



## 論文

## 速記反訳システム\*

絹川 博之\*\* 筒井 健嗣\*\*  
小田切 郁夫\*\*\* 木村 睦子\*\*\*\*

## Abstract

A stenograph to Japanese translation system, which is one of the Japanese Information Processing systems, is introduced in this paper. Input data to this system are phonetic symbols expressing spoken Japanese and are not divided into Japanese words.

The main feature of this system is that it translates the above-stated input string into normal Japanese sentence written in Kanji and Kana. Then, we have adopted a method of morphological analysis as translating algorithm and have made the dictionary by investigating spoken Japanese words.

After construction of the program system, we estimated mistranslating rate and processing-speed.

The mistranslating rate was 4% and the translating rate of Chinese borrowed phonetic words into Kanji was 67%. The processing-speed was 155 msec/word.

## 1. ま え が き

近年、コンピュータシステムの適用範囲の拡大、それともなうシステムの入出力情報に関する人間の増大から、より人の側に歩みよったシステムが要請されてきた。この一環として漢字を情報伝達媒体とする情報システム、漢字入出力機器が、種々開発されてきている。漢字出力装置の水準は、各メーカーの努力でかなり高いが、文字種の多さから入力装置・方式については、まだ決定的なものはない。このような状況に対処すべく、筆者らにおいても数年前から日本語情報処理システムの検討を行ってきた。本稿では、昭和46年以降、筆者らが試作を行なった速記用特殊タイプライタ：ソクタイプを用いた速記反訳システム（反訳：速記を元の言葉に直すこと）の紹介を行なう。なお、類似システムとして、九大<sup>1)</sup>、沖電気（株）<sup>2)</sup>、

NHK<sup>3)</sup>、<sup>4)</sup>、JIPDEC<sup>5)</sup> 等で、カナ漢字変換システムが開発されている。本システムは、これらに比し、会話速記という話し言葉を対象にしている点、言葉としての区切りが明確でない表音記号列から漢字かな混り文へ変換する点が、特徴であると考えている。また、将来音声認識と結合して、音声のハードコピー作成システムの基礎となる技術と考えられる。本文中2では、速記の特性と日本語情報処理、3では、反訳方式、4では、本システムの構成、5では、システムの実験結果について述べる。

## 2. 速記特性と日本語文出力

## 2.1 ソクタイプ速記の特性

ソクタイプによる速記記号列は、基本的な音節（例、「ア」、「シャ」、「コウ」etc.）を、区切りなく羅列的に並べた音連続である。この音連続は、速記効率を上げるため次のような打鍵規則がある。

- (1) 慣用句、繰り返し語の略表現
- (2) 1個の記号が複数個の音を表わすこと（例、「ジョ」、「ゴゾンジ」を同一の記号が表わす）
- (3) 濁音の清音記号表現（例、寝言→「ね」+「こ

\* Stenograph to Japanese Translation System by Hiroshi KINUKAWA, Kenji TSUTSUI (Systems Development Lab. Hitachi Ltd.), Ikuo ODAGIRI (Facom Hitac Ltd.) and Mutuko KIMURA (Institute of Behavioral Sciences).

\*\* (株)日立製作所システム開発研究所

\*\*\* ファコム・ハイタック(株)システム本部

\*\*\*\* (財)計量計画研究所言語情報研究室

と) ]

- (4) 表わす訳の一部に括弧書き部分を有する(例.  
((シ)テ)クル: 前後の連なり方から「シテ  
クル」, 「テクル」, 「クル」のいずれかを表わす)
- (5) 「動詞・形容詞+助動詞」の原形略語羅列表  
現(例. 「開かざる」を「ヒラク」+「ザル」と表  
わす)

また、速記効率を上げるため、規定の略し打ちの他  
に、タイピストは、各々個人特有の略し打ちをする。

## 2.2 日本語文出力変換項目

- (1) 形容動詞・普通名詞の音読み語の漢字化
- (2) 外国語・外来語の片仮名表記化
- (3) 数字データの漢数字化
- (4) 助詞〈オ・エ・ワ〉の〈を・へ・は〉化
- (5) 句読点の挿入
- (6) 前節 2.1 の打鍵規則を処理して正常な日本語  
文に直すこと

