K-042

外国人の初級日本語学習時における単語の読み仮名誤りの DP マッチングによる訂正

Reading Kana Error Correction Using DP Matching in Foreigner's Basic Japanese Language Learning

細田 裕樹† 谷之口 優人† 杉野 勝也† 絹川 博之†
Yuki Hosoda Yuto Taninokuchi Katsuya Sugino Hiroshi Kinukawa

１． はじめに

外国人を対象とした日本語学習においてコンピュータが多く利用されるようになってきた．しかし,外国人日本語学習者が作成した文章を添削するシステムはほとんど見られず,日本語教師等の人手によって添削されているのが現状である．そのため,学習者が独学で文章作成を学習することは困難である．

そこで我々は外国人学習者が独学で文章作成を学習できることを目標として日本語学習支援システムを開発している．現段階では,対象を初級日本語にしぼり,学習者の作成した文の読み仮名誤りをＤＰマッチングを使い訂正する方法を研究している.本稿では,その訂正方式について報告する．

２． 外国人向け初級日本語学習支援システム

2.1 外国人向け初級日本語とその誤り

本研究では外国人のための初級日本語を研究対象にしている.初級日本語とは日本語能力試験の N3レベルに相当しており,漢字は 300字程度,語彙は 1,500語程度が必要とされている． 初級日本語において外国人学習者は,平仮名の誤りが多く,平仮名による読み書きは正しい発音ができているかの指標になる.また,電子辞書やパソコンのキーボード入力が正しく行えることに繋がるため,重要とされている本研究では, 実際に日本語を学習している日本語学校の外国人が行った漢字テストの漢字の読み仮名誤りを訂正することを目的としている．

漢字の読み仮名誤りを以下の表音文字誤りと音訓読み誤りの *2*つに分ける．

・表音文字誤り
　- 濁音・半濁音,拗音,長音,促音に関する誤り
　- 文字が余分にあるもの,文字が抜けているもの
・音訓読み誤り
　- 音読みと訓読みとを誤選択しているもの

2.2 外国人向け初級日本語学習支援システム

本システムは学習者が文章を平仮名で入力すると,システムが誤り検出,訂正を行い,学習者に誤りの指摘と正解を提示することを目指している． 単語の正解候補が複数入力される場合,学習者に正解を選択させる形をとる．

誤り検出については報告されており,本研究では上記の読み仮名誤り訂正について述べる．

†東京電機大学大学院　未来科学研究科
Graduate School of Science and Technology for Future Life, Tokyo Denki University

３． 読み仮名誤りの訂正方式

3.1 表音文字誤りの訂正

(1)　表音文字誤りを確認するには,対象とする単語と正しい単語候補群を比較し,その相違度を求めることで正しい単語の訂正候補を求めることができる．

本研究では，ＤＰマッチングという動的計算手法を用いる．ＤＰマッチングは 2つの文字列を,先頭から並び順を崩さずに 1文字ずつ比べ,一致しているときは”0”，1字ずれるごとに”1”,1文字誤るごとにコストの値を誤りの種類によって変えることで,正しい訂正候補の相違度が低くなるようにする.その後,2つの文字列のコストを加算していき,コストの一番低くなる経路を最終的なコストとして,それを 2つの文字列の相違度の結果とする．

本研究では誤り単語と,正しい漢字とその読み方が記載されている単語辞書の単語を ＤＰマッチングですべて比較し,相違度が小さいものから 5番目まで出力する．

以下の**図１**にコストの算出方法を示し,**図２**に訂正方式の処理の流れを,コスト判定に必要な各パラメータの値を**表１**に示す．

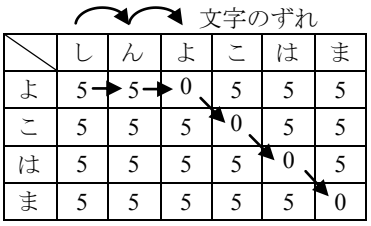

図１． ＤＰマッチングの最短経路の例

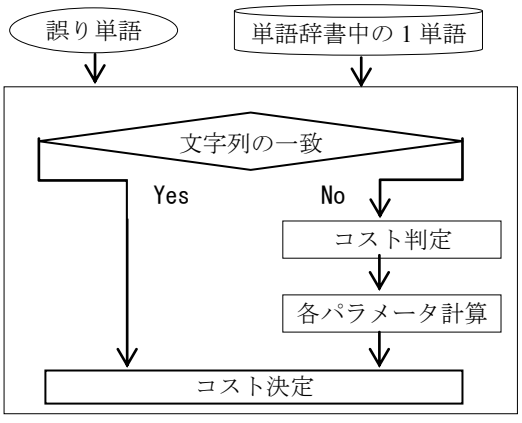

図２． ＤＰマッチングの訂正方式

(1) 誤りに対するコストは,漢字テストのデータから誤った人数が多いものを”4”とし,人数は多くないが特徴的

な誤りのものを”3”として設定した.
各パラメータの値を以下の**表１** に示す.

