

解 説カナ漢字変換入力法<sup>†</sup>牧 野 寛<sup>††</sup>

## 1. まえがき

吹米では、言語情報といえども、アルファベットを基本とする少數の文字種で済むところから、その入出力装置に対する技術的問題が比較的少なく、種々の言語データ処理システムが実用化されてきた。一方我が国では漢字が意味の伝達に大きな役割を果しており、したがって日本語の通常の表記法である漢字仮名混り文を処理することが重要な課題となる。

漢字は字種数が多く、その数4万～5万くらいあるといわれ、通常我々が用いている漢字に限っても最低2,000字程度は必要であり、人名、地名などに用いられる漢字までも含めると6,000～8,000字程度は必要だといわれている。このような字種の多さが漢字の入出力装置に与える影響は深刻で、日本語の機械処理上の隘路となっていた。しかし近年のハードウェア技術の進歩により、低価格で漢字出力装置が実現されるようになつた現在では、残された問題は漢字の入力であると考えられる。

数千もの漢字を平面上に配列した漢字鍵盤装置（いわゆる漢テレ）、邦文タイプライタ、漢字タブレットなどの入力装置では、人間が字形を見て直接漢字を出すという方法（サイト法）のために、一般の人が高速に入力することは容易ではなく、その操作は専門のオペレータの手に委ねられてきた。さらにこの方式の欠点として、目視によるため打鍵速度が遅い、鍵盤配列の異なる装置では経験が生かせないなどが挙げられる。一方漢字を何らかの記号列に割り当て、通常のタイプライタ装置等で入力する、すなわちタッチ法による入力方法も実用化されているが、漢字に対する記号列は全て人間の記憶に頼らねばならず、その入力実行は必ずしも容易ではない。以上の漢字入力方法は、いずれも人間の負担として入力をうるために、相当の訓

練を要し、誰もが容易に実行するわけにはいかない。このような専門家向けの入力法とは異なり、誰もが容易に漢字仮名混り文を入力できる方法として、文章を仮名（またはローマ字）で入力し計算機を用いて漢字仮名混り文に変換する仮名漢字変換の方式が注目され、研究が行われてきた。

本稿では、仮名漢字変換の研究の流れを踏まえ、変換上の問題点とその処理方法を中心にその概説を試みる。

## 2. 仮名漢字変換の研究の流れ

仮名漢字変換は、九州大学の栗原らのグループによって研究が始められ、入力カナ文の文節分かち書き、自立語辞書の検索法、単位分析法、辞書構成法など仮名漢字変換の基本的な手法が提案され<sup>1)</sup>、同時に辞書の編さん人が開始された。現在も九州芸工大の稻永らに引き継がれて編さんが行われている<sup>2)</sup>。ここでの提案を基に、沖電気の松下らはカナ漢字変換システムを試作した<sup>3)</sup>。さらにNHKの相沢らは、ニュース文を対象として局所文法処理、単純意味処理の手法を用い、実験システムを作成した<sup>4)</sup>。ここでは文節内の単位接続の検定に必要な文法情報がかなりコンパクトな分かち書き単位分析表にまとめられた。これら一連の研究はその後の仮名漢字変換の研究に大きな影響を与えた。また、共同通信社では外電のローマ字電文を漢字混り文に変換するシステムが作成された<sup>5)</sup>。さらに日立は、速記文から漢字混り文へ変換する速記反記システムを作成した<sup>6)</sup>。その後、電電公社の木村らは、同音異義語の解析に関連語情報や意味分類情報を利用したシステムを試作した<sup>7)</sup>。阪大の牧野らは品詞判定を容易にし、辞書検索を効率化できる自立語付属語分かち書きを提案している<sup>8)</sup>。以上の実験システムはいずれも汎用計算機上で実現されていたが、東芝の河田らは、ワードプロセッサのひな型とみられる実験システムをミニコンピュータ上で実現した<sup>9)</sup>。その後も分かち書きの柔軟さを求めて、べた書き入力方式の研究

<sup>†</sup> A Japanese Input Method by Kana-Kanji Translation by Hiroshi MAKINO (Faculty of Engineering Science, Osaka University).