## 2.3 反訳の際の問題点

上記 2.1 の特性を有する速記記号列から日本語文に  
変換する際の問題点をまとめると次のようになる。

- (1) 単位語の認定: 区切りのない音連続表現から  
単位語の認定・切り出しを行なうこと。
- (2) 同形語の判別: 同一の速記記号列で表わされ  
る言葉のことをいう。漢字化における同音異義  
語のみならず速記の特性故に発生する同形異義  
語の判別。
- (3) 語彙の収集: 本システムは話し言葉を処理の  
対象としており、辞書収録の言葉と、その速  
記記号による表わし方(略し打ちを含む)の調  
査を目的として、事前に総計約 15 万語の語彙  
調査を行なった。言葉の調査としては、本調査  
では、不十分であるので、国立国語研究所の調  
査結果<sup>9)~11)</sup>を併用し、その速記記号表現法は、  
規定の表現法に応じて定めた。

## 3. 反訳方式

### 3.1 単位語の認定

一般に区切りのない表音記号例(カナ文字列、速記  
記号列 etc.) から、言葉を切り出す方式として、辞書  
の最長一致探索法が考えられる。この方法は音連続が  
長くなれば言葉として一意に決まりやすくなること  
から考えられることである。しかし、音連続に対する  
単位分割は複数存在しうること、および同形異義語の  
存在から辞書中の見出しの長い言葉で単純に置き換え

る方式はおのずと処理精度に限界がある。したがって  
何らかの情報を利用して前後の接続関係を分析し、適  
正な単位分割を選択することが必要となる。一般に  
は、言語の規則性を利用し、言語の規則性には、文節  
内の形態上の規則性から、文節間の構文上の規則性ま  
で存在する。筆者らは、形態上の規則性に着目して処  
理することにし、これを“形態素分析法”と呼び、以  
下にその内容について述べる。

#### 3.1.1 形態素分析法

形態素とは、品詞分類、語形変化を扱かう形態論に  
おいて意味を有する最小単位である。言い換えると、  
形態素は接続の際の規則的变化に着目して分類した言  
葉の単位ということができる。この接続時の規則性の  
例としては次のようなものがある。

- (1) 用言の活用形と助動詞の接続: たとえば助動  
詞「ない」は、動詞の未然形に接続する。
- (2) 自立語と付属語の接続: たとえば格助詞「を」  
は、名詞・代名詞におもに接続する。

筆者らは、この形態に関する規則性を利用して修正  
最長一致を行なうことにし、ソクタイプ速記の表わす  
音連続をすべて形態素列でうまく表現できるよう形態  
素を定めた。

#### 3.1.2 品詞の設定

日本語文法、ソクタイプの打鍵規則および語彙調査  
結果に従って形態素を定めたが、その際の原則を以  
下に述べる。なお筆者らは、この形態素を品詞と呼ん  
でいる。

- (1) 国文法で設定されている品詞を基本に前後に  
来うる語との接続形態がともに等しいもののみ  
を集めて語類をつくり、これを品詞とする。
- (2) 前との接続が等しくても、後続語が異なるも  
の(たとえば、五段動詞の未然形と連用形)、  
後続語とは接続の仕方が等しくても、ソクタイ  
プの打鍵規則を処理する上で、前との接続の仕  
方が異なるもの(たとえば、上一段活用動詞を  
表わしうる略語で、「(シ)スギル」, 「((シ)テ)  
ミル」, これは括弧書き部分の処理が異なる)  
については、別の品詞を設定した。
- (3) 同形異義語を少なくするため、原理的には、  
品詞 a, b の 2 つで事足りる場合でも、a かつ  
b という語が多数存在すると予想される場合  
“a∩b” という品詞を設定した。この便宜は 3.  
1.5 で述べるが形態素分析の *delimiter* (以下  
*del.* と略す) としての言葉を認定するためのも

