U.D.C. 681.322.06＋681.327.1]:003.324.2

漢字情報処理システムの動向

Trends in KANJI Information Processing System

松岡　濶*　*Matsuoka Hiroshi*
及川　巖**　*Oikawa Iwao*
吉田浩三***　*Yoshida Kôzô*

我が国での情報処理機器として，コンピュータシステムが漢字情報処理機能を具備することは当然であり，不可欠であると言える。そして漢字情報処理システムの既稼動システム数，計画中のシステム数とも年々増大してきている。

本稿では，漢字情報処理システムの動向について述べ，これにこたえる日立漢字情報処理システムの概要及び特長，並びに高レベル処理技術を利用した研究開発中の自然言語処理システムの一例を紹介する。

幅広い研究開発を基盤として，日立漢字情報処理システムは更に拡充を重ねるものであり，拡大・多様化するニーズに対応していくことができるものと考える。

1 緒　　言

我が国の一般社会で流通している情報の表現は，普通，漢字仮名交じりの日本語文であるから，情報処理システムの中核的存在であるコンピュータが，漢字仮名交じりの情報を扱うことは不可欠であり，また当然であると言えよう。そして，これを十分に達成しないまま放置されるならば，コンピュータの処理情報と一般流通情報との間の不連続性を人力で解消しなければならず，社会全体に果たすコンピュータの役割は欧米に比べて著しく小さい範囲にとどまる恐れがある。

近年，コンピュータによる漢字情報処理への関心と要求が強まっているのは，この意味から当然であり，これら要求にこたえることはコンピュータメーカーに課せられた義務であると考える。

コンピュータによる漢字処理の実現は，新聞の編集にかかわるものから始まり，現在一般事務処理，情報管理検索の分野に広がっている。本稿では，これら漢字情報処理システムの動向を述べるとともに，これにこたえる日立漢字情報処理システムについて概説する。

2 漢字情報処理システムの動向

2.1 漢字情報処理システムの発展とその背景

漢字情報処理システムは，この十余年の間に先駆的なユーザーによる試行的段階から，多数ユーザーによる実利用の段階にまで発展してきた。この背景には多様なユーザーのニーズと，それを支える数多くの技術開発とが相まって発展を遂げてきたことがある。その過程は，目的，原因が複雑に絡み合っているが，その中で顕著な事柄として次のものが挙げられよう。

2.1.1 特殊分野から多種分野への適用の拡大

初期の漢字情報処理システムは，新聞社での漢字テレタイプセッティングとコンピュータ処理との結合であると言われ

表1　漢字情報処理システムの適用対象分野のタイプ分けとシステムの特長　漢字情報処理の対象分野は，現在，一般事務処理，編集・印刷，情報管理検索の三分野に大別できる。

項番	対象分野（業種）	システム内容	適用業務例	主要機能構成	今後の課題
1	一般事務処理分野（一般企業，計算センタ）	従来の英・数字，仮名文字の事務処理システムとほぼ同じ範囲の情報処理システムに，漢字処理機能を導入したもの	あて名印刷 在庫品リスト作成 各種名簿作成 経営資料作成	・入力編集機能 ・簡易出力編集機能 ・校正支援機能	・漢字ユーティリティ・プログラムの質・量の完備化 ・入出力操作の簡易化と低コスト化 ・データベースとの結合
2	編集・印刷分野（新聞・報道，印刷・出版）	書籍，新聞などの編集，印刷，発行を行なうためのシステム	新聞編集 書籍編集 書籍原稿管理	・入力編集機能 ・校正支援機能 ・割付指示処理機能 ・索引作成機能 ・写植機能	・割付面積の拡大化，高速化 ・記事編集と図表処理の一元化
3	情報管理検索分野（官公庁，一般企業）	漢字仮名交じりのドキュメントの検索や内容検索を行なうシステム	各種文献検索 特許情報検索 図書・資料館業務	・蓄積データの入力編集・校正機能 ・検索要求（漢字キーワードを含む）受付解読機能 ・検索機能 ・シソーラス管理機能 ・検索結果の出力編集機能	・オンライン検索システムの標準化 ・自動インデクシング