<sup>††</sup> 大阪大学基礎工学部情報工学科

表-1 仮名漢字変換の研究

年代	研究機関	システム及び特徴
1967	九大 (九州芸工大)	基本アルゴリズム提案、辞書編纂。 引き続き辞書編纂、豊富な辞書項目、収容語数23万語。
	沖電気 共同通信 N H K	字種指定式、接辞テーブルによる接辞処理。 ローマ字から漢字混り文へのシステム。 文節内の単位分析表のコンパクト化、単純意味処理。
	UNIVAC 日立 阪大 東芝	姓名のカナ漢字変換システム。 速記反訳システム。 自立語付属語分かち書き入力。 ミニコンによる実験システム、訂正、編集機能強化。
	電電公社 阪大 東芝	IROHAシステム、関連語、連語解析。 べた書き入力。 仮名漢字変換方式のワードプロセッサ発売、文法処理有り。
	東工大 電電公社 富士通	マクロ方式、特定個人向け。 IDIOM方式、字形情報付加入入力。 ワードプロセッサ発売、文法処理なし。 引き続き、各社のワードプロセッサ発表、発売。
1981		

が、電総研<sup>10)</sup>、阪大<sup>11)</sup>、富士通研<sup>12)</sup>などで試みられてきた。一方、より処理の簡単さ、実用的立場から東工大<sup>13)</sup>、電電公社<sup>14)</sup>など多くの大学、研究所において仮名漢字変換またはその応用システムの研究が行われている。また人名、企業名などの固有名詞の仮名漢字変換システムも試みられている<sup>15)</sup>。このような成果を踏まえて、数年前に仮名漢字変換方式の日本語ワードプロセッサが東芝により商品化され<sup>16)</sup>、その後も漸次仮名漢字変換方式のワードプロセッサが各社によって発売され、今日に至っている。以上の研究の流れを表-1にまとめて示す\*。

### 3. 仮名漢字変換上の問題

仮名漢字変換方式には、漢字単位で変換する方法と上述のほとんどのシステムで採られた単語(熟語)以上の単位で変換する方法がある。前者の方法では漢字の音訓表程度の辞書を用意すれば良いが、もし入力文の漢字の読み通りに入力するとすれば、多くの場合は音読みなので、同音異字数が13.64と多くなり<sup>3)</sup>、漢字選択の負担が大きくなる。通常の文章入力よりは単なる漢字データなどの入力に適していると考えられる。したがってここでは主として後者的方式について述べることにする。

仮名漢字変換の主要な問題点として

- i) 入力文の分かち書き
- ii) 同音異義語の処理

\* この表は、手許の論文、資料等に基づいているので実際のシステム作成時期とは必ずしも一致しないかもしれません。

が挙げられる。

#### 3.1 分かち書き

日本語では通常漢字仮名混じ表記が用いられる関係上、アルファベットを基本とする英語などのような分かち書きの習慣がなく、カナ文またはローマ字文の入力に際し、分かち書きの問題が生じる。分かち書きは人にとっても容易で、なおかつ機械処理上も容易であることが要請されることから、種々の方法が提案されてきた。主な方法をまとめると

- 1) 単語分かち書き……単語単位に空白(区切り記号)を挿入する分かち書き
- 2) 自立語付属語分かち書き……自立語と付属語列との間に空白を挿入する分かち書き
- 3) 文節分かち書き……立節単位に空白を挿入する分かち書き
- 4) 文字種分かち書き(漢字指定式)……漢字部分を空白または括弧などの区切り記号で囲む分かち書き
- 5) べた書き……漢字混り文をそのまま表音表記する方式

