

ローカリゼーションの未来



デジタル変革における言語と文化の活用

制作協力:

LIONBRIDGE

ビジネス展開の核

ローカリゼーションとは、製品、コミュニケーション、サービスなどを特定地域の顧客の期待に添うように適応させる取り組みです。

世界のグローバル化が進む中、ローカリゼーションによって海外市場に参入し、スケールメリットを活かして収益を増やすことができます。このため、製品の海外展開を視野に入れたビジネス戦略にローカリゼーションは欠かせません。たとえば「Designed by Apple in California」という Apple 製品についての約束のように、ブランディングを通じて特定の地域と結び付いている企業であっても、国境を越えた顧客にアピールするためにはローカリゼーションへの取り組みが不可欠です。

ローカリゼーションの中心には翻訳があり、この eBook で取り上げていくように、従来、この作業にはプロの翻訳者がかかわる場合がほとんどでしたが、近年においては最先端の機械翻訳 (MT) ツールを活用する場面も増えています。また、人間による翻訳と機械翻訳の両方を採用する「拡張翻訳」(5 ページを参照) と呼ばれるハイブリッドアプローチを採用している企業もあります。

しかし、真のローカリゼーションとは言語だけにはとどまりません。ある市場から別の市場へのサービスまたは製品の展開を成功させるには、各地域の微妙な違いを理解することが大切です。「基本的に、エクスペリエンスを別の言語で再構築するのです」と NI (National Instruments 社) で言語サービス シニア マネージャーを務めるイネスルビオ氏は AI Business に話しています。「これは言葉の問題ではなく文化の問題です」

翻訳とローカリゼーションを専門とする当社ライオンブリッジがお届けするこの eBook では、ビジネスの海外展開で直面する課題、そしてローカリゼーションを適切に行なった場合のメリットをご紹介します。また、それぞれの専門家にインタビューし、Zynga、Royal Caribbean、Volvo、ウィキメディア財団など、当社のお客様の見解も取り上げます。この eBook を通じて、真にグローバルな製品やサービスを構築するために必要な知識をご提供できれば幸いです。

Luke Dormehl | 編集委員 | AI Business

eBook 制作協力:

LIONBRIDGE

ローカリゼーションの未来 | www.lionbridge.com



3

現代のローカリゼーション サービスの概要
言語と文化

4

市場の概況
ローカリゼーションが主流に

5

AI の実用化
拡張翻訳とは?

7

ケース スタディ
Royal Caribbean がライオンブリッジの
支援によりローカリゼーションの
取り組みを全面的に見直し

8

グローバル オーディエンスのために
ローカリゼーションを通じて数百万人の
ゲーマーの心を捉える Zynga 社の戦略

10

ローカルで考える
言語サービスを活用して
ウィキペディアを向上

12

機械翻訳の進化
転換点

14

ケース スタディ
ライオンブリッジが Volvo Cars の
全 Web サイト構築を一元的にサポート

15

次なる未来は?
言語テクノロジーの未来

17

アカデミックな視点
パリティの問題

言語と文化

翻訳は重要 – しかもっと重要なのは対象のオーディエンスを理解すること

これは言葉の問題ではなく文化の問題です。

たとえば、

スペイン語について語る

場合、どこを対象として

いるのでしょうか？

世界中のスペイン語

でしょうか？

ヨーロッパのスペイン語

でしょうか？

対象のオーディエンスに

自分事と捉えてもらうには

どうすればいいのでしょうか？

単なる言葉の置き換えでは

済まないのは明らかです。

もっともっと**深く詳細な**

対象への配慮です。

一般的に、ローカリゼーションとは、Web サイト、使用マニュアル、ソフトウェア アプリケーションなどの言語資産の翻訳を指しており、これらを他言語のユーザーが読んで理解できるようにする取り組みを意味します。

「ローカリゼーションの目的は、実際に製品やサービスを現地に適応させることで、その重要な要素は、最終顧客の言語にすることです」とライオンブリッジで製品リーダーとストラテジストを務めるカジェタン マリノフスキーは AI Business に語ります。しかし、言語は全体のほんの一部に過ぎません。世界で最も収益性の高い市場では、その文化に関する専門知識が必要となります。

翻訳を超えて

皆さんは「言っていることは正しいが、言い方がよくない」というセリフを聞いたことがあるでしょうか？ローカリゼーションが単純に翻訳と呼ばれない理由がここにあります。ローカリゼーションは、現地のコミュニケーションの幅広い側面を考慮することを意味します。言語分析はその一部ですが、地域的な違いも同様です。たとえば、計測単位は地域によって基準が変わるわかりやすい例です。地域ごとに異なる要素として、ほかには規制や技術的な面があります。さらに「声のトーン」、つまり声域すらも異なります。

「基本的に、エクスペリエンスを別の言語で再構築するのです」と NI (National Instruments 社) で言語サービス シニア マネージャーを務めるイネス ルビオ氏は AI Business に話しています。「これは言葉の問題ではなく文化の問題です。たとえば、スペイン語について語る場合、どこを対象としているのでしょうか？ 世界中のスペイン語でしょうか？ ヨーロッパのスペイン語でしょうか？ 対象のオーディエンスに自分事と捉えてもらうにはどうすればいいのでしょうか？ 単なる言葉の置き換えでは済まないのは明らかです。もっともっと**深く詳細な対象への配慮**です」

ローカリゼーションの範囲としては、対象のオーディエンスの好みに合わせるために、インターフェイスのアイコンなどのグラフィックの翻訳が含まれる場合もあります。そのようなアイコンは、必ずしも普遍的または明白な意味を持っているわけではありません。たとえば、アイデアを表す電球は米国・英国市場ではよく使われる「普通」のアイコンですが、他の地域では必ずしも同じ意味で捉えられるとは限りません。また、親指を立てた「サムズアップ」の絵文字は米英では肯定的な意味合いですが、ラテン アメリカ、中東、西アフリカ地域では攻撃的な意味を持っています。色が持つ意味もさまざまです。インドではオレンジは愛と勇気を意味しますが、中東の文化では喪失や追悼に関連付けられるのが一般的です。

適切なローカリゼーションによるメリットは計り知れません。一方、ローカリゼーションが不適切であった場合や、現地の習慣の重要性を理解していなかった場合は、逆に大きなダメージとなる可能性があります。これはすべての企業が今後重視すべきツールの一つと言えます。



ローカリゼーションが主流に

ローカリゼーション市場の規模とその成長要因

「ビジネスを始めたばかりのときにはローカリゼーションについて考えることもないでしょう」と彼女は言います。「ギリギリになってようやく考える程度で、立ち上げの段階ではまったく考慮されていません。間違いから学び、後からその種の変更を行うことでどれだけコストがかさむかを認識するに至って、企業とチーム全体が**そのありがたみを知る**ようになるのです」

L と N の間の文字数を数字の 10 に置き換えて「L10N」と略されることも多いローカリゼーション (Localization)。現代のローカリゼーションの誕生は 1980 年代にまでさかのぼります。コンピューター ソフトウェアが世界中のオフィスや家庭に普及し始めたのが 80 年代であるのは偶然ではありません。企業は成長する市場機会を捉えるために、新しい文字セットのサポートであれ、特定の市場に合わせた機能の追加や変更であれ、ローカリゼーションを進める必要がありました。当初は社内で行われていた作業ですが、その規模と重要性が相まって、本格的なローカリゼーション業界の急速な発展につながりました。

今までにない規模

過去 10 年間で、言語サービス市場 (インストラクション、翻訳、幅広いローカリゼーションを含む) の規模がほぼ倍増しました。言語サービス業界に関する出版大手の Slaton による最近のレポートでは、2020 年の翻訳、ローカリゼーション、通訳市場の規模は 238 億ドルでした。その顧客の業界を見るとテクノロジー、ゲーム、メディア、そしてライフサイエンスが急成長しており、業界全体としては 2021 年に最大 10% の成長が見込まれました。

