

機械翻訳が広げるグローバルコミュニケーション

Changing Global Communication with Machine Translation



豊橋技術科学大学情報メディア基盤センター センター長・教授

井佐原 均

通商産業省工業技術院電子技術総合研究所、郵政省通信総合研究所、独立行政法人情報通信研究機構を経て、現職。産業日本語研究会世話人会代表。

1 はじめに

本稿は平成 29 年 3 月に開催された第 7 回産業日本語シンポジウムで行った基調講演「人工知能が拓く多言語ビジネスの世界」と、4 月に開催した機械翻訳シンポジウム「AI で変わるグローバルコミュニケーションー身近になった機械翻訳ー」¹で行った基調講演「機械翻訳が広げるグローバルコミュニケーション」をもとに加筆したものであり、個々のニーズに沿った多言語サービス構築の必要性と、そこで必要となる言語データ共有の重要性について述べる。

2 グローバルコミュニケーション

本稿では多言語によるグローバルコミュニケーションを産業界における国際競争力の強化と、観光分野におけるインバウンドの増加の二つの観点から論ずる。

2.1 産業界での国際競争力の強化に向けて

産業界での国際競争力の強化にはノウハウ文書や取扱説明書などの産業文書の精度の高い多言語化を効率よく行うことが必須である。これにより海外拠点などの企業の海外展開の支援や、製品の輸出の増大、知財貿易の増大などが可能となり、波及効果は膨大である。産業に関

1 4月24日(月)東京にて開催。

主催：国立大学法人豊橋技術科学大学

共催：日本マイクロソフト株式会社、株式会社ブロードバンドタワー、株式会社イーアイスクエア

後援：文部科学省

する情報を表現する日本語をこれまで以上に、迅速かつ正確に翻訳することが重要である。

近年の機械翻訳システムの性能は利用できるデータの量に依存している。同種の企業間のデータ共有により対象分野の大量のデータが利用できるようになれば、そのデータを利用して翻訳システムの性能が向上する。これにより機械翻訳を用いた翻訳プロセス全体の完成度を高めることができる。効率良い多言語情報発信が実現され、国際競争力が向上する。

言い換えると、既に翻訳した文書をデータとして活用することにより、新たな文書の翻訳コストを削減できる。さらには既存文書を活用した文書生成と後編集の効率化も可能となる。

日本には企業の垣根を超えて、文書データを共有するという考えが根付いていない。共有による効果を啓もうする必要がある。個々の企業のメリットとしては、翻訳の内製化・半自動化によって生まれる、新たなビジネスチャンスがある。具体的には、時間削減によるビジネスチャンスの増大や、コスト削減によって、これまで訳していなかった大量の文書が翻訳可能になることが挙げられる。製品のライフサイクルが短くなってきた現在、文書翻訳にかかる時間を削減することは製品が競争力を維持できる期間を長くすることに直結し、各製品からの収益の向上が期待できる。また、今まで日本語でしか情報提供してこなかった製品の情報を多言語で提供することにより、海外への販路を広げることが期待できる。

自社文書の提供（公開）が、より大きなビジネスチャンスに結び付く。個々の企業は、その製品の魅力で戦っ

ており、その製品に付属するマニュアルの翻訳の質で戦っているわけではない。非競争領域で協力することにより、競争領域での個々の競争力の強化に集中できるという視点が必要である。

2.2 観光分野でのインバウンドの増加に向けて

海外からの誘客推進に必要なものも産業界と同様に積極的な多言語情報発信である。東京オリンピック・パラリンピックなどで、日本の知名度が向上し、評価・関心が高まったタイミングを活かすことが必要である。

来日後の情報収集についていえば、現地に来てしまえば、何とかするという面はある。対面コミュニケーションでは言語による情報伝達だけでなく、身振りや表情による情報伝達や、周囲の状況による情報理解が大きな役割を果たす。さらには、日本人の持つ「おもてなしの心」が重要であろう。

一方、自国で事前に旅行先を決めるときに、ウェブ等で情報を得たいというニーズが存在する。この時、自分の母語（あるいは英語）で書かれていない情報は存在しないのと同じである。日本人にもあまり知られていない観光地が、どこかの国のブログで紹介され、急にその国の人であふれることがあるという。これはまさに多言語情報発信の重要性を示しているわけである。この場合は「偶然に」紹介された場所が多く訪日観光客を得たことになる。しかし日本にはこのような場所と同等の魅力のある場所は数多くあり、それらからも多言語で情報発信することができれば、同じように海外からの来客を勧誘できよう。もちろん、すべての場所が人手での翻訳を依頼できるわけではなく、機械翻訳などの安価な多言語化手段が必要となる。

