

# 日本語自動点訳ソフトウェアの 技術的問題点と改良の指針

栗原 亨, 河原 正治, 斉藤 正夫, 斎藤 玲子, 長岡英司

〒305 茨城県つくば市春日4-12

筑波技術短期大学視覚部

石川 准

〒422 静岡県静岡市谷田

静岡県立大学国際関係学部

河村 宏

〒113 東京都文京区本郷7-3-1

東京大学総合図書館

小山 智史

〒036 青森県弘前市文京町1

弘前大学教育学部

日本語文書を点字に変換するソフトウェアの改良の指針を示すために、人手による点訳作業の過程を解析し、ソフトウェア技術の達成度を評価する。現在公表されているソフトウェアは、文科系の文書の点訳に関して、正しく使えば実用的である。しかし、数式を含む理工系文書の点訳を対象とするものではなく、漢字をかなに変換する精度も不十分である。自動点訳の達成には、点字記号体系の整備と自然言語の意味解析や2次元表現処理の開発が必要だが、当面は、点訳作業を人手による部分と計算機による部分に分割し、それに合わせたソフトウェアの開発と前後処理をする専門的点訳エディタの養成が必要である。

Technical Problems of Japanese Automatic Braille  
Transcription Software and a Guideline to the Improvement

Tooru Kurihara, Masaji Kawahara, Masao Saito, Reiko Saito,  
Hideji Nagaoka

4-12, Kasuga, Tsukuba, 305 Japan Tsukuba College of Technology

Jun Ishikawa

Yata, Shizuoka, 422 Japan University of Shizuoka

Hiroshi Kawamura

7-3-1, Hongo, Bunkyo, 113 Japan

General Library, University of Tokyo

Satoshi Koyama

1 Bunkyo-machi, Hirosaki, 036 Japan Hirosaki University

The process of Braille transcription by human transcribers is analyzed to improve the software which transcribe Japanese text including Kanji and Kana into Braille. The programs at present are effective if they are properly used for the text in human sciences. None of them can treat the text including mathematical or scientific notation. The precision of conversion from Kanji to Kana is insufficient. The completion of the Braille code system and the technical advances in the analyses of natural language and 2-dimensional expression. However, the process should be divided into the one by human transcribers and the one by computers as a realistic solution.

## 1. はじめに

点訳は、点字での読み書きを必要とする視覚障害者のために、墨字（すみじ、普通の文字）の文書を、点字という6ビット記号体系の触知文字に翻訳する作業である。近年、視覚障害を持つ大学生数の増加や点字の普及に伴い、学術文書を含む様々な文書の点訳の需要は増大する一方だが、点訳は高度技術で点訳者数も少なく、点字文書の供給は不十分である。点訳は人手によって行われて来たが、作業の効率化と技術水準の維持発展のために、コンピュータを用いた自動化が必要である。

アメリカ合衆国では、1980年代にMS-DOSが普及するにつれてパーソナルコンピュータ上で動作する英語自動点訳ソフトウェアが開発され、数年前には実用の域に達していたが、日本では独特な言語体系と精神風土のために、点訳の自動化は遅れていた。当時日本では、一般に「点訳ソフト」と呼ばれる点字文書を編集するソフトウェアが開発され、利用されていたが、これには墨字文書を点字に変換する機能はなかった。

1990年前後から、パーソナルコンピュータのMS-DOSで動作する漢字かな混じり文を点字に変換するソフトウェアがいくつか発表され、視覚障害を持つコンピュータ利用者に知られるとともに、点字に関係する学校、施設、企業、団体などで利用されるようになった。これらのソフトウェアは点訳の対象を、数式や数学・理科記号を含まない文科系の文書に限定して適当な前後処理とともに使えば、十分に実用的である。

しかし、点字を知らない人々の点字に対する安易な考えと、ソフトウェアに対する過信とから、視覚障害者に判読困難な点字文書が届くといった混乱も起こっている。自動点訳は点字を知らない者にも点字が書けるという魅力があり、点字の普及に有用だが、現在の技術は未熟で、使い方に注意を要する。

