

テキストエディタを利用した日本語文書の自動点訳

Automatic Japanese-to-Braille Translation using Text Editor

小 山 智 史*

Satoshi KOYAMA

1. はじめに

視覚を失うことは情報の入手に大きな障害である。視覚障害者にとって、点字印刷物は、晴眼者¹にとって印刷物が重要であるのと同様の意味で重要である。

視覚障害者における文字情報の利用の仕方は、概ね次のようであると考えられる。(1)点字印刷物を入手して読む。(2)電子化された情報を入手し、パソコンのディスプレイの音声読み上げなどで読む。(3)電子化された情報を入手し、点字プリンタに出力し、点字印刷物として読む。本論文は、(1)および(3)を支援する技術に関するものである。

最近は文字情報の電子化が著しい。視覚障害者は、電子化文書を入手すると、音声読み上げで読むことができるが、(a)自分のペースで読める、(b)「読み返し」や「読み飛ばし」など読み方の自由度が高い、(c)特殊な装置がいらないなどの理由で、点字で読みたいという要求は強い。

点訳、すなわち「墨字²から点字への翻訳および点字印刷物の作成」は、その多くが点訳者の手作業により行われているが、点字本の供給は圧倒的に不足しており、点訳の「省力化」と「生産性向上」が切望されている。

情報機器を活用して点訳を支援する試みも検討され、一部実用化されてきたが[1][2][3][4]、日本語文書の自動点訳の実用性が確認されたのはごく最近である[5][6][7]。なお、ここでは、「電子化された漢字かな混じりの日本語文書（以後単に日本語文書と言う）を分かち書きされたカナ文字列に相当する点字データ（以後単に点字データと言う）に変換し、点字プリンタで印刷すること」を「日本語文書の自動点訳」と呼ぶ³。

筆者は、非対話型のテキストエディタを利用して文字列の単純な置換を行うことにより日本語文書を点字データに自動変換し、点字印刷⁴する自動点訳システムを開発した。多少の誤りを含む点字であることを承知でそのまま点字印刷して利用するか、あるいは点字データの誤りを修正してから点字印刷する。いずれの場合においても実用的であると判断された。

本論文では、点訳とその方法について述べた後、開発した自動点訳システムについて説明し、点訳事例を紹介する。

¹ 視覚障害者に対し健常者のことを晴眼者という。

² 点字に対し普通の印刷文字を墨字（すみじ）という。

³ 他に、文字の読み取りを含めて自動点訳という場合や、点字の漢字を用いる場合も自動点訳と呼ぶことがある。

⁴ TPR, TDC, ブレイルスター、コータクンなどのソフトを用いて点字プリンタで印刷する。使用するソフトにより点字データの表現方法に多少の違いがあるが、多くの場合相互に変換できる。

2. 点訳とその方法

一般に普及している日本点字表記法[8]に従った点字は「カナ点字⁵」と呼ばれている。漢字の混じった文書は、点字の漢字⁶を使わない限り直接点字で表現することはできないので、日本語文書をカナ点字に変換することは重要な意味を持つ。

点訳の主要な規則は、「漢字を正しい読みに直すこと」と「分かち書き」である。また、助詞の「は」「へ」は「わ」「え」に、う列、お列に続く長音「う」は長音符「ー」に書き換える。例えば、

日本語文書：私は今日学校へ行ってきた。

点字データ：ワタクシワ□キョー□ガッコーエ□イッテ□キタ。

となる（□はスペースを示す）。

このような点訳処理をどのように行うかについて次に述べる。

(1)従来の点訳

従来、点訳は、専門の教育を受けた点訳者が、日本語文書を点字データに翻訳しながら、点字器（点字板と点筆）や点字タイプライタを用いて点字本を作ってきた（図1 a）。しかしながら、

