

仮名漢字変換システムの貢献

高本條治 (たかもと じょうじ)

● 仮名漢字変換システムの歴史

初の日本語ワープロ、東芝JW-10が発売されたのは一九七八年九月だった。このワープロ機に搭載されていたのが

「仮名漢字変換システム」であった。キーボードからの入力を解析しながら単語辞書と照合し、漢字仮名交じりの文字列に変換していく日本語入力方式である。

引き続き、各メーカーがワープロ機を次々に発売し、また、パソコン上で利用するワープロソフトも次第に充実してい



く。いずれも日本語入力には仮名漢字変換の方式を採用するのが一般的であった。現在、仮名漢字変換システムは、携帯電話をはじめとする小型通信端末や各種の家電機器などにも組み込まれており、最も主流の日本語入力方式となっている。

仮名漢字変換システムの開発の歴史は一九六〇年代にさかのぼる。九州大学工学部の栗原俊彦氏らによる研究が最初であるとされ、その後、大学や企業の研究や技術者によって研究と開発が進めら

れた。当時は、それ以外のさまざまな日本語入力方式も研究・開発されていたが、どれも仮名漢字変換ほどには普及しなかった。

仮名漢字変換システムは、漢字一文字ごとに変換する最も素朴な方式から、単語（熟語）ごとの変換、文節単位での変換を経て、複数の文節連続を一気に変換する方式へと進化してきた。それに伴って、文節区切り位置をどのようにして決定するか、文節内部の形態解析をどのように行うか、同音異義語候補の中からいかにして優先候補を決めるか、など技術的問題を次々にクリアしていく必要があった。もちろん、そうした処理アルゴリズムの問題だけでなく、変換用の辞書をどのように構成すればよいかという問題もあつた。

● 日本語研究への期待と失望を経て

仮名漢字変換システムが高度な処理を

目指そうとすればするほど、それを実現するには、日本語についての体系的で網羅的な情報が必要とされた。例えば、文節内部の形態解析を正しく行うためには、自立語と付属語との接続規則、付属語相互の連接規則、活用語の活用規則などが精密かつ体系的にまとめられている必要がある。また、文節区切り位置の妥当性を検証したり、変換候補の優先度を評価したりするためには、語句同士の共起関係や係り受け関係について明確かつ包括的に記述された情報が必要となる。

その他にも、どのような範囲から収集した語を、どのような単位で切り分けて、またどのような付加情報とともに変換用辞書に登録すればよいのかという点も頭の痛い問題であった。また、妥当な表記を提供するにあたって、個々の語をどのように表記すべきなのかを判断するためのよりどころを何に求めるかという問題もあった。

仮名漢字変換システムの研究者や開発者は、当然のことながら、日本語学を中心とする言語学の研究成果にまず期待をかけた。しかし、残念ながら、形態論・統語論・意味論・語彙論・文字表記論などの各分野で蓄積されていた成果は、そうした期待に十分に応えるだけの実用性をもっていなかった。十分に体系的でなく、また、包括的でもなかったのである。

そのため、彼らは実用的な処理システムを実現するために、従来の日本語研究の固定観念や常識の枠にこだわることなく、独自の発見的手法によって日本語の記述を試みた。そのことは、仮名漢字変換システムで使われている基本用語の使い方にも表れている。例えば、「単語」「文節」などの用語は概念が拡張されているし、変換用の「辞書」も従来の辞書のイメージとはずいぶん異なるものである。なかでも重要なのは、「品詞」や「活用」の種類を分類するにあたり、処

理アルゴリズムに応じて、大胆な見直しと再整理、さらに精密化を行っていることであろう。

そうした新しい日本語の記述方法やその成果は、当時若手の日本語研究者を大いに刺激した。私もその一人であった。

すでに、機械翻訳や大規模コーパス構築なども含めた自然言語処理研究全般について、工学畑の人たちと言語畑の人たちとの相互啓発や共同研究も加速されていた。産学協同的な動きも進んでいた。そんな中、私も「ATOK監修委員会」の委員となり、すでに十年にわたって変換用辞書の監修作業を中心に、仮名漢字変換システムの開発現場の人々と直接関わってきた。

● ATOK監修委員会で経験したこと
「ATOK（エイトック）」というのは、ワープロソフト「太郎」で知られるジャストシステム社が開発した仮名漢

字変換システムの名称である。一九八三年に「K T I S」という名称で開発され、その後、一九八五年発表の「A T O K 3」以降、A T O Kと改称された。

A T O K監修委員会は、同社デジタル文化研究所所長の小林龍生氏の尽力により、一九九二年に作家・評論家の紀田順一郎氏を座長として発足した。私を含めた日本語学者数名が委員として参加している他、辞典編纂、国語教育、新聞、出版など、各専門分野の人たちが加わっている。