Table 1 Verbal Categories

項番	名称	例	項番	名称	例
1	サ変動詞語幹名詞	案内, 勉強	13	カ, タ, ラ, ワア行五段連用形②*	走っ, 扱っ
2	サ変動詞語幹副詞	はっきり, ぼさっと	14	ナ, マ, ガ, バ行五段連用形②*	引きつい, 盛ん
3	五段動詞未然形①*	扱わ	15	ラ行五段語幹	帰
4	五段動詞形容詞未然形②**+ウ	書こう	16	動詞終止連体形	払う, 捨てる
5	コ (動詞・補助動詞)	来 (カ変動詞)	17	動詞, 形容詞仮定形バ	取れば, 探せば
6	上一段, 下一段未然連用形	受け, 起き	18	動詞命令形	置け, 立て
7	セ (サ変語尾, サ行五段語尾, 使役)	せ	19	オイデ, ゴラン, オ+動詞連用形	お会い
8	シ (サ変・サ行五段語尾, 使役, 接続動詞)	し	20	一般活用動詞基本形**	考える, 決める
9	スル	する	21	ラ行五段活用動詞基本形	入る, わかる
10	サ行五段連用形	返し, 押し	22	一段, ラ行以外の五段動詞基本形	うかがう, 望む
11	サ行五段語幹	返	23	キ (動詞, 補助動詞)	カ変「くる」の連用形
12	五段動詞 (サ行除) 連用形①**	通い, 泳ぎ	24	サ変動詞語幹	開知, 開

\*1 「ない」に連なる形      \*2 「う」に連なる形      \*3 「ます」に連なる形      \*4 「た」に連なる形  
\*5 「だ」に連なる形      \*6 基本形は略語として存在するものをいう

のである。この例は、「サ変動詞語幹名詞」,  
「サ変動詞語幹副詞」の2つである。

(4) ソクタイプの変換規則や略語を考慮して品詞を設定する。たとえば、略語に「カット」を表わすものが存在するので、形容詞連用形 (例。「美しかった」) を設けず、形容詞語幹という品詞を設けた。

このようにした結果、品詞は全部で 203 個となった。そのうち動詞関係の品詞を Table 1 に示す。他は、スペースの都合上割愛する。なお、ここで動詞基本形という品詞は、3.3 で述べる語尾変化処理の都合上設定したものである。

### 3.1.3 文法表

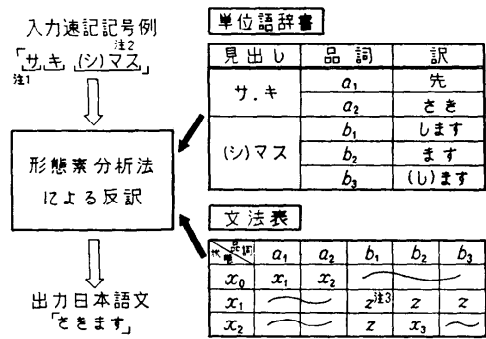
各品詞間の接続の可否を定めた規則は、有限状態文法であり遷移表の形式にまとめた。終端記号としては 3.1.2 で定めた品詞とし、非終端記号としての状態は文節内の文法的状態を表わすものとした。たとえば、「名詞」の後に助詞「ガ」がくると、文節は『「名詞」+「ガ」』という状態となる。なお、遷移後の状態として句読点挿入状態を設定し、出力の要求に対処した。状態数は、全部で 99 である。

### 3.1.4 単位語辞書

単位語辞書は、速記記号列に対して言葉を割り当てるための辞書である。本論文では、形態素分析法の効果の見極めを目的としているので、各タイプ固有の略し打ちについては、辞書に収録していない。

### 3.1.5 形態素分析アルゴリズム

形態素分析のアルゴリズムについて Fig. 1 を例として説明する。入力速記記号列が「サ・キ・(シ)マス」であるとす。単位語辞書を調べると、速記記号列「サキ」と「(シ)マス」には、それぞれ同形異義語