自動点訳は、自然言語の意味や紙面上の記号の2次元配列の解析を伴うので、目標の完全達成は困難である。しかし、それを幾万もの視覚障害者が必要としているので、その開発はガンやエイズの制圧と同様に総力を尽くして解決すべき人類の課題である。自動点訳がガンやエイズの制圧に比べて困難なのは、その開発の経済効果が期待できないと思われるためか、開発者の数が少なく、分野全体の開発能力が低いことである。しかし、この技術は完成すれば必ず一定の人数の利用者が常に存在する「本物の技術」である。本物の技術は応用も無限で、経済効果も期待できる。

自動点訳技術開発の問題は、まず、点訳という技術そのものの存在と意義が社会に認識されていないことにある。幸い、近年のパーソナルコンピュータやワークステーションの発展にはめざましいものがあり、点字プリンタや点字表示器も使いやすいものが発表されて、ソフトウェアの開発環境は整備されてきた。ここに、点訳作業の実態とソフトウェアの現状を紹介して、広く一般の理解を求め、今後の開発に有益な情報収集の契機としたい。

## 2. 点訳の手順と自動点訳ソフトウェアの機能

### 2. 1. 点訳の手順

自動点訳は、優秀な点訳者の作業をコンピュータに代行させるものである。ソフトウェアの機能を検討するために、人手による点訳作業を解析してみよう。

点訳の手順を示す。

【墨字原本の下読み】

### (1) 対象の評価

点訳も翻訳である。いかなる翻訳も100パーセント翻訳可能なものはない。最大効率の翻訳を追究するのが翻訳の使命である。最大効率の点訳を行うには、点訳対象の評価が必要である。

点訳は依頼人の要請に基づいて行われるものであるから、原本の評価に先だって、依頼人の意図を把握する必要がある。

評価項目：

- ◇ 緊急性： 納期の妥当性。
- ◇ 点訳精度： 不特定多数が使う教科書か？、個人的な資料か？。
- ◇ 点字書式の精度： 依頼人の特別な要望の有無を考慮する。
- ◇ 要求項目の優先順序： 納期を優先するか？、精度を優先するか？。

以上は原本の内容とは独立の評価基準。

以下は原本の内容に関する評価。

- ◇ 分野： 分野別点字記号体系の選択、局所的な記号体系の切り替え。
- ◇ 点訳困難箇所： 図表、特殊記号の対策。
- ◇ 難易度： 点訳者の技能に応じた分担の調整。
- ◇ 作業量： 緊急度と精度を調整した作業計画の立案。

### (2) 方針の決定

前述の各項の評価に基づいて作業計画を立案する。点訳者の判断を要する項目は以下の通り。

- ◇ 納期
- ◇ 記号体系の選択
- ◇ 作業の体制： 点訳者数、分担等。
- ◇ 作業手順： 全て人手によるか、一部ソフトウェアを利用するか。

### (3) 作業の準備

整備項目：

- ◇ かな訳資料： 漢字→かな変換のための、国語・人名・地名辞典、百科事典、専門分野の辞書・事典、教科書。
- ◇ 点訳資料： 点字記号表、点訳マニュアル。[1, 2, 3, 4]
- ◇ 点訳用機器： 点字盤、タイプライター、パーソナルコンピュータ、点字プリンター。
- ◇ ソフトウェア： エディタ、点字エディタ、自動点訳ソフトウェア。

### (4) 点字原稿の作成

作業内容の要素：

- ◇ かな訳： 固有名詞・専門用語の読み。
- ◇ 分かち書き： 点字表記法に基づくマスあけの挿入。
- ◇ 日本語と外国語の書き分け： 外国文字・綴り表示と外国語単語・文章表示。
- ◇ 特殊記号の変換
- ◇ 数式表記： 添え字・行列表現処理。
- ◇ プログラムリスト表記