- ・点字本は一度に1部しか生産できない
- ・修正や編集が困難である
- ・作業能率が低く、点訳者の負担が大きい

などの問題があった。

(2)パソコン点訳

最近になって、パソコンを使って点字データを入力し、点字プリンタで印刷する方法が普及してきた（図1 b）。この方法はしばしば「パソコン点訳」と呼ばれている。点字データの電子化により、データの再利用が可能となり、従来の「点字本が1部しか生産できない」という状況が改善されるようになった。また、点字データの修正や編集も簡単にできるようになった。しかしながら、操作には点訳の専門知識が必要であるため、点訳者の支援にとどまっている。

(3)自動点訳

日本語文書を点字データに自動的に変換し点字プリンタで印刷する方法が「自動点訳」である（図1 c）。文書が電子化されていない場合は、ワープロや文字読み取り装置（OCR）を使って文書を入力する。

日本語の点訳規則を、考え得るあらゆる日本語について厳密に記述することは難しい。従って自動点訳の変換精度には限界がある。点字に多少の誤りが許されるならば、自動変換された点字データをそのまま点字プリンタで印刷すれば良い。電子化された文書を視覚障害者が独力で点字で読むことができる。誤りの無い点字印刷物を作る場合には、点訳者が点字データの誤りを修正する。この場合も、点訳者の作業は誤りの修正だけで良いため、負担は大幅に軽減される。

⁵ 目的によっては JIS-X-0201 (旧 JIS-C-6220) に対応した「情報交換用点字」が用いられることもあるが、一般に「カナ点字」が広く用いられている。

