

ふしぎの国のスウガク使い

確率と統計の科学でヒトのことはの謎を解く

内村直之

▶最近のコンピュータは、私たち人間のことをずいぶん理解するようになりました。スマホに「今日の天気は？」と聞けば、天気情報を書いてあるホームページを見て答えてくれます。Googleでわからないことがらを調べれば、「ここに説明してあるでしょう」とばかりにたくさんの文書を提示してくれます。私たちのことをコンピュータで扱うのに、実は確率統計的な考え方がとても役に立っています。今回は、統計数理研究所の持橋大地准教授にことを確率的統計的に扱う最先端の言語研究について、お話を聞きました。「数学でことを？」とちょっとびっくりですが、機械による翻訳や情報探索など、これからの私たちの情報生活について不可欠なものとなりそうです。



持橋大地さん。
統計数理研究所（東京・立川）のロビー壁に刻まれた「数」の字を前にして。

「ことばというものには、子供のときから興味があったようです。幼稚園のときに『宇宙戦艦ヤマト』と漢字で書いて両親をびっくりさせた」と持橋さん。もちろん、高校でも英語と国語は大の得意科目で、当然、大学は東京大学の文系を受験しました。同時に数学も好きで、「大数へ数学」誌の熱心な読者だったといいます。

「言語学の研究をやりたかったんですが、大学へ入って知った文学部の言語学の内容は、〇〇語の音韻規則はどうしたこうしたとか、細かいマニアックな話ばかり。もっと抽象的・一般的な『ことばの問題』の研究に憧れていたもので、こりゃいかんと……」。

そこで決めたのが、抽象的な学問のできる理科に進むことでした。理科から文科へ鞍替えする「文転」は珍しくないけれど、文科から理科への「理転」はめったにいません。一年留年し猛勉強の末に「基礎科学科第二」という理系学問なら何でもやれるところへの進学に成功しました。

ことばの勉強は「センスが勝負」と考えがちだけれど、そうではなく、実はたくさん触れてパターン認識をうまくできるようになることがポイント、と高校時代から考

えていたそうです。「書き換えとか要約とか問題演習をいっぱいしないと英語はできるようにならない。なんか、語学って機械的だな、と思っていた」。それが今の持橋さんの研究の原点のようです。

コンピュータでヒトのことはを理解したい

コンピュータが開発されて、「ヒトのことはわかる電子頭脳」があるといいなあ、と思った研究者は多かったようです。もちろん、コンピュータはコンピュータ用に設計された人工の言語（たとえば今ならC言語とかJavaとかのプログラム用言語）を「理解」することができます。その命令に従って「プログラム」通りに処理を進めます。ヒトのことも同じように処理できるだろうか、と機械翻訳などを含めたいろいろな試みが1950年代からなされました。ヒトがきっかけの文を入力すると、あたかもそれに答えるような「会話ソフト」も作られました。これは、あらかじめたくさん用意された文を相手の入力文に合わせて出力するだけでとても「理解」というわけにはいきませんでした（これは、後に「人工無脳」というあだ名がついたこともあります）。

ヒトのことはには、たくさんの語彙に加え、それを組み合わせて文を作るルール「文法」があります。これにそってきちんと組み立てればヒトのことはをコンピュータで扱えるのではないかと……こういう試み「自然言語処理」の研究が始まったのですが、これはとてつもなく難しかったのです。

大学時代に持橋さんはコンピュータになじみました。2年生のときから、プログラムの相談を受ける指導員を務めていたほどです。ことばへの興味もあって、当然、その眼は自然言語処理に向かったのです。

「ルールを基にして僕たちの使う言語をコンピュータ内で組み立てようとすると、例外があまりにもたくさんあったり、一つの単語を知らないだけで処理が破綻した