Statisa によると、世界人口の 10% にも満たない欧州が、世界の言語サービス市場のほぼ半分を占めています。これは、この地域で話される言語の多様性によるものです。一方、市場規模では米国がトップを維持しています。

不均衡が見られるのは国単位の話だけではありません。ローカリゼーションは一部のセクターでは成熟した分野となっていますが、他の (特に新興の) 業界や企業では、その重要性はまだ認識されつつある状況です。

NI で言語サービス シニア マネージャーを務めるイネス ルビオ氏は、ローカリゼーションに対するコミットメントで成熟度を測れることを示唆しました。「ビジネスを始めたばかりのときにはローカリゼーションについて考えることもないでしょう」と彼女は言います。「ギリギリになってようやく考える程度で、立ち上げの段階ではまったく考慮されていません。間違いから学び、後からその種の変更を行うことでどれだけコストがかさむかを認識するに至って、企業とチーム全体が**そのありがたみを知る**ようになります」。ローカリゼーションがプロセスに組み込まれるのは、ようやくそれからです」

ローカリゼーションに関わる技術進歩のおかげで、この業界の力となるツールは日々向上しています。ニューラル機械翻訳 (NMT) などの新たなイノベーションが登場したのはこの数年の話ですが、すでに大きな影響を与えています。このような画期的な進歩が、顧客とエンド ユーザーの双方にとってのローカリゼーションを大きく変えていきます。





拡張翻訳とは？

ライオンブリッジのウィル ローランズ-リースと ジェイミー プニシルに聞く、人間と機械との関係

「機械翻訳は AI の基本的な応用例の一つです」と、ライオンブリッジの最高製品責任者であるウィル ローランズ-リースは AI Business にこう話しました。「最初に機械翻訳が行われたのは 1950 年代初めでした。当時でさえも翻訳者は必要なくなるだろうと予想されていました。その予想は少し外れていたかもしれません」

それから 75 年近くたちましたが、人はまだ翻訳作業の重要な部分を占めています。これは、機械翻訳が利用されているところでも言えることです。誰でも利用できる無料の翻訳エンジンは、特定のシナリオでは有効かもしれませんが、実務に適していることははまれです。

「何か文を貼り付けて出てきた訳についてはおもしろい話があるありましたよね」と、ライオンブリッジの最高マーケティング責任者であるジェイミー プニシルは AI Business に話しました。「これはトレーニングされていない公開エンジンの話です。商用環境では少しましですが、それでも 2 つの言語間でさまざまな言語

的問題が生じます。英語を韓国語に翻訳するのは簡単ではありませんし、ドイツ語から韓国語はなおさらです。機械がおかしな訳を作り出します」

優れた翻訳の重要性

話の大筋さえわかればいいのであれば、機械翻訳されたニュースの記事でも許容範囲かもしれませんが、しかし、実際に「十分」では不十分な状況があります。投資商品の目論見書、医学研究のラベル、法的な文書などは完璧な翻訳が要求される例です。

「主要なスポーツカーメーカーのマニュアルを作成する際は、誤った指示が及ぼす身体と安全性への影響を考慮します」とローランズ-リースは言います。「または、臨床研究や、その他の生死にかかわる「問題」とみなされ得る分野について考えてください。企業の観点から言えば、すべてがブランドにも影響します。このため、翻訳が正確であることを確認したいはずで、そうでなければ、ブランドの評判にかかわるからです」

従来から、このプロセスは「ポストエディット」と呼ばれています。このプロセスでは、最初に機械翻訳ツールによって初回の翻訳が行われ、次に機械翻訳済みのセグメントが翻訳者に渡されてコンピューターによるエラーが修正されます。理論的にはこの手法で時間を短縮できますが、翻訳者がかかわるのがプロセスの最終部分のみであるといった制約もあります。

拡張翻訳の活用

CSA Research による造語である「拡張翻訳 (Augmented Translation)」は、幅広いさまざまなテクノロジーを利用して、翻訳者が一貫性と正確性の高い高品質な翻訳を短時間で作成できるようにします。拡張翻訳では、人間とテクノロジーがうまく組み合わせられ、全体的な翻訳エクスペリエンスが向上します。これらの手法の多くは以前から試されてきましたが、それらを結合することで、より包括的な枠組みを構築することが可能となっています。付加価値は低くとも時間がかかる多くのタスクが自動化されるので、翻訳者はその専門性を最大限に活かすことができます。

拡張翻訳がメリットをもたらす
7つの主要領域

翻訳メモリ

アダプティブ ニューラル

機械翻訳

品質評価

自動コンテンツ強化 (ACE)

用語集管理

完全自動のプロジェクト管理

翻訳管理システム (TMS)

これらはすべて密接に機能します。たとえば、翻訳メモリは反復的な翻訳タスクに役立ちます。「同じ製品のマニュアルを毎年翻訳する場合があります」とブニシルは言います。「そのような製品マニュアルのほとんどは、前年の内容とほぼ同じです。それでもかつてのアナログの世界では、その翻訳のためにお客様の企業は2回分のコスト、もしくはそれ以上のコストをかける必要がありました。翻訳メモリは、ソース テキストを確認して合致するものがあれば通知するように設計されています」

完全一致のテキストであれば修正なしで次の処理に送ることができます。一方、一部が異なる場合には、後で翻訳者が更新できるようにフラグを付けることができます。翻訳者は、前後の語句や文をチェックしてその変更がテキストの文脈に及ぼす影響を確認できます。

拡張翻訳には、特定の翻訳ジョブに最適な人材と機械のリソースを自動的に決定できるという面もあります。つまり、仕事を翻訳者に依頼すべきか、またはどの機械翻訳エンジンなら最適な成果が生まれるのかを把握できることを意味します。

強みを活かす

「この分野でわかるのは、機械ごとに得意・不得意があることです」とブニシルは続けます。「英語からドイツ語、スペイン語から韓国語、製薬関連のコンテンツ、マーケティングのコンテンツなど、機械によって得意ものが異なります。このため、MT エンジンはうまく組み合わせる使用

必要があります。使用する翻訳システムを強化するには、機能の適切な組み合わせが求められます」

「ライオンブリッジでは、企業によるシステムの設定を支援し、企業に代わってこれらのシステムを運用します。当社の支援は、エンジンのトレーニングから、全体の運用に必要なブランドやボイスなどの各ガイドライン、用語辞書、その他すべての言語資産の維持管理、さらにそのスムーズな運用にまで及びます。ただ適切に翻訳するというだけではなく、たとえば Apple らしい、またはマイクロソフトらしい表現やトーンで実際に翻訳を行うのです」

ローランズ-リースが指摘します。「当社は、『このドキュメントは人間の手による翻訳がベストです。理由は…』、または『これは MT で処理しましょう。MT が適していると思う理由は…』というように、お客様に当社の提案について明確な説明を常に提供します。これは、当然ながら当社がそれぞれのケースで何が最適であるかを把握しており、提示案を通じて当社の見解をお客様にご説明し、大切なコンテンツを最適な方法で処理することができる能力を備えている

ためです。これは当社にとって効率的であるだけでなく、最も重要なことですが、お客様にとってもかなりのメリットとなります」

拡張翻訳では、テキストがうまく翻訳される部分を測定する予測分析の機能など、他のさまざまな方法でも翻訳作業を支援できます。これによってコンテンツ制作プロセスの上流と下流の両方の工程を向上できます。さらに、他の言語での検索エンジン最適化 (SEO) の分野で、包括的な評価も実施できます。これは、オンラインでの発見可能性にも大きく寄与します。

