

ジャストシステムデジタル文化研究所
JUSTSYSTEM Digital Culture Research Center

概要

本稿では、日本の代表的な仮名漢字変換システム ATOK の歴史に沿いながら、仮名漢字変換技術の変遷を概観する。全体を、黎明期（フレームワーク模索の時代）、規範性追求期（ATOK 監修委員会の時代）、多様性追求期（多様なデバイスへの対応と方言対応の時代）に分けて記述する。記述の順序は一部時間的に前後関係が逆になる場合と重複する場合がある。

黎明期

日本最初の国語普通語辞書『言海』を編者の大槻文彦から贈呈された福沢諭吉は、「寄席の下足札が五十音でいけますか」と、その五十音配列を暗に批判したという。それに対する大槻の慨嘆。「小学でも八や二十年来五十音を教へて居ることに思ひ至らなかつたのであらうか」[1]

日本語普通語辞書を五十音配列にしたのは、当然のことながら『言海』を嚆矢とする。しかし、今では、誰もが疑うことのない「国語辞典は五十音順」という仕組み＝フレームワークが絶対的ではなかった時代があったことに思いを馳せる人は少ない。

この伝でいけば、今では多くの人が疑うことなく漢字仮名交じり文の入力方法として用いている「べた書き文の仮名漢字変換システム」が、当初から絶対的であったわけではない。

ATOK の歴史は、1982 年 10 月にデータショウで発表された CP/M 用日本語処理システム「KTIS」に遡る。市場に投入された製品としては、1983 年 10 月に発売された、NEC の PC-100 にバンドルされた「JS-WORD」の仮名漢字変換部分を嚆矢とする。1979 年に設立されたジャストシステムとしての最初の一般民生用製品である。

ATOK の歴史を振り返る前に、その前史としての 1979 年から 1983 年の前後を、日本語情報処理技術、仮名漢字変換技術という観点から概観してみよう。

1978 年 01 月：「情報交換用漢字符号系 JIS C 6226-1978」

（現在の「7 ビット及び 8 ビットの 2 バイト情報交換用符号化漢字集合 JIS X 0208:1997」の前身）発行

1978 年 06 月：牧野寛・木澤誠「べた書き文の分かち書きと仮名漢字変換」（情報処理学会論文誌）

1978 年 12 月：森健一氏らの開発による日本最初の日本語ワードプロセッサ JW-10 発売

1979 年 03 月：NEC の 8 ビットパソコン PC-8001 発表

1979 年 10 月：森健一・河田勉「かな漢字変換」（情報処理 解説）

1982 年 10 月：NEC の 16 ビットパソコン PC-9801 発売

1983 年 09 月：「情報交換用漢字符号系 JJIS X 0208-1983」（JIS C 6226-1978 の改正版）

上述の森・河田が言及しているように 1967 年に栗原俊彦らによって始められ、主として研究室において、また、特定分野用として連綿として研究開発が続けられてきた仮名漢字変換研究が、この時期に、研究の進展、実装環境双方が整い、一気に大学の研究室や一部の特殊な分野向けの専用システムから、民生分野への適応が始まったことが見て取れる。

ジャストシステムは、このような時代の空気の中で、創業された。

創業者の一人である浮川初子の話によると、パーソナルコンピューター用の仮名漢字変換システム、ワードプロセッサを手がける前のジャストシステムは、JIS C 6226-1978 に準拠した漢字ディスプレイ、漢字プリンターを用いて、オフコン用の専用システムで漢字を表示、印字するための簡単な入出力システムを作っていたという。[2]

こうした背景があった上で、16 ビットの CPU と汎用的な OS（MS-DOS）の登場とほぼ同時に、同社の最初の仮名漢字変換システム「KTIS」が、市場に投入されている。

