

自動点訳編集システム ibukiTenCEdit2

江本倫基, 高田和典, 池田尚志

岐阜大学工学部

1 はじめに

我々の研究室では視覚障害者の読書支援・点訳ボランティアの活動支援となる自動点訳システム ibukiTenC を開発し、公開してきた [1][2]。

本論文では、ibukiTenC の現状について述べるとともに、ibukiTenC の構成機能のうち編集・校正機能を受け持つ ibukiTenCEdit の問題点の指摘と、より効率的な編集・校正機能を目指して新規作成した ibukiTenCEdit2 について述べる。

2 ibukiTenC について

ibukiTenC は墨字から点字への自動変換を行うことを目的としたシステムである。我々の研究室で開発中の日本語解析システム ibukiC をベースとしており、入力された墨字文書に対して、点字規則に沿って分かち書きを行い、漢字かな混じり表記を点字規則に沿ってかな表記に変換する。また数符など必要な点字符号を挿入することなどの点字用処理を施して、点字文書を得ることができる。さらに得られた点字文書を編集する機能も持っており校正作業の支援が出来るようになっている。また点字プリンタや点字ピンディスプレイに出力する機能も備えている。

ibukiTenC は web で一般公開しており、2000 年の公開開始以来延べ数で約 15000 件のダウンロードがある。

また、点訳エンジンは独立した DLL として作成されており、本システム以外からもその機能を利用することができる。最近では点字カラオケのための点訳エンジンに使うという利用例 [3] もある。

2.1 システム概要

システム構成を図 1 に示す。

点訳の流れとしては、日本語文が入力されると、日本語解析システム ibukiC によって文節解析と複合語解析を行う。その後 ibukiTenC の点訳処理 (分かち書き処理、点字表記変換) を行う。編集・校正機能を持つ ibukiTenCEdit では、さらに後編集処理 (点訳誤り検出、かな表記修正、分かち書き修正) を行い点字を出力する。

2.2 点訳エンジン

点訳エンジンは点訳用辞書から辞書情報や点訳規則などをメモリ上にロードし、それを用いて点訳を行う機能を担当している。点訳時には点訳誤りの可能性指摘などの処理も行う。

2.3 点訳用辞書

辞書には解析用の情報の他に、点訳用の点訳表記の項目がある。点字で分かち書きを必要とする箇所には “/” を記述してある。自立語は現在のところ約 24 万語登録されている。機能語は ibukiC の特徴である長単位での登録となっており、約 5 千語登録されている。

辞書の例を表 1 に示す。複数の読みを持つ単語には読み優先フラグを設定し、フラグが true の読みを優先して出力する。

表 1: 点訳用辞書

見出し語	品詞説明	点訳表記	読み優先フラグ
会社	名/一般	かいしゃ	
今日	名/時	きょう	○
今日	名/時	こんにち	
作物	名/一般	さくもつ	○
作物	名/一般	さくぶつ	
作物	名/一般	つくりもの	
民主主義	名/一般	みんしゅ/しゅぎ	
日	尾/名/時	にち	
にちがいな	機/用 ナイ	に/ちがい/な	
に違いな	機/用 ナイ	に/ちがい/な	

2.4 ユーザーインターフェイス

ibukiTenC のインターフェースとしてはダイアログ形式の ibukiTenC (図 2) と点訳後の校正支援機能を備えた ibukiTenCEdit (図 3) がある。

前者は視覚障害者がスクリーンリーダー (画面上の文字列を音声読み上げるソフトウェア) を用いて利用することを想定し、ボタンの数や点訳設定などを考慮して作成されたものである。ただし点訳はテキストファイルから点字ファイルへの変換しかサポートしていないなど、機能は単純化してある。

後者は点訳ボランティアの方々の利用を想定して開発したシステムであり、自動点訳の機能に加えて校正支援機能が搭載されている。自動点訳では分かち書きや正しい漢字の読みを完全に正確に選択することは難しく、点訳結果はどうしても誤りを含んでしまう。そのためユーザーは自動点訳した結果を見直し、誤った箇所を見つけて校正する必要がある。ibukiTenCEdit では点訳結果をエディタ部に表示しユーザーが自由に編集できるようになっているほか、システムが誤りの可能性があると判断した箇所を色分けして強調表示したり、漢字の読み誤りはその他の読み候補リストを表示して校正作業を効率よく行う工夫がされている。

