこと：菊地 大悟氏

実際に ISO18587 の供給者自己適合宣言を行った事例について説明する。

供給者自己適合宣言を行った動機は、お客様とのサービスレベル共通認識、提供するサービス品質の証

明、入札や選定時の得点源、といった観点で有利と考えたからである。弊社は ISO17100 認証を取得済であったため、規格の審査や内部監査の面で、必要な知識や人材がある程度揃っていた。これらは、ISO18587 の供給者自己適合宣言に有利に働いた。また、社内のオペレーションとお客様からの仕様を組み合わせると、既に ISO18587 に準拠していると言えるレベルのプロジェクトが存在したこともあり、短期間での ISO18587 の供給者自己適合宣言が可能であった。

今後の課題としては、社内教育や継続的な内部監査員の確保・要請、お客様における規格認知度向上があると考えている。

◎AAMT ガイドライン作成 WG に参加して：小野 眞帆氏

弊社では医薬専門の翻訳サービスを行っており、主なお客様は製薬会社や大学、研究機関である。翻訳対象は、新薬申請書類や論文など多岐に渡る。特に新薬申請書類では高品質を保ちつつ納期短縮のニーズが高い。そのため、2017 年から医薬専門の AI 翻訳エンジン開発に取り組み始め、MT エンジンの提供と、MT と PE を組み合わせた翻訳サービスを提供している。

2022 年に ISO18587 の供給者自己適合宣言の審査を行った。発表者は ISO18587 の審査対象部門とは別の部門に所属しており、内部監査担当として審査に関わった。内部監査では、制作プロセス、ポストエディタの力量及び資格、フル PE の要求事項について審査を行った。その結果、ポストエディタの力量に軽微な不適合があり、12 月中の再内部監査と供給者自己適合宣言を目指すこととなった。

◎ディスカッション

浜田氏：

会社の方針にも依るが、ISO17100 では審査先社内です事前に適合チェックを行った案件を審査対象案件とすることも多い。アスカコーポレーション様の ISO18587 審査時には、案件選定は行わなかったのか。

小野氏：

事前の情報共有と、手順書の記述が不十分だったことが原因と考えている。手順書は是正し、経営層含めて認識を合わせた。

森口氏：

ガイドラインはフル PE が対象だが、お客様からフル PE やライト PE の指定を受けたことはあるか。

菊地氏：

最初「ライト PE で良い」という依頼で来た案件が、最終的にフル PE 以上の要求になってしまうこともあった。ISO18587 という基準があるとありがたい。

小野氏：

案件ごとにお客様と進め方を決めるので、案件がフル PE かライト PE かの区別はしていない。

森口氏：

PE の案件で、MT の品質が問題となることはあるか。

菊地氏：

PE を行う際に、MT の品質をチェックしている。MT の品質が悪く PE を断った事例は、確認できておらず、存在したとしても多くはないと思われる。

小野氏：

同様の状況である。MT の品質が悪い場合は、MT を使わずに翻訳を行うので、問題にはならない。

森口氏：

PE 作業のレベルやプロセスなど、PE 作業者に説明する資料は用意しているか。

菊地氏：

プロジェクトごとに PE ガイドラインを策定する。自社基準やお客様指定の基準など、状況に応じて策定する。

小野氏：

基本的な PE ガイドラインは社内にあるが、お客様と相談の上、案件ごとに定め、PE 作業者に周知する。

森口氏：

浜田氏に伺いたい点として、ガイドラインではスコープとして言語対を指定した方が良く、としているが、具体例までは述べていない。どう考えればよいか。

浜田氏：

適合性評価のスコープは、お客様に範囲を明確にする上でも非常に大事である。原則として、ISO18587に準拠できる範囲を明確にするために、言語対を指定した方が良い。ただし、会社の方針等により、どの言語対でもISO18587準拠で受ける体制ができていれば、限定は不要である。

森口氏：

ガイドラインでは「自社が提供するサービスすべてが規格に適合しているわけではない」と宣言することを推奨している。例えば、ライトPEの案件も受ける場合、ISO18587に準拠しない案件に該当する。

オンライン視聴者：

PEの依頼を受ける際、料金を安価にしてほしい、という要望が来ることがある。どのような方針で、クライアントに対応しているか。

菊地氏：

PE案件では料金割引の依頼も同時に来ることが多い。サンプルを用いて作業レベルをお客様と合意することが多い。

小野氏：

お客様に予算が無い場合、MT結果のサンプルをどこまでPEで編集するか説明し、料金を抑えることがある。今回のISO18587は、フルPEが対象となるので、そういった調整は宣言の範囲外になると考える。

