



漢字・日本語処理技術の発展： 仮名漢字変換技術

小林 龍生

ジャストシステムデジタル文化研究所

tlk@kobaysh.com

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。また、仮名漢字変換技術も含め、日本語処理技術は、特に初期においては、パーソナルコンピュータ用ワードプロセッサソフトウェアに比して、ワープロ専用機がかなり先行していた時期があるが、この点に関しては別稿で森健一氏が論じられると思われるので、本稿では深く触れることをしない。

黎明期

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市場向け製品である。

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1月：「情報交換用漢字符号系 JIS C 6226-1978」（現在の「7ビットおよび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6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）

年月	事項	備考
1979年7月	ジャストシステム創業	
1981年	最初のパソコン用日本語処理システムの開発に着手	
1982年10月	8ビットマイコンの標準OS「CP/M」用日本語処理システム「KTIS」をデータショウで発表	
1983年10月	[KTIS] （「JS-WORD」に搭載）	ATOKの前身。KTISとはKana-kanji Transfer Input Systemの頭文字。先読み単語・熟語変換方式を採用した。 「JS-WORD」はNECのPC-100にバンドルされたジャストシステム開発、ASCII販売のワープロソフト。
1984年5月	[KTIS2] （「JS-WORD Ver.2」に搭載）	文節仮名漢字変換システムを採用。
1985年2月	[ATOK3] （「jx-WORD 太郎」に搭載）	初代ATOK - ATOK。ATOKは「エイトック」と読む。Advanced Technology Of Kana-kanji transferの頭文字。 連文節仮名漢字変換システム採用。
1985年8月	[ATOK4] （「一太郎」に搭載）	MS-DOS上の他ソフトに組み込んで使えるようにFEP化。一太郎での日本語入力環境を他ソフトでも実現。
1986年5月	[ATOK5] （「一太郎 Ver.2」に搭載）	自動変換、複合連文節変換、複合単文節変換、一括変換の4つの変換方式を採用。
1987年6月	[ATOK6] （「一太郎 Ver.3」に搭載）	
1989年4月	[ATOK7] （「一太郎 Ver.4」「一太郎 dash」に搭載）	EMS（拡張メモリ）に対応。環境に応じた辞書（S、M、L）プログラムを提供。

表 -1 黎明期の ATOK



- 1978年9月：森健一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3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9月：「情報交換用漢字符号系 JIS 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、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、一気に一般企業でのビジネス利用や個人利用の分野へと適応が広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1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¹⁾。

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」は、市場に投入されている。

以後、表-1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 文節仮名漢字変換

1985年：ATOK3 連文節仮名漢字変換

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」²⁾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いたが、CPUのパワーや、メモリ、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ころで)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。ちなみに、現在のATOKは、二文節最長一致法の発展形であるN文節最長一致法をベースにさまざまな解析手法が加味された変換アルゴリズムが用いられている。

もう1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ことできよう。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年のデバイスドライバ化(後に一般にFEP: Front-End Processorと呼ばれるようになる)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できるデバイスドライバという構造をとることにより、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



1992年1月	ATOK 監修委員会発足	辞書編纂から変換候補の順位付け、AI 変換用例など、日本語をコンピュータで扱うためのさまざまなテーマを時代を見据えながら議論し、製品に反映していくための組織として発足。ATOK8 の AI 用例は、この委員会の議論が反映されている。
1993年4月	[ATOK8] 〔「一太郎 Ver.5」に搭載〕	AI 変換（共起）機能搭載。
1995年1月	[ATOK9] 〔「一太郎 Ver.6.3」に搭載〕	ローマ字入力時のアルファベットの入力ミスや、仮名入力時の濁点・半濁点などの入力ミスを自動修正する自動入力補正機能「JAC」を搭載。
1996年9月	[ATOK10] 〔「一太郎 7」に搭載〕	慣用句や仮名遣いの誤り、ら抜き表現などを入力変換時にリアルタイムで校正する「JUST MEDDLER」を搭載。
1997年2月	[ATOK11] 〔「一太郎 8」に搭載〕	変換確定した単語の意味情報を元に、現在入力中の話題（分野）を抽出し、話題にあった最適な同音語を使い分ける文脈解析変換を搭載。
1998年9月	[ATOK12] 〔「一太郎 9」に搭載〕	日本語変換の基本性能を見直し、「SuperATOK エンジン」を新たに開発。文脈解析処理をさらに強化し、文中の「それ」や「これ」の指す意味を理解し、より適切な変換候補の提示を可能とした。
1999年11月	[ATOK13] 〔「一太郎 10」に搭載〕	漢字や記号のマウスによる手書き入力可能な「手書き文字入力」、文脈から判断が難しい同音語の使い分けを変換候補窓に表示する「同音語選択支援」機能を搭載。さらに、変換履歴データを参照し、入力する単語を予想してポップアップ提示する「推測変換機能」、指定したファイルやフォルダ、クリップボードから未登録語や AI 用例を抽出してユーザー辞書に自動的に学習させる「AI 辞書トレーナ」機能等により、変換効率・速度が向上。
2001年2月	[ATOK14] 〔「一太郎 11」に搭載〕	また、「ことば」を入力するだけで目的のホームページにアクセスできる「ATOK Direct」、サーバ上に設定されたさまざまな辞書やコンテンツを参照できる「ATOK Channel」、インターネットに個人の辞書を保存して、異なるパソコンからも常に自分の ATOK 環境が利用可能な「ATOK Sync」の 3 つから構成される「iATOK」により、インターネットとの融合をさらに強化。

