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

博　士　論　文

日本語学習者の文法的誤用文検出に関する研究

平成 27 年 6 月

神戸大学大学院国際文化学研究科
山本　卓司

博　士　論　文

日本語学習者の文法的誤用文検出に関する研究

審　査　委　員：康　　敏　教　　授
水　野　マ　リ　子　神戸大学名誉教授
田　中　順　子　教　　授
森　下　淳　也　教　　授

平成 27 年 6 月

神戸大学大学院国際文化学研究科
山本　卓司

博　士　論　文

日本語学習者の文法的誤用文検出に関する研究

所属専攻・コース：グローバル文化・情報コミュニケーション
氏　　名：山　本　卓　司
指導教員氏名　：森　下　淳　也

要旨

本論文は，電子化処理がされた日本語学習者のテキストコーパスデータから，文中に日本語として誤りがある誤用文を検出する手法を提案したものである．

近年，日本語学習者の誤用に関する研究においては，日本語学習者コーパスのデータを活用したものが少なくない．日本語学習者コーパスは Web 上などに公開されている．本研究において調査したところ，公開されている日本語学習者コーパスは，作文など学習者が産出した日本語をデータとし，多くがテキストファイルなどの電子化処理のみを施したものであり，またそれらファイルのデータ中に存在する学習者の誤りは訂正されずに電化処理が行われていた．日本語教育の研究者らが，これらテキストコーパスのデータを活用して誤用を分析する際には，誤用を直接目視で確認するか，あるいは，できるだけ多くの誤用文を想定して文字列検索をするといった方法で誤用文の取り出しを行っている．しかし，このような方法は，効率的ではない．そこで，本研究においては，日本語学習者のテキストコーパスデータから多様な文法的誤用文をできるだけ簡便に取り出す手法を提案した．

第 2 章では，本研究における提案について述べた．本研究の提案は，日本語学習者のテキストコーパスの文を基にして独自に定義した構造化テキストを作成し，その定義に基づて作成した表現式を用いて構造化テキストに検索をかけ，検索にマッチした文を取り出す手法である．構造化テキストは文によって構成され，文には形態素情報を埋め込み，形態素情報にはラベルを付けて定義することを述べた．また，その定義に基づいて作成して検索をかける表現式について述べた．表現式は形態素情報を用いて検索をすることから，文の集合にマッチして多様な文が取り出せることを述べた．

第 3 章では，構造化テキストと誤用文を検出する表現式を用いて，品詞間に接続の誤り

がある誤用文を検出する実験を行い，提案手法の有効性を確認した．検出例として，形容動詞の名詞修飾表現の誤りに着目して検出実験を行った．また，この表現における誤りを定め，定めた誤りを基に表現式を作成して日本語学習者のテキストコーパスに検索を行った．その結果，誤用文が検出され，有効性が確認できた．

第 4 章では，人が目視で確認した分類で同類とした誤用文を提案手法で再現できるか検証実験を行った．従来の分類の方法は，人が誤用の基準をめてその基準に従って誤用文を分類している．本研究で提案した手法は，検索して見つけたものに取り出しを行っているが，これは機械が自動で誤用文を分類していると考えることができる．そこで，目視で同類であると分類した誤用文を提案手法で再現する検証実験を行った．再現の例には，動詞のテ形活用誤りとし，その誤用と表現式を一対一に対応させるのではなく作成して再現を行った．実験の結果，提案手法を用いて目視で同類である誤用を再現することができ，有効性が確認できた．

第 5 章では，提案手法を用いて効率的に誤用文を検出し，検出したデータを活用して学習者が誤りを原因の調査と分析を行い，提案した手法が誤用研究に有効であるか確認した．ここでの有効とは，誤用文の検出が目視や文字列検索よりも効率的であったか，さらに表現式検出法で取り出したデータが学習者の誤用を分析することができたかである．提案手法を用いて二つの日本語学習者コーパスから誤用文を検出した．誤用文を検出する表現式は，第 3 章で作成したものを再利用した．その結果，提案手法で取り出したデータは学習者が誤用を生み出す原因を分析することができ，また表現式が再利用できたこと，さらに二つの異なる日本語学習者のテキストコーパスから誤用文が検出できたことから，目視や文字列検索よりも効率的に誤用文を取り出すことができ，有効性が確認できた．

第 6 章では，本論文を総括し，今後の課題を述べた．

目次

要旨 i

第 1 章　序論 1

1．1　研究の背景と目的 1

1．1．1　近年の日本語学習者コーパスの特徴 2

1．1．2　学習者の誤用と日本語学習者コーパスにおける誤用の扱い 3

1．1．3　日本語学習者コーパスにおける誤用文検出の課題 4

1．1．4　データの活用を考慮した検出と多種多様な誤用文の検出 5

1．2　本研究における提案 6

1．2．1　形態素解析器を利用した誤用文の検出と文の構造化 7

1．2．2　単語における文法的属性の抽象化レベルについて 10

1．2．3　日本語学習者のテキストコーパスを構造化した文に処理 14

1．2．4　構造化テキストを検索する表現式について 17

1．2．5　本提案手法の命名 21

1．2．6　表現式検出法の環境について 21

1．2．7　表現式検出法の有用性と誤用研究に貢献の可能性 22

1．3　関連研究 28

1．3．1　形態素解析と形態素解析器について 28

1．3．2　日本語学習者の誤用文自動検出について 30

1．4　本論文の構成 37

第 2 章　構造化テキストの定義とその定義に基づく表現式 39

2．1　本章のあらまし 39

2．2　コーパスの文に形態素情報を付加した構造化テキストについて 39

2．2．1　一文単位の文と単語の順番を保持した形態素解析器の処理 40

2．2．2　形態素解析器から出力された情報の選択と情報へのラベル付け 42

2．2．3　構造化テキストに対する検索の可能性 45

2．3　構造化テキストを基に作成する表現式について 46

2．3．1　表現式における属性の選択と組み合わせによる作成例 46
2．3．2　表現式を作成するための形態素解析器への入力 54
2．4　表現式検出法の簡便性について 55
2．7　本章のまとめ 57

第 3 章　品詞の接続法に誤りがある誤用文の検出実験 60
3．1　本章のあらまし 60
3．2　品詞の接続法に誤りがある誤用文とは 60
3．2．1　形容動詞の名詞修飾表現と学習者の誤用文 60
3．2．2　実験に使用するデータ 61
3．2．3　正しい形容動詞の名詞修飾表現の調査とその結果 61
3．2．4　誤用表現とその判別基準 61
3．2．5　目視で形容動詞の名詞修飾表現の誤用文を判別した結果 62
3．2．5　誤用表現を分類した結果 62
3．2．6　機械による自動分類器としての表現式 63
3．2．7　表現式を使った誤用文の検出結果 70
3．2．8　目視による判別と表現式による検出の比較 71
3．2．9　表現式で検出されなかった誤用文の分析 72
3．3　品詞の接続法の誤りがある誤用文検出の検討 74
3．4　本章のまとめ 75

第 4 章　従来の分類において同類である誤用文の再現 77
4．1　本章のあらまし 77
4．2　表現式検出法の分類と目視による分類 77
4．2．1　動詞の活用誤りとテ形における学習者の誤用 77
4．2．2　実験に使用するデータ 78
4．2．3　正しいテ形表現の調査とその結果 78
4．2．4　誤用表現とその判別基準 78
4．2．5　目視でテ形の誤用文を判別した結果 79
4．2．6　表現式作成のためのテ形活用語尾の参照 80

4．2．7　目視の分類を再現するテ形誤用文の表現式 80
4．2．8　テ形の誤用文を検出する表現式を用いた検出結果 127
4．2．9　表現式で検出されなかった誤用文の分析 127
4．3　誤用文の分類としての表現式検出法の検討 130
4．4　本章のまとめ 130

第 5 章　誤用研究における手法の有効性 133
5．1　本章のあらまし 133
5．2　日本語学習者の誤用分析と表現式検出法によるデータの活用 133
5．2．1　接尾辞「的」と日本語学習者の「的」関する誤用 133
5．2．2　本研究で着目する「的」の誤用と母語の影響を調査する方法 134
5．2．3　使用するデータと文の解析 135
5．2．4　誤用表現とその判別の手順 135
5．2．5　誤用文を検索する表現式 136
5．2．6　単語と単語が共起する強さ 136
5．2．7　再利用した表現式による日本語学習者コーパスの検索 137
5．2．8　学習者が翻訳した誤用文の検証 137
5．2．9　中国語における共起の強さの検証 138
5．2．10　学習者の母語の影響の分析 139
5．3　誤用研究における手法の有効性について 140
5．4　本章のまとめ 140

第 6 章　結論と今後の課題 141
6．1　結論 141
6．2　今後の課題 143

参考文献 144

業績リスト 150

謝辞 153

図表目次

図 1. 1 単語を分割して単語に文法的属性を付けた例 10
図 1. 2 単語における抽象化のレベル 11
図 1. 3 品詞だけで検索した例と品詞と活用形で検索した例 13
図 1.4 JUMAN によるラティス構造(笹野ら[53]から) 28
図 1. 5 茶筌と ipadic で返された入力文 29

図 2. 1 テキスト上のラベル付け処理を行った例 44
図 2. 2 ラベル付き単語を一文にした例 45
図 2. 3 ラベル付き属性の単語と単語の境界 47
図 2. 4 表現式作成のための形態素解析器への入力例と得られる情報 54

表 1. 1 「大きいでした」の形態素情報 8
表 1. 2 タブで区切られた単語の情報 30

表 2. 1 句点を基準とした文の例 40
表 2. 2 形態素解析器から出力された情報と単語の順序 41
表 2. 3 選択処理を行った形態素解析からの単語情報の例 42
表 2. 4 ラベル付け処理を行った単語の文字列と文法的属性の例 44

表 3.1 目視で判別した形容詞名詞修飾表現の誤用文と誤用表現の件数 62
表 3.2 目視で判別した形容詞名詞修飾表現の各誤用表現の件数 62
表 3.3 形容動詞名詞修飾表現の表現式で使用した正規表現 70
表 3.4 形容動詞名詞修飾表現の表現式で検出された文と表現の件数 70
表 3.5 目視で判別した件数と表現式で検出した件数 71
表 3.6 表現式で検出されなかった文の表現 72
表 3.7 正しく書かれていない単語の形態素情報 73

表 4.1 目視で判別したテ形の誤用文と誤用表現の件数 79

表 4.2 「て」に関する活用語尾の誤用表現 80
表 4.3 テ形の誤用文を検出する表現式の種類 80
表 4.4 テ形の表現式で使用した正規表現 127
表 4.5 テ形の表現式で検出された文と表現の件数 127
表 4.6 表現式で検出されなかったテ形の誤用表現 128
表 4.7 常用漢字表の送り仮名を基にした正しいテ形と辞書の登録 128

表 5.1 翻訳付きデータと台湾人学習者データの検出件数 137
表 5.2 翻訳付きデータの誤用表現 137
表 5.3 台湾人学習者データの誤用表現 138
表 5.4 共起の強さを測った表現と t スコアの値 139

第 1 章

序論

1. 1　研究の背景と目的

近年，日本語学習者の誤用に関する研究においては，日本語学習者コーパスのデータを活用したものが少なくない．日本語学習者コーパスは Web 上や CD-ROM などを配布する形で公開されている．コーパスのデータは，学習者の作文などをテキストファイルに打ち直して電子化処理を行い，電子化ファイルを収集して一つのコーパスと定義したものが多い．日本語学習者コーパスから誤用の取り出しを行っているユーザーは，学習者が産出した誤りに関心を持つ教師や研究者が想定される．それらのユーザーは，日本語を教えた経験から学習者が誤りやすい誤用に関する知見を持ち，また日本語の教え方に関する書籍や日本語学習者の誤用に関する論文・辞典から誤用の傾向の知見を得ている．そして，経験と文献から得た知見から学習者の様々な誤用に着目し，データを研究や日本語の指導に活用している．着目される誤用には，単語の誤りや動詞・形容詞などの文法・表現の誤りといった種類の異なるタイプの誤りと，「見る」，「読む」，「する」のような同じ動詞ではあるがバリエーションが異なる誤りがある．このような誤りのタイプとバリエーションが異なる多種多様な誤りについて，ユーザーは公開されている日本語学習者コーパスからの取り出しを行っている．

しかし，多くのユーザーが行っている誤用文の取り出し方法は，着目した誤用を目視で一つずつ確認するか，あるいはできるだけ多くの誤用文を想定して文字列検索をするといった方法を行っている．これは，多くの日本語学習者コーパスが電子化処理しかされておらず，誤用文の取り出しに対応した処理がされていないからである．多くのユーザーが用いている方法で，目視による確認は，多様な誤用文を取り出すことができるが，作業には手間がかかり効率的に取り出すことができない．文字列検索を用いた取り出しは，目視の確認よりは簡便な取り出し方法ではあるが，ユーザーが知見から得ている誤用の文字列を検索するか，または知見を基に想定した文字列の検索しか行うことができない．このようなユーザーが用いている方法を考慮すると，多種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が必要であると考える．多種多様な誤用文を簡便に取り出すことがで

きれば，誤用に関する研究において，電子化処理しかされていない日本語学習者コーパスをさらに活用することができるとともに，効率的に研究を進めることができると考える．

そこで，本研究では電子化された日本語学習者テキストコーパスから多種多様な文法的誤用文をできるだけ簡便に検出する手法の提案を目的にする．

さらに，本研究の提案手法を用いてテキストコーパスから誤用文を検出し，誤用の生じる原因を調査する分析を行い，手法の有効性を確認する．

1．1．1　近年の日本語学習者コーパスの特徴

近年，日本語教育において，日本語学習者コーパスを利用した研究は少なくない．たとえば，学習者が使用した単語・文法などの実態調査や文法的誤りの原因を分析するデータに日本語学習者コーパスが利用されている[1～4]．

学習者が使用した日本語の調査や分析において，コーパスの利用が増えるなか，近年，日本語学習者コーパスの構築と公開が進められている[5～12]．

■ 本研究で調査した日本語学習者コーパスの特徴

本研究において，日本語学習者コーパスを調査したところ，Web 上にデータを公開，CD-ROM または DVD を配布しているものが 17 件あった[7，13～28]．調査した日本語学習者コーパスを分析したところ，次のような特徴があった．

1. 学習者が産出した日本語のデータを提供している
2. 電子化ファイルをコーパスと定義している
3. 検索して文が検出できる
4. 文中の誤りを訂正していない
5. 文中に誤りのある文を検出できる

特徴 1 は，学習者の会話・作文・日本語文の作例をデータとして提供していた[7，13～28]．

特徴 2 は，特徴 1 のデータをテキストファイル，PDF ファイル，Microsoft 社の Word または Excel ファイルに打ち直して電子化処理し，ファイルを一つまたは複数集めたものを一つのコーパスと定義していた［7，13，15～17，20，21，24，27，28］．

特徴 3 は，単語や表現を検索して，文を検出できるコーパスであった[14，18，19，22，23，25，26]．文の検出は，ユーザーが文字列の入力または単語や表現を選択して検索し，システムは入力または選択にマッチした文を検出した．

特徴 4 は，文中に学習者による単語や文法的な誤りがあっても，正しい日本語に訂正していなかった[7，13～28]．電子化されたファイルにおいては，誤りは訂正せずに打ち直していた．教師や日本人の添削をシステムに反映させたコーパスにおいては，別窓に訂正内容を表示する[18, 22]，または誤っている単語や表現の周辺に訂正内容を表示していた[19, 26]．

特徴 5 は，誤りのタイプを定めて分類し，分類ごとのタグを準備してそれらを誤りの箇所に付けていき[11，12，29，30]，タグを付けた文をシステムに収めて誤りの検索と検出を可能とするものであった[18，19，23，26]．これらのシステムは，ユーザーが誤りのタイプを選択し，検索して文を検出する方式[18，23]と，文字列を入力して検索し，誤りのタグとマッチした文を検出する方式[26]と，文字列入力と誤りタイプの選択を組み合わせて検索して文を検出する方式[19]であった．

■ 調査した日本語学習者コーパスのまとめ

現在公開されている日本語学習者コーパスを分析した結果，その特徴がわかった．その特徴は，すべてのコーパスは学習者が産出した日本語をデータとし，データ中に存在する学習者の誤りは訂正されていなかった．また，多くの日本語学習者コーパスは，学習者が産出データをテキストファイルなどに電子化処理しただけのものであった．さらに，システムに収められた学習者コーパスには検索機能があり，誤りのタイプを分類してタグを付けたシステムは，文中に学習者の誤りがある文の検索と検出が可能であった．

1．1．2　学習者の誤用と日本語学習者コーパスにおける誤用の扱い

特徴 4 の学習者による誤りは，一般に誤用と呼ばれている．誤用は，学習者が書いたり話したりしたとき，読み手・聞き手であるネイティブが違和感を覚える単語の使い方や表現であるとされる[31]．また，日本語学習者コーパスにおいて，誤用は訂正を行わず処理されている．

■ 学習者の誤用について

誤用には，単純に言い間違えたなどのミステイクと，学習目標とする言語の知識がまだ十分でないために引き起こされるエラーがあることが指摘されている[32]．エラーは学習者が繰り返し引き起こす誤りである．しかし，学習者のエラーは，目標言語を習得する過程に現れる必然的な現象であり，習得のための方略であると考えられている [32]．また，学習者の誤用には共通性がみられ，不十分ながらも学習者独自の体系的な言語であり，目標言語を習得する過程における母語と第二言語との中間的な言語であると考えられている [33].