⁶ 全ての漢字について点字符号を定義したもので、現在「六点漢字」と「漢点字」の2種類の体系が提唱されている。修得は必ずしも容易ではない。

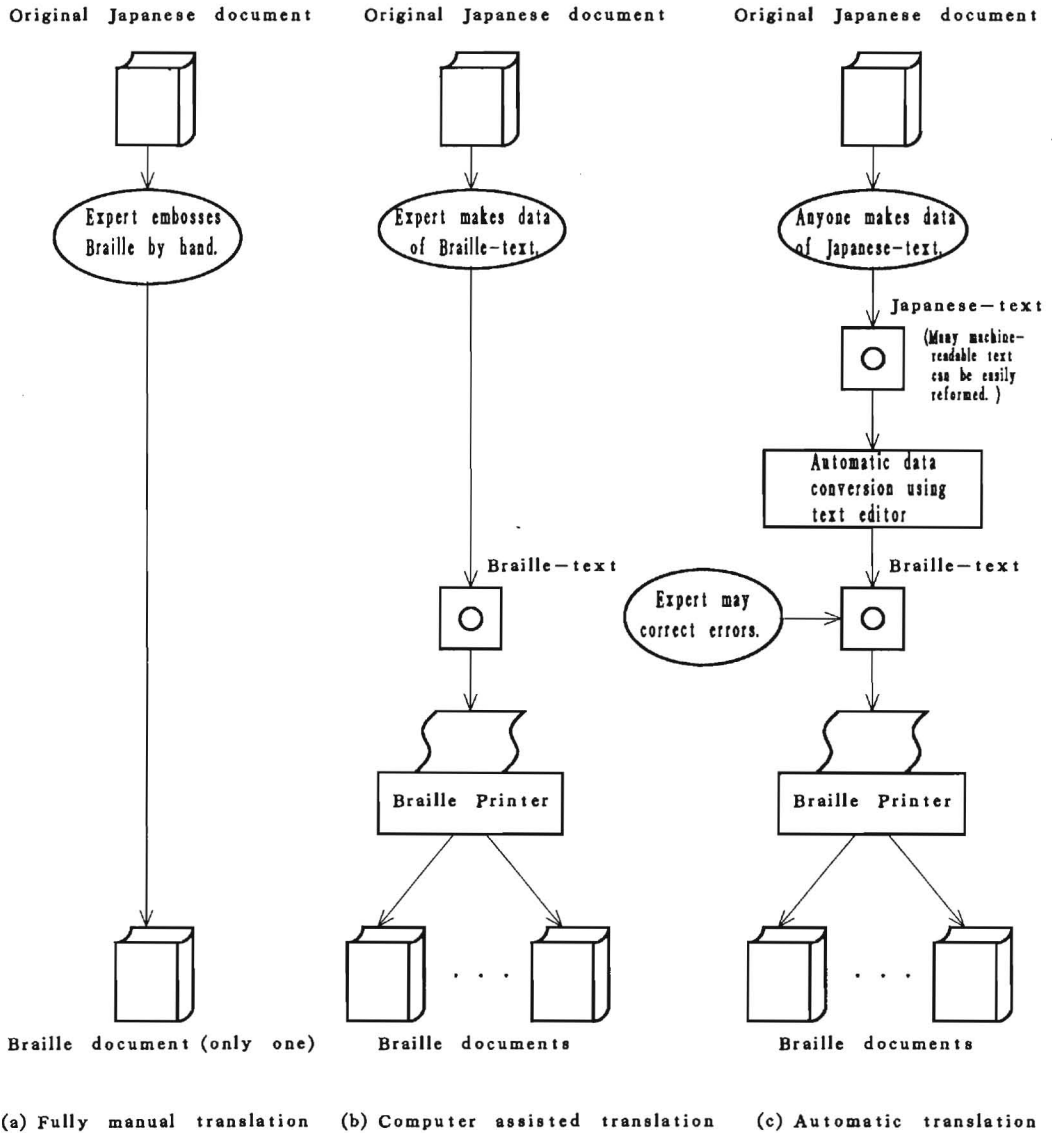


図1 点訳とその方法

Fig. 1 Methods of Japanese-to-Braille translation

3. 自動点訳システムの構成

図2に開発した自動点訳システムの構成を示す。(1)~(3)で用いる変換規則 81.sub, ztoh.dic, 82.sub が、新規に作成したプログラムで、MS-DOSをOSとするパソコンで動作する。(4)は任意のテキストエディタを利用する。

