

論 文

漢字かな混り文変換システム*

松下 温** 山崎 晴明** 佐藤 文和**

Abstract

A Kana alphabet to Kanji converting system, which is currently at the footlights in the Japanese information processing, is introduced in this paper. A lot of problems appearing in the design of this system, are discussed.

Especially, this system has the following features in order to obtain machine independence of this software system. 1) Data structure of the Kanji dictionary which can be easily updated and enlarged. 2) Compact homonym processing. 3) Machine independence obtained by used of COBOL.

Furthermore, system estimation concerning to mistranslation rate is described.

The result was extremely good as expected.

1. まえがき

近年、報道、出版、人事情報管理などの関係各方面において、急激に漢字情報の計算機処理への要求が高まってきている。コンピュータシステムの普及に伴ない氾濫し始めたカナ、ローマ字表記の伝票や帳票、ダイレクトメールなどは、今ようやく問題としてとりあげられ、日本語情報から漢字を除くことのむずかしさ、それに伴う弊害などから、情報の伝達媒体としての漢字の機能が再認識され、各方面で、種々の日本語情報処理システムや、漢字の入出力デバイスが開発されるに至った。

このような情勢に対処すべく、筆者などにおいても数年前から、主として通信電文の翻訳処理を対象としたシステムの検討を行ない、いくつかの実例をみたが、本稿では昭和47年、情報処理振興事業協会(IPA)からの依頼により筆者等が作成した漢字かな混り文変換プログラム(以下本システムと呼ぶ)の紹介を行なう。なお通信電文の漢字翻訳処理を扱ったシステムとして、昭和46年に藤井などの論文¹⁾が報告されているが、本システムはこれに比し、特定業務に依存しない

点、またソフトウェアの汎用性を特に重視した点が大きく異なると考えている。本文中2では日本語情報を計算機で扱う際に問題となっていた事項を、3では本システムの構成を述べる。また、4では変換処理の方式を同音語処理を中心として述べ、最後に5ではその同音語処理の評価を行なう。

2. 日本語情報を扱う際の問題点

2.1 同音語

一般に、カナ文字表記の語と、漢字記述の語とは1対1では対応しない。1つの音に対しては複数の同音異字が、1つの読みことばに対して複数の同音異語が対応し、その判別は、カナ漢字変換最大の難点となる。

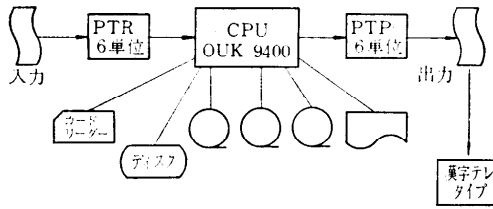
2.2 漢字変換部の抽出

書きことばとしての日本語の特性から日本語文は、分かち書きをしないのが習慣となっている。したがって変換処理を行なう場合には入力記号列中から漢字相当部分を抽出する技術が必要となり、これのために生ずるあいまいさも避けられないものとなる*。

* Kana Alphabet to Kanji Converting System, by Yutaka Matsushita, Haruaki Yamazaki and Fumikazu Sato (Software Division, Oki Electric Industry Co., Ltd.)

** 沖電気工業(株)ソフトウェア事業部

* 本システムでは、漢字相当部分をカッコでくくる方式を採用した。これによって、抽出アルゴリズムによる曖昧さがなく、原稿に忠実な漢字混り文を期待できる。等の利点があった。



使用機器	用途
紙テープリーダ	入力データ読取り
紙テープパンチ	出力漢字混り文打出し
ラインプリンタ	エラーメッセージの出力
磁気ディスク (8 Mbyte/バック)	漢字辞書の格納
漢字プリンタ	出力データの印字
磁気テープ	ワーキング用
カードリーダ	辞書データの入力

