

# 多言語情報発信とその効果

Multilingual Information Outbound and Its Effect for Societies



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

先日、出張でマレーシアにある本学のペナン校に行ったときに、タクシーの運転手が状況を説明しようとして、スマホに英語を入力し、日本語に機械翻訳して見せてくれた。翌月に中国に出張した時には、ちょっとした買い物しようとしたら、お店の若い女性店員がものすごいスピードでスマホに中国語を入力して、日本語に機械翻訳して、コミュニケーションをしようと頑張ってくれた。外国人とのコミュニケーションにスマホのテキスト翻訳を使うということが思っていた以上に日常化していることに驚かされた。

マレーシアの場合は、日本語出力も悪くはなかったが、英語が理解できる我々としては英語の原文を見るだけでも十分だったかもしれない。一方、中国の場合は日本語出力はあまり良くなって、半分くらいしか意味が分からなかったが、こちらは中国語が分からないのだから、無いよりは、はるかにましだった。いずれの場合もアプリの翻訳精度は不十分であったが、最近注目されている Neural Machine Translation による機械翻訳の性能向上は驚くべきものであり、今後の翻訳精度の向上が期待できよう。

スマートフォンやパソコンを使って、テキストを入力したり、画面上のテキストを読むということが日常化していることと、機械翻訳の精度向上によって、テキスト翻訳による情報の受発信は急速に社会に普及していくと思われる。

## 2 訪日外国人の増加に必要なもの ～地域の情報の多言語化のすすめ

いわゆるインバウンドの増加のためには、当然のことながら海外に向けての積極的な多言語情報発信が必須である。海外の人に情報提供をして、日本に来たいと思わせなくてはならない。特に東京オリンピック・パラリンピックなどで日本の知名度が向上し、評価・関心が高まっているタイミングを活かすことが重要である。

日本国内の必ずしも有名ではない観光地が急に特定の国からの観光客でいっぱいになるといった話を耳にする。その国のテレビ番組でその観光地が紹介されたり、そこを訪れた人が母国語でブログを書いたりしたことがきっかけになるという。そのようなきっかけが起こった場所に限らず、日本には同じように素晴らしいところが山のようにある。このようなところにもインバウンドの波を届かせるにはどのようにすれば良いだろうか。

海外の人々に向けて、日本の魅力について母国語で情報を提供し、海外旅行先として日本を選んでもらう。さらには、日本の中で今まであまり知られていなかった場所を訪れたいと思ってもらう。訪日経験がある人にも、まだまだ行っていない魅力あふれる場所が日本にあることを知ってもらう。このようなことが必要である。

我々が国内外を問わず、旅行先についての情報を事前に得る場合、インターネットで検索することは当然のように行われる。インターネット社会の現在、ウェブを介しての情報発信は必須である。一方、日本語は多くの人に理解してもらえらるわけではなく、日本語を介さない人にとっては日本語だけで書かれている情報は無いのと同

じである。できれば各国の母国語で、少なくとも英語で情報発信をすることによる日本の魅力の「見える化」が必要である。

海外から見えるようになった日本の魅力によって、主要な観光地に限らず、日本国内の各地域の知名度が向上し、評価・関心が高まり、旅行先として日本が世界から選ばれるようになることを期待したい。一旦、日本に来てもらうことが出来れば、対面のコミュニケーションは、簡単な言葉や身振り、そしておもてなしの心で対応することが可能であろう。

インバウンドの増加に向けて、観光情報をはじめとする地域の情報の多言語化は重要であり、日本を旅行先として選択してもらうために、場所的・質的・量的に出来るだけ多くの情報を発信しておきたい。しかし、人手による翻訳を行うことはコストとスピードの点で困難が伴う。そこで機械翻訳の活用が必須であろう。機械翻訳システムを用いることができれば、観光ガイドブックに載っているような主要な観光地に関する部分的かつ静的な情報だけではなく、小規模の旅館、神社、土産物店などからもリアルタイムの情報発信が可能になる。

