布上俊一**

漢字情報処理システムの動向

Trends in Kanji Information Processing System

漢字情報処理は、従来の大形機による特定アプリケーション分野の適用から、中・小形機やオフィスコンピュータをも含む一般的な適用へと、対象範囲が一挙に広まった。これは、メモリ技術の発展を背景としたプリンタ技術の革新に負うところが大きいが、関連するシステム技術、ソフトウェア技術も見のがせない。問題の多い入力方法には、仮名入力により漢字に変換する技術も実用化された。

本稿では,漢字情報処理の現状と主要技術課題について概説し,併せて今後の発 展方向についても言及する。 木村 温* Atsu Kimura

松岡 濶*** Hiroshi Matsuoka

Toshikazu Nunokami

11 緒言

我が国でコンピュータが使用されるようになってから,既 に20年を越える年月が経過している。当初は、コンピュータ とのかかわりをもつ人は、コンピュータの専門技術者に限ら れていたが、最近ではコンピュータ技術にはなじみの薄い人 にまで、利用層が広がってきている。また、コンピュータの 利用法も, 当初は数値処理に限定されていたが, 文章, 図形, 画像などの非数値処理にまで拡大してきている。このため, コンピュータ処理で英数字処理から漢字, 図表を含む日本 語処理にまで機能を拡張することが必要となっており、また この方向は,今後ますます拡大していくものと考えられる。 日立製作所では、まず、大形機による汎用システムとして、 昭和52年11月「日立漢字情報処理システム」を発表した1)。 次に, 低価格化, システムの操作性の向上に重点をおいて技 術開発を行ない,中・小形機,オフィスコンピュータ,各種 端末装置でも漢字を扱えるようにし、昭和55年1月、拡張さ れた「日立漢字情報処理システム"KEIS"*1)」として発表し た。また、最近ではオフィスオートメーションの一環として、 文書処理の機械化のニーズが高まってきている。より高次の オフィスオートメーションを実現するには, 文書処理とデー タ処理との結合が必要であり、漢字・日本語処理は、この面 からも重要な技術である。これらの点を踏まえ,以下に日本 語情報処理の展開の背景とKEISの考え方を述べ、次に、漢 字・日本語処理に関する基本技術, ハードウェア技術及びソ フトウェア技術の現状と今後の動向について概説する。その 後に、漢字処理の標準化の動向と将来の発展方向について併 せて述べる。

② 漢字・日本語情報処理の動向と日立漢字情報処理システム"KEIS"の考え方

LSI技術の進展により、産業用、家庭用など、あらゆる面での利用機器のインテリジェント化が進められている。コンピュータそのものをとってみても、大形機の高性能化、小形・低価格化はもちろんのこと、オフィスコンピュータ、更にはパーソナルコンピュータの普及により、だれでもが利用できるコンピュータへと変わりつつある。同時に、コンピュータ

* 1) KEIS: Kanji Processing Extended Information System

周辺機器のインテリジェント化,高機能・低価格化も図られ, だれでもが利用するための使いやすさの追究が進められてきている²⁾。この傾向は今後もいっそう進展するものと予想される。

この使いやすさを支える一つの大きな要素として、漢字・日本語処理技術がある。オフィスオートメーションに関連して、この技術を用いた漢字入出力端末、オフィスコンピュータ、日本語ワードプロセッサなどが市場に出始めている。将来は、文書処理を中心とした文字記号、図形、画像、音声など多数の情報媒体の統合的データ処理分野に展開するものと子想され、パターン情報処理、人工知能の研究成果がそこに採り入れられてゆくものと考えられる。また同時に、光伝送を中心とした通信・交換技術も研究開発されており、多種情報媒体を含んだ分散処理が進むものと考えられる。

日立製作所では、このような背景を踏まえて、日立漢字情報処理システム"KEIS"を開発し、「日本人のための日本語による情報処理」の実現を図ろうと考えている。その考え方を次に示す。

- (1) 小形システムから大形システムまで、漢字処理共通のア ーキテクチャによる支援
 - (a) KEISコード(日立標準漢字コード) JISコード準拠
 - (b) テーブルウェア
- (2) 従来の英数字・仮名文字システムを共存可能とするシステムの拡張
 - (a) 対話処理, バッチ処理, オンラインリアルタイム処理, 分散処理などの各種処理形態の支援
 - (b) 漢字システムでも従来のソフトウェア, ファイルが扱 える。
- (3) 多様なシステムの構成を可能とする漢字機器の整備
 - (a) 低価格漢字VDT(Video Data Terminal)
 - (b) 低価格漢字プリンタ
 - (c) 漢字入力装置
 - (d) 日本語ワードプロセッシング機能付き分散処理用各種 装置
- (4) OS(オペレーティングシステム), 言語プロセッサからアプリケーションプログラム(情報検索, エンドユーザー言語, ワードプロセッシングを含む文書編集など)に至るまで, 一貫した思想に基づくソフトウェアの支援
- (5) 開発支援システムに対する日本語導入によるシステム建