図 1 ハードウェア構成

Fig. 1 System Organization

2.3 収録語彙*

たとえば、新聞紙面上のスポーツ欄と政治欄とを比較すると明らかのように、業務種別の違いにより用語の使われ方が非常に異なる²⁾。日本語における熟語の数は膨大なものであり、それらの語をすべて変換用の辞書に収録するのは不可能に近い。したがって多くの変換システムは、業務ディペンデッドなものとならざるを得なかった。

3. システム構成³⁾

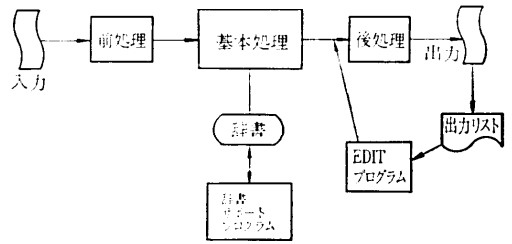
3.1 ハードウェア構成

本システムの開発は、OUK 9400 計算機を用いて行われた。使用言語はコボルおよびアセンブラである。図1にハードウェア構成を示す。なお入出力には6単位紙テープを、中間結果の記録には磁気テープおよび磁気ディスクを使用している。

3.2 ソフトウェア構成

本システムのプログラム構成は、主として利用者オリエンティッドな手続きを取扱った前処理および後処理、漢字混り文への変換を行なう基本処理、漢字辞書

* 本システムでは、熟語を、比較的高頻度で使用され、かつ特定の専門分野に偏らない一般語と、業務独特の固有名詞、専門用語、略語から成る特殊語の2種に分け、それぞれ汎用辞書、私用辞書と呼ばれる媒体に収録した。これにより辞書の汎用化を計ることができた。なお一般語の選定に当たっては国立国語研究所の語彙調査⁴⁾を参考にした。また私用辞書は頻繁な削除、更新が予想されるため辞書保守用のユーティリティーを充実させた。また本論文の草稿作成中に、甲賀等⁵⁾によるカナ漢字変換システムが発表された。このシステムにおいても、辞書の区分が試みられているが、本システムの汎用辞書に該当するものがないかわりに、私用辞書をさらに分野別に構成したものとなっている。



モジュール	ステップ数 (K step)
前処理プログラム	1
基本処理プログラム	5.3
後処理プログラム	0.35
辞書サポートユーティリティー	4.8
エディットプログラム	2.5

図 2 プログラム構成

Fig. 2 Software System Organization

の保守運営を行なう辞書サポートユーティリティー群、誤字訂正の機能を備えたエディットプログラムから成る。

3.2.1 前処理プログラム

利用者の入力コードから、基本処理で扱う標準コードへの変換、ローマ字文からかな文への変換を行なうプログラムで、コボル記述のメインプログラムと、紙テープデータの読取りを行なうアセンブラ記述のサブプログラム、およびコード変換テーブルから成る。

3.2.2 基本処理プログラム

かな漢字変換処理そのものを実行するプログラムで全処理がコボルで記述されている。プログラムは各機能単位で4つのフェーズに分割され、各フェーズは、実行時にオーバーレイされる。

1) 制御プログラム

基本処理内の各フェーズプログラムの制御、オーバーレイの実行を行なう。

2) 入力文チェックプログラム

前処理プログラムにより磁気テープに出力されたコード変換済の入力文を読み込んで、データ形式のチェックを行なう。エラー出現ときにはメッセージを出力し、継続不可能なエラーの場合には処理を中断する。

3) 辞書索引プログラム

入力文中のカナ文字をキーとして、ディスク上の木構造に展開された見出し部を検索、該当する同音熟語を全て出力する。

4) 同音語処理プログラム

辞書索引プログラムにより引き継がれた同音熟語のうち、入力文中で最も良く適合するものを選択を行なう(詳細は後述する)。

5) 出力編集プログラム

変換された漢字混り文、同音漢字のリストを、漢字プリンタの出力形式にあわせて編集する。

出力データは磁気テープに格納され、後述する後処理プログラム、エディットプログラムの入力となる。

3.2.3 後処理プログラム

基本処理で出力されたデータを読み込み、利用者の使用している漢字プリンタコードへの変換を行なう。

出力は、紙テープにさん孔され、オフラインで漢字プリンタを使用、出力文の印字が行なわれる。

プログラムは、コボル記述のメインプログラムと、紙テープのさん孔を行なうアセンブラサブプログラムおよび出力コード変換テーブルより成る。

3.2.4 辞書サポートユーティリティ

カードベースで作成された辞書データのディスクへの登録、更新などを実行するユーティリティ群で、次の3つの独立したプログラムより成る。なお、記述言語は全てコボルである。