■ 日本語学習者コーパスにおける誤用の扱い

調査した日本語学習者コーパスにおいて学習者の誤りは，ミステイク，あるいはエラーに関わらず，正しい日本語に訂正されていなかった．学習者の誤用が訂正されていない理由は，使用した日本語の実態を反映させるることで，調査や誤りの分析などに活用できるからである．

1．1．3　日本語学習者コーパスにおける誤用文検出の課題

一方，特徴 2 のように，多くの日本語学習者コーパスは電子化処理がされているものの，直接資料として提供されている．この直接提供された資料から誤用をどのように検出するかが課題となっている[34].

■ 公開されている学習者コーパスにおける誤用文検出の問題点

1.1 節で述べたように，電子化処理しかされていないテキストコーパスは，取り出しに対応した処理がされていないため，多くのユーザーは一つずつ目視で確認していくか，文字列検索の方法で誤用の取り出しを行っている．また，特徴 5 のように，タグを付けて誤用文を検出するシステムもあるが，タグ付けの作業は人手で行われるため，同じ誤用に対して同じタグが付けられているかといった揺れの問題などが存在する[5，35]．タグは単語・文法・意味などの誤用に対して柔軟に付けられる利点はあるが，検出された誤用文は揺れがなく客観性に判別されたものかという点については問題がある．そのため，客観性のある誤用文の検出が必要であると考える．

■ 機械を利用した誤用文の検出手法とユーザーのデータ活用

客観性の疑問を解消するためには，機械を利用した誤用文の検出が考えられる．機械を利用した日本語学習者の誤用文検出については，主に辞書と照合する手法と機械学習の手法が提案されている．これらの提案手法の多くは格助詞の誤りを検出し，訂正することを目的にしている．

検出された誤用文データについてのユーザーの活用を考えると，誤りは訂正するのではなく，提示する方式での取り出しを行うべきではないかと考える．ユーザーは取り出されたデータに対し，さらに任意に定めた基準で確認を行い，確認したデータをケースとして誤用の原因などの分析を進めていくからである．

また，格助詞は学習者が誤りを生じやすい文法的な誤用の一つであるため，検出対象にされている．しかし，文献には格助詞以外の誤りも取り上げられている．たとえば，「試験は難しいでした」のような形容詞過去の表現の誤り[36]や，「必要の注意を怠る」のような形容動詞の名詞修飾表現の誤り[37]などが学習者の誤りやすい文法的な表現として指摘されている．そのため，1.1 節でも述べたように，ユーザーはこのような文法的に誤った表現にも着目してデータを活用しているが，提案されている主な手法は，検出されたデータをユーザーが活用できるほどの有用な検出は難しいと考える．格助詞の検出は誤りのバリエーションの予測ができ，予測を基に辞書への登録または機械学習を行うことができるが，文法的に誤った表現は学習者が多様なバリエーションを生み出すために予測が難しい．そのため，登録または機械学習が難しく，それらを行ったとしても有用な検出データは得られないと考える．

1．1．4　データの活用を考慮した検出と多種多様な誤用文の検出

機械を利用した誤用文の検出は客観性があり揺れのない検出が期待できるが，ユーザーが誤用に関する研究にデータを活用し，それに考慮した検出は行われていない．検出されたデータが，その後主に研究に活用されることを考慮すると，誤用文検出の方式はユーザーに対する配慮が必要であると考える．

また，提案された主な手法は，格助詞の誤用検出が対象とされていたが，文法的に誤った表現など他の異なるタイプの誤用検出が行えることが望まれる．そして，他のタイプを検出の対象とした際にも，多様なバリエーションの誤用が検出されなければ効率的ではない．これらの課題を解消する手法に文法的な属性を組み換え可能な検索を行い，文法的な属性を集合として検出する方式が考えられる．

■ ユーザーに対して誤用文を提示することが必要

1.1.3 項で述べたように，提案されている主な検出手法の処理は，システムが誤りを見つけると，それを訂正していた．これは，ユーザーに対して，検出した誤用文を取り出して提示する方式ではない．誤りを取り出して提示することは，ユーザーに対する重要な処理であると考える．1.1.3 項で述べたが，誤用の研究において，ユーザーは任意に定めた基準で誤用を確認して研究を進めるからである．そのため，最終的な判別はユーザーの目に委ねるべきである．その最終的な判別を支援するために，検出した文を取り出して提示する方式は重要であるとともに必要な処理であると考える．

■ 集合による多様な誤用の検出と異なる誤りタイプの検出

文法的な表現の誤りのように多様なバリエーションが存在し，作文などでどのように書かれれいるか予測が難しい誤用には，特定の誤りのタイプを集合として検出する手法が考えられる．特定の誤りのタイプを集合とすることで，多様なバリエーションの誤用が検出できるからである．たとえば，1.1.3 項の文献に示されていた「難しいでした」の誤りを形容詞過去の表現誤りは，その集合には「難しいでした」の他に「楽しいでした」や「いいでした」などが存在し，その集合を対象として検出することで多様なバリエーションの誤りが検出できる．また，誤りの集合を検出する手法には，品詞や活用形などの文法的な要素を検索する手法が考えられる．品詞や活用形などの文法的な要素は，単語の抽象化した要素であるため，これらの要素を検索でマッチさせることで集合を得ることができる．さらに，集合を得ることで，多様な誤用の文を効率的に収集できる．

加えて，誤用の検索において，文法的な属性を柔軟に組み合わせて検索キーワードにすれば，誤りのタイプが異なる誤用文が検出できる．1.1.3 項の「難しいでした」と「必要の注意を怠る」における「必要の注意」のそれぞれの品詞および活用形を例にすれば，「難しいでした」の検索においては「形容詞・基本形＋助動詞＋助動詞」のように検索が行え，「必要の注意」においては「形容動詞＋助詞＋名詞」のようにキーワードが異なる検索を行えば，検出される誤用文も異なりる．つまり，誤りのタイプが異なる誤用文が検出できる．

1．2　本研究における提案

多様なバリエーションの誤りに対しては，集合として検出する手法が考えられ，集合を

検出する手法には，品詞や活用形などの文法的な要素を検索する手法を考えた．また，文法的な要素を柔軟に組み合わせて検索キーワードにすることで，誤りのタイプが異なる文を検出することが可能になると考えた．このような考えに基づき，本研究では，日本語学習者テキストコーパスから文法的な誤用文を検出する手法を提案する．

本研究における提案は，日本語学習者のテキストコーパスの文に形態素情報を付加して構造化テキストを作成し，それに基づいて誤用文を検出するように設定した検索式を用いて検索を行い，マッチした文を取り出す手法である．

1．2．1　形態素解析器を利用した誤用文の検出と文の構造化

本研究では，形態素解析器から出力された形態素情報を検索して誤用文を検出することを考えている．先の 1.1.3 項の文献には，学習者が誤りやすい表現の例に「難しいでした」が示されていた．この誤用文は「形容詞自立語・基本形」に「でした」が接続している．ここで，「難しいでした」の文字列を使わずに，「形容詞自立語・基本形」と「でした」の組み合わせで日本語学習者コーパスに検索をかけると，「難しいでした」以外に「悲しいでした」や「楽しいでした」などの多様な文が検出される．これは，単語の文字列ではなく形態素情報を検索するため，検索のキーワードにした形態素情報の集合にマッチするからである．このように，単語がわからなくても形態素情報を用いれば，誤用文が検出できる．この手法により，日本語学習者のテキストコーパスの文を一つずつ目で見て誤用を確認しなくてもよく，また誤用のパターンが一つ見つかれば多様な誤用文が検出できる．

この手法を用いるには，文に形態素情報を埋め込んで構造化し，構造化した文に基づく検索式を用いて検索を行う．文には形態素情報が埋め込まれていることから検索で形態素情報にマッチさせることができ，検索式は構造化した文に基づくことで形態素情報を埋め込んだ文に対応することができる．

■　形態素情報を用いた誤用文検出の実験

ここでは，「難しいでした」の文の形態素情報を例にして，誤用文を検出する実験を行う．まず，形態素解析器に誤用文を入力すると形態素情報を付けるかどうか確認する．次に，構造化した文に形態素情報の検索式を用いて誤用文を検出する．

- **形態素解析器への**日本語学習者コーパスの入力

形態素解析器の茶筌[38]と ipadic[39]の組み合わせに，日本語学習者のテキストコーパス[22](135 作文, 2,185 文)を入力した. 入力したテキストコーパスを目視で確認したところ，文中には誤用が含まれていた．形態素解析器は誤用が含まれた文に対しても形態素解析を行い，単語には単語の表層に関する属性の文字列(表層)・読み方・基本形と，文法的属性の品詞・活用型・活用形の情報を付けていた.

- **属性を組み合わた検索と文の検出**

形態素解析されたテキストコーパスに対して，文法的属性の品詞と活用形を選択して検索をかけた．着目したのは自立語の形容詞とその基本形である．この二つの文法的属性を選択して検索したところ，707 件の文が検出された．検出された文を目視で確認すると，誤用文も含まれていた.

- **検出された誤用文の形態素情報を確認**

検出された文のなかに「ケーキは今まで一番大きいでした。」という誤用文があった．この誤用文の「大きいでいた」の箇所の形態素情報を確認した．確認した形態素情報を表 1.1 に示す.

表 1．1 「大きいでした」の形態素情報

単語	文字列(表層)	読み方	基本形	品詞	活用型	活用形
大きい	大きい	オオキイ	大きい	形容詞-自立	形容詞・イ段	基本形
でし	でし	デシ	です	助動詞	特殊・デス	連用形
た	た	タ	た	助動詞	特殊・タ	基本形

(□ は検索で選択した属性を表す)

表 1.1 の四角は, 検索で選択した品詞と活用型の形態素情報が付けられていたことを表す. これらの形態素情報とマッチしたため，この誤用文が検出された.

検出された文は，人が見て「大きいでした」の箇所に誤りがあり，誤用文であると判別できる．これは，「大きい」と「でし」および「た」の接続法に誤りがあるからである．これら接続法に誤りがある誤用に対して，形態素解析器は辞書の見出しに載っている文法的属性を付けた．つまり，形態素解析器は，人が見て誤用と判別される文であっても単語の

識別を行い，辞書に載っている文法的属性を付ける．

また，入力した文は 2,185 文であったが，検出されたのは 707 件であった．これは品詞と活用形を選択および組み合わせて検索したことで，検出された文が絞り込まれた．仮に，品詞だけを選択して検索しても元の文数よりも絞り込まれるであろうが， 707 件よりも多くの文が検出されることと予想する．選択の数を複数にし，複数を組み合わせて検索したことで，さらに検出される文の絞り込まれた．

- **形態素情報を用いた誤用文の検索**

1.1.3 項には，学習者や誤りやすい形容詞過去の表現の例として「難しいでした」があげられていた．また，表 1.1 の「でし」と「た」の単語に分割されいる．そこで，自立語の形容詞・基本形に二つの文字列「でし」と「た」を加えて検索を行った(「形容詞-自立・基本形＋でし＋た」で検索)．その結果，24 件が検出された．うち，目視で確認したところ，15 件が形容詞過去の表現における誤用文であった．検出された誤用文のうち，3 例を示す．

- 私たちは悲しいでした。
- P はかわいいでしたが、よく食べすぎますから、ちょっと太って見ました。
- 苦しいでしたが、楽しいでした。

検出された文は，それぞれ「悲しいでした」,「かわいいでした」「苦しいでした」,「楽しいでした」の箇所に誤用があると判別できる．

このように検索に形態素情報を用いることで，誤用文を検出することができ，誤用のパターンが一つわかかれば多様な誤用文を検出することが可能である．

■ 誤用文を検出するための文の構造化と構造化文の検索式

先に述べたように，特定の文法的属性を検索るためには，文に文法的属性を埋め込んで構造化しなければならない．文に文法的属性を埋め込むことで，検索で文法的属性にマッチさせることができるとともに，多様な文を検出できる．また，検索式を構造化した文に基づいて組み合わせることで，形態素情報を埋め込んだ文に対応するとともに，様々な誤用文を検出することができる．

1. 2. 2　単語における文法的属性の抽象化レベルについて

単語の文法的属性には，単語が抽象化された情報であり，抽象化された情報には，レベルが存在する．抽象化レベルは品詞の下位に存在する詳細化の情報である．

ここでは，単語の抽象化された情報についてと，その抽象化のレベルについて述べる．また，検索において，抽象化のレベルに存在する品詞の詳細化の情報を組み合わせることで，正しい文と誤用文を分けられることを述べる．

■ 単語の抽象化された情報

文は単語によって構成される．単語は品詞や活用形などの文法的属性を持っている．それらの文法的属性は，形態素解析を行うことで得ることができる．

また，文法的属性は単語の抽象化した情報である．ここで，1.1.3 項の文「難しいでした」を例に形態素解析をして，単語の分割と，分割した単語に文法的属性の品詞・活用型・活用形を付ける．文を単語の分割をし，文法的属性を単語に付けた図 1.1 を示す．

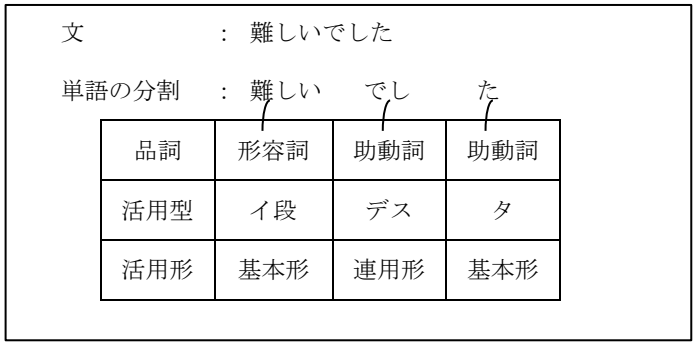

図 1. 1 単語を分割して単語に文法的属性を付けた例

図 1.1 の文「難しいでした」は，「難しい」，「でし」，「た」を，それぞれ一つの単語に分割した．また，それぞれに品詞・活用型・活用形の文法的属性を付けた．「でした」と「た」は，それぞれ単語の文字列は異なるが，品詞は同じ「助動詞」のグループに属している．また，「難しい」と「た」も単語の文字列は異なるが，活用形は同じ「基本形」のグループに属している．つまり，品詞や活用形などの文法的属性は単語の抽象化された情報である．

本研究では，これら単語の抽象化した情報を検索して集合を得る．

■ 単語における抽象化のレベル

抽象化した情報には，レベルが存在する．レベルとは，品詞の下位に存在する詳細化の情報である．抽象化のレベルを図 1.2 に示す．

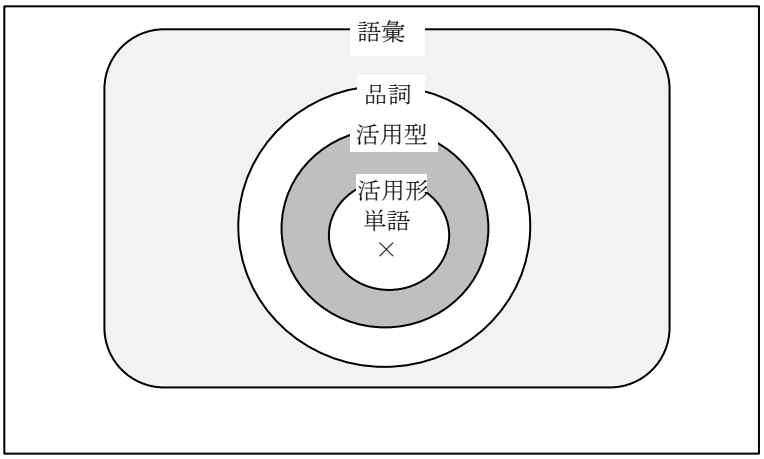

図 1．2 単語における抽象化のレベル

図 1.2 の語彙が全体集合である．ある単語は「×」があったとして，その単語は語のなかの一点として存在する．そして，品詞・活用型・活用形の品詞の情報は抽象化レベルが存在する．品詞レベルの集合の下には活用型レベルの集合があり，活用型レベルの集合の下には活用形レベルの集合がある．これらを動詞にたとえると，品詞レベルには動詞・形容詞・助動詞などの集合があり，その下の活用型レベルには一段・五段・サ変・カ変などの集合があり，その下の活用形レベルには未然形・連用形・連体形などの集合がある．そして，単語「×」が存在する．

このように，抽象化した情報には，レベルが存在している．

また，検索において，品詞は抽象化されて情報であるが，集合として大きいため，正しいものも誤ったものもこの集合に存在する．そのため，品詞の詳細化の情報である活用型と，活用型のさらに詳細化の情報である活用形を入れて検索することで，正しい文と誤用文を分ける．

■ 抽象化レベルの組み合わせによる文の検索

以上のように，品詞だけで検索した場合には，正しい文も誤用文も検出される．そこで，品詞に加えて正しい文とは違う詳細化を入れて検索することで，誤用文だけを検出する．

その例として，次の 5 つの文があるとする．これらは，1.1.3 項の文「難しいでした」を基に，筆者が作例した文である．また，これらの文の 3 と 4 は正しい文である．

- 文 1(誤用文)　: 難しいでした
- 文 2(誤用文)　: 楽しいでした
- 文 3(正しい文) : 雨でした
- 文 4(正しい文) : 難しかったです
- 文 5(誤用文)　; いいでした