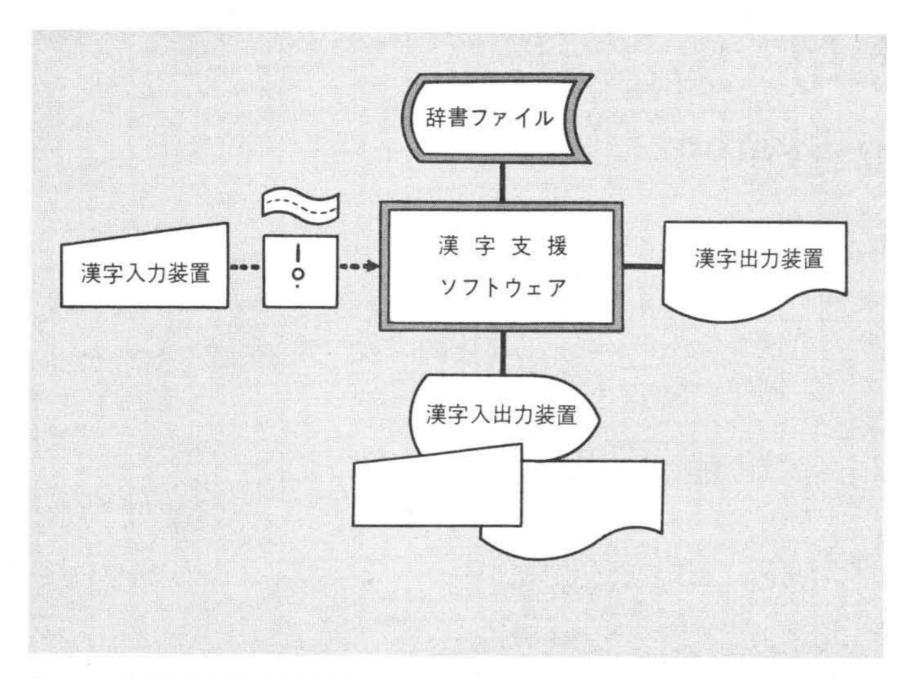


図 | 漢字情報処理技術 漢字を扱うコンピュータシステムは,漢字入力技術,出力技術及び処理技術に大別される。

設の効率化

以上の考え方に則して、以下の章で述べるようにKEISの開発を進めている。

3 技術の現状

漢字を扱うコンピュータシステムを技術的に大別すると,

- (1) 漢字情報をコンピュータで扱えるように、コード化する技術。すなわち、漢字入力技術。
- (2) 漢字を含む情報をファイルし、各種の変換、分類、更新など論理操作を行なう狭義の処理技術。
- (3) 処理結果を編集し、漢字まじりの情報として、印刷・表示を行なう漢字出力技術。

の三つのサブ技術になる(図1)。(2)については、現状ではハードウェアは、処理装置本体やファイル装置が主体で、漢字用の特別な装置はないが、主にソフトウェアに漢字・日本語処理技術開発が行なわれている。(1)及び(3)には、漢字用の各種ハードウェア技術、ソフトウェア技術が開発、実用化さ

表 | 漢字の特性と技術課題 漢字がもつ種々の特性は,漢字情報処理を複雑にし、多くの技術課題を与える。

特	性	技術課題	解 決 方 法
1.字種が非常に多く, 字形 も複雑。		●漢字の入力方	●入力漢字数を限定し,不足
		法	分は外字処理。
		●コード化に13	● 2 バイトコードの採用
		~14ビット必要	LSIメモリ, フロッピーディ
		●文字発生機構	スク, 小形固定ディスクの採
		に大容量のメモ	用によるコストパーフォーマ
		リが必要	ンスの改善
		●高精度の印刷・	レーザビーム、インクジェ
		表示技術	ットプリンタなど、新印刷技
			術の開発。
2.複数の読み方がある。 (同音異字,同訓異字)		●仮名漢字変換	●表示選択, 構文解析, 意味
		時の漢字の特定	解析, 頻度学習など。
		技術	
3.順序づけ(配列)が複雑。 (部首別画数順,総画数順, 代表読み順,国語辞典順 など。		●用途に応じた	●辞書ファイルに文字属性(部
		配列順とするこ	首, 画数, 読み方など) を記録。
		と。	●漢字ソートプログラム
.複数の字形	をもつものが	●コード化方法	●標準的な扱いの決定
ある。	-	●文字種の増加	
(異体字)			

れた。また、(1)~(3)を通じて、相互に密接に関連する基本技術も多く、「漢字情報処理は総合技術である。」とも言われる。

3.1 基本技術

3.1.1 漢字の特性と情報処理

漢字情報処理システムは、日本語表記に使われる漢字そのものの特性に大きく左右される。技術的に問題となる主な特性と、その解決方法の現状を表1に示す。文字種が非常に多く、1個1個の字形も複雑な漢字の基本的特性は、対象文字数、字種選定など、装置の設計に限らずシステムの設計にも大きな課題となり、文字選定方法そのものが、一つの技術分野として存在する。現在は、昭和53年1月に制定された「JIS漢字符号系3」の文字セットを基準に設計される場合が多い。同音異字など読み方が複数ある特性は、仮名による読み入力だけでは漢字を特定できないことになり、入力時に装置側から候補漢字を提示して再選定したり、構文解析などの言語処理的手法を採り入れたソフトウェアの支援が必要となる。漢字の配列順序に画数順や読み順など多数の方法がある特性は、漢字によるソートを複雑なものにし、また入力時の索字方法を困難なものにしている。