オンライン視聴者：

PEに適した人材と、翻訳に適した人材は異なりそうですが、PE作業を依頼する際、適した作業実施者を限定するために、適切な文言はありますか。

浜田氏：

ISO17100でも同様だが、適格な作業者については、ISO18587では規格に定められた要求事項の他、自社の運用ルールに従っていれば良い。

菊地氏：

PEに特化した作業者というのは募集していない。翻訳者として募集した方にPE作業も依頼している。

小野氏：

PE作業者の募集はしているが、現時点でPEのみの応募は無く、翻訳と作業者は共通している。MTが流暢になった結果、MTの誤りに気づきにくい作業者もいる。作業依頼時にその点を考慮することはある。

森口氏：

アウトソースする言語対にも、供給者自己適合宣言ができるか、という観点で、外部の委託先をこのガイドラインで評価する用途は考えているか。

菊地氏：

そのような使い方もできると気づきを得た。アウトソースする業者を選定する際は、ISO18587を取得済みの業者を選定するので、ガイドラインを使って委託先を評価することは考えていなかった。

小野氏：

同じ状況である。社内で使い方を追加検討したい。

浜田氏：

ISO17100では委託元の責任として、作業を委託する際は、委託元が委託先に対してISO17100準拠を求め、ISO17100準拠を確認する必要がある。ISO18587でも同様にすべきと考えられる。

5. おわりに

3年ぶりに会場での開催があり、参加者間での活発な議論が行われた年次大会となりました。コロナ禍を機に始まったオンライン開催も継続されました。オンライン化はMTにとって追い風になる、と言われていきます。今後の年次大会にもご期待ください。

イベント報告

第6回自動翻訳シンポジウム 参加報告

東山 翔平

国立研究開発法人情報通信研究機構

1. はじめに

2023年2月16日に、「第6回自動翻訳シンポジウム～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～」が品川ザ・グランドホールにて現地開催され、389名が参加した(写真)。また、過去最多の24企業・団体により、多言語翻訳関連技術・サービスを中心とした展示が行われ、出展者と参加者との活発な交流が行われた。2019年以降の実現となった現地開催・リアル展示の盛況な様子を通じて、人々の自動翻訳への関心の高さが窺え、「アフターコロナ」に向けた自動翻訳含む様々な経済活動の活性化への期待も感じられるようであった。本稿では、講演2件、パネリスト発表3件、パネルディスカッションの内容を報告する。



2. 基調講演「インバウンド観光の最近の動向・今後の見通し、多言語翻訳技術に期待すること」

日本政府観光局（JNTO）理事・中山理映子氏による基調講演の内容を紹介する。

JNTO は、訪日外国人旅行者の来訪促進のため、1964年に政府が設置した独立行政法人である。日々の業務として、訪日旅行市場に関するマーケティングや、海外現地旅行会社等の訪日ツアー企画・販売促進、訪日者向けプロモーションなどを行っている。

インバウンド観光は、日本が今後直面する人口減少に伴う課題解決の観点からも重要である。2052年に国内人口が1億人程度まで減少するとの見通しがあり、これにより、国内消費額が30兆円以上消滅する計算になる。日本の定住人口1人当たりの年間消費額約130万円（2019年）は、訪日外国人観光客約8人分の旅行消費額に相当することから、インバウンド観光は、今後の日本の人口減少による消費額減少を補う効果が期待できる。訪日外客数や1人当たり消費額を更に増やすことができれば、更に効果的である。

日本のインバウンド観光におけるこれまでの過程として、2003年のVisit Japan キャンペーン開始時では訪日外客数521万人であったのに対し、2019年には3,188万人に増加した。新型コロナウイルス感染症の影響により2021年は25万人にまで減少し、2022年は383万人となっている。10月に外国人観光客受け入れを本格再開した後、11月約93万人、12月約137万人、2023年1月には約149万人と増加した。なお、国連世界観光機関（UNWTO）の専門家によると、国際観光が2019年水準へ回復する時期は2024年以降とする意見が過半数の約63%を占めている。

日本での観光に対し、世界からの期待の高まりも感じられる。2021年の世界経済フォーラム（WEF）で発表された旅行・観光競争力ランキングでは、ホテル・交通網などの観光客向けインフラ、観光資源の豊富さなどを理由に、日本が初めて1位を獲得した。米国大

手旅行雑誌 Condé Nast Traveler では、2021 年の「世界で最も魅力的な大都市ランキング」の読者投票で、東京、大阪、京都が上位 3 位に選出された。その他、Travel+Leisure、Lonely Planet、New York Times のランキングでも京都、福岡、盛岡といった日本の地域が選ばれている。

