



# 日本語学習者の文法的誤用文検出に関する研究

山本, 卓司

---

(Degree)

博士 (学術)

(Date of Degree)

2015-09-25

(Date of Publication)

2017-09-25

(Resource Type)

doctoral thesis

(Report Number)

甲第6495号

(URL)

<https://hdl.handle.net/20.500.14094/D1006495>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



(別紙様式3)

## 論文要旨

氏名 山本 卓司

専攻 グローバル文化

指導教員氏名 森下 淳也 教授

論文題目

日本語学習者の文法的誤用文検出に関する研究

### 論文要旨

近年の日本語学習者の誤用に関する研究においては、作文などの学習者コーパスのデータが使用されることが少なくない。Web 上などに公開されている学習者コーパスは、何も処理がされていないテキストファイルを集めたものが多い。誤用研究などに関心を持つ研究者や教師をユーザーと仮定すると、ユーザーは、文献や日本語を教えた経験から得た誤用に注目し、テキストからひとつずつ目で確認するか文字列検索をして誤用が書かれた文を取り出していることが考えられる。このような方法に対して、ユーザーは多様な誤用文をできるだけ簡便に取り出すことができる手法を求めていることが考えられる。本研究では、ユーザーが注目するような日本語学習者の文法的な誤用文をテキストコーパスから簡便に検出する新たな手法の提案を目的にした。さらに、提案手法に基づいて検出した誤用文が誤用研究などにおいて、有効に利用できるか検討した。

本研究における提案は、文に形態素情報を付けて表現式で検索し、誤用文を検出する手法である。形態素情報とは、形態素解析器で文を単語に分割し、単語に付けた品詞などの情報である。表現式とは、形態素情報を組み合わせて作成した文字列式である。つまり、形態素情報を付けた学習者コーパスの文に対して、多様な誤用文を検出するように設定した表現式を使って検索し、日本語学習者の文法的な誤用文を効率よく検出する手法である。

本研究における文法的な誤用文とは、単語の品詞は正しいけれども、品詞の接続は正しくない文である。たとえば、「難しい(形容詞)でし(助動詞)た(助動詞)」という文は、単語の品詞は正しいけれども、「単語(品詞)+単語(品詞)」は日本語として正しく接続していない。本研究においては、このような品詞間に接続の誤りがある文を文法的な誤用文とする。

本研究においては、ユーザーが注目するような誤用文を有効に検出できるか実験した。実験において、検出されなかった誤用文を分析し、どのような誤用文が有効に検出できないか考察した。また、実験結果と考察をふまえて、提案手法が誤用文を検出する手法として有効であるか検討した。さらに、形態素解析器は誤用文を入力してもルールに基づいて単語を分割する特性があることを利用して、誤用文の検出実験および考察と検討を行った。

第1章では、本研究の背景と目的および提案の概要と関連研究について述べた。

第2章では、形態素情報と表現式を用いた日本語学習者の誤用文の検出手法を提案した。はじめに検出手法の概要について述べ、次に形態素情報を付けた文と表現式の概要について述べた。また、提案手法における簡便性と有用性について述べた。

第3章では、提案手法を用いて品詞間に接続の誤りがある誤用文の検出実験を行った。品詞間に接続の誤りがある誤用文の例として、文中に誤って書かれた形容動詞の名詞修飾表現に注目した。実験の結果、誤用文が検出された。また、実験で検出されなかった文を分析し、提案手法では検出されない誤用文を考察した。実験の結果と考察をふまえて、提案手法が誤用文を検出する手法として有効であるか検討した。その結果、品詞間に接続の誤りがある誤用文において、有効であると判断した。

第4章では、形態素解析器の特性を利用した誤用文の検出実験を行った。実験では、動詞の活用のひとつであるテ形の誤用文を例にして、形態素解析器の特性である単語の分割パターンと表現式を用いた検出を行った。実験の結果、誤用文が検出された。また、実験で検出されなかった文を分析し、特性を利用しても検出されない誤用文を考察した。実験の結果と考察をふまえて、形態素解析器の特性を利用して、誤用文が有効に検出できるか検討した。その結果、動詞など活用に誤りがある誤用文において、有効であると判断した。

第5章では、提案手法に基づいて検出した誤用文が誤用研究などにおいて、有効に利用できるか検討した。提案手法に基づく例として、中国語を母語とする学習者が品詞間の接続を誤った接尾辞「的」を検出した。その結果、誤用文が検出された。検出された誤用文を利用して、学習者が誤りを生じる原因に母語が影響しているか検証した。その結果、母語が影響していることがわかった。さらに、検証結果をふまえて、誤用の原因の考察した。提案手法に基づいて検出した誤用文を利用して学習者の誤りの原因を検証できたことから、本研究の手法で検出した誤用文が誤用研究などにおいて、有効に利用できるか検討した。その結果、有効に利用できるかと判断した。

