

## 自動点訳システム開発の動機と経緯

一般教育等 黒川哲宇

**要 旨：**自動点訳システムは視覚障害教育を実践していく上で重要な位置を占めてきている。本稿では、本学の創設準備時代に検討・実験されていたファイル作成形式と点訳プログラムについて述べた報告書を掲載し、今後の参考に供する事にした。

本学が創設される準備段階で、当然のことだが、教科書や参考書などの教材をどのようにして用意すべきかという議論が起きた。特に、視覚障害を持つ学生に対しては点字教材、弱視学生に対しては拡大教材が必要となるにちがいない。当時、すでにコンピュータもだいぶ進歩してきており、パソコンもかなり性能のよいものが登場してきた時期であった。そこで、コンピュータ利用というものを視点においた教材開発・作成こそ、本学の特性を発揮できる有望な分野ではないかという認識に到ったのである。

著者が準備室に赴任したときには、教材は点字データベースとして用意しておき、端末を通して学生に提供するという方式が想定されていた。データベースの中にある教材は点字コードで記述されるわけだから、そのテキストファイルは、墨字の教材を直接点字で入力するか、仮名分かち書きで入力して、それを点字コードに変換するかの方式が採用されようとしていた。

ところが、その方式だと後で不便なことが起きてくる。第1に、仮名コードや点字コードで記述された文書だと、元はどんな漢字が使われていたのかという情報がカットされてしまう。文書のある箇所の漢字と変換された仮名や点字との間に何らかのリンクを引いておき、対応する漢字情報にアクセスできれば便利ではないか。すなわち、点字情報から漢字辞典や国語辞典にアクセスできないかということである。

第2は、弱視学生用の教材提示である。教材のファイルが漢字仮名コードで記述されていれば、それらの文字を画面上に拡大して表示できれば弱視学生にも便利ではないか。ただ、この時は、今のような優れたプリンタが低価格で出回っていなかったため、今のような拡大印刷方式は考えつかなかった。

第3は、世の中に流布している情報をそのままの形で視覚障害学生が取り入れることができないかということであった。当時、すでにワープロは一般的でありワーブ

ロで入力された、つまり電子化された漢字仮名文書をそのまま視覚障害者が利用できないだろうか。弱視者の場合は前述の方式で読めばよいが、全盲者の場合は何らかの変換手続きが必要になってくる。自動点訳や合成音声出力がこれである。その当時は合成音声出力装置がまだ開発されたばかりで、現在7万円で購入できるものが150万円もした。開学時に、これを20台も揃えるのは不可能である。そこで自動点訳プログラムがあれば、漢字仮名交じりのテキストを点字に翻訳して学生に提供することができる。一方、パソコン通信がかなり盛んになってきた時でもあり、またCD-ROMの話題も登場し始めていたから、自動点訳システムを利用して、これらの情報も点字使用者が利用できるように是非したいものだと考えたわけである。

ところが、問題は教材作成の予算要求の方式であった。文部省には仮名分かち書きで入力したファイルを、文字変換テーブルで点字コードに変換するというように説明していたので、漢字仮名ファイル作成はまかりならぬというのである。文部省はともかく、まず準備室内を説得する必要があった。そこで、石田助教授（当時は助手）に頼んで、自動点訳システムのプロトタイプを作ってもらったのである。このプログラムが動いて、うまく点訳ができなければ、漢字仮名ファイルからの入力は認知されなかった。結果的には、不十分ながら自動点訳らしきシステムが動き、現在のような漢字仮名ファイル教材をオリジナルな形式にしようということになったのである。その時、そのための説明資料を作成して内部的に検討してもらったのだが、公表するには到らなかった。

そこで、本稿では当時作成したレポートをそのまま掲載させていただき、当時検討されていた問題と、解決の手だての素案を記録しておきたかったのである。ここでは、分量の関係から図と表とその説明は削除したが、その他の記述は最初のままにしてある。本学では、学外の人の手になるEXTRAが一般的に使われているが、OCR