1) 辞書リストプログラム

カードベースの辞書データのフォーマットチェックおよび辞書リストの作成を行なう。

2) 辞書登録プログラム

チェック済みの辞書データをディスクに登録、見出し部の木構造展開と、熟語の格納を行なう。

3) 辞書更新プログラム

ディスクに登録された辞書に対する熟語の追加、削除、一部変更などを実行する。

3.2.5 エディットプログラム

印字された漢字混り文中に、同音語の誤りや、文の変更箇所、そう入文の必要箇所などが発見された場合、修正データにより、基本処理の出力磁気テープに修正を加えるプログラムである。修正済磁気テープは再度後処理プログラムにより処理され、完全な出力文を得ることができる。

4. 変換方式

一般に、漢字変換の方式には、語単位変換* と、熟語単位変換** とが有る。前者は辞書の容量も少なく、

* 漢字1文字を変換の対象とするもの。

** 日本語として意味をなす漢字語単位で変換するもの(従って1文字の熟語も有り得ると考える)。

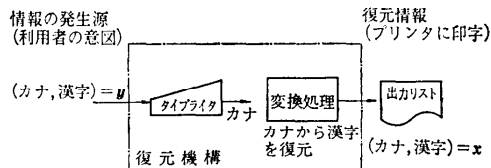


図3 カナ漢字変換システムのモデル

Fig. 3 A model of Kana-Kanji conversion system

処理も簡単であるため、漢字入力端末装置などで採用されている方式である。これに比し、後者は、辞書も大規模となり、処理も複雑化するが、漢字変換の最大の難点である同音異字(語)の出現については、大幅に減少することが予想される⁹⁾。

以下では、この同音語の出現頻度、語単位変換と、熟語単位変換との効率の比較について述べる。

4.1 語単位変換と熟語単位変換との比較

漢字変換処理を、図3のようにモデル化する。このように、変換効率を評価するため、復元機構と名づけたモデルを想定しているのは以下の理由による。

まず、情報の発生源での利用者の意図する情報とは必ずカナと漢字の対でなくてはならない。なぜならこれをカナのみとするとたとえばコウギョウセイヒンのような表現の場合「鉱業製品」と「工業製品」のうちどちらが正しく変換された語であるかを判断する明確な基準がなくなる。つまり、漢字変換あるいは同音語選択の成否は、利用者が初期に意図した語が達成されたか否かによってのみ判断されるといえる。このようにして、カナ漢字変換を、データ発生源例えば原稿における利用者の意図情報を印字用紙の上に復元する機構であると再定義することができる*。したがってこの復元機構には、原稿に基づいてカナ表現データを作成し変換処理にインプットするという人手を介した手続きも含まれる。なお変換処理とは、カナから漢字への変換を行なうハードウェア、ソフトウェアの総体をいう。こうして同音語によって生ずる復元情報と発生情報の不一致は復元機構内部での雑音と考えることができる。もちろんこのとき復元情報もカナと漢字の対から成るものとする**。

* この定義により従来、入力データの一様性の仮定の基に想定されていた同音語数の平均値等の統計量が、より正確になる。なぜなら一様性を仮定し、入力はカナのみ、出力は漢字のみの変換処理の部分だけを考察の対象としたのであれば、例えば「コウ」、「ショウ」のような対応する漢字が60程度にも及ぶカナ読みと、「ミナモト」(源)のような1漢字のみが対応するカナ読みと同一の重みを持つという不合理が生ずる。従って一様性は漢字とカナの対の上で仮定するのがより適切であると考えられる。