以後、年表にもあるように、1989 年の ATOK7 まで、毎年のようにバージョンアップが繰り返され、そ

の仮名漢字変換機能は、進化発展していく。

そして、この年表だけからも、仮名漢字変換システムの技術的、市場的発展にとって重要な、いくつかのポイントを読みとることが可能である。

最初に指摘すべきポイントは、仮名漢字変換のプロセスの最初の段階である文節区切り（形態素解析）のアルゴリズムの進展である。

1983年：KTIS 先読み単語・熟語変換

1984年：KTIS2 文節かな漢字変換

1985年：ATOK3 連文節かな漢字変換

1986年：ATOK5 自動変換、複合連文節変換、複合単文節変換、一括変換

特に、ATOK5においては、文節の区切り位置を推論するためのアルゴリズムとして、二文節最長一致法を採用している。

二文節最長一致法は、先に挙げた牧野・木澤の論文からも見て取れるように、1979年時点で、すでにべた書き文を文節ごとに区切っていく方法としては、非常に有効であることが示されていた。しかし、一方で、牧野の「カナ漢字変換」[3]では、べた書き文を自動的に文節に区切って入力する方式だけではなく、単語分かち書き、自立語・付属語分かち書き、文節分かち書き、漢字指定式などの方法が列挙されている。ATOK5において、二文節最長一致法が採用された1986年という時代においては、この方法が、仮名漢字変換方式としては学界における一定の評価を獲得しつつあったと同時に、まだまだ、他の入力方法も検討対象となっていたことが見て取れる。

ジャストシステムの開発担当者の間では、かなり早い段階から二文節最長一致法が注目されてはいたが、CPUのパワーや、メモリー、フロッピーディスクの辞書容量などの制約で、ATOK5以前のシステムに採用することを見合わせていた。

その時点での市販ハードウェアを性能限界ぎりぎりのところで（場合によっては、若干限界を超えると）フルに利用して、市販ソフトウェア製品に実装していくというジャストシステムの開発手法は、ハードウェアの長足の進歩、価格低下を考慮に入れると、結果的には、ソフトウェア製品の市場競争力を高め、市場生命を長くする効果が大きかった。1989年のEMS（拡張メモリー）の採用や1993年のハードディスクを前提とした変換辞書の採用なども、この流れに沿ったものであり、一時的には、ハードウェア市場そのものの拡大にも貢献した。

もう一つのポイントとして、最初期から汎用的なOS環境の上で製品を開発していたことを挙げるのが出来よう。8ビット時代のプロトタイプ製品である「KTIS」からCP/Mを用い、最初の市販製品がMS-DOS環境の上で稼働していたことが、その後のATOKの市場における競争力を高める上で、非常に役立ったと考えられる。1985年のFEP(Front-End Processor)化も、この流れの中で重要な転機である。MS-DOS上でさまざまなアプリケーションから自由に呼び出して利用することが出来るFEPという構造を取ることで、使い慣れた仮名漢字変換システムをさまざまなアプリケーションにおいて共通に利用できる環境が実現したのである。

このことを逆の面から捉えると、仮名漢字変換システムの進展に伴い、仮名漢字変換システム自体が、アプリケーションとは切り離されて、ユーザーインターフェースの善し悪しも含めて、独立に評価される状況となってきた、とも言うことが出来よう。

ジャストシステムは、1987年には、他社のアプリケーションからもATOKをコントロール可能とするために、ATOKのAPI(Application Program Interface)を公開している。

プロセスの独立性を高め、汎用的なOSに対応することにより、ATOKは、一般ユーザーの日本語入力インターフェースを確立することに成功した。

べた書き文からの仮名漢字変換をアプリケーションとは独立したプロセスとして実行する、という現在の仮名漢字変換システムのフレームワークは、おおむねこの時期までに形作られたと考えられる。

ちなみに、この時期は、ATOKのみならず、多くの仮名漢字変換システムが市場での覇を競っていた。参考までに、箭内平成6年に挙げられている調査対象仮名漢字変換システムを列挙する。