一方で、解決すべき課題もあり、3 点挙げる。一つ目は地方への誘客である。延べ宿泊者数の上位 5 都道府県は東京都、大阪府、北海道、京都府、沖縄県であり、大都市や一部の地域に偏っている。二つ目はインバウンド消費額である。2015 年の爆買いブームで 1 人当たり消費額が約 17 万円に増えたものの、その後は 15 万円程度で伸び悩んでいる。三つ目は言語障壁である。観光庁の調査結果では、「旅行中に困ったこと」として、多言語表示の少なさやわかりにくさ、施設スタッフとのコミュニケーションのとりにくさなどが挙げられている。

最後に、インバウンド観光の今後の方向性について述べる。2022 年 10 月 11 日、岸田首相は、インバウンド消費額 5 兆円超達成を速やかに目指すことや、2025 年の大阪・関西万博開催に向けて新たな観光立国推進基本計画を同年度末に策定する要望を述べ、また、同日、和田観光庁長官は、今後の観光の方向性として、「消費額の拡大」、「地方誘客促進」、「持続可能な観光」の三つのキーワードに留意するよう述べた。こうした状況の下、JNTO は、高付加価値旅行¹、サステナブルツーリズム、アドベンチャートラベル²の三つの重点取り組みを推進しており、今後のインバウンド観光が「モノ消費」から「コト消費」へ、つまり体験型観光へシフトすると捉えている。体験型の観光においては、これまで以上にコミュニケーションが重要となるが、各地方で多言語のガイド人材を育成・確保することは容易ではない。また、JNTO のプロモーション対象市場も、最近、中東地域やメキシコが加わり、更に、今後北欧も加わる予定であり、更なる多言語化が必要であ

る。そのため、多言語翻訳技術に期待している、と中山氏は結んだ。

3. 講演「同時通訳を支える翻訳バンク」

国立研究開発法人情報通信研究機構（NICT）フェロー・隅田英一郎氏の講演の内容を紹介する。

AI による音声認識、音声合成、翻訳がいずれも高精度になり、これらを組み合わせた AI 同時通訳が実現可能になってきている。現在、総務省のグローバルコミュニケーション計画 2025 の下、官民で委託を受けて同時通訳技術の研究開発を進めている。これまでの進捗として、2020 年に明示的な区切りがないテキストの文分割が可能となり、2021 年には文より短い単位であるチャンク分割が可能となり、2022 年にはこれらの分割技術を英語・中国語・韓国語・ベトナム語に多言語化し、技術移転も実施した。2023 年には商用化、2024 年には技術の完成を目指し、2025 年の大阪・関西万博の時点では広く使われる状態になっていると考えている。デモ動画で示したように、現在のプロトタイプシステムでは、日本語の合成音声及早口である点が気になる方もいるかもしれないが、発話中の重要な点に絞って翻訳する技術も開発している。

人間による同時通訳でも、日英は、文法が大きく異なることから特に難しいとされている。たとえば日本語では文末に述語や否定が来ることが多い。通訳者は、原発話者がまだ言っていないことを予測したり、一度訳した後に訳を言い直したりすることで対処している。我々は違う方法での対処を考えており、即時にチャンク単位で翻訳しながら、文単位での訳し直しも合わせて行うことで、チャンクと文、音声と文字という複数の方法で訳を提示する。コンピュータによる新しい同時通訳の方法であり、それを実現できる未来がすぐそこまで来ていると考える。

そのためには、テキストを適切な位置で区切るチャンクの技術に加え、翻訳システム自体を多分野で高精度化する必要がある。深層学習に基づく AI 翻訳シス

¹ 1 人当たり着地消費額が 100 万円以上の旅行を指す。

² アクティビティ、自然、異文化体験のうち 2 つ以上を含む旅行を指す。

テムを構築するには、原文と訳文を対にした対訳データが必要で、対訳データをたくさん集めるほど高精度にできる。1 組織で大量のデータを集めることは難しいため、翻訳バンクの枠組みの下、多組織でデータを集め、これまでに金融や製薬分野での専用モデルを構築した。まだまだデータが足りない分野があるため、引き続き皆様のご協力をお願いします、と隅田氏は結んだ。

4. パネリスト発表

パネルディスカッションの前に、JNTO 理事・中山理映子氏、ポケット株式会社代表取締役社長兼 CEO・松田憲幸氏、名古屋市役所観光文化交流局参事・伊藤毅氏の 3 名のパネリストによる発表が行われた。各発表の内容を紹介する。

(1) 中山理映子氏「インバウンド観光における多言語対応状況等」

多言語による情報発信・提供の現状を紹介する。JNTO においては、「旅マエ」において、訪日観光の魅力を伝え、訪日意欲を喚起する観点から、その国の言語で訪日観光情報発信することを重要視しており、重点市場 23 か国・地域向けに、計 13 言語でウェブサイト及び SNS を運用し、情報発信を行っている。

