

ウェアラブル点字リーダにおけるスペルコレクタの改良

柳澤 陽葉[†] 伊藤 祥一[‡]

長野工業高等専門学校 電子情報工学科^{†‡}

1 はじめに

視覚障害者が情報を入手する手段のひとつとして点字があげられる。現在、点字は社会の様々な場所に普及し、点字利用者は十分に情報を入手することができるようになった。しかし、点字習得には点字独自の文法の理解と指先の感覚で凸点を読み取ることが必要であり、これは中途視覚障害者にとって特に困難なことである。厚生労働省が行った調査¹⁾によると、視覚障害者のうち「点字ができる」と回答したのは12.7%となっており、点字が利用できる視覚障害者が少ないことがわかる。

そこで我々は、指に装着した感圧センサで点字をなぞり、その翻訳結果を音声として出力できる「ウェアラブル点字リーダ」の開発を行っている。このデバイスを使用することで点字を利用できない視覚障害者が簡単に点字を利用することができるようになることを目指している。現在のウェアラブル点字リーダでは、読み取ったデータを処理した際に点字を構成する点が欠落してしまったり、存在しない点が認識されてしまったりすることがある。図1では感圧センサで「トショカン（図書館）」という点字を読み取ったつもりが「ノショカン」と翻訳されてしまっている。

先行研究²⁾ではこのような問題を解決するため、翻訳処理部において誤入力を検出し修正するスペルコレクタシステムの開発が行われてきた。本研究ではスペルコレクタの開発・改良を行い、実際に点字を感圧センサが読み取った際の誤入力がどの程度修正可能なのか検証し、システムの評価を行う。

2 点字に特化したスペルコレクタ

2.1 システムの流れ

本システムの処理は、以下に示す7ステップで構成されている。

1. 1つの点字を2進数6ビットで表した数値が本システムに任意の個数だけ入力される。
2. 入力された向きの点字（以降、正点字と示す）

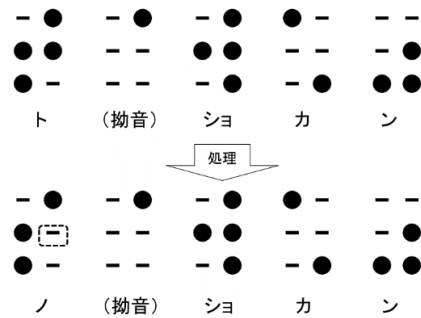


図1 圧力データの処理による誤入力

と、正点字を上下左右逆さまにした向きの点字（以降、逆点字と示す）を生成する。

3. 助詞・助動詞が含まれていた場合は取り除く。
4. 正点字と逆点字のそれぞれについて、一致する語が辞書にあるかどうか調べる。
5. 一致する語がなかった場合は辞書の中から最も近いと思われる語を決定する。
6. 取り除いておいた助詞・助動詞を付け足す。
7. 正点字と逆点字のそれぞれについて、決定された語を比較し入力により近いものを修正結果として出力する。

本研究では固有名詞へ対応するため、オリジナルの辞書を作成した。辞書はIPAdic³⁾とIPAdicの後継であるNAIST-jdic⁴⁾、さらにmecab-ipadic-NEologd⁵⁾の3つを統合し、重複する語を削除して作成した。

2.3 Mecabを用いた助詞判別

本システムで用いる辞書には自立語のみ収録されているため、自立語に助詞や助動詞が付属した場合はそれらを取り除いてから語の訂正を行う必要がある。先行研究では分かち書きされた文節の末尾の文字が「は」や「です」のような付属語であった場合に付属語を分割する処理を行っていた。しかし、この方法では例えば「おから」という語の本来助詞ではない部分が助詞と判別されてしまい、「オ・から」のように分割されてしまう。これを解決するため本研究ではMecab⁶⁾を用いて助詞判別を行う機能を開発した。これを用いることで品詞IDという品詞の分類を表す数値を取得することができる。品詞IDが13以上25以下、または32であればその語を助詞・助動詞と判別し、分割を行う。手順は以下のようである。

1. 分かち書きをした点字データを日本語に翻

Improvement of a Spell Correction Algorithm for Wearable Braille Reader

[†] Akiha Yanagisawa: National Institute of Technology, Nagano College

[‡] Shoichi Ito: National Institute of Technology, Nagano College

訳する。

2. 日本語に翻訳したデータを Mecab によって解析し、品詞 ID(id)と語(surface)を取得する。
3. 品詞 ID が 13 以上 25 以下、または 32 であったら助詞・助動詞があったと判断する。このときに surface に格納されている文字が助詞・助動詞である。
4. 翻訳前の点字データの末尾を助詞の文字数分だけ削る。
5. 助詞を取り除いた点字データが辞書に存在するか調べ、訂正を行う。
6. 訂正後の語に手順 2 で取得した助詞または助動詞(surface)を付け足す。