で読み取ったテキストの校正用や、点字ディスプレイに接続して、漢字仮名情報を同時に点字で表示するシステムのためには、分かち書きと点訳用のシステムの開発がいまも望まれている。

## 漢字仮名まじり文書の自動点字変換システムの試作 (1987.10)

### 1. はじめに

視覚障害者、特に全盲のディスアビリティは主として2つあるといわれている。それは文字コミュニケーションと歩行に関する障害である。歩行については従来いろいろな方策が試みられてきたが、十分な対応ができるまでに到っていない。一方、文字コミュニケーションの障害は、Louis, Braille (1809-1851)が19世紀初めに点字を考案して以来めざましい改善を見た。点字の使用によって、視覚障害者は書きやすかつ読みやすい文字を手に入れることができた。

また、点字を書く(打つ)道具も進歩した。点字は、当初は点字版と鉄筆で打たれていたが、1892年にFrank, Hall (1841-1911)が点字タイプライターを発明し、100wpm以上の速度で点字を打つことを可能にした(Hendrickson, 1955)。Hallはまた、亜鉛版を用いた点字印刷機も発明している。これら2つの革命的な機器は以後100年近く基本的な構造を変えずに使われ続けてきた。

一方、一般の文字(墨字)を点字になおす作業は目の見える人(正眼者)が担当してきた。点字翻訳には膨大なマンパワーが必要である。そのために点字のユーザーは自分が読みたい本を、読みたい時に、読みたいだけ読むという状況を獲得するには到っていない。

最近、電子機器、特にコンピュータのめざましい進歩と普及によりこれらの技術を点字翻訳に利用していこうとする試みが多数みられる。パソコンを利用した点字ワープロ、合成音声による文書やメッセージの出力、点字プリンタやペンディスプレイの開発などはコンピュータを最大限に利用したシステムであるといえる。今回は、点字や点字翻訳のもつ問題点を整理して、コンピュータを利用した自動点字翻訳システムの開発についての中間的な報告を行うこととした。

#### 1. 点字の問題点

確かに点字は視覚障害者にとって便利な文字であるが、日本語を点字に直すということで生じる問題点もある。日本語の場合、使用される文字は漢字、平仮名、片仮名、

アルファベット、数字などであるのに比べて、点字の場合では仮名と数字によって表記する。文書を仮名で書き表すので、英語のように語と語の間をマス空けしなければならない。ところが、仮名だけで表現することで時として表現や意味が不明確になることがある。

「ここは しどうなので しでは しゅうり できません。しで しゅうり できるのは しどうだけです。」とか、「きしゃの きしゃが きしゃで きしゃした。」というような文章は仮名で書いた場合には大変分かりにくい。日本語では漢語を使うことと、歴史的に音節を単純化してきた経緯があるので、仮名にしたときの同音異字あるいは同音異語が多いのである。上の文は漢字仮名まじり文で書くと、それぞれ、

「ここは私道なので市では修理できません。市で修理できるのは市道だけです。」 「貴社の記者が汽車で帰社した。」となる。

もう一つは、点字文書・図書はかさばるということである。300ページほどの本1冊を点訳すると10分冊程度の分量になってしまう。1冊のコンサイス英和辞典を点訳すると100分冊になり、それを入れるとリヤカーいっぱいになってしまうということになる。

### 2. 点字翻訳の問題点

次は文書・図書を点字になおす場合の問題である。

まず、点訳者の資質の問題をあげねばならない。点訳はたくさんのボランティアの努力に支えられてきた。これらの人たちが点字の本を作り、点字図書館では点字図書貸出サービスを行なってきた。点訳する場合には点字表記の種々の決まりがあり、それに基づいて墨字を点字になおすのである。また、点訳者は点字表記上の知識の他に、点訳する本の専門領域についての素養も持っていないなければならない。古文を点訳するには古文をよく知っている必要がある。物理学を知らなければ物理学の点訳はできない。また、左翼関係の本を右翼が点訳してはくれそうもない。仏典をキリスト教徒は点訳してくれないかも知れない。どの本を点訳するかは点訳者の知識と考え方に影響される。点字本のユーザー側のニーズは自分が読みたい本をなるべく早く手に入れたいということであろう。そのためには適切な素養を持った点訳者が適切な数だけそばにいななければいけない。一般大学に進学している全盲学生には点訳ボランティアのグループが一体となって勉学を支援しているのが現状である。