「旅ナカ」において緊要性の高い医療に関しては、観光庁による日本での医療機関受診に関する多言語検索・情報サイト³があり、診察を受けられる言語などを指定して医療機関を検索できる。自然災害等の非常時向けには、JNTO が年中無休・24 時間対応の多言語コールセンター「Japan Visitor Hotline」を運営しており、多言語チャットボットも運用している。

次に、多言語による観光案内の現状を紹介する。「旅ナカ」で訪日観光の満足度を高めてくれる「全国通訳案内士」の資格試験は、10 言語で実施されているものの、大部分は英語（2019 年合格者数 505）であり、中

国語（同 47）、フランス語（同 21）、スペイン語（同 18）、韓国語（同 17）と続く。一方、国・地域別訪日外国人観光客数では、中国語、韓国語、英語が主要な使用言語である各国・地域の他に、東南アジア諸国からの訪日客数が多く、全国通訳案内士の人数とバランスが取れていない。

JNTO では、日本各地の観光案内所をカテゴリー（1～3）ごとに認定し、その活動支援を行っているが、このうち半数近くを占めるカテゴリー1 には、英語対応人材の常駐を求めているが、電話通訳サービスや多言語翻訳システム等による英語対応も可としている。観光庁による調査結果では、カテゴリーを問わず、6 割前後の認定観光案内所において、多言語機械翻訳ツールを導入しており、特に、英語以外の外国語で導入率が高い。多言語翻訳ツールを導入していない理由としては、導入予算の不足、ツールを使用すると案内しにくい、ツールを使いこなせていないなどの理由が多くを占めた。

基調講演でも述べた通り、今後のインバウンド観光の方向性として地方への誘客を重視しており、その地域に関して豊富な知識を有する質の良いガイドの確保が重要となる。しかし、特に地方部において多言語に対応できる人材を確保することは容易ではなく、多言語翻訳技術を活用する余地は極めて大きいと考える。

(2) 松田憲幸氏「POCKETALK」

ポケット株式会社は「言葉の壁をなくす」をミッションとして事業を行っている。2017 年にポケット端末を販売開始してから、累計出荷台数が 100 万台を超え、日本国内での翻訳機のベンダー別販売金額シェアのうちポケットシリーズが 99.5%を占めている。国外でも、米国や欧州で多く使われている。現在はスマホ用ポケットアプリと、PC 用ポケット字幕も提供しており、2022 年 10 月にはポケット同時通訳も発表した。

AI 通訳・翻訳機に求められる観点として翻訳速度や使いやすさもあるが、翻訳精度が最も重要と考えてい

³ https://www.jnto.go.jp/emergency/jpn/mi_guide.html

る。ポケトークが高精度である理由は、音声認識、翻訳、音声合成の各処理で、また対象言語に応じて、最も高精度のエンジンを選択しているためである。たとえば、日本語に関する翻訳では NICT、英語などの言語では DeepL というように言語ごとの最適なエンジンを選んでいく。セキュリティについても、翻訳機で唯一、米国連邦規制基準 HIPAA と EU 一般データ保護規則 GDPR に準拠した製品であり、特に欧米企業・団体に選ばれる最大の要因の一つとなっている。

AI 多言語翻訳の社会的な役割として、観光、グローバル展開に加え、労働現場や人道支援への貢献が挙げられる。米国では年々移民の割合が増加し、在住者の 20%程度が十分に英語を話せていない印象がある。米国でのポケトークの導入は全米の企業・団体に広がり、病院・救急車、学校、警察、運転免許センターなどの労働現場へ導入が進んでいる。学校では 1000 台単位といった大規模な導入実績があり、学校側と非英語話者の生徒や保護者とのコミュニケーションツールとして活用されている。人道支援としては、ウクライナ大使館にポケトークを 1000 台寄贈し、ポーランドへの避難民にもポケトークを寄贈した。ポケトークは、Newsweek の「コロナで世界に貢献したグッドカンパニー50 社」にも選出された。

(3) 伊藤毅氏「地方自治体における自動翻訳の活用可能性～名古屋市の事例を通じて～」

日本国内の在留外国人数は約 300 万人に達し、またコロナ禍で行政の支援が必要な場面も増え、地方自治体における「言葉の壁」が顕在化している。そのため、多言語での行政サービスの充実が必要である。自治体での AI を含む翻訳・通訳の利用場面には、ウェブサイト翻訳、窓口対応、文書翻訳がある。特に文書翻訳については、年間の起案文書約 50 万件に対し、翻訳されている文書は 8 言語で各約 400 件に留まっている。AI 翻訳には、言語による情報格差を解消する役割が期待される。AI 翻訳導入の効果としてコスト削減の側面