3.1.2 文字セットとコード

コンピュータで扱う漢字セット(漢字集合)は、当用漢字表や文章に使用される漢字の使用頻度調査などに基づき、昭和53年1月に制定された「JIS C6226情報交換用漢字符号系」が標準になっている。これは第1水準漢字集合2,965字と第2水準漢字集合3,384字の合計6,349字の漢字集合で、文字配列順は、第1水準は代表音訓順による五十音配列、第2水準は部首別画数順配列である。国立国語研究所の調査によると4)~6)、新聞3種からサンプリングした延べ約99万字の異なり文字数は3,213字、雑誌5部門90種からサンプリングした延べ約28万字の異なり文字数は3,323字と報告されている。使用頻度累積率では、500字で75~80%、1,000字で90%を超える。

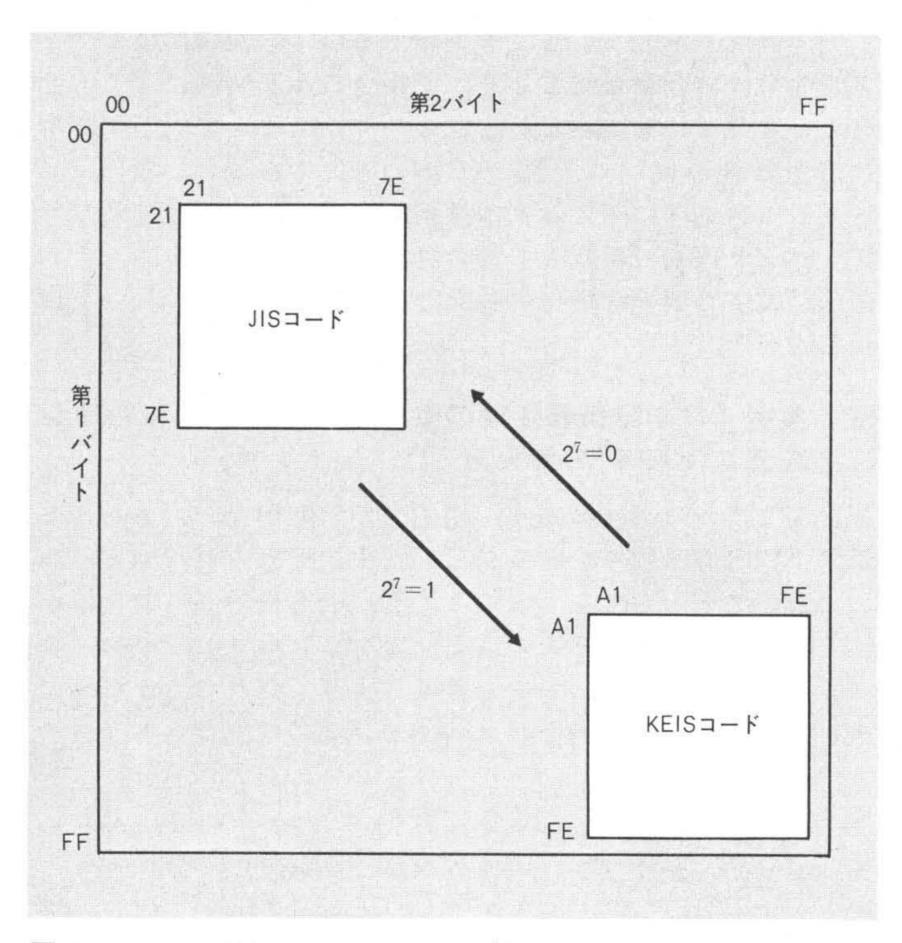


図2 JISコードとKEISコードの関係 KEISコードは、JISコードに 準拠したコードになっている。

したがって、日常使用の範囲内では、第1水準漢字集合でほぼ用が足りることになるが、住所、人名及び専門用語の入る適用分野では、第2水準まで必要となる。それでも完全な人名の扱いには不足するケースも生じている。

JIS漢字セット6,349字の符号化(コード付け)には、13ビットが必要になるわけであるが、英数字や片仮名用に制定された7単位及び8単位符号、並びにその拡張方法に基づく符号体系とするため、7単位又は8単位の2バイト符号を制定している。2バイトの各々に、94個の文字エリアを確保し、94×94=8,836字が割り当てられている。したがって、第1・第2水準漢字集合、非漢字集合(平仮名、片仮名、ローマ字、ギリシア文字など453字)の合計6,802字に対して、2,034字分は予備及び空き領域となっている。

JIS漢字符号系は、各種システム相互間で情報交換を行なう場合の標準として扱われるものとなるため、各種の漢字装置やシステムの開発は、このコードが基準となっている。日立製作所のKEISコードは、従来のEBCDIKコードとの制御符号のぶつかりを配慮して、JISと同一コードとはしていないが、コード変換を一率に行なえるよう1ビットの操作でで

表 2 テーブルウェア(辞書ファイル) 辞書ファイルの作成は、漢字情報処理の一つの技術分野を形成し、テーブルウェアと呼ばれる。

種類	内 容	主な用途
文字パターン辞書	ドット文字パターン	印刷・表示ハードウェア
漢字辞書	漢字コード, 部首, 画数, 音訓ほか。	漢字ソート, 異体字処理
固有名詞辞書 (地名,人名,法人名)	読みと漢字表記	仮名漢字変換
専門用語辞書 (事務用,技術用ほか。)	読みと漢字表記	仮名漢字変換
国語辞書	読みと漢字表記,品詞, 活用形	仮名漢字変換, 文章解析ほか。