などが挙げられる。

入力文に分かち書きによる制限を付加すればするほど機械処理上は容易となるが、入力者側の負担がそれに応じて重くなり、打鍵速度、誤入力などに影響を及ぼすことになる。比較的容易な分かち書き規則と考えられている文節分かち書きにおいても、付属語の認定についての個人差、また自立語および付属語の間での同形語の存在などに起因して、入力の際に詳細な文法知識が要求される場合が少なからず生じる。特に補助用言(いる、くる等)、形式名詞(こと、もの等)、不完全自動詞(くれる、やる等)など判断に迷うことが多い。

原文：人類は勝れた目と指先の感覚を持たなければならない。

- 1) ジンルイ ハ スグレ タ メ ト ユビサキ ノ  
カンカク ヲ モタ ナケレ バ ナラ ナイ。
- 2) ジンルイ ハ スグレ タ メ ト ユビサキ ノ  
カンカク ヲ モタ ナケレバナラナイ。
- 3) ジンルイハ スグレタ メト ユビサキノ カンカク  
ヲ モタナケレバ ナラナイ。
- 4) (ジンルイ) ハ (スグ) レタ (メ) ト (ユビサキ) ノ  
(カンカク) ヲ (モ) タナケレバナラナイ。
- 5) ジンルイハスグレタメトユビサキノカンカクモタナ  
ケレバナラナイ。

- 1) 単語
- 2) 自立語付属語
- 3) 文節
- 4) 漢字指定(文字種)
- 5) べた書き

図-1 入力分かち書きの例

## (例)

- (1) 行かなければならぬことはない  
 • イカナケレバ ナラナイ<sup>注1)</sup> コトハ ナイ\*\*  
 • イカナケレバナラナイ コトハ ナイ  
 • イカナケレバナラナイコトハ ナイ

注 1) \*は助動詞、 \*\*は形容詞

- (2) 私にも写せます

- ワタシニモ<sup>注2)</sup> ウツセマス

注 2) 「ニモ」は助詞

- (3) 持ってきたにも拘らず

- モッテキタニモ カカワラズ → (誤)  
 • モッテキタ ニモカカワラズ<sup>注3)</sup> → (正)

注 3) 「ニモカカワラズ」は接続詞

また、「語構成」、「漢字仮名混り文」などの語を一語と見なすか複数の語から成ると見なすか個人によって異なることは十分に考えられる。以上述べてきたように正書法の定まっていない現状では、単語や文節に対して我々の持っている漠然としたイメージのみで入力しても正しく変換が行われないという不満が入力者側に残ることになりかねない。このような状況に対処するために、現行のワードプロセッサでは、文節の定義を緩やかにし、派生語、複合語処理など種々の工夫が施されているが、完全には対処しきれず、ユーザによる語の登録機能などの側面的サポートを行っている<sup>16)</sup>。一方漢字指定式では、漢字混り文の原稿があれば、機械的に入力可能となり、漢字になるべき箇所が指定されているので、同音異義語のチェックにも利用できるなどの利点があるが表記上のゆれなどに災いされて正しく変換されない場合も起り得る。さらに「物価安定対策補助金」などの長い漢字列の語を処理することは必ずしも簡単ではなく、何らかの工夫が必要となる。

## (例)

- (1) (ナ)イヨウニ→{無いように ○  
 内容に ×}  
 (2) ハ(アク)スル→{把握する × (漢字部分)  
 は悪する ?}

べた書き入力では、分かち書きなどの知的負担がないため、仮名入力方式の本来の利点であるタッチ法による高速打鍵が可能となるが、その反面自動分かち書きの問題が生じる。べた書き文の分かち書きの方法として、カナ文字中に連続して二つの文節が抽出されたとき、初めて前半分の文節を正しいと認定する二文節最長一致法が提案されている<sup>17)</sup>。図-2 に解析例を示

入力文字列: ソウイウザッシヲ...

