

自動点訳システム IBUKI-TEN の改良と現状

服部剛士 高松大地 伊佐治和哉 松本忠博 池田尚志
岐阜大学工学部

1 はじめに

IBUKI-TEN は、我々が開発している文節解析システム IBUKI を自動点訳システムに応用したものである。本システムは、入力されたテキスト文書に対して点訳規則に沿って分かち書きを行い、漢字かな混じり文を点字かな表記に変換し、数符など必要な点字符号を挿入するなどの点字用処理を施し、点字文書を得るソフトウェアである。

IBUKI-TEN のインターフェースとしてはダイアログ形式の IbukiTen と点訳後の校正支援機能を備えた IbukiTenEdit がある。IbukiTen は、視覚障害者の方が使用しやすいように、シンプルなダイアログで、容易に多様なファイル出力ができるようになっている。また、IbukiTenEdit では、点訳誤り箇所指摘機能や、原文とひらがな分かち書き、点字を同時に出力するなど、点訳後の校正作業が容易にできるよう設計されている。

また、点字プリンタや点字ペンディスプレイに出力する機能も備えている。

IBUKI-TEN は WWW 上で無償で公開しており、全国で多くのユーザの方に使われている。ユーザから多くの質問や要望・問題箇所などが寄せられており、改良・改版の契機になっている。

本報告では、こういったユーザの方から寄せられた要望・問題への対応を含め、行った機能追加・整備の状況、WWW 上での公開状況について述べる。

2 機能追加・改良状況

2.1 点字コードから日本語テキストへの変換

IbukiTenEdit には、BASE ファイルを読み込み、点字コードを対応した墨字の文字へ変換させる機能がある。この機能の整備を行った。この機能は墨字文字を見ながら点字を編集する際に使われる。BASE ファイルは、書式情報と NABCC コードのデータから作成されている。書式情報には、全ページ数、1 ページの行数、1 行の点字数、ページ数表示の有無があり、エディタのパラメータに設定される。NABCC (北米点字) コードとは 6 点点字の凹凸の状態 (64 種類) を 8 ビットの中に符号化する標準的な点字コードであり、ASCII コー

ドの 0x20 ~ 0x7F を使用する (0x40 ~ 0x5F と 0x60 ~ 0x7F には同じ点字を表す)。清音ひらがなや、英数字などは、6 点 1 マス (1 バイトの NABCC コード) で表され、濁音や「『」などの記号は、複数のマス (複数バイトの NABCC コード) で表現される。

BASE ファイルの読み込みの流れは次のようになっている。システムは BASE ファイルが入力されると、先ほど述べた書式情報を取得し、IbukiTenEdit の書式パラメータを設定する。以下 NABCC コードのデータが並ぶが、ページ数表示有りの場合、各ページの先頭にはページ数を表す NABCC コードがあるので、それを取得し墨字に変換する。その後には、スペースで区切られた NABCC コードのブロックが並ぶ。各ブロックは、文節の単位に対応している。それらのブロックを墨字へ変換していき、全てのブロックの変換が完了した後に、変換した墨字を分かち書き結果表示部に表示する。

NABCC コードは墨字と 1 対多対応であり、例えば「0x41」という NABCC コードは「あ」「1」「a」「A」の 4 通りを表現する。NABCC コードを墨字へ変換する場合、どの文字へ変換するかという変換モードを決めなければならない。変換モードの判断には外文字 (0x38) や外国語引用符 (0x38 ~ 0x30)、大文字 (0x2c)、数字 (0x23) を用いる。これらの点字記号の出現に応じて変換モードが決まり、同じ NABCC コードでも異なった墨字へ変換することが可能となる。

変換モードには、「あ」などのひらがなへ変換するひらがなモード、「1」などの数字へ変換する数字モード、「a」などの英小文字へ変換する英小文字モード、「A」など英大文字へ変換する英大文字モードの 4 種がある。変換モードが決まれば、表 1 の墨字変換テーブルを参照して、墨字への変換を行う。

外文字は、日本語文中に英文字がある場合に前置されているので、先頭バイトに外文字のあるブロックは、英小文字モードで墨字へ変換する。大文字の場合は、その 1 文字に対して、さらに大文字が前置されるので、その場合は、英大文字モードとなる。ただし、ブロック内の英文字が全て大文字の場合は大文字が 2 つ (2 重大文字) 前置されるので、そのブロック全体が英大文字モードとなる。ブロックをまたいで英字が続く

表 1: NABCC コードとモード毎の墨字変換テーブル

NABCC コード	ひらがな モード	数字 モード	英小文字 モード	英大文字 モード
0x61	あ	1	a	A
0x62	い	2	b	B
0x63	う	3	c	C