- ◇ 本文中の数学・理科記号の表記
- ◇ 標準書式作成： 見出し行のインデント、改行のインデント、ワードラップ等。
- ◇ 特殊書式作成： 和歌、俳句、詩等の文学形式に固有な書式表現。
- ◇ 臨時書式作成： 内容の特殊性による臨時の特殊書式の設定。
- ◇ 図表の代替表現
- ◇ 局所的特殊表記法の設定
- ◇ 点訳者注の挿入： 局所的特殊表記法の指示と原書表記に関する補注。
- ◇ ページ番号処理： 原書ページとの対応表示等。
- ◇ 目次・索引作成： 原書目次、点訳書目次。
- ◇ 印刷、製本

### (5) 点字原稿の校閲

点字使用者自身による校閲が重要。

## 2. 2. 自動点訳ソフトウェアの機能

### 【主要機能】

- ◇ 電子化された墨字文書記録を、完全な点字文書の電子ファイルに変換する。

### 【具体的機能】

- ◇ かな訳
- ◇ 分かち書き
- ◇ 外国語処理
- ◇ 専門分野別記号処理
- ◇ 数式処理
- ◇ 書式処理
- ◇ 印刷制御

## 3. 日本語点訳ソフトウェアの現状

### 【基本機能】

現在、日本語自動点訳ソフトウェアが共通に備えている機能は、以下の通り。

- ◇ 対象分野： 文化系文書。
- ◇ かな訳
- ◇ 辞書登録： かな訳や点訳のためのユーザー辞書登録。
- ◇ 分かち書き
- ◇ 書式処理： ワードラップ、開業コード処理、ページ番号付け。
- ◇ プリンター制御

### 【公共されている日本語自動点訳ソフトウェア】

現在、少なくとも4種類のソフトウェアが公表され、利用されている。

- ◇ EXTRA（（株）アメディア）
- ◇ がってんだ（（株）言語工学研究所）
- ◇ 点訳ぶうちゃん（松山市視力障害者友の会）
- ◇ 80点（福祉システム研究会）

### 【亜種】

視覚障害者がパーソナルコンピュータを利用するために、モニター画面へ出力される文字情報を合成音声で読み上げる、画面読み上げソフトウェアというものがある。現在、日本電気(株)や富士通(株)のパーソナルコンピュータの日本語MS-DOSに対応するものがいくつか発表されているが、それらには、かな訳機能がある。実際、前述の自動点訳ソフトウェアの中には、画面読み上げソフトウェア開発の副産物として開発されたものもあるが、一般に画面読み上げソフトウェアは自動点訳ソフトウェアの予備軍になっている。

#### 【かな訳精度の評価】

前述4種の日本語自動点訳ソフトウェアのかな訳の精度については、福井哲也氏の報告[5]がある。その結論は、多少の差はあるものの、日常的な数学・理科記号を含まない文書については、ソフトウェアによる漢字かな変換が実用の域に達したことを示している。

この結論は、本論著者の使用経験のない「点訳ぶちゃん」を除いて、本論著者の経験的印象に大きく異なるものではない。

#### 【分かち書きの精度の評価】

福井氏の報告[5]は、前述4種のソフトウェアの分かち書きの性ども実用の域に達したことを示している。これこれも、使用経験のない「点訳ぶちゃん」を除いて、本論著者の印象と一致している。

#### 【達成度の評価】

現在、次のことはできる。

- ◇ 数学・理科記号を含まない日常的な文書の、実用的精度での、かな訳と分かち書き。
- ◇ オフライン辞書(変換表)登録。
- ◇ ある程度の書式処理。
- ◇ ある程度のプリンター制御。

現在、次のことはできない。

- ◆ 完全自動点訳。
- ◆ 理科系文書の処理。
- ◆ 数式処理。
- ◆ プログラムリスト処理。
- ◆ オンライン辞書登録。
- ◆ オンライン対話型の再変更可能な書式処理。
- ◆ オンライン対話型再変換。