とはいえ、現時点では汎用の機械翻訳システムは、特に日本語からの情報発信の観点からは、精度が十分とは言えない。より多くのデータを共有するなどの枠組みが必要である。

### 3 観光から産業全般へ

前項で述べたことは観光に限定されるものではない。同じ枠組みで観光に限らず、他の地域情報の海外発信も可能となろう。国際競争力強化にはウェブ等での地域からの情報発信の多言語化が有効であり、観光客のインバウンドの増加はもちろん、地域の特産品の輸出など、波及効果は膨大である。このためには地域に関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は大量の対訳データを前提としている。関連組織間のデータ共有によって、大量のデータを利用することが可能となれば、その分野での機械翻訳システムの精度は向上する。それによって、効率よい多言語情報発信が実現され、国際競争力が向上する。

このような動きは、地域全体の底上げになるとともに、

関わる個々の組織のメリットにもなる。既に社内外のスタッフによる翻訳を行っている企業においては、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。翻訳に関わる時間削減により、製品の海外投入までの時間が短縮されることによるビジネスチャンスの増大や、翻訳コストの削減によって、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による海外に向けた新たな商品展開が可能となろう。

個々の組織の文書の提供（公開）が、機械翻訳の精度の向上につながり、個々の組織や地域のより大きなビジネスチャンスに結び付く。別の見方をすれば、多くの企業は翻訳の速度や精度を競っているわけではないのだから、企業同士が翻訳という非競争領域では協力しあう（データを提供しあう）ことにより、それぞれの企業が競争領域での個々の競争力の強化に集中できる。

### 4 豊橋技術科学大学の取り組み

既に述べたように、我が国の国際競争力強化に向けて、多言語による情報発信がその重要性を増している。我が国産業の市場拡大や、年々増加傾向にある海外からの来訪者へのサービス向上、そして一層の海外来訪者増のためには、多言語による様々な情報へのアクセスを可能にすることが重要となる。一方で情報通信技術（ICT）がより高度化するにつれ、情報の更新頻度は高くなり、事前に情報を翻訳しておく静的な翻訳では現代の様々なニーズを満たすことは時間だけでなく、コストの面からも現実的ではなくなっている。また、日本は頻繁に自然災害に見舞われる国でもあり、防災・災害に関する情報などにおいては、リアルタイムに多言語に翻訳し、正確な情報を提供できることが重要である。

このような課題に対する解決策の一つとして、豊橋技術科学大学では、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と協力し、機械翻訳システムを中心とする自然言語処理システムによって、品質の高いリアルタイム翻訳サービスを実現し、社会実装することを目指している。三者がそれぞれのテクノロジーを持ち寄り、AI・機械学習による多言語コミュニケーションの実現に向け協働を開始した。

この内容は2016年6月21日に東京で記者説明会を開催して広報した。当日は40名近い参加を得て、日



刊紙6紙への記事掲載や30件を超えるネットニュースへの掲載を得た。共同研究のイメージを図1に示す。我々の取り組みの全体構想を図2に示す。

この協働では、2020年に開催される東京オリンピック・パラリンピック競技大会に向けて大幅な増加が見込まれる海外からの来訪者を対象に、観光情報をはじめ、滞在中に必要な医療、災害など、言語による様々な情報をリアルタイムに提供する翻訳サービスを実現し、

2020年までにインターネット上でこれらのサービスを利用可能にすることを目指している。

三者は、実社会における様々なシーンでのAI・機械学習の活用促進を目的に、機械学習の品質向上に必要な情報の収集やビッグデータの構築を協働で推進する。様々な業種の多くの組織の協力を得ながらビッグデータを構築し、データ収集と分析を行う事で、翻訳サービス品質の向上を図る。また、これらの成果を用いた新

AI・機械学習による多言語コミュニケーションの実現に向け協働  
～2020年東京オリンピック・パラリンピック競技大会における海外からの来訪者を対象にリアルタイム翻訳を提供～