ACE((株)大塚商会) AJIP((株)エクセル) ASK((株)アクセス) ATOK((株)ジャストシステム) DANGO((株)スズキ教育ソフト) DFJ((株)デジタルファーム) E1(イースト(株)) EGBridge((株)エルゴソフト) FIXER(シティソフト(株)) JJ((株)リードレックス) Katana((株)サムシンググッド) 松茸((株)管理工学研究所) NECAI(日本電気(株)) OMAC((株)モーリン) VJE(株式会社バックス) WAVE((株)クレオ) WPFEP(エルゴソフト・ワードパーフェクト) WX(エー・アイ・ソフ

ト(株)) [4]

プロセスの独立性が高まり、汎用的な OS に対応することで、一時的に多様な仮名漢字変換システムのある種の繚乱期が形成されていたことが見て取れよう。この後、Microsoft Windows の普及・寡占化と OS に仮名漢字変換システム(入力支援プロセスとしての Input Method)が標準で組み込まれるようになることにより、この繚乱期はある種徒花的な状況として、時代の背景に消えていくことになる。

そして、仮名漢字変換システムの開発の焦点は、同一の枠組みの中で、どのようにして変換精度を上げていくか、ユーザーにとってストレスのない処理を行うか、といった段階に入っていく。

では、精度の高い仮名漢字変換とは、ストレスの少ない仮名漢字変換とは、どういったものなのか。議論の焦点は、単なるエンジニアリングだけの問題ではなく、ユーザーの日本語観、言語能力といった問題に移っていくこととなる。

規範追求期

1992 年 1 月、ジャストシステムは批評家・作家の紀田順一郎を座長に、近藤泰弘、栗岩英雄、矢澤真人、高本條治らの国語学者、教育者をメンバーとする「ATOK 監修委員会」を設置する。この背景には、エンジニアリングという観点からはある程度の枠組みが固まってきた仮名漢字変換システム全般特に変換用辞書に対する世上の不満と要求があった。

例えば、紀田順一郎は、「ATOK 監修委員会」座長に就任する以前「しにか」1990 年 5 月号(大修館書店刊)において下記のように記している。

「OA 文具としてのワープロの現況に見る限り、漢字処理能力にはまだ不満が残る。一つは辞書の貧弱さであり(中略)。辞書の充実度はコスト要因に左右される面もあるが、普及機種では達意の日本語入力に不足を感じるケースが多い。これはいまだに辞書の編者が明らかでないことも関係がある。そこには辞書編纂に必要な編者の人格性(思想や言語生活の体系)が存在せず、言語生活における定見を有しない係員が、かなり恣意的に既成の紙辞書を孫引きしたり抜粋したりするだけという弊害が一向に改まっていないようだ。」[5]

また、箭内の前掲書には、

「電脳辞書の編集責任は必ずしも明確ではない。(中略)しかも電脳辞書の作成には、国語学者たちが参画している気配もない。ほとんどの辞書はコンピュータ関係の技術者が作ったものらしい」[4]との記述もある。

このころ、市井の電脳国語学者箭内敏夫は、各社の仮名漢字変換システムの変換辞書をさまざまな観点から縦横に分析し、その一貫性のなさを批判していた。

さらに、作家の井上ひさしは、「常用漢字しか変換しない辞書を載せた器機、人類学者用の漢字をたくさんたくわえた器機など、ワープロの辞書を切りはなす工夫が必要だ。」と記して、暗に当時のワープロ専用機搭載の仮名漢字変換辞書への不満を表明している。

これらの批判に共通していることは、仮名漢字変換システムがプログラムとしてはある一定の水準に達していながら、その変換用辞書の構築が一貫した方針に基づいてなされたものではなく、事務文書用途への偏りを持ち、かつ、かなり恣意的な語彙選択によってなされていたという当時の状況への批判である。

1989 年、雑誌編集者の経験を経てジャストシステムに入社した筆者が抱いた ATOK に対する不満も、上記各氏の指摘と通底するものであった。

要は、メカニズム、フレームワークとしてはある程度成熟してきた仮名漢字変換システムの変換精度、ユーザビリティを大きく向上させるためには、すでに工学的な手法のみでは限界に達しており、紀田が述べたように「ある人格性(思想や言語生活の体系)と言語生活における定見を持った専門家が恣意性を排除して」仮名漢字変換辞書の編纂に当たる必要があるのではないかと、という考えであった。