* 日立製作所システム開発研究所　** 日立製作所神奈川工場　*** 日立製作所ソフトウェア工場

ている。言うまでもなく，我が国の新聞は漢字仮名交じり文によって編集されており，短時間内での情報収集と印刷とを必要としている。漢字情報処理の機械化の先駆となる必然性をもっていたと言えよう。

このような特殊な分野で開発された機器や技術が，その後他の出版関係や，金融機関での証券代行業務などへと拡大適用され，更に一般事務処理へと発展を遂げてきている。その過程で，量的，質的に処理内容の多岐化と広範化をもたらし，かつ高度な技法の適用を実現するという変化をもたらしてきている。**表1**に現段階及び近い将来での漢字情報処理システムの適用対象分野とその特長を示す。今後の発展の方向としては，情報処理のあらゆる分野に漢字情報処理が採り入れられ，かつ高度利用が図られるものと考えられる。

なお，漢字情報処理システムを処理要素別にみるならば，その処理内容によって，次のように大別することができる。

(1)　データ処理

対象が漢字であることを特に考慮する必要はなく，従来の英・数字，片仮名処理と同様のロジックで処理するものである。住所，氏名，商品名などの分類，印刷などがその例である。これらの情報は処理上，レコード内の特定のアイテムであり，かつそのアイテムの内容が漢字コードで入力され，アイテム単位の情報の複写や転送が行なわれた後，ハードウェアによって漢字の形で出力されるものである。

アイテムの内部にまで立ち入って処理をしないと言う意味で，漢字入りドキュメントを一つの画像として蓄積，転送などを行なうこともこの部類に入れることができる。実際長大なドキュメントを漢字コードとして入力することは大きなコストを要するから，その内部の処理を必要としないならば，フィルムなどの画像媒体に記録するほうが安く大量に扱うことができる。これは漢字の画像としての扱いであり，画像処理技術の進展とともに実利用化が期待される。

(2)　編集処理

漢字を含む文字列を日本語の文章として処理を行なうもののうち，文書の体裁を整えるための編集処理範囲にとどまるものである。出力編集処理での行頭・行末の禁則処理，ページ内の割付けなどの処理がその例である。

(3)　言語処理

漢字を含む文字列を日本語の文章として扱い，言語認識への接近を目指すものである。これらは，たいていシンタックス（構文解析，文法的解析）やセマンティックス（意味論的解析）などの言語解析を必要とする。仮名漢字相互変換，情報検索システムでの自動キーワード抽出，機械翻訳などがその例である。

一つの分野の情報処理システムをとって考えるならば，上記3種の処理それぞれを部分的に含むシステムとなる。しかし，利用の動向としては，まずデータ処理と簡易な編集処理が普及し，次いで高度な編集処理，言語処理の順に普及の輪を広げてゆくものと考えられる。

2.1.2　リアルタイム処理の発展

情報処理システム全般の傾向であるリアルタイム処理の普及は，漢字情報処理でも同様であり，漢字ビデオデータシステムを利用したデータ校正や，システムへの指示，照会などが，システム構成上ますます重要度を加えてきている。検索や照会などに対する回答を，漢字を用いた表示とすることによって，漢字の視覚に訴える大きな速読性が有効に利用される。また仮名表示としたときの同音語の判別困難化を防止することができる。校正やテキストエディティングに対しては，結果を確認しながらの作業となるため，正しい結果への到達が速い。

2.1.3　漢字情報のデータベース化と広域オンライン化

既に幾つかの官公庁では，各種ドキュメントや申請情報を漢字仮名交じり文によってコンピュータシステムに入力し，膨大なデータベースを作成している。またデータベース化計画中のものも多い。一般企業でも顧客情報，取引先情報などのデータベース化が進展している。一方，漢字入力の能率の低いこと，漢字端末が高価であることなどから，漢字情報処理システムの広域オンライン化の普及は，一般の情報処理システムのオンライン化よりも遅れている。しかし，官公庁，金融機関などで，漢字情報処理システムの広域オンライン化，又は広域オンラインシステムへの漢字処理の導入のニーズは高いものがある。ことに，前述の漢字データベースが建設されているところでは，これを遠隔地から検索可能とする必要性が強い。システム技術，ハードウェア技術の進展に伴い，ここ数年のうちには実用化が広がるものとみられる。