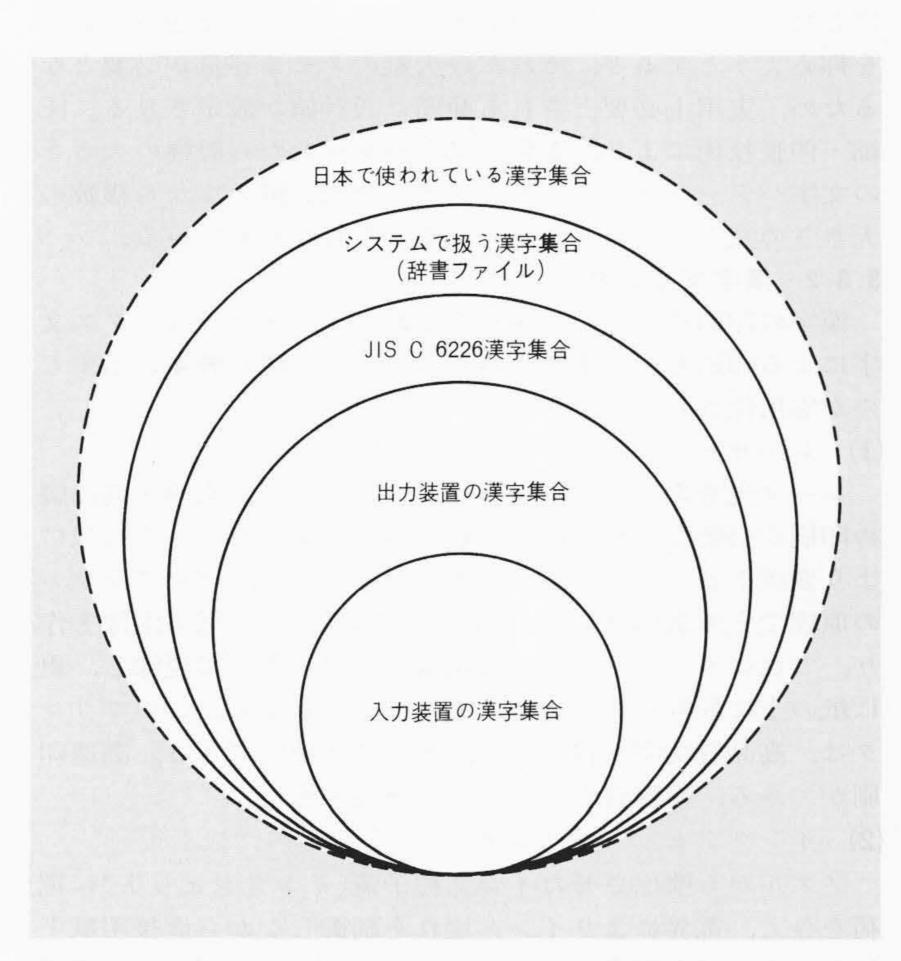


図3 コンピュータシステムで扱う漢字の範囲 入力装置,出力装置,辞書ファイル各々の漢字集合の相互関係を示す。

きるものにしており、JISコードに準拠している(図2)。

3.1.3 テーブルウェア(辞書ファイル)

漢字情報処理には**表1**で示した漢字の特性を吸収するため、文字ごとにかなりの量のデータが必要であり、また、漢字表現の熟語、用語なども使用する。これらは、多量のデータであり、それぞれの作成や配列技術などには専門的な知識や多くの作業を伴うため、漢字処理の一つの大きな技術分野を形成し、「テーブルウェア」とか「辞書ファイル技術」とか呼ばれる。辞書ファイルの種類や、その内容及び主な用途を**表2**に示す。

3.1.4 外字処理

入力装置、出力装置及び辞書ファイルのそれぞれに内蔵される文字容量には、図3に示すように差があり、当該装置にない文字の扱いを可能にすることを「外字処理」と呼んでいる。

入力鍵盤上にない文字を入力する場合を「入力外字処理」と言い、外字を表わす制御記号(符号)と4文字の16進数により、漢字1字分を表わす。これをプログラムにより2バイトの漢字コードに変換する。