2.4 助詞判別方法の比較

改良前は「どんより」という語が「ドン・より」のように「より」という助詞で分割されることがあったが、Mecab を用いることで助詞を含まない言葉の末尾を助詞として分割してしまうことがなくなった。しかし、Mecab による助詞の判別では誤りを含む語は正しく品詞を判別できないことがある。例えば、「えいざいがくは」という語が入力された場合、期待される訂正結果は「ケイザイガク・は」であるが「えいざいがく」という言葉は存在しないため、Mecab は助詞の「は」を判別することができない。対して、先行研究²⁾で開発された「語の末尾の文字で判別する方法」では本来は助詞ではない文字も助詞と判別してしまう点が問題であった。しかし、「えいざいがくは」のように語に誤りが含まれていたとしても、末尾の文字で助詞を判別することが可能である。どちらの方法も助詞・助動詞自体が誤っていると正しく訂正することが困難になる。

2.5 翻訳結果の選択

入力された語によって Mecab を用いた方法と末尾の文字で判別する方法とで、結果が異なることがあり、片方の方法では正しく訂正されるがもう一方では正しく訂正できない場合もある。そのため両方の方法を用いて、訂正の精度を高めた。手順は以下のようである。

1. 点字データを表す数値が入力される。
2. 逆点字を作成する。
3. 正点字と逆点字をそれぞれ末尾の文字で助詞判別を行い、辞書を用いて訂正を行う。
4. 正点字と逆点字の訂正結果を比較し、より入力されたデータに近いほうを決める。選んだ訂正結果を A とする。
5. 正点字（逆点字）のほうが入力に近ければ正点字（逆点字）のみを Mecab を用いて助詞判別を行い、辞書を用いて訂正を行う。この訂正結果を B とする。
6. A と B を比較し、入力されたデータにより近いほうを最終的な訂正結果として表示する。

3 実行結果

助詞を含む入力を行った場合の結果を表 1 に示す。出力で平仮名になっている部分が助詞・助動詞として判断された文字である。

表 1 実行結果

入力	出力	正誤
ばすていえ	バステイエ	○
ねてばかり	エテばかり	×
どんより	ドンヨリ	○
がっこーでわ	ジャッコウドわ	×

単純な「自立語+助詞・助動詞」という組み合わせの入力については正しく助詞の判別を行うことができるようになった。しかし、「ねてばかり」の「て」と「ばかり」のように複数の助詞の組み合わせや「では」の「は」のみが助詞と判断されるなど、正しく判別できない場合がある。また、訂正前の点字データにおいて助詞・助動詞にあたる部分に誤りが含まれていた場合には訂正を行うことができない。

4 まとめ

本システムでは判別できる助詞・助動詞が増えたため、以前よりも正しい入力を誤った結果に改変してしまう可能性が低くなった。複数の助詞の組み合わせに対応したり、「では」などの最後尾の「は」だけでなくその手前まで考慮すべき助詞に対応したりすることができれば、さらなる精度向上につながると考えられる。

参考文献

- [1] 厚生労働省社会・援護局障害保健福祉部企画課：平成 18 年身体障害児・者実態調査結果、<http://www.mhlw.go.jp/toukei/saikin/hw/shintai/06/index.html>
- [2] 下平啓太, 伊藤祥一：圧力センサから読み込まれた点字のスペル訂正アルゴリズム, 情報処理学会第 82 回全国大会講演論文集, pp.4-395-4-396, 2020.
- [3] 浅原正幸, 松本裕治: ipadic version 2.7.0 ユーザーズマニュアル, <https://ja.osdn.net/projects/ipadic/>, 2003.
- [4] 浅原正幸, 松本裕治: NAIST Japanese Dictionary version 0.4.0 ユーザーズマニュアル, <https://ja.osdn.net/projects/naist-jdic/>, 2008.
- [5] 佐藤敏紀, 橋本泰一, 奥村学: 単語分かち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討, 言語処理学会第 23 回年次大会発表論文集, pp.875-878, 2017.
- [6] Mecab: Mecab 公式ウェブサイト <https://taku910.github.io/mecab/>