以降発生データ y は、使用頻度による偏りがないこと（一様性）を仮定し変換処理を語単位変換によったとき、熟語単位変換によったときのそれぞれにおける復元機構の定量評価を行なう。

以下の記法を用いる。

i) X で漢字とそれに対応するカナ読みとの対の全体をあらわす。したがってたとえば（マツ，松）と（ショウ，松）とは異なるエレメントとして表現される。

ii) X の任意のエレメントを z としたとき z は (z_1, z_2) のように表記され、 z_1 は z のカナ部、 z_2 は z の漢字部をあらわす。

iii) $\forall x \in X$ に対し $E_x (C X)$ を次のように定義する。

$$E_x = \{z | x_1 = z_1\} : x \text{ の同音語全体}^*$$

iv) $\forall Y \subset X$ に対し、 $|Y|$ で Y に属するエレメントの総数をあらわす**。

出力 x が得られたと仮定したとき発生データが y である確率 $P_x(y)$ は、一様性から

$$P_x(y) = \begin{cases} \frac{1}{|E_x|} & (y \in E_x) \\ 0 & (y \notin E_x) \end{cases}$$

となる。

故にその条件付きあいまい度は

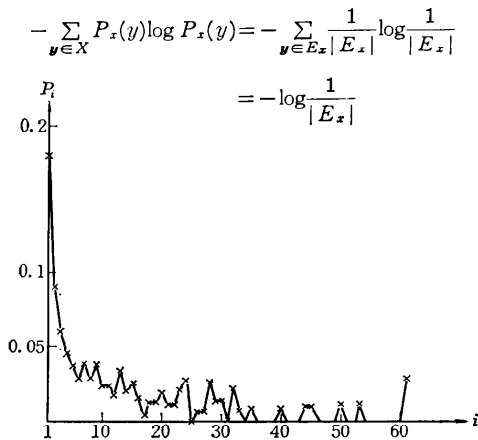


図 4 語単位変換法による同音字分布

Fig. 4 A distribution of homonymous characters

** このように仮定しても一般性を失わないことは、例えば漢字を印字後、必ずその後へ入力された読みがなを印字するという機能を変換処理のソフトウェアに付加した場合を考えれば明らかである。ただし復元機構の内部でカナ表現が変化する確率は0であるとす。

* a と b とが同音語であるとは $a_1 = b_1$ であることと定義する。

** 従って $|E_x|$ は x の同音語数をあらわす。

表 1 語単位変換時の同音漢字出現度

Table 1 The rates of appearances of homonymous characters.

調査対象 延べ漢字総数 4354 (当用漢字含む)
発音種別 1327 通り (発音と漢字のペア)
 $i=30$ 以上で $P_i=0$ のものは省略。

i	P_i	i	P_i	i	P_i
1	0.178686	16	.014699	32	.22049
2	.092329	17	.003904	33	.007579
3	.062012	18	.012402	35	.008039
4	.045935	19	.013091	40	.009187
5	.036748	20	.018374	44	.010106
6	.027561	21	.009646	45	.010335
7	.038585	22	.010106	50	.011484
8	.027561	23	.021130	53	.012173
9	.037207	24	.027561	61	.028020
10	.022967	25	.000000		
11	.022738	26	.005972		
12	.016537	26	.006201		
13	.032843	28	.025723		
14	.019293	29	.013321		
15	.024116	30	.013780		

あいまい度: 3.53 bit 同回同音数平均: 13.64 語

表 2 熟語単位変換時の同音語出現度

Table 2 The rates of appearances of homonymous words.