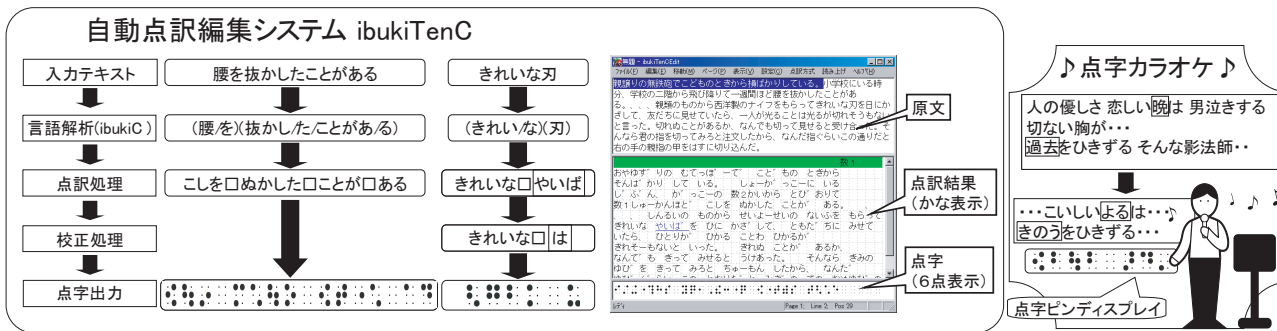


図 1: システム構成

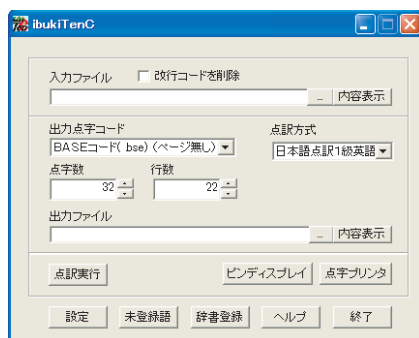


図 2: ibukiTenC

3 ibukiTenCEdit の問題点

ibukiTenCEdit は先に述べたように、ibukiTenC の点訳エンジンを使った自動点訳機能と誤り指摘機能が搭載されている。これらの機能を使うことで一から手作業で点訳作業を行うよりはるかに効率的に点訳された文章を得ることができる。

しかし ibukiTenCEdit は自動点訳の点訳誤りを校正をするために用意されたシステムであり、テキストおよび点字の編集システムとしての機能には不十分な点があった。

1. 点訳は全文章を一括して行うしかない (部分的な点訳ができない)
2. 選択文字列のコピー・貼り付けなどの編集機能が不足している
3. 誤り指摘機能を利用した校正支援機能の実行速度に難がある

1. の問題点については、点訳したい文章の入力と校正を平行して行うような編集方法に対応できず、原文を入力→全文点訳→点訳結果を校正という手順を強制するこ

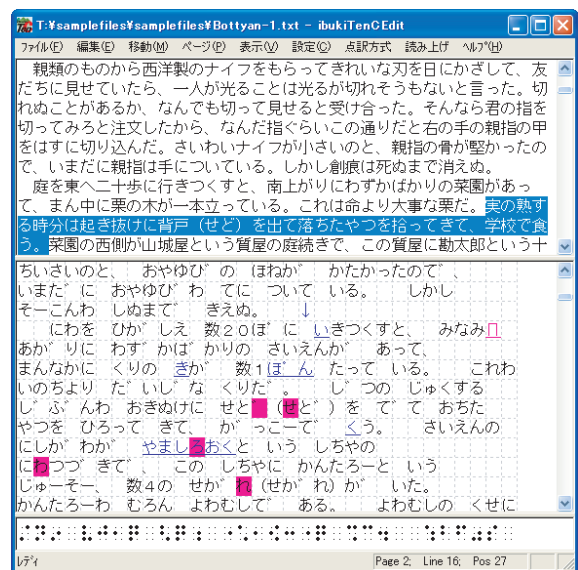


図 3: ibukiTenCEdit

とになってしまう。校正支援としては十分かもしれないが、編集システムとして使いたい場合は不十分である。

2. に関しても通常のテキストエディタと比べて機能的に劣っており、文章の編集に支障が出る場合がある。ユーザーからも要望としてあがってきている。

3. の問題点については誤り指摘機能により機械的な変換による誤りの発見には貢献しているが、その後の修正の支援で課題を残しており、やはり改善の余地がある。

このような問題があるためそれぞれの改善が望まれているが、システムの規模が大きいという、長年にわたる部分的な改修が引き継がれてきたため、現在はさらなる改良が困難な状況になっている。