ジャストシステムが「ATOK 監修委員会」を設立した理由は、まさに上記につきよう。この成果は、1993 年 4 月発売の、ATOK8 に如実に表れる。

1992 年 1 月から 1992 年 10 月までの間に、アドホックの委員会も含めると都合 8 回の委員会が開かれ、この間に、変換用辞書の徹底的な見直しが行われる。

座長の紀田自らが、百科語彙の選定を担当し、『ブリタニカ国際大百科事典』の全目次項目を通読し、まさに常住坐臥、電車の中においてまでもマーカー片手に語彙の要不要の検討を行ったという。[7]

ATOK 8 には、技術的にも共起情報（「蝉は鳴く」であり「赤ん坊は泣く」といった前後の語彙との関係で同音異義語の判別処理を行うための情報）が取り入れられており、このための共起辞書の開発にも ATOK 監修委員各氏に並々ならぬ協力があつた。

浮川初子、阿望博喜は、この時期（規範追求期）の技術的達成度と課題をまとめた論文[8]で、郡司隆男の「仮名漢字変換は、理論などあるのかと思われる世界を力づくで実用化しているという感じなのだ」[9]という謂いを引用し、「この「力づく」のところに、どれだけ傾注できるかが、実用化を推進できるか否かの分かれ目になっているような気がしてならない」[8]と受けている。その上で、紙幅の多くを「同音語処理と用例数」、「同音語処理と表記法」、「表記法と辞書」、「弊害対策」、「変換の洗練、日本語データの整備」といった項目に割いている。このことから、この時期の仮名漢字変換システムの製品レベルでの実装開発作業の多くの時間が、膨大な日本語事象を収集吟味し、多くの例外処理を含むルールベースの記述と膨大な辞書の構築に充てられていたことがうかがえる。

そうした作業過程の中で、当然のことながら、自然言語一般が持つあいまいさや、文書の目的や個々人の個性による表現のゆれをふまえた上で、製品の初期出荷仕様を決定することが非常に重要な要件として認識されるようになっていた。

先の浮川・阿望の論文においても「記述性を求める仮名漢字変換において、そのもととなる辞書に、ある種の規範性を求めてきた」[8]との記述がある。パーソナルコンピューターによる日本語文書生産の拡大に伴った仮名漢字変換システムの普及により、社会全体が仮名漢字変換システムを提供する側に対して、そのシステムが変換表記する日本語の規範性に対して、ある種の社会的責務を持つことを求める状況を招いた。

実際、多くのユーザーが ATOK や MS-IMS がデフォルトで変換する日本語をそのまま無批判に用いるようになり、その結果として、伝統的、一般的には稀であるような同音類義語の使用例が、Web サイトの調査において頻発するような事態も発生している。

こうした中で、ジャストシステムが ATOK15 で導入した「話し言葉関西モード」は、仮名漢字変換システムに対してユーザーが求める規範性はそれとして受け止めながら、多様な言語表現への対応を打ち出すことによって、ユーザーに対しても自らの言語表現に対して自覚的であることを求めた、一歩踏み込んだメッセージが込められたもの、ということが出来よう。

多様性追求期

2002 年 2 月、ATOK15 において、ジャストシステムは「話し言葉関西モード」を搭載し、いわゆる方言対応の第一歩を踏み出した。引き続き、2003 年には、「話し言葉東北モード」と「話し言葉九州モード」を追加、さらに 2004 年以降も多様な方言への展開が計画されている。また、ATOK16 では、研究者を中心として要望が多かった「文語モード」も追加された。

これらの動きは、多くの日本人に共通して受け入れられる規範性を持ったツールとしての ATOK から、さらに多くの日本人がその多様な言語文化的背景やさまざまな使用状況において、より自由な自己表現を可能とするツールへの脱皮もしくは飛躍を意味している。多くの人が満足する一つの ATOK から、あらゆる人々のためのそれぞれの ATOK へ、といったスローガンで置き換えても良いかもしれない。

