

日 本語やその処理を趣味や専門の1つにしてきた。ここでは日本語への思いをいくつか脈絡なく述べよう。

日本語はとても不思議な言語だとよく思う。まずその起源だ。英語・仏語など欧米の言葉はほとんどみな、印欧語族に属し、8000年前の1つの祖語から発展した。日本語はそうした仲間がいない。韓国語は文法や構造が日本語と酷似するのに、語彙の違いが著しい。日本語の起源の謎を、ITを駆使して解けないか。

1500年ほど前に日本語に漢字が入ってきた。しかし今日でも使われ方は混沌としている。呉音と漢音と訓とだけではない。「生」という字の読みを収集した知り合いがいた。658通りとのこと。これでよいのかという思いもあるが、文化だと開き直っても楽しい。

日本語の入力には40年以上の歴史がある。最初はタブレット入力といって机ほどの盤の上の漢字を指して文字を入力した。その後仮名漢字変換技術が確立してタブレット入力は減じた。でもスマホや(今日の)タブレットでは画面を見て文字を押している。先祖返りだ。文字入力鍵盤や日本語入力方式は、効率的なアイデアと、非効率なデファクトスタンダードとの戦いだった。そして残念ながらたいい後者が勝つ。スマホの文字入力もまた同じ轍を踏んでいる。

スマホの文字入力では指を1本しか使わない。指で画面を押す場所を見ないといけない。指を10本活用すべきだ。片手でも指5本。考えるスピードを超える文字入力が必ず実現できるはず。どんな楽器でも指元を見たりしない。頭の中の曲が反射的に指の動きになり、楽器に伝わり、速い指の動きができる。口も(音声認識器への)日本語入力デバイスだが、舌の位置や唇の位置を確認しながらしゃべりはしない。

通常のキーボードでは、一方の手に子音を、他方の手に母音を割り当てると速い。どの言語も、子音と母音が交互に連なる確率が高い。特に日本語はそうだ。

小谷善行 Yoshiyuki KOTANI

東京農工大学

[正会員] kotani@cc.tuat.ac.jp

1977年東京大学工学系大学院博士課程修了、東京農工大学工学部講師。1994年東京農工大学教授(情報工学)。2014年同大定年退職。現在同大学名誉教授・シニアプロフェッサー。(株)テンソルコンサルティング技術顧問。本会フェロー。

右左右左……という動作になり、速く打てる。分かっているのに DVORAK やオアシス鍵盤が QWERTY に勝つことがなかった。故山田尚勇先生の2ストローク漢字入力もある。

自分の日本語入力の工夫としては、昔からローマ字でなく仮名入力を使っている。このほうが速い。また単語に対して2ストローク(仮名2文字)を単語登録している。濁点は使わない。コンピュータはコヒ、将棋はシキ、情報はシホ、人工知能はシチ、アルゴリズムはアコ、開き括弧はカコなど。東京農工大学(情報)なら、トキノコタカカコシホコカと入れる。速い。

20年ほど前、日本語プログラム言語を設計し、論文



[シニアコラム]

IT好き放題



[No.50]

日本語 IT 雑念つれづれ草

を書いた。プログラムは仮名漢字交じり文だ。Prolog 的な論理型言語だが、変数は表面に出てこず、非決定的関数型言語に見える。「孫とは子の子。」というのが孫を定義するプログラムだ。データ「太郎」の子は「太子。」、「太子」の子は「玉子。」があれば、「太郎」の孫。」と問い合わせると「玉子」が出る。助詞の意味を吟味し、それに一致するように用語を設計した。数学的にも美しい系ができあがった。ただこれを普及させ維持するのは困難なことだった。

近年の日本語技術を振り返ると、それにかかわりながらも、その発展には驚く。形態素解析が確立し未知語処理も何とか可能になった。これは巨大コーパスのテキストに内在する情報に基づく方法だ。さらに語義の多義性解消など意味の世界に発展している。ビッグデータへのテキストマイニングとして、実社会に大きい変革が生まれている。

鉄腕アトムの誕生は2003年だが、実際にはあんな日本語をしゃべれるシステムができなかった。しかし近頃では意味理解を伴った質疑ができるシステムに一步一步近づいている感がある。行く末をぜひ確認したいものだ。

(2014年12月28日受付)