一方では点字図書の複製の問題がある。貴重な努力の結晶である点訳本は普通1冊である。点字紙にコツコツ打ち出すのであるから、1回の点訳は1冊に限られる。点字印刷すると原版作成などに莫大な費用がかかり、と

ても個々のユーザーの需要に応じているわけにいかない。

### 3. コンピュータを利用した点字翻訳

以上述べてきたように、点字や点訳の問題点は主として次の4つに整理される。すなわち、仮名表記によって生ずる意味の不明確さが除去できないこと、かさが張ってしまうこと、点訳者の資質が不可欠だということ、複製が手軽にできないということである。

ところが、都合のよいことにこれらの問題は、実はコンピュータが得意としている仕事なのである。

点字図書の情報をつロッピーディスクに格納しておき、必要な箇所を取り出して読めばリヤカー1台分の点字辞書は必要ない。大きい辞書でも数枚のつロッピーディスクに納まってしまうだろう。1MBのつロッピーディスクというのは1枚の中に約1,000,000字分の点字情報が貯蔵できることを意味している。今ではこのつロッピーディスクよりも容量の多い装置がいろいろ開発されている。例えば、20MBのハードディスクにはつロッピーの20倍の貯蔵能力があり、これがCD-ROMになると500MB、光ディスクだと1,500MBという一種天文学的とも思える能力になる。一方、普通の本の文字数は20万から30万であるからバイトになおすと0.5MB程度になる。光ディスクを利用した場合には、なんと3,000冊の本が1枚のディスクに入ってしまう。

文書の複製はコンピュータではいたって簡単である。文書ファイルの複製はたとえば、「COPY」という命令だけで遂行できる。複製された文書情報は、それがつロッピーディスクに入っていればカバンに入れて持ち運ぶことができる。

ところで、最近では人工知能を利用したエキスパートシステムの開発がさかんである。エキスパートシステムというのは、コンピュータに専門的な知識を記憶させておいて、その知識を利用して種々の判断を下そうとするシステムである。自動車整備のためのエキスパートシステムでは、故障箇所の点検や実験の指示をエキスパートシステムが出し、その指示に基づいて人間が作業を進めていけるようになっている。点訳の場合では、点字表記上の知識や専門領域の知識をエキスパートシステムとして構築しておき、それを利用して点訳を進めていくことになる。

仮名表記点字の不明確さを除去するためには少し工夫を要する。点字文書を仮名の形で保存しておくという方法では、この不明確さの問題の解決はむずかしい。「しどろ」や「きしゃ」を仮名のままデータとして貯えても、それが「私道」なのか「市道」なのか「指導」なのか、あるいは「記者」なのか「汽車」なのか、仮名の単語だ

けからは判断できない。そこで、漢字仮名まじり文をそのままデータとしてファイル化しておき、エキスパートシステムを介してその文書を自動点訳する方法がよいと思われる。このシステムによって自動点訳された仮名点字文書に、たとえば「しどろ」という語があり、意味が不明確であれば、ユーザーは処理をさかのぼって、「私道」という漢字熟語に到り、その漢字そのものやその意味などを、別に電子ファイルとして用意されている漢和辞典や国語辞典などにより検索できるのである。

原文書を通常の漢字仮名まじり文で入れておくことの利点は他にもある。ひとつは、漢字仮名まじり文書から容易に漢点字に変換できることである。仮名でなく、漢字を点字で表記する方式は最近はかなり普及しつつある(黒川, 1984)。ユーザーは、その単語を漢点字に変換して読むことにより、どんな漢字が用いられているかを知ることができる。