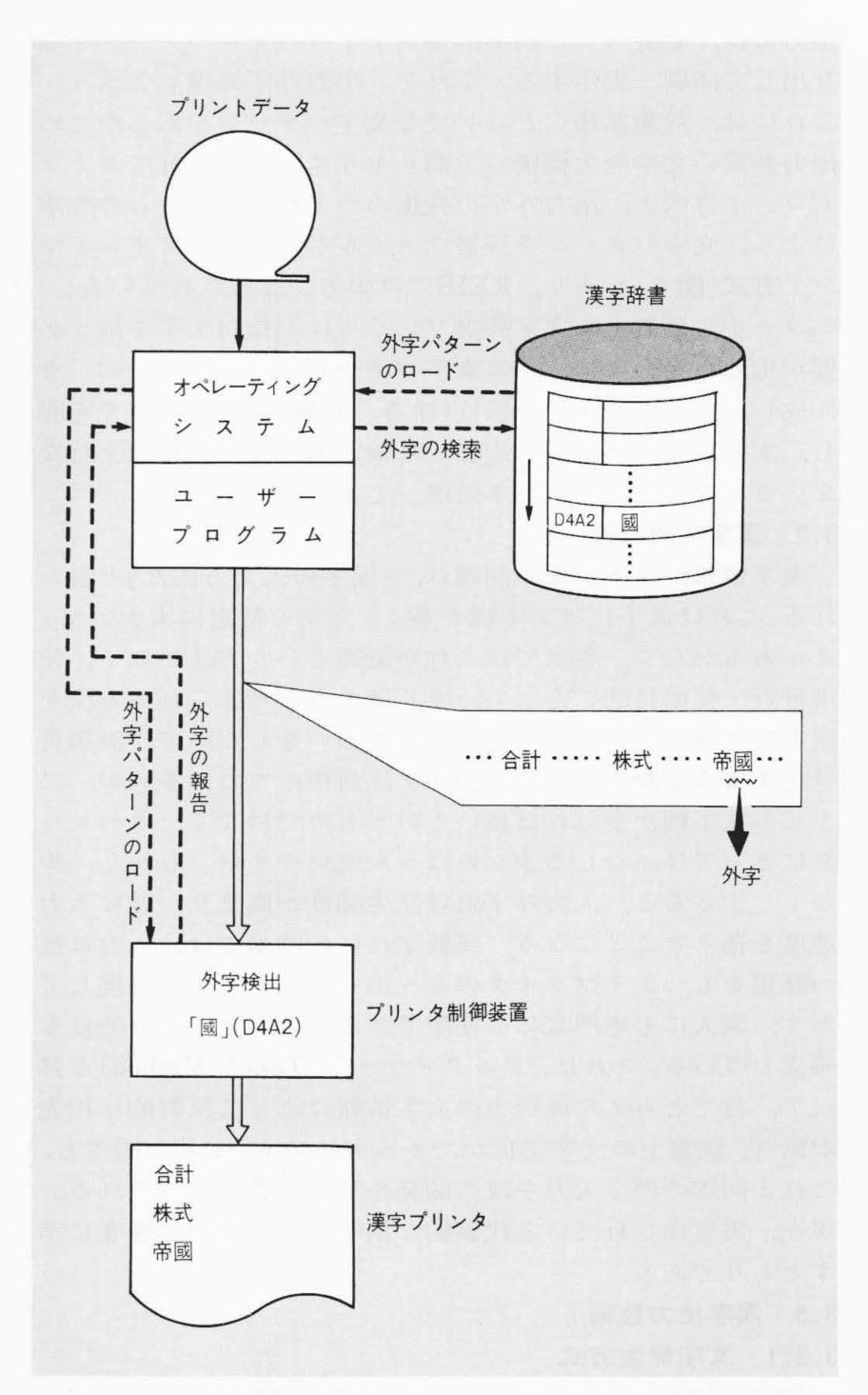


図 4 オンデマンド出力外字処理 「國」を外字として検出したプリンタ制御装置からの割込みにより、臨時に漢字辞書から「國」に相当する文字パターンをプリンタ制御装置に転送する。

表 3 代表的な漢字入力方式 適用分野,目的に応じて最適な入力方法が選択される。

9 "	入 力 方 式	入 力 方 法	備考		
直接入力方式	多段シフト方式 (漢字テレタイプ方式)	キー上に4~15個の複数の漢字を割り当て、シフトキーで該 当漢字を選択する。	習熟すると、かなりの入力速度が得られる。		
	タブレット方式	タブレット上にマトリックス状に配列された文字をペンで触 れて入力する。	文字配列を代表読み順などにできるので,素人にとって索字 が行ないやすい。反面,入力速度はあまり上がらない。		
仮名キーボード入力方式	仮名漢字変換方式	読み入力した仮名データをソフトウェアで漢字に変換する。 仮名キーボードが使える。	変換率を100%にするには校正をするか、変換内容をあらか じめ1:1に設定する。		
	連想記憶方式	漢字の意味や字形などから連想するものを仮名2文字に置き換え,暗記して入力する。	熟練すると最も速い入力速度が得られる。連想方式や訓練方 法に各種のノウハウがある。		
	漢字コード方式	漢字コード 2 バイト分をそのまま 16進数 4 桁で入力する。コード表を見ながらの入力となる。	入力速度は遅い。外字入力や少量のデータ入力に使用される。		

プリンタやディスプレイ装置の文字発生機構にない文字の出力には、必要のつど漢字辞書ファイルの文字パターンを取り出して印刷・表示する。これを「出力外字処理」と言う。これには、対象業務ごとに必要な文字パターンをあらかじめ出力装置の文字発生機構の予備メモリエリアに入れておくプレロード方式と、出力外字の発生のつど出力装置からの要求により、文字パターンを辞書ファイルから取り出すオンデマンド方式(図4)があり、KEISでは両方が支援されている。

ファイル装置上の文字辞書ファイルにもない文字を扱う必要が生じた場合には、その文字パターン(ドットパターン)を作成し、コードや文字の属性(読み、画数など)について定義し、文字辞書ファイルに追加した後、入力や出力処理を行なう。これを「システム外字処理」と言う。

