コンピュータと教育　19－2
（１９９１．１１．１５）

ＯＣＲと点字プリンタによる点訳支援

石田久之
（筑波技術短期大学）

本報告は、ＯＣＲと点字プリンタを中心として、点訳支援システムを構成し、点訳に要する時間を短縮するための方法を検討した。収集された典型的なＯＣＲの文字読み取り誤りから作成された編集コマンドによる自動編集を検討し、その有効性を示した。

しかし、点訳過程全体をみると、墨字修正と点字書式修正という２ヶ所の修正過程があり、前者を省略するために、墨字自動修正のための読み取り誤りデータの収集とともに、以下の点を検討する必要があることを論じた。（１）墨字修正を省くことによる、点字書式への変換の精度の低下、（２）その結果としての、点字書式修正のための時間の増加。

Support for the Braille Translation by the OCR and the Braille Printer.

Hisayuki Ishida
Tsukuba College of Technology

This paper reports a system for supporting the braille translation by the OCR and the braille printer. Typical misreadings of the OCR were gathered and used as commands for the automatic editing by the computer. The results indicated the effectiveness of the method.

The process of the braille translation contains two parts of corrections of data, the printing characters and the braille formats. To speed up the braille translation, the possibilities of omitting the correction of the printing characters is mentioned and the problems of further examination are described, especially the increase of the time needed to edit the braille formats.

1．はじめに

点訳といわれる、正眼者が通常用いる活字（以下、墨字という）を、点字に移し変える作業は、現在、主として、点字板と点筆、あるいは、点字タイプライタを用いて、一つ一つ行われている。これらの器具は、視覚障害者が、日常的にメモをとり、文書を作成する際に、用いるものであり、正眼者のボランティア等が用いて点訳作業を行う場合にも、何ら問題なく利用出来るものである。

しかしながら、このような点訳作業には、幾つか問題点を指摘できる。速さ、複数部作成の簡便さ等である。

これらの問題に対して、コンピュータを利用した、点訳システムがあり、市販されているものとしては、点字入力編集システム、全自動点字製版機械（小林鉄工所製）がある。これは、編集システムによって6点点字入力された、データを後者の点字製版機械で亜鉛板に出力し、この亜鉛板を用いて、印刷するものである。

更に、ＯＣＲ（光学式文字読み取り装置）を用いて、入力部分もコンピュータの制御下で行おうとする、試みもある（島田ら、１８８９、１９９０）。

2．目的

点訳を短時間に行おうとする装置・システムはいくつか有り、使用上の条件を持ちながらも、それぞれの領域で有効に機能している。

そこで、本報告は、「長文文字データの入力、及び、複数部数の点字出力」という条件内での、コンピュータによる点訳支援システムを、最初に紹介する。

次に、異なった書物間の読み取り誤りの特徴を比較し、ある図書での誤りを他の図書の誤りの自動修正に、適用できるか否かを検討する。

具体的には、以下の点を検討することを目的とする。

１）　ＯＣＲによる文字読み取り誤りの分類。特に、頻度について

２）　自動修正による文字読み取り誤りの特徴の変化について。

3．方法

ａ）　点訳支援システム

図１は点訳支援システムの構成を示している。パーソナルコンピュータ（ＮＥＣ製：ＰＣ９８０１ＶＸ４１）を制御用として中心に据え、入力装置として、ＯＣＲ（富士電機製：ＸＰ－５０Ｓ、イメージスキャナを含む）、出力装置として、点字プリンタ（ティール製：ＢＥＴＡ－Ｘ３）でシステムを構成している。ワークステーション、及び、デュプリケータについては以下の項に示す。

