

機械翻訳で

広がる

知の世界

国立国会図書館長 長尾 真

私は長年機械翻訳の研究をやってきました。機械翻訳というのは、機械つまりコンピュータによって文章を他の言語の文章に翻訳する技術です。たとえば日本語で書かれた文章をコンピュータに与えると、同じ内容の英語の文章が出てくるようにするわけです。

機械翻訳には二、三の方式がありますが、基本的には原文の文法的な構造を明らかにして、それに対応する相手言語の文の構造を作り、そこから相手言語の文を作り出すのです。たとえば「私は本を買いました」という文の主語は「私」であり、「本」は目的語、「買う」は動詞で述語であると分析し、日本語の「主語＋目的語＋動詞」は英語では「主語＋動詞＋目的語」という構造に対応するから「I bought a book」となる、というようにするわけです。

文の構造を明らかにするためには、日本語と英語の文法体系をできるだけ正確に作ることが必要です。そして日本語のある構造が英語のある構造に

応するという、文法構造の対応関係を明確にしなければなりません。しかし言語表現は非常に複雑でまた自由なところがありますから、これらを完全に説明できる文法体系は、日本語でも英語でも作ることは至難のわざです。

文の構造は文法に支えられているとともに、意味によっても支えられています。川端康成が行ったノーベル賞受賞スピーチのタイトルは「美しい日本の私」というものでした。ここで美しいのは「日本」だと思われそうですが、ひょっとして「私」であるのかもしれないと話題になったことがあります。

川端はわざとどちらにもとれるあいまいな表現を企てたのかもしれませんが。

機械翻訳の立場からすると、明晰で一度読んだらすぐわかるといった文章を書くことが望まれます。「美しい着物の娘」だったらどうでしょう。もし娘が美しいことをはつきりさせたいならば、「着物の美しい娘」とか「着物姿の美しい娘」と書くべきでしょう。

長い文を書くのも考えものです。長い文の解析は機械が最も苦手とするところですが、人間にとっても理解しにくかったり、場合によっては間違った解釈をすることがあります。

この直前の文は一見何ら問題のない文に見えますが、「人間にとっても理解しにくかったり、（人間にとっても）場合によっては間違った解釈をするところがあります」という並列の構造だとすると、後半は文法的に少しおかしいものとなっていることがわかります。そこで文を区切って「人間にとっても理解しにくいものです。場合によっては人間も間違った解釈をすることがあ

ります」と主語を挿入すれば明確になるわけです。短い文を書くよう心がけていれば、このような気のつかない誤りは避けることができます。

過去半世紀近くの悪戦苦闘の結果、今日では日本語と英語の間の機械翻訳は、かなり良質の結果を出すところまで来ています。しかしこれも翻訳の対象となる原文が文学的なものでなく、あいまいで複雑な構造をした長い文でないことが必要です。ビジネスで用いる文章や種々の製品の取扱説明書などはこのような条件を満たすものが多いので、機械翻訳がよく用いられています。

日本語と韓国語の間の機械翻訳はかなりよくできます。それは日本語と韓国語の文の構造が非常に近いからだといわれています。文章表現はその背後にある文化や慣習に支えられています。が、日韓両国の文化や慣習には相似性が強くありますので、それも翻訳の質を高めるのに役立っているわけです。同じことがフランス語とイタリア語の

間の翻訳に言えます。両言語ともにラテン語が背景にあるからです。英語とフランス語は成り立ちが少し違いますが、文化的には近い関係にあることもあって、両言語間の機械翻訳の精度も高く、EUで実用しています。

国立国会図書館には膨大な図書・資料の蓄積があり、これは広く国民の利用に供されていますが、さらに世界各国の関心のある人たちにも使ってもらいたいと考えてきました。その手始めに同じ漢字文化圏に属す中国の国家図書館と韓国の国立中央図書館と協力し、機械翻訳を通じて図書・資料の相互利

長尾 真
なが まこと

1936年生まれ。国立国会図書館長。京都大学名誉教授、工学博士。京都大学総長、独立行政法人情報通信研究機構理事長を経て、2007年4月から現職。専門は自然言語処理、画像処理、パターン認識、電子図書館。主な著書に「わかる」とは何か」「電子図書館」「情報を読む力、学問する心」など。

用の道を開きました。日本語のキーワード（検索語）を入れると、これが中国語や韓国語のキーワードに変換されて、それぞれの図書館の図書・資料の検索をします。該当するものが出てくると、それを日本語に翻訳して提示するというシステムです。

日本語と中国語の間の翻訳の質はまあ良くありません。中国語は孤立語に属し、語形変化がないので、その構文解析が非常に難しいからです。しかし、これから何年もかけて解析の質を高めれば実用にもってゆけるでしょう。日本語と英語の間の機械翻訳はまずまずですが、フランス語、ドイツ語、ロシア語、アラビア語、スワヒリ語などとの間の翻訳は将来の課題です。

いずれにしても、このような努力をすることにより、日本語の貴重な資料が広く外国に知られるようにすることにも、それらの国々の貴重な情報を日本でもよく利用できる環境を作ることが大切であると考えています。