

日本語学習者の作文における格助詞の誤り検出と訂正

石川裕司[†] 河合敦夫[†] 多田直人[†] 永田亮[†] 榎井文人[†]

三重大学 工学部 情報工学科[‡]

1 はじめに

近年，第二外国語として日本語を学ぼうとする人々の数が急激に増加している．しかし，日本語教師の育成は，短期間では難しい．そこで本論文では，日本語学習者を対象として，書き誤る可能性の高い格助詞についての誤りを検出・訂正するシステムを考案し，日本語学習者の学習能力の向上を目指した．

本論文では，従来のシステムと異なり，文中のどの位置にある格助詞に対しても，もし誤りが存在すれば検出し，訂正するシステムを提案する．また，入力文となる作文には実際の日本語学習者の書いた作文[1][2]を利用した．

2 システムの概要と構成

2.1 システムの概要

本システムでは，日本語学習者が入力した文に対して，その文が正しければそのまま正文を表示し，誤っていれば誤り箇所と訂正結果を表示する．ただし，訂正候補文が複数ある場合には，それらを全て訂正候補文として出力する．

2.2 システムの構成

本システムの構成を図1に示す．

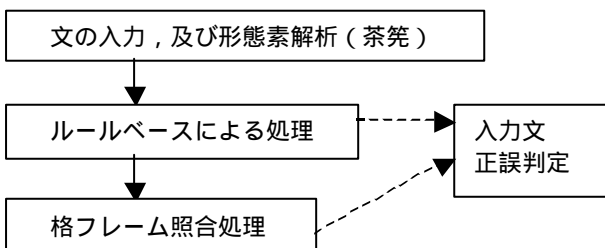


図1: システムの流れ

文の入力及び形態素解析

入力文として，正文・誤文を使用する．制約として，単文のみを扱い，誤りが存在する場合には，それは格助詞の部分で，一箇所だけであると仮定する．また，形態素解析器には茶筌[3]を用いる．

ルールベースによる処理

での出力結果と，品詞を利用したルールを用いることにより，誤りを検出する．ルールベースを用いた処理の具体例を以下に示す．

- ・助詞「を」が2つ含まれている文は誤りとする．
例) 私 は、テニス **を** 練習 **を** する。
- ・隣接できない助詞がある時，誤りとする．
例) 僕 は、それ **しか** **を** 食べない。

格フレーム照合処理

この処理の説明の前に，まず，用語の説明をする．

- ・意味属性
名詞の持つ意味を表すもの．例を図2に示す．
- ・格フレーム
用言の取り得る格（助詞）の情報と，その格の取り得る名詞の意味属性を記述したもの．例を図3に示す．

名詞	意味属性番号(意味属性)
私	8(自称(単数)) 37(私)
日本	30(独立国) 98(大学・高専)
	:

図2: 意味属性の例

暮らす(くらす)
(1)N1がN2に/で 暮らす
N1 4(人)
N2 388(場所) 2610(場)
(2)N1がN2で 暮らす
N1 4(人)
N2 *(全ての名詞)
:

図3: 格フレーム“暮らす”の例

格フレーム照合処理では，始めに，入力文の名詞の意味属性を決定する．意味属性の決定には，NTT 日本語語彙体系 1 意味大系[4]を用いる．次に，入力文の格フレームを作り，辞書の格フレーム（正しい格フレーム）と照合する．辞書の格フレームには，NTT 日本語語彙体系 5 構文大系[5]を用いる．照合の結果，助詞に間違いがあれば，訂正し，訂正結果を出力する．照合方法を図4に示す．

比較の際に，図5の名詞意味属性の木構造を用いる．

Error detection and correction of case particles
in Japanese Learner's composition