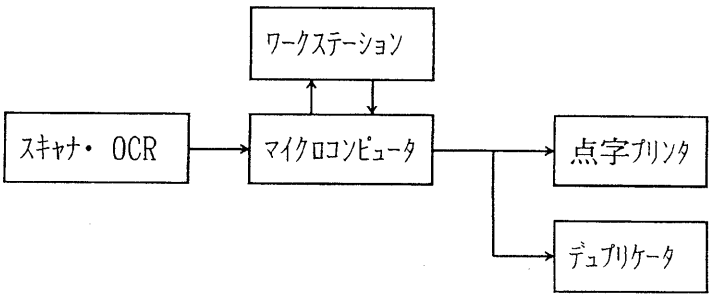

図1　点訳支援システム

ｂ）　データの種類とその保存

本システムでは２種類のデータが作成される。　一つは墨字データであり、他の一つが、点字書式データである。

墨字データは、通常の文字データなので、ワープロ等で内容を見、修正するこ

とが可能である。「点字書式」データとは、点字と同様の、符号の使い方、仮名遣い、分かち書き等がなされた、仮名だけの文字データである。点字そのものではないが、点字と1対1の対応があるので、「点字書式」とした。点字出力（図書・資料）は、このデータを、点字プリンタのコードに直して、プリンタへ与えれば得られる。

点字書式データは、点字独特の特別な記号が含まれるので、専用のソフトウェアが必要である。

この2種類のデータは、全過程が終了した後、デュプリケータ（アバールデータ製）と呼ばれるデータコピー機に保存される。この装置は、名前の通り、フロッピィーディスクのコピー機であるが、データ保存用としても用いることができる。必要に際して、この装置からデータをフロッピィーに出力し、提供する。

c）　点訳過程

図2は、点訳過程を示している。ＯＣＲによる墨字の読み取り、読み取りデータの修正、点字書式への変換、点字書式修正、点字出力の5つの段階が基本である。

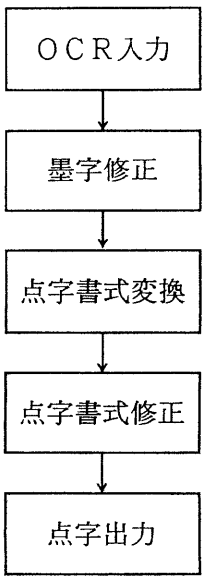

図2　点訳過程

なお、「2．目的」の条件で述べたように、本報告では、テキスト部分の点訳支援を目的とし、図の触図への移し変え等には、言及していない。また、ここでいう点字とは、6点の仮名点字である。

d）　データの収集

点訳の対象には、市販の図書4冊を選んだ（表1）。

表1　点訳図書

図書	出版社	総文字数
a	a出版	１０８，７００
b	b出版	７２，７００
c	〃	８８，４００
d	〃	９２，２００

図書a、bについては、上記の点訳過程にそって点訳作業を行った。ＯＣＲで墨字を読み取り、ワープロ、或いは、エディタを用いて、誤りをチェック、修正を行った。図書c、dについては、ＯＣＲ読み取りの後、ストリームエディタ（次項参照）による自動編集を追加し、以下、同様の作業を行った。

e）　ストリームエディタ

自動編集のためのエディタである。編集コマンドに従って、ファイルの内容を編集するものであり、ワークステーションには、標準に装備されている命令である。

図書c、dの点訳においても、基本的には、パソコンを用いているが、ＯＣＲで読み取った墨字データを一度、ワークステーション（東芝製、ＡＳ４３３０）に移し、ストリームエディタの特に、文字列置換コマンドによって自動編集し、このデータを再び、パソコンに移して、人間の手によ

る修正を開始した。

編集コマンドに記述する置換のパターンは、図書cについては図書bから、図書dについては図書b、cから、OCR文字読み取りの誤りをチェック、抜き出して、作成したものである（付図1）。

なお、以下の結果の記述においては、誤りの実数が示されている。作業時間は、そのなかに、実際の修正に要した時間と、誤りの抜き出しなどに要した時間が含まれているので、今回は、検討の対象とはしなかった。

4．結果

（1） 文字読み取りの誤り

表2は、図書a、bにおける、文字読み取りの誤りを分類したものである。

表2　文字読み取りの誤りの分類（回数）

分類	図書a	図書b
1→1	303	143
1→2	62	42
1→3	1	0
2→1	49	0
2→2	44	6
2→3	2	0
3→3	1	0
合計	462	191