(1) 分かち書き処理

非対話型のテキストエディタ (sed[9]) とそのコマンド群 (81.sub) によって、漢字かな混じり文を分かち書き処理する。

例えば、

s /¥([^ あーん] [がにのはもや] ¥) /¥ 1 □ /g

というコマンドは、「かな以外」の文字の後に「が」「に」「の」「は」「も」「や」のいずれかが

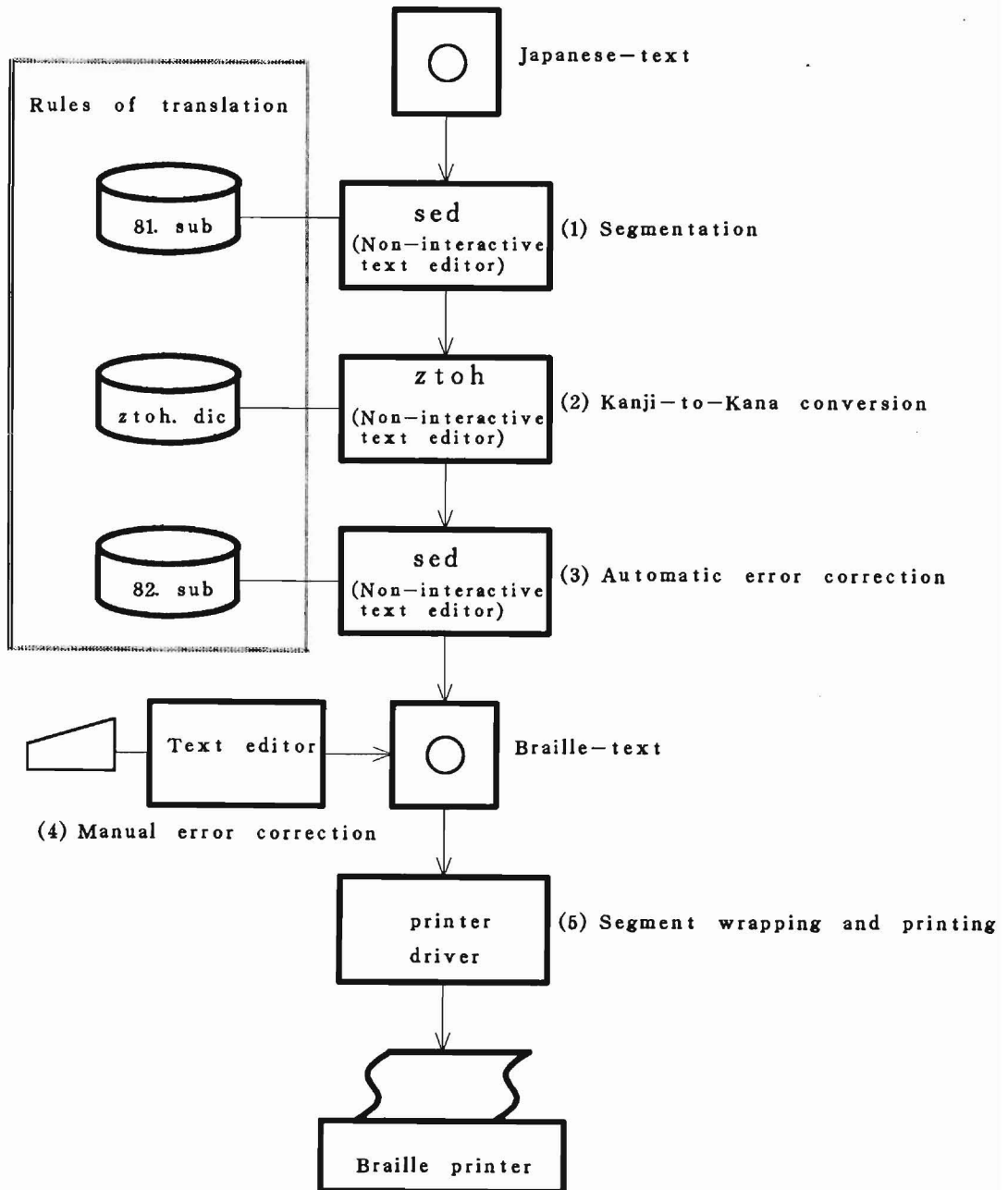


図2 自動点訳システムの構成

Fig. 2 Automatic Japanese-to-Braille translation system

来たら、その後にスペース (□) を挿入する」ことを意味している。また、

s / は ¥ ([^ あーん] ¥) / わ □ ¥ 1 / g

というコマンドは、「は」のあとに「かな以外」があれば、「は」を「わ」に置き換え、その後にスペース (□) を挿入する」ことを意味している。これらは、「助詞の「は」を「わ」に換えその後で分かち書きする」という点訳規則を表現したものである。また、

s / ¥ (て ¥) ¥ (き [たてま] ¥) / ¥ 1 □ ¥ 2 / g

というコマンドは、「て」のあとに「きた」または「きて」または「きま」が来たら、その間にスペース(□)を挿入する」ことを意味している。入力文書の各行について、196個のコマンドが順に適用される。これにより、

私は今日学校へ行ってきた。

→私わ□□今日学校え□行って□きた。

となる。複数の規則が適用されると、スペースは余分に挿入される。

コマンドは利用者が追加修正できる。

(2)漢字かな変換処理

(1)で分かち書きされた全角の漢字かな文書を、半角カナ文字に変換する。変換プログラム(ztoh)は、漢字で始まる文字列の置換を高速で実行する[10]。変換辞書(ztoh.dic)の一部を付録に示した。辞書は逆順にソートされている。漢字で始まる文字列について、辞書の先頭から一致を調べ、一致したら対応するカナ文字列に置換する(最長一致置換)。次に漢字が現れるまでの文字は、対応する半角文字に変換する。2文字以上漢字が続く場合は、原則として前にスペースを挿入するように辞書を用意し、(1)の分かち書き処理を補っている。これにより、

私わ□□今日学校え□行って□きた。

→ワタクシワ□□□キョー□ガッコーエ□イッテ□キタ。

となる。「今日」と「学校」は、カナに置き換えられると同時に、その前にスペースが挿入される。「行って」は「イッテ」に置き換えられているが、「...を行って」という文章が現れたときのために、(1)のコマンド群の中には、

s / を行っ / を□おこなっ / g

というコマンドが用意されている。

変換辞書には約3万4千語が登録されており、利用者が追加修正できる。

(3)補正処理

非対話型のテキストエディタ(sed)とそのコマンド群(82.sub)によって、余分なスペース(複数の分かち書き規則が重複して適用されることによる)の削除など補正処理を行う。