[†]Yuji Ishikawa, Atsuo Kawai, Naoto Tada,
Ryo Nagata, Fumito Masui

[‡]Department of Information Engineering,
Faculty of Engineering, Mie University

例えば意味属性 3 は, 3 から 387 までの意味属性を指す。

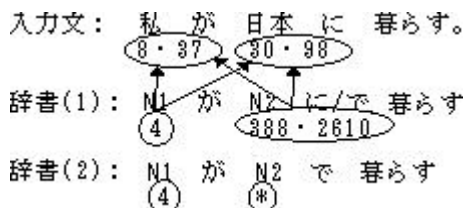


図 4: 入力文と辞書の格フレームとの照合方法



図 5: 名詞意味属性の木構造

図 4 の例では, 辞書(1)の意味属性 4 と入力文の意味属性 8・37 を最初に比較する。意味属性 4 は, 意味属性 8・37 を共に含む。次に辞書の意味属性 388・2610 と入力文の意味属性 30・98 を比較する。このようにして, 全ての入力文と辞書の意味属性を照合する。図 4 から得られた照合結果を表 1 に示す。

(1)

辞書	入力文	私		日本	
		8	37	388	2610
N1	4	○	○	×	×
N2	30	×	×	×	×
	98	×	×	×	×

(2)

辞書	入力文	私		日本	
		8	37	388	2610
N1	4	○	○	×	×
N2	*	○	○	○	○

表 1: 入力文と辞書の照合結果

表 1 の 印は, 入力文と辞書の名詞の意味属性が一致した事を示している。意味属性同士がそれぞれ一致したら, それに付随する助詞を比較する。比較した結果, 助詞が同じ場合は何もせず, 異なっていた時に変換し, 訂正候補文として出力する。

この例では, 表 1 の(2)で, N1 の意味属性 4 と入力文の“私”の意味属性 8・37 が一致し, N2 の意味属性 * と入力文“日本”の意味属性 388・2610 が一致する。次に図 3 の格フレーム(2)の助詞を比較し, “日本”に続く助詞, “に”を“で”に変換する。この流れで得た訂正候補文は, 格フレーム(2)からの「私が日本で暮らす。」となる。また, この例の場合, N2 と“私”の意味属性も一致しているが, N1 と“日本”の意

味属性が一致しないため, 訂正候補文の出力はない。

3 実験

3.1 実験データと実験条件

実験データは, 参考文献[1][2]の中の誤文 80 文と, それを人手で正文とした 80 文の合わせて 160 文を用いた。また, 入力文は全て単文とした。

3.2 評価方法

評価は, 本研究が日本語学習者に対する誤り訂正支援であることを考慮し, 検出・訂正の定義をした。

まず, 誤入力文の誤り検出について定義する。訂正候補文中に入力文と同じ文がなければ, 元の文が誤っていると判断でき, この場合, 検出成功とする。

また, 正入力文の正解検出の場合, 入力文と同じ文が訂正候補文中に含まれていれば検出成功とする。

次に, 誤入力文の訂正について定義する。訂正候補文中に元の文がなく, かつ, 候補の 1 番目に正しく直された文があるとき, 訂正成功とする。

最後に, 正入力文の訂正成功の定義をする。この場合, 正解が検出できれば元の文が正しかったと分かるので, 検出が成功ならば訂正も成功となる。

2.2 で扱った誤入力文「私が日本に暮らす。」は, 検出・訂正ともに成功した例である。

4 実験結果

実験結果を, 3.2 で定義した検出・訂正を用いて表すと, 検出率は, 75.6%(121 文/160 文), 訂正率は, 62.5%(100 文/160 文)となった。

5 まとめ

本論文では, 日本語学習者の書いた作文を題材にして, 格助詞の誤りを検出し, 訂正する手法を提案した。実験の結果として, 75.6%の検出率, 62.5%の訂正率を得ることができ, 本手法が格助詞の誤り検出と訂正に有効であることが示せた。

今後の課題として, 精度を上げるための新たな解析手法の提案とシステム化, 異なる入力文を用いての実験, 日本語学習者が運用できるようなインタフェースの構築, なぜ誤りであるかの情報の付与などがある。については, システム化が終了しつつあり, については, 実験を終え, 分析している最中である。新しい実験結果については, 講演で述べる。

参考文献

[1] 明治書院企画編集部: “日本語誤用分析”, 明治書院
 [2] 市川保子: “日本語誤用例文小辞典”, 凡人社
 [3] 松本祐治, 北内啓: “日本語形態素解析システム「茶筌」 version2.0 使用説明書”
 [4] 池原悟, 他: “日本語語彙大系 1 意味体系”, 岩波書店
 [5] 池原悟, 他: “日本語語彙大系 5 構文体系”, 岩波書店