拡張翻訳は、成果物の大幅な品質向上を可能にする最先端の機械翻訳です。翻訳と品質予測分析 (機械翻訳で実現できる品質を統計的に評価して提供) のような各テクノロジーを最新のアダプティブ ニューラル機械翻訳ツールと統合することで、拡張翻訳は、あらゆるローカリゼーション チームにとって理想的な手法となります。今後数年の間に「拡張翻訳」はごく一般的な手法、そして用語になるでしょう。



Royal Caribbean がライオンブリッジの支援によりローカリゼーションの取り組みを全面的に見直し

大手客船運航会社がグローバル オーディエンスを引き付けるための方法を模索

世界中を航海するクルーズ客船は、当然ながら国際的であると言えます。毎年、世界中から 530 万人もの旅行者が Royal Caribbean のクルーズ船であちこちの魅力的な目的地に出航しようと列をなします。利用客に希望どおりのエクスペリエンスを体験をしてもらい、Royal Caribbean のブランドとしての統一性を維持するために、同社の Web サイトである royalcaribbean.com には高品質な翻訳が不可欠でした。だれにとってもわかりやすい Web サイトであれば、利用客はバケーションを予約するだけでなく、Royal Caribbean のクルーズ船、アクティビティ、目的地などを見て回って楽しむことができます。

解決すべき課題

Royal Caribbean のような一流ブランドの場合、場当たりの非効率なコンテンツ翻訳ワークフローによって、コンテンツの正確さやサイトのわかりやすさや、使い心地などが損なわれることがあります。コンテンツ プラットフォームのリニューアルを決めた際に、同社はローカリゼーションのエキスパートとして定評のある企業の支援を仰ぐことにしました。そしてお選びいただいたの

が、言語サービス業界で 25 年以上の実績を持つ当社ライオンブリッジでした。Royal Caribbean の目標は、すべてのコンテンツを、一元化されたサイトで閲覧数と売上の両方を伸ばし、翻訳ワークフローを改善して、優れたユーザー エクスペリエンスを提供することでした。

導入したソリューション

ライオンブリッジが提供したソリューションは、Lionbridge's Connector™ と Adobe Experience Manager™ の 2 つのツールで構成されたサービスの導入でした。これにより、Royal Caribbean のコンテンツの適切な更新と各国語での検索の最適化を実現する上で欠かせない、効率的なワークフローと多言語サポートの両方を提供しました。高品質な翻訳を妥当な期間内に妥当なコストで用意できるようになり、わずか 4 か月で、既存の Royal Caribbean コンテンツの 80% 以上、合計で約 1,500 ページを 7 つの言語に翻訳できました。

「ライオンブリッジの Adobe Experience Manager 向けコネクタを利用して Web サイトの翻訳プロセスを自動化した結果、数千もの Web ページ

1,500
ページ

ライオンブリッジが
4か月で7つの言語に翻訳

を簡単に翻訳できました」と Royal Caribbean Cruise Lines でプロダクト コンテンツ最適化マネージャーを務めるアリアドーナ カストロ氏は話します。「ライオンブリッジのテクノロジーによって、当社は効率的な翻訳ワークフローを実装できました。当社を担当したライオンブリッジのアカウント チームも非常に優秀で、当社の目標達成に向けて必要なときにはいつでもすぐに相談に乗ってくれました」

ローカリゼーション プロジェクトでは、何よりもまず、顧客に優れたユーザー エクスペリエンスを提供する必要があります。しかし、それが企業にとってなぜそれほどメリットのある投資なのかを示すような具体的な結果に結び付けることも重要です。Royal Caribbean の事例では、その効果がすぐに明らかになりました。導入後の数年間で、オーガニック クリック数は 35% 上昇、オーガニック 検索のインプレッション数はもっと多い 37% 上昇、そしてグローバル クリックはなんと 3 倍にもなったのです。

世界中の顧客にパーソナルなエクスペリエンスを届けることができ、ローカリゼーションの全面見直しによる恩恵を関係者全員が十分に享受できたのです。



ローカリゼーションを通じて 数百万人のゲーマーの心を捉える Zynga 社の戦略

Zynga 社でパブリッシングを統括するベルナルド キム氏と
戦略的パートナーシップを統括するヴィニート ラジヨシ シャーマ氏による、
世界中のゲーマーを魅了するゲームの開発についてのアドバイス

「社内では『ローカリゼーション』ではなく『カルチャリゼーション』と呼んでいます」と Zynga 社でパブリッシングを統括するベルナルド キム氏は言います。「作品の言語を切り替えるだけではなく、ゲームプレイや魅力的な機能に関してゲームを対象市場のそれぞれの状況に適応させるよう努めています」

ゲームは世界共通であり、国境を越えて多くの人々を魅了しています。現在のゲーム業界で成功を収めるために、大手ゲーム開発企業のお客様は先例を踏襲する必要があります。2007 年創業の Zynga 社は、世界有数のソーシャル ゲーム開発企業の一つ、世界で毎日数百万ものプレイヤーが同社のゲームを楽しんでいます。最も有名なタイトルの一つが『FarmVille』です。2009 年 6 月に Facebook で公開され、数週間のうちに一日当たりのアクティブ ユーザー数が 1,000 万人に達しました。10 年以上たった今、同社はその多くをモバイル プラットフォームに移行して、『Zynga ポーカー』、『Words With Friends 2』、『Game of Thrones Slot Casino』、『ハリー・ポッター：呪文と魔法のパズル』などのタイトルで、月間数千万人にも上る 150 か国以上のプレイヤーたちを魅了し続けています。

大きな成功を収める「フォーエバーゲーム」(同社では永続的な資産をこう呼びます)を制作するカギの一つは、製品を世界中の市場に適応させ

ることに重点を置くことです。第一言語または第二言語が英語という人は、世界人口の 12 分の 1 以下に過ぎません。英語のみのゲームでは、潜在的なユーザーベースのほとんどをみすみす逃すこととなります。つまり、ゲームを多言語で販売することは商業的に手堅い選択なのです。とは言え、Zynga 社の経営陣は、ローカリゼーションとはゲーム内のテキストをただ翻訳すればよいということではないと AI Business に語っています。

「カルチャリゼーション」と呼ぶ意図

「社内では『ローカリゼーション』ではなく『カルチャリゼーション』と呼んでいます」と Zynga 社でパブリッシングを統括するベルナルド キム氏は言います。「作品の言語を切り替えるだけではなく、ゲームプレイや魅力的な機能に関してゲームを対象市場のそれぞれの状況に適応させるよう努めています。たとえば、韓国のゲーマーは一つのゲームに夢中になって、一日のうちに何度もプレイする傾向にあります。このようなプレイヤー向けに、市場に合わせてカルチャライズしたオファーやチャレンジを作成します」

Zynga 社で戦略的パートナーシップを統括するヴィニート ラジヨシ シャーマ氏は、まさに自分のために作られたゲームだとプレイヤーたちに感じさせることが重要だと AI Business に語りました。「私がユーザーとしてアプリを使用したときに、『これは自分のために作られたものだ』と感じさせるようなものでなければなりません」



複数の国や地域の数百万にも及びユーザーを同時に魅了する一方で、それぞれがパーソナライズされていると感じさせるゲームを作るのは容易ではありません。ゲーム業界で長い経験を持つキム氏は、適切に翻訳されていない IP の課題に直面した経験は数多いと言います。彼が学んだ一つの教訓は、特定の市場に向けたタイトルのローカリゼーションを「付け足し」のような気持ちで扱ってはならないということです。

「このゲームが長くプレイできるものかどうかは**ゲーマーであればわかります**」とキム氏は言います。「操作性、言語、表現。ターゲットとする市場に向けて、これらのいずれかでも不十分なことがあった場合は、それがエンゲージメントの数字に**悪影響となって表れます**」