上の表において、「1→1」は、1文字を、他の1文字に誤ったことを示している。

最も多い誤りは、1文字を他の文字に誤って読み取るというもので、aでは、全誤りの、65．6％、bでは74．9％である。

次は、1文字を2文字に、分けて、読み誤るもので、aでは13．4％、bでは22．0％である。

更に、図書aでは、2文字を、合わせて、別の1文字、あるいは、2文字に読み間違えるものがある（それぞれ、10．6％、9．5％）。

表3、4は、それぞれの図書の誤りの中で頻度の多い（10回以上）ものを示している。

表3　読み取り誤りの典型（図書a）

正		誤	回数
—	→	…	45
が、	→	ゕぶ	23
が、	→	ゕ斗	12
ま、	→	北	11
—	→	：	10
心	→	、D	10

表4　読み取り誤りの典型（図書b）

正		誤	回数
ロ	→	p	32
す	→	ず	29
合	→	台	14
ン	→	ソ	13
憶	→	億	12
は	→	ぱ	12

回数の多いもので見るかぎり、図書aとbで、同じ誤りは、認められない。また、図書aの特徴として、1つの文字を異なった文字に誤る場合が見られる（「が、」　→　「ゕぶ」、「ゕ斗」、「—」　→　「…」、「：」）。

（2） 自動編集後の誤り

表5は、ストリームエディタによる自動編集を行った後に見られる、文字読み取りの誤りを示し

ている。

表5　文字読み取りの誤りの分類（回）

分類	図書c	図書d
1→1	96	81
1→2	13	14
1→3	0	0
2→1	0	0
2→2	0	2
2→3	0	0
3→3	0	0
	109	98

誤りの総数には、減少がみられる。

誤りのタイプは、図書c、dともに、1文字を他の1文字、あるいは、他の2文字に読み誤るというのがほとんどである。

表6　読み取り誤りの典型（図書c、d）

正	誤	回数
（図書c）		
ン	→ ソ	55
順	→ □頂	8
す	→ ず	4
（図書d）		
す	→ ず	23
ン	→ ソ	14

表6は、図書c、dの誤りの典型例を、多い順に示している。ともに、「ン → ソ」、「す → ず」の誤りが、多い。

表7は、図書b、c、dに共通して多く見られる、「ン → ソ」、「す → ず」の、誤りについて、文字総数に対する割合を示したものである。

図書b、c、dの順に自動編集の編集コマンドは、増加しているのであるが、誤りの割合は、必ずしも、順に減少しているわけではなく、変動がかなりあるといえる。

表7　誤りの割合（%）

図書	す→ず	ン→ソ
b	3.0	3.6
c	0.4	6.5
d	2.4	2.5

5．評価

点訳作業には、幾つかの問題点を指摘できる。容易さ・速さ、そして、同一の資料を数多く準備したい場合の、複数部数作成の簡便さ等の問題である。

点訳とは、墨字を単に点字に変換すればよいというものではない。点字は仮名分かち書きでなければならない。見出し等の書きかたには、決まったいくつかの約束がある、等、始めての者がすぐに、点字をうてるというものではない。このような人間の側の問題とともに、用いられる器具についても、点筆と点字板、点字タイプライタ等の器具では、修正する場合に時間がかかる、一度におなじものを、複数作成することが出来ない、あるいは、簡単にコピーが出来ない等の問題がある。点字タイプライタでは、一度に、2～3枚はできないこともないが、それ以上は無理である。

この、コピーを簡単に出来ないという点は、同一の資料・図書等を大量に使用する場合に、大きな問題点となってくる。特に視覚障害者を対象とする教育機関等では、作成したい、あるいは、点訳したい資料を即座に点字化できること、また、長文の図書・資料を労少なくして、複数作成できること等が、重要な課題となっている。

このような課題に対して、コンピュータを用いて、点訳を行う試みがなされている。市販されているものには、点字入力編集システム、全自動点字製版機械（小林鉄工所）がある。これは、前者によって6点点字入力された、データを後者で亜鉛板等に製版・出力し、これを用いて、印刷するものである。

この機器を用いると、一度データが入力されてしまえば、大量印刷という点についての問題は解決する。しかし、データの入力という点に依然として、問題は残る。また、1、2部ではないとはいえ、せいぜい数十部程度でよい場合、複数部作製が容易という理由だけで、必ずしも安価とはいえない単用の機器を準備しておくということは、あまり賢明な方法ではない。