2.1.4　漢字コード等の標準化

従来，漢字コードは漢字機器により異なっていた。また，ユーザーによっては独自の漢字コードを設定して，漢字情報処理システムを構築していた。このため，ユーザー間での漢字情報の交換が円滑には行なえない状態であった。昭和53年1月，JISによって情報交換用漢字コードの標準[1]が設定されたことは大きな前進と言えよう。既開発の漢字情報処理システムでは，この標準コードへの移行になお問題があろうが，今後開発されるシステムでは，全面的にこのJISコードを採用するか，そうでないにしてもこのJISコードとのコード変換の手段を具備することにより，漢字データベースの流通が円滑化されるものと期待される。

字形（「曽」と「曾」など）など，漢字コード以外についても標準化が望まれ，実現されることが期待される。

2.2　日立製作所の漢字情報処理システム

2.2.1　システムの構成

日立製作所は，**表2**に示す漢字情報処理システムを開発した。このシステムは，HITAC　MシリーズVOS2（Virtual Storage Operating System 2）及びVOS3によってサポートされる。

漢字情報処理システムを構成する要素は，システム化の度合について基本的な要素から業務処理システムそのものまで各段階が考えられ，それらは複雑な階層構造を形成するが，基本エレメント，標準サブシステム，業務処理システムの3段階に大別して考えることができる。高度なサブシステムまで標準化することにより，業務処理システムの構築を容易化することができる。今後の開発を含め，漢字情報処理システムの概略の構成要素を**表3**に示す。既開発のものに加えて順次整備拡充してゆく計画である。

2.2.2　システムの特長

このシステムは，次のような特長をもっている。

(1)　幅広い処理形態に対応できる製品群である。

既に述べたように，コンピュータの利用形態はバッチ処理からリアルタイム処理，分散処理へと多様化してきているが，日立漢字情報処理システムは，このような各処理形態に対応できるよう，各種の製品を用意している。

センタバッチ処理では，H-8196漢字プリンタサブシステムによる高速出力機能や，帳票様式を同時に印刷できる点などが有効に利用できる。また，各種印刷物の版下作成にはH-8195漢字プリンタサブシステムが有効である。

表2　開発した日立漢字情報処理システムの構成　　集中処理，分散処理，リアルタイム処理など幅広い処理形態に対応でき，業務処理プログラム作成を容易化する豊富なユーティリティを備えている。

種別	名称	概要
ハードウェア	H-8196 漢字プリンタサブシステム	毎分 7,000行の印字速度をもつ集中出力処理向けのレーザビーム電子写真方式のプリンタで，連続折畳み普通紙を使用する。
	H-8195 漢字プリンタサブシステム	鮮明な印字品質と豊富な編集機能をもったレーザビーム電子写真方式のプリンタで，カット紙を使用する。 （A4サイズ，12枚/分）
	T-560/40 漢字ビデオデータシステム	問合せ/応答，情報検索などのリアルタイム処理向けのビデオデータシステムである。
	H-1811 漢字入力装置	日立標準漢字コードを紙テープにせん孔するペンタッチ方式の入力装置
ソフトウェア	漢字入力編集	H-1811漢字入力装置で作成した漢字入力データを，指定された形式に従って編集し，漢字データセットへ出力する。モニタプリントも可能。
	漢字編集プログラム	豊富な編集機能をもち，漢字データセットのデータを与えられた制御パラメータに従って漢字プリンタ出力形式に編集し，H-8196又は磁気媒体に出力する。
	MAP/5435	業務処理プログラムと連結して使用され，簡単な画面のマッピング情報を与えることにより漢字ビデオデータシステムでの漢字データの操作を可能とするプログラムである。
	外字処理ルーチン	業務処理プログラムと連結して使用され，出力外字を漢字辞書から取り出し出力装置へ登録できる形に編集するサブルーチンである。
	書式オーバレイゼネレータ	H-8196漢字プリンタサブシステムの書式オーバレイ機構で印刷する帳票フォーマットを，簡単なパラメータの記述によって作成するプログラムである。
	漢字ライブラリ保守	漢字辞書ライブラリの作成及び修正を行なう。漢字辞書ライブラリを入力として，ユーザーシステム上に漢字辞書を作成することもできる。
	漢字辞書ライブラリ	漢字の音訓の読み，部首，画数などの属性情報と，漢字プリンタ，漢字ビデオデータターミナルで使用されるドットパターンの情報とを磁気テープに収容したものである。

