

## 会員の声



### 情報科学・工学, 私はこう考える

## 自然言語処理, 私はこう考える

荻野 紫穂†

最近の研究会などの発表をみると、自然言語処理の流れは、大きく二つに分化しているように見える。非常に実用化に近い方向と、実用化よりは言語処理そのものの理論や技術などに重点を置く方向との、二つである。と言ってもこの二分化傾向は、この分野に限ったことでも今に始まったことでもない。自然言語処理は現在、研究している機関の多くが、「言語処理をしよう」という、いわば「掛け声段階」を終え、「いかに実用化するか」という問題に突入している時期にある。そのため、実用化に対する期待が大きくなっているのに、実用も基礎研究も混沌としていた「掛け声段階」に比べ、二分化傾向が際立って見えるのだろう（もっとも、一般に言う「工学系」の学問は、純粋理論に近い少数の分野を除くと、皆、根本は「実用」にあるから、研究方向であっても必ず実用の香りを保っているところが面白い）。

その、いわゆる実用の方向では、実例を使って解析する、用例ベースシステムに関する研究発表が多くなってきた。この方法は、正解付きの用例をベースにできるし、「教えた結果を覚えていて、用例と完全に一致するデータに対して同じ正解を引き出し、似たようなデータに対して似たような正解を導き出す」ことも可能であり、「素人にもできるカスタマイズ」につながる可能性をもつ。従来行われてきた規則ベースの方法は、専門家でなければ規則を書きにくいという側面があるため、簡単なパラメータの設定と辞書登録のほかには、一般使用者がシステム全体の正解率向上に関与する手段はほとんどなかった。が、「正解を教える」という形でデータを蓄積することが正解を導き出すことにつながるなら、正解率向上に関与できる人数は広がることになる。一般使用者が、対象となる言語に関して正否を判断できることは一般的だからだ。個々の使用者が根気良く正解を教えることで、個々の使用環境に合わせてシステムが「学習」していくことが可能になる。現段階の自然言語処理では、知識獲得がボトルネックとも言えるが、この「学習」機能は、裏返すと知識獲得機能でもある。システム自身を書き変えることなく細かい「学習」が可能であり、また知識獲得機能が高いという点が、恐らくは用例ベース隆盛の理由であり、その「学習」機能は、特に実用システムにおいては必須の部分ともなるだろう。

ここで問題なのは「根気良く正解を教える」という部分である。従来、一般商品になってきたアプリケーションソフトウェアは、大部分が「箱売り」である。パラ

メータ設定が必要な場合もあるが、基本的にはスイッチのオン/オフや初期設定の類がほとんどで、システム自体を作り変えようとするのでなければ、カスタマイズに2年も3年もかかるものはまずない。ところが、上に述べたように、一般使用者にもカスタマイズができるようにするために用例ベースを使うとすると、そのカスタマイズの作業は、まず2〜3日では終わらないだろう（解析にある一定の用例ベースを適用することは、作成者側が提供できるだろうが、それだけでは個々の使用環境に合ったカスタマイズ、とは言えない）。使用者は長時間かけてカスタマイズする形態に慣れていないため、「そんなに面倒なことをするなら使わないほうがよい」といった意識をもちやすいのではないだろうか。あまりに変換を覚えないうえに漢字変換にイライラした覚えがある方も多いと思うが、用例ベース作りに費やされる時間はその比ではあるまい。正解を教える「だけ」の手間であっても、何カ月、何年も先の精度向上を目指して作業を続けていくのは、箱売りのソフトウェアを使うのと同じようにはいくまい。

用例ベースをカスタマイズに使うには、このほかにも、たとえばシステムは、使用者ができる以上のことは結局できないのではないかなど、さまざまな困難や問題点があるだろうが、こういった使用者の意識とどのように折り合いをつけて使いやすいものを作っていくか、あるいは、どのようにこの意識を変えていくかなども、一つの大きなポイントがあるだろう。これらをいかに解決するかによって、自然言語処理システム全体のあり方も変わっていくと思われる。

(平成4年4月6日受付)



荻野 紫穂 (正会員)

1963年生。1986年東京女子大学文理学部日本文学科卒業。1988年同大学院文学研究科修士課程修了。同年日本アイ・ビー・エム(株)入社。

東京基礎研究所に勤務。日英機械翻訳の研究に従事。計量国語学会会員。

† 日本アイ・ビー・エム(株)東京基礎研究所