表１． 誤りによるコスト

誤りの種類	点数
濁音に関する誤り 半濁音に関する誤り 促音に関する誤り	3点
余分な表音文字がある 表音文字が抜けている 長音に関する誤り	4点
上記以外の誤り	5点

3.2 音訓読み誤りの訂正

単語の音訓読み誤りを確認するには, 文字の音訓の組み合わさり方を把握する必要がある. そこで, 3.1**節**で使用した単語辞書にほかの読み方の項目を作る.

辞書に新たに追加する項目は全ての文字の組み合わせとし, 日本語で主に使われている音音読み, 訓訓読み, 音訓読み, 訓音読みの 4つの項目を追加した.

単語の誤りの訂正時にほかの組み合わせの読み方がある場合, その項目と比較し, 一致した場合正しい読み方に導く. 音訓読み誤りが他の正しい単語の読みになってしまうものは, 正しい単語が音訓読み誤りと表音文字誤りのどちらかわからないので, 両方の訂正方式を行う必要がある. それらの単語にはあらかじめフラグを付け, フラグのついている単語の場合はＤＰマッチングによる訂正と両方行うことで, 音訓読み誤りの候補と表音文字誤りの候補が両方出せるようにした.

以下に音訓読み誤りの訂正方式を示す.

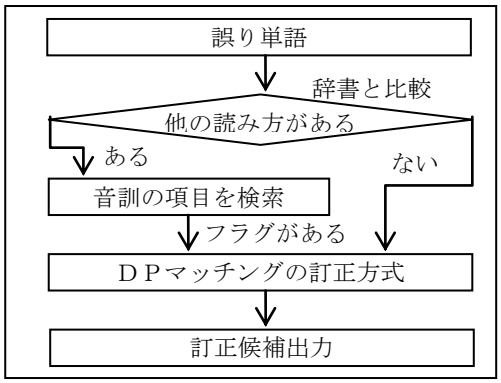

図３． 音訓読み誤りの訂正方式

４． 実験評価と考察

4.1 実験方法と目的

漢字テストより学習者が実際に誤った単語を入力として, 訂正候補の出力数と, その中の正しい単語が何番目に表示されるかを調べた.

4.2 使用データ

今回実験に使用したデータは,漢字テストから収集した表音文字誤りを含む単語, 203単語, 音訓読み誤りを含む単語 28単語を使用した.

4.3 実験結果

表音文字誤りと音訓読み誤りの訂正結果を**表２**に示す

表２． 表音文字誤りと音訓読み誤りの訂正結果

	正しい単語の順位	表音文字誤り	音訓読み誤り
1	1 番目に出力	170	28
2	2 番目	7	0
3	3 番目	4	0
4	4 番目	1	0
5	5 番目	1	0
6	出力されない	20	0
7	合計 (=1+2+3+4+5)	183	28
8	全体 (=6+7)	203	28
	訂正率 (=7/8)	90.0%	100%

4.4 考察

正しい単語を出力できなかった誤りとして,「でんごん(伝言)」を「たいげん」,「さけぶ（叫ぶ)」を「よろぶ」のような相違度が極端に低い単語がある. これは別の単語が訂正候補の上位に出力されるため, 正しい訂正候補が出力されなかったと考える. このような単語に対応できるようにするために文の構文を解析し, 誤り単語の前後関係を調べることで, 正しい単語を推測できるようになるとと考える.

3.2**節**で説明した両方の訂正処理を行う条件は, 単語にあらかじめフラグを付けるという条件に決めた. 他にもＤＰマッチングによる訂正結果の相違度が 5以下の場合に音訓読み誤りの訂正を行うなど, いくつか候補があった. しかし, 他の条件では, 片方の訂正処理を行う条件としては有効だったが, 両方の処理を行う条件としては最適ではなかった為, 現段階では, この条件の結果が最善だと考えた. しかし, この条件では, 未知の単語が出現した場合には対応ができないという問題点があるので, 両方の訂正処理を行う条件を新たに考える必要がある.

５． おわりに

外国人学習者が作成した単語を対象とする, 読み仮名のＤＰマッチングによる訂正方式を提案した.

表音文字誤りが 90.0%,特殊な誤りが 100%という訂正率得られ,この結果から訂正方式として有効だと考える.

今後の課題として, 考察で述べたパラメータの追加と検討事項の改善を行うことで, 誤りの訂正ができなかった単語や処理時間の短縮に対応する.

謝辞

ＤＰマッチングプログラムを作成した開発者の方々に感謝いたします.

参考文献

[1] ＤＰマッチングのプログラム, ソースコード
http://staff.aist.go.jp/toru-nakata/dpmatching.html
[2] 杉野勝也,谷之口優人,絹川博之:外国人の日本語学習時における単語の仮名表記誤りの訂正方式,第 11 回情報科学技術フォーラム (FIT2010) 第 3 分冊 (2010)