島田ら（１９８９、１９９０）は、パーソナルコンピュータと市販のスキャナでシステムを構築し、点訳の自動化を報告している。

正変換率８６．５１％を示しているが、点訳対象を小説本と限定し、更に、点字書式の詳細については、不明な部分がある。

そこで、本報告は、「長文文字データの入力、及び、複数部数の点字出力」という条件内での、コンピュータによる点訳支援システムを紹介し、自動修正による、点訳作業のスピードアップの可能性を検討しようとしたものである。

（１）　文字読み取りの誤りの分類

表2及び5より、4冊の図書に共通して、最も多い誤りは、1文字を他の1文字と読み誤るというもので、次は、1文字を2文字に間違えるものである。

更に、図書aでは、2文字を1文字、ないし、2文字に読み誤っている。このように、図書aと他の3冊には、特徴に違いがみられる。これは、図書（あるいは、出版社）による印刷形態等の違いが影響しているのであろう。図書aに見られるように、2文字を1文字に間違えるのは、文字切り出しの誤りであり、となり合う文字の間隔が小さいというような原因が考えられる。

読み取りの誤りは、表3、4、6から分かるように、また、当然予測されたことではあるが、似た文字と、間違える場合が多い。また、図書b、c、dでは、「す　→　ず」、「は　→　ぱ」等のような間違いがあるが、これは、それら3冊に共通した活字の特徴のためであろう。つまり、横線の右端が少し丸まっているために、濁点とみてしまうのが原因である。

一方、誤り方については、表3から見られるように、必ずしも一定していない。「が、」を「ゕぶ」や「ゕ斗」と読み誤ったり、「一」を「…」や「：」と誤ったりする。また、「ロ」を「p」と誤る場合が多いが、「q」とする場合も少なからずある。

以上をまとめると、活字の特徴により、誤り方が異なる（図書aと他の3冊）。類似した活字を用いている場合は、それが必ずしも、全く同じではなくても、異なった書物間で、類似した誤りの傾向を示す（図書b、c、d）。

（２）　自動編集の効果

表2及び5から、図書b、c、dと順に誤りの総数は、減少していることがわかる。文字総数全体（表1）を考慮すると、その傾向はより顕著であり、かなり有効な方法であると思われる。しかしながら、問題はある。表からもわかるように、どの図書においても、同じ様な誤りがその数は大小するが、見られる。例えば、「す」を「ず」と誤ることが多い。しかし、「ず」を全て「す」に置き換えるわけにはいかない。そこで「ず」を含む文字列の置換となるが、「ず」を含む文字列といっても、図書によって様々である。完全に置換するには、数多くの図書・文字列をチェックしなければならず、そのためには、多大な労力が必要となる。

結果として、どの程度のチェックが効率的であるかという議論になるが、今回の検討によると、2、3冊の図書をチェックすれば、同様な種類の図書については、このストリームエディタによる自動編集は、有効に機能すると思われる。

（３）　点字書式変換と点字プリンタ出力

点字書式変換には、専用のソフトウェアを用いる。前述のように、墨字の点字への変換は、必ずしも１対１の対応ではないので、単純ではない。意味内容を考慮した変換が必要であるが、本点訳支援システムで用いている変換ソフトウェアは、単純な文字列のマッチングを基本においている。

そこで、ここにも、ストリームエディタの自動編集機能を使用している。

墨字から点字書式への変換後の、データの修正の際に、変換ソフトウェアで生じる典型的な誤りをチェックし、これを次の図書の、変換の後に、ストリームエディタの編集コマンドとして、使用している。付図２は、この修正コマンドのリストである。これは、別の言葉で言えば、変換ソフトウェアの不十分度を示すリストである。

しかしながら、このようなリストを用いる自動編集機能は、十分な自然言語処理を装備していない、本システムにおいて、結果は検討中で、現在まだ、数量化されてはいないが、ある程度、修正のための時間を短縮するものと思われる。

ただ、ここで考えねばならないのは、点字書式データの修正とは、仮名分かち書きの誤りを訂正するというだけのものではなく、前述のように、見出しや注その他を、点字に特有のものに表示し直すという処理もあり、これにもかなりの時間がかかっているということである。この部分の自動化が必要と思われる。