場合は、その英文字の区間が、外国語引用符（開き）と外国語引用符（閉じ）で囲まれるので、その区間が英小文字モードとなる。数字符が現れると、そのブロック全体、またはつなぎ符が出現するまでの区間が数字モードとなる。つなぎ符とは、あ行、ら行の点字が数字の1~0の点字に対応しているため、数字かひらがなかを区別するために挿入される点字記号である。

2.2 IBUKI-TEN のファイル出力機能

ファイル出力には、以下に示すいくつかのフォーマットがある。これらについて整備した。

NABCC コードフォーマット [* .brl]

NABCC コードによる点字データのみを出力する。

BASE フォーマット [* .bse]

NABCC コードによる点字データに加えて、文字数、行数、全ページ数などのファイルの書式情報を出力する。IbukiTenEdit を含め、BASE ファイルに対応しているエディタで読み込むことができる。

点字かなフォーマット [* .kana]

点訳結果のかな表記を*.kana 形式のテキストファイルとして出力する。IBUKI-TEN 独自の出力であり、IbukiTenEdit やその他テキストエディタで編集することができる。

点字プリンタ用フォーマット [* .brp]

NABCC コードによる点字データに加えて、点字プリンタを動作させるコードや改行コード、改ページコードを、NABCC コード列に付加したフォーマットであり、多くのプリンタが対応している。

IbukiTenEdit 編集用フォーマット [* .ibk]

原文テキストと点訳結果のかな表記及び誤り可能性指摘のデータを合わせた IbukiTenEdit 独自の編集用フォーマットである。

点字かなフォーマットのファイルを IBUKI-TEN に入力すると自動的に NABCC コードが生成され、図 1 に示すようにフォーマットを変換して出力することもできる。

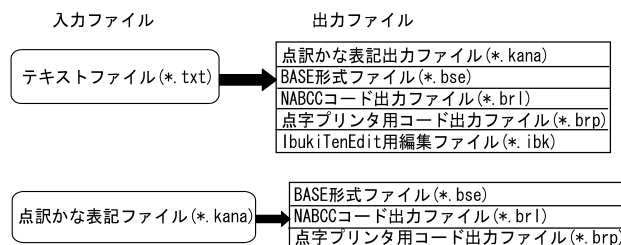


図 1: IbukiTen のファイル出力

2.3 点訳モード

IBUKI-TEN の点訳モードとして、表 2 の 4 種類を整備した。

表 2: 点訳のモード

日本語点訳	1 級英語
	2 級英語
英語点訳	1 級英語
	2 級英語

1 級英語とは、英文字を清音ひらがなと同様にそのまま点字に 1 対 1 で対応させたものである。2 級英語とは、約 200 個定められている短縮表記規則（例えば、"about" は"ab"となる）を用いて短縮表記した英文テキストである。2 級英語は 1 級英語に比べて、文章量の約 30 % を削減することができ、視覚障害者が文章を読むスピードを上げることができる。

日本語点訳モードとは、漢字かな混じり文の日本語テキストを点訳するモードである。文中に現れる英語は、「A さん」などのような文字として書き表す場合には外文字を前置し、語や文として書き表す場合には外国語引用符でその英語の前後を挟む。

英語点訳モードとは、英文テキストを点訳するモードである。従って、英語表記に対して外文字や外国語引用符を使用することはない。

2.4 墨字印刷機能

IbukiTenEdit の、墨字プリンタへ出力する機能を整備した。出力できる内容を以下に示す。

点字

点字は印刷用紙上に、誘導点と呼ばれる下地と共に縦3点、横2点の6つの点を1マスとして印刷する。点字の大きさは紙の大きさによって変わり、印刷用紙1枚には点字プリンタで出力する場合の1ページ分が出力される。

点字かな表記

点字出力と共に、かな表記を対応して出力する。この内容を出力することで、どの点字がどの文字に対応しているかを紙面で確認できる。

行数

点字、点字かな表記の左側に行数を出力することができる。この機能は印刷後に点字資料を使用したり、点字文書の校正を行う際の使用性を上げるためのものである。

点字を印刷する方法には、凸面表記、凹面表記の2種類があり、互いに鏡像の関係にある。凸面表記は、通常の点字文書と同じ表記であり、晴眼者が校正作業のために点字出力を確認する際に使用する。凹面表記は、それを見て点字タイプライタで書き写して（打ち写して）点字文書を作成するためのものであって、行の方向も右から左へと向かう。点字タイプライタで書き写した（打ち写した）文書を裏返せば、通常の凸面点字が現れる。