3.2 漢字入力技術

漢字情報処理の最大の問題は,「漢字の入力方法だ」と言わ れる。これは漢字には文字種が多く、文字の特定に大きなネッ クがあるからで、現状ではこれが最適という方法がなく、使 用分野・使用目的に応じて、より適すると考えられる方法を 選択して使用される。入力外字処理扱いをしないで、直接符 号化する文字数としては、3,000字種前後のものが多いが、必 ずしも文字種が多ければ良いというものではなく, オペレー 夕にとっては、むしろ少ないほうが使いやすい。しかし、少 なくしすぎると, 入力外字処理発生頻度が高まり, 逆に入力 速度を落とすことになる。英数字だけの欧米では、入力は長 い歴史をもつタイプライタのキーボード入力技術が発展して おり、素人にも専門家にも使用できるものとして社会的にも 確立している。これは、タッチメソード(Touch Method)と言 って, 目でとらえた原稿上の文字情報によって反射的に指先 が動き,鍵盤上の文字選択ができる方法であり,我が国でも, これと同等の漢字入力手段の開発や社会的普及が望まれる。 現在, 実用化されている代表的な漢字入力方法は, 表3に示 すとおりである。

3.3 漢字出力技術

3.3.1 文字発生方式

漢字の印刷·表示には,文字発生装置(機構)が必要である。 これには,ディジタル式と字母式に大別される各種方式が開 発されてきたが,大勢はディジタル式に属する「ドット文字 方式」が主流を占めるようになってきた。ドット文字方式は、モノクロのドットマトリックスパターンで文字を表わす方式で、モノクロに対応する2値情報をメモリ上に記憶し、印刷・表示の際に、このメモリ情報を読み出して文字パターンとして使用する。1文字のドットマトリックスは、通常の8~10ポイントの大きさの文字で16×16ドット、24×24ドット及び32×32ドット(縦×横)によるものが一般的となりつつあるが、写真植字機の分野では64×64ドットあるいは128×128ドットまで必要と言われている。

ドット文字パターン用メモリは、約8,000文字種の場合、 1文字24×24ドットでは約500kバイトもの容量が必要であり、 装置のコストに大きな影響を与える。文字品質を上げたり、 各種書体(明朝体、ゴシック体など)による印刷・表示を可能 にしたり、出力外字の出現頻度を少なくして印字速度の低下 を抑えようとすると、それだけ大量のメモリ容量が必要とな るため、実用上必要とされる範囲に設計値が設定される。圧 縮・伸長技術により、1個の文字パターンから数種の大きさ の文字パターンを発生させ、メモリ増加を抑えながら複数の 大きさの文字を印刷・表示することも行なわれている。

3.3.2 漢字プリンタ

漢字の印刷技術は、前述したとおりドットマトリックス文字による印刷方式が主流を占めつつあり、現在表4に示すものが実用化されている7,8%。

(1) レーザビームプリンタ

レーザ光を応用した電子写真式プリンタで、高速・高品質の印刷が必要な分野で使用されている。文字パターン情報により変調されたレーザビームをスキャンして、ゼログラフィの原理で光導電体材料(セレンドラムなど)上に電子潜像を作り、これにカーボントナーと現像剤を付着させて現像し、更に紙の上に転写して熱定着、又は圧力定着する。このプリンタは、高品質文字が得られる、普通紙が使用できる、高速印刷ができる、衝撃音がないなどの特長をもつ。

(2) インクジェットプリンタ

ノズルから噴出させたインク粒子流(インクジェット)に電荷を与え、電界によりインク流れを制御しながら直接用紙上にドット文字を形成するもので、インク流れの制御には電界によるものと粒子に与える電荷によるものとの二通りがある。

表 4 主要漢字プリンタ ドットマトリックス文字による漢字プリンタが主流になった。

L	種	類	用 #	紙	漢字印字速度
ノンインパクト式 インパクト式	レーザビームプリン	9	普通統	紙	~15,000LPM
	インクジェットプリ	ンタ	普通	紙	~85CPS
	サーマルプリンタ(ラ	インプリンタ形)	感熱	紙	~120LPM
	サーマルプリンタ(シ	/リアルプリンタ形)	感熱	紙	~20CPS
	ワイヤドットプリン	タ(ラインプリンタ形)	普通	紙	~100LPM
	ワイヤドットプリン:	タ(シリアルプリンタ形)	普通	紙	~60CPS

注:LPM(行/分) CPS(字/秒)

このプリンタは、現像・定着が不要な一次発色方式である点及び衝撃音が極めて少ないため音が静かである点などに特色をもつ。

(3) サーマルプリンタ

薄膜,厚膜,半導体などの熱抵抗体で作られたプリントへッドを感熱記録紙に接触させ、記録紙上に顕色させる原理によるものである。印字速度は遅いが、比較的良い印字品質が得られ、特殊紙が必要というハンデはあるものの、簡単な機構による小形化が可能な点、衝撃音がないなどの点に特色がある。

(4) ワイヤドットマトリックスプリンタ

ワイヤをインクリボンを介して用紙上に打ち当て、1ドット分を印刷する原理のもので、縦に1文字分のワイヤを並べた印字へッドを作り、これを横方向に動かして印字するシリアルプリンタ形と、横方向に一定間隔で並べた複数のワイヤを一斉駆動しながら横方向に移動させ、1ドットライン分の印字が終了後、用紙を1ドット分送り出して次のドットラインを同様に印刷するラインプリンタ形のいずれもが実用化されている。シリアル形は16~24ワイヤのヘッドが開発され、毎秒50字前後の印字速度が得られている。

ラインプリンタ形では、1ドット行のワイヤ数を増やせば、1ワイヤの印字担当区域が狭くなり、それだけ印字速度は向上するが、ワイヤの駆動機構が増えるため高価になるので、実用上の最適設計が行なわれている。衝撃音が出る難点はあるが、カーボンにより複写印刷ができるので根強い需要がある。ワイヤの直径は0.2~0.3mmであり、細くすることに限度があるため、24×24ドット文字が限界と言われる。