注1. ーが1個の速記記号の表わす音  
注2. 訳の一部に括弧書きを有することを示す。  
注3. z印: 接続不能を示す。

Fig. 1 Algorithm of Morphological Analysis

が存在する。品詞は、 $a_1, b_1$  で表わし、添字の小さいものほど優先順位が高いとする。 $x_2$  は状態を表わし、 $x_0$  は初期状態を示す。

- 辞書探索で「サキ」に対して「先」を表わす「 $a_1$ 」を得る。 $a_1$  に対する分析は、  
 $x_0 \rightarrow a_1 x_1, x_1 \rightarrow b_1 z, x_1 \rightarrow b_2 z, x_1 \rightarrow b_3 z$   
となり、「先」には接続するものがない。
- 次に、品詞  $a_2$  の「さき」について調べると  
 $x_0 \rightarrow a_2 x_2, x_2 \rightarrow b_1 z, x_2 \rightarrow b_2 x_3$   
となり、品詞  $b_2$  の「ます」が接続する。その結果として、出力の日本語文は、「さきます」となる。

形態素分析アルゴリズムの概要を示したが、処理を連続的かつ効率的に行なうために次のような工夫をした。まず、速記記号列の表わす音連続に対する言葉の区切り方は複数個存在しうること、および同一の速記

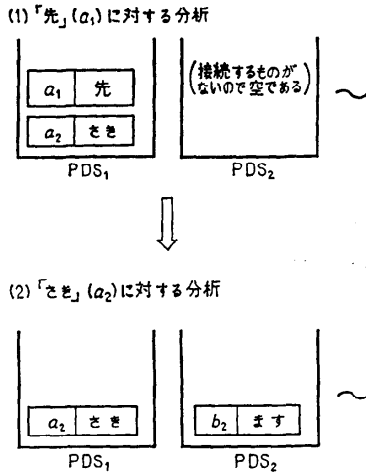


Fig. 2 An Example of Push Down stack

記号列については、上例にあるように同形異義語が存在しうる。したがって、ある言葉に対して接続しうるものは複数個ある。これら接続可能な言葉をすべてPDS (プッシュダウスタック) に蓄積させることにした。上記(1), (2)について蓄積例を Fig. 2 に示す。

この結果、上例のように優先順位の高い言葉 (PDS では上に蓄積される) について接続するものがないとき、次の順位の言葉についてすぐに分析が行えるようにするとともに、辞書探索で見出しの先頭を同じくするものについて一度の探索ですむようにした。またこれを実現するために、単位語辞書の言葉は見出しの先頭を同じくするものの中では、優先順位の低いものが、先に現われるよう配列されている。言葉の区切り方を正しいとして決定とみなすのは次のいずれかの時点で、*del.* となりうる言葉 (単位語辞書にその旨表示してある) が、PDS の最上位に存在する場合である。すなわち、

- (a) PDS に、見出しの先頭を同じくする言葉で接続可能な言葉をすべて蓄積し終った時点
- (b) バックトラック時、以後に接続可能な言葉がなくて PDS の最上位の言葉を取除いた時点

また、*del.* となりうる言葉とは、速記記号列の表わす言葉が一意的に定まり、かつその言葉の意味も明確であるような言葉のことをいう (例. 弁護士, 取締役 etc.). この *del.* となりうる言葉は、語彙調査結果 (2.3 参照) により決めた。したがって形態素分析法による単位語への区切り分析は、*del.* となりうる言葉

から次の *del.* となりうる言葉の範囲で行なっているといえる。またどのような区切り方を行なっても接続不可のときは、次の *del.* となりうる言葉を表わす速記記号列まで形態素分析をしないで、各速記記号の本来もつ訳を羅列することにした。なお、3.1.2 で述べた「 $a \cap b$ 」なる品詞は、*del.* となりうる言葉を多くし、分析効率を高めるために設定したものである。PDS は、本システムでは 10 個用意したが、約 3,500 語 (ローマ字式の分ち書き単位による語数) のデータ 8 件について行なった筆者らの実験によると、11 個以上必要となるのは、1 件あたり 1~2 度であり、単位語辞書の言葉の補強で避けることができる程度であった。