そういった状況に陥らないよう、Zynga 社ではすべての新規プロジェクトの開始時点で、ローカリゼーション(またはカルチャリゼーション)戦略も検討しています。大切なのは、潜在的なプレイヤーに「機会を広げる」ことです。

「当社のフォーエバー シリーズを見ただければわかりますが、いずれも世界に通用する魅力があります」とシャーマ氏は言います。コンセプトの選定については「カルチャライズが最優先事項です」と付け加えました。「アプリの表現はどの側面をとっても対象市場で理解されるものでなければなりません」

言語は重要だがそれがすべてではない

米国、カナダ、インド、アイルランド、トルコにオフィス拠点をもち、世界各地にスタジオを擁する Zynga 社にとって、言語は当然ながら重要な要素です。シャーマ氏は、ローカリゼーション プロセスをシンプルにするためのカギとして、テキストの量を減らすことを挙げました。「基本的に、ゲーム内のテキストが多いと仕事はその分大変になります。それに文字をたくさん読まされるのが好きなプレイヤーもいないでしょう」とシャーマ氏は言います。特にモバイル デバイスの小さな画面ではなおさらです。

さらに厄介なことに、すべてのテキストは、オリジナル タイトルが持つ

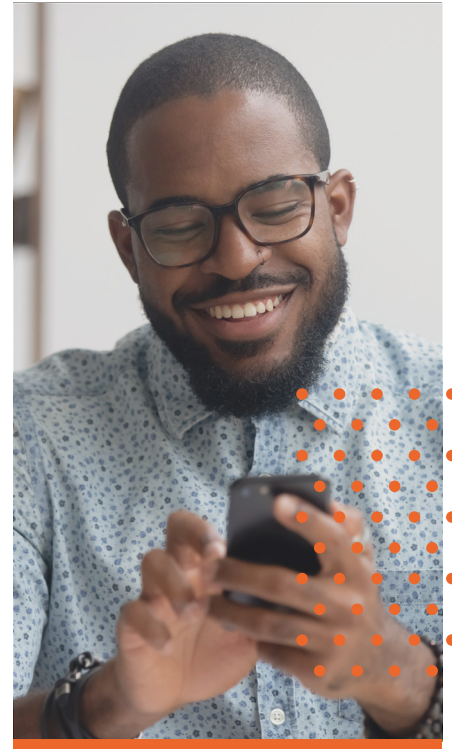
「ボイス」を伝えるように翻訳する必要があります。翻訳対象はゲーム内のテキストと表記だけでなく、通貨、測定単位、日付形式など作品世界のあらゆる要素が含まれます。次に、アプリ ストアの説明、マーケティング資料など、付随するその他のテキストについても翻訳が必要です。「社内には翻訳者がおり、アウトソーシングもしています」とシャーマ氏は説明します。「当社が力を入れて磨き上げているのが**翻訳の品質**です。社内にはローカリゼーション QA チームが存在し、私たちは『インターナショナル QA』と呼んでいます。翻訳された内容が理解しやすく矛盾がないことをここで確認しています」

書き言葉と話し言葉という要素は、幅広いカルチャリゼーションの取り組みのほんの一部分に過ぎません。たとえば、考慮すべきものとして、特定の記号、性別の呼び方、役職があります。現地で人気のソーシャル メディア プラットフォームの統合についても、マーケティング目的か、認証のためか、またはソーシャル共有のためかを問わず、同様に考慮が必要です。キム氏は、市場によって異なる細部で、注目すべきものの一つがゲーム内の「アプリ内課金パッケージの形」であり、特定の地域ではプレイヤーをより引き付けるためによく変更されると述べています。過去に、同社は特定のアジア市場で最適なエクスペリエンスを保証するために、ハリー・ポッターなどのゲームのリリースを遅らせたことがありました。

「当社が翻訳者を抱えている理由はまさにこれです。翻訳者たちは長年にわたって当社のゲームを担当しており、ゲームについて膨大な知識を持っています」とキム氏は述べています。

基本無料プレイにおけるローカリゼーションの重要性

Zynga 社の経営陣は、基本無料プレイ ゲームのセクターの方がローカリゼーションはずっと重要だと話しました。購入に一定の料金がかかるプレミアム タイトルとは異なり、「フリーミアム」ゲームでは、繰り返しプレイするプレイヤーを呼び戻すことで収益化につなげています。これは、うわべだけのローカリゼーションに注力するだけでは足りないということを意味しており、たとえ現地マーケティング キャンペーンをうまく実



施したとしても、ゲーム内の変更が最小限では不十分となります。キム氏は、フリーミアムで高品質を確保するのは、多数のゲームがひしめくマーケットプレイスにおいて、ゲーマーが繰り返し選ぶのは親しみを強く感じるタイトルであるからだと述べています。そして、そのような親しみを覚えるゲームとは、適切なローカリゼーションが施されたゲームであるケースがほとんどです。Statistaによると、現在、世界中のスマートフォン所有者は38億人で、これは世界人口の約48.33%にあたります。それを考えれば、ローカリゼーションはモバイル ゲーム スタジオにとって賢明な投資と言えるでしょう。

「全体的に、カルチャリゼーションとは創造的な活動だと考えています」とキム氏は言います。「カルチャリゼーションは、ユーザーが当社のゲームについて広告や現地でのマーケティング キャンペーンで初めて見聞きしたときから、後の CRM (顧客関係管理) を通じたアウトリーチ活動を実施するまでに至る全体的なサイクルすべてにかかっています。リーチする際にメールと通知のどちらを使用すべきか、睡眠時間中にプッシュ通知を送ったり不適切な言葉を使ったりしないように現地調査を行うなど、すべての要素がカルチャリゼーションの一環なのです」

言語サービスを活用して ウィキペディアを向上

ウィキメディア財団のマシン ラーニングディレクターを務める クリス アルボン氏が語るプロジェクトのミッションと現地データの重要性

「マシン ラーニングが採用されている領域では、(中略) 読者と編集者のエクスペリエンスを向上するためのモデルが使用されています」とアルボン氏は語ります。「たとえば、編集の品質を監視して予測しようとするモデルがあります。私たちがその編集を有害なものと考えるかどうにかかわらず機能するモデルです」

ユーザーが作る世界最大の百科事典「ウィキペディア」には 300 もの言語で 4,000 万件を超える記事が含まれており、毎月 5 億人にもユーザーが利用しています。この有益なリソースをより多くのユーザーが利用できるようにして、さらなる向上を目指す上で、ローカリゼーションサービスを通じて支援することは可能でしょうか？ これはウィキペディア運営団体からの質問で、答えは「もちろん可能」です。

各国語版のウィキペディアの包括性を可能な限り高める上で最も簡単なのは、それらをすべて集めて自動翻訳を行う方法です。たとえば、現時点で最大の記事数を誇る英語版を、利用者の少ない別の言語に翻訳するのです。

ウィキメディア財団はこの手法を限定的に採用しており、同財団が支援するコンテンツ翻訳ツールを使用して記事を機械翻訳し、それを編集者が必要に応じて修正したものを最終的に公開しています。これは、未トレーニングの機械翻訳では問題が生じる可能性があるため、極めて重要なプロセスです。機械翻訳の有名なエラーとしては、英語の「village pump (村のポンプ)」が「bomb the village (村を爆破する)」という意味のポルトガル語に翻訳された例がありますが、「人間参加型」にすることでこのようなエラーを最小限に抑えることができます。この自動翻訳ツールは、これまでに 50 万件以上の記事の翻訳に使用されています。

しかし、ウィキペディアのマシン ラーニング ディレクターを務めるクリス アルボン氏は、AI Business に自動翻訳に全力を尽くすよりも「現地コミュニティの意見を実際に考慮に入れる方がはるかによい」と語っています。