文 1～5 を形態素解析して抽象化の情報をつけ，検索例とヒットする情報を表した図 1.3 を示す．

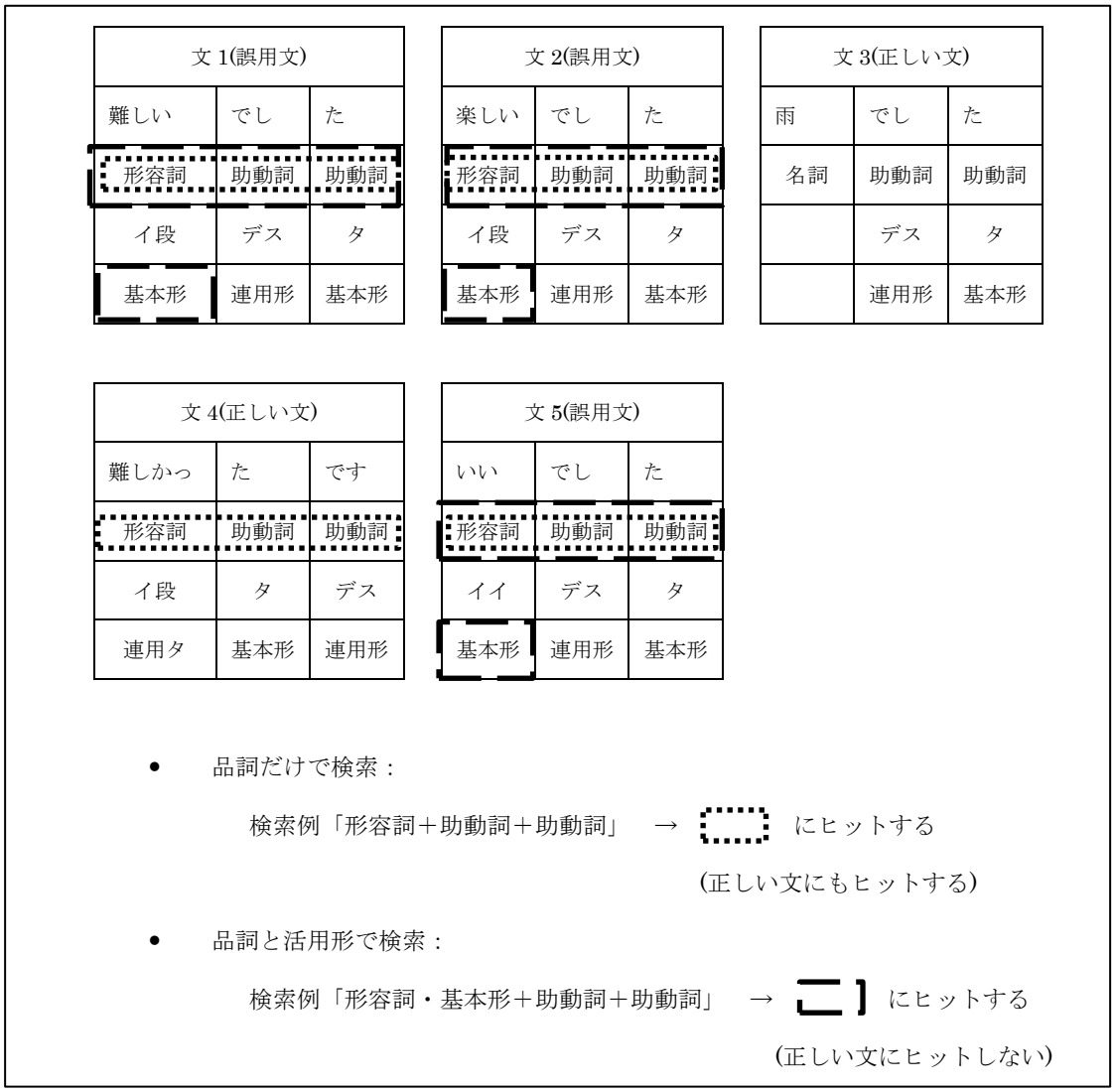

図 1. 3 品詞だけで検索した例と品詞と活用形で検索した例

図 1.3 の品詞だけの「形容詞＋助動詞＋助動詞」で検索を場合は, 正しい文にもヒットする. 品詞と活用形で「形容詞・基本形＋助動詞＋助動詞」で検索を場合は，正しい文にはヒットしない.

このように，抽象化した情報である品詞で検索すると，文字列よりも大きな集合を得ることができるが，検出に正しい文が含まれる可能性がある．そのため，抽象化の詳細を組み合わせで検索する．詳細化と組み合わせて検索することで，正しい文と誤用文を分けることができる．つまり，抽象化レベルの組み合わせを利用して，誤用文の集合を得る.

1. 2. 3　日本語学習者のテキストコーパスを構造化した文に処理

1.2.1 項で述べたように，本研究では，文に形態素情報を埋め込んで構造化する．これを日本語学習者のテキストコーパスの文に対して行う．

構造化した文に対して検索を行うことで，多様な文にマッチできるようにする．そのために，必要であると考えているのは，次の二つである．

- 文を構成する単語に文法的属性が埋め込まれていること
- 単語の文字列と文法的属性を区別するためにラベルが付いていること

文中の単語に文法的属性が埋め込まれていれば，検索で文法的属性を捜すことができる．また，ラベルが付いていれば，単語の文字列と各文法的属性を識別でき，さらに単語の境界も識別できる．

本研究では，以上のような考えに基づき，形態素情報が付いた文単位でテキストを作成する．これを構造化文と呼ぶ．さらに，テキストコーパスのすべての文を構造化文したものを構造化ファイルと呼ぶ．

ここでは，構造化テキストについての簡単な説明および作成の流れを述べる．詳細については 2 章で述べる．

さらに，本研究で使用する形態素解析器・辞書，日本語学習者コーパスと，形態素解析器を使用する環境について述べる．

■ 形態素解析器を用いた単語情報付加と形態素情報について

1.2.1 項で述べたように，本研究では，形態素解析器を利用して単語に情報を付ける．

形態素解析器は文を入力すると単語に分割する．分割された単語には，単語の表層に関する属性の文字列(表層)・読み方・基本形などと，文法的属性の品詞・活用型・活用形などと，そして文の終わりには文末記号が付けられて出力される．この単語に関する情報と文法的属性を付加していく処理を，電子化テキストの日本語学習者コーパスの文に対して行う．

また，本研究では，単語に付加された読み方などの単語の表層に関する属性と文法的属性のうち，誤用文の検出に重要であると考えるものを選択する．選択するのは，単語の文字列，品詞，活用型，活用形である．本研究では，選択したこれらの属性の一つまたは複

数を形態素情報と呼ぶ.

■ 構造化テキストの概要と作成の流れ

形態素解析器から出力された単語の属性に選択処理を行い，選択した単語の属性に本研究で定義した独自のラベルを付加して形態素情報付き単語の文にしたものが構造化文である.

選択した属性である文字列，品詞，活用型，活用形だけを単語に付加する処理を，分割された全ての単語に対して行う．そして，それぞれの属性には，属性ごとに異なるラベルを付けていく．そして，単語一つずつをラベル・形態素情報付き単語にする.

ラベル・形態素情報付き単語を文末記号を基に一文に整形する処理を行う.

以上の日本語学習者のテキストコーパスの文に対して，形態素情報の選択処理・ラベル付け処理・一文への整形処理を全ての文に施し，構造化テキストを作成する.

■ 使用する形態素解析器と辞書について

本研究では，次の形態素解析器と辞書を使用する.

- 茶筌 2.3.3[38]
- ipadic version 2.7.0[39]

茶筌は一般によく使用される形態素解析器であることから，本研究においても茶筌を利用して形態素情報を付加するをする．また，茶筌は初期設定で辞書にない単語を未知語として返す．未知語で返された単語は誤用である可能性があるため，誤用文を検出するには使い勝手のよい形態素解析器であると考える．そのため，茶筌を形態素解析器に使用する．ipadic は茶筌で使用可能な辞書であり, version 2.7.0 は 2015 年 6 月現在の最も新しいバージョンである.

■ 構造化テキストにする日本語学習者コーパスについて

本研究では，次の日本語学習者コーパスを形態素解析したのち，構造化テキストにする．

- 台湾人日本語学習者コーパス[22]
  135 作文，2,185 文：第 3 章と第 4 章で使用する．
  581 作文，9,358 文：第 5 章で使用する．
- 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[24]
  台湾人学習者の 57 作文，1,046 文：第 5 章で使用する(文中には段落や作文の終わりを表す記号などが含まれている[40])．

これらは日本語学習者の作文を電子化したテキストコーパスである．また，これらは，それぞれ一つのファイルにまとめたものである．

これらの日本語学習者コーパスを使用する理由は，筆者が[22]の開発に関わり，データを所有しているからである[41]．さらに，母語による日本語学習者数を考慮したからである．文化庁が行った調査によると，2013 年 11 月現在，国・地域別による国内の日本語学習者は，1 位が中国，6 位が台湾であった[42]．この結果は，母語別からみると，中国語を母語とする学習者が最も多いことになる．したがって，中国語を母語とする学習者コーパスを使用して誤用文を検出できたなら，母語別による日本語学習者の大多数を検出できるであろうと考えたからである．

■ 形態素解析器を使用する環境について

茶筌が Unicode を扱うことができる次の環境で形態素解析を行う．

- OS：MacOS 10.6.8

使用する[22]のコーパスは，作文中に日本語と母語(中国語)が使用されている箇所があることから，Unicode で電子化されている．そのため，形態素解析器が Unicode を扱える以上の環境で形態素解析をする．

1. 2. 4　構造化テキストを検索する表現式について

1.2 節で述べたように，本研究では，多種多様な文法的誤用文の取り出しを行う手法である．1.2.3 項では，多様な文にヒットさせるためには，単語に関する情報が付いており，また単語に関する情報にラベルを付け，それらの単語を文にして構造化することで，品詞や活用形などで検索した場合に，特定の集合にマッチして，多様な文検出が可能との考えを述べた．

構造化文に対しては，構造化の定義に基づいた検索の表現形式が必要である．また，形態素情報，一つ以上あるため，それを組み合わせることができる方式を考えている．加えて，組み合わせを変えることで，誤りのタイプが異なる文の検出を考えている．

このような考えに基づき，構造化文に基づいて作成した組み合わせ可能な表現形式の検索キーワードを作成し，構造化テキストに対して検索を行う方式をとる．そして，構造化文を検索するために作成した検索式を表現式と呼ぶ．

表現式は形態素情報を組み合わせと正規表現を利用して作成した文字列式である．正規表現とは，文字列のパターンを記述できる検索式である．構造化文に対して，形態素情報の組み合わせと正規表現を挿入した文字列式で検索し，検索にマッチした文の集合を取り出す．

形態素情報における単語と文法的属性は，自由に組み合わせて設定することができる．

本研究においては，形態素情報の組み合わせたものをユニットと呼ぶ．ユニットの数と組み合わせも自由に設定することができる．本研究の表現式設定におけるこれらの自由とは，単語については文字列と文法的属性を任意に選べるという意味を表し，ユニットについては数と組み合わせを任意に決められるという意味を表す．

このように柔軟な設定ができる表現式を用いて，多種多様な文法的誤用文の検出を行う．

ここでは，表現式の作成例をあげ，組み換えを行うことで検出される文がどのように変わるか，文の検出例について述べる．構造化文に基づき作成した表現式の詳細は 2 章で述べる．

■ 組み合わせ自由な表現式の作成例と想定通りに形態素情報を付けた文の検出例

ここでは，表現式の作成例と，それを用いて検出される文を示す．ここで示したいことは，表現式は自由に組み合わせ検索できることと，自由な組み合わせで検索することにより，検出の範囲を広げたり狭めたりすることが可能であるということである．つまり，柔

軟な検索ができることを示したい.

表現式の作成例で検出される文がどのうなものかを示すため，文には想定した通りに形態素情報が付けられたと仮定する．ここでの想定とは，本研究の環境で形態素解析器が単語に ipadic と同じように単語を分割し，文法的属性を付けたということである.

また，ここでの文は，いずれも筆者による作例である．二つの文のみ，1.1.3 項の文献に取り上げられていた誤用文「難しいでした」と「必要の注意」を参考に作例する．そして，これらの文が構造化テキスト内に作例文が存在していると仮定する．実際の構造化テキストにおける文は，形態素情報付き単語の文であるが，ここでは文をわかりやすくするため，形態素情報を付けずに示す.

文を検索する表現式の形態素情報は，簡単な形で「品詞　―　活用型　―　活用形」と示す．単語の文字列も含めた場合は，品詞の前に挿入した形で示す．ipadic の文法的属性についても，詳細を省略した簡単な形，および「形容詞の自立」のように示す．さらに，ユニットを組み合わせた例においても，「形容詞ユニット　＋　助動詞ユニット」のように示す.

さて，次の 10 の文が構造化テキストに存在するとする.

- 文 1：食べます.
- 文 2：雨らしいです.
- 文 3：雨でした.
- 文 4：難しいです.
- 文 5：おいしいです.
- 文 6：赤いです.
- 文 7：食べてほしいです.
- 文 8：難しかったです.
- 文 9：難しいでした.
- 文 10：必要の注意を怠りました.

「～ます」の文に着目し，検出を行うとする．「ます」は，品詞が「助動詞」，活用型が「マス」，活用型が「基本形」である．そのため，(1)を作成して検索すると，文 1 が検出される.

- 助動詞　—　マス　—　基本形　　(1)

ここで，(1)の「マス」を「イ段」にかえて，(2)で検索すると，文 2 が検出される．

- 助動詞　—　イ段　—　基本形　　(2)

さらに「助動詞」を「形容詞」にかえて，(3) で検索すると，文 4，5，7，9 が検出される．

- 形容詞　—　イ段　—　基本形　　(3)

ここで，片仮名の文字集合を抽象化した正規表現「[ァ-ヶ]」，それが 1 回以上を表す「+」を活用型に設定して(4)を作成して検索する．

- 形容詞　—　[ァ-ヶ]+段　—　基本形　　(4)

すると，(3)の検出結果(活用型が「イ段」)の形容詞文と，形容詞「アウオ段」の文 6 が検出される．

また，ipadic 品詞と活用型は階層構造が存在するものがある．ここでは，形容詞における品詞の下位階層を例にする．形容詞は下位階層に「自立」,「非自立」などがある．これを(4)の形態素情報の組み合わせに正規表現の *OR* を表す「|」と，範囲を表す「()」とする．さらに, 活用形に, Unicode で日本語・中国語・韓国語で使用される漢字の正規表現「[㐀-䶵]」[43]と，片仮名の正規表現を挿入し，それが 1 回以上の「+」を設定した(5)を作成する．

- 形容詞の(自立|非自立)　—　[ァ-ヶ]+段　—　[㐀-䶵ァ-ヶ]+形　　(5)

(5)で検索すると，(4)で検出された文に加えて，文 8 が検出される．

また，形容詞の自立語が基本形で使用されている文のなかでも，「～しい」という文字列に平仮名だけが使用されいる文を対象に検出する．この場合，品詞の前に文字列を挿入して設定する．また，ここでは，二つの表現式を作成する．平仮名の正規表現「[ぁ-ん]」が 1

回以上と，文字列「～しい」を設定した(6)，または平仮名「[ぁ-ん]」が一つ以上の正規用言を設定した(7)を作成する．

- ［ぁ-ん］+しい　―　形容詞の自立　―　イ段　―　基本形　　(6)
- ［ぁ-ん］+　―　形容詞の自立　―　イ段　―　基本形　　(7)

(6)または(7)で検索すると，文 5 が検出される．

次に，形態素情報の組み合わせを並べたユニットの作成例を示す．

助動詞が並んだ(8)の組み合わせを作成して検出する．すると，文 2，3，8～10 が検出される．

- 「助動詞のユニット」＋「助動詞のユニット」　　(8)

ここで，「でした」の表現に着目して，(9)を作成して検索する．

- 「“でし”の助動詞のユニット」＋「“た”の助動詞のユニット」　　(9)

すると，文 3，9 が検出される．

また，形容詞に助詞が接続した文に着目して，(10)を作成する．

- 「形容詞のユニット」＋「助動詞のユニット」　　(10)

(10)で検索すると，文 4～9 が検出される．

ここで，(10)に助動詞のユニットをさらに増やして，(11)を作成する．

- 「形容詞のユニット」＋「助動詞のユニット」　＋「助動詞のユニット」　(11)

(11)で検索すると，文 8，9 が検出される．

さらに，名詞以外で，助詞が並んだ文に着目して，検出する．正規表現の *NOT* である「(?!)」

を設定し，(12)を作成して検索する．

- (?!「名詞のユニット」)＋「助動詞のユニット」 ＋「助動詞のユニット」 (12)

(12)で検索すると，文 8～10 が検出される．
ここで，形容詞で「でした」と表現した文に着目して，(13)を作成して検索する．

- 「形容詞のユニット」＋「“でし”の助動詞のユニット」
  ＋「“た”助動詞のユニット」(13)

すると，文 8 が検出される．
さらに，形容動詞と「の」で名詞を修飾した表現に着目して，(14)を作成し，検索する．

- 「形容動詞のユニット」＋「助詞のユニット」＋「名詞のユニット」

すると，文 10 が検出される．

このように表現式は自由な組み合わせで検索することができる．自由な組み合わせによる検索は，検出の範囲を広げたり狭めたりすることが可能になる．つまり，柔軟な検出を行うことができる．

1．2．5　本提案手法の命名

以上の本研究における提案手法を形態素情報付き構造化表現式検出法と名付ける(以下，短く表現式検出法と呼ぶ)．

1．2．6　表現式検出法の環境について

表現式検出法における構造化テキスト作成と文検出の実験は，次の環境で行う．

- OS：Windows 7
- プログラミング言語：Strawberry Perl 5.14.2.1[44]
- IDE：Padre([44]が組み込まれている)[45]

