

ショートノート

漢点字を用いた自動点訳システム†

下村有子^{††} 水野舜^{†††} 長谷川貞夫^{††††}

点字は仮名のみで触読文字であるが、視覚障害者は漢字の学習を希望している。点字に漢字を導入することは多くの利点があると考えられる。本論文では、漢点字を使用した自動点訳システムについて述べている。システムは自動点訳と自動編集、そして付加機能に分けられる。自動点訳では、変換表を用いて点字コードに変換する。変換表は、基本変換表とそれぞれの点訳物専用変換表を要する。自動編集では、点字表記法に従って編集を行う。これらを行うことによって実際に使用できる点字文を生成する。また付加機能では、漢字出力、点字パターン出力、および JIS コード、点字コードの CRT 表示、修正を可能としている。

1. ま え が き

近年印刷物は増大し、それに比例して計算機で利用できる文字データは増えつつある。これとは逆に、点訳は点訳奉仕者の手作業で行われるため、点訳物と印刷物の差は広がる一方である。われわれは計算機で利用できる文字データを自動点訳し、視覚障害者に点訳物を提供するシステムの作成を試みた。

仮名点字の自動点訳システムは、計算機処理による分ち書きの精度が低いため、現在実用化の可能性は低い。本試作システムでは分ち書きの必要のない漢点字を採用している。漢点字を採用することにより、多くの視覚障害者の願望であったコンピュータによる点訳が可能となった。このことにより、仮名点字では不可能であった同音異義語の理解や文章の意味を忠実に理解させることが可能になる。

2. 点字と漢点字

点字は指先で触読される凸点の符号化された文字である。これは6点で構成(これを1マスという)されており、日本では仮名50音を6点に調和させ使用している。近年、点字界においては漢点字の必要性が唱えられ、試作もなされてきている¹⁾。

仮名点字が表音文字のため、本論文において作成した漢点字も表音表を基本として漢点字体系を構成²⁾することとした。漢字は部首で構成されているので、それを主体とした漢点字の構成も考えられるが、日常会話で使用する点を考慮し、漢字の音と訓を取り入れた3マスの漢点字を考案した。例外として4マスも用いている。漢点字は漢字符号(7種類)と、音の読みの一音と訓の読みの一音の3マスで表す。このように構成することにより、学習していない漢字が現れても順次読み続けることができ、その結果、文章をより深く理解することが可能となる。また仮名点字で表せなかった点字も漢字で表すことができ、それらの1対1の対応づけが可能となった。

3. システムの構成

本システムのフローチャートを図1に示す。本システムはたんに点訳のみではなく、次の機能も含んでいる。

- (1) JIS コード、点字コードともに文字として CRT 上に表示でき、修正が可能である。
- (2) 漢字出力や点字パターン(墨点字)出力が可能である。その出力例を図2に示す。
- (3) 点字表記法³⁾に従い自動編集を行っているが、点訳作業とは独立である。

点訳対象となる入力データ⁴⁾としては、三省堂の国語辞典を用いてみた。原データは1文字が18ビットコードである。それを JIS で設定されている情報交換用漢字符号系コードに変換する。入力データは1行が128バイト(64文字)であり、1レコードは27行で構成されている。本論文では図3に示すように、1

† The System of Automatical Translation into Braille by Using Kanji Braille by YUUKO SHIMOMURA (Kanazawa Women's College), SHUN MIZUNO (Kanazawa Institute of Technology) and SADA O HASEGAWA (School for the Blind Attached to Tsukuba University).

†† 金沢女子短期大学情報処理工学

††† 金沢工業大学工学部情報処理工学

†††† 筑波大学附属盲学校

* 現在 金城短期大学

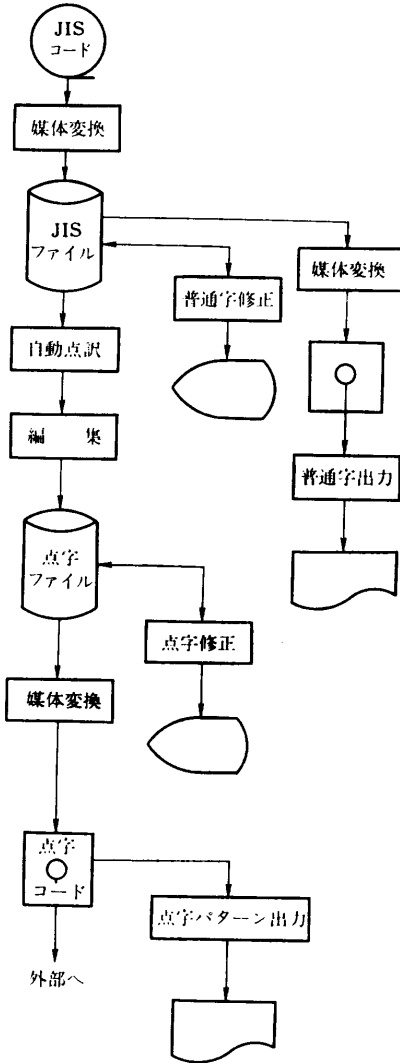


図1 システムフローチャート
Fig. 1 System flowchart.