ローカライズされた マシン ラーニング

ウィキペディアには、多くのユーザーが想像するよりも遙かに広い範囲でマシン ラーニング モデルが使用されています。「マシン ラーニングが採用されている領域では (中略)、読者と編集者のエクスペリエンスを向上するためのモデルが使用されています」とアルボン氏は語ります。「たとえば、編集の品質を監視して予測しようとするモデルがあります。私たちがその編集を有害なものと考えようかどうかにかかわらず機能するモデルです」

ウィキペディアでは、さまざまな複雑性に対応する数百の「ボット」を使用しています。これらのボットは記事のトピックの提案や荒らしの取り締まりなど、さまざまなことを実行する目的で開発され、膨大な数の記事の品質維持に貢献しています。しかし、アルボン氏と彼の同僚 (コンテンツの執筆は行わない関係者) は、無料の百科事典としてベストなものを作りたいと考えてはいましたが、構築した AI モデルに自身のバイアスが組み込まれるリスクを冒すことは望みませんでした。

ウィキペディアの場合、バイアスとはどのような意味を持つのでしょうか？ これは、たとえばウィキペディアに必要なものや、優れた記事に対するサンフランシスコ在住の人の考えを、文化的背景が異なるケニア在住のユーザーに適用することはできないということです。

「私が読むウィキペディア記事の中で最も多いのは、もちろん英語版の記事です」とアルボン氏は言います。「このため、優れた記事とはどのようなのか、特定のトピックの記事がどうあるべきか、その記事をどう構成すべきか、その記事でどのようなトーン

(語調) を使用すべきかなど、すべてに関する私の考えは、ウィキペディア プロジェクトという一つのコミュニティにおける私自身の体験に基づいています。それが多種多様なコミュニティすべてに完全に適用できるわけではありません。私たちは全員、ウィキペディア プロジェクトの一員であるという総意の下に行動していますが、良質の編集とは何か、または有害な編集とは何か、また、**特定のトピック**をどう扱うべきかということに関しては、それぞれの地域ごとの『ローカルな問題』になります。このようなことは、現地のコミュニティによって現地の言語で規定されるべきことです」

ウィキペディアの ミッションに組み込まれた ローカリゼーション

現在、ウィキペディアでは現地コミュニティのデータを使って完全にトレーニングされた約 100 個のマシンラーニング モデルが稼働しています。これらは、約 30 種類のウィキに配備されています。世界中のあらゆる人々にそれぞれの言語で無料の百科事典を提供するという目標に言及しながらアルボン氏はこう述べました。「ローカリゼーションが当財団にとって非常に重要であるのは、当財団のプロジェクト全体にローカリゼーションが組み込まれているからです」

各モデルを現地のデータでトレーニングする理由は、優れたウィキペディア記事を構成する要素に関するユーザーの期待が、国や地域によってまったく異なる可能性があるためです。

「これが重要になる例をお話ししましょう。記事の中で重要なことや重要でないこと、リンクすべき重要なこと、またはあまり読まれないトピックに関するスワヒリ族のコミュニティの考えは、当然ながら現地の『コンテキスト』、つまり現地の文化や状況に基づくため、英語版ウィキペディアで想定している考えとは異なります」とアルボン氏は言います。「たとえば、実際に英語版ウィキペディアでは、マタトゥ (民間のマイクロバス) のようなものは**重要とは見なされません**。しかし、スワヒリ語版ウィキペディアでは注目度が高くなります。それは、これが東アフリカ、南アフリカ、および西アフリカの人々



にとって日常生活に欠かせないものだからです。(中略) 私たちは、**自分たちの見方**を押し付けるようなまねはしたくありません。英語版ウィキペディアを使ってモデルを作成し、それをスワヒリ語版ウィキペディアに適用するというのはそういうことです」

ウィキペディアで使用されているマシンラーニングモデルの一つに、記事内の単語を別のページにリンクする必要がある場合に、その旨を提案するように設計されているものがあります。なぜなら、ウィキペディアでは名詞の大多数にそれ自体のページがあるため、全体がハイパーリンクであふれた記事を作成することも実際は可能だからです。しかし、このような雑然としたページでは、読むべき関連リンクから読者の注意がそれてしまいます。このため、ウィキペディアでは「常識」、つまりごく基本的なことで読者がほぼ確実に知っていると思なされる情報にはリンクすべきでないというスタンスを取っています。ただし、これには問題があります。何が常識かは地域によって異なる能性があるのです。このため、このタスクに配備する AI モデルでは、常識の想定にローカリゼーションの手法を採用することが極めて重要なのです。

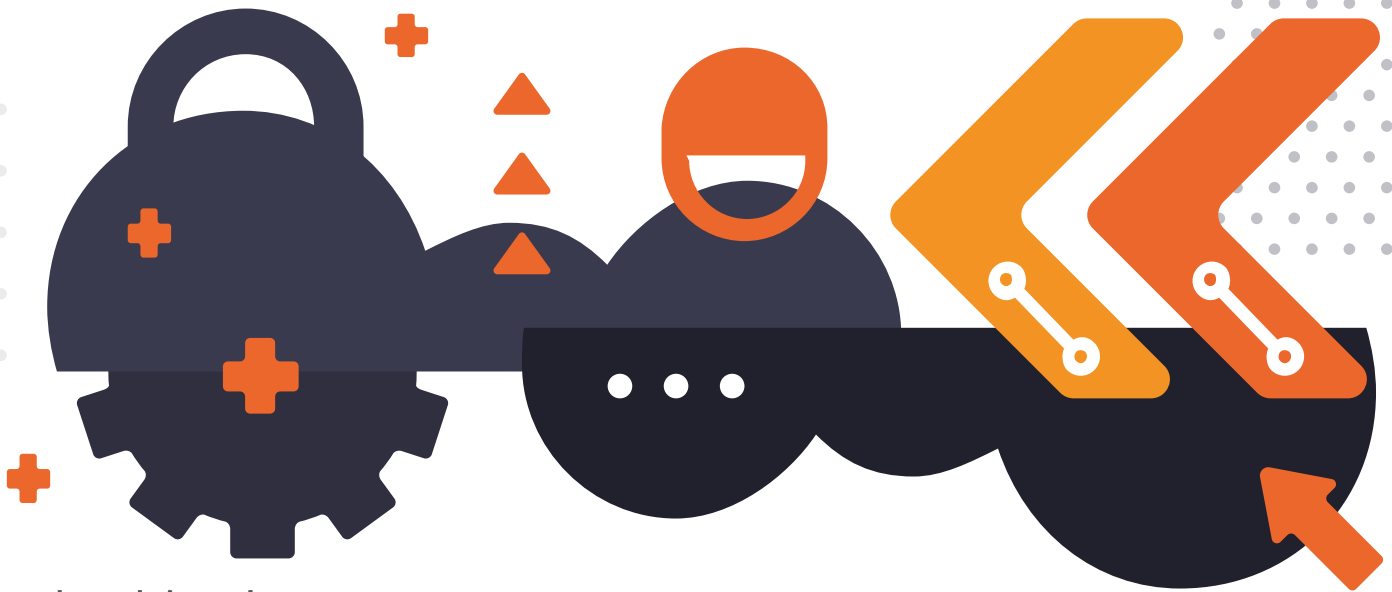
絶対的なルール

「モデル作成における絶対的なルールとして、『マシンラーニングモデルをトレーニングする際には、モデルが利用されるコミュニティのデータを使用する』ということが挙げられます」とアルボン氏は話します。「たとえば、スワヒリ語版ウィキペディア

チームから『編集品質を検出するモデルが欲しい』という要望があった場合、私たちは、優れた記事に関する判断のトレーニング データをそのコミュニティから集めるのです。その後、モデルをトレーニングしてスワヒリ語版で利用できるように上げます」