現在、一般には弱視の人には本をテレビカメラで拡大してモニターテレビに投影する、いわゆる拡大読書器が利用されていたり、本を複写機で拡大するという利用がされている。しかし、このようにしてパソコンに入力された漢字仮名まじり文書をディスプレイ上で拡大してやれば弱視の学生も読みやすくなる。

文字を電子的な情報に変換して保存しておき、後でそれを検索するという方式は、今後さらに一般的になっていくと考えられる。電子ファイルとしての本の出版は、辞典、図書・文献情報データベースなどとして盛んになりつつある。また、新聞の記事もパソコン通信を利用して読めるようになっている。これを点字に自動変換できれば視覚障害者はいながらにして新聞を読むことができる。また、今ではほとんどの本は電子ファイルとして入力されてから印刷工程にかけられている。もし、印刷工程に入る前のつロッピーを手に入れることができれば、それを自動点訳して読むことができる。先生が自分でワープロを叩いて入力した参考資料や課題も自動点訳システムを介することで容易に点字で読むことができる。電子ファイルという形で文字を貯えることにより、視覚障害者と正眼者との間の文字コミュニケーションがきわめて容易になるのである。

自動点訳システムも含めて、コンピュータを用いて漢字仮名まじり文を仮名書きに変換したり、単語で区切るシステムは従来大型のコンピュータが用いられてきた(田中, 1969; 長尾他, 1978; 荒木、板山, 1979; 藤崎, 1981; 脇田、嘉手川, 1984; 藤崎、亀田, 1985)。これらのシステムでは数万字から数十万字の文書を1分以内で処理している。しかしながら、視覚障害者が個人のレベルで自動点訳装置を持ち、それを利用して本を読んだり、

辞書を引いたり、データベースにアクセスしたりするためには、いわゆるパソコン規模での自動翻訳システムを構成することが必要である。現在のパソコンは大変進歩しているのだから、処理時間の点以外では十分満足のいくシステムを実現できると考えられる。

## II. パソコンを利用した自動点訳システムの概要

### 1. 点訳の書式

墨字の文書を点字に翻訳する場合には、漢字仮名まじり文を単に仮名書きするだけでなく、さらに点字のフォーマットに変換しなければならない。点字では助詞の「は」は「わ」と、「へ」は「え」と表記する。また、オ列とウ列の長音は片仮名で用いているような長音記号で表す。したがって、「今日は学校へ」という節は「キョーワ ガッコエ」というように書き表す。また、仮名書きであるので、英語の様に分かち書きしなければならない。この分かち書きの単位は文節である。文節の単位は自立語であるから、助詞、助動詞などの付属語は自立語に続けて書く。

「我輩は猫である。名前はまだない。」という文では、「ワガハイワ ネコデ アル。 ナマエワ マダ ナイ。」というように書く。点字に翻訳するのではなく、墨字用の漢字-仮名変換システムの開発（長尾他, 1978；荒木・板山, 1979；藤崎, 1981；藤崎・亀田, 1985）では助詞も一つの単語として処理しており、点字の分かち書きとは書式が異なる。点字の場合は正書法に近い分かち書きの表記基準ができており、その規則に従って書き表されている。今回開発したシステムもこの点字表記の基準にいかにか正確に従えるかというのが重要な観点であった。

### 2. 翻訳用辞書の構成

本システムは人工知能を用いたエキスパートシステムではなく、漢字仮名まじり文書を辞書と対応しながら点訳処理をしていく方式を採用した。その意味では、長尾ら(1978)、荒木・板山(1979)、藤崎・亀田(1985)らのシステムと基本的には変わらない。点訳の主な任務を辞書が担っているということは、処理判断の基準や点字の規則を辞書に持たせなければならないということである。言い換えれば、エキスパート的な専門知識を辞書そのものが持っているということである。つまり、辞書は構造化されていなければならない。

我々は、まず、本システムの辞書に基本的な4万語程度の語彙を入力した。たとえば、「利用」というような語彙を4万語選択して辞書に登録しておいた。これらが

自立語である。これは国語辞典に掲載されている語彙の形と同じものである。

ところで、文節を決定する場合には、自立語あるいは自立語プラス付属語というのが一般的な単位になる。どういう文字の並びが自立語であり、どういう文字並びが付属語であるかを辞書に登録しておかねばならない。このシステムでは辞書そのものが判断を下すので、基本的には自立語だけでなく付属語をも含んだ文節を、どのようにして辞書に登録しておくかがカギであった。