リアルタイム処理のためには，T-560/40漢字ビデオデータシステム及びMAP/5435が用意されている。H-8195漢字プリンタサブシステムには，漢字ビデオデータターミナルを8台まで接続できるので，窓口業務などを処理する分散処理システムを構成することができる。

(2)　業務処理プログラムを容易化するユーティリティ

既に**表2**に示したように，漢字入力情報の編集，ビデオデータシステムによるリアルタイム処理，各種漢字プリンタサブシステムへの編集出力などを容易化するためのユーティリティプログラムが用意されており，これらを組み込むことによって，業務処理プログラムシステムを容易に編成することができる。

(3)　帳票出力の環境改善

H-8195及びH-8196漢字プリンタサブシステムは，レーザ光を用いた電子写真方式の非衝撃形のプリンタであり，事務用の電子写真機と同様な静かな運転を行なう。また，H-8196漢字プリンタサブシステムでは，帳票フォーマットを同時に印刷でき，業務処理プログラムの指示により，ある制限内で複数個の帳票フォーマットを取り替えて印刷することができるので，フォーマット紙かけ替え作業や印刷後のページ合せ作業が省力化される。

2.2.3　自然言語処理への接近

漢字情報処理システムは，全般にデータ処理及び簡易な編集処理の要素から成るものが普及してゆくにつれ，言語処理への要求が強まるものと考えられる。言語処理的な技法が一歩一歩開発されることにより，漢字情報処理の柔軟性，有効性が格段に拡大されるものと期待される。日立製作所が研究・開発中の自然言語処理システムの一例として，日本語文情報の検索用インデックスの自動付与システムについて紹介する。

情報検索システムでは，蓄積情報作成作業をどのように省力化するかが大きな課題となっている。この課題に対して，国際関係ニュース文の情報検索システムを対象として，個々のニュース文に蓄積・検索用インデックスを自動付与する方式の研究を進めてきた。ここでのニュース文は，漢字仮名交じり日本語文であり，インデックスとは，内容を表わす重要語（＝キーワード）と，それらの重要語の文構成上の役割〔＝ロール：5W（Where, What, When, Why, Who）1H（How）に相当〕を言い，これらを自動付与するのがこの課題の特徴である。この自動付与方式を，(1)キーワード自動抽出と文節構成語の認定，(2)文構造解析によるロール自動付与から構成

表3　システム要素の概要　　システムの要素は，既開発の要素に加えて，順次整備拡充してゆく。

	基本エレメント	標準サブシステム	業務処理システム
構成要素	1. 漢字入力装置，漢字ビデオ，データターミナル，漢字プリンタなどのハードウェア 2. 漢字データ管理プログラム，漢字入力編集，漢字編集プログラム，漢字画面編集，書式オーバーレイゼネレータなどの入出力サポートユーティリティ 3. 漢字辞書ライブラリ，漢字ライブラリ保守などの各種ユーティリティ	固有名詞処理・変換サブシステム，文章仮名漢字変換サブシステム，構文解析サブシステム，情報検索サブシステム，会話型文書編集校正サブシステム，自動インデックスサブシステム など，業務処理システムの主要な要素となるサブシステムである。	一般事務処理，編集印別，情報管理検索，その他の分野の各業務処理システムである。