辞書に登録すべき語彙の選定から始まり、所属品詞の決定、表記バリエーションの検討、変換候補の順位づけ、個々の語同士の共起関係の記述など、作業内容は多岐にわたった。そうした作業を進めつつ、常に変換効率の検証を行って、その成績を次の作業にフィードバックさせていく。どの作業をとっても、常に例外がつきまとい、理屈通りにはなかなか

まく運ばない。

例外があるかどうかは一つ一つ丹念に調べるしかない。一般に、処理の効率性と適切性とはトレードオフの関係にあるので、そのバランスをとることも大切である。ルールを優先させると例外が救えず、例外にばかり目を奪われるとバランスを崩す。ある問題を解決するために打った手が、逆に他のところで思わぬ副作用を起こすこともある。全体として高い変換効率を保持するには、日本語の使用実態に対する根気のいる調査が必要であった。

監修委員会では、「仮名漢字変換システムとはいかなる道具なのか」という問題もよく議論された。この道具は、多様な使用実態をそのまま映し出す記述主義に立つべきなのか、それとも、何らかの伝統的基準によって正当性を主張できるような規範主義に立つべきなのか。答えは簡単には出なかった。こうした議論に

は、言語に対するスタンスの差が如実に表れるからだ。具体的には、「ら抜き言葉」をどう扱うか、表記の揺れや誤用をどう扱うか、方言や古典語にはどこまで対応するのか——そうした問題の一つ一つに対して活発な意見を聞かせながらできるだけ見通しの良い基準づくりが進められていった。

仮名漢字変換システムの社会的責任についても盛んに議論を交わした。「ワープロを使うと漢字がやたらに多い読みにくい文章になる」とかつて言われたことがあった。何でも漢字に変換しさえすればよいのではない。逆に、漢字に変換したくても辞書に登録されていないからあきらめるというケースもありうる。語彙、文法、表記などの各面で、仮名漢字変換システムは日本語の運用に制約や影響を与えているのかもしれないのだ。私自身はこうした経験を積み重ねる中で、自分自身のそれまでの日本語に対す

る研究姿勢を大いに反省させられた。仮名漢字変換システムの研究開発現場は、日本語と生々しく向かい合う現場そのものでもあった。

●利用の中で日本語への意識を磨く

しかし、開発に関与する立場にいたなくても、仮名漢字変換システムを利用しながら日本語の構造や運用について考えることはいくらでもできる。

例えば、仮名漢字変換システムへの入力には、通常キーボードが用いられる。しかも、仮名漢字変換を利用する人の大半が、ローマ字入力方式を用いている。キーボードからローマ字で入力するとき、私たちは頭の中で、ラテンアルファベットによる綴り、仮名文字による綴り、漢字仮名交じりによる綴りという三種類の綴りを交互に意識していなくてはならない。そのことはわれわれの表現行為のプロセスに何か影響を及ぼしている

だろうか。

また、仮名漢字変換システムで採用されているローマ字の綴り規則は、従来の訓令式ローマ字やヘボン式ローマ字を部分的に拡張している。例えば、「C」や「X」という組み合わせを小文字の「c」に対応させたり、「Z」を「ん」に対応させたり、「H」を「てい」に対応させたりしている。こうした拡張はなぜ必要だったのだろうか。逆に、訓令式やヘボン式の場合にはなぜ必要でなかったのだろうか。

変換処理の重要な部分を占める形態解析方法としては、「文節最長一致法」や「文節数最小法」といったアルゴリズムが採用されているが、こういうやり方なぜほどほどに成績のよい解析結果が得られるのだろうか。その理由については、少なくとも言語学的にはいまだ実証的に解明されていない。

入力文字列を解析する際、仮名漢字変換システムは、活用語については活用形

を派生させたり、また、接辞・助詞・助動詞・補助的文節などが接続した形式を派生させたりしながら妥当な適合パターンを探す。仮名漢字変換システムが古典語や各地の方言に対応するには、自立語を登録した変換用辞書も必要だが、付属語や活用に関する精緻で体系的な記述も不可欠である。それらを共通語の変換と問題なく併存させるにはどうすればいいのだろうか。

その他、変換用辞書をどのように整備すべきか問題も依然として大きい。実際に日本語入力をするとき、自分自身と仮名漢字変換システムとの間にかなる相互作用が起きているのかを反省的に観察してみよう。そうした自覚的使用を通して、日本語の構造や運用に対する分析的意識を向上させることができるはずである。それもまた仮名漢字変換システムの貢献の一つであると考えたい。

(上越教育大学／日本語学)