### 3.2 漢數字化処理

ソクタイプ速記においては、数字データも、たとえば「イッポン」(一本の意味) のように、音の連なりとして表現されるが、本システムでは漢数字表現に変換することが必要である。漢数字変換を形態素分析法に組み込むこと (数詞の細分と単位語辞書への追加) も考えられる。しかし数字を表わす音連続は短かく、数詞以外の言葉との同形異義語が存在することが多いことから、形態素分析法による単位語の認定率、漢數字化率をともに低下させると考えられる。したがって、筆者らは数詞の認定は別に行なうことにし、その方式として、数字の表音表現が、接続する助動詞によって規則的に音韻変化することを利用することにした。これを筆者らは、数詞音韻分析法と呼んでいる。例をあげて説明すると次のようになる。すなわち、

- (1) 数字を <イチ>, <ニ>, <サン> とする。
- (2) 助数詞を <バン> (番), <フン> (分), <ホン> (本) とする。
- (3) <イチ> は <バン> のときは変化しないが、<フン> <ホン> のときは、<イツ> と音韻変化し、<フン> は <ブン> に、<ホン> は <ボン> に変化する。
- (4) <ニ> <サン> と <フン> <ホン> との接続では <ニ> についてはどちらも変化しないが、<サン> については <フン> は <ブン> に、<ホン> は <ボン> に音韻変化する。

この数詞音韻分析法によれば、数字・助数詞は 32 のグループに分類できる。本システムでは、一兆未満の整数、小数、分子 2 桁以下の分数、概数および筆者らの行なった語彙調査に出現した助数詞、数詞に接続する接頭語・副詞・名詞・副助詞を処理の範囲とし

た。数字、助数詞を表わす速記記号列に対して、漢字訳とグループ名を対にした数詞辞書を作成した。数詞辞書の見出し数は、420語である。また各グループ間の接続の可否を遷移表(33×20)の形をした数詞構成規則表にまとめた(3.1.3 文法表と同様の形式をしている)。形態素分析の途中、数詞の先頭を表わしうる速記記号が現われ、かつそれ以後を数詞と仮定して接続可能なとき、本処理に分岐することにした。

漢数字化処理では、数詞辞書探索により属するグループを知り、数詞構成規則表により接続の可否を判定し、可の場合は数詞辞書の見出しにない速記記号が現われるまで分析を続ける。接続がすべて可の場合、数字データと判定し漢数字訳を与え、形態素分析法に復帰する。接続不可の場合は、形態素分析に復帰する。一兆以上の数、あるいは新しい助数詞の出現に対しては、グループ名と対にして数詞辞書に収録することにより処理可能となる。

3.3 文章平滑化処理

3.3.1 語尾変化処理

話し言葉でよく使われる動詞(68種、例: 思う、いる etc.)、形容詞(16種、例: よろしい、少ない etc.)を表わす速記記号が略語として定められている。これらに助動詞が接続する場合、2.1の例に述べたように原形略語を羅列して表わす。これを出力の際通常の日本語文としてなめらかにする必要があり、これを語尾変化処理と呼ぶ。動詞・形容詞は、語尾変化の仕方により16種に分類した。また後続の助動詞は、語尾部分の変化のさせ方で、34種に分類した。これらの分類を、形態素分析による単位語の認定過程で判別し、通常の日本語文に変換することにした(Fig.3参照)。

3.3.2 括弧はずし処理

表わす訳の一部に括弧書き部分を有するものの処理が本処理である。括弧書き部分は、先行語との関係で選択の仕方が定まる。たとえば「(ノ)」については先行語が、名詞類(名詞、数詞 etc.)であれば、「ノ」を採用し、動詞であれば採用しない。この点に着目し