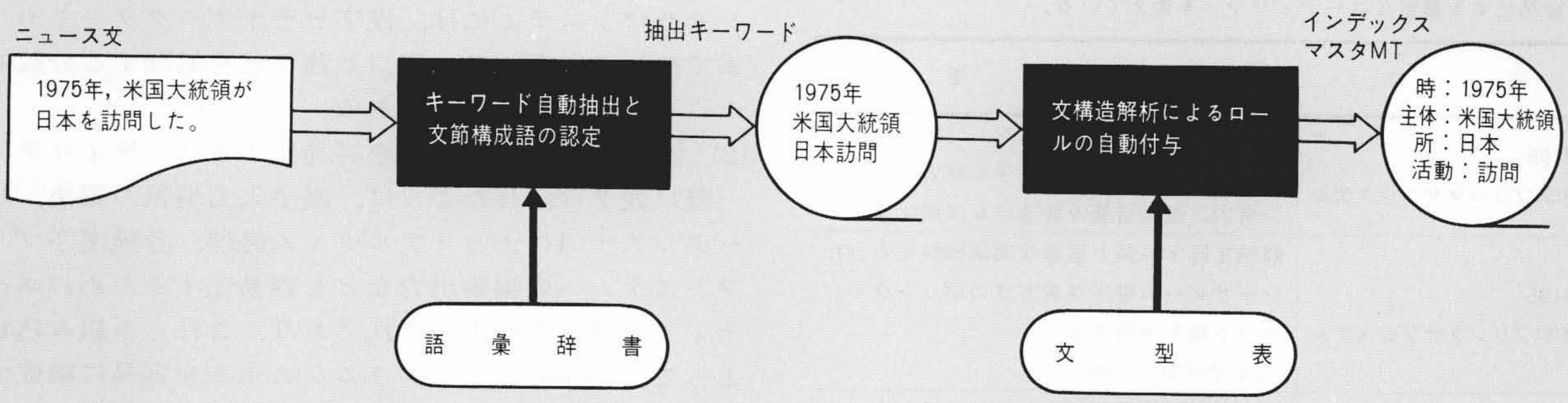

図1　インデックス自動付与実験システムの概要　　文節解析と文章解析との2段階の解析によって，インデックス自動付与が行なわれる。

７７０４０６カーター大統領、政府機関改組法に署名。

1　カーター	1　大統領
6　政府機関	6　改組法
5　署名	3　７７０４０６

７７０１１７フォード大統領、７８会計年度の予算教書を議会に送付。

1　フォード	1　大統領
6　７８会計年度	6　予算教書
6　議会	5　送付
3　７７０１１７	

７７０１１４上院外交委員会、バンス次期国務長官を承認。

1　上院外交委員会	2　バンス
2　次期	2　国務長官
5　承認	3　７７０１１４

７７０１１２フォード大統領、上下両院合同会議で一般教書をろう読。

1　フォード	1　大統領
6　一般教書	5　ろう読
3　７７０１１２	6　上下両院
6　合同会議	

注：抽出されたキーワードの左側の数字は，それぞれ
1：(有意志の)主体
2：(有意志の)客体
3：時
4：場所(＝地名)
5：活動
6：その他の主題
の意味のロールを表わす。

図2　インデックス自動付与結果の例　　各ニュース文で，キーワードの自動抽出とロールの自動付与とが正常に行なわれている。

する(**図1**)構想を固め，それぞれのステップの実験を行ない，その詳細なアルゴリズムを確立した。キーワードの自動抽出とロールの自動付与とを行なった実験結果の四つの例を，それぞれの元のニュース文とともに**図2**に示す。付与されたロールは，同図の注で示したように数字で出力した。実験の結果，キーワード抽出精度は，再現率で8割程度，ロールの自動付与精度も8割程度を実現できる見通しが得られた。100%に満たない部分は，新出キーワードの処理とともに人手校正にゆだねる計画であり，校正手段も含めたシステム化の検討を進めている。

3　結　　言

以上，最近の漢字情報処理システムの動向について述べた。この動向にこたえるため，日立製作所は漢字情報処理システムを開発した。このシステムは，(1)幅広い処理形態に対応でき，(2)業務処理プログラム作成を容易化する豊富なユーティリティを備えており，(3)帳票出力環境を大幅に改善するなどの特長をもっている。また，自然言語の認識や処理をも含め，幅広く漢字情報処理技術の研究開発を進めている。それらの結果をも駆使して，今後更に日立漢字情報処理システムの拡充と改善に努力を重ねてゆく考えである。

終わりに，本稿の執筆に際し資料の提供など，御協力をいただいた各位に対し，深謝の意を表わす次第である。

参考文献

1)　日本工業標準調査会審議：情報交換用漢字符号系，JIS　C 6226，日本規格協会（昭53-1）