一方、1993 年の Macintosh 対応を皮切りに、ATOK は Windows、Solaris、hp-ux など、さまざまなハードウェア、O.S. 対応している。さらに、最近では、PDA 用の O.S. や携帯電話、セットトップボックス、ゲーム機に至るまで、さまざまなハードウェアデバイスで ATOK が使用できる環境が整いつつある。

これらの多様なプラットフォーム、ハードウェアへの対応は、ある意味で、現今のデジタル装置における日本語利用の拡がり、と捉えることも可能であろう。携帯電話の電子メール機能でやりとりされる言葉もまた、日本語であることには変わりがない。そして、そこから、新しい日本語の使われ方、立ち現れ方が生まれてくる。

また、2001 年 2 月発売の ATOK14 以降、さまざまな専門辞書がオプションとして用意されるようになり、

利用機会、利用目的によって、仮名漢字変換辞書を切り替えながら利用することが出来るようになっている。

PDA 用 ATOK や携帯電話用 ATOK には、べた書き文変換とともに、推測変換と呼ばれる新たな変換方式が併用されている。インクリメンタルサーチと同様の方法で、途中までの入力を参照して、類似の入力履歴を引き出してくるにより、入力効率をアップさせようという試みである。この方式は、もともと PC 用の ATOK で先に開発されたものであるが、限られた入力デバイスによって、限定された目的のための定型的情報を入力する際に特に有効であるため、PDA や携帯電話などへの適応によってさらなる展開が期待されている。

パーソナルコンピューターによる日本語文書入力需要の増大とともに発展してきた ATOK のべた書き文変換、二文節最長一致法というフレームワークもしくはパラダイムが、携帯電話を初めとするさまざまなデジタル機器による多様な利用状況の進展とともに、新たな多様なフレームワークへと展開する兆しが現れつつある。21 世紀の ATOK の歴史は、仮名漢字変換機能の歴史としてのみならず、多様な日本語入力、日本語処理の歴史の中で、さらに相対化して語られるべきものとなるろう。

謝辞・参考文献

謝辞：

浮川和宣氏、浮川初子氏を初めとする、ジャストシステムで ATOK 開発に係わってこられた方々、紀田順一郎氏を初めとする歴代 ATOK 監修委員会委員諸氏、米国サンマイクロシステムズの樋浦秀樹氏、マイクロソフトプロダクトディベロプメントリミテッドの阿南康宏氏からは、仮名漢字変換技術に係わる当事者にしか知り得ない情報を多く提供していただいた。浦城恒雄氏を初めとする「情報処理」編集関係者、bit 編集部、月刊アスキー編集部には、過去に掲載された仮名漢字変換機能に係わる論文に目を通す便を図っていただいた。

参考文献：

- 1) 高田宏：言葉の海へ，新潮社，東京(1978)
- 2) 篠原一：電脳日本語論，月刊アスキー，Vol.24，No.7 (2000)
- 3) 牧野寛：「カナ漢字変換，ワープロと日本語処理」，bit 別冊，共立出版(1985)
- 4) 箭内敏夫：電脳辞書の国語学，おうふう(1994)
- 5) 紀田順一郎：「コンピュータと漢字文化」，しにか 1990 年 5 月号、p9ff
- 6) 井上ひさし：「ワープロは日本語を変えたか」，文藝春秋 1992 年 3 月号 p337
- 7) 紀田順一郎：「日本語の歴史と現実に根ざして 初期 ATOK 委員会の思い出」，ATOK 監修委員会 10 周年記念シンポジウム [ケータイ言葉から方言まで デジタル時代の日本語] ジャストシステム (非売品) (2001)
- 8) 浮川初子・阿望博喜：「仮名漢字変換の実用化技術と今後の課題」，人工知能学会誌，Vol.11 No.6 (1996)
- 9) 郡司孝夫：二円と煮えん，言語，Vol.17，No.1 (1988)