調査対象 延べ漢字語総数 14,200 語
見出総数 11,465 語

i	P_i	i	P_i	i	P_i
1	0.704225	11	.002324	24	.001690
2	.137465	12	.002535	26	.001831
3	.057465	13	.002746	30	.002113
4	.029014	14	.003944	44	.003099
5	.001408	15	.002113		
6	.006338	16	.001127		
7	.004437	17	.004789		
8	.005634	18	.001268		
9	.005070	19	.002517		
10	.002113	21	.001479		

あいまい度: 1.32 bit 同音語数平均: 2.32 語

したがって、復元機構の持つあいまい度は、 x が実現する確率 $P(x)$ で平均して

$$-\sum_{x \in X} P(x) \log 1/|E_x| = -\sum_{x \in X} 1/|X| \log 1/|E_x|$$

(出力 x の一様性を仮定) さらに $|E_x|=i$ なる x の集合を A_i とすれば上式は $-\sum_{i, x \in A_i} 1/|X| \log 1/|E_x| = -\sum_i |A_i|/|X| \log 1/i = -\sum_i P_i \log 1/i$ となる。

ここで P_i は、一様性の仮定のもとで任意のカナ、漢字の対を指定したとき、それが i 個の同音語を持つ確率をあらわす。また復元機構により生ずる同音語数の期待値は $\sum_i i P_i$ となる。図 4、図 5、表 1、表 2、

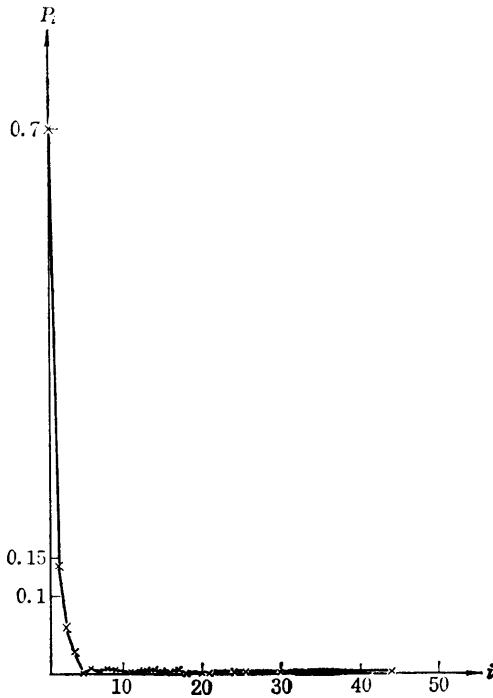


図5 熟語単位変換法による同音語分布
Fig. 5 A distribution of homonymous words

に語単位変換，熟語単位変換の場合のこれらの値を示す。

4.2 本システムでの変換方式

本システムでは，システム利用者にかかる負担の削減，変換能率の向上を考慮して，語単位変換方式を採用している。さらに同音語については，使用頻度により重み付けがされ，また，送り仮名解析および，一種の複合語処理である接辞処理により一層，同音語によるあいまいさの減少を計っている。

以下に同音語処理の概略を述べる。

本システムにおける同音語処理は，辞書索引によって得られたいくつかの同音語のうち，該当語の属する文法的なカテゴリーを決定し，もし同音語中に同一カテゴリーのものがあれば，使用頻度の高いものを出力する。カテゴリーとしては，概略国語文法の品詞分類に従っているが，特に語数の多い名詞を細分割することが試みられている。

4.2.1 後続文字の属性

一般に，漢字熟語が文中に出現する際には，後続する文字は，次の3つの属性を取り得る。

表3 文法カテゴリー
Table 3 Grammar Categories

文法カテゴリー	例
数	一, 二, 三, 百二十
人名	佐藤, 山田
地名	日本, 北野
普通名詞	山, 机, 動作
サ変名詞	運動, 実行, 愛 (「する」がつくもの)
ザ変名詞	感(ずる)
人名接辞	氏, 君, 殿
地名接辞	市, 町, 村, 路
数詞接辞	日, 分, 円
他接辞	会, 長, 落
形容詞	美しい, 明か(い)
形容動詞	静か(だ)
変形動詞	見(る), 居(る)等の上一もしくは, 下一語用の動詞
カ五動詞	行(く), 引(く)
ガ五動詞	研(ぐ), 泳(ぐ)
サ五動詞	話(す), 殺(す)
タ五動詞	打(つ), 持(つ)
ナ五動詞	死(ぬ)
バ五動詞	飛(ぶ)
マ五動詞	苦し(む)
ラ五動詞	走(る), 戻(る)
ワ五動詞	言(う), 向(う)
その他	連体詞, 副詞等の他の品詞