「利用」という言葉は名詞にもサ変動詞にも使うので、「利用は」や「利用を」など格助詞を含めたものや、「利用する」、「利用しない」、「利用すれば」などの活用語尾もすべて登録する必要があるかなどを検討した。長尾ら(1978)の辞書では46,279語のうち、「利用」などのサ変動詞が6,760語(11.07%)あり、もしこれらのサ変動詞に格助詞や活用語尾をすべてつけると辞書のボリュームが膨大になり得策でない。そこで、「利用 リヨウ」と登録することに加えて、2字以上の仮名並びが辞書にない場合は、これを助詞などの付属語としてみなすことにした。つまり、「を」は1字として辞書に登録しておく。また、ある長さの文字列を処理している時、「は」と「へ」が語尾にくる語彙が辞書にない時には付属語の助詞であるとして、それぞれ、「わ」、「え」に変換するように辞書に登録した。つまり、「をヲ」、「はワ」、「へエ」などと登録した。動詞は、「利用す」、「利用せ」、「利用しな」、「利用さ」などとして入力した。この際、活用語尾の1部だけを登録したのは、「利用す」を登録しておけば「利用する」の終止形や「利用すれば」の仮定形が出てきた時には、まず「利用す」を辞書との対応で自立語とみなし、「る」、「れば」などの語尾を付属語として判断する。こうすれば辞書の容量がコンパクトにできる。

文節分かち書きでは、「利用する」は続けて書くが、「利用をする」では、「利用を する」と「利用」と「する」の間を区切るのである。辞書には区切るかどうかという情報の他に、その語彙の読み方も登録しておくので、たとえば、「利用す」は「利用す リヨース」というように語のまえを1マス空けて登録されている。また、「る」は「るル」と、「せる」は「せるセル」というように前を空けずに登録しておくことで付属語としてみなされる。したがって、「利用する」というデータに対しては「リヨースル」と変換する。これは、辞書の内容が「利用す リヨース」、「るル」となっているためである。また、「利用をする」というデータに対しては「リヨウヲ スル」と間を区切る。これは「利用 リヨウ」「をヲ」、「する スル」と登録してあるためである。

### 3. 点字変換用プログラムの構成

ファイルの入力してある漢字仮名まじり文書を自動的に点字に変換するプログラムの概要を述べる。

まず、ファイルから1センテンス分の文字列を読み込んでくる。今、例として「我輩は猫である。」という文をファイルから読んできたとする。次にその文字列の最初の文字を切り取る。すなわち、「我」の1字を切り取る。その文字（我）が辞書に登録されていれば、文字列の長さを1文字分プラスして、先頭から2文字の並びを切り取ってくる。つまり、「我輩」という文字列を切り取ることになる。この「我輩」という文字列も辞書に登録されていたときには、文字の長さをさらに1字加えて、「我輩は」という3文字を切り取ってくる。この3つの文字列は辞書にないので、「我輩」をひとつの文節として決定し、点字の規則である助詞の「は」を「わ」にかえて、「ワガハイワ」というデータを新しいファイルの書き込む。その次に、切り取るべき文字列の先頭を移動させて、「猫である」の先頭に移動させる。この処理の次に、読み込んできたセンテンスが終わりであるかどうか、さらに、ファイルにある文書データが終わりであるかどうかをチェックするステップが加わっている。文節を決定する場合の細かい処理は別に用意しなければならないが、プログラムの処理の流れはだいたいこのようになっている。

文節切りの方は、荒木(1979)が最長照合といい、藤崎(1985)が最長一致法としているものと基本的には同じである。すなわち、「その場かぎり」という文節の場合、変換用の辞書には「その」も「その場」も「その場かぎり」も用意されているが、文字列の1番長い「その場かぎり」が優先されるようにした。