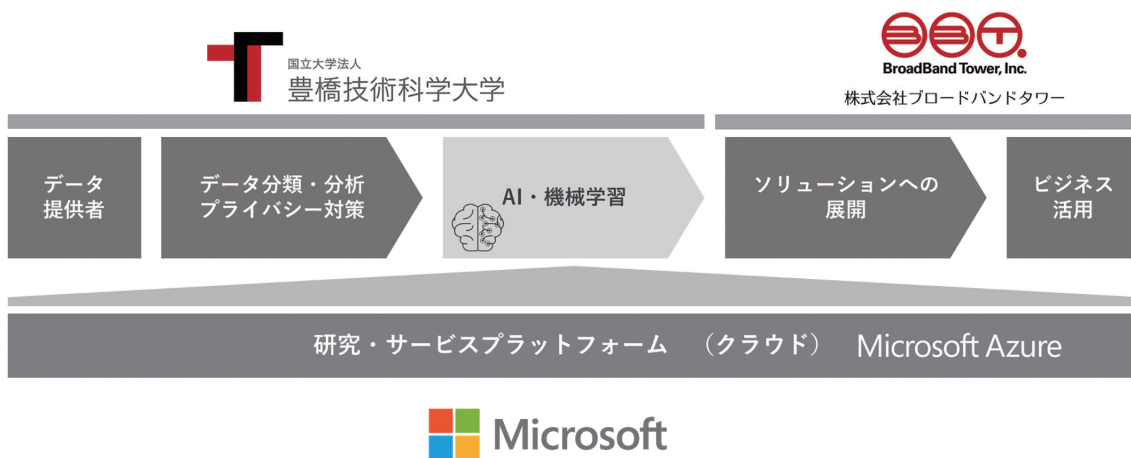


図1 三者共同プロジェクトのイメージ

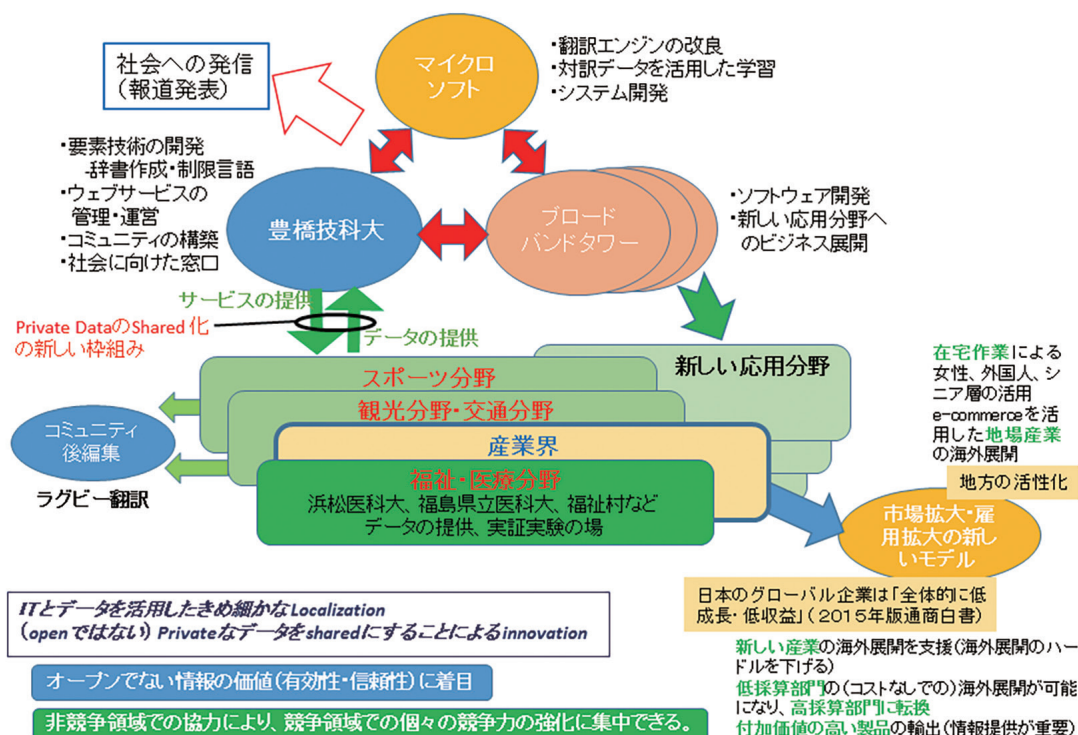


図2 プロジェクトの全体構想

サービスによる新たな社会基盤を構築することで、継続的なビッグデータの構築、AI・機械学習、そしてビジネスへの展開といった取り組みを可能にし、社会インフラにおける幅広い活用を目指している。

本協働の最初の取り組みとして、対訳コーパスデータベースの構築、および収集した対訳コーパスを活用して、高度な機械翻訳サービスを実現するとともに実サービスへの活用を行い、多言語による高品質の情報発信・コミュニケーションの実現に向けて活動を進めている。

## 5 後編集活動を通して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しよう

機械翻訳の利点は新しい情報が書かれたテキストをリアルタイムに翻訳出来ることであるが、出版物のように高品質の翻訳が要求されるなど、機械翻訳の精度が必要なレベルに達していない場合、後編集が行われる。通常、後編集は翻訳者あるいは後編集者が行うが、コストが高くなること、必ずしも分野知識が十分ではない場合があることなど、課題もある。我々はこれに対する解決策の一つとして集合知後編集（Crowdsourcing Post-edit）を提案した。本学の英語ホームページを機械翻訳によって多言語に翻訳できるようにし、得られた訳文を本学の留学生が後編集した。対象とした言語はその留学生の母語であり、また学内組織の役割など、本学に関する知識も持っていることから、プロによる後編集と同レベルの結果が得られることを示した。

この枠組みを広く一般に適用することを考えている。たとえばラグビーに関する文書の翻訳を考えてみよう。2019年にラグビーワールドカップが日本で開催されることから、ラグビーに関する国内外の情報の流通が盛んになると思われる。我が国では以前から大学ラグビーをはじめとして、ラグビーが盛んであったこともあり、シニア層のラグビー経験者が一定量存在する。