- 1) 後続の文字が漢字 例 (佐藤) 氏
- 2) 後続の文字が仮名 例 (明) かるい
- 3) 後続の文字が英数特殊文字 例 (三百) m
なお3)には，文章が漢字熟語で終了する場合も含まれる。

この後続の文字属性を判別するのは，本システムにおいては，特に容易である。本システムへの入力データには，漢字，かたかな，英字などを指定するため，必ず，制御符号がそう入される。すなわち，制御符号“<”でくくられた部分が漢字に変換されるべき部分であり，“((”のときかたかな，“/”のとき英字，制御符号がない場合は，ひらがな，数字，特殊記号などである。

図6に文法カテゴリー決定までの論理ツリーを示す。

4.2.2 送りがな解析処理

後続の文字が，カタカナもしくはひらがなであった場合，文法ツリーの索引を行なう。文法ツリーのさくいん形式を図7に例示する。この例は次のことをあらわす。まず漢字熟語に続くカナ文字が“カ”であった場合，該当漢字語がカ行五段活用をする動詞である可能性が強いことを示す。このときのカテゴリー“カ五”はワーエリアにセーブされる。次に“カ”に続くカナ文字が“ッ”，“ロ”，濁点以外のものであればマッチ

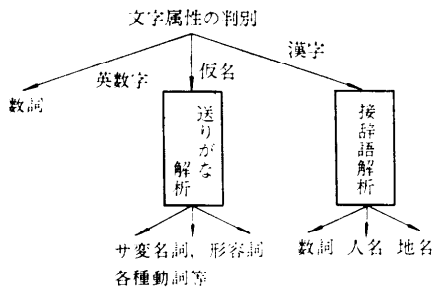


図 6 文法カテゴリーの決定アルゴリズム
Fig. 6 The character category decision process

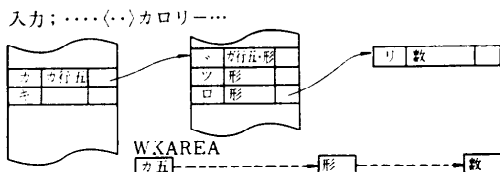


図 7 文法辞書さくいん例
Fig. 7 An example of the grammar information search.

ングはとぎれ、該当熟語の文法カテゴリーは、カ行五段動詞であると決定され、同音語中でカ行五段動詞がある場合は、第1に優先される。この例では、“カロ”と続いたと仮定する。この場合には、形容詞の活用語尾である可能性がより強く、この情報は先のワークエリアにオーバーレイされる。ここでマッチングがとぎれれば、決定されるカテゴリーは、形容詞である。さらに“カロ”に続く文字が“リ”であれば、これは、カロリーという単位を意味し、数詞に接続する。したがって該当熟語は数詞と判断される。(例 百カロリー)

4.2.3 接辞語処理

本システムでは、名詞(数詞, 人名, 地名, 他)と接辞語(接数, 接人, 接地)との連結関係の処理が行なわれる。図8に優先的に処理される連結関係を表わす接辞テーブルを示す。たとえば、図9に示すような入力文に対し、接辞テーブルの参照により最適のペアを見つける訳である。ただし2つ以上のペアが接辞テーブル上に存在する場合はテーブルの上位のものが、いいかえれば結合の強いと考えられるものが優先する。

5. 評価

本システムの一般文章における変換効果は、前述し

<例>

結合の強いもの	数詞	接数	三百・円
	数詞	数詞	三百・五十二
	人名	接人	佐藤・様
	人名	人名	(徳川・家康)
	地名	接地	鹿児島・県
	地名	地名	東京都・港区
書名	接書	国際・会議	

図 8 接辞テーブル
Fig. 8 Contents of a 'setsuji' table.