各国語版のウィキペディアの品質向上により、ユーザーにとって記事がさらに有用なものになり、より多くのユーザーを引き付けることにつながります。新しいユーザーがフィードバックを提供してくれば、現地でトレーニングされたマシンラーニングモデルがさらに賢くなり、ローカリゼーションも進みます。これは、ローカリゼーションにとって唯一無二の「好循環」の手法であり、有益なオンラインリソースがさらに向上されていきます。

「これが、品質向上についてコミュニティの関与が非常に重要である理由です」とアルボン氏は語ります。「そうしなければ、ウィキペディアの在り方について特定の考えを持つ特定の地域出身のグループが、その意思を**他の全員**に押し付けることになるからです。マシンラーニングモデルでバイアスを浄化するという話もありますが、それこそ私たちが望まないことです。コミュニティには自分たちのモデルについて意見を言い、その作成プロセスや適用プロセスにも加わってほしいと考えています。結局のところ、私たちにとって重要なのは、これらの AI モデルを使ってコミュニティのエクスペリエンスを向上することなのです」



転換点

ライオンブリッジのラファ モラルとカジェタン マリノフスキーに聞く、マシン ラーニングと言語との関係

長年の間、翻訳は人工知能の一部でした。AI がその独自の分野を確立する 20 年以上も前の 1933 年、ソビエト連邦の科学者、ピーター トロヤンスキー氏は彼の『machine for the selection and printing of words when translating from one language to another (ある言語から別の言語に翻訳するときの単語の選択と印刷のための機械)』をソビエト連邦科学アカデミーで発表しました。1964 年には、ニューヨーク万国博覧会で IBM が原始的な方法で英語とロシア語間の翻訳ができる機械翻訳ツールを展示しました。

機械翻訳は AI の下に連なる多くの学問分野の下位区分の一つと見られがちですが、ライオンブリッジのイノベーション担当バイス プレジデントを務めるラファ モラルは、機械翻訳がこのテクノロジーの中核となるテーマだと主張しています。「AI は要するに人間の知能の模倣であり、言語ほど人間らしい要素はありません」とモラルは AI Business に語ります。「AI の研究者が人間のように振る舞う何かを作ろうとする場合、それは、概念上、言語を扱うことを示唆しています。これが、AI にとって言語が最初から、おそらくは最も重

要なトピックの一つであったであろう理由だと思えます」

性能向上が著しい機械翻訳

AI に関する他の研究分野と同様に、機械翻訳にもその歴史を通じて多くの変化があり、それらがこの数年間の驚くべき進化につながりました。機械翻訳テクノロジーが有望な R&D の対象から、ついに本格的なツールへとレベルアップしたのです。

画期的な開発の一つが「単語の埋め込み」という概念です。マシン ラーニングとディープ ラーニングのツールセットの一部である「単語の埋め込み」とは、同じような意味を持つ単語や語句は類似した方法で表現できることを意味する単語表現の手法です。単語の埋め込みのプロセスでは必然的に単語や語句の意味をキャプチャしますが、この方法を使って翻訳プロセスを大きく向上することができます。意味が似ていたとしても異なる単語や語句は互いにまったく異なるように表現されていたかつての手法とは明確に異なるものです。

「埋め込み以前は、単語や語句はテキスト文字列として扱われていました」とモラルは言います。「埋め込

みの概念とともにベクトル化の概念がもたらされ、単語や語句を数値に変換できるようになりました。それらの数値には意味論的意味を含めることができます。つまり埋め込は、単語や語句に関連付ける数値またはベクトルを文脈に基づいて特定することです」

ニューラル MT: 明らかな転換点

近年で特に画期的だったこととして、「ニューラル機械翻訳」の登場が挙げられます。モラルはこれを言語サービス業界の「明らかな転換点」と呼んでいます。ニューラル機械翻訳 (NMT) は、以前使用されていたルールベース MT (RbMT) や統計的 MT (SMT) とは異なります。これは、完全に自動化された機械翻訳の手法で、人間の脳に触発された「ニューラルネットワーク」ツールを活用します。これによって、現在のマシン ラーニングの多くが強化されています。NMT では、個々の単語や語句を切り離して分析する代わりに、使用されている単語や語句の文脈を加味することで、より正確な優れた翻訳を提供します。また、NMT では個々の単語や語句だけではなく、それら同士の関係にも注目します。

NMTのようなテクノロジーは、当社のようなローカリゼーションプロバイダー企業が提供するサービスの品質向上に役立っており、ドキュメントや資料などの高品質な翻訳を提供するための強力なツールとして使用されています。

「コンテンツの量は膨大になっており、その量は世界中の翻訳者が処理可能な数の数百倍以上にもなっています」とライオンブリッジのプロダクトリーダー兼ストラテジストであるカジェタン マリノフスキーはAI Business に述べています。「その意味で、機械翻訳によって業界に変革をもたらし、対応可能なコンテンツの量を増やすことができます。こうすることで、人間は自らの付加価値を加えられる仕事により注力できるようになります」

マリノフスキーは、モラルと同じように、NMTのような手法の登場が自動翻訳に極めて大きな影響をもたらしたと言います。しかし、高品質な翻訳を実現する上では、どの手法を採用すべきかということも把握することも重要です。コンテンツのタイプごとに適切な手法が異なるためです。

最高の成果を得るための手法を知る

「コンテンツは目的も種類もさまざまです」とマリノフスキーは言います。「たとえば、利用期間が非常に短く注目度も低いコンテンツがあるとします。このようなコンテンツに超高品質の翻訳は必要ありません。一方で、マーケティングコピーやCM、または非常に特殊な種類の法律、医療、または規制関連のドキュメントといったコンテンツもあり、これらには最高品質の翻訳が必要になります。機械翻訳によって可能になるのは、処理可能なコンテンツ量を増やすことと、より『付加価値の高い』タスクに人間を割り当てることができるようになることです」

たとえば、ライオンブリッジでは翻訳に取り掛かる前に、AIベースのツールを使用して最適な成果を得るための手法を判断します。

「当社には、最適な手法を選び、語彙の密度と多様性、複雑性と変動性、品質などに関連するすべての問題を特定するハイブリッドプロセスがあり



お客様のコンテンツを明確に定義できたら、次にその種のコンテンツに対して最適なプロセス、つまりワークフローを特定します。これには、最適な翻訳メモリ、対象の言語またはドメインに最適な機械翻訳エンジン、特定のコンテンツの種類に適したツールセットが含まれます。

ます」とマリノフスキーは言います。「これが当社のプロセスにおける出発点です。お客様のコンテンツを明確に定義できたら、次にその種のコンテンツに対して最適なプロセス、つまりワークフローを特定します。これには、最適な翻訳メモリ、対象の言語またはドメインに最適な機械翻訳エンジン、特定のコンテンツの種類に適したツールセットが含まれます」

人間参加型

このeBookの5ページにある拡張翻訳のセクションで述べたように、多くの場合は、機械翻訳と翻訳者を組み合わせたアプローチが最適なオプションとなります。

「機械翻訳エンジンの性能が上がったので、注目度が低いタイプのコンテンツもすべて処理することができます。この場合は、ほどほどの品質であれば十分だからです」とマリノフスキーは言います。「次に、もう少し高い品質が求められる場合は、翻訳者を加えてコンテンツを修正します。こうしたプロセスでは、翻訳者はより注意が必要な部分にいつでも集中でき

るようになります。このようなコンテンツの翻訳は読みやすく、適切なストーリー性が維持されている必要があります。対象読者の感情に働きかけるものでなくてはなりません。機械翻訳では『まだ』それができません」

最適な機械翻訳ツールであっても、言語の主観性と、優れた翻訳の構成要素が何であるかについてのあいまいさはまだまだ難しい課題ですが、この分野の進化は現在の翻訳に多大な恩恵をもたらしています。「人間が生成するコンテンツの量は、ほんの1～2年前と比較しても飛躍的に増大しており、さらに増え続けています」とマリノフスキーは言います。「機械翻訳と、当社が所有するAIツールやマシンラーニングツールのおかげで、作成できるコンテンツの量が増え、より多くのオーディエンスと対話し、対象のコンテンツをより適切に処理できるようになっています」