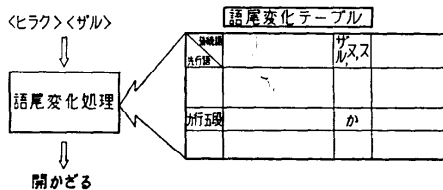


Fig. 3 Conjugation processing Algorithm

Table 2 Pattern of expressed Sound containing Parenthses

項番	パターン	例
1	(ノ)	(の)間、(の)とき、(の)ように
2	(ニ)	(に)従って、(に)なり、(に)入る
3	((ノ)デ)	((の)で)ございます、((の)で)ある
4	((ノ)デハ)	((の)では)ございません、((の)では)ありません
5	(デ)	(で)ございましょう、(で)あろう
6	(コトガ)	(こと)ができる、(こと)ができない
7	((シ)テ)	((し)て)おる、((し)て)いた、((し)て)やる
8	(シ)	(し)ます、(し)たる、(し)なく
9	(ス)	(す)べき、(す)べく、(す)れば
10	(セ)	(せ)ねば、(せ)しめ、(せ)られ
11	(ト)	(と)思う

た処理方式とした。括弧書き部分のパターンをTable 2に示す。

3.4 入力エラーの修正

速記データ作成時に、ソクタイプの打鍵ミスが、熟練者で全データ中2~3%存在するため修正する必要がある。エラーの修正は、計算機処理を行なう前に、オフラインで処理することとし、そのため、カセットテープ-to-カセットテープ機能と表示機能を有するエラー修正装置を用いた。

3.5 各タイピスト固有の略し打ちの処理

実際にシステムを運用する場合には、各タイピスト固有の略し打ちの処理が、重要となる。これは、辞書への収録により解決できる問題である。3.4のエラー修正時、修正済テープの冒頭に、タイピスト固有の略し打ちを、規定の打ち方と対比させて登録し、単位語辞書に帰着する方式、外部記憶装置に余裕がある場合は、略し打ちを個人辞書として収録し、共通の単位語辞書に先だって探索する方式が考えられる。

4. 試作システム構成

4.1 ハードウェア構成

本システムの実験は、HITAC-8400 計算機を用いて行ない、使用言語はアセンブラである。Fig.4(次頁参照)にハードウェア構成を示す。

4.2 ソフトウェア構成

本システムのプログラム構成を、Fig.5(次頁参照)に示す。

4.2.1 反訳サブシステム

(1) 入力処理プログラム

カセットテープ上に作成された速記データを読み込み、以後処理しやすいように速記記号 NO. を付与し、磁気ディスクあるいは磁気テープに出力する。選択機能として、入力速記データを、

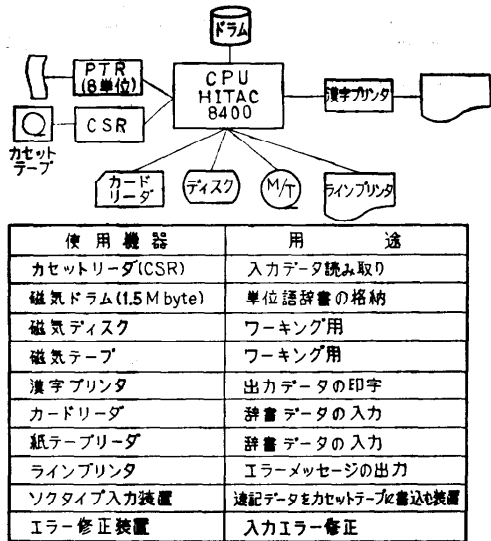


Fig. 4 System Organization

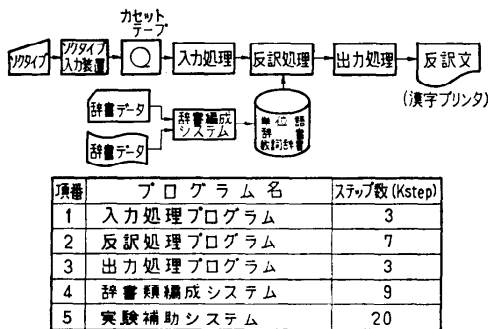


Fig. 5 Software System Organization

ソックタイプの打鍵形式にラインプリンタに出力する機能をもつ。この時、各速記記号に、本来割り当てられている訳をカナ訳で出力する。ここで速記記号 NO. とは、速記記号ごとに番号付けされたコードである。

