

1F-03 事例を用いたオンライン日本語点字翻訳システム

浜田佳延 高木喜次 小野智司 水野一徳 西原清一
筑波大学 電子・情報工学系

1 はじめに

日本語を点字翻訳(点訳)するためには、漢字をかなに変換するとともに、意味の理解を助けるために単語間に区切りを入れる必要がある(分かち書き)。この作業には、複雑で曖昧な多くの規則があり、全ての規則を計算機上で表現することは困難である。

近年、日本語の点訳作業は、パソコンを用いて行われることが多い。しかし、市販されている点訳ソフトには、精度改善や修正の方法に不十分な点も多い。

本研究では、Web 上において対話的に点字翻訳を行うシステムを開発する。本システムは、ユーザの修正の手間を軽減するインタフェースを持ち、ユーザとの対話から学習を行うことができる。

2 従来の点訳システム

一般的な既存のシステムでは、入力された文章(原文)から点字までの変換を一括で行う。しかし、全く誤りのない点訳結果を得ることはできないため、ユーザは点字エディタを用いて見直し・修正を行う必要がある。点字エディタによる修正では以下のような問題点がある。

- (1) 一度行った修正を、同一文書内の他の箇所や、以後の文書に反映することができない。
- (2) 原文と点訳結果を比較しながら見直しを行うことが難しい。

問題点(1)より、同じ修正を繰り返さなくてはならなかったり、ユーザ辞書への単語登録が必要となる。また、修正の見逃し等により点訳結果の揺れが起こり得る。

上述の問題の他にも、市販されているシステムには、使用可能な OS が限定されている、高価である等の問題も存在する。

一方、既存の Web 上の点訳システムは、開発段階のものが多く、実用的な精度の点訳を実現することは困難であると思われる。

A Case-Based System for Japanese-to-Braille Translation on the Net
Yoshinobu Hamada, Yoshitsugu Takagi, Satoshi Ono, Kazunori Mizuno, Seiichi Nishihara
Institute of Information Sciences and Electronics, University of Tsukuba

3 提案するシステム

本研究では、ユーザの修正の手間を軽減することを目的とした Web 上の点訳システムを開発する。以下では、提案するシステムの詳細について述べる。

3.1 方針および特徴

本システムの方針と特徴を以下に述べる。

- (1) 点訳規則集から抽出したルールに加えて、ユーザの修正から獲得した事例を用いることによって、点訳を行う。
→ユーザの修正を学習し、以後行われる点訳に利用できる。
- (2) 見直し・修正を行う画面において、原文と点訳結果を対応させて表示する。
→原文と点訳結果を比較しながら、見直し・修正を行うことができる。
- (3) 修正が行われた際、類似する箇所を一覧で提示し、ユーザに修正・確認を促す。
→ユーザの修正を同一文書内の他の箇所に反映でき、点訳結果の揺れを防止することができる。

3.2 処理の流れ

本システムの処理の流れを図 1 に示す。はじめに、原文に対して自動点訳を行う。次に、ユーザが自動点訳の結果に対して見直し・修正を行う。最後に、点字表記のテキストを出力する。出力されたテキストは、点字エディタを用いることで、点字ペンディスプレイ、点字プリンタ等に出力することが可能である。

自動点訳処理では、原文に対して形態素解析を行い、属性情報を付加する。そして、その属性情報に基づき、ルールベース推論および事例ベース推論を行う[1]。

見直し・修正処理は、基本修正機能と複数一括修正機能の 2 つの機能を用いて行われる。基本修正機能では、ユーザが修正したい箇所をマウスで指定することにより、区切り方、読み方を修正できる。複数一括修正機能については次節で述べる。ユーザは、両機能を切り替えながら修正を行う。また、修正した結果は事例として学習し、以後の点訳で利用する。

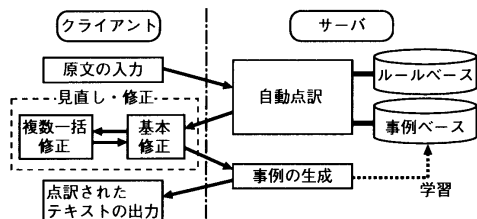


図 1: 処理の流れ

3.3 複数一括修正機能

本システムは、修正を行った際、類似した箇所を一覧表示させて、修正・確認を行う複数一括修正機能を持つ。これにより、修正の見逃しや点訳結果の揺れを回避することが可能となる。