第6章では、本論文を総括し、今後の課題を述べた。

論文審査の結果の要旨

氏名	山本 卓司		
論文題目	日本語学習者の文法的誤用文検出に関する研究		
判定	合格 不合格		
論文チェックソフトによる確認	<input checked="" type="checkbox"/> 確認 <input type="checkbox"/> 未確認 理由：		
審査委員	区分	職名	氏名
	委員長	教授	康 敏
	委員	教授	田中 順子
	委員	神戸大学名誉教授	水野 マリ子
	委員	教授	森下 淳也
委員			印
要 旨			
<p>本論文は、日本語学習者の作文などを収集した学習者のテキストコーパスデータ(学習者コーパス)から誤用文(日本語として誤りのある文)を検出する手法に関する研究である。日本語教育の研究者達が学習者の誤用を分析する際には、学習者コーパスの作文中の誤用を直接目視で確認する、或いは誤用文を想定して文字列検索にて学習者コーパスから抽出するといった方法が取られている。これに対して、著者である山本卓司は学習者コーパスから多様な誤用文を出来るだけ簡便に取り出す手法を本研究で提案し、誤用研究を効率化することを目指している。</p> <p>第2章においては、学習者コーパスの文に対する構造化テキストの定義とその定義に基づく誤用文の表現式を定めている。まず学習者コーパスの文を形態素解析器にかけて、文中の語句に対する文法的属性(品詞、活用型、活用形など)を抽出する。抽出した属性を、文中には出現しない特殊な記号と文字列を用いてラベルを定義し、そのラベルと抽出した属性値の組み合わせを、文の語彙の順序</p>			

を保ちながら元の文に埋め込むことで構造化テキストを定義する。構造化テキストからラベルと属性を除けば、元の文に戻る。また、特定の属性に着目すれば、抽象的な文の文法表現を得ることが出来る。この構造化テキストは文を多角的な視点で見ることの出来る表現である。構造化テキストを検索する表現式には正規表現を用いる。正規表現は文字列のパターンを記述できる検索式である。構造化テキストに埋め込まれたラベル・属性値の特定の値を検索で調べることができ、また、順序関係も検索に指定できる。直接の語句ではなく誤用文の形態素解析結果から得られた文法的属性値を表現式に組み込むことで、属性値から誤用文を検出する表現式ができる。この表現式を用いて学習者コーパスから抽出された文は、特定の語句に依らない、より広範な誤用文の集合となる。

第3章では、この構造化テキストと誤用の表現式を使って、品詞間の接続に誤りがある誤用文として、「～な」、「～の」、「的な」、「的の」といった誤用表現を台湾人学習者の学習者コーパスを対象に調べている。その結果、有効性が確認された。

第4章では、動詞の活用形の一つであるテ形の誤用文を調べている。目視による誤用判定で、研究者に同列のものとして扱われる誤用(いわば誤用の文型)は、必ずしもここで定義されている文法的誤用の表現式と一対一に対応するものではなく、多くの異なる文法的誤用表現式の集合として記述される。ここでは21件の表現式を生成して第3章と同じ学習者コーパスを対象に調査を行ない、有効性を確認している。

第5章には、第3章で作成した誤用表現式を用いて、台湾人学習者の学習者コーパスとWeb上に公開されている「日本・韓国・台湾の大学生による日本語意見文データベース」を対象として誤用の生じる原因を調査する分析を行なった。手法の有効性が確認された。

第6章に結論と将来への課題が述べられている。

文を形態素解析器からの構造化テキストに直して分析に用いる方法は、既に多くの研究者によって行われている。しかしながら、これを誤用文を対象として、誤用の発見に用いるという発想は今までにない斬新なものと評価されている。従来とは異なる文法的誤用の分類など、誤用の見方に新しい視点を与えるものともなる。将来が大いに期待できる研究である。

なお、著者は博士後期課程在籍中にこの研究に関連して、査読付きの2件の学会発表と査読付きの2件の国際会議報告、計4件を発表している。①山本卓司・大月一弘・森下淳也(2013)「形態素の素性情報を利用した日本語学習者の誤り共起表現の検索—中国語を母語とする日本語学習者の形容動詞表現—」、計量国語学会 第五十七回大会 予稿集、pp.7-12。②山本卓司・清光英成・大月一弘・森下淳也(2014)「日本語学習者によるテ形誤りの自動検出—形態素解析器を利用した誤り文の検出—」、計量国語学会 第五十八回大会 予稿集、pp.13-18。③山本卓司(2014)「日本語作文の翻訳からみる台湾人日本語学習者の接尾辞「的」の誤用」、2014年国際学術研討會—應用日語教育の理論と実践—大会論文集、pp.41-48。④山本卓司(2014)「共起頻度からみる台湾人日本語学習者の接尾辞「的」の誤用」、2014年日語教學國際會議 大會手冊、pp.1-11。

よって学位申請者、山本卓司は、博士(学術)の学位を得る資格があると認める。