3.3.3 漢字ディスプレイ

漢字プリンタには種々の方式が開発されているのに比べ、 漢字ディスプレイはCRT(Cathode Ray Tube)による方法 だけ実現されている。4,000字種以上の漢字の表示には16×16 ドット程度以上が必要なため、従来の仮名文字のものに比べ、 解像度を2倍以上に向上させたCRTが使用される。モノクロ表示では24×24ドット文字を1,000字種程度まで表示できる。カラーによる表示も16×16ドット文字程度までは可能になっているが、モノクロ表示と同程度になるまでには達していない。

今後は文書形式で、文字はもちろん、図形などの表示も必要とされるため、より高精度のモノクロ・カラー表示CRTの開発が技術課題となっているが、CRT以外の表示技術についても開発努力が続けられるであろう。しかし、CRT以外のものについては、漢字表示以前の技術課題が多い。

3.4 漢字処理ソフトウェア

漢字・日本語処理のソフトウェア技術は、制御プログラムからアプリケーションプログラムまで多岐にわたっている。

(1) ソフトウェアシステム開発のための考え方

OS, ユーティリティなどのベーシックソフトウェアからアプリケーションソフトウェアまで,一貫した思想に基づいて開発する必要がある。その際,従来の英数字・仮名文字に限定されたソフトウェアとも共存併用ができるように拡張する必要がある。具体的には,スーパバイザ,データ管理,通信管理,ジョブ管理,サービスプログラム,データベース管理システムに関して,対話処理,オンライン処理,バッチ処理,分散処理などの各種処理形態で,漢字・日本語情報処理機能を支援する必要がある。

(2) 言語処理プログラム

プログラム言語仕様に、漢字定数の定義機能やリテラル作成などの機能追加支援が行なわれる。プリンタやディスプレイの書式・画面様式⁹⁾に漢字項目や罫線を入れる機能もサービスプログラムで支援される。

(3) 漢字入出力支援

コード変換,編集,外字処理など,漢字入出力装置の特殊性を吸収し,従来の英数字・仮名文字体系での運用と同等にまで近づけるための支援が行なわれる。1バイトコード体系と2バイトコード体系との併存,混在運用の支援を行なっている。

(4) システム開発支援システム

システム建設の容易化、プログラムの開発費・保守費の低減を図るため、KEISでは日本語形式のソフトウェア開発支援システムCORAL*2)が備えられており、更にこの面の研究開発も進められている¹⁰⁾。

(5) 仮名漢字変換支援

ファイル作成, 言語, データベース, 対話式プログラム入力, ファイル変換ユーティリティで仮名入力漢字変換機能を支援するため開発が進められている。氏名, 法人名の変換¹¹⁾ ツールとテーブルが支援されており, また, 日本語文の仮名漢字変換に関しては, 日本語文書編集¹²⁾の構成機能として開発が進められている。

(6) アプリケーション支援

従来のアプリケーションプログラムに、まず第一歩として 漢字表現機能を追加することが、必須の条件であり、現在開 発が進められている。また、情報検索で、仮名キー、漢字キ ーによる検索方式¹³⁾の支援が必要であり、現在研究が進めら れている。また、分散形データベースの実現¹⁴⁾には、キーワ ード自動¹⁵⁾抽出、自動インデクシング^{16),17)}の具備が必須条件 であり、そのための基礎技術の開発が行なわれている。

日本語ワードプロセッシングなどの文書編集出力機能^{12),18)}についても、スタンドアロン形のもの、インテリジェント端末形のもの、TSS(Time Sharing System)形のものなど、各種の処理形態のソフトウェアが開発されている。

4 漢字処理適用分野と標準化の動向

4.1 処理内容と適用分野

漢字情報処理には,処理内容的に大きく「漢字データ処理」, 「文書編集処理」とでも言うべき二つの形態がある。

「漢字データ処理」とは、出力帳票で言えば、伝票などに人

※ 2) CORAL: Customer Oriented Application Program Development System

名,住所,品名など漢字項目を含む処理で,従来の事務処理の大部分の処理形態と同様の内容に,漢字を含めた処理である。人事,販売,在庫,資材,財務,会計,統計などの一般事務管理業務や,保険契約業務,不動産取引あるいは地方自治体での住民記録,税務,保健,就学などの住民情報サービス業務,更に,医療機関でのレセプト発行・診療記録及びこれらの業務を受託している計算センターなどで実施されている処理業務の大部分は,この漢字データ処理によるものである。

「文書編集処理(ワードプロセッシング)」は、漢字まじりの文章を編集し、文書形式で表示・印刷する内容のもので、簡単なものでは社内用の通達文の作成管理、社内用資料の作成管理などがある。この機能は、オフィスオートメーションを構成する重要な要素となるもので、大形から小形まで各システムにそれぞれ適した形で実現するため開発を進めている。このほかに、高速な編集を伴うものとしては、出版物原稿編集、印刷版下作成などがあり、このための研究開発も進めている。また、文献の抄録などをもとにした情報案内、検索サービスなども実用化されている。文書編集処理は、漢字データ処理のあとを追って開発されてきており、現状では両者の間に不連続点がある。今後は、文書編集処理と漢字データ処理との有機的な結合を実現し、幅広いニーズに応ずる必要がある。

