



コーパスに魅せられて

小磯花絵 国立国語研究所 副所長



—どのようにして言語の研究をするようになったのですか。

学校の教科では国語が一番苦手で、数学が好きでした。最初に興味を持ったのは、ことばではなく、音なんです。ピアノ、バイオリン、フルート、パーカッションなど、子どものころからいろいろな楽器をやっていたことも関係しているのでしょうか。

私たちが会話をする際、ことばの音声も大きな役割を果たします。例えば、「うんうん」という相づちは時に共感なども示しますし、こうしたことばのリズムが合うと、心地よく感じ話も弾みます。そういうことに興味があったので、大学では初め心理学・認知科学を勉強していました。

言語学の真ん中の道ではなく、心理学、認知科学、情報工学と変わりながら、言語学との境界領域を歩いてきました。専門は？と聞かれたら、今はコーパス言語学と答えています。

—コーパス言語学とは？

コーパスとは、ことばを大量に収集してさまざまな情報を付けて検索できるようにした、ことばのデータベースのことです。コーパスを使った言語研究をコーパス言語学と呼んでいます。どういうことばをいかに集めるか、どういう情報を付けるかというコーパス設計や構築法についての研究も対象となります。

私がコーパスと出会ったのは学生だった1990年代前半です。当時は、British National Corpusという1億語から成るイギリス英語のコーパスが完成した一方、日本語の大規模なコーパスはありませんでした。特に話しことばのコーパスは遅れていて、共有化の動きがようやく始まったところでした。コーパスは、ことばを集めるだけでなく、公開して活用できるようにすることが重要です。学生のころに日本語コーパスの構築と共有化に取り組み始め、国語研に入った今も延々と続けています。

—2021年から副所長に。最近の成果を教えてください。

2016～21年度は国語研の第3期中期計画の期間で、「多様な言語資源に基づく日本語研究」をテーマに掲げてきました。日常会話や方言、日本語学習者のことば、奈良時代から明治・大正時代までカバーするものなど、新たなコーパスを構築したり拡充して研究を進めるとともに、研究・教育・産業など社会

で広く使ってもらえるよう、こうしたコーパスを一般に公開する取り組みを行ってきました。

私は「日本語日常会話コーパス」の構築に携わりました。家族や友達との食事、仕事など、さまざまな場面における日常会話を200時間分集めたもので、映像も公開しています。うなずきや身振り手振りも会話に重要な役割を果たしていますから。この規模で映像付きの日常会話コーパスは、世界に例がない画期的なものです。手前みそながら自慢させてください。

—コーパスは言語研究だけでなく広く使うことができますか。

コーパスを社会で広く使ってもらおう、というのには、2種類あります。1つ目は、コーパスを用いた研究成果が私たちの生活に還元されることです。例えば国語研で構築・公開している「日本語話し言葉コーパス」を学習データとして使った結果、音声認識の精度が飛躍的に上がったのです。コーパスは産業界からも注目されていて、生活に還元される例が増えています。

2つ目は、一般の人にもコーパスを使ってもらおうことです。国語研のウェブサイト「少納言」では「現代日本語書き言葉均衡コーパス」のデータを検索できます。ぜひ、気になることばを検索してみてください。そのことばはブログではよく使われているけれども新聞では使われていない、といったことなどが分かります。インターネットの検索エンジンの結果とは違うものが見えてきて、ことばの選び方の参考にもなると思います。

—今後、どのような研究を進めていこうとお考えですか。

日常会話コーパスを拡充していきます。成人の会話を中心だったので、子どもの会話を増やしたいのです。ことばは、人の成長によっても変わりますよね。どのように変わっていくかを分析し、教育にも役立てたいと思っています。

また、コロナ禍で高齢者のコミュニケーション不足による認知機能の低下が問題になっています。そこで、高齢者の会話をコーパス化し、それに基づく研究を始めました。認知機能を活性化させる会話やことばが明らかになれば、問題の改善に貢献できるかもしれません。コーパスは、私たちの生活に役立つ多くの可能性を秘めていると思います。