(2) 反訳処理プログラム

磁気ディスク上のコード付けされた速記記号列を読み込み、3. で述べた方式により漢字かな混り日本語文に変換する処理を行なう。プログラムは、各機能単位で、次の5つのモジュールからなっている。

(a) 制御モジュール

反訳処理内の各モジュールの制御を行なう。このモジュールにより、単位語への分割、再試行に

よる backtrack 時のモジュール移行の制御を行ない、入力データを連続的に処理することができ

- (b) 速記データ入力モジュール
  - (c) 単位語認定モジュール
  - (d) 漢数字化モジュール
  - (e) 文章平滑化モジュール
- (3) 出力処理プログラム

反訳された漢字かな混り文を、漢字プリンタの出力形式にあわせて編集し、出力する。

4.2.2 辞書類編成システム

本システムにおいては、反訳文を漢字プリンタに出力するための漢字コードが必要である。漢字コードは漢テタイプにより紙テープに穿孔して入力する。品詞、付加情報等のアイテムについては、カードより入力することにした。辞書類の編成は次のような部分からなっている。

(1) 入力・エラーチェック処理

EBCDIK 部分、漢字コード部分の辞書データのフォーマットチェック、辞書リストおよび各マスターテープの作成を行なう。

(2) 更新処理プログラム

辞書に対する更新処理を行なう。

(3) マッチング処理

EBCDIK 部分と漢字コード部分とのつき合せを行なう。

(4) 構成処理

単位語辞書については磁気ドラム装置、数詞辞書については磁気ディスク上に構成を行なう。なお、反訳処理実行時、単位語辞書のインデックステーブルと数詞辞書は、コア常駐である。

4.2.3 実験補助システム

語彙調査用 KWOC/KWIC システム、反訳結果分析用プログラム等の実験をサポートするプログラム群である。

5. 実験結果とその検討

本システムの反訳精度を評価するために、実際の会話データについて実験を行なった。

- (1) データ量: 30分データ×4件(約3,500語×4件、この語数は、ローマ字表記の際の分ち書き単位数である)。
- (2) 単位語辞書の収録語数: 約 12,700 語

Table 3 Translation program efficiency

精度項目	形態素分析法	単純最長一致法	
正確性	96%	94.7%	
表記変換率	漢字化	67%	58%
	漢数字化	78%	57%
	固有名詞	60%	20%
	カタカナ化	61%	38%
	助詞<オ・エ・ワ>の<を・へ・は>化	80%	63%

$$\text{正確性} = \frac{\text{正処理語数}}{\text{総語数}} \times 100$$

$$\text{表記変換率} = \frac{\text{正処理数}}{\text{処理対象総数}} \times 100$$

実験結果を Table 3 (次頁参照) に示す。正確性 96% とは、残り 4% が文字の抜け、単位語認定誤り、括弧書き部分の未処理、助詞<オ・エ・ワ>の未変換等であることを示している。この誤りの原因について分析すると、単位語辞書の語数不足 60%、プロセディアの adhoc な修正で救えるもの 15%、文法表がデータに十分なじんでいないことによるもの 5%、同形異義語の存在 20% である。

この結果から、プロセディア、文法表の適正化、単位語辞書の補強で、正確性は 98% に達すると推定できる。漢字化率は 67% である。本システムでは形容活動詞、普通名詞の音読み語を漢字化の対象にしており、訓読み語については、カウントしていない。漢字化できていない部分を分析すると、約 1/5 が、単位語辞書の収録不足であり、残りは同形異義語の存在のためであった。この結果、本方式での限界性能は、75% 程度と推定される。

漢数字化については、数詞類を表わす速記記号列が短かく、他の品詞と同形異義語が存在するもの(例。イチ「一」と「市」、ゴ「五」と「後」etc.)である。

片仮名化および固有名詞については、言葉としての出現頻度が低いため単位語辞書に収録されておらず、正しく変換できなかったものである。これらの点からさらに精度の高い日本語文を得るには、現時点では何らかの形で人手の介入が必要であることは言うまでもない。