もあるが、住民へのより良いサービスの提供が第一目標である。

名古屋市国際交流課は、全庁の翻訳・通訳を主に担当しており、2020 年より全国初の行政文書 AI 翻訳プロジェクトを開始した。ICT 活用による情報格差の解消と、全国自治体の多言語化推進を目標に、次の五つを行ってきた。(1) 翻訳資源として対訳用語集と対訳フレーズ集を 5 言語について構築した（今後 8 言語に拡大予定）。(2) 翻訳エンジン選定のために機能比較と効果測定を行い、当初の六つの候補から NICT の「みんなの自動翻訳」を選定した。(3) AI 翻訳の試験導入と効果検証を行い、翻訳に要する 1 文書当たり平均日数が減少したことで有用性を確認した。(4) AI 翻訳活用ガイドラインとして、市役所内で AI 翻訳を上手に使うための 4 つのルールを定めた。最後に、(5) 名古屋市役所内の各局へ、多言語対応状況と AI 翻訳のニーズ調査を行った。調査結果として、行政文書の多言語化状況は不十分という回答が多かったものの、AI 翻訳ツールの使用希望については、気軽に利用可能となることを期待する回答の他、翻訳の精度や正確性の担保を懸念する回答などがあり、意見が分かれた。

自治体への AI 翻訳導入における今後の課題として、品質保証の仕組み、導入予算の確保といった問題と、翻訳精度やシステムを使いこなすことへの不安に対する理解・受容の促進が挙げられる。これらの課題に対し、利用者側が現在の技術水準を理解し、知識をアップデートしていくことも必要である。今後の展望としては、名古屋市におけるベストプラクティス、翻訳資源、ガイドラインといった知見を日本・世界へ発信・展開していくことで、1 事例として参考になれば良い。AI 翻訳システムには、行政文書への適応や、希少言語への対応を期待するが、自治体として、技術の適正な利用や、開発者へのフィードバックも行いながら、AI 翻訳を活用していきたいと考えている。

5. 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」

NICT 上席研究員・内山将夫氏をモデレータとして、前述のパネリスト 3 名によるパネルディスカッションが行われた。以下、中心的な内容を紹介する。

a. 翻訳結果の正しさについて

松田氏：特にマイナー言語について翻訳結果の正しさを判断することは難しい。しかし、逆翻訳の機能を使うと適切に伝わっているかどうか確認できる。名古屋市での翻訳精度の懸念についても、逆翻訳を活用されるとどうか。

伊藤氏：会話ではそのような使い方もできそう。行政文書では、翻訳内容の正確さに慎重にならざるを得ない。誤訳を減らすためにプリエディットをするなどのガイドラインも作成し、対処している。

b. 言語内バリエーションへの対応について

中山氏：英語でもアメリカ人、イギリス人、インド人、シンガポール人の英語はかなり違って聞こえる。日本語でも各地域の方言があり、アクセントなどが異なる。ポケトークではどの程度対応可能か？

松田氏：英語ではアメリカ英語、オーストラリア英語、シンガポール英語、インド英語などに対応している。スペイン語やフランス語でもメジャーな方言に対応している。日本語でもいろいろトライシ、関西弁はかなり対応できている。

c. 翻訳機の備蓄について

伊藤氏：人道支援としてのポケトーク寄贈の話があったが、国内でも災害時向けに予め翻訳機をストックしておくことが大事だと気付かされた。

松田氏：ウクライナからの避難民へ提供されるパッケージにポケトークも含まれている。特に希少言語では翻訳機が本当に役立っていると聞いている。

伊藤氏：インバウンド観光などの短期在留者向けにも、翻訳機の備蓄や提供が有効だと思う。

d. 即時的な情報伝達への対応について

中山氏：災害時など、情報は時々刻々変わっていくが、それを逐一人手で翻訳していたのでは間に合わない。そのような場合に、即時に正確な情報を提供できる自動翻訳技術は有用なのではないか。

松田氏：病院受診時などは即時性が必要だが、希少言語については通訳者を見つけるのが困難。日本でもアメリカでも、病院での利用に重宝されている。

e. ガイドライン作成について

フロア質問者：伊藤さんのお話で、翻訳ツールを選ぶ際にガイドラインがなく、ご自身で作られたことは素晴らしい取り組みだと思う。翻訳ツールのガイドラインがもっと整備される必要があると思うが、こんなガイドラインや取り決めがあると良いという期待や意見はあるか？

伊藤氏：翻訳ツールを選ぶ際のガイドラインは作成していない。翻訳エンジン選定時に何をベースにするかの目安にするための基準のようなものは作った。作成したのは運用時のガイドライン。こちらは自治体間で共有できるが、エンジンの選定は各自自治体で決めるものだと思う。自治体としてはどこかと契約しないとイケないので、いろんなエンジンの中からフリーに選べるサービスがもしあれば使いやすい。