形態素解析を行った 1.2.3 項のテキストコーパスは，以上の環境で構造化テキストを作成するとともに，3 章～5 章で検出実験を行う．

1．2．7　表現式検出法の有用性と誤用研究に貢献の可能性

ここでは，表現式検出法の有用性と，表現式検出法を用いた検出により，誤用に関する研究に貢献できると考える，その可能性について述べる．

なお，表現式検出法の簡便性については，2.4 節で述べる．

■ 形態素解析器が正しく形態素情報を付けた誤用文に有用

表現式検出法による誤用文検出は，形態素解析器が正しく形態素情報を付けた誤用文を有用に検出することができると考えている．これは，単語の文法的属性は正しいが，品詞などの文法的属性の接続法は日本語として正しくない文法的な誤用文である．形態素解析器は，綴りを誤った単語は正しく形態素解析を行わないが，綴りを誤っていない単語に対しては，正しく単語に分割して形態素情報を与えるからである．

また，文法的属性の接続法が日本語として正しくない文法的な誤用文は，1.1.3 項の誤用文「難しいでした」と「必要の注意」以外にも，日本語学習者の誤用辞典[46]や日本語の指導に関する文献[37，47，48]，さらに誤用検出に関する雑誌[34]をみると，このような誤用文は少なくない．たとえば，以下のような誤りがあった．

- 伝聞表現の「そうだ」に関する誤り：
  「どくしんそうです」，「どくしんなそうです」，「忘れにくいだそう」[46]
- 形容詞や動詞において「て」で活用させた表現に関する誤り：
  「高かっても買います」，「日本語を話してはおもしろいです」[46]
- 形容詞の名詞修飾表現に関する誤り：
  「あの背が高いの人」[47]
- 助動詞「な」と共起した表現に関する誤り：
  「善意な人々」[37]，「なんで日本語はこんなに難しいなの？」[34]
- 単語の使い方に関する誤り：
  「用事ぶかい」[48]

これらの誤用文は，品詞の接続法が日本語として正しくない．そのため，これらの誤用文に対しても，表現式検出法は有用な検出が期待できると考えている．

したがって，表現式検出法は，品詞の接続法に誤りがある誤用文おいて，有用に用いることができ，後述する効率的な支援が行えると考えている．

■ 少ない誤用文例から多様な誤用文の検出が可能

1.2.1 項で示したように，1 例からでも多様な誤用文が検出できる．つまり，数少ない誤用例から多様な誤用文を検出できる．

文献や辞書などには，誤用文が数例しか取り上げられていない．また，日本語を教えた経験から想定できる誤用文についても，数例であるのが一般的であろう．表現式検出法は，このような少ない誤用例から，多様な誤用文を検出することができる．これは，検索に形態素情報を利用しているからである．一般的な文字列検索は，文字列のみを対象としているが，表現式検出法では文法的属性に対して検索を行っている．文法的属性は単語の抽象化された要素であるとともに，単語がどの集合に属しているかを示すものである．つまり，文法的属性の検索は，単語における集合の検索を行っているのである．そのため，少ない例で検索しても，多くの文にマッチさせることができ，多様な文を取り出すことができる．

この少ない例からでも多様な誤用文を検出できることは，2.4 節の簡便性においても述べる．

■ 予想外の誤用文発見の可能性および誤用研究発展への可能性

表現式検出法は，多様な誤用文が検出されることから，ユーザーが想定していなかった誤用文が検出される可能性がある．この可能性は，日本語教育におけて，誤用に関する研究が発展する可能性を秘めている．作文などにおいて，学習者がなぜそのように誤って書いたかという原因は，知り尽くされていないのが現状であろう．表現式検出法は，予想をしていなかった誤りが発見される可能性があるため，その誤用に対する分析によって，新たな誤用の原因を究明できるかもしれない．また，予想をしていなかった誤用は，研究に対する新たな角度からの分析を提供する可能性もある．このように，表現式検出法を用いた誤用文の検出は，誤用研究発展への可能性が期待できる．

■ 異なる誤りタイプの誤用文を検出することが可能

表現式は作り方次第で，誤りのタイプが異なる誤用文を検出することができる．1.2.3 項で示したように，形態素情報は自由に組み合わせて設定でき，ユニットの数と組み合わせを自由に設定できるという柔軟性を持っている．柔軟であるということは，表現式の作り方次第で，タイプが異なる誤用文を検出する式を作ることができるということになる．したがって，1 種類の誤りタイプの検出に限られたものではなく，様々な種類の誤用文を検出することができる．

■ 機械が誤りの種別を自動で選別している可能性と誤用タイプ選別への貢献

表現式は，様々な種類の誤用文を検出するように作成でき，種類の異なる誤用文を検出するということは，機械が誤りの種別を自動で選別している可能性がある．たとえば，特定の誤用を意図せず作成した表現式が誤用文を検出し，検出した誤用文があるタイプの誤用だと判別できたなら，それは機械によって誤りが選別されたと捉えることができる．公開されている学習者コーパスは，コーパスごとに誤りの種別が異なる[11，29]が，表現式検出法で検出した誤用文が誤用タイプ選別に貢献できる可能性がある．

■ 誤用文だけではなく正しい文を検出することも可能

柔軟に表現式が作れるということは，正しい文を検出する式を作ることもできるということを意味する．正しい文についても，タイプの異なる文を検出する式を作成ことができ，作成して検索をすれば，多様な正しい文が検出される．

■ 表現式検出法は文を取り出して目視で確認することが可能

1.2 節で述べたように，表現式検出法は，ユーザーに提示する形での取り出しを行う．これは，1.1.4 項で述べたように，誤用文の最終的な判別はユーザーに委ねるべきだと考えるからである．その理由についても，1.1.4 項で述べたように，ユーザーごとにどのデータをどのように活用するかは異なるからである．そのため，表現式検出法の最終処理は取り出しを行うことにしている．そして，ユーザーは取り出された誤用文を目視で確認することができる．

■ 表現式検出法で取り出した誤用文は誤用分析データに活用が可能

形態素解析器を利用する表現式検出法は，先にものべたように，綴りの誤りは正しく形態素解析されないことから，綴り誤りについて検出は難しいと考える．また，形態素解析器は文単位で形態素解析をすることから，一文以上の関係における誤りの検出も難しいと考える．加えて，意味の誤りに関する検出についても難しいと考える．しかし，先に示したように，品詞の接続法の誤りの検出は可能であり，そのような誤りは誤用辞書や文献などに取り上げられていた．これらのことから，表現式検出法で検出した誤用文データを誤用に関する研究に活用することは可能である．また，仮に，検出されたデータが誤用の分析などに活用できるほどの件数でなかったとしても，使用する学習者コーパスを増やすなどの対応をとることができる．

この表現式検出法で検出したデータを誤用研究に活用し，その有効性を確認することは，5 章で改めて検討する．

■ 他の形態素解析器と辞書でも表現式検出法は可能であると予測

本研究では，形態素解析器に茶筌と辞書に ipadic を利用したが，他の形態素解析器と辞書を利用しても表現式検出法を用いることは可能であると予測する．辞書については，文法的属性と品詞などの名称が ipadic と異なっていることが予想されるが，それらを把握さえすれば，表現式検出法を用いることが可能であると予測する．

■ 表現式検出法はツールでも利用でき，他のユーザーでも使用可能

本研究ではプログラミング言語で表現式検出法を用いるが，表現式検出法はツールを用いて使用することが可能である．たとえば，ChaKi[49]は，形態素解析器 MeCab[50]と辞書が組み込まれているとともに，検出において正規表現の利用も可能である．また，KH Coder[51]は，茶筌と辞書が組み込まれており，単語の文字列・品詞・活用形の検索ができる．したがって，ユーザーが正規表現を正しく使え，辞書の文法的属性を把握することさえできれば，ツール上においても表現式検出法を用いることができる．

ツールを利用した表現式利用法は，2.4 節の簡便性においても述べる．

■ 誤用研究を効率化：誤用例から多様な誤用文が取り出せて同じ環境なら再利用も可能

先に述べたように，表現式検出法は，少ない例からも誤用を取り出しが可能なことから，

誤用文の取り出しにおいて，研究を効率化することができる．

また，表現式は，構造化テキストを作成した環境と同じであれば，再利用することができる．たとえば，コーパス A を構造化テキストにし，表現式 A を用いて誤用文を取り出したとする．そして，コーパス B も，コーパス A と同じ環境で構造化テキストにすれば，表現式 A を用いて誤用文を検出することができる．つまり，同じ環境であれば，異なるコーパスでも表現式が利用でき，効率的に誤用文を取り出すことができる．

さらに，同じ環境であれば，ユーザー同士が表現式を持ち寄ちよることで，効率的に誤用文の取り出すことが期待できる．

表現式検出法を用いた研究の効率化については，2.4 節の簡便性においても述べる．

■ 教師と研究者であれば，直感的に扱えると予測し，試行も繰り返すことが可能

1.1 節で述べたように，教師や研究者は，学習者に日本語を教えた経験から，誤用に関する知識を持っている．経験と知識があるということは，辞書と正規表現を理解していれば，直感的な表現式作成を行うことができると考える．以下のユーザーが行うであろう手順で説明する．

- 手順 1：特定の誤用に着目する．
- 手順 2：構造化テキストに処理する(または手順 5 において行う)．
- 手順 3：着目した誤用を形態素解析器に入力する．
- 手順 4：形態素解析器から返された情報の観察および選定を行う
- 手順 5：表現式を作成する(ここで手順 2 を行う場合もある)．
- 手順 6：検出実験を行う．
- 手順 7：表現式を改良する．
- 手順 8：改良した表現式で検出実験を行う．(最良と考える検出結果が得られた場合は終了)
- 手順 9：手順 7，8 の繰り返し．
- 手順 10：検出実験を終了する．

手順 1 は，ユーザーがどのような誤用の種別を対象にするか，誤用の種別の選定を表す．

手順 2 は，テキストコーパスを構造化テキストに処理を行うことを表す．この処理は手

順 4 で行われる場合もある．

手順 3 は，着目した誤用の少ない例を形態素解析器に入力することを表す．ここで，誤用を形態素解析に入力することで，手順 4 で表現式作成のための概要を知ることができる．

手順 4 は，形態素解析器に入力した誤用は，どのような形態素情報が付けられて返されたか観察し，その情報を把握することを表す．観察は表現式を作成することを想定しながら行っている．選定は，表現式で用いる形態素情報およびユニットを選んでいる．

手順 5 は，手順 4 で選定した形態素情報およびユニットを用いた表現式を作成する．また，構造化テキストは，ここで作成を行う場合もある．

手順 6 は，構造化テキストに対し，手順 5 で作成した表現式を使った検索を行う．この際にも，目視による確認が行われる．

手順 7 は，検出された文の多寡，または目視で確認して，想定通りの検出結果が得られなかった場合に，表現式の改良を行う．ここでユーザーが最良と考える検出結果が得られた場合は，実験を終了する．

手順 8 は，改良した表現式を用いて，再度検出実験を行う．この際にも，文の多寡の判断と目視による確認が行われる．

手順 9 は，ユーザーが最良と考える検出結果が得られるまで，手順 7 と 8 を繰り返す．

手順 10 は，ユーザーが最良と考える検出結果が得られたので，検出実験を終了する．

以上の想定される手順 7，9 において，ユーザーは経験による直感を用いると考える．たとえば，「この誤用の後ろには，単語 A が接続して間違えやすい，その単語の品詞や活用形は・・・」など，経験を基にした直感を働かせて表現式の改良と試行を行うであろう．この教師や研究者の直感を用いて誤用文を検出できるという点が，表現式検出法における有用性の一つであると考える．さらに，改良が行えるのは，表現式が検出の範囲を広げたり狭めたりすることができるからである．

また，試行は直感を利用して行う組み換え作業(形態素情報およびユニットの組み換え)だけであるから，比較的簡単にできるのはないかと考える．

手順 3，4，8，9 については，2.3.2 項および 2.4 節でも述べる．

1. 3　関連研究

本研究では，形態素解析器を利用する．関連研究において，まず形態素解析器の処理について述べる．次に，ワープロソフトの文法チェックについて述べる．そして，日本語学習者の誤用文を対象とした自動検出の研究について述べる．

1. 3. 1　形態素解析と形態素解析器について

ここでは，形態素解析の処理過程と，形態素解析器に日本語学習者の文を入力するとどのように処理されたかを述べる．

■ 形態素解析とは

形態素解析とは，文を単語単位である形態素に分割して，各単語の品詞などの統語的役割を決定する処理である[52]．通常，形態素解析は次の手順で行われる[53]．

- 入力された文に対し，文中の各位置から始まる可能性のある形態素をすべて検索
- 形態素の候補を列挙したグラフ構造(ラティス構造)を作成
- 形態素同士の組み合わせのなかから，文として最も確からしい形態素の並びを決定

たとえば，「掘り炬燵になっている」という文がある[53]．この文を入力すると，図 1.1 に示すラティスが作られ，最終的に太字で記されている組み合わせに決定される．

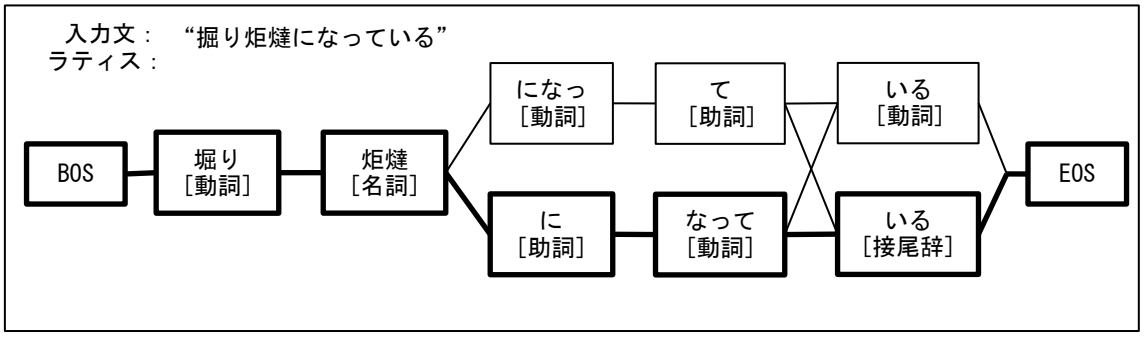

図 1.4 JUMAN によるラティス構造(笹野ら[53]から)

■ 形態素解析器とは

形態素解析器とは，入力文を単語単位に分割して品詞を付与するツールである[38]．一般

的な形態素解析器は，ルールにしたがって単語を分割し，辞書を参照して単語の読み方や品詞に関する情報を与える．たとえば，[22]の学習者コーパスに「私は今 W 大学の学生です。」という文があった．この文を茶筌と ipadic の組み合わせでに入力すると，図 1.2 のように返された．

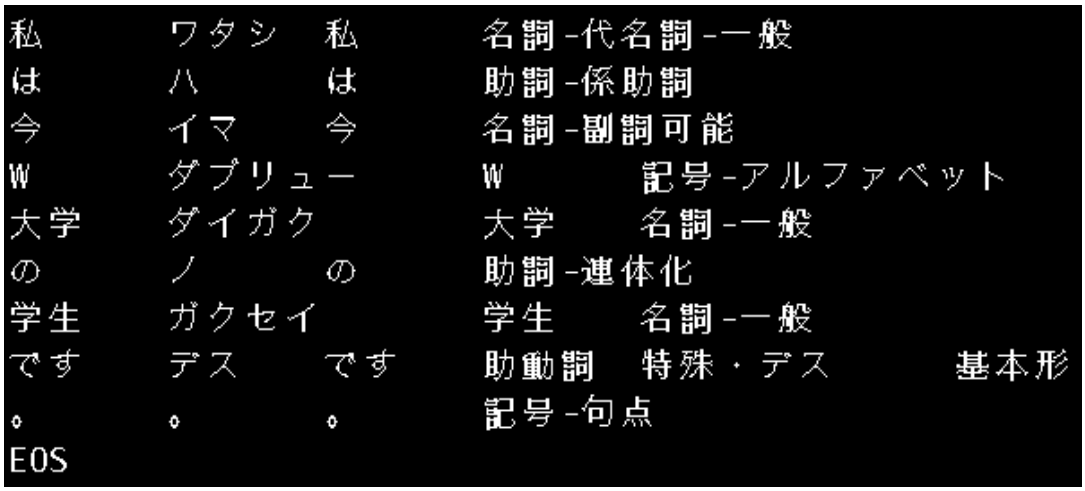

図 1. 5 茶筌と ipadic で返された入力文

図 1.2 は「私は今 W 大学の学生です。」が形態素解析された文である．入力した文の単語が縦に連なった状態で出力されている．

入力した文は単語に分割され，それぞれの単語に情報が付けられる．また，それぞれの情報はタブで区切られている．タブで区切られた単語の情報を表 1.1 に示す．最後の「EOS」は，文末を表す記号である ．

表 1. 2 タブで区切られた単語の情報

文字列	読み方	基本形	品詞	活用型	活用形
私	ワタシ	私	名詞-代名詞-一般		
は	ハ	は	助詞-係助詞		
今	イマ	今	名詞-副詞可能		
W	ダブリュー	W	記号-アルファベット		
大学	ダイガク	大学	名詞-一般		
の	ノ	の	助詞-連体化		
学生	ガクセイ	学生	名詞-一般		
です	デス	です	助動詞	特殊・デス	基本形
。	。	。	記号-句点		
EOS					

表 1.1 に示した単語の情報には，単語の文字列・読み方・単語の基本形と，単語の品詞・活用型・活用形がある．各情報間には，タブが挿入されている．活用を持たない単語(「です」など)には，活用型と活用形の情報が付けられる．活用を持たない単語(「私」など)と記号(「W」や句読点など)には，活用型と活用形にタブのみが付けられて返されている．