類似箇所提示のアルゴリズムを以下に述べる。

(1) 区切り方について修正された場合

文章全体について、修正された箇所との類似度計算を行い、閾値を越える箇所を提示する。

(2) 読み方について修正された場合

修正された箇所と同じ文字列を含む箇所のうち、修正前、または修正後と同じ読み方を持つ箇所を提示する。

4 実行例

システムの実行画面を図 2、3 に示す。図 2 は自動点訳が終了し、ユーザが読み方を修正している状態である。この修正後、複数一括修正ボタンを押すことで、図 3 のように基本修正画面の下半分に複数一括修正画面が現れる。該当箇所が一覧表示され、ユーザは、各箇所を選択的に修正することができる。

複数一括修正における提示内容を調べた結果、文全体の提示を要する箇所の約 8 割が提示され、提示された箇所のうち約 9 割が実際に提示を要する箇所であった。

5 おわりに

ユーザの労力を軽減する Web 上の点訳システムの開発を行った。本システムは、点訳規則集から抽出したルールベースに加えて、ユーザの修正を事例として利用するものである。また、本システムの複数一括修正によって、ユーザの修正を同一文書内の別の箇所に反映し、点訳結果の表記の揺れを防ぐことが可能となる。

今後の課題としては、複数一括修正における提示の再現率と適合率の両立、複数クライアントに対応する場合を考慮した事例ベースのメンテナンス機能の開発、システムの評価を行うことが挙げられる。

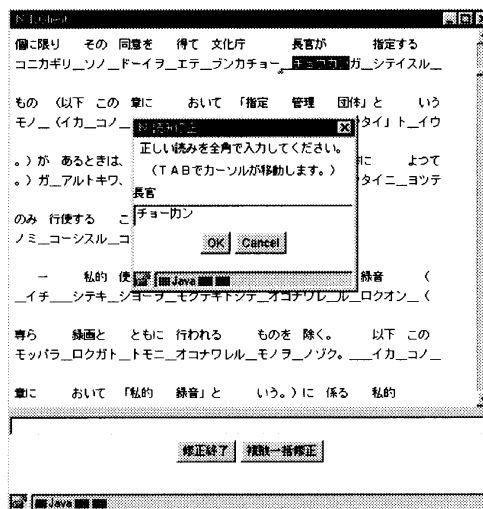


図 2: 基本修正の実行画面

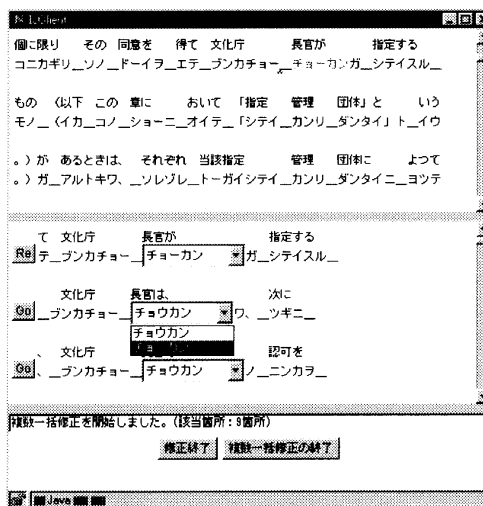


図 3: 複数一括修正の実行画面

謝辞

音声合成形態素解析プログラムおよび韻律付与プログラムを御提供頂いた日本電信電話株式会社 情報通信研究所殿、本研究に御協力頂いた本学 電子・情報工学系 福井幸男教授に感謝の意を表します。

参考文献

- [1] 中居好美, 高木喜次, 小野智司, 西原清一: 点字用分かち書きへの表層解析と形態素解析による事例ベース推論の導入, 情報処理学会第 58 回全国大会, 4M-1(1999).
- [2] 鈴木恵美子, 小野智司, 狩野均: 点字翻訳ボランティアのための対話型分かち書き支援システム, 自然言語処理学会誌, Vol.5, No.4, pp.95-110(1998).
- [3] Golding, A.R. and Rosenbloom, P.S.: Improving Rule-Based Systems through Case-Based Reasoning, *Artificial Intelligence*, Vol.87, pp.215-254(1996).