6. おわりに

各講演および議論の内容について、報告者によるまとめを述べる。JNTO 中山氏からは、インバウンド観光における現状と今後の展開、それらを踏まえた多言語翻訳技術の必要性について講演いただいた。NICT 隅田氏からは、AI 同時通訳の研究開発状況を中心に、翻訳バンクの取り組みについて講演いただいた。ポケトーク株式会社松田氏からは、移民割合が増えている

米国労働現場を中心としたポケトーク導入実績の拡大について講演いただいた。名古屋市伊藤氏からは、自治体行政における多言語翻訳の必要性和、名古屋市での AI 翻訳導入に関する先進的な取り組みについて講演いただいた。このように、インバウンド観光、行政サービス、米国労働現場といった、社会生活の様々な場面で自動翻訳の導入・利用が広がっていることを確認できた。

一方で、自動翻訳技術に対する課題・期待として主に次の 2 点が挙げられた。1 点目は、様々な言語への多言語化の重要性であり、特に翻訳・通訳者の確保が難しい希少言語において、自動翻訳の必要性・有用性が高いことが指摘された。2 点目は、自動翻訳結果の翻訳品質の担保の課題であり、技術に関する利用者側の理解や利用ガイドラインの整備なども含めて運用していくことの重要性が指摘された。

ニューラル機械翻訳におけるエヌビディアのハードウェアと SDK

呉先超

エヌビディア合同会社

1. はじめに

本稿では、大規模深層学習をサポートするエヌビディアの H100 GPU や自然言語処理とニューラル機械翻訳を含むオープンソース SDK NeMo（訓練用）や Riva（推論用、応用向け）を紹介する。

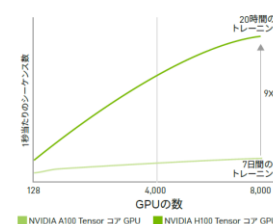
2. NVIDIA H100 Tensor コア GPU

アクセラレーテッド コンピューティングを桁違いに高速化：NVIDIA H100 Tensor コア GPU は、比類のないパフォーマンス、拡張性、セキュリティをあらゆるワークロードに提供する。NVIDIA® NVLink® Switch System との組み合わせで、最大 256 基の H100 GPU を接続して、エクサスケールのワークロードを高速化する。さらに専用の Transformer Engine で、数兆個のパラメーターを持つ言語モデルや翻訳モデルにも対応している。H100 は、NVIDIA Hopper™アーキテクチャが実現する画期的な変革を利用して業界最高水準の対話型 AI を提供し、大規模言語モデルの処理を前世代の 30 倍以上の速さで実行する。

エンタープライズ AI への対応は？ 標準サーバー向け NVIDIA H100 Tensor コア GPU には、エンタープライズサポートを含む 5 年間の NVIDIA AI Enterprise ソフトウェア スイートのサブスクリプションが付属しており、その非常に優れた性能によって、AI 活用を簡素化する。それにより、企業や組織は、AI チャットボット、レコメンデーション エンジン、ビジョン AI など、H100 で高速化する AI ワークフローを構築するために必要な AI フレームワークおよびツールが活用できるようになる。NVIDIA AI Enterprise ソフトウェア サブスクリプションやこれに関連する NVIDIA H100 サポートの詳細については、こちら¹をご覧ください。

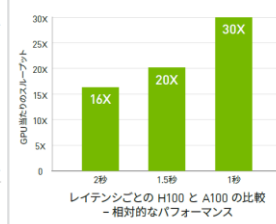
エンタープライズからエクサスケールまで、さまざまな規模のワークロードをセキュアに高速化：NVIDIA H100 GPU は、第 4 世代の Tensor コアと FP8 精度の Transformer Engine を搭載し、大規模言語モデルを使うトレーニングにおいて前世代比で最大 9 倍高速化、推論の速度を最大 30 倍と驚異的に高めることで、AI 市場における NVIDIA のリーダーシップをさらに確かなものにする。ハイ パフォーマンス コンピューティング (HPC) のアプリケーションについても、FP64 による FLOPS（1 秒間に実行できる浮動小数点演算の回数）を 3 倍に高め、新しい DPX 命令は、動的計画法のアルゴリズムを最大 7 倍まで高速化する。第 2 世代のマルチインスタンス GPU (MIG)、組み込みの NVIDIA コンフィデンシャル コンピューティング、NVIDIA NVLink Switch System を装備する H100 は、エンタープライズからエクサスケールまで、さまざまな規模のデータセンターのあらゆるワークロードを安全に高速化する。

最大規模のモデルを使う AI トレーニングを最大 9 倍に高速化
Mixture of Experts
(3,950 億個のパラメーター)