## 4. 自動点訳の問題点

### 4.1. なぜ難しいのか

#### 【技術的問題点】

- ◇ 点字記号体系の不備：点字表記法が存在しない部分があり、翻訳不能な場合がある。
- ◇ 自然言語処理技術の未熟：文章の意味を解析する技術の開発が必要。
- ◇ 数式処理技術の未熟：紙面上に2次元的に配列された記号の式を、コンピュータの内部表現に適した形式に変換する、取扱いの容易な公認された手続きの規格がない。これは、原本の電子ファイルの作成の障害になっている。

#### 4. 2. 点字記号体系の問題点

##### 【点字の宿命】

点字は指先の触覚で判読する文字である。点字の点と点の間隔と1マスの点の数は、平均的なヒトの生理的限界によって定められ、縦3点、横2点、1マス6点の表現形式が定着している。これは墨字の1文字に相当する1マス記号の数が、わずか64個に限られることを意味するが、この制約が点字記号体系の困難を運命づけている。

##### 【墨字の自由度／点字の自由度】

現在、人類社会全体に存在する墨字記号の数は、漢字の数よりやや多い程度であろうから、約10万個であろう。これを冗長性のない2進数で無駄なく表現した場合、少なくとも17ビットの記号体系が必要である。1マスの点字は6ビットであるから、この二つの記号体系の間には少なくとも11ビット、約2千倍の大きさの隔りがある。

##### 【記号体系の簡約】

この隔りを埋めるために、Louis Braille（ルイ・ブライユ）を元祖とする多くの点字記号制定者は、以下の方針で記号体系の簡約を断行した。

《誰にも当たり前なこと、誰も知らないことは表現しない。》

- ◇ 点訳対象の限定： 大部分の点字利用者が利用しない墨字記号は、点訳の対象としない。（漢字や特殊記号の無視など。）
- ◇ 第二義的な視覚情報の削除： 音読したときに直接表現されない視覚情報は削除する。（ひらがなとカタカナの共用、書体の区別の無視、各種カッコ類の共用など。）
- ◇ 前置フラグの採用： いくつかの前置符号を定義して、その後の記号を読み分ける。前置フラグによって生じたシフト状態をリセットする規則がある。（大文字符、外字符、数符、外国語引用符など。）
- ◇ 直後の記号による判別： 直後の記号の種類、特に、直後がマス空けであるかどうかによって、その記号を読み分ける。（句読点記号と特殊音前置符との区別など。）

##### 【点字と暗号】

墨字から点字への変換規則は、その点字を使用する一定の人間集団の中で刻々と変化する常識を前提として制定されている。このようにして制定された記号体系は、常に解読表を必要とする暗号に似ている。暗号文は解読表が変わればその意味が変わる。そして、日本の点字表記の実際を見ると、文書内容の分脈や点訳者の注によって、適宜変換表を切り替えている。厳密には、日本の点字文書は、途中から読んだのでは正確に読めないし、墨字の原稿を途中から与えられたのでは、正確に点訳できない。このような記号体系への翻訳を、普遍的な単一の規則で行うことは困難である。

##### 【最適の簡略】

点字記号体系の制定者は、常に個人的効率性と社会的普遍性というジレンマの間で苦悶してきた。個人的には、専門分野を限定すれば、64個の記号で十分な場合もあるだろう。1マス記号のみを用いて作業ができれば、その利用者の作業効率は最高だが、そのような体系が利用者ごとに全て異なれば普遍性を欠き、情報交換の手段とはならない。逆に、全ての墨字と一対一対応を持つ点字記号体系を作れば、墨字記号の1個に相当する点字記号のマス数が多くなって誰にも使いにくいものとなり、普遍的な情報交換の手段とはならない。

