

■宮尾祐介 コンテンツ科学研究系 准教授

言語の根っこには、人間特有の規則性がきつとある

賢い動物であるサルも、人間のような言語を持ってはいません。なぜ人間だけがことばを使いこなせるのかと考えると、やはり脳の構造の中に、人間特有の何らかの規則性を備えているんじゃないだろうか？ 私たちが、あって当たり前だと思って過ごしている言語の"根っこのところ"のルールを知りたいんです。

●リンゴが木から落ちるのを見て、式を思いつくように

どんな言語でも、名詞や動詞があったり、主語や述語があったりと、結構似たような構造があります。すると単語がどういう順で並び、どう組み合わせさせて文全体の意味ができるのか、そこに何らかの規則性があるはずなんです。そこで文法学者の理論をひもとき、いざ式に書き起こそうとするのですが、これまでの理論には矛盾が潜んでいることも多く、コンピュータが理解できるように形式的な規則にまとめようとする、実は大変な難題であることがわかります。

自然言語処理という分野では、最近ではむしろ膨大なデータを利用してコンピュータで統計的に処理し、意味を獲得する方法が広く行われています。ところがデータからすべて自動的に学習させようとしても、それはまだコンピュータにはできないところなんです。自然言語もニュートン物理学と同じように、まずは自然現象のように観察し、やっぱり自分の手で言語の規則性を明らかにしていかなければならないと考えています。

●実証研究としての「深い解析」アプリケーション

採り出した言語の規則性は、プログラムに書いて走らせることで、それが正しいかどうかを検証することができます。私が開発した構文解析器「Enju」では、英文を入力すると英語の構造を解析し、その意味を計算します。たとえばこの他動詞はこの単語を目的語にする確率が高いといった表層的な関係性を算出するのではなく、それぞれの単語の性質とそれらが組み合わさった場合の論理的な構造を取り出して計算する「深い解析」が特徴です。「Enju」を使った分子生物学の論文検索アプリケーション「MEDIE」「Info-PubMed」は、一般の検索エンジンとは異なり、たとえばタンパク質 A とタンパク質 B の相互作用について書かれた論文といったように、知りたい内容をピンポイントに引くことができ、現在広く活用されています。

●構文解析というタテ糸と「意味」というヨコ糸の融合

構文解析を縦糸とすれば、横糸は「意味」でしょう。英語、日本語、中国語を対象に、この2つをいかに融合させるかというのも、私のチャレンジです。しかし「意味」とはそもそも、どんな枠組みで記述したらいいのでしょうか？ 個々の文にはその内容以前に前提なり、主張なり、提案といった何らかの伝えたい基本姿勢のようなものがあります。そのような意味の土台となる枠組みが作れば、現代文の試験によくある「筆者の主張はなんですか？」といった問題に、コンピュータが答えられる日が来ると考えています。

(取材・構成 池谷瑠絵)