予想されるパフォーマンスは変更される可能性があります。入カシーケンス長を 128、出力シーケンス長を 20 としたときの Mixture of Experts (MoE) Transformer Switch-XSL のバリエーション H100 クラスター H100 推論 ネットワーク H100 クラスター H100 クラスター NVLink Switch System, H100 推論

最大規模のモデルを使う AI 推論を最大 30 倍に高速化
Megatron チャットボット推論
(5,300 億個のパラメーター)



予想されるパフォーマンスは変更される可能性があります。入力シーケンス長を 128、出力シーケンス長を 20 としたときの Megatron 5,300 億パラメーター モデル チャットボットによる推論 H100 クラスター H100 クラスター H100 クラスター NVLink Switch System, H100 推論

図 1、トレーニングと推論の高速化の比較 (A100 vs. H100)

3. NVIDIA DGX H100 AI インフラの金字塔

人工知能 (AI) はビジネスにおける困難な課題を解決するための有力な手段になった。顧客サービスの向

¹ <https://www.nvidia.com/ja-jp/data-center/h100/activate/>

上、サプライチェーンの最適化、ビジネス上の知見や洞察の獲得、ほぼすべての業界における最先端の製品やサービスの設計など、AI は組織にイノベーションを実現するための仕組みを提供する。AI インフラの先駆者となった NVIDIA DGXTM システムは、このような重要な概念を実現するための最も強力に完結した AI プラットフォームを提供する。

NVIDIA DGX H100 はビジネスのイノベーションと最適化を支援する。NVIDIA の伝説的な DGX システムの最新モデルであり、NVIDIA DGX SuperPODTM の基盤である DGX H100 は、画期的な NVIDIA H100 Tensor コア GPU を搭載し、AI の活用を押し進める。AI スループットを最大化するように設計されており、高度に洗練され、システム化された拡張可能なプラットフォームを企業に提供し、自然言語処理、リコメンデーション システム、データ分析などにおいて、飛躍的な進歩を達成できるように支援する。NVIDIA DGX H100 は、オンプレミスに導入できるほか、さまざまなアクセスと展開の選択肢があり、企業の大規模な課題を AI で解決するために必要なパフォーマンスを実現する。

4. NVIDIA NeMo²

NVIDIA AI プラットフォームの一部である オープンソースの NVIDIA NeMo は、新しい最先端の会話型 AI モデル、自然言語処理、音声処理、機械翻訳などの深層学習モデルの訓練や推論を構築するためのツールキットである。NeMo には、自動音声認識 (ASR)、自然言語処理 (NLP)、およびテキスト読み上げ (TTS) 合成モデル用の個別のコレクションがある。各コレクションは、データのトレーニングに必要なすべてを含む事前構築済みのモジュールで構成されている。すべてのモジュールを簡単にカスタマイズ、拡張、構成して、新しい会話型 AI モデル アーキテクチャを作成できる。

通常、単言語・多言語会話型 AI アーキテクチャは大規模であり、トレーニングのために大量のデータと

計算を必要とする。NeMo は PyTorch Lightning を使用して、簡単で高性能なマルチ GPU/マルチノード混合精度トレーニングを実現する。

すでに公開されている事前トレーニング済みの NeMo モデルは、Transformer などのネットワークに基づいて、10 数個言語の間の翻訳モデル³と使い方⁴が含まれている。

5. NVIDIA Riva⁵

エヌビディアの Riva は、スピーチ AI、自然言語処理、文章翻訳、音声翻訳などの アプリケーションのための シンプルなエンドツーエンドの ワークフローです。

Riva の機能: 学習済みのスピーチ AI SOTA モデル: データセットに合わせて完全にカスタマイズ可能な ASR モデル、TTS モデル、翻訳モデルが 200 個以上揃っている。分野ごとに特化されたモデルの開発を 10 倍加速する。

高性能推論: 推論は NVIDIA TensorRTTM 最適化と NVIDIA TritonTM Inference Server によって強化されており、いずれも NVIDIA AI プラットフォームに含まれている。

Riva サービス: 低遅延ストリーミングや高スループットのオフラインでの利用に向けて、gRPC ベースのマイクロサービスとして提供される。

高い拡張性: Riva は、完全にコンテナ化されており、数百および数千の並列ストリームに容易に拡張できる。

NeMo と Riva は、すでに日本語音声認識モデルやアジア言語に対する、訓練済みの BERT、GPT、翻訳モデルが豊かにそろっていて、エヌビディア NGC⁶で公開している。

² <https://developer.nvidia.com/nvidia-nemo>;
<https://github.com/NVIDIA/NeMo>

³ https://docs.nvidia.com/deeplearning/nemo/user-guide/docs/en/v1.0.0/nlp/machine_translation.html#

⁴ <https://docs.nvidia.com/deeplearning/nemo/user-guide/docs/en/stable/starter/tutorials.html>