## (a) 文節候補抽出

- (a-1) ソウイウ△ザッシヲ...  
 (a-2) ソウイ△ウザッシヲ...  
 (a-3) ソウ△イウザッシヲ...

## (b) 二文節抽出

- (b-1) ソウイウ△ザッシヲ△  
 (b-2) ソウ△イウ△

## (c) 文節決定

- (c-1) そういう△ザッシヲ

(注) △は区切り記号

図-2 二文節最長一致法による分かち書き

す。図-2 a) に示されるごとく、文字列が文節となり得るかどうかの解析を試み、複数の可能性がある場合には、全てを取り出す。この段階では (a-1)~(a-3) の可能性が存在する。さらにそれぞれの後続の文字列に対して、同様な処理が行われる。今の場合、(b-1), (b-2) の二つの解析が成功し、(a-2) に対する解析が失敗するので捨てられる。次に連続する二文節と解釈される文字列の長さを比較し、最長な二文節の解釈を与える区切りを二文節間の境界と決め、順次同様な処理によって変換を実施していく。また、この他にも構文解析の手法によるもの<sup>11)</sup>や単語の頻度による評価で分かち書きおよび変換を行うもの<sup>12)</sup>などが報告されている。完全な自動化は困難といえるが柔軟な分かち書き入力を許容するにはこのような能力が機械側にも要求されることになる。

## 3.2 同音異義語の処理

仮名漢字変換において同音異義語の問題を考える際の第一の問題点は辞書の見出し単位と同音異義語の発生との関係である。たとえば「情報処理の機器構成」という場合、「情報処理」「機器構成」などの語を見出し単位に採れば、カナ文字列の一致のみでは一意的に変換が可能で同音異義語の処理が非常に簡単となる。しかしその反面辞書の膨大化は避けられないところとなる。ちなみに国立国語研究所の語い調査では<sup>18)</sup>、上例のような長い単位（長単位）の語の場合、延べ数約 68 万に対して異なり数は約 10 万に及ぶ。一方同一の標本に対して単語単位（短単位）では、延べ数 94 万に対し異なり数は約 4 万 8 千となっている。このような理由で従来の仮名漢字変換システムの多くは、原則として短単位の見出しを採用してきた。上例の場合では、ジョウホウ（上方、乗法など）、キキ（危機、既記など）、コウセイ（公正、校正など）などに同音異義語が発生することになる。このような前提に立った場合、どのような同音異義語が発生するのかを考えてみ

ると、

1) 自立語の品詞、活用が同じ

(例)

ハナ→ 花, 鼻, …

カイタ→ 書いた, 欠いた, 撃いた

2) 自立語の品詞属性は異なるが一部の活用形が同じ

(例)

イッタ→ 言った, 行った, 炒った

終止形……言う, 行く, 炒る

3) 接辞、派生語として同じ

(例)

ゼンセカイ → 全世界, 前世界, …

ニホンブン → 日本文, 二本分, …

4) 文節として同じ

(例)

ヒトハ → 火とは, 人は, …

などに分類される。以下ではすでに採られてきた同音異義語の処理について述べる。

#### (I) 形態素分析の利用

日本語文は文節が有限個連続したものと考えられ、さらに多くのシステムでは文節分かち書きを採用していることから、文節を単位とした形態素分析法が用いられる<sup>19)</sup>。

文節は、[(自立語)+(付属語)\*] という形式で表現できる(\*印は繰り返しを許すという記号である)。さらに動詞、形容詞、形容動詞などの用言では活用語尾が各活用形で定まっており、それに接続する助動詞なども規則化されているという性質を利用して、文節の解析を行うことができる。分析の方法としては、最長一致法<sup>20)</sup>と総当たり法<sup>21)</sup>が考えられている。前者は文節の文字列の左端を基準に辞書照合を行い、最長の見出しが一致する単語(同音語があれば複数)を採用し、後続の文字列の解析を行う。もし文字列が解析されずに残った時のみ、次長の文字列一致を引き起こす単語を自立語と設定する方法で、後者は辞書照合の段階で、最長一致の単語のみならず、全ての文字列一致を引き起こす単語を自立語の候補とする方法である。最長一致法では自立部の長さが異なる文節単位の同音語は解析されないので、現在は総当たり法を採用しているシステムが多いようである。これら両者の違いは主として辞書照合の方法にあり、解析法そのものは、辞書照合の後、各候補語について付属語検定を行うというもので両者に基本的な差はない。ところが辞書の照合