海外からの「見える化」は必須であり、これにより各地域の知名度が向上し、海外からの評価・関心を高めることができる。各地域がその魅力によって、観光やビジネスなどの様々な分野で世界から選ばれるようになる。

現地に来てからでも、今、役に立つ情報を得たい、明日の観光に必要な最新情報を得たい、というニーズは存在する。宿泊施設や観光施設の側にもイベントなど最新の情報を提供したいというニーズが存在する。このような情報の多言語での提供には、やはり機械翻訳による低コスト・高品質の多言語化が必須となる。情報の提供手

段としては、スマートフォンやデジタルサイネージを介しての情報提供など、いくつかの媒体が考えられよう。

このような考え方のもと、地域の情報を多言語で発信することを支援する枠組みとして、「地域情報翻訳」を提案する。機械翻訳を活用することによるコストとスピードのメリットを活かし、出来るだけ多くの場所の多様な情報を発信可能としたい。しかしながら現時点では、いくつかの課題が残る。人間の翻訳はコストと時間が課題である。一方、汎用の機械翻訳システムは精度が十分ではない。これらの課題を克服する実用的な機械翻訳システムの開発が必須である。

実用的な多言語機械翻訳システムが実現すれば、低コストかつリアルタイムでの情報発信が可能になる。コストの負担に耐えうる大規模な組織のみではなく、小規模の旅館、神社、土産物店などからも情報発信が可能になり、海外からの観光客の誘致が可能となる。これまで海外市場を想定していなかった中小規模の県産品情報の海外発信も可能となり、対外ビジネスへの展開が可能となる。

3 機械翻訳の精度向上

機械翻訳の精度は、かなり向上したが、まだ完ぺきではない。欧米言語間の翻訳精度は相当に向上したが、日本語に関する翻訳、特に日本語から英語への翻訳には難しい課題が残る。これを解決するには機械翻訳システムの性能向上のみならず、翻訳プロセス全体としての性能の底上げが必要となる。具体的には、翻訳エンジンの改良はもちろん、対訳データの収集と活用、文書作成・後編集の効率化による機械翻訳の実用性の向上が必要である（図1）。機械翻訳による翻訳精度は対象分野の対訳文の量に依存するので、言語データの蓄積も重要である。

機械翻訳エンジンの性能は、ニューラル機械翻訳の

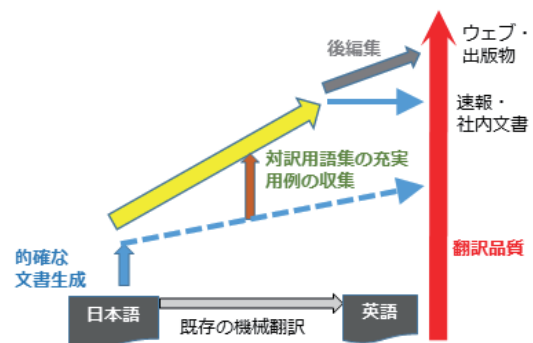


図1 機械翻訳を用いた翻訳プロセスの性能向上

導入により急速に向上した。このシステムは大規模なテキストデータと機械学習に基づくシステムであり、Google や Microsoft によるサービスへの導入により、広く一般に利用可能となった。忠実度に難がある場合もあるが、出力の流暢度は完ぺきに近く、平均的な日本人よりも「翻訳がうまい」と言われている。一方で、訳の抜けや誤りが発生することがある、低頻度の語句については失敗が多い、同じ単語を繰り返し出力することがある、といった問題がある。また、1つの文を1つの文に置き換える、前後の文の情報は使わない、といった、より大きな課題もある。

これらは情報発信には致命的となる場合もある。これを克服して情報発信に活用できる機械翻訳システムを実現するには、まずは個々の分野での学習が必要となろう。個々の分野というのは、例えば「医療」や「自動車」「観光」あるいは「特定の企業」「特定の観光地」を想定する。しかし、分野ごとの文章量は多くはない。これに対応するためには用語の対訳辞書では十分でなく、重要な語句の対訳を抽出し、利用することにより、比較的文章の量が少ない分野の翻訳の精度を向上することを考えている(図2)。低頻度の分野固有の語句に対応することにより、公式文書にもなりうる高品質の翻訳を目指している。