⁵ <https://www.nvidia.com/ja-jp/ai-data-science/products/riva/>

⁶ <https://www.nvidia.com/en-us/gpu-cloud/>

編集後記

編集後記

園尾 聡

AAMT 編集委員会

世間では Open AI 社から発表された ChatGPT/GPT-4 の性能に衝撃を受け、一部の業務スタイルが激変するかもしれないと話題になっています。GPT の元となっているのは NMT のアーキテクチャである Transformer であり、翻訳業界においても技術動向を注意深く把握する必要があると感じております。

巻頭言では、永田昌明様から「二度目の嵐」は吹き荒れるか? という題で、ChatGPT の登場によって機械翻訳研究が今後どのようなようになるかについて、ご自身の経験を踏まえてご考察をいただきました。

堂本秋次様からは「機械翻訳らしさを軽減するための手法の妥当性と MTPE の効率化の条件について」という題で、訳文からだけでは機械翻訳文と人間が翻訳した訳文とは区別できないという興味深い実験結果を踏まえ、翻訳者が PE においてどう工夫すべきかについてご提言いただきました。

上田有佳子様からは「機械翻訳活用で顧客の期待に応える」(英文タイトル: Case Study: How NetApp is Utilizing Machine Translation to Meet Customer Needs and Expectations)という題で、情報提供スピードが求められる業界において、PE なしの Raw NMT を活用した情報発信の成功事例についてご報告いただきました。

影浦峽様からは「人間の翻訳と機械の翻訳(7): 社会的な受容の一側面」という題で、MT にできること・できないことについて、MT を受け入れる社会的基準の観点からの考察をいただきました。

AAMT からは「ISO18587 供給者自己適合宣言のススメ」および「MT ユーザーガイドへのお誘い」を紹介しました。ISO 供給者適合宣言支援ガイドラインは AAMT 会員であれば無料で利用できるため、ぜひ活用

を検討いただければ幸いです。また、MT ユーザーガイドは MT を正しく使うために必要な知識 (MT リテラシー) を提供することを目的に一般公開しており、ぜひご一読いただきたいと思います。

西山聖久様からは「理工系の英語論文執筆における課題と機械翻訳の活用」と題し、第 2 回 AAMT セミナーに関するご報告をいただきました。3C(Clear, Collect, Concise)の観点からテクニカルライティングにおける課題と実際の MT 活用法についてのご解説をいただきました。

西村賢治様からは「海外投資家を引き付ける英文 IR 〜第 3 回 AAMT セミナー要旨」と題し、第 3 回 AAMT セミナーに関するご報告をいただきました。株式市場の要となる IR 情報を英語で発信することの重要性ならびに MT の可能性についてのご解説をいただきました。

第 2 回目となる温故知新では、1991 年 11 月号と 1992 年 1 月号からの記事を紹介しております。機械翻訳システム開発を F1 の世界 (マシンの性能とドライバーの技能) に例えられているのは 30 年後の現在にも通じるところがあるように思えます。また、Mu プロジェクトから始まった機械翻訳がビジネスとして儲かるかの対談も紹介されています。

出内将夫様からは AAMT 2022, Tokyo の参加報告をいただきました。本イベントでは、2 件の招待講演、MT Summit2023 の紹介、7 件の公募セッション、パネルディスカッションがあり、3 年ぶりの会場開催もあったことから非常に活発な議論が行われ、盛況となりました。

東山翔平様からは第 6 回自動翻訳シンポジウムの参加報告をいただきました。会場には 389 名の参加者が

Editor's note

Satoshi Sonoh

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/>

集まり、2 件の講演、パネルディスカッション、企業展示が行われ、盛況だったとのこと。MT (AI 同時通訳) が様々な場面で使われている一方、希少言語を含む多言語化や翻訳品質の担保が課題であると、まとめられています。

法人会員であるエヌビディア合同会社様からは「ニューラル機械翻訳におけるエヌビディアのハードウェアと SDK」という題でご寄稿いただきました。

今年 9 月には AAMT が主催する MT Summit 2023 が中国・マカオで開催されます。現地およびオンラインで活発な議論、多数のコミュニケーションが行われることを今から楽しみにしております。

AAMTジャーナル「機械翻訳」No. 78

- 【 発 行 日 】 2023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 】 〒619-0289
京都府相楽郡精華町光台3-5
国立研究開発法人情報通信研究機構 先進的翻訳技術研究室内
【 編 集 委 員 会 】 内山将夫 後藤功雄 中澤敏明 新田順也 園尾聡 森口功造 隅田英一郎
山田優 石川弘美 早川威士 出内将夫
【 表 紙 デザイン 】 泉谷東十郎
【 題 字 】 長尾真
【 事 務 局 】 石川弘美
【 印 刷 所 】 株式会社 プリントバック

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