4.2 標準化の動向19)

漢字処理関連事項の標準化は、現在のところ前述した文字セットと符号についての下記JISが制定されているだけである。

- (1) JIS C6225 情報交換用漢字符号系のための制御文字符号
- (2) JIS C6226 情報交換用漢字符号系
- (3) JIS C6228 情報交換用符号の拡張法

その他については、標準化の必要が叫ばれているものもあるが、標準化の可能性についての検討すら進んでいないものも多い。以下に標準化の対象と考えられる項目を列記する。

- (1)「漢字入力用鍵盤(文字セットと配列方式)」
- (2) 「漢字入力方式」,「出力用字体(ドットパターン)」
- (3) 「文書処理用制御機能」,「外字コード」
- (4) 「專門用語辞書」

5 漢字システムの将来展望

漢字を取り込んだコンピュータ情報処理が、経済、社会、 文化のあらゆる面で活用され始めた。これに伴い解決されな ければならない技術課題も多い。

まず、入力技術について大きな課題が残されている。究極的には欧米でのタイプライタで実現されているタッチメソードと同程度の入力ができるキー入力手段や、手書きOCR(光学文字読取り装置)、音声認識などの実用化が課題であるが、ここしばらくは適用分野・目的ごとに、より合理的な入力方法、漢字セット、キー配列について研究開発が行なわれるであろう。仮名漢字変換については、辞書技術・プログラム技術の向上により、自動変換率の大幅向上も達成されるものと考えられる。

出力技術については、メモリ技術のよりいっそうの進展に伴う大容量化・低価格化とあいまって、より高品質・高速・高機能・小形・低価格化が進展するであろう。特にプリンタは、ノンインパクトプリント技術を中心に、マルチフォント化、添え字・ルビ表現などの機能向上、図形・画像を伴う印刷及びそれらのカラー化なども可能となり、現在の出版物印刷品質に近いものも部分的に出現すると考えられる。

各種のアプリケーションを支えるソフトウェアも長足の進歩を遂げるであろう。文書形式の処理を行なうアプリケーションでは、しだいに日本文という言語そのものとのかかわりを深めることになり、自然言語処理に関する研究開発が進められている。これらの成果は、情報検索、文書処理、自動翻訳などの進展や実用化を促すものと考えられる。

漢字処理に伴う技術発展は、図形・画像処理面での処理技術とともに、現在大きくクローズアップされているオフィスオートメーションの課題にも大きく寄与するものとなるであろう。

6 結 言

長い間の懸案であったコンピュータシステムでの漢字の取込みが実現した。しかし、まだまだ解決されなければならない多くの技術課題もあり今後の研究開発が期待されるが、ようやく「日本人のためのコンピュータ」ができ上がり、欧米諸国と同一環境に置かれたとも言える。今後、漢字を使う情報処理のあらゆる分野で、アプリケーション開発が急速に展開されようとしている。

参考文献

- 1) 日立製作所:小特集,漢字情報処理システム,日立評論, 60, 327~358 (昭53-5)
- 2) 三浦:最近の計算機技術の動向,電気学会雑誌,100,981~985 (昭55-11)
- 3) 日本工業標準調查会審議: JIS C6226, 情報交換用漢字符号系, 日本規格協会
- 4) 国立国語研究所:電子計算機による新聞の語彙調査(I)~(IV), 国立国語研究所報告(1976)
- 5) 国立国語研究所:現代新聞の漢字,国立国語研究所報告(1976)
- 6) 国立国語研究所:現代雑誌九十種の用語用字,国立国語研究 所報告(1961)
- 7) 日本電子工業振興協会:日本語情報処理技術の動向(昭和53 年3月)
- 8) 日本電子工業振興協会:日本語情報処理の研究調査(昭和54 年3月)
- 9) 鳥居,外:レーザプリンタにおけるプログラム制御方式のフォームズオーバレイ,情報処理学会第19回全国大会,311~312 (昭53-8)
- 10) 横山,外:PSL/PSAにおける図形入力方法の考察,情報処理学会第22回全国大会,309~310(昭56-3)
- 11) 筒井:宛名の仮名漢字変換システム,情報処理学会第15回大会,615~616 (昭49-12)
- 12) 絹川,外:オフィスコンピュータを用いた日本語文書編集システム,情報処理学会第22回全国大会,855~856(昭56-3)
- 13) 田中,外:カナ・キーワード入力による漢字情報の検索方式の提案,電子通信学会総合全国大会,6-163(昭55-3)
- 14) 瀬谷:分散型データベースの将来予測,情報処理, 22, 186~195 (昭56-3)
- 15) 絹川:情報検索のための日本語解析,情報処理,**20**,907~910 (昭54-10)
- 16) 絹川,外:日本語文構造解析による自動インデクシング方式, 情報処理学会論文誌, 21, 200~207(昭55-5)
- 17) H. Kinukawa et al.: Japanese Sentence Analysis for Automatic Indexing,; Proc. of The 8th International Conference on Computational Linguistics (COLING80) '80. Sep.
- 18) 武市,外:分散型オフィス情報システムのためのオフィスワークステーション,情報処理学会第22回全国大会,1027~1028 (昭56-3)
- 19) 日本電子工業振興協会:日本語情報処理の標準化基礎調査 (昭和55年3月)