バイトに1マスを割り当て、制御用ビットは、マスあけ、改ページ、改行等に使用した。

4. 自動点訳⁵⁾と自動編集

点字という特殊性のため、以下に示す三つの基本変換表を用いる。(1) JIS 点字対応表は JIS コード1文字に対する変換であり、変換文字は記号、数字、英字、ひらがな、カタカナ、漢字である。(2) 拗音表は拗音の変換、(3) 特殊音表は特殊音の変換である。(2)、(3)ともに JIS コード文字で構成されており、2文字目が小文字という特徴を生かし、小文字列と大文字行の組合せのインデックスを作成して変換する。しかしながら、点字にない文字が出現した場合は、よく似た意味をもつ点字を割り当てることとする。

あ (副) 「あのような」の口
形。「— いう [= あのような] — した
[= あんな] — いつも— だ から困 ってしまう」
あい [愛]
① 愛 情。「— の手 をさしの へる。 —
の告 白。 — を誓 う。 友 — 博
— 恋 — 母 性 — 人 類 —
— 欲 —」
② そのものに尽 くすことこそ生 きが いと考
え、自 分 をその中 に没 入 させる心
「学 問 への—」

図2 パターン出力例
Fig. 2 Example of translated braille.

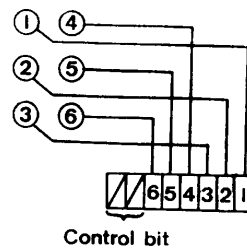


図3 点字とバイトの対応
Fig. 3 Association for braille and byte.

自動編集については点字表記法に従い、次のような編集を行っている。

- (1) 1行を30マスとする。
- (2) 1マスから6マスで構成される1文字が行末から行の始めにかけて分離しないようにする。
- (3) カギ類、カッコ類の閉じ記号、句読点、つなぎ符類、捉音符等が行の始めに来ないようにする。
- (4) 同様に開き記号が行末に来ないようにする。
- (5) 3マス以上の余分な空白を削除する。

(6) 段落をつける。

(7) ページ付けをする。

この結果、以下に示す評価が得られた。

(1) 漢点字：仮名点字の場合と比較⁶⁾すると、点字量は30%の増加が認められた。また読む速度は、半年間漢点字を学習した人で30%の低下であった。しかし国語辞書の使用上の問題はないと思われる。

(2) 処理時間：本システムで辞書を点訳すると、1分間当たり5万字が生成される。

(3) 変換精度：仮名点字への変換精度は86%⁷⁾であるが、本システムの変換精度は自動編集も含めて100%である。

本システムにはまだいくつかの問題点が残されている。まず原データそのものに問題がある。たとえば、「- (ハイフン)」の代わりに「- (ダッシュ)」が使用されている。このコードの間違いは目で認識するうえにおいてはあまり問題としなくてもよいが、点訳上では大きな間違いとなる。

また原データには JIS コード表に規定されていない文字や記号が多く、JIS コード表に存在していても点字にない文字が多い。そのため国語辞書専用の特殊記号、漢字のみを変換する変換表をパッケージ形式でシステムに組み込み、点訳を行った。

5. む す び

パッケージ形式の変換表は辞書のほかに、JAPAN/

MARC、電算写植で生成された新聞、書物などを完成し、実用化した。今後の課題として、(1) パッケージ形式ではなく点訳システムに変換表を組み込むこと、(2) 現在点字は、1行に30字または32字と定められているが、将来を考慮してどのような形にでも適用できるように自動編集を改良する、などがある。

謝辞 本研究を進めるにあたり、三省堂刊新明解国語辞典磁気テープの使用を承諾して下さった(株)三省堂の各位に深謝いたします。

参 考 文 献

- 1) 川上泰一：八点式漢点字システム，第9回 IBM ウェルフェアセミナー報告集 (1979)。
- 2) 長谷川：六点漢字解説一覧表，六点漢字協会，東京 (1982)。
- 3) 日本点字委員会：改定日本点字表記法，p. 108，東京 (1980)。
- 4) 金田一京助 (編)：新明解国語辞典第11刷2版，p. 1248，三省堂，東京 (1974)。
- 5) 下村，水野：Machine Readable Data の自動点訳，信学技報，E 82-9 (1983)。
- 6) 下村，水野：自動点訳による国語辞書の分析，信学北陸大，B-54 (1982)。
- 7) 嘉手川，脇田：日本語点訳システム，第27回情報処全大，6 H-8 (1983)。

(昭和59年4月23日受付)

(昭和59年10月18日採録)