### III. 自動点訳の実際

#### 1. 試作システムの点訳結果

一般教養書を点訳用として選択し、その文書の内4,549文節について点訳を試みた。その結果3,368文節が正しく点字に翻訳できたので、正当率は74.4%ということになる。自動点訳システムとして実際に使おうとする場合には、点訳正当率が98%以上は必要であると考えられる。したがって、74%程度では翻訳ミスを訂正する作業量が多すぎて効率的なシステムとはいえない。そこで、どのような種類の誤りがあり、その原因は何かということを分析してみることにした。

まず、誤りの種類別の相対的な比率を調べてみた。最初は「開けすぎ」と「詰めすぎ」のミスである。「開けすぎ」というのは、自立語の中をさらに分けてしまった

誤りである。つまり、「ワリキッタ」という文節を「ワリキッタ」と2つに分けてしまったようなケースである。「詰めすぎ」というのは、複数の文節をひと続きの文節として詰めてしまったものである。「アルコールヲツクッテ」というところを「アルコールヲツクッテ」とつなげてしまった誤りである。3番目は仮名ふりの間違いである。すなわち、「アナタガタ」とするところを「アナタホー」としたようなケースである。ここでは、「詰めすぎ」の誤りが最も多かった(52.33%)。2番目は開けすぎの誤りで27.77%あった。最後は仮名ふりのミスで19.9%であった。

次に、誤りの原因を4つに分けてみた。すなわち、第1は辞書に「未登録」であるためにミスをしたケースである。固有名詞を除いた語彙について、その語が辞書に登録されていなかったために誤って翻訳されたケースがこれに該当する。第2は「登録ミス」のために誤って翻訳されたケースである。第3は「補助プログラム」が必要な場合である。これはある特定のケースに遭遇した場合、特定のルーチンに飛んでいく補助的なプログラムが必要なケースである。最後は今回のプログラム構造(アーキテクチャー)では「対処が不可」であると思われるケースである。固有名詞や接尾語、接頭語などをこの中にいれた。頻繁に出てくる固有名詞は辞書に登録してあるが、多くの固有名詞は翻訳前にその読みを予測できない場合が多い。また、「～方」という字は「かた」「がた」「ほう」「ほう」などと読むが前にどの字がくるかで読みが変わってくる。この場合でも登録の仕方を工夫すれば解決が可能な部分を多いが、今回は対処が不可としてみた。4つの翻訳ミスの原因を相対的に比較してみると、「未登録」が断然多かった(63.01%)。他の3つはいずれも20%以下であった。辞書への登録を追加・修正すればかなりの部分が解決できることを示している。

文節として分かち書きすべきところを詰めてしまう間違いは、当該の自立語が辞書に登録されていないために起きたのであろうと思われた。逆に、切るべきところを開けてしまった誤りは、辞書の登録が適切でなかったからであると考えられた。そこで、「開けすぎ」、「詰めすぎ」、「誤仮名ふり」の3種のミスについて、その原因がどのような比率で発生したかを調べてみた。翻訳ミスの種類をその原因別にみると、語彙が未登録であることによって詰めすぎている誤りが顕著に多かった。全体の誤りのうち約44%が語彙の未登録による詰めすぎであった。他の2つの誤りでも未登録が原因しているものが相対的には多かった。登録ミスによる誤りは「開けすぎ」で数%みられた。これは付属語(助詞や助動詞)の登録の仕方が適切でなかったことに起因する。補助的なプログラ

ムが必要であるケースというのは、例えば、いかにしたら読点や中点など記号の後のマス開けを正しくできるかという問題である。

以上のことから、翻訳ミスのはほとんどは追加登録や修正登録で解決できることがわかった。今回の誤り全体の中で、「対処不可」が原因のものは13.78%であった。これは辞書やプログラムを修正すればかなり正当率を向上させることができることを意味する。すなわち、今回の翻訳正当率が74.04%であったので、誤りの比率（誤翻訳率）は約26%である。その内約14%が「対処不可」の誤りなのだから、「対処不可」の誤りは全体のわずか3.6%を占めるにすぎないということになる。つまり、辞書を修正したり補助的なプログラムを追加してやれば96%程度には正当率を上げることができるはずである。

