

# 我が国の国際競争力強化のための機械翻訳の実活用

Machine Translation for Intensification of the International Competitiveness

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**

**PROFILE**

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 表示にみる翻訳の誤り

英語教育を専門とする旧知のイギリス人の教授が、非英語圏での英語表示の間違いを集めるのが趣味だと話していた。日本に限らず、非英語圏ではとんでもない表示が氾濫しているのだろう。図1は筆者が国内で国際会議を開催した時の昼食会場で撮影したものである。海外からの参加者がほとんどであったために、食事に英語の説明を付けてほしいと依頼しておいたところ、このようになった。「がんもどき」⇒「癌・擬き」⇒“false cancer” というわけである。この表示を見た参加者が、美味しく食べていたかどうかは確認していない。



図1 翻訳の間違い

一方、我々も海外での日本語の表示の間違いを見ると、誰か日本人にチェックしてもらえなかったのだろうかと思うことがある。図2はレイキャビクであった国際会議の会場で撮影したものである。繁体字の中国語と共に日本語が表示されているのは嬉しいけれど、「ようこ（そ）」

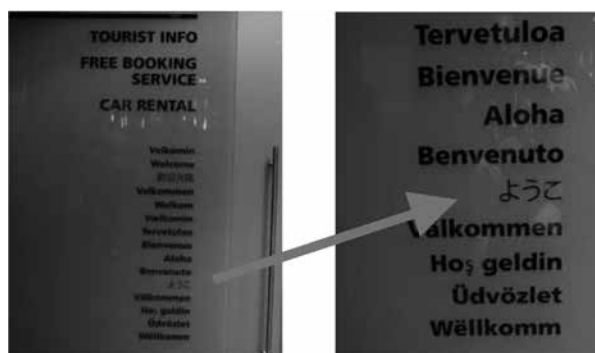


図2 日本語の間違い

とは。惜しい！まあ、実害の無い間違いではある。

オリンピック・パラリンピックの開催も決まり、多くの訪日観光客が予想される中、人的・時間的・金銭的なコストをあまりかけずに、このような表示をなくすことができれば、非常に有益である。

## 2 翻訳と情報の受発信

Globalizationの高まりの中、企業の海外展開、製品輸出、知財貿易など、様々な場面で産業文書の多言語化のニーズが高まっている。また、第2のデジタルデバイドとして、インターネット上に蓄積された大量の英語情報にアクセスできるかどうか、生活の質に直結する環境となりつつある。翻訳プロセスの効率向上は必須であり、機械翻訳等のICT技術の活用が求められる。

我々は外部から情報を受信し、頭の中でその情報を加工し、新たな情報として外部に発信している。ネットワーク社会における情報受信の典型例は、インターネット上の情報を検索し、表示されたページを読んで情報を得る

ことである。情報が自分の理解できない言語で書かれている場合には機械翻訳を用いることも日常的に行われている。一方、情報発信は個人の（あるいは組織の）考えや知見を記述したウェブページを作ることが典型例である。母語の異なる人々へも情報発信をしたい場合は、ウェブページを翻訳することになる。

多言語での情報受信の場合は、読み手が情報を欲しがっており、また必要な情報が手に入れば良い。例えば、ある場所に行きたいと思っていて、そこへの交通手段を調べるといった場面であり、翻訳出力（母語の文）は完ぺきな文である必要はなく、情報を得ることができれば良い。一方、情報発信の場合は、読み手は積極的に情報を欲しがっているとは限らず、読みにくい文は読んでもらえない。発信側は内容に誤解や誤りが無いことを求める。したがって、情報発信の場合には翻訳文は流暢度の点でも忠実度の点でも十分に高品質であることが求められる。

日本での通常の使用を考えた場合、情報受信は他言語から日本語への翻訳を伴い、情報発信は日本語からの多言語化となる。翻訳先の言語としては、まず第一には英語となる。機械翻訳の品質を考えると、欧米言語と日本語とは言語の形態が異なることから、欧米言語間の翻訳と比べて、日本語から、あるいは日本語への翻訳は難しい。特に日本語からの翻訳は、主語が省略されるといった日本語の特徴による解析の難しさもあり、品質が低い。さらには、日本語から英語に機械翻訳をされた文が正しいかどうかを判断して、必要な修正を行うことは、英語を母語としない人間にとっては非常に困難である。納得のできない機械翻訳の出力を見ても、それが間違った出力であるのか、自分の知らない流暢な英語表現であるのかを判断することは難しい。

その一方で情報発信での翻訳には有利な点がある。それは文章を作るのが自分自身であり、また対象とする分野も限定されていることである。情報受信の場合、ウェブ上の既存の文章を機械翻訳に適した文章に変換することは現実的ではないし、様々なトピックの文章が翻訳対象となるので、分野限定も行いづらい。情報発信の場合は、企業が社員に対して文書作成時に何らかの規格に従わせることは可能であろうし、その会社の製品で用いられる用語の訳語を事前に準備しておくことも可能とな

る。

既に欧米では産業界を中心に情報発信のために機械翻訳が実ビジネスで活用されている。これを受けて、日本でも外資系企業等で英日翻訳システムの利用が始まっている。このような実ビジネスの場面では、翻訳システム単体の能力向上ではなく、翻訳過程全体としての効率向上が必要とされる。文書の作成から最終の利用までのトータルサービスでの有効性の評価が重要となる。