なお、本システムには、同様な点字書式への変換ソフトウェアが他に一つある。こちらは現在、辞書の関係から、情報処理の分野の専用になっており、誤りのリストは、辞書として、変換ソフトウェアに随時登録され、使用されている。

本システムの最終段階は、点字プリンタによる出力である。点字プリンタは、通常ＮＡＢＣＣコードを用いている。前述の点字書式データとは、１対１の対応があるので、単にコード変換のフィルタを通して、出力している。

（４）　今後の課題

本報告は、点訳支援システムにおいて、点訳に要する時間を短縮するための方法の一つとして、自動編集機能を検討し、その有効性を示した。

しかし、点訳過程全体をみると、２ヶ所の修正のための過程がある。つまり、墨字修正と点字書式修正の部分である。ここで人間の手による墨字修正を省けないであろうかという問題提起ができる。

今後、墨字データ自動修正のためのデータの収集（コマンドリストの充実）、及び、墨字修正を省くことによる、点字書式への変換の精度の低下、点字書式修正のための時間の増加等の検討によって答えを得る必要があると思われる。

６．参考文献

島田恭宏、塩野充（１９８９）：パーソナルコンピュータを用いた文庫本小説の印刷漢字認識実験、テレビジョン学会誌、４３（８）、８１６～８２３

島田恭宏、塩野充（１９９０）：パーソナルコンピュータを用いた小説本の自動点訳システム、テレビジョン学会誌、４４（１１）、１５９６～１６０４

```
・・・・・・・・・・・・
s/.セ.ブラ/アセンブラ/g
s/.セン.ラ/アセンブラ/g
s/.ログ.¥([ムミ]¥)/プログラ¥1/g
s/.頂¥([次序番]¥)/順¥1/g
s/.頂編成/順編成/g
s/ゼＰ/ゼロ/g
s/[１１]勺/的/g
s/[すず]べ/すべ/g
s/[すず]べ/すべ/g
s/[デテ]ー[ヌタタ]/データ/g
s/[口ｐｑ][ーー]/ロー/g
s/¥([みしい]¥)まず/¥1ます/g
s/¥([ミモ]¥)ニタ/¥1ニタ/g
s/¥([降昇]¥).頂/¥1順/g
s/¥([場都問併具集]¥)台/¥1合/g
s/¥([明願文]¥)害/¥1書/g
s/１[主ばぱまよょ]/は/g
s/１[主ばぱまよょ]/は/g
s/１頂/順/g
s/ＣＰＯ/ＣＰＵ/g
s/丘ｌｅ/　ｆｉｌｅ/g
s/なし、。/ない。/g
s/のよらな/のような/g
s/[ぺぺ]き/べき/g
・・・・・・・・・・・・
```

付図１　墨字自動編集コマンドの一部

```
・・・・・・・・・・・・・・・・
s/によ¥([らりるれ]¥)/に　よ¥1/g
s/にたい¥([すしせ]¥)/に　たい¥1/g
s/とともに/と　ともに/g
s/につれて/に　つれて/g
s/にしたが/に　したが/g
s/において/に　おいて/g
s/にもと¥([づず]¥)/に　もと¥1/g
s/にあたって/に　あたって/g
s/¥([こそ]¥)のーち/¥1の　うち/g
s/¥([こそ]¥)の　ちゅー/¥1の　なか/g
s/¥([^がててとにわを]¥)　する/¥1する/g
s/¥([^がててとにわを]¥)　され/¥1され/g
s/ひとりひとり/ひとり　ひとり/g
s/¥([：・]¥)¥([^　]¥)/¥1　¥2/g
s/¥([^０-ｚ]¥)，¥([^０-ｚ]¥)/¥1、　¥2/g
#s/.　*/.　　/g
s/。　　¥n/。¥n/g
s/　*(/(/g
s/できるかいなか/できるか　いなか/g
s/にともな/に　ともな/g
s/¥([こそ]¥)のーえ/¥1の　うえ/g
s/なくなる/なく　なる/g
s/てくる/て　くる/g
s/¥([^　]¥)なくして/¥1　なくして/g
・・・・・・・・・・・・・・・・
```

付図２　点字書式自動編集コマンドの一部