

# 外国人の初級日本語文における助詞誤りの検出

細田 裕樹†

杉野 勝也††

絹川 博之†††

東京電機大学院 未来科学研究科‡

## 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、学習者の作成した文章から助詞誤りを検出する方法を研究している。本稿では、その中の助詞の検出について報告する。

## 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

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、文章時に助詞の誤りをすることが特に多く、助詞を理解することは文章作成を正しく行えることにつながるため、重要とされている。本研究では、日本語学校で学習している外国人が作成した文章中の助詞について自動検出を行うことを目的としている。

### 2.2 対象とする助詞

外国人学習者が作成した作文から文章を調べた結果、特に多かった6種類の助詞の中から主に4種類の「が」「を」「に」「で」について扱う。「は」「の」の処理には、本格的な構文解析が必要であるので、今回は対象外とした。

### 2.3 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。

現在単語の訂正については報告されており、本研究では上記の助詞誤りの検出について述べる。

## 3. 助詞誤りの検出方式

### 3.1 助詞誤りの判断

助詞の誤りを検出する方法として助詞によって接続されている名詞と動詞の関係性が重要であると考えた。

Detection of Particle Errors in Foreigner's Basic Japanese Text  
†Yuki Hosoda, ††Katsuya Sugino, †††Hiroshi Kinukawa  
‡Graduate School of Science and Technology for Future Life,  
Tokyo Denki University

例を挙げると「私は試験の合格したいです」

この例文を見てみると、「私は試験」と「合格したい」の接続関係を調べるとその間にある助詞が誤っていることがわかり、正しくは「に」を使用した「私は試験に合格したいです」であることがわかる。

このことから名詞の意味と動詞の意味から正しい助詞を導き出せるのではないかと考えた。

### 3.2 動詞辞書

動詞にはそれぞれ選択制限規則が存在する。

動詞には接続親和性の高い名詞の意味素性が存在し、さらに付接すべき助詞にも制限がある。このことから接続親和性の高い名詞の意味素性の情報と各動詞に付接すべき助詞を持つ辞書を作成する。

動詞辞書に載せる名詞の意味素性は IPAL 名詞辞書名詞辞書で使用されている 16種類の意味素性を使用する。その例を下記の表 1 に示す。

表 1. 意味素性分類表

略号	素性名	例
ANI	動物	犬・虫・貝
HUM	人間	学生・男性・女性
ORG	組織・機関	企業・国・県
PLA	植物	花・森・木
PAR	生物の部分	頭・腕・足
NAT	自然物	山・川・石
PRO	生産物	車・布・パソコン
PHE	現象	光・音・風
ACT	動作・作用	勉強・仕事
MEN	精神	心・気持ち・悩み
CHA	性質	美・青・赤
STA	状態	安定・静止・固定
REL	関係	原因・系列
LOC	空間・方角	公園・場所・駅
TIM	時間	昨日・明日・時
QUA	数量	3人・100円

作成した動詞辞書は初級日本語で使われている 100程度を対象とし、漢字とその読み、それぞれ接続親和性の高い名詞の意味素性と付接する助詞を加えた。下記にその例を示す

表 2. 動詞辞書の例

読み	漢字	が	を	に	で
いく	行く	HUM	—	LOC	PRO

### 3.3 助詞誤りの検出方法

助詞の誤りを確認するには前後の動詞と名詞の意味素性から正しいかどうか判断する必要がある。

本研究では形態素解析ツール Mecab を用いて文章の形態素解析を行い、動詞を取り出す。例外として動詞「する」は直前の名詞がサ変接続だった場合、「する」の直前の名詞を「名詞+する」の複合動詞として扱う。また助詞「を」はほとんどが他動詞にしか接続しないという規則から動詞が他動詞でないのに「を」が使用されていた場合には誤りと判断する。動詞の直前には助詞がある可能性が高く、さらにその前には名詞が来ると仮定して処理を行っていく。文章から動詞を取り出し、辞書から情報を読み込む。直前の名詞の意味素性と一致している助詞が使用されているかの判定を行い、一致している場合には正しく使われている、一致していない場合には誤っていると助詞を判定する。一連の処理の流れを下記に示す。