1. 3. 2　日本語学習者の誤用文自動検出について

ここでは，まずワープロソフトの文法チェックと，文法チェックに日本語学習者の誤用文を入力しても有用に機能し，取り出しを行えるか述べる．また，日本語学習者の誤用文を対象とした自動検出の研究に述べ，多種多様な誤用文の検出と取り出しが行えるか，さらに検出された誤用文の研究への活用について述べる．

■　ワープロソフトのスペルチェックは誤用文検出に有用に機能して取り出しもできるか

近年のワープロソフトにはチェック機能が内蔵されている．たとえば，Microsoft 社の Word は自動文章校正機能を内蔵し，スペルミス，入力ミス，入力した語の揺れ，文法的な誤りを修正候補に指摘する．

ワープロソフトは簡単な文法チェックができるが，人が見て違和感を覚える文を入力，

または貼り付けてもチェックされないことが多い．たとえば，1.1.3 項で取り上げられていた誤用文「難しいでした」を Microsoft Office 2013 Word に入力と貼り付けをしても，誤りと指摘されなかった．そのため，ワープロソフトを使って学習者の誤用文の有用な検出は期待できない．

また，ワープロソフトは誤りのチェックと指摘は行うが，誤りの取り出しは行わないため，学習者の誤用文を取り出すことができない．

■ 格フレーム辞書照合の助詞誤り検出は多種多様な誤用文を検出して取り出すか

動詞などの用言(活用する語)は，単語によって接続する格助詞が異なる．その性質を利用して，動詞の各単語と接続する正しい格助詞の組み合わせと照合し，誤りを検出する手法が提案されている．これらの提案手法では，格フレーム辞書が利用されている．格フレーム辞書とは，用言に関係する名詞とその格(格助詞)を整理したものである．

今枝ら[54]は，ルールに基づいた処理に加え，その処理において検出・訂正ができなかった格助詞に対し，格フレーム辞書と照合している．ルールに基づいた処理とは，茶筌を用いた形態素解析器の出力結果と品詞のルールを利用して誤りを検出している．その具体的例として，以下が紹介されている．入力文「人形を，お子さんあげる.」を形態素解析すると，「お子さんあげる」の箇所は，「お子さん(名詞-一般 2)」，「あげる(動詞-自立 47)」が返される．品詞の数字は数字番号を表す．ここで，人手でルール化した「用言(46-53)の直前に名詞(1-40)がある時はその間に助詞が省略されている」を適用する．このルールにより，「「お子さん」と「あげる」の間に助詞が脱落している」という誤りを検出する．そして，NTT 日本語語彙体系の名詞の意味属性と用言の取り得る助詞とを照合して，誤りを検出する．その具体的例の紹介として，入力文「冬は，雪を降ります.」の名詞「冬」と「雪」の意味体系を参照と，動詞「降る」の格フレームの照合処理を行う．参照と照合の結果，各名詞と動詞が取り得る格助詞の関係から，「冬は，雪が降ります」の訂正候補文を得る．

三浦ら[55]は，入力文を構文解析したのち，格フレーム辞書と照合して，格助詞の誤りを検出・訂正している．この手法では，入力文を構文解析して係り受け関係と品詞情報を得ている．例として「家の居間で木をあります。」があげられており，この手法では「ある」という動詞に対して，名詞「居間」と「木」が格フレーム辞書の検索対象になっている．検索から意味属性と名詞の取り得る格 ，係り受け，頻度のデータが付与される．頻度情報から格の修正候補を選択し，修正文「家の居間に木があります。」を出力する．

これら格フレーム辞書を利用した検出は，格助詞については多様な誤りを検出するが，格助詞以外の誤用文を検出することができない．これは，ユーザーにとって制約が大きいと考える．先にも述べたように， ユーザーは多種多様な誤用文に着目していることから，格助詞とその他の誤用文を検出する手法が望まれる．さらに，辞書と照合する手法は，辞書に存在する誤りは正しく検出できるが，辞書にない誤りは検出が困難である．また，これらの研究は検出後の誤りを正しく直しているが，誤りをユーザーに提示，つまり取り出す処理ではないことが本研究と異なる．

■ 動詞辞書照合の助詞誤り検出は多種多様な誤用文を検出して取り出すか

橋本ら[56] によると, 動詞には選択制限規則が存在し, 動詞と親和性のよい名詞があり，それに接続する助詞にも制限があるということである．これは橋本らも述べているように，意味解析における格フレームの概念と同様であるが，橋本らは動詞辞書を再構築した点が異なる．そして，助詞誤りの判断基準として動詞の選択制限規則を用いた．

橋本らは，日本語計算機用辞書 IPAL から新たな動詞辞書を構築し，助詞の誤りを検出している．この手法で構築された新たな動詞辞書は，見出しに動詞漢字，動詞の読み，語幹の読み，それぞれの動詞に親和性のある助詞と名詞の基本文型，接続可能な助詞，親和性のある名詞の意味素性(名詞に対する制限)を付加している．この新しく構築した動詞辞書と文を照合して，助詞の誤りを検出する．

橋本らの手法においても，助詞の誤りを検出対象としているが，格フレーム辞書で述べたことと同様に，制約が大きいと考える．つまり，助詞以外の誤用文は検出できない．また，格フレーム辞書と同様に，辞書にない誤りは検出が困難である．さらに，今枝ら[54]が指摘しているように，IPAL の辞書は，登録語数が重要基本単語のみで少ないこと，対処としている助詞が動詞の直前の助詞であることから，それ以外の助詞の誤りには対応できない．

■ 文節内特徴による助詞誤りの検出・校正は多種多様な誤用文を検出して取り出すか

南保ら[57]は，文節内の特徴を用いて，帰納的学習を行って助詞の誤りの検出と自動校正をする手法を提案している．南保らは，構文解析によって得られた特徴から特徴スロットを抽出し，特徴スロットと助詞を組み合わせてルールとしている．特徴スロットは，三つのカテゴリーから構成されている．「私はみかんもりんごも好きです.」の例において，特

徴スロットの対象文節を「みかんも」としたとき，

- 対象文節カテゴリー：
  助詞を除いて最後に来る語(「みかんも」の「みかん」)に関する特徴で，この語の品詞情報・文節位置・係り受けの種類などの情報.
- 係り受け先文節カテゴリー：
  対象の文節が係る先の文節に関する特徴(文節内に最初に来る語「りんご」)で，この語の品詞情報・文節位置・係り受けの種類などの情報.
- 近傍文節カテゴリー：
  対象の文節からみて前後一文節ずつの助詞(前文節「は」と次文節「も」)

となる．さらに，帰納的学習(学習によって獲得したルールを用いて帰納的に実例中に内在しているルールを獲得し，ルール辞書の更新を行う[39])を用いて得られた助詞選択ルール辞書を用いて誤りの検出・校正を行っている．この手法による実験の結果，格助詞以外の助詞に対しても誤り検出と校正ができることを示した.

南保らの手法は格助詞だけではなく，それ以外の助詞も検出を行い，助詞の誤りの検出と校正の更なる性能向上を目指している．その点において，多様な助詞誤り検出と校正は高い精度が期待できる.

南保らが述べるように，学習者にとって助詞の習得は日本語の中で最も難しい文法項目ではあるが，『日本語誤用辞典』[45]をみると，助詞以外の見出しが多く列挙されている(たとえば，「なる・ようになる」，「ても」など)．南保らの手法は，助詞については多様な検出が行えるが，それ以外のこれらの誤用文の検出もできれば，ユーザーにとってさらに有用であるが，助詞以外の検出は対象とされていない．また，検出結果については，ユーザーに対して，取り出したものを示すのではなく，校正をしていることから，本研究で重要であると考える最後の処理が異なり，ユーザーへ提示する形での取り出しは行っていない.

■ 機械学習による誤用文検出は多種多様な誤用文を検出して取り出すか

機械学習による手法においては，SVM (Support Vector Machines : 線形二値分類器．クラス数が二つ(正クラスと負クラス)であるような問題に用いられる．タグ付きコーパスを訓練データとして与えてクラスを学習させ，実際のデータを適用すると，分類器はそのデータに対して，ク

ラスを出力する．) [59，60] を用いて助詞の誤りを検出する手法が提案されている．

Oyama [61]は SVM を用いて，格助詞のなかで最も識別されやすい「を」の誤りを対象とした検出手法を提案している．大山の実験では，新聞を訓練データとして，クラスのパターンを学習させた．パターンは格助詞を中心に前後三つの単語の品詞などが用いられた．最も識別されやすい「を」に対して，学習者の文 100 件(内，誤用文 27 件)と 200 件(内，誤用文 43 件)を適用したところ，適合率はいずれも高い精度で識別された(100 件：92.6%，200 件：95.2%)．

大山[62]は, SVM を用いて, 日本語学習者の文の正誤を判定するシステム開発に取り組んだ．システム開発において行われた実験では，訓練データとして 10,000 文のデータ用いられた，これらの半分は新聞や日本語教師が学習者の文を添削した正しい文であり, 残り半分は日本語学習者による誤りを含む文である．学習させて分類器には 978 文が用いられ，これらも半分ずつ訓練データと同じようにしてある．実験で検出した誤用文を 4 種(活用の誤り，文法の誤り，表記の誤り，語彙選択の誤り)に分類したところ，いずれも比較的高い割合で誤用文が検出された(活用の誤り：約 93%，文法の誤り：約 86%，表記の誤り：約 84%，語彙選択の誤り：約 75%)．

Oyama による手法は「を」に対して，適合率は高い精度を示したが，助詞の「を」1 語の検出では，制約が大きい．つまり，ユーザーが着目する多種多様な誤用文の検出は十分に行えない．大山による手法においても，比較的高い割合で誤用文が検出されていたが，誤用文と同時に正しい文がどれくらい検出されたかが述べられていない．そのため，ユーザーがデータを誤用に関する研究に用いる際に効率化されたかどうかを判断できない．また，分類した結果によって，高い割合が示されたが，ある種の誤用文を対象として検出したものではない．たとえば，先の形容詞過去の誤り表現を例とすると，この種の多様な誤用文をはじめから対象として検出できたのであれば，ユーザーの利用において有用性が期待できるが，検出された誤用文がどの種のものかわからないのであれば，有用に用いることが期待できない．この点について，機械学習の手法は，多様な助詞誤りの頑健な検出は期待できるが，助詞以外の分類結果の誤用文および，さらに詳細な誤用文をはじめから対象として検出することは難しいのではないかと考える．また，Oyama と大山においても，ユーザーへ提示する形での取り出しは考慮されていない．

■ コーパスを構築して統計的機械翻訳を使った手法は他の学習者コーパスにも有効か

Mizumoto ら[63]は，語学学習 SNS の添削ログから大規模な学習者コーパスを構築し，統計統計的機械翻訳の手法を用いて様々な種類の誤りを訂正する手法を提案している．

語学学習 SNS から構築した学習者コーパスには四つの特徴があり，学習者の文と添削文のペアを入手できること，添削は 1 人以上で行われることから学習者の一文に対し一文以上の添削文が付くこと，学習者の文には文の内容を意図する多言語の翻訳が付くこと，添削者は学習者の文に対してコメントを入れることができること，があげられている．

統計的機械翻訳については，奥村[60]を参考に述べる．出力言語の文 Tが，雑音のある通信路(単語の品詞についての出現確率と，品詞の並びにおける出現確率)を通って伝えられた結果，入力言語の文 Sとして観測された．この観測された S から Tを推測する問題として機械翻訳を考える．このとき，S が与えられたときに T が生じる確率 $P(T|S)$ を最大にする Tを求めたい訳文と考える．ここで，ベイズの定理を用いると，$P(T|S)$ は，

$$P(T|S) = \frac{P(T) \times P(S|T)}{P(S)} \tag{1.1}$$

となる．分母の $P(S)$ は Tを変化させても変化しない定数であるため無視すると，

$$P(T) \times P(S|T) \tag{1.2}$$

を最大化すればよいことになる．$P(S|T)$ は翻訳モデル，$P(T)$ は言語モデル表し，一般に言語モデルは単語 N-gram モデルが用いられ，翻訳モデルは一文単位での対訳データベースから学習される．

Mizumoto らにおいては，S が学習者の文となり，T が添削文となる．翻訳モデルは学習者文と添削文の 1 対 1 のコーパスから学習されている．言語モデルは単語よりも細かい文字単位に分割した手法が提案されている．これは，単語の誤りなどを含む学習者の文を形態素解析すると，うまく単語に分割されないことがあるためである．単語単位と文字単位の評価実験では，文字単位のほうが高い精度を示した．

Mizumoto らの手法は，日本語学習者の多種多様な誤りに対して対応するように提案された手法であり，形態素解析がうまく行われない点を考慮している．本研究では形態素解析器を利用して誤用文検出を行うことから，単語の綴りを誤った誤用文の検出においては，単語よりも細かい単位で検出を行う Mizumoto らの手法のほうが有効であることが予想される．しかし，Mizumoto らの行った実験は，Mizumoto ら自身が構築した学習者コーパス

であったため，教育研究向けに構築された学習者コーパスにおいても有効であるか，確認されていない．近年の学習者コーパスは，一定の条件を満たしたものであるとされる．その条件とは，

- 話し言葉や書き言葉の大規模なデータ
- 収集デザインの基準が決められた網羅的・代表的なデータ
- コンピュータで処理できること
- 言語研究に利用するもの

である[64，65]．Mizumoto らが構築した学習者コーパスは，構築のデザインと研究利用の目的とする条件が満たされていないと考える．そのため，近年の条件を満たした学習者コーパスにおいても，単語綴りの誤りを含む学習者の多種多様な誤りに対して有効であるかは不明である．

さらに，Mizumoto らが構築したコーパスにおいて，検出された誤用文データは，ユーザーが誤用に関する研究に用いる際に有用に活用できるか確認されていないため，テータの有効性は不明である．検出は添削文を基に行われるが，ある誤用に対して添削者によって添削されたりされなかったりするケースや，文の対訳があっても書き手が何を意図しているかわからず，修正するかしないか判断が分かれるケースがあることが報告されている[66]．このようなケースがあることから，検出後の誤用文データが誤用文を活用するユーザーにとって有効であるか確認が必要であるとともに，テータを有効に活用できるかは不明である．

また，Mizumoto らの手法は，添削文(統計的機械翻訳においては対訳)があることで検出が可能になるが，本研究の手法は添削文を必要としないことから，検出法において，対訳の必要の有無が異なる．加えて，本研究で調査した学習者コーパスの特徴 2 において，対訳がないものは少なくなかった[13，15～17，20，24]．したがって，これらの添削文がない学習者コーパスにおいては，統計的機械翻訳の手法では誤用文が検出できないが，表現式検出法では可能である．

また，Mizumoto らの処理は誤りの訂正であり，本研究では誤りの取り出しである点において異なっている．

1．4　本論文の構成

本論文の第 2 章以下は，次のように構成する．

第2章　構造化テキストの定義とその定義に基づく表現式
第3章　品詞の接続法に誤りがある誤用文の検出実験
第4章　従来の分類において同類である誤用文の再現
第5章　誤用研究における手法の有効性
第6章　結論

第 2 章では，日本語学習者のテキスト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について述べる．さらに，表現式検出法の簡便性について述べる．

第 3 章では,表現式検出法を用いて品詞の接続法に誤りがある誤用文の検証実験を行い，有効性を確認する．品詞間に接続の誤りがある誤用文の例として，形容動詞の名詞修飾表現の誤用文を検出する．また実験の結果と分析を踏まえて，品詞の接続法に誤りがある誤用文における表現式検出法について検討する．

第 4 章では，人が目視で確認した分類した誤用文を表現式検出法で再現できるか検証実験を行い，有効性を確認する．従来の分類の方法は，誤用の基準を人が決めて目視で判別して分類を行う．表現式検出法は誤用文を検索して見つけたものに取り出しを行っているが，これは機械が自動で誤用文を分類していると考えることができる．第 4 章では，従来の分類で誤用と判別した文を表現式検出法で再現する．また，機械による分類の視点から表現式検出法について検討する．

第 5 章では，提案手法を用いて効率的に誤用文を検出し，検出したデータを活用して学習者が誤りを原因の調査と分析を行い，提案した手法が誤用研究に有効であるか確認する．ここでの有効とは，誤用文の検出が目視や文字列検索よりも効率的であったか，さらに表現式検出法で取り出したデータが学習者の誤用を分析することができたかである．提案手法を用いて二つの異なる日本語学習者コーパスから誤用文を検出し，第 3 章で作成した表現式を再利用して誤用文を検出する．そして，提案手法によって作成した再利用の表現式が効率的に誤用文を検出するか，そして検出したデータは日本語学習者のが誤りを生じさる原因を分析できるか，手法の有効性を確認する．

第 6 章では，本論文を総括し，今後の課題を述べる．

第 2 章

構造化テキストの定義とその定義に基づく表現式

2．1　本章のあらまし

1.2 節において，日本語学習者の文法的誤用文を検出する手法を提案した．提案はテキストコーパスの文に形態素情報に基づいて誤用文を検出するように作成した表現式を用いて誤用文を検出する手法である．この手法を 1.2.5 項において，表現式検出法と名付けた．また，1.2.3 項では，構造化テキストの概要を述べ，1.2.4 項では表現式の概要を述べた．

本章では，構造化テキストの定義を行う．この定義は，形態素解析器の処理を基に，本研究で選択を行う属性と，テキストコーパスには存在しない特殊な記号と文字列を結合したものをラベルとして属性に付けていき，ラベルと属性が組み合わさった形の単語を，元の文の単語順序を保持しながら文に埋め込んでいく．本研究では，この一連の処理によって作成したラベル・属性付き単語の文から成るテキストを構造化テキストと定義する．その詳細を作成手順に従って述べていく．

また，その定義に基づいて作成を行う表現式の詳細について述べる．表現式においては，属性の選択と正規表現の組み込みと，属性およびユニットの組み合わせについての作成例を示しながら詳細を述べていく．さらに，表現式作成の概要を知るための形態素解析器への入力例についても述べる．加えて，ユーザーにとっての表現式検出法の簡便性を述べる．

2．2　コーパスの文に形態素情報を付加した構造化テキストについて

1.2.3 項で述べたように，特定の文を検出するために，テキストに対して処理を行う．処理とは，テキストコーパスから構造化テキストを作成する処理を表す．また，構造化テキストは，文を単位とし，文の単語には，形態素情報を付けてある．形態素情報を付けるには，形態素解析器を利用する．

形態素解析器は文を形態素解析すると，文と単語の連続性を保持しながら，辞書を参照して各単語に形態素情報を付加する．形態素解析器が付加する情報は，単語の文字列・読み方・基本形・品詞・活用型・活用形の情報である．

また，1.2.3 項で述べたように，形態素解析器によって付けられた単語の情報のうち，単

語の文字列と文法的属性(品詞・活用型・活用形)を選択する．そして，これらにラベル付け処理を行ったのち，一文の形に整形して構造化した文にする．

以下では，日本語学習者コーパス[22]の文を例に，構造化の詳細を作成手順にしたがって説明する．

2. 2. 1　一文単位の文と単語の順番を保持した形態素解析器の処理

ここでは，形態素解析器が文の順番を保持しながら文単位で処理を行っていることを述べる．また，文単位で処理された単語は，その順番が保持されていることを述べる．

■ 文の順序を保持しながら文単位で処理

形態素解析器は，句点を基準に一文単位で処理する．複数の文を入力すると，入力された文の順序を保持しながら処理していく．テキストコーパスを入力しても，順序を保ちながら一文単位の処理を行っていく．

日本語学習者コーパスには，次のような文の集合があった(「下略」は筆者によるもの，「…」は文が続くことを表す).