例えば、

s / □□ * / □ / g

は、2個以上連続するスペースを1個にする。これにより、

ワタクシワ□□□キョー□ガッコーエ□イッテ□キタ。

→ワタクシワ□キョー□ガッコーエ□イッテ□キタ。

となる。また、

s / ¥ ([ラー] ¥) ¥ ([a-z A-Z] ¥) / ¥ 1 □ ¥ 2 / g

は、カナとアルファベットの間にスペース(□)を挿入する。このように、点字の表現に必要な処理もここで行っている。

コマンド群は69ステップからなり、利用者が追加修正できる。

以上3つの処理によって、漢字かな混じりの日本語文書が、分かち書きされたカナ文からなる点字データに変換される。

(4)点字データの修正

(1)~(3)の処理で、必ずしも誤りの無い点字データが作成できるわけではない。これは、上記のような文字列の置換コマンド群が、点訳規則を完全には記述していないためである。従って、

誤りの無い点訳を行いたい場合は、点訳規則に習熟した人がテキストエディタで点字データを修正する。

4. 点訳事例と変換の精度

本システムを使用して自動点訳した例を次に示す。

日本語文書（朝日新聞「天声人語」より）：

そのアセビの自生するさまを見たくて、伊豆の山を歩いた。
標高1400メートルの万三郎岳やその周辺の山はまだ枯れ葉色で、
早春の風が吹いていた。点在する豆桜の花が淡いとき色にかす
んでいる。キジらしい大きな鳥が一瞬、ササをよぎった。

点字データ：

□□ソノ□アセビノ□ジセイシルサマヲ□ミタクテ□イズノ□
ヤマヲ□アルイタ。□□ヒョーコー□1400メートルノ□
マンザブローガクヤ□ソノ□シューヘンノ□ヤマワ□マダ□
カレ□ハイロデ□ソーシュンノ□カゼガ□フイテ□イタ。□□
テンザイスル□マメサクラノ□ハナガ□アワイ□トキ□
イロニ□カスンデ□イル。□□キジラシイ□オオキナ□
トリガ□イッシュン□ササヲ□ヨギッタ。

表1は、天声人語10件を本システムによって点訳した結果について、その点訳誤りを分類したものである。

分かち書きの誤り（表中1～3）は2158箇所中150箇所（約7％）であるが、文意を損なう恐れの大い読み間違い（表中4～7）は1755箇所中40箇所（約2％）と少なかった。また、天声人語1件（平均で日本語782文字）につき約20ヶ所、点字用紙1枚（32文字×22行）につき約8ヶ所の誤りであった。誤りのひとつひとつは「事例を規則に加える」方法で改善可能であるが、規則の数が爆発的に増大することを考えると、なるべく一般化した規則にとどめるのが良い⁷。

天声人語10件（日本語7816文字）を点字データ（カナ13277文字）に変換するのに要した時間は、パソコン（FMR-50）で175秒であった。その印刷には、印字速度が毎秒20文字の点字プリンタを用いたとしても600秒以上かかるので、相対的に満足できる変換速度であると考えて良い。

利用者の評価⁸を通じて、「ある程度誤りを含む点字であることを承知でそのまま利用する」あるいは「点字データの誤りを修正してから点字印刷する」いずれにおいても十分実用的であると判断された。