そのような人達は英語がある程度でき、ラグビーの知識がある。さらにリタイアした方々は、時間があり、社会貢献をしたいという熱意がある方も多い。このような方々に後編集をしてもらうことにより、高品質の訳文が作成可能となる。後編集を受けた訳文と原文をペアにして機械翻訳エンジンを学習させることにより、翻訳システムの精度が上がる。これにより、文あたりの後編集の負担が減少

する。同じ人数で後編集できる機械翻訳出力の文が増加し、対訳の蓄積速度が高まり、得られた対訳がさらに翻訳の精度を向上するといったループが実現できる。

この枠組みはもちろん、ラグビーに限らず、他分野へも、そのまま適用可能である。茶道や華道などの日本の伝統や、アニメ、料理、カメラ、鉄道など、様々な分野での実現が考えられる。さらに多言語への展開を行う場合は、国内のボランティアだけではなく、海外のボランティアによる後編集作業も必要となろう。後編集作業を介して世界中のファン同士の交流が起これ、新しいコミュニティが構成できる。

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とともに、多言語化するプロセスによって、世界の人々の距離を近づけることが出来る。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。これにより、科学技術が牽引する新たな社会システムの実現が可能となる。社会・文化のイノベーションに向けた提案であると考えている。

## 6 おわりに

以上述べてきたように、地域や企業の情報を世界に発信することにより、地域や企業のステータスを向上させ、地域や企業と世界をつなぐグローバル社会の実現を目指している。

企業・産業においては翻訳データの共有によるネットワークの強化と、非競争領域での協力による、競争領域でのグローバル化が実現できる。

個人においては、海外への情報発信に関わる人々が増加し、ボランティアベースのグローバルコミュニティが実現される。情報発信活動に参加しているという実感や、さらには国内外の参加者との一体感が醸成されることにより、参加者の生活の満足度の向上につながる。