ライオンブリッジが Volvo Cars の全 Web サイト構築を一元的にサポート

Volvo Cars ではかつて 100 あまりの Web サイトを管理・運営

私たちはこの統一感のない外観と印象から脱却したいと考え、一つのブランドと一つのプラットフォームを構築する必要があると考えました。

2030 年までにセールスをすべてデジタル/オンライン化し、自動車をすべて電気自動車に移行することを決定した際に、Volvo Cars は、ビジネスサービスとクラウド ソリューションのプロバイダーである Avanade およびライオンブリッジと連携して必要な変更を加えることにしました。これは、オンライン プレゼンスを全面的に見直し、Web サイトを単なるマーケティング ツールではなくデジタル マーケットプレイスとしてイメージし直す作業が含まれていました。そのためには世界中の潜在顧客に情報を提供し、優れたユーザー エクスペリエンスを構築する必要がありました。

解決すべき課題

全面的に見直しを行う前の Volvo Cars は、Web 上でしっかりとしたプレゼンスを持っていたものの、極めて断片的でした。また、45 の言語で 100 あまりの Web サイトを抱えており、それぞれのページは一つのチームによって作成されていたものの、その翻訳は各市場のチームに委ねられていました。結果的に各市場の Web サイトの構成や販売戦略が異なるものになり、全体的なブランドの統一性が損なわれていました。

Volvo Cars は、Web サイトを集中管理することでそれぞれの翻訳を中央でモニタリングし、極めて重要な作業の担当者には直接連絡することで、ローカリゼーションの取り組みをトップダウンで管理することを希望していました。「オンライン販売に移行するにあたり、当社のメッセージを再構築する必要がありました」と語るのは Volvo Cars でグローバル オンライン デジタル部門の制作責任者を務めるセシリア エンビー氏です。「私たちはこの統一感のない

外観と印象から脱却したいと考え、一つのブランドと一つのプラットフォームを構築する必要があると考えました」

導入したソリューション

Volvo Cars では、翻訳プロセスを最適化するために Smartling 翻訳管理システム (TMS) を採用し、翻訳タスクの管理に費やされていた時間をおよそ 1,000 時間削減できました。この中には、以前はそれぞれの市場のみで使用されていた翻訳メモリの一元化も含まれていました。さらに、翻訳ワークフローを改善し、翻訳されたコンテンツが現地の検索エンジン最適化 (SEO) 要件に合わせて最適化されていることを確認するステップも追加しました。

一元化とローカリゼーション：侮れない力

見直しの結果、同社 Web サイトのデザインはすべての市場で統一されたものになりました。つまり、英国、ドイツ、ベルギーで製品ページを見ている顧客は、言語は異なるものの、すべて同じカスタマー ジャーニーを体験できるようになりました。また、これによって複数の地域でのキャンペーンや製品発表が容易になりました。

「数多くのキャンペーンを同時に開始しました。以前ならば、そんなことは絶対にできなかったでしょう」とエンビー氏は述べています。

このケース スタディは、現地の顧客が求めるメリットを何一つ犠牲にすることなく、ローカリゼーションと一元化で統一感のあるユーザー エクスペリエンスを生み出すことのできた貴重な事例です。

1,000
時間

新システムで実現した
工数削減

言語テクノロジーの未来

ウェアラブル端末、拡張現実 (AR)、膨大な言語モデルの可能性を考察する

昔から機械翻訳の限界について、さも真実かのように語られている話があります。それは初期の MT ツールで英語の文「The spirit is willing, but the flesh is weak (心は熱しているが肉体は弱い)」をロシア語に翻訳すると、「The whiskey is strong, but the meat is rotten (ウイスキーは強いが肉は腐っている)」という結果になったというものです。

機械翻訳と言語テクノロジーは業界の黎明期から大きな発展を遂げてきましたが、言語テクノロジーにはまだまだ発展の余地があります。その信奉者がずっと願ってきたように、機械翻訳が私たちの生活に欠かせない要素になってきているからです。

次世代の言語テクノロジーで最も印象的だったデモのいくつかは、ハードウェアを基盤とするものでした。これらのイノベーションは、急成長するウェアラブル端末テクノロジーにかかわるもので、エンドユーザーが意識する必要すらないほど翻訳がスムーズで当たり前になることを期待させるものです。SFの世界で昔から登場するどんな言語も瞬時に理解できる「万能翻訳機」に少しずつ近づいているのです。

万能翻訳機への道

このテクノロジーがすでに実用化されている例が、イヤホン型の翻訳機です。これは、音声処理と機械翻訳を組み合わせ、話している文章をリアルタイムで翻訳します。よく知られているものでは、Pilot Smart Ear Buds (会話の書き起こしも提供可能)、Google Pixel Buds、WT2 Plus イヤホン型翻訳機があります。

一方、視覚的な面で改良されたのが Google 翻訳の拡張現実 (AR) 機能で、Google Lens と呼ばれるものです。

ユーザーが他の言語の看板やメニューなどのテキストにスマートフォンのカメラを向けると、画面に翻訳が重なるように表示されます。また、多くの場合は元のテキストと同じ書体になります。これはすべてスマートフォンに搭載されたマシンラーニング ツールで実行されます。つまり、Google Lens を使用する際にインターネット接続は必要ないのです。2021年1月、Google は、Google Play アプリ ストアでのスタンドアロンの Google Lens アプリのダウンロードが 5 億回を超えたと発表しました。Google Lens は Google 翻訳が提供する大半の言語をサポートしており、現在その数は 100 を超えています。

Google Lens は先進的だと思われるかもしれませんが、これは第一世代の AR をベースにした翻訳ツールです。Google Lens の初回リリースは 2017 年後半でしたが、すぐに幅広い

支持を得られたわけではありませんでした。しかし、スマート グラスのような人気のヘッドアップ ディスプレイの登場で事態は劇的に変化しました。これらのウェアラブル端末によって、電話やタブレットの小さい画面に閉じ込められていた AR が解放され、端末を装着した人がどこにいてもその周囲にシームレスに溶け込めるようになります。ローカリゼーションの観点で言うと、翻訳により簡単にアクセスできるようになることを意味します。

内部の進化

一般公開されるデモは言語テクノロジーの進化を認識できる印象的なものですが、高品質な翻訳を必要とする人々にとって、本当のイノベーションは、スマート イヤホンやスマート グラスよりも、今だソフトウェアの内部に存在するものとされています。



5 ページの「拡張翻訳とは?」セクションでも説明したように、近い将来の翻訳テクノロジーは、ほぼ確実に AI と翻訳者の両方がかかわるハイブリッド型の拡張手法を利用するものになるでしょう。知的財産、法律、ライフ サイエンス分野に特化した言語サービス プロバイダーである Absolute Scientific Translation (ASTV) の CEO を務めるドメニコ ロンバルディーニ氏は、人間対機械という通俗的な概念化は間違っていると考えています。

「新しいテクノロジーは、人間の関与をある程度抑制または削減できますが、新しいプロフェッショナルの姿やワーキング スタイルを生み出すのも事実です」とロンバルディーニ氏は AI Business に語りました。「昨今、翻訳者から機械翻訳のポスト エディターになる人が増加しています。機械翻訳にはエラーが含まれるため、人間の関与が不可欠だからです。リンギストの育成は優れた翻訳を生み出すための必須条件です」