表-2 規範追求期の ATOK

ケーションとは切り離されて、ユーザインタ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、とも言うことができよう。

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、筋内³⁾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)ボックス), 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))³⁾

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。

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。

たとえば、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。1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ろう。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」⁴⁾



また、筋内の前掲書には、
「**「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」³⁾

との記述もある。このころ、市井の電脳国語学者筋内敏夫は、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。

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るながら、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。

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に如実に現れる。

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片手に語彙の要不要の検討を行ったという⁵⁾。

ATOK8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⁶⁾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」⁷⁾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どうか、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」⁶⁾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、膨大な日本語事象の収集吟味、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、文書の目的や個人個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」⁶⁾の記述がある。パーソナルコンピュータ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ついて、ある種の社会的責務を持つことを求める状況を招いた。

④ 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。2003年以降もATOKにおいては、さらに多様な方言への展開が計画されている。

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1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よいかもしれない。

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、OSに対応している。さらに、最近では、PDA用のOS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



2002年2月	[ATOK15] (「一太郎12」に搭載)	多彩な日本語表現への対応第一弾として「話し言葉関西モード」を搭載。「ATOK LAN Extension」は、LAN環境内で公開された他のパソコンの辞書を利用して、社内の専門用語や社員名簿、商品リスト等が共有でき、ATOKからデータベースへのアクセスが可能となる。また、言葉で関数を入力できる「Excel関数入力支援」なども搭載、単なる日本語変換を超えた機能を付加。
2003年～		多様な方言への対応、古文への対応などを計画。

表-3 多様性追求期の ATOK

1993年11月	ATOK8 for Macintosh
1994年4月	ATOK8 for Windows
1994年	ATOK8 for Solaris
1994年	ATOK8 for hp-ux
1998年10月	Voice ATOK
1999年4月	ATOK Pocket for WindowsCE
1999年7月	ATOK12 SE for Linux
2000年9月	ATOK X for Linux
2000年9月	ATOK Pocket for Palm OS (ATOK for Palm OS)
2001年3月	ATOK Pocket for Mobile Phone (ATOK for Mobile Phone)
2001年10月	ATOK for SONY Airborad
2002年4月	ATOK for PS2

表-4 ATOK の環境, OS への対応

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換えながら利用することができるようになってきている。

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すことにより、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な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(とその発展型)というフレームワークもしくはパラダイムが、携帯電話を始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

となろう。

謝辞 浮川和宣氏、浮川初子氏を始めとする、ジャストシステムでATOK開発にかかわってこられた方々、紀田順一郎氏を始めとする歴代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かかわる当事者にしか知り得ない多くの情報を提供していただいた。浦城恒雄氏を始めとする「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かかわる論文や記事に目を通す便を図っていただいた。

参考文献

- 1) 篠原 一: 電脳日本語論, 月刊アスキー, Vol.24, No.7 (2000).
- 2) 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).
- 3) 箭内敏夫: 電脳辞書の国語学, おうふう (1994).
- 4) 紀田順一郎: コンピューターと漢字, 月刊しにか, Vol.1, No.2 (1990), p9ff.
- 5) 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).
- 6) 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).
- 7) 郡司隆男: 二円と煮えん, 言語, Vol.17, No.1 (1988).

(平成14年9月10日受付)