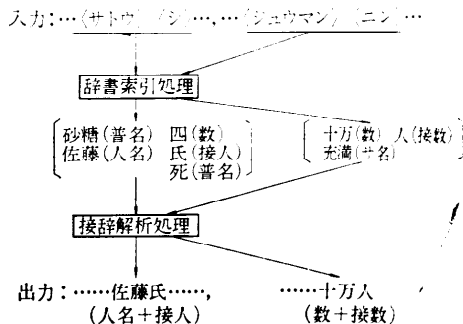


図 9 接辞解析の例
Fig. 9 An example of a 'Setuji-shori', which discriminates an optimum pair of words using a Setuji-Table.

た種々の同音語処理の総合的結果によるものと考えられる。この同音語処理の効果を測定するために、もっとも標準的な日本語文と考えられる新聞から、ランダムに記事を抽出し、変換処理を行なった。漢字変換標本数 360 のうち、正しく変換された漢字は 87% である

表 4 変換率

Table 4 Conversion program efficiency.

同音語数 (n)	使用率 (u)	正答率 (c)	[u×c]
1	64.8%	99%	0.642
2	14.5%	87%	0.126
3	6.5%	80%	0.052
4	7.1%	52%	0.037
5	0.7%	50%	0.004
6以上	6.4%	50%	0.032

標本数: N
同音語数 n の出現個数: d_n
同音語数 n の正当数: t_n

$$u = \frac{d_n}{N}$$

$$c = \frac{t_n}{d_n}$$

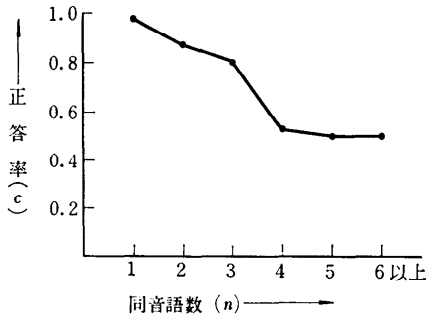


図 10 同音語数と正答率の関係

Fig. 10 Relation between the number of homonymous words and the number of correct conversion.

った。(正答率)

次に同音語数と正答率の関係、さらには正答率を使用頻度で重みづけを行なった場合の同音語数との関係が評価された。表 4, 図 10, 図 11 にこれを示す。

但し、使用した辞書は汎用辞書のみであり、また未変換語 (対応する漢字が辞書に登録されていない) は除いて標本とした。また同音語数が1のもの (対応する漢字語が1熟語のみ) に対し、正答率が100%となっていないのは、実際には同音語が有るにもかかわらず、それがきわめて低頻度の特殊な用語であるため、一般語として、汎用辞書に収録されておらず誤変換に

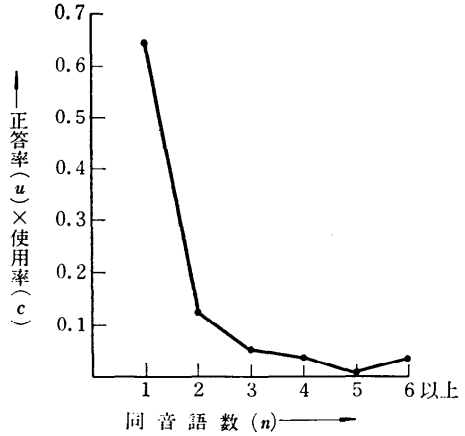


図 11 同音語数と正答率使用率の関係

Fig. 11 Relation between the number of homonymous words and the product of correct conversion and reference rate.