⁷ 分野を限って専用の辞書や規則を用意することは実用化に際して有効な方法である。医学用語や法律用語など、利用者の手でそのような整備が行われつつある。

⁸ 本論文の自動点訳システムを構成するソフトウェア（名称「80点」）は、1989年に配布を開始し、以後、利用者の協力を得て評価と改良を繰り返している。

表 1. 点訳誤りの分類（天声人語10件分）

	誤りの分類	誤りの例	誤りの頻度 (誤り数/出現数)
1	分かち書きの誤り (ひらがな列中)	ほおがくぼんで →ホオガクボンデ	88/493
2	分かち書きの誤り (漢字列中)	労働省編 →ロードーショーヘン	46/212
3	分かち書きの誤り (その他)	枯れ葉色 →カレ□ハイロ	16/1453
4	漢字の読み間違い (固有名詞)	田英夫さん →タ□ヒデオ□サン	7/86
5	漢字の読み間違い (上記以外)	糧を →リョーヲ	15/1506
6	濁音／半濁音の抜け	豆桜 →マメサクラ	15/17
7	助詞の変換の誤り	などへの →ナドヘノ	3/146

活用事例を2件紹介する。

N氏（千葉市鍼灸マッサージ師会会長，全盲）は，次のような方法で独力で墨字と点字の印刷物を作成している。まず，音声化したワープロソフトで晴眼者用の墨字文書を作成し，視覚障害者用としては，その文書を本システムで自動点訳した後，音声読み上げを利用して校正してから点字プリンタで印刷している。

また，U氏（青森県立盲学校教師，全盲）は，次のように会議の配布資料の点訳に活用している。晴眼者の教師が作成した資料（墨字）のワープロ原稿を本システムで点訳し，そのまま（校正の時間的余裕が無い場合），あるいは校正してから点字プリンタで印刷している。会議で，「晴眼者と「同時に」印刷物を読む」ことが可能となった。また，視覚障害者教師宛の給与明細書は，本システムで点訳したものを校正し，点字印刷物として渡されるようになった。

5. むすび

テキストエディタを利用して文字列の置換だけで自動点訳を行うシステムを開発した。誤りを含んだ点訳データをそのまま印刷する場合でも，点訳者が校正して印刷する場合でも，実用に供することが確認できた。

当初想定していた「本を1冊点訳する」というような典型的な点訳の支援にとどまらず，これまで多くの場合困難であった晴眼者と視覚障害者の情報の共有に有効であることが，活用事例を通じて明らかになった。即時性が重視される場面は，視覚障害者の社会参加が進むにつれて，更に増えると考えられる。

本システムでは，点訳規則を表現するコマンド群や変換辞書を「経験的に作成」したが，日本語文書と点字データから「規則を自動生成」するシステムについても今後検討を行っていききたい。

謝辞 本研究は、飯塚潤一氏と長谷川貞夫氏の協力を得て行った。深く感謝します。漢字かな変換プログラムを開発した川野勇氏と、性能改善に協力していただいた永戸久雄、石田透両氏に感謝します。また、日本語変換辞書の利用を快諾された(株)ボックスに感謝します。

文 献

- [1] 長谷川貞夫：“視覚障害者に必要な点字情報処理”，昭52連大，180（1977）.
- [2] 森健一：“点字翻訳”，計測と制御，vol.20，no.12，pp.58-59（1981）.
- [3] “第25回 IBM ウェルフェアセミナー報告集”，日本IBM(株)，（1986）.
- [4] 加藤俊和：“視覚障害者とコミュニケーション機器”，視覚障害研究，no.26，pp.14-26（1987）.
- [5] 小山智史，飯塚潤一，長谷川貞夫：“テキストエディタを利用した日本語文書の簡易自動点訳”，第15回感覚代行シンポジウム発表論文集，pp.92-95（1989）.
- [6] 小山智史，飯塚潤一，長谷川貞夫：“自動点訳システム「80点」”，視覚障害，no.106，pp.24-29（1990）.
- [7] 国分芳宏：“墨字から音声・点字へ”，視覚障害，no.106，pp.19-24（1990）.
- [8] 日本点字委員会編：“日本点字表記法1990年版”，日本点字委員会（1990）.
- [9] 福崎俊博，山田伸一郎：“MS-DOS を256倍使うための本 Vol. 3”，アスキー出版局（1988）.
- [10] 小山智史，野島秀夫，太田茂，川野勇，城戸勝康，長谷川貞夫：“合成音声による視覚障害者用多目的日本語処理システム”，Human Interface，vol.3，no.2，pp.113-116（1988）.