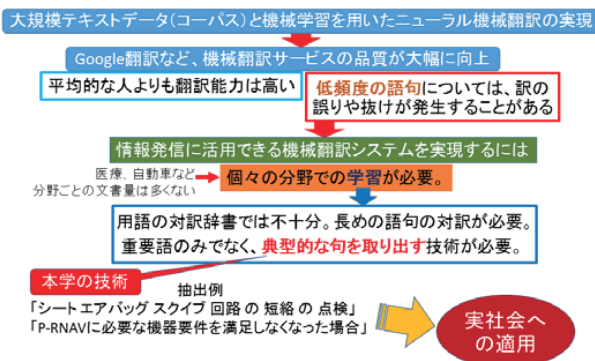


図2 機械翻訳技術の現状と将来

翻訳プロセスでの機械翻訳の活用のためには、翻訳エンジンの改良はもちろん、分野対応や、誤訳・訳抜けへの対応が必須であるが、将来的には相手に応じた情報発信も必要となろう。国や嗜好に応じて翻訳すべき情報を変更する枠組みが研究テーマの一つとなる。文を翻訳するのではなく、伝えたい情報や意図を翻訳するという考え方である。現時点では荒唐無稽かもしれないが、将来の意味翻訳、意図翻訳、文芸翻訳へとつながる考え方の一つと思われる。

4 後編集を介してのコミュニケーション

図1に示したように機械翻訳を含む翻訳プロセスの性能向上には、機械翻訳エンジンの性能向上はもちろん、制限言語による文書作成の定型化、対訳データの充実による翻訳精度の向上が重要である。これらによって、多くの場面で有効な機械翻訳システムが実現するが、それでもなお、印刷して出版する場合など、誤りを許容できない場合があり、後編集が必要となる場合が残る。

我々は集合知後編集 (crowdsourcing post edit) という考え方を示し、後編集を翻訳の専門家によって行うのではなく、その分野や内容をよく知っている人にやってもらうことにより、翻訳専門家と同等の、場合によっては専門家より適切な後編集結果を安価に実現できることを示した。

この枠組みを様々な分野に広げることにより、個人においては後編集に関わるアクティブ・シチズンの増加が見込まれる。すなわち、ボランティアベースのグローバルコミュニティが実現する(図3)。ウィキペディアの拡大が示すように、社会貢献を望むボランティアは多数存在する。社会的活動に参加することにより、参加者のQOLの向上にもつながる。各分野における熱狂的なファンに参加していただくのも一案である。ボランティアベース翻訳は、社会・文化のイノベーションに向けた提案であるが、これは多くの分野に適用可能である。茶道や華道といった日本の伝統文化、アニメ、料理、カメラ、鉄道など、多くのファンを持つ分野が対象となる。

このような活動による国内外の参加者の一体感の醸成は重要なポイントである。従来から、多言語化された文書によって、世界の人々の距離を近づけることが言われているが、さらに多言語化するプロセスを共有することによって、世界の人々の距離を近づけることが出来れば素晴らしい。情報技術による多言語コミュニケーションが世界を繋ぎ、その影響は社会・文化に広がる。ポラン

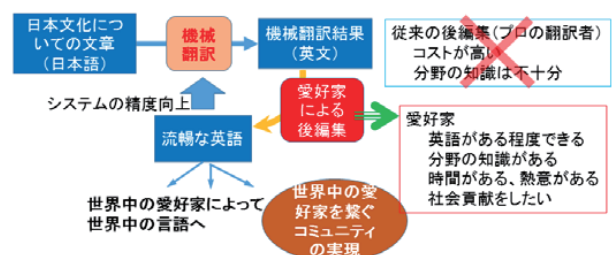


図3 集合知後編集のコミュニティへの適用

ティアベースでの国際コミュニティ構築により、新しいタイプのグローバル社会を実現する。これは科学技術が牽引する新たな社会システムの実現である。

地域の情報を世界に発信することにより、地域のステータスを向上し、地域と世界をつなぐグローバル社会が実現される。これは単に外国人のインバウンドの増加に貢献するだけではなく、国内在住の外国人への母語での情報提供にも適用でき、そのような人々のQOL向上のサポートともなる。