なったものが存在するためである。

代わりに、本システムでの変換例を図 12 に示す。

6. あとがき

本プログラムの開発は、41 年度までに九州大学で行なわれた基礎的研究、およびそれをベースにした当社の研究所での実績をもとに¹⁰⁾汎用化を目的としてソ

<p>▽ 2 ◇ はんだんりよく 終</p>	<p>▽ 1 ◇ うちゆうせん 了</p>	<p>◇ 14 ◇ 時期 ◇ 13 ◇ 作 ◇ 12 ◇ 策 ◇ 11 ◇ 科学 ◇ 化学 ◇ 制作 ◇ 化学</p>	<p>◇ 10 ◇ 伸 ◇ 9 ◇ 乗 ◇ 8 ◇ 機 ◇ 7 ◇ 動 ◇ 多 ◇ 兵 ◇ 器 ◇ 6 ◇ 破 ◇ 5 ◇ 城 ◇ 4 ◇ 国 ◇ 3 ◇ 家 ◇ 2 ◇ 製 ◇ 1 ◇ 品 ◇ 工 ◇ 業 ◇ 多 ◇ 豊 ◇ 富</p>	<p>◇ 10 ◇ 乗 ◇ 9 ◇ 軌 ◇ 8 ◇ 道 ◇ 覆 ◇ 平 ◇ 7 ◇ 氣 ◇ 破 ◇ 6 ◇ 城 ◇ 国 ◇ 5 ◇ 歌 ◇ 4 ◇ 漁 ◇ 3 ◇ 業 ◇ 2 ◇ 覆 ◇ 抱 ◇ 1 ◇ 負</p>	<p>◇ 野 ◇ 延 ◇ 棋 ◇ 併 ◇ 記 ◇ 欽 ◇ 業 ◇ 興 ◇ 業</p>	<p>◇ 本 ◇ 文 ◇ 1 ◇ 1 ◇</p>	<p>◇ 漢 ◇ 字 ◇ かな ◇ 混 ◇ り ◇ 文 ◇ 変 ◇ 換 ◇ シ ◇ ス ◇ テ ◇ ム</p>
--------------------------------	-------------------------------	---	---	---	--	--	---

図 12 変換例
Fig. 12 An example of conversion

ソフトウェアシステムとして開発することができた。

本プログラムは、従来各方面で開発されたカナ文字漢字変換システムの中で、プログラムの記述言語、辞書形式などの方向から特定業務、特定分野に依存しないよう配慮した点が、新しい試みであったと考えている。現在の辞書は汎用辞書だけが作成されているが、さらに業種対応に私用辞書を設けることにより、変換率の一層の向上が見込まれる。

なお、汎用辞書には一般語が約 15,000 語収録されているが、一般語の定義そのものがはなはだあいまいであること、また汎用辞書に収録されていないもので、一般語と考えられる熟語の有無など、いまだ未検討の部分が残っているし、またこのことは、本システムの作成に当って最も労苦の多かった部分でもあった。

おわりに筆者等がカナ漢字変換を検討する際、多くの御教示、御指導をいただいた九州大学工学部の故栗原教授に、また本システムの開発に当って多大の援助と励ましを与えていただいた沖電気の黒崎悦明氏に、また情報処理振興事業協会および関係各位の方々に、深く感謝する次第である。

参考文献

- 1) 藤井他：カナから漢字への変換，昭和 46 年電気四学会連合大会講演，No. 177.
- 2) 国立国語研究所：現代雑誌九十種の用語用字(1)，総記および語彙表，国研報告 21.
- 3) 国立国語研究所：現代雑誌九十種の用語用字(2)，漢字表，国研報告 22.
- 4) 国立国語研究所：現代雑誌九十種の用語用字(3)，分析，国研報告 25(特に語の優先度に関連するものとして，pp. 7~54).
- 5) 国立国語研究所：電子計算機による新聞の語彙調査，国研報告 37.
- 6) 国立国語研究所：電子計算機による新聞の語彙調査(II)，国研報告 38.
- 7) 甲賀他：漢字カナ文字変換の事例研究，昭和48年日本情報処理開発センター調査研究発表会.
- 8) 松下，山崎：漢字情報処理システム，電子通信学会，資料番号 EC-59，1972年1月.
- 9) 黒崎，松下：かな漢字変換システムの評価，電気学会，資料番号 IP-71-16，1971年10月.
- 10) 栗原，黒崎：カナ文の漢字かなまじり文への変換，九州大学工学集報.

(昭和 48 年 8 月 11 日受 付)

(昭和 48 年 10 月 4 日再受付)