4 ibukiTenCEdit2 について

以上の問題点をふまえ、本研究では点字編集システムを新規に開発した。

4.1 概要

- 名称:ibukiTenCEdit2
- 開発言語・環境:C# + Visual Studio 2008
- 動作環境:WindowsXP 以上

点訳エンジンは ibukiTenCEdit と同様に ibukiTenC のエンジンを利用し、ユーザーインターフェース部分を新規に作り直した。その外観を図 4 に示す。

ibukiTenCEdit2 は大きく分けると、自動点訳を行いたい文章を入力する原文エディタ、自動点訳された結果の表示・編集を行う訳文エディタ、キャレットがある行の内容を点字フォントで表示する点字ビューの 3 つのパーツから成り立っている (それぞれ図 4 の上から順に対応する)。ユーザーはファイル読み込みもしくは直接キーボードなどから原文エディタに点訳を行いたい文章を入力し、訳文エディタに表示された点訳結果を編集して、最終的に点訳された文章を得ることができる。

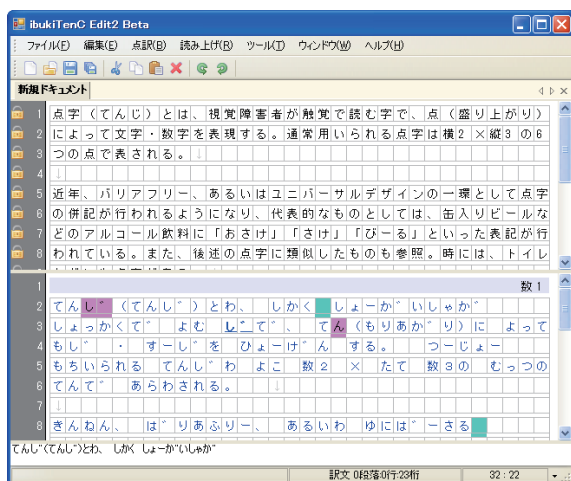


図 4: 自動点字翻訳編集システム ibukiTenCEdit2

4.2 ibukiTenCEdit2 の作成にあたって

本研究の最終的な目的は、墨字から点字への変換・編集を従来のシステムである ibukiTenCEdit よりも効率的に行えるシステムを構築することである。

効率的な作業を行うためには以下の 2 点を重点的に意識することが必要だと考えた。

- 文章を適切な内容・タイミングで機械的に変換すること
- 機械的に変換した誤りを素早く発見・修正できるようにすること

また、このシステムは無償で一般公開し、実際の点訳の現場で利用されることを期待している。

4.3 翻訳に関する改良

ibukiTenCEdit2 では自動点訳を行う範囲と、自動点訳を行うタイミングを ibukiTenCEdit から大きく変更した。

第 4 章で指摘した点訳の適切なタイミングに対する改善策として、段落単位でのリアルタイム点訳機能を実装した。この機能はユーザーが文字を入力した場合に、原文・訳文のそれぞれ対応する段落の内容を自動的に再点訳して更新する機能である。図 5,6 のように、入力した文字が確定された段階で自動的に点訳が実行され、適切な位置に挿入される。



図 5: 文字を入力する



図 6: 確定と同時に点訳も実行される

この機能によるユーザーのメリットは、変換という処理を意識しなくても文章を入力するだけで点訳結果を得ることができるという点、入力と同時に点訳結果を知ることができるため正しく点訳できたかのチェックを行いやすい点、部分的な点訳が可能になったため文章の入力と校正を平行して行うことができるようになった点などが挙げられる。