これらの出力内容や出力方法は設定ダイアログにチェックを入れるだけで変更することができ、使用者の必要な情報だけを出力できる。

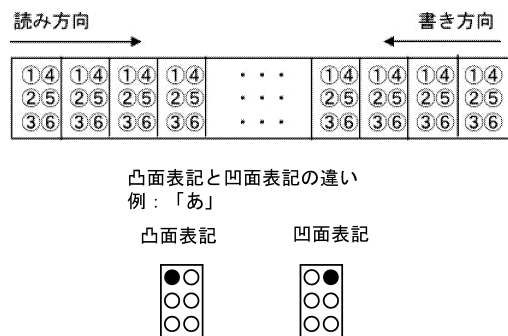


図 2: 点字の読み書き方向と凹面、凸面の例

2.5 6点入力

点字の6点入力の方法を整備した。6点入力とはパソコンのキーボードの6つのキーを点字入力キーとして使う入力方式である。よく使われる方式に図3に示す3つの方式がある。

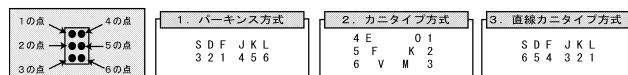


図 3: キーボードと対応する6点

さらにこの3方式以外に、ユーザの好みに応じて任意の6つのキーを選んで、6点に対応させる新しい方式を定義することもできる（図4のファイルKeyAssign6.txtで設定する）。

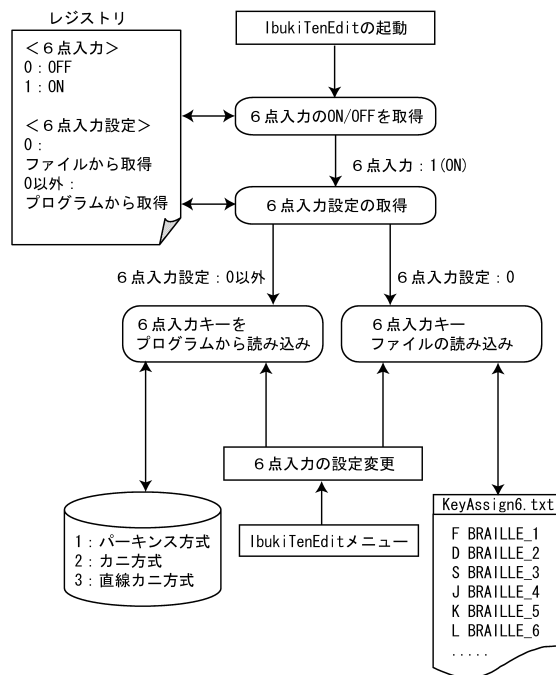


図 4: 6点入力設定のシステム

3 IBUKI-TENの公開状況

IBUKI-TENは2000年9月にWWW上にバージョン0.1を公開し、ユーザからの指摘・要望を取り入れながら多くの改版・機能追加を行ってきた。途中で一時公開を中断したが（2001.12～2002.11）、現在バージョン0.55になっている。

これまでに我々が把握している 2005 年 1 月までのダウンロード数は、延べ数 5,428 件、異なり数はおよそ 3,200 件である (ログを取っていない時期がある 2002.12 ~ 2003.4)。

また、ユーザの方から寄せられた鍼灸の専門用語辞書を公開している (およそ 1,500 件ダウンロードされている (2003.10 ~ 2005.1))。

以下に現在把握できているダウンロード後の利用状況を紹介します。

ボランティア・サークル

把握できるボランティア・サークルの団体数は 23 団体である。その主な使用目的は、パソコン点訳作業・点字資料作り (書籍の点訳など) である。点訳ボランティアやサークルからのメーリングリストで紹介され、興味を持った方からもダウンロードされている。

公共施設

公共施設では 18 施設からダウンロードされており、講習会や研修に使用したり、広報などの出版物や図書館の書籍の点訳にも使われている。

盲学校

盲学校からは現在 38 校からダウンロードされている。授業や授業で使用する資料の作成などに使われている。また、盲学校の先生方で家庭の PC で点訳作業などを行うためにダウンロードしている方も多くいる。

個人

家族に視覚障害者の方がいたり、また、視覚障害者自身がダウンロードする場合もあり、点訳や点字を勉強するために使われている。点訳者養成講座の勉強や「ASKK てんてん」、「ブレイルメモ」などで使用するためにダウンロードする方もいる。

その他

大学や企業で点字関連の研究を行っている方からのダウンロードもあった。専門学校などからのダウンロードもある。また、中学・高校では総合学習の授業教材として用いられているケースもある。

今後も IBUKI-TEN の問題点への対応や点訳精度の向上など改良を進めていく予定である。

参考文献

- [1] 自動点字翻訳編集システム IBUKI-TEN ,
<http://www.ikd.info.gifu-u.ac.jp/IBUKI-TEN/>
- [2] 福井哲也：初歩から学ぶ英語点訳：日本点字図書館 (1991)
- [3] 兵藤，横平，早川，村上，池田，誤り箇所指摘機能をもたせた点字翻訳編集システム IBUKI-TEN，電子情報通信学会論文誌 VOL.J84-D-I No.7，pp.1102-1111，2001。

4 終わりに

自然言語処理に基づく自動点訳システム IBUKI-TEN の概要と改良点，公開状況について述べた。