付 録

漢字かな変換辞書 (ztoh.dic) 変換プログラム (ztoh) で用いる辞書の一部を付表 1 に示す。漢字で始まる文字列とその読みの対からなり、逆順にソートされている。登録語数は約34000。

付表 1 漢字かな変換辞書 (ztoh.dic) の抜粋 (□はスペースを表す)

...		行司	□ギョージ	行き方	□イキカタ	学績	□ガクセキ
...		行使	□コーシ	行き届	□ユキトド	学生	□ガクセイ
行路	□コーロ	行在所	□アンザイショ	行き倒れ	□イキダオレ	学殖	□ガクショク
行列	□ギョーレツ	行在	□アンザイ	行き渡	□ユキワタ	学術	□ガクジュツ
行李	□コーリ	行弘	□ユキヒロ	行き止まり	□イキドマリ	学習者	□ガクシユージャ
行雄	□ユキオ	行広	□ユキヒロ	行き交	□ユキカ	学習	□ガクシユ
行目	□ギョーメ	行幸	□ギョーコー	行き詰ま	□ユキヅマ	学者	□ガクシャ
行木	□ユキキ	行啓	□ギョーケイ	行き掛け	□イキガケ	学識者	□ガクシキシャ
行末	□ギョーマツ	行軍	□コーグン	行き掛かり	□イキガカリ	学識	□ガクシキ
行方郡	□ナメガタグン	行橋	□ユクハシ	行き過ぎ	□ユキスギ	学校長	□ガクコーチャウ
行方	□ユクエ	行宮	□アングー	行きずり	□ユキズリ	学校	□ガクコー
行彦	□ユキヒコ	行脚	□アンギャ	行き	イキ	学区	□ガクク
行年	□ギョーネン	行儀	□ギョーギ	行か	イカ	学級	□ガクキュー
行徳	□キョートク	行間	□ギョーカン	行お	オコナオ	学究	□ガクキュー
行動	□コードー	行楽	□コーラク	行え	オコナエ	学期	□ガクキ
行灯	□アンドン	行革	□ギョーカク	行う	オコナウ	学外	□ガクガイ
行田	□ギョーダ	行員	□コーイン	行い	オコナイ	学界	□ガクカイ
行程	□コーテイ	行為	□コーイ	行	コー	学会	□ガクカイ
行単位	□ギョータンイ	行わ	オコナワ	...		学課	□ガクカ
行川	□ナメカワ	行な	オコナ	...		学科	□ガクカ
行跡	□ギョーセキ	行っ	イッ	学力	□ガクリョク	学位	□ガクイ
行盛	□ユキモリ	行こ	イコ	学用品	□ガクヨーヒン	学ん	マナン
行政	□ギョーセイ	行け	イケ	学用	□ガクヨー	学ば	マナボ
行数	□ギョースー	行く末	□ユクスエ	学友	□ガクユ	学べ	マナベ
行水	□ギョーズイ	行く年	□ユクトシ	学問	□ガクモン	学ぶ	マナブ
行進	□コーシン	行く先	□ユクサキ	学費	□ガクヒ	学び舎	□マナビヤ
行松	□ユキマツ	行く手	□ユクテ	学年	□ガクネン	学び	マナビ
行商	□ギョーショー	行く行く	□ユクユク	学長	□ガクチャウ	学ば	マナバ
行書	□ギョーショ	行く	イク	学卒	□ガクソツ	学	ガク
行事	□ギョージ	行き来	□ユキキ	学窓	□ガクソー	...	
行子	□ユキコ	行き戻り	□ユキモドリ	学説	□ガクセツ	...	

Abstract

Braille documents are of great importance for the blind, as we read printing materials every day. However, those documents are not sufficiently supplied to them, because Japanese-to-Braille translation must be made by hand of the expert. We developed an automatic Japanese-to-Braille translation system using non-interactive text editor, in which automatic segmentation and Kanji-to-kana conversion can be successfully carried out. This system is shown to be quite useful for the blind. Fully automatic translation can be done by them, if some errors are allowed. Before printing, those errors may be removed by the experts.