入力文字例: ヒトハ

- 最長一致 ⇒ 人 [名詞]→は [助動詞]
- 右方向解析 ⇒ 人 [名] →は [助]  
火 [名] →とは [助]
- 総当たり ⇒ 人 [名] ←は [助]  
火 [名] ←とは [助]
- 左方向解析 ⇒ 人 [名] ←とは [助]

図-3 解析方法の違い

が、外部記憶への参照を伴い実行速度の点で不利になるという判断から、文節の後端、すなわち付属語部分の解析を実施し、その前方にくるべき自立語の品詞推定を行った後、自立語辞書の照合を行うという左方向の解析方法を採用したシステムも考えられている<sup>19)</sup>。

解析方法の違いを図-3 に示す。このような形態素解析を行っても、文節単位の分かち書きでは一意には定まらず、漢字部が指定されるか、自立語と付属語に分かち書きされないと同音語を排除できない場合が生じるが、少なくとも辞書見出し上の見かけの同音語を排除することができる(通常、用言などの活用語は辞書見出しとして語幹または不変化部分が採用されることから、国語辞典の見出しに比べて辞書見出し上の同音語が増加する)。たとえば

イキティル

→(息 [名]+ている … (接続不可)  
→生き [動・連用]+て [助動]+いる [補助用言]…  
(可)

#### (II) 頻度情報の利用

我々が、通常国語辞典に記載されている単語全てを使用しているわけではなく、その中でもごく少数の語しか使用していないことに気付く。また新聞や雑誌の語の調査においても、対象となる分野での単語の使用頻度が異なることがよく知られている<sup>20), 21)</sup>。たとえば、「ハイケイ」という言葉が事務取引文書の中で出現した場合には、多分「拝啓」であろうし、また論文等の中ではむしろ「背景」の方が妥当と考えられる。したがって処理の簡単さも手伝って、多くの仮名漢字変換システムにおいて、単語の使用頻度が同音語選択の際に利用されてきた。さらに現行のワードプロセッサでは、使用者ごとの頻度を記録し、使用された単語の頻度を変更することによって使用者向きの辞書を構成していく機能や編集中に選択された単語を優先させる等の工夫を行って、単なる頻度による同音語選択の不備を補っているものが多い。

#### (III) 複合語・関連語情報の利用

日本語では、すでに述べてきたように他の単語と一緒にまとまりとなって複合語として用いられる。したがっ

て連接しやすい語を辞書に登録しておき、同音語の選択に利用する。また慣用句なども同様に処理される。

(例)

シンブン	キシャ	→	新聞	{記者 … (可) (汽車 … (不可)}
イッセンヲ	カクス	→	一線を	{画す … (可) 穏す … (不可)}

#### (IV) 構文・意味的情報の利用

単語の意味的な分類に基づいた同音語処理方法として、単語を数個のグループに分け、その間の距離を定め、距離計算によって語を選択する方法<sup>21</sup>や、単語の意味的な階層関係によって同音語を判定する単純な方法<sup>22</sup>も試みられているが、文章中の各単語間の構文的関係を基礎に置かなければ、その効果はあまり期待できない。

(例)