- その顔を見た後十秒くらいの時間で私はやっと思い出しました。あのおじさんは、一年前で病気で、もう亡くなりました。ライブの後は(下略)　…　(文集合 1)

文集合 1 を例に，句点を基準に文にしたものを表 2.1 に示す.

表 2. 1 句点を基準とした文の例

文番号	文
1	その顔を見た後十秒くらいの時間で私はやっと思い出しました。
2	あのおじさんは、一年前で病気で、もう亡くなりました。
3	ライブの後は(下略)
⋮	⋮

表 2.1 の文番号 1 は「～思い出しました。」の「。」で一文，文番号 2 は「～亡くなりました。」の「。」で一文となる．文番号 2 以降も同様である.

形態素解析器は，文集合 1 に対し，「。」を基準に一文と判断する．そして，一文単位での処理を行う．表 2.1 を文例にすると，文番号 1 が一文，文番号 2 が一文，…となる．処理の際，文の順番は保持され， 1 番目の文が文番号 1 が一文，2 番目の文が文番号 2，…と順番に処理していく．これは，テキストコーパスが入力されても，文の順番を保持しながら，一文単位で処理していく．

以上の処理はテキストコーパスのすべての文に対して行われる．

■ 単語の順序を保持しながら処理

形態素解析器で解析された文は，図 1.2 で示したように，縦に連なった状態で単語が出力される．この際，文における単語の順序は保持される．

表 2.1 の文番号 1 を形態素解析器に入力した．出力された情報と単語の順序の一部を表 2.2 に示す(「…」は筆者によるものであり，単語が続いていることを表す).

表 2．2 形態素解析器から出力された情報と単語の順序

単語番号	文字列	読み方	基本形	品詞	活用型	活用形
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	やっと	ヤット	やっと	副詞-一般		
2	思い出し	オモイダシ	思い出す	動詞-自立	五段・サ行	連用形
3	まし	マシ	ます	助動詞	特殊・マス	連用形
4	た	タ	た	助動詞	特殊・タ	基本形
5	。	。	。	記号-句点		
EOS						

形態素解析器から出力された単語情報は横に並んでいる．たとえば，表 2.2 の単語番号は，文字列，読み方，基本形のように，情報は横に並んでいる．単語の順序は ，単語番号 1 が「やっと」，単語番号 2 が「思い出し」，単語番号 3 が「まし」，単語番号 4 が「た」，単語番号 5 が「。」と，縦に連なっている．つまり，この縦の状態は，元の文「…やっと思い出しました。」の単語の順序を保持して出力している．これは，テキストコーパスを入力しても，文の単語の順番を保持しながら処理していく．

以上の処理は，入力したテキストコーパスのすべての文に対して行われる．

2．2．2　形態素解析器から出力された情報の選択と情報へのラベル付け

形態素解析器にはテキストコーパスを入力後，形態素解析を行い，その後ファイルを出力するように設定する．

ここでは，形態素解析器から出力された単語情報から，表現式検出法で選択する情報の処理について述べる．また，選択した情報には，ラベル付けを行う．ラベルの詳細およびその処理と，ラベルを付けた単語と文の詳細を述べる．

形態素解析器にはテキストコーパスを入力後，形態素解析を行い，解析後のファイルを出力するように設定する．

■　形態素解析器から出力された情報から単語の文字列と文法的属性を選択

1.2.3 項で述べたように，構造化テキストは，形態素解析器が単語に付けた情報のなかから，単語の文字列と，文法的属性である品詞・活用型・活用形を選択する．

選択はプログラムで処理する．表 2.2 を例にして，選択処理を行った形態素解析器からの単語情報を表 2.3 に示す．

表 2．3 選択処理を行った形態素解析からの単語情報の例

単語番号	文字列	品詞	活用型	活用形
⋮	⋮	⋮	⋮	⋮
1	やっと	副詞-一般		
2	思い出し	動詞-自立	五段・サ行	連用形
3	まし	助動詞	特殊・マス	連用形
4	た	助動詞	特殊・タ	基本形
5	。	記号-句点		
	EOS			

表 2.3 のように，形態素解析器が付けた単語情報から，文字列，品詞，活用型，活用形の情報を選択する．1.2.3 項で述べたように，各情報間は，タブで区切られており，単語番号

1，5 など，活用を持たない単語と記号にも，活用型と活用形にタブが挿入されている．

■ 単語の文字列と文法的属性に付けるラベルの定義

単語と各文法的属性に，それぞれ異なるラベルを付ける．ここでは，それぞれのラベルがどのようなものか示す．

- 単語の文字列に付けるラベル：*W_*
- 品詞に付けるラベル　　：*POS_*
- 活用形に付けるラベル　　：*CT_*
- 活用型に付けるラベル　　：*CF_*

選択処理した単語情報に対して，以上のそれぞれ異なるラベルを付ける．

■ ラベルの付け方とラベルを付けた属性の状態

ラベルはプログラミングで処理して付ける．処理は，単語の文字列の先頭に「*W_*」を挿入し，1 番目のタブを「*POS_*」に置換，2 番目のタブを「*CT_*」に置換，3 番目のタブを「*CF_*」に置換する．このようにして，それぞれの単語情報にタブを付けていく．

なお，文字列の先頭に「*W_*」を挿入るるように処理することから，「EOS」は「*W_*EOS」に処理される．そのため，「*W_*EOS」を「EOS」に置換する．

表 2.3 を例にして，ラベル付け処理を行った単語の文字列と文法的属性を表 2.4 に示す．

表 2. 4 ラベル付け処理を行った単語の文字列と文法的属性の例

単語番号	文字列	品詞	活用型	活用形
⋮	⋮	⋮	⋮	⋮
1	*W_*やっと	*POS_*副詞-一般	*CT_*	*CF_*
2	*W_*思い出し	*POS_*動詞-自立	*CT_*五段・サ行	*CF_*連用形
3	*W_*まし	*POS_*助動詞	*CT_*特殊・マス	*CF_*連用形
4	*W_*た	*POS_*助動詞	*CT_*特殊・タ	*CF_*基本形
5	*W_*。	*POS_*記号-句点	*CT_*	*CF_*
EOS				

ラベル付け処理を行うと，表 2.4 の単語番号 1 のように，単語の文字列の頭に「*W_*」，品詞情報文字列の頭に「*POS_*」, 活用型情報文字列の頭に「*CT_*」, 活用形情報文字列の頭に「*CF_*」が付けらる．また，単語番号 1, 5 のような活用を持たない単語と記号は，活用型に「*CT_*」，活用形に「*CF_*」だけが付けられた状態になる．加えて，表 2.4 は，説明のために，表の形にしたが，単語番号 1 を例にテキスト上の形を示すと，図 2.1 になっている．

*W_*やっと *POS_*副詞-一般 *CT_ CF_*

図 2. 1 テキスト上のラベル付け処理を行った例

図 1 のように，テキスト上では，一つに繋がった形になっている．これは，表 2.4 の単語番号 2～5 も図 1 と同じような形になっている．

さらに，未知語は，単語の文字列と品詞情報(「未知語」という品詞情報が付けられる)の後ろにタブが一つしか存在しない．そのため，品詞情報「未知語」と後ろのタブをターゲットに「未知語 *CT_CF_*」に置換する．たとえば，[22]の学習者コーパスには，「メーロン」という単語に「未知語」の品詞情報が付けられていた．「メーロン」に先のターゲットで置換して，次のように処理する．

- *W_*メーロン *POS_*未知語 *CT_CF_*

この処理により，活用を持たない単語と同じ形に処理する．

以上のラベル付け処理を形態素解析器から出力された単語情報に対して行う．

■ ラベルを付けた単語を一文ずつに処理

ラベルを付けた単語は，文末記号「EOS」を基点に一文ずつに処理する．

先に示した表 2.4 を例として説明する．ラベルを付けた単語 1～4 および「EOS」までの単語を，図 2.2 のように，一文の形に処理する．

> … *W_*やっと *POS_*副詞-一般 *CT_ CF_ W_*思い出し *POS_*動詞-自立 *CT_*五段・サ行 *CF_*連用形 *W_*まし *POS_*助動詞 *CT_*特殊・マス *CF_*連用形 *W_*た *POS_*助動詞 *CT_*特殊・タ *CF_*基本形 *W_*。*POS_*記号-句点 *CT_ CF_* EOS

図 2．2 ラベル付き単語を一文にした例

図 2.2 は，「*W_*」から次の「*W_*」の前までがラベルを付けた一つの単語である．表 2.1 の単語番号 1 は，図 2.1 と同じものである．単語番号 1 は，活用を持たないため，「*CT_CF_*」は，次の単語開始ラベル「*W_*」に接続する．活用を持つ単語番号 2～4 は，「*CT_*」と「*CF_*」に，それぞれ属性が付いた形で次の「*W_*」に接続する．単語番号 5 の記号も単語番号 1 と同様の形で続く文末記号に接続する．このように，各単語が繋がっていき，最後の文末記号「EOS」に接続し，一文の形をとる．

以上の一文にする処理をすべたのラベル付き単語に施す．そして，すべての文を一文ずつに処理し，構造化テキストを作成する．

2．2．3　構造化テキストに対する検索の可能性

日本語学習者のテキストコーパスを構造化テキストにすることで，有効な検索が行えると考える．ここでの有効な検索とは，検索を行うことで多様な文にヒットすることを意味する．1.2.3 項で述べたように，多様な文にヒットさせるためには，単語に品詞の情報が付いており，それを文にして構造化することであると考えた．2.2.1 項で述べたように，形態

素解析器は文と単語の順序を保持しながら処理を行う．そのため，文の形に整形しても，それらの順序は保持されていることから，元の文と変わらない状態を保っている．元の文と変わらない状態に対して，形態素情報の一つまたは複数をキーワードとして検索を行えば，品詞や活用形など，それぞれの集合に属する情報にヒットする．そのため，多様な文を検出することが期待できる．したがって，構造化テキストに対して，それを基にして作成した表現式を用いて検索することで，多様な誤用文を有効に取り出せる可能性がある．また，単語の順序は保持されていることから，取り出した文のラベルと文法的属性を取り除けば，元の文に戻すことができる．

2．3　構造化テキストを基に作成する表現式について

1.2.4 項で述べたように，構造化テキストに対して検索を行うには，それに基づいた表現式を作成して検索することで，構造化テキスト内の集合にマッチして多様な誤用を取り出すことができる．また形態素情報は一つ以上であることから，それらの組み合わせおよび組み換えを行うことで，異なる誤りのタイプを検出することもできる．

ここでは，ラベル付け処理を行った形態素情報付き単語を基に表現式の作成例を述べる．作成例では，属性の選択と組み合わせについて述べ，それらを選択および組み合わせることで，どのような集合を得るための検索となるか説明する．

2．3．1　表現式における属性の選択と組み合わせによる作成例

表 2.4 を基に，文「…やっと思い出しました。」の「やっと思い出しました」が単語に分割され，分割された単語にラベルを付けた図 2.3 を示す．

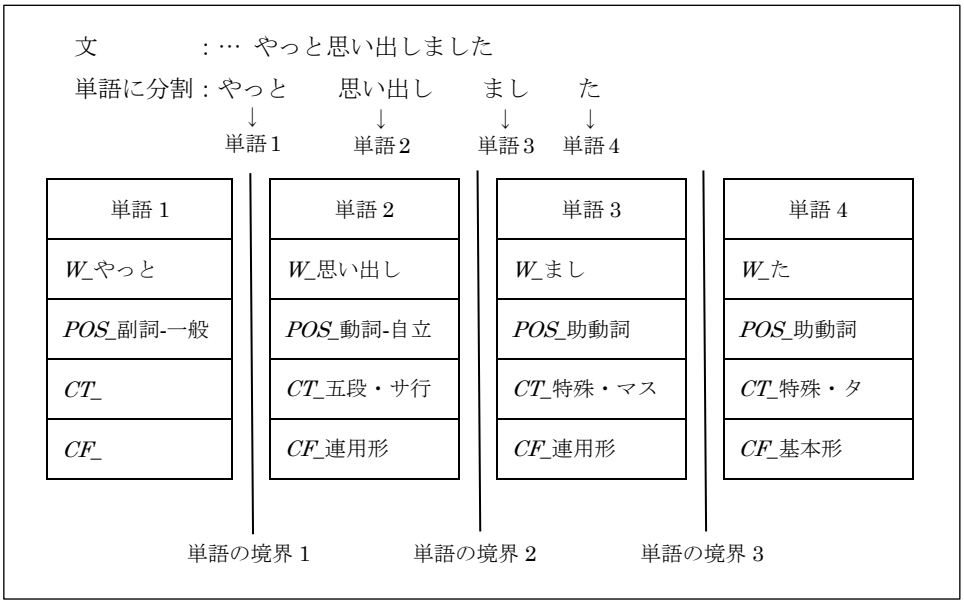

図 2. 3 ラベル付き属性の単語と単語の境界

図 2.3 の「やっと思い出しました」は，単語 1～4 がそれぞれ分割された単語を表し，単語の表 1～4 がラベルを付けた単語の文字列・品詞・活用型・活用形の属性を表す．また，単語の境界 1～3 は，単語 1～4 それぞれの単語の切れ目となる境界を表す．

以下においては，図 2.3 の単語 1～4 のラベルおよびラベル付きの属性と，単語の境界 1～3 を例にして，単語の境界の詳細，文法的属性の階層構造，属性の選択，ラベル付き属性に正規表現を組み込んだ例と表現式の作成例を述べる．さらに，表現式を組み込んだラベル付き属性の追加と省略について述べる．

■ 単語間の境界

2.2.3 で述べたように，単語間の境界は，次に出現するラベル「*W_*」までである．図 2.3 では，単語の境界 1～3 が単語間の境界になる．つまり，「*W_*」から始まり，次に出現する「*W_*」の前の文字列が一つの単語になる．

■ 階層構造を持つ文法的属性

1.2.4 項で述べたように，文法的属性の品詞と活用型には，階層構造が存在するものがある．品詞では，助動詞・感動詞・フィラー以外に存在し，活用型では，動詞のカ変と一段以外に存在する．品詞の下位階層は「-」で表され，活用型の下位階層は「・」で表されて

いる．たとえば，品詞は単語 1，2 の「副詞-」と「動詞-」に後続している「一段」と「自立」がそれぞれの下位階層の品詞であり，活用型は単語 2～4 の「五段・」と「特殊・」に後続している「サ行」,「マス」,「タ」がそれぞれの下位階層の活用型である．

これら階層構造も選択して表現式に設定することが可能である．

■ ラベル付き属性を自由に選択して集合を取り出し

ラベルの付いた単語の文字列の属性と文法的属性は，自由に選択して表現式に設定することができる．選択した属性を検索対象とすることで，その属性の集合が得ることができ，得られた集合が取り出される．ここでは，選択と設定を行った属性が，どのような集合を検索する対象となるかについて述べる．

- **単語の文字列を対象に検索**

単語の文字列を対象に検索する場合，ラベル「*W_*」の後ろに，任意の文字列を入力して設定する．たとえば，単語 1 を例にすると，「*W_*やっと」を入力して設定と，ラベル付きの単語の文字列「*W_*やっと」の集合が検索対象になる．「や」だけの単語(並列を表す助詞など)を検索するなら，「*W_*や」を入力して設定すると，「*W_*や」の集合が検索対象になる．

- **品詞を対象にした検索検索**

品詞において，単語 1 のラベル付きの品詞「*POS_*副詞-一般」を選択して設定すると，品詞ラベルに「副詞-一般」の属性が付いた「*POS_*副詞-一般」の集合が検索対象になる．また，単語 2 の「*POS_*動詞-自立語」を選択して設定すると「*POS_*動詞-自立語」の集合が検索対象になり，さらに，単語 3 または 4 の「*POS_*助動詞」を選択して設定すると「*POS_*助動詞」の集合が検索対象になる．加えて，下位階層を省略して「*POS_*副詞」または「*POS_*副詞-」を設定すると，品詞ラベル付きの副詞の集合が検索対象になる．同じように「*POS_*動詞」または「*POS_*動詞-」を設定すると，品詞ラベル付きの動詞の集合が検索対象になる．

