



新しいことを！

前川喜久雄 国立国語研究所 所長

—2023年4月、所長に就任されました。まず、これまでどのような研究をされてきたのかを教えてください。

専門は、音声学と言語資源です。音声学は、人が話しているときに、どのように音をつくり出していて、その音がどのような性格を持ち、どのように聞き取られるかを研究します。最近は、医療用のMRI（核磁気共鳴画像）装置で唇や舌や喉などが動く様子を撮影したリアルタイムMRI動画を用いた研究に力を入れています。音声学の研究は学生時代からなので40年以上になり、言語資源は20年ほど前からです。

—言語資源とは？

ことばの研究のために体系的に収集したデータのことです。研究者一人一人が自力でことばを集めて分析するのが、従来のやり方です。しかし一人の力では限りがあります。そこで、ことばを体系的かつ大規模に集め、コンピュータの力を借りて複雑な検索ができるようにしたコーパスがつくられるようになりました。コーパスは典型的な言語資源です。ただし、誰でも使えるように公開されているものでなければ言語資源とは言えない、というのが私の考えです。「日本語話し言葉コーパス」や「現代日本語書き言葉均衡コーパス」などの開発に携わり、最近では「リアルタイムMRI調音運動データベース」を公開しました。

—国語研にはどのようなミッションがあると、お考えですか。

国語研は新しいことをやっていかなければいけません。では新しいこととは何か？20年前は言語資源をつくって公開することが新しいことでした。それを続け、世界的にも充実した巨大なデータが得られています。今やらなければいけない新しいことは、巨大なデータを料理する方法の開拓です。巨大なデータを扱うには、情報科学の力を借りなければなりません。特に必要なのがモデリングの技術です。言語のモデル化は昔から行われていますが、頭の中で考えるものがほとんどでした。そうではなく、データに基づいてモデル化し、シミュレーションを行ってその結果をデータと照合し、モデルの正しさを検証できる。そういうモデリングの技術がないと、量に負けてデータに振り回されてしまいます。巨大なデータに対応できる情報科学と結び付いた、新しい言語学をつくらなければいけないのです。

—そのために必要なことは？

アメリカでは言語学とコンピュータサイエンス両方の学位を持つ人がいて、言語学研究を引っ張っています。日本ではまだ少数です。国語研は2023年春から総合研究大学院大学に日本語言語科学コースを開設し、日本語をデータに基づいて客観的・定量的に分析できる次世代の研究者の養成を目指しています。

—国語研が直面している課題はありますか。その解決策は？

日本の財政が厳しく研究所の運営が難しい中、いかに生き残っていくかを考えなければいけません。それには、研究者に限らず、ことばを使う人々にとって、なくてはならない研究所であることが不可欠です。皆さんは、スマートフォンなどで音声入力を使っているでしょう。音声認識の研究が始まったのは1950年代ですが、使い物にならないとずっと言われていました。そうした中、音声のデータベースに基づいて新しい音声処理・認識の技術を開発することを目指してつくられたのが「日本語話し言葉コーパス」であり、それによって認識精度が向上したのです。また国語研が開発した一連のコーパスを検索できる専門家向けの「中納言」というウェブサイトは、4万人以上のユーザーに年200万回以上利用されています。一般向けに「現代日本語書き言葉均衡コーパス」のデータを公開している「少納言」も、年60万回以上利用していただいています。

私たちの研究が皆さんの生活に直接利便性をもたらすことはなかなか難しいのですが、どういう研究をしているかを伝え、またコーパスを気軽に使っていただくことで、まずは国語研を知っていただきたいと思っています。また、研究の下支えとなる言語資源などの整備を継続することで、なくてはならない研究所だと、研究者から評価してもらえることを目指します。

—所長になることが決まったとき、どう思われましたか。

実は、組織のマネジメントには大学生のときから興味があり、本を読んだりしていました。国語研で室長になったときには、研究室運営のノウハウを教えてもらおうと、国内の研究室をいくつか訪問しました。一番大事なことは何かと尋ねると、答えはみんな同じで、人だと。私も同意します。マネジメントについてこっそり勉強してきたことや、音声学研究や言語資源の開発で実験的な手法や情報科学を取り入れてきた経験と知見を、活かしていければと思っています。