椰子の	ハヤシ	{林 … (可) 難子 … (不可)}
-----	-----	------------------------

動詞には、その主語（主格）、目的語（目的格）に特定の意味範囲を要求する場合が多く、そのような情報を構造化して辞書に埋め込むことにより、同音語の選択に利用することができる<sup>23</sup>。実際の処理では、構文的な係り受け解析を行う必要があるので処理が複雑化することは否めない。さらに語の意味的分類、動詞の格による分類など辞書作成の労力が大きいといえる。

(例)

ホンヲ	ヨンダ	→ 本を 読んだ
ヒトヲ	ヨンダ	→ 人を 呼んだ

### 3.3 特殊文節の処理<sup>16</sup>

単語単位（短単位）の辞書構成を採用している変換システムでは、分かち書きの柔軟さに対処するために複合語、派生語の合成処理、さらに数字、助数詞からなる数詞文節や固有名詞、人名・地名接辞からなる文節相当語句などの合成処理が行われている。

- (1) 複合語処理 (例) 健康・調査
- (2) 接辞処理 (例) 大型化 新時代
- (3) 数詞処理 (例) 第123回の
- (4) 固有名詞処理 (例) 情報処理学会 様

複合語処理は、通常の文節の処理が失敗した時に、行われる。すなわち自立部を一つしか持たないと仮定した場合の処理が不成功に終ると、続く文字列が自立語の可能性が高いとしてさらに自立語辞書を検索することによって行われる。

数	詞	接	数	例
数	詞	接	数	百・円
人	名	人	名	百・五十二
人	名	接	人	佐藤・様
人	名	接	名	徳川・家康
地	名	接	地	鹿児島・県
地	名	接	名	東京都・港 区
普	名	接	普	国際・会議

図-4 接辞テーブル

接辞処理は、頻繁に用いられる接頭語、接尾語を用意し、それらと自立語辞書の見出しとを合成することによって行われ、意味的、用法的なチェックはなされない。したがって、複合語処理、接辞処理によって通常は存在しないような語の組合せが合成されることがある。

固有名詞は特殊な読みを持つものが多いので、自立語辞書とは別の辞書が用意される。固有名詞の接辞や助数詞なども特有のものが多いために、人名、地名接辞や助数詞にグループ化したり、単語と接辞の連結関係の強弱によって優先的に処理がなされるなどの方法が試みられている。図-4 にその例<sup>23</sup>を示す。

### 3.4 辞書

辞書の構成法と変換アルゴリズムとは表裏一体の関係にあり、個々のシステムで異なるが、形態素解析を行うシステムでは、自立語辞書中には、単語見出しと対応漢字の他に、各単語の品詞、動詞、形容詞、形容動詞など用言の各活用型に関する情報、さらに用言の活用語尾表、付属語辞書、付属語間の接続規則表などの単位分析のための文法辞書が必要とされる。また派生語や特殊文節の処理を行うために、接頭語、接尾語、助数詞表など接辞辞書、固有名詞辞書などが用意される。また意味的な同音語処理を行うシステムでは意味的関連語、分類語い表の意味分類<sup>25</sup>、格支配関係などを含む意味用法辞書などが用意される<sup>22</sup>。

自立語辞書の収容語いをどのように決めるかは、システム設計上の最も大きな問題であり、個々のシステムの目的によっても異なるが、おおむね、国語辞典や新聞の語い調査、分類語い表の共通集合などから収容語いを決定している場合が多い。実用を目指したシステム、ワードプロセッサなどでは、2~3万語を基本語としているようである。一方長単位の語を採用した辞書では、基本語い数約6万語、複合語、派生語、慣用句などを含めると23万語に及ぶものがすでに作成されている。

従来の実験システムでは、固定式の辞書構成が採られ、逆にワードプロセッサでは、ユーザによる新しい語の登録削除が行える可変式の辞書構成を探ることによって、利用者自身によるアップデートを可能にし、分野による出現単語の偏りを吸収しようと図っている。さらに文章入力中に定義した語や同音異義語中の選択語を登録する暫定辞書や最初備わっていた辞書から頻度情報を使用に応じて変更し、利用者向きの辞書を自動的に構成する方法など辞書構成に関しても種々の方法が採用されている。