### 3 情報発信における翻訳の精度向上

機械翻訳を活用した多言語情報発信（情報提供）の実現のためには、書き手の協力を得て入力文をコントロールすること、対象分野の用語集（辞書）を自動獲得により追加すること、（翻訳者ではないボランティアや留学生などの）対象分野の知識を持つ人による後編集を行うことなどが有効である。これらにより、実用レベルの翻訳（多言語化）を高速に実現できる。また、書き手の自己努力、自動獲得、非専門家の活用により、コスト面でも負担は少なくて済む。

企業内での書き手の協力という点では、二つの場合がある。一つは書き手が執筆を本業としない場合である。たとえば、企業で開発を担当する人が、開発した製品の説明書を書くといった場合である。このような人を対象とする場合には、執筆における制約はあまり複雑なものには出来ない。100ページの規約集を読んで、初めて執筆を開始できるというのでは受け入れられない。このような場合は、質・量の双方で受け入れやすい制約を設定し、可能な範囲の成果（翻訳効率の向上）を目指すことになる。もう一つの場合は、プロの書き手（テクニカルライター）による執筆である。この場合は、文書執筆の専門家であり、かなり複雑かつ詳細な制約でも受け入れ可能であろう。精緻な制約の下、多言語化しやすい（機械翻訳システムにかけて、誤りの出にくい）文書を作成することが可能となる。これにより、機械翻訳システムの入力としての制限言語の効果を最大限に発揮できる。

いずれの場合も、制約に沿った文書作成によって、元の日本語文の読みやすさが損なわれてはならない。そうでないと、国内向けに日本語で書かれた文書と、多言語

化の元原稿として使いやすい、しかし日本人向けのマニュアルとしては成り立たない文書を作ることになりかねない。

対象分野の用語集の自動獲得や、効率の良い後編集の実現も合わせて、筆者らは、多言語化したい文書を持つユーザ企業、翻訳会社と共に、多言語情報発信を支援する環境を構築するために、コンソーシアムを設立することを検討している（図3）。

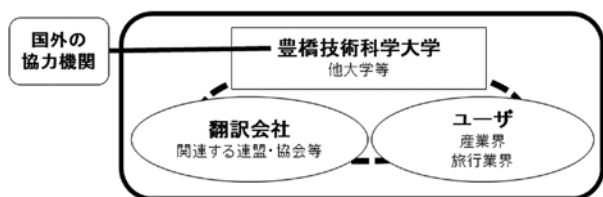


図3 多言語翻訳コンソーシアム

## 4 東京オリンピック・パラリンピックでの機械翻訳の活用

ここまで議論してきた枠組みは、ビジネス文書を対象に我が国産業の国際競争力向上を目指すものであるが、この枠組みはそのまま、来日観光客への情報提供にも利用できる。これは観光立国における我が国の国際競争力の強化につながる。

オリンピックにおいては、多くの外国人観光客が見込まれることから、多言語での情報提供が必要とされる。2008年に開催された北京オリンピックにおいては、組織委員会の公式システムに、筆者が研究代表者を務めた科学技術振興調整費・重要課題解決型研究等の推進「日中・中日言語処理技術の開発研究」の下で開発が行われた、京都大学黒橋研究室の用例翻訳システムが組み込まれた（図4）。

訪日観光客への言語サービスを考えた場合、観光情報

（観光地、文化財、宿泊施設、交通など）の提供、非常時の情報（災害時の緊急情報）の提供、競技の結果や交通情報などの速報、さらには日本に関する知識を得る学習のための情報の提供といった内容が考えられる。そのような情報の媒体としては、ウェブ、標識、文書などがあげられる。たとえば（中小規模のものも含めて）国内の旅館や商店が簡単に多言語情報発信ができる環境が必要とされる。しかしながら、人手による翻訳はコストが高く、機械翻訳技術は完ぺきではない。ここで、機械翻訳を用いた翻訳支援環境の実現が必要となる。

スポーツや観光業界の文書は専門的な表現は多いが、比較的統一されている。このため、競技、観光、宿泊、防災などの案内の表現をコントロールし、さらに語彙を蓄積することで、機械翻訳の精度が一層向上し、訪日観光客に向けた文書やウェブページを英語または観光客の母語で安価で高品質に作成することが可能となる。

## 5 国際標準化の動き

本項で述べた翻訳支援に関連する国際標準化の動きとしては、ISO TC37において、以下の標準化が進められている。制限言語に関しては、SC4において、24620-1 Language resource management—Controlled natural language—Part 1 : Basic concepts and general principlesが標準化に向けて進行中である。辞書に関しては、SC3において、ISO 30042 : 2008 Systems to manage terminology, knowledge and content—TermBase eXchange (TBX) が国際標準化されている。後編集に関しては、SC5において、18587 Translation services—Requirements for machine translation (MT) and post edition levels



図4 北京オリンピックでの翻訳システム

が標準化に向けて進行中である。

## 6 おわりに

ここで述べた機械翻訳の利活用の研究開発は、機械翻訳システム（翻訳エンジン）を開発するものではない。多言語情報発信あるいは翻訳というプロセス全体の効率化を目指して、プロセス中での機械翻訳の有効利用を目指すものである。多言語情報発信を必要とする組織・企業を対象とする問題解決型のアプローチであり、サービス工学の一環である。翻訳対象文書を持つユーザ企業や翻訳会社と協働することによって、具体的なニーズに基づく着実な問題解決を図ることができる。