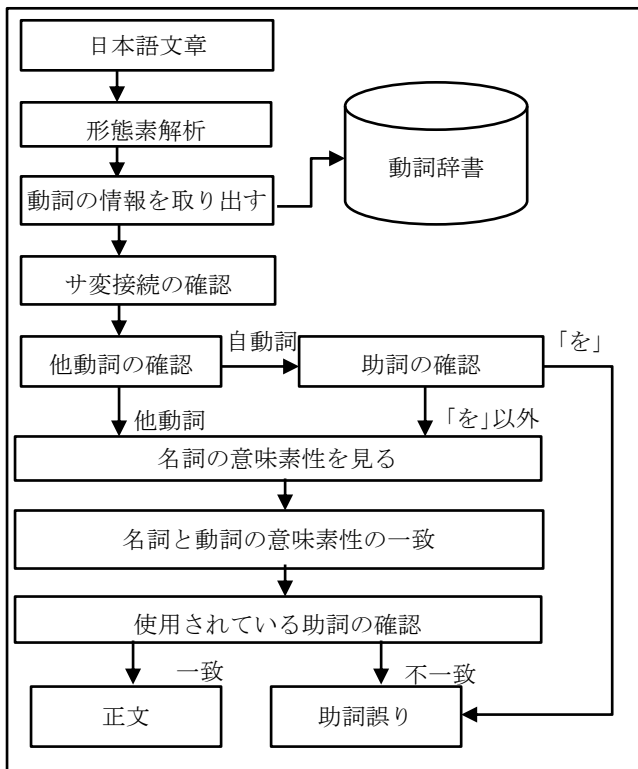


図1. 助詞誤りの検出方法

## 4. 実験評価と考察

### 4.1 実験方法と目的

外国人学習者が作成した文章を入力として、誤った助詞を正しく検出するかの検出率を調べた。

### 4.2 使用データ

今回実験に使用したデータは、実際に外国人学習者が作成した作文約100枚から誤った助詞が35カ所の誤文と誤つてる部分を人手で直した35カ所の正文を使用した。

### 4.3 実験結果

助詞誤りの検出結果を下記に示す

- (1) 誤り助詞35カ所を含む誤文: 誤り助詞と判定した数34
- (2) 正しい助詞35カ所を含む正文: 誤り助詞と判定した数7

### 4.4 考察

正しく誤りを検出が出来なかった文は、複合動詞「発明できる」が使われていたため、名詞「発明」と動詞「出来る」に分けて判断されてしまったためうまく助詞を検出することが出来なかった。そのため「する」だけでなく他にも複合動詞になる例を探し対応していく必要があると考える。

正しい文で誤り助詞と判定されてしまった7カ所は、下記の表3のように動詞辞書に登録されている意味素性が2つ以上助詞に存在する場合誤りの可能性があるとして判定されてしまった。

対応策としては、2つ以上の候補があった場合には格フレームのように基本文型を作成し、それに適応する場合としない場合に分けることで、候補を1つに絞れるのではないかと考えている。さらに授受表現「くれた」、「もらった」といった文には前の助詞によって後につく助詞は決まるので、このような例を多く見つけ出すことでさらに選択肢を減らせるのでは無いかと考える。

表3. 意味素性が重複する動詞の例

読み	漢字	が	を	に	で
とる	撮る	HUM	HUM	HUM	PRO

## 5. おわりに

外国人学習者が作成した文章を対象とする、助詞の誤り検出方式を提案し、実験評価した。

この結果から本方法を発展させていけば、有効な手法になるのではないかと考える。今後の展開としてはより多くのデータを集め、考察で述べた複合動詞として使われる動詞の発見と、複数の意味素性が見つかった場合の処理について対応していく。

### 謝辞

形態素解析ツールMecabを作成した開発者の方々、IPAL名詞辞書を作成した方々に感謝いたします。

### 参考文献

- [1] Mecab のプログラム, ソースコード  
<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>
- [2] 今枝恒治,河合敦夫,石川裕司,永田亮,榎井文人:日本語学習者の作文における格助詞の誤り検出と訂正,情報処理学会研究報告 コンピュータと教育研究会報告 (2003.2.7)
- [3] 南保亮太,乙武北斗,荒木健治:文節内の特徴を用いた日本語助詞誤りの自動検出・校正,情報処理学会研究報告 2007-NL-181
- [4] 橋本利典,島田静雄:外国人の書いた文章の助詞使用誤りの抽出,情報処理学会研究報告 自然言語処理 (1997.1.20)
- [5] 情報処理振興事業協会:計算機用日本語基本名詞辞書 IPAL(1996)