### 3.5 その他の変換方式

仮名漢字変換の多くは、自動的な変換を目指して、できる限り機械側で処理を行おうとするものであり、それらは一種の単語認識技術の集成化システムと見なすことができる。したがって、高度な処理が要求され、処理過程が複雑化するとともに、豊富な辞書情報が必要になり、変換システムの規模が大きくなる。これに対し機械側では、単語の辞書引き程度にとどめ、入力者側の負担によって同音語を減少させようという試みがある。

電電公社の IDIOM 方式<sup>14)</sup>は、图形としての漢字の特徴を符号化し、それらの符号を単語区切り記号として入力を行おうとする方法である。以下の例に見られるように漢字を縦横、上下、その他の三つのグループに符号化し、それらの符号の付加によって同音語を減少させている。しかしながら、全ての漢字に対し、瞬時に图形符号を付加することが一般的に可能かどうか疑問のあるところである。さらにこの符号を付加しても同音語が一意に定まらないなどの問題も残されている。

(例)

□ カンジ曰ノ□トクチョウ□ヲ□チュウシユツ□スル

## 漢 字 特 徵 抽 出

東工大のマクロ方式は、同音語に入力者側の自由な表現を付加し、計算機と入力者との約束表現（マクロ表現）によって一意的な変換を得ようとするものである。この方法では、変換は基本的に文字列の一致のみでよいことから機械処理上の負担が軽いという利点を持つが、同音語の処理を入力者側が背負い込むことになる。すなわち同音語を全て書き分ける必要が生じるわけでその負担は大きく、また個人または小人数の合意のもとでしか有効ではないという欠点を合わせ持

つ。しかしこの方法の強みは、訳書を一冊分仮名漢字変換したことである<sup>23)</sup>。

(例)

書く → カク／WRITE

掛る → カカル／テヘン

また市販のワードプロセッサにも、用言などは、通常の国語辞典の参照と同様に終止形で入力し、変換後文章中の語尾変化に入力者側が変更するという方法を採用し、文節分析を行わず、機械側の負担を軽減しようと意図したものもある。

## 4. ま と め

仮名漢字変換における問題点とその処理方法を中心にして解説を試みた。この他にも、システムの処理速度に大きな影響を与える辞書検索法などが挙げられるが、仮名漢字変換には限らず、ファイルを取り扱う情報検索、データベースなどに共通した問題であるので割愛した。さらに本稿で取り上げなかった問題に、入力鍵盤の問題がある。従来の仮名漢字変換の研究では主としてソフトウェア、テーブルウェアの研究として行われてきたために、人間工学的な色彩の強い入力鍵盤およびその配列に関する研究は、ほとんど行われなかつた。したがって JIS 鍵盤に従うキーボードを使用しているワードプロセッサが多いが、ローマ字形式の入力鍵盤の方が高速に打鍵可能という報告もある<sup>24)</sup>。また富士通のワードプロセッサでは、独自の鍵盤配列が採用されている。入力鍵盤、配列の問題は、入力速度、疲労度ひいては職業病の発生など、様々な要因を含み、今後さらに研究されるべき課題である。

仮名漢字変換の性能の向上は、複合システムとしての観点からの、個々の技術的要素の改良向上を考えいかねばならないが、とりわけ、同音異義語選択の方法は重要となる。同音異義語の問題は、言葉の認識、理解と基本的には等価であり、その解決には言語理論の基礎的な研究が不可欠となろう。と同時に、従来人手に頼っていた辞書作成の労力を軽減し、豊富な語い情報を附加するために、国語辞典のデータベースから有効な情報を機械的に抽出する方法などもさらに研究される必要があろう。