##### 【現状】

現実には、前述の両極端の間に位置する記号体系が、それぞれの地域、それぞれの分野で、必要に応じて制定されていて、時には、一つの国の一つの分野の中でも、複数の記号体系が互いに優劣を競っているのである。分野が異なれば、一つの墨字記号に対して複数の点字記号が存在することも日常茶飯で、とりわけ、理工系の分野では、混乱を極めている。

#### 4. 3. ソフトウェア技術の問題点

##### 【ソフトウェア以前の問題】

- ◇ 原本の電子ファイルの不足。
- ◇ 原本の電子ファイル作成の困難。
- ◇ 入力手段の未熟

##### 【変換の問題】

- ◇ 自然言語処理の困難。
- ◇ 2次元配列処理の未熟。

##### 【使い勝手の問題】

- ◇ 書式処理の未熟。
- ◇ ユーザー辞書登録の柔軟性の不足。

#### 4. 4. 人手作業の必要

人手作業を必要とする理由：

- ◇ ソフトウェア技術の未熟： 当面は人手で行った方が効率的な部分がある。
- ◇ 前処理： 分野別記号体系の選択や図表処理の指定等、点訳者の判断が必要。
- ◇ 後処理： 出力書式の調整等、点訳者の判断が必要。
- ◇ 校閲： 点字利用者自身による点字の校閲が必要。

#### 5. 今後の開発の指針

三段構えの開発体制が必要である。自動点訳の開発は大規模事業で、目標を長期、中期、短期のそれぞれについて設定する必要がある。

##### 5. 1. 長期計画

###### 【入力手段】

- ◇ 点字記号体系の整備。
- ◇ 電子ファイル書式規格の標準化。
- ◇ 文字コードの標準化。
- ◇ 計算機コードによる数式表記法の標準化。
- ◇ 出版用電子ファイルの提供の制度化。
- ◇ 光学的文字認識技術の改良。

###### 【変換技術】

- ◇ 自然言語処理技術の開発。
- ◇ かな訳用データの整備。
- ◇ 点訳用データの整備。

###### 【製作体制】

- ◇ 点字文書製作施設の整備、拡充。
- ◇ 点字編集専門職の確立： 計算機技術に習熟した点字利用者の育成。

## 5. 2. 中期計画

### 【変換技術】

- ◇ かな訳・分かち書き精度の改善： 掛かり解析技術改善と変換用辞書の増強。

### 【編集機能】

- ◇ オンラインかな訳用辞書。
- ◇ オンライン点訳マニュアル。

## 5. 3. 短期計画

### 【編集機能】

- ◇ 対話型前処理機能： 原本データに分野別記号体系の選択の指示を埋め込む。
- ◇ 対話型ユーザー辞書登録機能。
- ◇ 対話型書式編集機能。

## 6. おわりに

公の文書は誰にでも読める必要がある。目の見えない人にも読める必要がある。目の見えない人も人間である以上、公の文書の中に目の見えない人だけに読めないものが存在することは、本当は異常なことである。現在の目が見える人に都合のよい書式は、人類社会が選択した方式であって、絶対ではない。今、目の見えない人に一般の文書が読めないのは、社会のシステムがそうになっていないから仕方がないのであって、目が見えないから仕方がないのではない。

文書記録の標準書式は人類の自由意志で変更可能である。そして、それを可能にする技術の開発が、どうしても必要である。

広範な理解を切望する。

## 謝辞

本研究は、財団法人日本失明予防協会の助成を受けて行われた。関係各位に深く感謝します。

## 文献

- [1] 日本点字委員会：日本点字表記法（1990）
- [2] 日本点字委員会：点字数学記号解説（1981）
- [3] 日本点字委員会：点字理科記号解説（1983）
- [4] 日本点字委員会：コンピューター用語の6点式点字表記—相互変換用点字専門委員会報告—，日本の点字，第9号（1981）
- [5] 福井哲也：日本語自動点訳ソフト4種の精度の比較，第2回視覚障害リハビリテーション研究発表大会論文集，視覚障害リハビリテーション協会，， P P . 114-117（1993）