## 2. 辞書を修正した後の点訳結果

辞書の修正は何種類かに分かれる。ひとつは仮名辞書の修正である。試作的な翻訳辞書には仮名の語彙は当初あまり含まれていなかった。形容詞、接続詞、副詞などは仮名で表記される場合が多い。このような語彙は自立語であるから、前を1マス開けて登録することになる。すなわち、「しかしながら」という接続詞は、「しかしながら シカシナガラ」というように登録される。最初の辞書では、このような仮名の語彙が未登録であったために翻訳ミスをしてきたケースがかなりみられた。また、助詞や助動詞などの付属語を登録する場合は、その単語の前を詰めて登録しなければならない。つまり、「よゆうな」という助動詞は、「よゆうなヨーナ」と登録しておかなければならない。このマスあけで修正する必要があるケースがいくつかみられた。

もう一つは漢字熟語の問題である。特に専門用語はその語彙が登録されていない限り正しく変換してくれない。ただ、今回翻訳の対象とした文書は一般的な教養書であったので特殊な専門用語は少なかった。医学や鍼灸あるいは情報処理関係の専門書を点訳する場合には独自の辞書を用意する必要がある。また、固有名詞のうち人名は登録が問題になる。英語のスペルチェッカーでは、処理中に辞書にない語彙に遭遇すると、その語をディスプレイに表示して登録すべきかどうかを聞いてくる。日本語の点訳プログラムもこのような処理を挿入すれば問題がある程度解決されると思われる。

漢字の辞書で厄介なのは接頭語・接尾語の問題である。これらの語は漢字1字であるので、まずいくつかの漢字列を分割し1字を取り出すという処理がある。「鉛」という字は、「黒鉛」では「エン」と読み、「酢酸鉛」では「ナマリ」という。したがって、「鉛」の読みをどのよ

うにして辞書に登録しておけばよいか課題になる。荒木・板山(1979)のように「2通り以上の読みがある場合、種類の少ない方を辞書化する」のも一つの方法であるが、これだけでは完全ではない。また、分割の精度は漢字列の数が多くなると落ちることが知られている（長尾ほか, 1978）。さらに、「2本」と「3本」では「本」の読みがちがってくるが、これは「2本 2ホン」、「3本 3ボン」というように「10本」まで登録すればよい。

今回は、仮名だけについて辞書に新たな語彙を追加してみて、翻訳の精度がどのくらい向上するかをみた。約200程度の形容詞、接続詞、副詞を追加登録して処理した結果、正翻訳率が85.51%まで向上した。最初のシステムより10%以上向上したのは、これら追加登録した語彙が出現する頻度が高かったためである。

## 文献

荒木啓介, 板山和彦(1979) : JICSTの実用的全自動漢字-カナ変換システム, K-KACSについて. 情報処理, 20, 917-923.

藤崎博也, 亀田弘之(1985) : 新聞記事データを対象とする自動単位切り処理とそれに基づく語彙調査. 情報処理学会研究報告, NL51-2.

藤崎哲之介(1981) : 動的計画法による漢字仮名混じり文の単位切りと仮名ふり. 情報処理学会研究報告, NL28-5.

Hendrickson, W.B. (1955): The beginnings of the braille writer. New Outlook for the Blind, 49, 299-304.

黒川哲宇(1984) : 点字問題をめぐって. 特殊教育学研究, 21(4), 44-47.

長尾真, 山上明, 辻井潤一, 建部周一(1978) : 国語辞書の記憶と日本語文の自動分割. 情報処理, 19, 514-521.

田中章夫(1969) : 漢字かなまじり文を全文カナ書き・ローマ字書きに変換するシステムについて. 国語研究所報告, 34, 17-22.

脇田修躬, 嘉手川繁三(1984) : 視覚, 聴覚障害者援助のためのコンピュータ利用. 情報処理, 25, 462-470.