Table 3 から、形態素分析法は、単純最長一致法と比べて正確性というより、表記変換率(漢字化、片仮名化 etc.)の向上に寄与しているといえる。

処理速度は、7,000 語(1時間分の速記データ)のデータについて、カセットテープを読み始めてから、漢字プリンタに出力終了するのに、HITAC-8400 シ

入力速記記号	音訳	反訳結果
KSA I T I O S H T 7 B Y I O A T 7 7 7 O I S H 7 7 7 AIO I S H 7 7 7	イソノ イソノ イソノ イソノ	それは、何時ごろからですか。 四時二十...四時三十分前だったと思います。
YTH OTK A KH 7 7 7 7 7 Y S I Tn ASK TY ((1))7 7 7 7 7 Y S I Tn N K 7 7 7 SAI INO ASHMT 7 7 7 7 7	イソノ イソノ イソノ イソノ	
S OTK I NOJA K 7 7 7 H SA N I S T 7 7 7 T K. OTK OIA K T 7 7 7 THKSA OTK A K 7 7 7	イソノ イソノ イソノ イソノ	
Y H S I O I S T 7 7 7 N I S T 7 7 7 N I S T 7 7 7 N I S T 7 7 7	イソノ イソノ イソノ イソノ	
Y U I S T 7 7 7 THK SAI TK OI S TV 7 7 7 Y THK N 7 7 7 Y THKSA OTK 7 7 7 THKSA Y N A T 7 7 7 AI Tn IM ASK ((1))7 7 7 N H 7 7 7	イソノ イソノ イソノ イソノ イソノ イソノ イソノ	

Fig. 6 An Example of Translation

ステムで約 18 分、すなわち 155 msec/語であった。このうち反訳処理には、7 分 10 秒程度を要していた。本システムの反訳結果例を、Fig. 6 に示す。

### 6. むすび

本システムは、話し言葉である速記を対象とし、かつ言葉としての区切りがない表音記号列から漢字かな混り日本語文へ変換する点が、各方面で開発されている類似のカナ文字漢字変換システムと比べて、新しい試みであったと考えている。現在、単位語辞書は、約 12,700 語であるが、漢字化・片仮名化等の表記変換率を向上させる(これには限界がある)には、収録語数を増大させる必要がある。単位語辞書中の収録語は、語彙調査を行なうことにより決めたが、話し言葉を処理するシステムとして、十分なものであるとはいえない。これらの点、および本方式をカナ文字列に適用した場合の問題点の解明が、未検討部分として残っている。また、各タイピスト固有の略し打ちについては、エラー修正時に登録する方式、さらには個人辞書と共通辞書を併用する方式を考えているが、全文のうちで、タイピスト固有の略し打ちの出現する頻度と種類の調査の結果を踏まえて運用システム設計の際には検討していくべきものと考えられる。

おわりに臨み、本システムの開発に際し、御指導いただいた(株)日立製作所の藤本謹也氏、伊藤祐太郎氏、霜田忠孝氏に感謝いたします。

## 参考文献

- 1) 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
- 2) 松下温他2: 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)
- 3) 相沢輝昭: 文法情報を利用したカナ漢字変換, *NHK* 技研月報, Vol. 13, pp. 516~522 (1970)
- 4) 相沢輝昭他1: 計算機によるカナ漢字変換, *NHK* 技研月報, Vol. 25, pp. 261~298 (1973)
- 5) 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 (1973)
- 6) 湯沢幸吉他1: 要解国語・国文法辞典, 東京堂出版 (1970)
- 7) 水谷静夫: 言語と数学, p. 162, 森北出版, 東京 (1970)
- 8) 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 (1962)
- 9) 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 (1963)
- 10) 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 (1974)
- 11) 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 (1970)
- 12) 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 (1971)

(昭和49年9月3日受付)

(昭和49年12月24日再受付)