ただし常にこの機能が働くと不必要な点訳・更新が発生してしまう可能性があるため、ユーザーが任意にこの機能を使用するかしないかを段落単位で設定できるようにしてある。

4.4 校正に関する改良

ibukiTenCEdit と同様、点訳エンジンの機能を利用した誤り指摘と、誤り指摘箇所の校正支援機能を実装した。具体的には分かち書きの誤り (過剰に区切ってしまった可能性、区切り忘れの可能性)、漢字の読み誤りの可能性の指摘を行い、画面上で特別な色でマーキングされて表示される。読み誤りの場合はシステムが用意する読み候補一覧から正しい読みを選択することで、素早く誤り箇所を修正することができる (図 7, 表 2)。

ibukiTenCEdit ではこの読み誤りの修正を行った時に、長い文章 (小説 1 冊相当など) が対象の場合に数十秒~数分の処理時間がかかっていたのに対し、ibukiTenCEdit2 では同じ文章を使用してもほぼ待ち時間なしで処理することが可能になった。ibukiTenCEdit では読み誤りの修

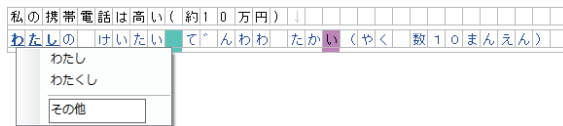


図 7: 誤り指摘と校正支援機能の表示例

正を一カ所行うたびに画面の再レイアウトを行っていたことが速度低下の主な原因であり、そのような余分な処理を行わないことで十分な高速化となった。

また誤り指摘は視覚的に点訳誤りの可能性を表示するが、音声読み上げではこれらの情報は一切出力されず利用できないという問題もあった。そこで音声出力時に謝り指摘のある箇所ではそれを示すように読み上げを行うようにした。たとえば図7では「わたし 2個の候補の」「けいたい 区切りすぎの可能性」「でんわは たかい 区切り忘れの可能性」といったように読み上げられる。

表 2: 誤り指摘の種類

誤りの種類	表示	解説
読み誤り	しりつ の がっこー	漢字の読み方が複数候補あり、一意に決められない場合
区切り不足	まつ なみき	区切る必要があるかもしれない場所
区切り過ぎ	とし こっか	区切る必要がないかもしれない場所

4.5 編集機能に関する改良

不足している機能として挙げていたコピー・貼り付けの機能を実装し、テキスト編集機能はメモ帳 (notepad.exe) 程度の機能を備えた。

4.6 その他現状

4.6.1 タブエディタ化

1つのウィンドウで複数の文章を開いて編集できるようにした。

4.6.2 スクリーンリーダーの標準添付

エディタの内容を読み上げるためには、これまではユーザーが自分で読み上げソフトを用意しなければならなかったが、テキスト音声合成ミドルウェア AquesTalk[5]を同梱することで特に準備をしなくても読み上げ機能を利用できるようにした。

また、これまでは各スクリーンリーダーごとに個別に対応しており、新しいスクリーンリーダーを利用する

ためには専用の処理を記述しなければならなかった。そこで音声認識・合成のための API である SAPI5 を利用し、その仕様に対応したスクリーンリーダーならば追加処理なしで読み上げを行えるようにした。

4.6.3 エディットコントロールの自作

点訳エンジンの校正支援機能をフルに利用できるようにするために、標準のテキストボックスは使用せずにテキストエディタ機能を持つコントロールを自作した。

このコントロールはクラスライブラリとして作成したので、ほかのプログラムからこのエディットコントロールを利用することもできる。

5 おわりに

本論文では ibukiTenCEdit の問題点の指摘と、それをふまえて ibukiTenCEdit2 の新規作成について述べた。ibukiTenCEdit にあった問題点を概ね解決できた。

ibukiTenCEdit2 はユーザーからのフィードバックを目的にベータ版として一般公開している [6](2008/10/16 公開開始)。

参考文献

- [1] ibukiTenC 公式サイト
<http://www.ikd.info.gifu-u.ac.jp/ibukiTenC/>
- [2] 横平貫志, 自動点訳システム IBUKI-TEN に関する研究, 岐阜大学大学院 工学研究科 池田研究室 修士論文 (2001)
- [3] 日本テレソフト
<http://www.telesoft.co.jp/>
- [4] 高田和典, 江本倫基, 脇田貴之, 山田佳裕, 池田尚志 (岐阜大学) 自動点字翻訳編集システム ibukiTenC-点字毎日との比較による精度評価- NLP2007 (2007)
- [5] AquesTalk - テキスト音声合成ミドルウェア
<http://www.a-quest.com/aquestalk/>
- [6] ibukiTenCEdit2 beta
<http://www.ikd.info.gifu-u.ac.jp/ibukiTenC/Edit2Beta/>