5 データ共有と活用に向けて

機械翻訳をはじめとする自然言語処理の研究開発には研究用の共用データが必要である。AI・機械学習を活用したシステムの開発の重要性が高まっているが、実世界の課題に対応した研究開発に自由に利用できる大規模データはない。実社会のテキストを大規模に蓄積し、システムの開発と実証を行える環境（プレイグラウンド）を整備することが必要である。誰もが実際のデータでサービスを開発することができ、それをビジネスとして実証できることが望ましい。大企業や大学等での実用的研究開発から中小企業や学生によるユニークな研究開発

まで、幅広い研究開発の実施が求められる。このようなデータの活用を通して、事前には想像もできなかったような発想の応用が生まれ、新しいビジネスを日本から発信できることが期待される。

また、地域や分野のデータという考え方が重要となる。地理的に分散（階層化）していたデータが、インターネット社会では相互に直結する。また、コミュニティが相互に直結する社会では、情報の伝達能力の向上による言語障壁の撤廃が重要な課題である。データの作成・編集・蓄積・管理における地域住民の知の活用を通して、QOLの向上を目指すことが必要である。

6 おわりに

本稿で述べた考え方をもとに豊橋技術科学大学ではイノベーション協働研究プロジェクト「多言語情報発信支援の社会実装に関する研究」を実施している（図4）。精度が大幅に向上した機械翻訳システムが実社会に普及していく一助となれば幸いである。

情報技術が拓くグローバルコミュニケーション —IT技術を用いた多言語情報発信プロセスの高度化—

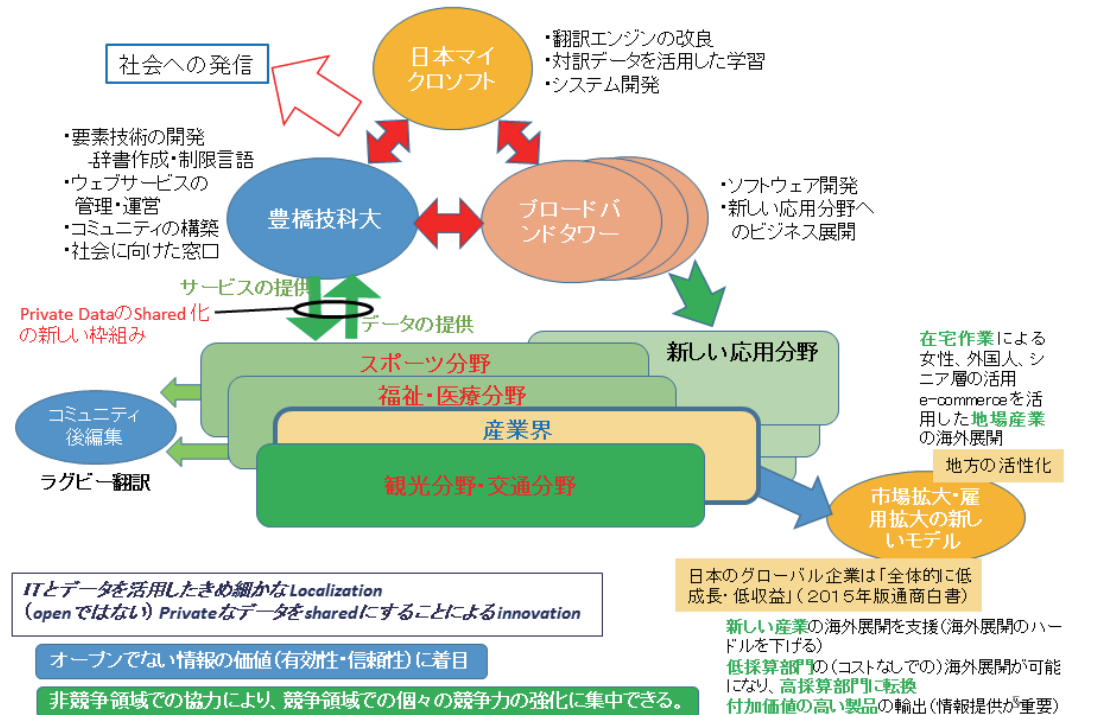


図4 多言語情報発信支援の社会実装に関する研究