- **活用型を対象にした検索**

活用型において，単語 2 のラベル付きの活用型「*CT_*五段・サ行」を選択して設定すると，活用型ラベルに「五段・サ行」の属性が付いた「*CT_*五段・サ行」の集合が検索対

象になる．また，単語 3 の「*CT_*特殊・マス」を選択して設定すると「*CT_*特殊・マス」の集合が検索対象になり，さらに，単語 4 の「*CT_*特殊・タ」を選択して設定すると「*CT_*特殊・タ」の集合が検索対象になる．単語 1 の活用を持たない品詞の集合を得るなら，次の活用形を対象にした検索で述べる「*CF_*」と組み合わせて「*CT_CF_*」を設定する．また，活用型においても，下位階層 を省略して「*CT_*五段」または「*CT_*五段・」を設定すると，活用型ラベル付きの「五段」の集合が検索対象になる．同様に，「*CT_*特殊」または「*CT_*特殊・」設定すると，「特殊」の集合が検索対象になる．

- **活用形を対象にした検索**

  活用形において，単語 2，3 ラベル付きの活用形「*CF_*連用形」を選択して設定すると，活用形ラベルに「連用形」の属性が付いた「*CF_*連用形」の集合が検索対象になる．さらに，単語 4 の「*CF_*基本形」を選択して設定すると「*CF_*基本形」の集合が検索対象になる．加えて，活用を持たない単語の集合を得るなら，活用型「*CT_*」と活用形「*CF_*」を組み合わせて「*CT_CF_*」を設定する．

以上のように，ラベルの付いた属性は，自由に選択することができる．選択した属性を検索対象とすることで，属性の集合を得ることができる．

■ 正規表現を組み込んだ検索

ラベル付きの属性に正規表現を組み込んで設定することができる．ここでは，表現式の検索において，よく使用すると考えるいくつかの正規表現をラベル付き属性に組み込んだ例を示す．

- **任意の一文字を組み込み**

「.」は任意の一文字を表す正規表現である．単語 1 のラベル付き単語の文字列「*W_*やっと」を例として，「*W_*.っと」と組み込むと，「*W_*やっと」や「*W_*きっと」などが検索対象となる．

- エスケープの組み込み

エスケープとは，使用された記号が正規表現ではなく文字列であることを示す．エスケー

プには「\」を使用し，正規表現と同じ記号の前に置かれる．たとえば，ラベル付き単語の文字列「*W_*」に「.」と「\」を組み込んで，「*W_*\.」とすると，検索において，「.」は正規表現ではなく，文字のピリオドとして扱われ，ピリオドにラベルが付いた「*W_*.」の集合が検索対象になる．

- **0 回以上の繰り返しを組み込み**

「*」は文字などが 0 回以上の繰り返しを表す正規表現である．単語 2 の「*W_*思い出し」を例として，このラベル付き単語の文字列に「.」と「*」を組み合わせて「*W_*思い.*」と組み込むと，「*W_*思い」，「思い出」，「*W_*思い出し」，「*W_*思い出す」などのラベル付き単語の文字列集合が検索対象となる．

- **1 回以上の繰り返しを組み込み**

「+」は文字などが 1 回以上の繰り返しを表す正規表現である．単語 2 の「*W_*思い出し」を例として，このラベル付き単語の文字列に「.」と「+」を組み合わせて「*W_*思い.+」と組み込むと，「*W_*思い出」，「*W_*思い出し」，「*W_*思い出す」などのラベル付き単語の文字列集合が検索対象となる．

- ***OR* の組み込み**

有効範囲を表す「( )」と *OR*(「または」の意味)を表す「|」の正規表現を組み込むことで，「( )」の値を選択して返す．単語 2 の「*POS_*動詞-自立」を例として，このラベル付き品詞に「( )」と「|」を組み込み，さらに動詞と形容詞を組み合わせて「*POS_*(動詞|形容詞)-自立」と設定すると，自立語の動詞または形容詞の集合が検索対象となる．

- ***NOT* の組み込み**

「(?!)」は *NOT* を表す正規表現である．検索において，*NOT* は先読みして設定した値を返さない．単語 1 と単語 2 のラベル付き活用型「*CT_*特殊・マス」と「*CT_*特殊・タ」を例に，「(?!)」を組み込んで，「*CT_*特殊・(?!マス) タ」とすると，検索において「*CT_*特殊・マス」は返されず，「*CT_*特殊・タ」の集合が得れれる．

- **文字列を抽象化した正規表現の組み込み**

計算機(コンピュータ)で使用される日本語の文章は，漢字，平仮名，片仮名，全角と半角による大文字と小文字アルファベット，全角と半角のアラビア数字が使用される．これらの文字列を抽象化した正規表現がある．文字列を抽象化した正規表現は,

- 一-龥：漢字(日本語・中国語・韓国語で使用される Unicode 漢字)
- ぁ-ん：平仮名
- ァ-ヶ：カタカナ
- Ａ-Ｚ：全角大文字アルファベット
- ａ-ｚ：全角小文字アルファベット
- A-Z：半角大文字アルファベット
- a-z：半角大文字アルファベット
- ０-９：全角アラビア数字
- 0-9：半角アラビア数字

である．これらの正規表現は有効範囲を表す「[ ]」に入れて使用する．単語 1～4 のラベル付き単語の文字列を例にすると，平仮名「ぁ-ん」と「+」組み合わせて「*W*_[ぁ-ん]+」と設定すると，単語 1，3，4 においては,「*W*_やっと」,「*W*_まし」,「*W*_た」のラベル付き単語文字列の集合が検索対象となる．また，漢字「一-龥」も組み込んで「*W*_[一-龥ぁ-ん]+」と設定(「一-龥」と「ぁ-ん」の並び順はどちらでもよい)すると，単語 1～4 においては，すべてのラベル付単語文字列の集合が検索対象となる．さらに，すべての抽象化した文字の正規表現を組み込んだ

- [一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚA-Za-z０-９0-9]+

は，計算機上で使用される日本語の文字の全体集合が検索対象になる．

以上のように，ラベル付きの属性に正規表現を組み込んで設定することができる．正規表現を用いることで，属性の文字列を抽象化した検索と，検索範囲の広狭を設定することができる．

■ ラベル付き属性を組み合わせた表現式の作成例

ここでは，図 2.3 の単語 1～4 を例にして，表現式の作成例を述べる．作成例では，ラベル付き属性の選択と組み合わ，および正規表現を組み込んだ例を示す．さらに，ユニットを組み合わせた作成例を示し，ユニット間またがる表現式が作成可能であることを述べる．これらの選択および組み合わせと正規表現の組み込みにより，検索範囲を広狭にすることができる．

- 副詞を検索対象にした作成例

単語 1 を例に，「副詞-一般」の集合を対象に検索するなら，ラベルの付いた品詞「*POS_*副詞-一般」またはラベル付きの活用型と活用形も組み合わせて「*POS_*副詞-一般 *CT_ CF_*」を設定する．

「副詞-一般」のなかでも，文字列「やっと」の集合を検索するなら，ラベル付きの単語の文字列と品詞を組み合わせて「*W_*やっと *POS_*副詞-一般」，またはラベル付きの活用型と活用形も組み合わせて「*W_*やっと *POS_*副詞-一般 *CT_ CF_*」を設定する．

- 平仮名の副詞を検索対象にした作成例

単語 1 を例として，単語の文字列が平仮名の「副詞-一般」の集合を検索するなら，「*W_*」に正規表現の平仮名と 1 回以上繰り返しの組み込みとラベル付き品詞を組み合わせて「*W_*[一-纔]+*POS_*副詞-一般」，またはラベル付きの活用型と活用形も組み合わせて「*W_*[一-纔]+*POS_*副詞-一般 *CT_ CF_*」を設定する．

- 動詞に関連する品詞を検索対象にした作成例

単語 2 のラベル付き品詞を例として，動詞の集合を検索するなら，下位階層を省略して「*POS_*動詞」，または「-」を含めた「*POS_*動詞-」を設定する．動詞の下位階層には，「自立」，「非自立」，「接尾」がある[39]．これらのうち，「自立」と「非自立」の集合を検索するなら，正規表現の *OR*「|」と有効範囲「( )」を組み込んで，「*POS_*動詞-(自立|非自立)」を設定する．さらに，「接尾」も含めた集合を得るなら，「|」と「接尾」を追加する．また，自立語の動詞または形容詞の集合を検索するなら，「*POS_*(動詞|形容詞)-自立」を設定する．

- 動詞の活用型を検索対象にした作成例

単語 2 を例として，動詞の活用型「五段・サ行」の集合を検索るなら，「*CT_*五段・ラ行」を設定する．また，辞書の動詞の活用型は，漢字と片仮名の文字列で構成されている[39]．そこで，「五段・サ行」を含めたすべての動詞の活用型の集合を対象にするなら，正規表現の漢字と片仮名が 1 回以上を組み込んで，「*CT_*[一-龥ァ-ヶ]+」を設定する．

- **単語間にまたがる表現（ユニット間）を検索対象にした作成例**

表 2.3 の単語 1～4 の単語間の接続を例に，単語間にまたがる表現式の作成例を示す．また，ユニットは，単語における一つ以上のラベル付き属性が 1 ユニットである．

単語 1，2 を例として，単語 1 をユニット 1，単語 2 をユニット 2 とする．ここで，品詞が「副詞-一般」と，単語の文字列が「思い出し」で品詞が「動詞-自立」の表現の集合を得る検索を行うとする．その場合，ユニット 1 の「*POS_*副詞-一般 *CT_CF_*」と，ユニット 2 の「*W_*思い出し *POS_*動詞-自立」を接続し，「*POS_*副詞-一般 *CT_CF_W_*思い出し *POS_*動詞-自立」と設定する．さらに，品詞が「副詞-一般」と，すべての「動詞-自立」の品詞が接続した表現の集合を得る検索を行うとする．動詞の単語の文字列は漢字と平仮名の連続であると想定される．そこで，この設定における「*W_*思い出し」を「*W_*[一-龥ぁ-ん]+」に変更，あるいは先の計算機上で使用される日本語の文字の全体集合を組み込んで設定する．

単語 2，3 を例として，単語 3 をユニット 3 とする．ここで，品詞が「動詞-自立」で活用型が「五段・サ行」で活用形「連用形」と，単語の文字列が「*W_*まし」品詞が「助動詞」の表現の集合を得る検索を行うとする．その場合，ユニット 2 の「*POS_*動詞-自立 *CT_*五段・サ行 *CF_*連用形」と，ユニット 3 の「*W_*まし *CT_*助動詞」を接続し，「*POS_*動詞-自立 *CT_*五段・サ行 *CF_*連用形 *W_*まし *CT_*助動詞」と設定する．さらに，この設定において，動詞の活用型「*CT_*五段・サ行」を「*CT_*[一-龥ァ-ヶ]+」にすると，すべての活用型の集合を得る設定になる．加えて，この設定において，動詞の活用形「連用形」を先の計算機上で使用される日本語の文字の全体集合を組み込みこむと，すべての活用形の集合を得る設定になる．

単語 3，4 を例として，単語 4 をユニット 4 とする．ここで，品詞の助動詞が二つ接続した表現の集合を得る検索を行うとする．その場合，ユニット 3 の品詞「*POS_*助動詞」と活用型「*CT_*五段・サ行」を「*CT_*[一-龥ァ-ヶ]+」にし，活用形に日本語の文字の全体集合を組み込む．そして，ユニット 4 の単語の文字列のラベル「*W_*」に日本語の文字の全体集合

を組み込み，品詞「*POS*_助動詞」を後続させる．これら変更を行った二つのユニットを接続させることで，助動詞が二つ接続した表現の集合を得る設定になる．

以上のようにラベル付き属性は選択と組み合わを行った設定ができる．また，表現式は正規表現を組み込んだ設定も行え，ユニットの組み合わせも行うことができる．設定におけるラベル付き属性およびユニットの選択と組み合わせ，また正規表現の組み込みより，検索範囲の広狭と，異なるタイプの文を検索対象にすること可能となる．

■ ラベル付き属性の追加上限と省略

正規表現を使うことで，ラベル付きの属性は，上限なく**追加**して設定することができる．加えて，「て」など，単語の文字列においても，上限なく入力して設定することができる．さらに，ユニット間の設定も上限なく**追加**設定ができる．

また，正規表現を用いて，ラベル間の属性およびユニット間の省略を行うことができる．

2．3．2　表現式を作成するための形態素解析器への入力

表現式の作成は，着目した少ない誤用例を形態素解析に入力することで，どのように作成するか，その概要を知ることができる．たとえば，1.1.3 項の誤用文の例「難しいでした」を形態素解析器に入力する．入力から得られたユニットと本研究で選択する形態素情報をを図 2.4 に示す(図 2.4 の下線は表現式検出法で選択する形態素情報を表す)．

図 2．4　表現式作成のための形態素解析器への入力例と得られる情報

図 2.4 から，入力した誤用文は三つの単語に分割されてことから，三つのユニットから構成であることがわかった．また，入力により選択する形態素情報についてもわかった．

ここで，1.1.3 項の文献に指摘されているように，形容詞に「でした」が接続している誤用文を検出する表現式を作成するとする．その場合，最初のユニットは品詞が「形容詞-自立」であり，活用型が「形容詞・イ段」，活用形が「基本形」であることがわかる．次のユ

ニットは単語の文字列が「でし」，品詞が「助動詞」，活用型が「特殊・デス」，活用形が「連用形」であることがわかる．最後のユニットは単語の文字列が「た」，品詞が「助動詞」，活用型が「特殊・タ」，活用形が「基本形」であることがわかる．そして，これらの形態素情報にそれぞれのラベルを付け，各ラベルを付けた形態素譲歩とユニットを接続させることで，形容詞に「でした」が接続した表現式となる．

このように，表現式作成にあたり，着目した誤用例を形態素解析器に入力することで，作成のための概要を知ることができる．

また，この着目した少ない例の誤用文の入力は，本研究では，表現式作成のための簡単な実験と捉えている．

2．4　表現式検出法の簡便性について

ここでは，表現式検出法の簡便性について述べる．これは，ユーザーにとって，どのように簡便であるか，というものである．

なお，1.2.7 項で表現式検出法の有用性を述べたが，簡便性にも関連がある内容は，2.5 節においても再度述べる．

■　文字列検索よりも多様な誤用文の取り出しが可能

1.2.4 項で述べたように,表現式検出法を用いた検索は,集合検索を行っていることから，多様な文にマッチする．また，少ない例から検索しても，集合にマッチさせることで，多様な文が取り出せる．この取り出し方式は，従来のユーザーが用いている文字列検索よりも簡便性が高い．

■　電子化されただけのコーパスから誤用文の検出が可能

1.1 節の調査から，日本語学習者コーパスの多くが電子化されただけのコーパスであった．表現式検出用は，このようなコーパスから誤用文を検出することができる．これは，ユーザーにとって，日本語学習者コーパスのデータを，さらに活用できる機会を提供できるものと考える．

■　形態素解析器を利用して形態素情報を付加が可能

形態素情報の付加は，人手で行わなくても，形態素解析器を利用して付けることができ

る．形態素解析器の利用により，構造化テキスト作成にまでの簡便な処理を行うことができる．

■ 機械を利用した検出による揺れのない誤用文の取り出し

表現式検出法は，機械を利用した検出法であることから，テキストコーパスから揺れのない誤用文が検出される．1.1 節の調査から，公開されいる日本語学習者コーパスで誤用文を検出するシステムは，人手でタグを付けたものであった．また，タグ付けは揺れがあることが指摘されていた．そのため，これらのシステムで検出された誤用文には，揺れの問題が存在する．表現式検出法による検出は，この問題を解消している．

■ 構造化テキストはツールを利用しても可能

1.2.6 項で述べたツール[49，50]を利用して，テキストコーパスを構造化テキストにすることができる．これらのツールは，辞書の違いや文法的属性などが本研究とは異なっていることが予想されるが，ツール上では文法的属性を用いて検索することができることから，テキストコーパスから構造化テキストを作成していることがわかる．つまり，ユーザーはツールを利用して，テキストコーパスを簡便に構造化テキストにすることができる．

■ 表現式検出法はツールを利用しても可能

ツールを利用した構造化テキスト作成と同様に，ツール上において，表現式検出法を用いることができる．ツールにテキストコーパスを入力すると，構造化処理が行われ，処理後，窓から単語の文字列や文法的属性などを入力する．そして，入力したキーワード(文字単語の文字列や文法的属性など)を検索すると，文が検出される．

本研究と比較して，KH Coder[51]は，使用できる文法的属性が品詞と活用型の二つであることと，組み合わせが行えるユニットは三つと制限があるため，柔軟性はやや低いが，検出される文は本研究に近いことが予想される．また，ChaKi[49]は，正規表現が使えるとともに，単語の文字列と文法的属性に加え，単語の読み方・発音・基本形も扱える．選択項目が多いことは，ユーザーに混乱を招き利便性が低下している可能性があるが，うまく選択を行うことができれば，検出される文は本研究とほぼ同じであろう．

つまり，ユーザーはこれらのツール上においても，表現式検出法を用いて簡便に誤用文検出を検出ことができる．

さらに，近年のテキストエディターは，正規表現を扱うことができることから，テキストエディターにおいても，表現式検出法を用いることができる．テキストエディターによる表現式検出法は，ユーザーの作業コストが増えることが予想されるが，テキストコーパスをうまく構造化テキストに加工し，表現式を作成して検索すれば，ラベルが付いたままのわかりにくい文ではあるが，検出が行える．