仮名漢字変換の過程で必然的に得られる単位分割、漢字仮名対応の情報は、日本語文（漢字仮名混り文）の計算機処理にとって不可欠なものであることを考え合わせると、仮名漢字変換入力法は、今後ますます重要な位置を占めると予想され、なお一層の発展が望ま

れる。

おわりに、日頃ご指導を頂く図書館情報大学木澤誠教授に深謝する。

## 参考文献

- 1) 栗原他：仮名文の漢字混り文への変換について，九大工学集報，Vol. 39, No. 4, pp. 659-664 (1967).
- 2) 稲永他：カナ文字文の機械処理のための辞書について，電子通信学会技報，EC 76-45, pp. 1-10 (1976).
- 3) 松下他：漢字かな混り文変換システム，情報処理，Vol. 15, No. 1, pp. 2-9 (1974).
- 4) 相沢他：計算機によるカナ漢字変換，NHK技術研究，Vol. 25, No. 5, pp. 261-298 (1973).
- 5) 藤井他：カナから漢字への変換，電気四学会連合大会予稿，177, pp. 619-622 (1971).
- 6) 絹川他：速記反訳システム，情報処理，Vol. 16, No. 6, pp. 484-491 (1975).
- 7) 木村他：日本語文入力用カナ漢字変換システムの試作，情報処理，Vol. 17, No. 11, pp. 1009-1016 (1976).
- 8) 牧野他：カナ漢字変換の一方法，情報処理，Vol. 18, No. 7, pp. 656-663 (1977).
- 9) 河田他：ミニコンピュータを用いたカナ漢字変換システム，電子通信学会技報，PRL 76-47, pp. 25-34 (1976).
- 10) 横山他：構文解析を用いたカナ漢字変換，情報処理学会全国大会予稿，102, pp. 199-200 (1977).
- 11) 岡田他：べた書き文のカナ漢字変換システム，情報処理学会全国大会予稿，5E-4, pp. 445-446 (1978).
- 12) 内田他：自由入力形式のカナ漢字変換，情報処理学会自然言語処理研究会，27-3 (1981).
- 13) 高木他：技術文書作成用カナ漢字変換システムの設計と試作，情報処理学会全国大会予稿，103, pp. 201-202 (1977).
- 14) 白鳥他：特徴分類形かな漢字変換方式，電子通信学会論文誌，Vol. J64-D, No. 8, pp. 773-779 (1981).
- 15) 田中：姓名のカナ漢字変換システム，情報処理，Vol. 16, No. 3, pp. 230-238 (1975).
- 16) 森他：計算機への日本語情報入力，電子通信学会技報，EC 78-23, pp. 33-41 (1978).
- 17) 牧野他：べた書き文の分かち書きと仮名漢字変換—二文節最長一致法による分かち書き一，情報処理学会論文誌，Vol. 20, No. 4, pp. 337-345 (1979).
- 18) 国立国語研究所：電子計算機による新聞の語彙調査IV, p. 530, 秀英出版, 東京 (1973).
- 19) 大河内他：仮名漢字変換のための文法解析，情報処理学会計算言語学研究会，25-4 (1981).
- 20) 国立国語研究所：電子計算機による新聞の語彙調査, p. 341, 秀英出版, 東京 (1970).
- 21) 国立国語研究所：現代雑誌九十種の用語用字(1), p. 321, 秀英出版, 東京 (1969).
- 22) 牧野他：べた書き文の仮名漢字変換システムとその同音語処理，情報処理学会論文誌，Vol. 22, No. 1, pp. 59-67 (1981).
- 23) 木村：マクロ方式のカナ漢字変換による訳書製作の経験と将来構想，情報処理学会全国大会，2I-3, pp. 869-870 (1981).
- 24) 渡辺他：漢字入力装置の操作性について，電子通信学会技報，EC 77-25, pp. 31-42 (1977).
- 25) 国立国語研究所：分類語彙表, p. 362, 秀英出版, 東京 (1973).

(昭和 57 年 2 月 1 日受付)