翻訳品質に最終的に影響を与えるのは人間のままですが、翻訳者のスキルに基づいて特定のプロジェクトに最適な翻訳者を選定する推奨システムの作成などの拡張翻訳タスクでは、AI ツールは非常に有用です。使い方としては、翻訳タスク全体をアルゴリズムに渡すよりも、AI でドキュメントのコンテンツを分析し、キーワードやスタイルの要件を抽出する

ことができます。また、この情報を相互参照する前にテキストのタグ付けと分類も可能です。

次世代の文法・スペルチェッカーのように、翻訳ジョブにおいて機械によるリンギスト支援はますます増えていきます。入力ミスの修正といったシンプルなものかもしれませんが、スタイルの不一致や例外、SEO ポイントにリンギストがよりうまく対応できるよう支援してくれる可能性があります。また、文化的に慎重な対応が求められる問題について、リンギストや関係者に通知を送ることもできます。

大規模な言語モデル

次に、ローカリゼーションの要件に関連する他のタスクに対する AI 言語モデルの使用について考えます。2020 年の最も著しい AI の進化の一つが「GPT-3」の登場でした。これは、OpenAI が開発した巨大な言語モデルです。GPT-3 のニューラル ネットワークは、1,750 億個のパラメータを誇り、モデル自体は 45 テラバイトのデータでトレーニングされています。これは、インターネットで利用できるコンテンツのほとんどの量を占める数値です。このような膨大なデータを取り込むことで、GPT-3 は、ある単語の次に統計的に出現する可能性が高い単語を予測するという点では、たとえば、多くの場合は「orange」の後には「juice」が続くことを認識するなど、驚くほど正確なジョブを遂行できます。

この巨大なコーパスをトレーニング データとして利用できるため、GPT-3 の応用範囲は非常に広がります。法律文を平易な英語の文に、またはその逆に変換するなど、異なる「ボイス」間の切り替えが求められる翻訳タスクにも利用できます。

GPT-3 のようなツールは既存のテキストを翻訳するだけでなく、完全に新しい文章を生成するなど、言語テクノロジーの新たな可能性を切り開いています。「AI の支援があれば、翻訳だけにとどまらず、言語に関してもっといろいろなことができるはずです」とライオンブリッジのプロダクト リーダー兼ストラテジストであるカジェタン マリノフスキーは言います。「コンテンツの執筆にも役立ちます。最新の GPT-3 モデルをテストしてきましたが、複数のユースケースで利用することができました。なかでもコンテンツ生成は最も興味深かったものの一つです」

より洗練され、パーソナライズされたメッセージが期待される業界では、大規模言語モデルのようなイノベーションによって、テキストを国や地域に合わせてローカライズするだけでなく、対象のユーザー一人ひとりに合わせてパーソナライズできる可能性があります。



パリティの問題

実際に機械は人間の言語能力の再現にどれくらい近づいているのか？

2020年の注目すべき論文である「**A Set of Recommendations for Assessing Human-Machine Parity in Language Translation (言語翻訳におけるヒューマン-マシンパリティ評価用推奨セット)**」で、この著者は「この数年間で、機械翻訳の品質は、多数の実証的調査においてプロの翻訳者による翻訳と区別できないレベルにまで向上した」と主張しています。

共著者のうちのチューリッヒ大学のサミュエル ラウブリ氏とダブリンシティ大学のシェイラ カスティリョ氏が機械翻訳の力と可能性についてAI Businessに語りました。

AIB: 論文で、機械翻訳がこの数年で著しく進化したと指摘されていましたが、そのような進化をもたらした主な要因は何でしょうか？

サミュエル ラウブリ氏 (SL): 主な要因は3つあると見ています。リソースの民主化、演算能力の向上、そして商業的な関心の増加です。「ParaCrawl」プロジェクトのような取り組みで、機械翻訳システムの開発に興味を持つ人なら基本的に誰でも利用できる、大量のトレーニングデータが作成されました。機械翻訳に対するニューラルの手法が、統計的手法といった以前の手法よりも概念的にシンプルになったことと、複数の実装を自由に行えるようになったことで、学術目的で、または商用でも機械翻訳システムの構築がかつてないほど容易になりました。

大量のトレーニングデータでニューラルモデルをトレーニングするには多大な演算能力が必要ですが、ほんの数年前と比べて、同じ金額で利用できる演算能力はずっと大きくなったので、このことはGoogle、マイクロソフト、Amazonといった巨大企業以外の組織や個人にとってもはや致命的な問題ではなくなりました。次に、商業的な関心の高まりです。

先行のテクノロジーを使った翻訳処理で、すでに機械翻訳では大幅な時間とコストの削減ができる可能性は示されていましたが、ニューラルモデリングでは一度の処理でかなり自然な文章を生成できるため、意思決定者を含む一般の人々の間でいわゆる「WOW効果」が生まれ、その成果に皆が驚嘆しました。

AIB: 翻訳精度の評価基準として最適なものは何でしょうか？

シェイラ カスティリョ氏 (SC): それは難しい質問です。翻訳の評価方法は数千以上あり、ある手法一つを取っても意見が合わないばかりか、翻訳品質の定義においても多様な意見があるからです。ただ、私にとって、翻訳品質の最適な測定方法とは、求められる翻訳品質を最初に定義することです。その後、それに基づいてプロセスやワークフローの方法論を検討・設計するのです。

SL: 私もシェイラと同意見です。ユーザー向けの翻訳で、ラップトップの問題を解決する内容であれば、プロの翻訳者を投入して翻訳内の言語的なエラーを見つけて分類してもらうのは得策ではないかもしれませんが、ここで一点申し添えたいのですが、どんな方法を採用するにしても、どこかの段階で人を関与させる必要があるということです。翻訳品質の向上で一つ示されたことがあるとすれば、それは**BLEU**のような翻訳品質の自動評価ツールの信頼度が徐々に低下しているということです。

AIB: 翻訳テクノロジーの今後の動きとして最も刺激的で期待できることは何ですか？

SC: 私としては、さまざまな手法の組み合わせすべてに興味を持っています。たとえば、インタラクティブなポストエディットとスピーチです。

SL: 混合イニシアチブの翻訳に加えて、すでに間近に迫っている変化が2つあります。まず、機械翻訳システムはより文脈を考慮できるようになります。つまり、ドキュメントを細切れにして文章に分け、各文を個別に翻訳して後でまとめるのではなく、ドキュメント全体を翻訳するのです。これに関する問題は、本質的には**研究環境**において解決されています。いったん商用化されれば、さらに一貫したものになるでしょう。

次に、機械翻訳システムの出力がより細かく管理できるようになります。機械ではすでに高い品質の翻訳を生成できるようになっていますが、たとえば訳調をユーザー自身の好みのものに調整するといったことはできません。将来的には、**翻訳の丁寧さの度合いを制御**するだけでなく、豊富な語彙の運用やジェンダー化などその他多くの性質も制御できるようになります。

上記のインタビューの内容はわかりやすくするため一部を割愛して編集しています。





AI Business
eBook シリーズ



LIONBRIDGE

ライオンブリッジはさまざまな企業と連携し、障壁を打ち破りあらゆる課題を解決に導くことで、世界中に架け橋を築いています。

25年にわたり、350を超える言語で翻訳とローカリゼーションのソリューションを提供し、お客様の企業が世界中の顧客とつながるための支援をしてきました。当社独自のワールドクラスのプラットフォームを通じて、世界各地で活躍する多数の専門家のネットワークを活用し、さまざまなブランド・企業と提携して文化的に豊かなエクスペリエンスを生み出しています。

言語のプロフェッショナルとして、ライオンブリッジは優秀な人材と洗練されたマシン インテリジェンスの能力を活用し、お客様の顧客の心に響くメッセージをお届けします。米国マサチューセッツ州ウォルサムに本社を置くライオンブリッジは、世界 23 か国にソリューション センターを設けています。

詳しくは、www.lionbridge.com をご覧ください。

詳しくはこちら

