

第5回自動翻訳シンポジウム 講演 ～自動翻訳の素材を蓄える翻訳バンク～

講演者：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

令和4年3月11日、「2025年に向けたグローバルコミュニケーション技術」と題し、「第5回自動翻訳シンポジウム」がオンラインにより開催されました。シンポジウムでは、翻訳技術の進化により「言葉の壁のない世界が見えてきた」をテーマに、有識者らによる講演や国立研究開発法人情報通信研究機構（NICT）が開発する翻訳エンジンを活用した国内企業（16社）の製品・サービス等のオンライン展示が行われ、NICTフェロー、隅田英一郎氏により、「自動翻訳の素材を蓄える翻訳バンク」と題した講演が行われました。



隅田氏は、「自動翻訳の研究が開始されてから75年となり、現在の技術は一般に使用できるレベルにまで到達した。DeepL翻訳、Google翻訳、ポケトークなどがその例として挙げられ、現在の自動翻訳のレベルはTOEICスコア900程度であり、多くの日本人にとって役に立つものとなっている」と、現在の自動翻訳技術の進化を紹介しました。自動翻訳は、近年、加速的に進化し、特に深層学習を適用してからは一気にそのスピードが上がりました。現在の自動翻訳は、第三世代ニューラル機械翻訳と呼ばれ、非常に精度の高いものになっています。

「深層学習とは入力と出力を真似するシステムであり、模範となるデータの質が高いほど良いシステムとなります。そのため、質の高いデータを得るためにはWEB上のデータだけでなく、信頼できる組織からもデータを得られたほうが良いでしょう。例えば、翻訳データを作った信頼の置ける組織から寄付していただくといったアプローチが良いのではないかと」隅田氏は話しました。

データの量は非常に重要であり、データが少ない分野については自動翻訳の精度も低くなります。Google翻訳でも、ある分野のデータが少ないために精度が低くなったり、またDeepL翻訳では別の分野の精度が低くなるなど、翻訳システムごとに弱い分野が出てきてしまいます。そのため、データ収集を全組織で協力することを目的として翻訳バンクが作られ、発足してから約4年経ちました。当初は「点」的な活動だったものが、現在は「領域」へと広がってきています。

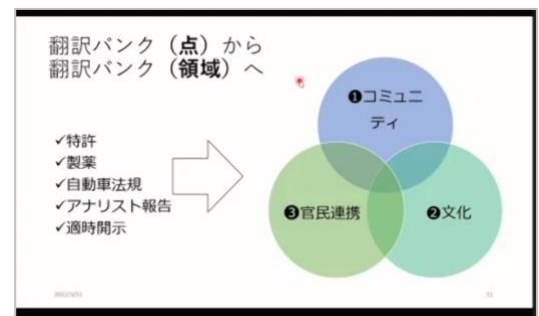
隅田氏はその領域を「コミュニティ」「文化」「官民連携」の3つであると説明しました。

1つ目の「コミュニティ」とはオープンソースコミュニティです。オープンソースソフトウェアである「The Linux

Foundation」や、「Libre Office」は、翻訳者と協力することにより、日々、翻訳例を増やしています。

2つ目は「文化」という非常に大きいエリアに、領域を拡大させることができたことです。茶道の裏千家 大宗匠・千玄室氏が、世界に茶道を広める活動の中で通訳や翻訳を必要としたことが、翻訳バンク支援につながったことを例として紹介しました。

3つ目の「官民連携」は、金融庁など官公庁との協力関係を築いたことです。これまで企業とNICTとは1対1の関係でしたが、こうした協力関係を築いたことで、膨大なデータが翻訳バンクに提供されることになりました。その結果、精度が大幅に向上し、プロの翻訳者レベルの翻訳が2割から5割程度増加した、と評価しました。



講演終了後に視聴者から、「人が普段の会話で主語を省略する「あれ」「それ」などの指示代名詞を、自動翻訳が文脈によって省略された主語を推測し、内容を理解することが可能なのか？」との質問がありました。これに対し隅田氏は「翻訳の事例を学習して主語を補完することは可能であり、1～2年以内に実用可能だと考えている。」と回答しました。また、「専門用語の学習では、業界が一体となる必要があるのか？」という質問には、「業界が一体となったほうが効率は良いが、事例さえあれば学習は可能である」と回答し、自動翻訳技術の進化について、「翻訳者による固有の訳し方の違いも、事例があれば学習できるところまで技術は高まっている」と話し講演を締めくくりました。

(令和4年3月作成)



問い合わせ先

主催：総務省、グローバルコミュニケーション開発推進協議会、
国立研究開発法人情報通信研究機構(NICT)

グローバルコミュニケーション開発推進協議会事務局
(NICT 内)
gcp-info@ml.nict.go.jp