■ 表現式の再利用と誤用文検出の効率化

1.2.7 項で述べたように，表現式は，構造化テキストを作成した環境と同じであれば，再利用が可能である．つまり，異なるコーパスにおいても，同じ表現式を利用して誤用文を検出することができる．これにより，誤用文検出の効率化を行うことができる．

表現式を同じ環境で用いて効率化を行うことについては，5 章においても述べる．

■ 簡単な実験によって表現式作成の概要を知ることが可能

ユーザーは，形態素解析器を使った簡単な実験を行うだけで，表現式作成の概要を知ることができる．1.2.7 項および 2.3.2 項で述べたように，表現式の作成にあたり，最初の実験として，経験や文献にある少ない例を形態素解析器に入力してみる．この簡単な実験だけで，ユーザーが着目した誤用文を検出する表現をどのように作成するか，その概要を知ることができる．そして，概要を基に，ラベル付け，文法的属性の選択，必要に応じて正規表現の設定を行い，表現式を作成する．さらに， 1.2.7 項で述べたように，ユーザーは学習者に日本語を教えた経験を生かして，直感的に表現式の改良を行うことができる．

2．7　本章のまとめ

本章では，ラベル・属性付き単語の文から成るテキストを構造化テキストの定義を行い，作成手順を基に詳細を述べた．また，日本語学習者のテキストコーパスを構造化テキストにすることで，検索において多様な文にヒットさせることが期待できること，構造化テキストは元の単語の順を保持していることから，元の文に戻すことが可能であることを述べた．

構造化テキストの定義に基づいて作成する表現式については，作成例に基づいて属性の選択と正規表現の組み込みと，ラベル付き属性およびユニットの組み合わせについての作成例を示しながら詳細を述べた．そして，ラベル付きの属性とユニットは自由な組み合わ

せがが可能であるとともに正規表現も組み込むことができ，検索範囲を広狭に設定できることを述べた．また，多様な誤用文の集合を検出できるとともに，異なるタイプの誤用文を検出する設定も可能であることを述べた．

さらに，少ない例を形態素解析器に入力することで表現式作成の概要を知ることができることと，表現式検出法の簡便性について述べた．

第 3 章

品詞の接続法に誤りがある誤用文の検出実験

3. 1　本章のあらまし

1.2.7 項で述べたように，表現式検出法が有用な検出を行えるのは，品詞の接続法に誤りがある誤用文であると考えている．本章では，それを検証するために，表現式検出法を用いて品詞の接続法に誤りがある誤用文の検出実験を行い，有効性を確認する．品詞の接続法に誤りがある誤用文には，形容動詞の名詞修飾表現の誤用表現を例とする．

3. 2　品詞の接続法に誤りがある誤用文とは

1.2.7 項で述べたように，品詞の接続法に誤りがある誤用文とは，単語の文法的属性は正しいが，品詞などの文法的属性の接続法は日本語として正しくない文法的な誤用文である．

また，同様に 1.2.7 項で述べたように，品詞の接続法に誤りがある誤用文は，学習者が誤りやすい傾向があるとして文献にも示されている．

3. 2. 1　形容動詞の名詞修飾表現と学習者の誤用文

3.1 節で述べたように，品詞の接続法に誤りがある誤用文の検出実験として，形容動詞の名詞修飾表現に着目する．ここでは，形容動詞の名詞修飾表現についてと学習者が産出するその誤用について述べる．

形容動詞は名詞に続くときに助動詞「だ」の連体形語尾「な」と共起させて修飾するという特色がある[37].

学習者が生じやすい形容動詞の名詞修飾表現の誤用文は,「善意な人々」のように形容動詞ではない単語に「な」を共起させた使用や,「必要の注意を怠る」のように形容動詞を「の」と共起させた使用が見られる[37]．このような誤用について，中国語を母語とした学習者を対象に調査したところ，文末で使用される形容動詞表現よりも誤用率が高く，中上級者となっても誤用の割合は減らずにむしろ増加していたことが報告されている[3]．また，接尾辞「的」の表現も形容動詞と同じように「な」と共起して名詞を修飾する．このことから，「的」の付いた表現を「的」付き形容動詞として調査したところ,「希望的な春」,「体格的

な体」のような誤った表現があることが報告されている[67]．この「的」付き形容動詞について，本研究においても形容動詞の名詞修飾表現として扱う．

3．2．2　実験に使用するデータ

1.2.2 項で述べたように，台湾人日本語学習者コーパス[22]の 135 作文(句読点を基準としたところ 2,185 文)を使用する．この作文は「思い出」というテーマで書かれた作文で，旧日本語能力試験の取得級が示されいるものである．

3．2．3　正しい形容動詞の名詞修飾表現の調査とその結果

実験で使用するデータにおいて，正しく使用された形容動詞の名修飾表現を調査した．この調査の目的は，正しく使用された例がなければ，学習者はこの表現を使用していないことになるため，使用されていることを確認するために行った．調査は目視で確認した．

調査の結果，正しく使用された表現は 203 件あった．うち，「な」の正しく使用された表現は 195 件，「的な」の正しく使用された表現は 8 件あった．

以上の結果から，実験で使用するデータにおいて，形容動詞の名修飾表現が使用されていることが確認できた．

3．2．4　誤用表現とその判別基準

3．2．1 項で指摘されているような誤った表現は，文中に一つ以上書かれていることがあるため，文中に書かれた一つの誤った表現を誤用表現とする．

誤用表現を判別する基準を設けた．この基準は誤用を研究する研究者ら設けるであろうと想定したものにした．以下に想定した基準を示す．

- 「善意な人々」など，「な」と共起しない単語を使用して後続する単語を修飾している文
- 「必要の注意」など，形容動詞の単語と「の」を使用して後続する単語を修飾している文
- 「ラッキな人」など，「な」または「の」の前後の単語が正しく書かれていない文
- 「巨大的なビル」など，「的な」と共起しない単語を使用して後続する単語を修飾している文

- 「伝統的の文化」など,「的」と「の」を使用して後続する単語を修飾している文

3. 2. 5　目視で形容動詞の名詞修飾表現の誤用文を判別した結果

基準にしたがって，目視で文を判別した．判別の結果を表 3.1 に示す.

表 3.1 目視で判別した形容詞名詞修飾表現の誤用文と誤用表現の件数

誤用文の件数	42
誤用表現の件数	44

表 3.1 の結果から, 使用したデータには形容動詞の名詞修飾表現の誤用文が存在していた. また，誤用表現は，一文中に一つ以上存在する文があったしていた.

3. 2. 5　誤用表現を分類した結果

表 3.1 の誤用文を基に目視で判別した結果を分類した．分類の結果を表 3.2 に示す.

表 3.2 目視で判別した形容詞名詞修飾表現の各誤用表現の件数

誤用表現	誤用表現の件数
「な」の誤用表現	31
「の」の誤用表現	11
「的な」の誤用表現	2
「的の」の誤用表現	0

表 3.1 を分類した表 3.2 の結果から，誤用文は 4 種に分類された．また「的の」の誤用表現は，使用したデータにはなかった.

誤用表現が 4 種に分類されたれたことから，この 4 種を基にして表現式を作成する．また,「的の」の誤用表現はなかったが，表現式を使った検索で検出される可能性があるため,「的の」の表現式も作成する.

3. 2. 6　機械による自動分類器としての表現式

誤用文のなかにある 4 種の誤用表現にマッチさせる表現式を四つ作成する．この 4 種を検出する表現式は，機械が自動で誤用文を分類して検出するものであると考えることができる．特定の文を捜して検出するという作業は，分類器にかけていると捉えることができるからである．

以下にそれぞれの表現式と機械による分類の内容を示す.

- 「な」の誤用文：
  形容動詞**ないのに**「な」を使って後ろの単語を修飾している文を検出する表現式
  例：善意な人々
- 「の」の誤用文：
  形容動詞を「の」を使って後ろの単語を修飾している文を検出する表現式
  例：必要の注意
- 「的な」の誤用文：
  形容動詞を使って「的な」で後ろの単語を修飾している文を検出する表現式
  例：巨大的なビル
- 「的の」の誤用文：
  「的の」を使って後ろの単語を修飾している文を検出する表現式
  例：伝統的の文化

これら機械による分類の表現式を次に作成する．なお，4 種の表現式の作成に使用した正規表現の一覧を表 3.3 に示す.

■ 形容動詞ではないのに「な」を使って後ろの単語を修飾している文を検出する表現式

「善意な人々」のように，形容動詞ではない単語を使用して「な」と共起させ，後続する単語を修飾している表現にマッチさせる表現式 3.1 を示す.

((?!*POS*_名詞-(形容動詞語幹|特殊-助動詞語幹|非自立-(助動詞語幹|形容動詞語幹)|接尾-(助動詞語幹|形容動詞語幹)))*POS*_(名詞[\-一-㍿ァ-ヶ]+([\-一-㍿ァ-ヶ]+)*|副詞-一般|未知語|動詞-自立|形容詞-自立)*CT*_([一-㍿ァ-ヶ]+)**CF*_(基本形)**W*_な *POS*_助動詞 CT_特殊・ダ *CF*_体言接続(?!*W*_(の|ん))*W*_[一-㍿ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_(名詞-|副詞-一般|未知語|動詞-自立|形容詞-自立) (3.1)

表現式 3.1 は，四つのユニットから構成されている．

■ **表現式 3.1 のユニット 1**

- ((?!*POS*_名詞-(形容動詞語幹|特殊-助動詞語幹|非自立-(助動詞語幹|形容動詞語幹)|接尾-(助動詞語幹|形容動詞語幹)))

は，「きれいな花」の「きれい」ように，「な」の前に正しく接続する品詞を「!?」(先読みして失敗させる，正規表現 *NOT* の意味)にした．これは，「!?」の後ろに続く品詞が「な」に接続する正しい日本語表現だからである．「名詞-(形容動詞語幹|特殊-助動詞語幹|非自立-(助動詞語幹|形容動詞語幹)」にある「名詞-」の「-」は，文字列ではなく品詞を表す文字列の一部であり，名詞の階層を表す．「-」に続く「( )」は正規表現の範囲を表し，「( )」のなかにある「|」は，正規表現の *OR*(または)を表す．よって，「名詞-(形容動詞語幹|特殊-助動詞語幹|…)」は，「名詞-形容動詞語幹 *OR* 名詞-特殊-助動詞語幹 *OR*…」の意味である．「名詞-形容動詞語幹」は「きれい」などの形容動詞，「名詞-特殊-助動詞語幹」は「ニュースによると，明日は雨が降るそうだ」の「そう」(伝聞)，「名詞-非自立-助動詞語幹」は「明日は雨が降るようだ」の「よう」など，「名詞-非自立-形容動詞語幹」は「明日は雨がみたいだ」の「みたい」など，「名詞-接尾-助動詞語幹」は「もうすぐ雨が降りそうだ」の「そう」(様態)，「名詞-接尾-形容動詞語幹」は「彼は忘れ物をしがちだ」の「がち」など，の品詞である[39].

■ **表現式 3.1 のユニット 2**

- *POS*_(名詞[\-一-㍿ァ-ヶ]+([\-一-㍿ァ-ヶ]+)*|副詞-一般|未知語|動詞-自立|形容詞-自立)*CT*_([一-㍿ァ-ヶ]+)**CF*_(基本形)*

の形態素情報は，「な」の前に接続している品詞にマッチさせる．

先頭の品詞,

- (名詞[\-一-龥ァ-ヶ]+([\-一-龥ァ-ヶ]+)*

は，名詞を表す．「[ ]」は正規表現の文字範囲を表し，そのなかの「\」はエスケープを表す．エスケープは，「-」を正規表現と区別させ，文字列であることを示す (1.2.4 項で述べたように，名詞は「名詞-形容動詞語幹」のように階層があるため).「[ ]」のなかの「一-龥」は，1.2.4 項で述べたように，日本語・中国語・韓国語で使用される Unicode 漢字の正規表現,「ァ-ヶ」は日本語のカタカタの正規表現を表す．名詞の階層で使用されている文字列は漢字と片仮名が使用されている[39]ため，「[一-龥ァ-ヶ]」に設定した．続く「+」は 1 回以上の繰り返しを表す．名詞の階層は一つ以上あるため，1 回以上の繰り返しを設定した．続く「[\-一-龥ァ-ヶ]+」も名詞の階層を表すが，名詞は階層が一つのものと二つのものがあるため，0 回以上の繰り返しを表す「*」を「( )」の範囲のなかに入れて設定した．

名詞に続く，

- |副詞-一般|未知語|動詞-自立|形容詞-自立

は，名詞以外の自立語と辞書にない単語を表す．「な」の前に接続する自立語は形容動詞以外であるため，これらの単語を設定する．「副詞-一般」は「多分」など必ず後ろで切れて(活用しない)連体修飾しない副詞[39]，未知語は辞書にない単語,「動詞-自立」は「買う」などの自立語動詞,「形容詞-自立」は「楽しい」などの自立語形容詞を表す．これらを *OR* で設定した．

自立語単語に続く，

- *CF_* ([一-龥ァ-ヶ]+)*

は，自立語動詞と自立語形容詞の活用型を表す．活用型の文字列は，漢字と片仮名が使用されている[39]ため「 [一-龥ァ-ヶ]」にし，階層も一つ以上あるため「+」を設定した．また，名詞・副詞・未知語は活用型がないため,「( )*」を設定した．

活用型に続く，

- *CT_*(基本形)*

は，自立語の動詞と形容詞の活用形を表す．「な」に誤って接続させた動詞と形容詞は，基本形で使われていると予測し,「基本形」を設定した．また，名詞・副詞・未知語は活用型がないため,「( )*」を設定した．

■ **表現式 3.1 のユニット 3**

- *W_*な *POS_*助動詞 CT_特殊・ダ *CF_*体言接続(?! *W_*(の|ん))

は,「な」の形態素情報を表す.

「な」の形態素情報は,

- *W_*な *POS_*助動詞 CT_特殊・ダ *CF_*体言接続

である．この形態素情報の意味は，単語の文字列が「な」，品詞が「助動詞」，活用型が「特殊・ダ」，活用形が「体言接続」を表す． ipadic において,「な」の品詞は「助動詞」に分類されている[39]．品詞「助動詞」は単独では意味を持たない付属語で，体言(活用しない語)と用言(活用する語)に接続する．「特殊・ダ」は断定の助動詞「だ」の活用型である[39]．ipadic における「体言接続」は，品詞「助動詞」の場合，後続する品詞が名詞など，体言と接続する活用形である．

「な」の形態素情報に続く,

- (?! *W_*(の|ん))

は，単語の文字列が「の」または「ん」を *NOT* で返すように設定した．これは,「～なのです」の「の」,「～なんです」の「な」など,「の」と「ん」が接続する正しい表現があると直感により予測したからである．

■ **表現式 3.1 のユニット 4**

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*(名詞-|副詞-一般|未知語|動詞-自立|形容詞-自立)

は,「な」の後ろに接続する形態素情報を表す.

単語の文字列,

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，正規表現の漢字，片仮名，平仮名「ぁ-ん」，全角大文字アルファベット「Ａ-Ｚ」，全角小文字アルファベット「ａ-ｚ」, 全角アラビア数字「０-９」, 半角大文字アルファベット「A-Z」, 半角大文字アルファベット「a-z」，半角アラビア数字「0-9」，を１回以上に設定した．

単語の文字列に続く,

- *POS_*(名詞-|副詞-一般|未知語|動詞-自立|形容詞-自立)

は，単語の品詞を「名詞-」，副詞，未知語，自立語動詞と自立語形容詞を *OR* に設定した．名詞は「名詞-」とすることで下の階層すべてにマッチするため，このように設定した．また，名詞修飾表現の検出であるから，名詞以外の自立は共起しないはずであるから，直感

を用いて名詞以外の自立語となる品詞を設定した.

■ 形容動詞なのに「の」を使って後ろの単語を修飾している誤用表現

「必要の注意」のように，形容動詞の単語を使用して「の」と共起させ，後続する単語を修飾している表現にマッチさせる式を作成した．作成した表現式 3.2 を示す.

(?!*POS*_名詞-(特殊-助動詞語幹|非自立-(助動詞語幹|形容動詞語幹)|接尾-(助動詞語幹|形容動詞語幹)))*POS*_名詞-形容動詞語幹 *CT_CF_W_*の *POS*_助詞-連体化 *CT_CF_*(?!*W_*(の|ん))*W_*[一-㷀ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_(名詞-|副詞-一般|未知語|動詞-自立|形容詞-自立) (3.2)

表現式 3.2 は，四つのユニットから構成されている.

■ 表現式 3.2 のユニット 1

- ((?!*POS*_名詞-(形容動詞語幹|特殊-助動詞語幹|非自立-(助動詞語幹|形容動詞語幹)|接尾-(助動詞語幹|形容動詞語幹)))

は，表現式 3.1 と同様の設定である.

■ 表現式 3.2 のユニット 2

- *POS*_名詞-形容動詞語幹 *CT_CF_*

の形態素情報は，形容動詞を設定した．正しい形容動詞の名詞修飾表現は，続く単語「の」と接続しないからである．つまり，「形容動詞＋の」の誤用文にマッチさせる設定である．また，ipacdic の形容動詞は名詞に分類している[39]．名詞は活用しないため，活用型と活用形は設定を行わない.

■ 表現式 3.2 のユニット 3

- *W_*の *POS*_助詞-連体化 *CT_CF_*(?!*W_*(の|ん))

は，助詞「の」の形態素情報である.

この形態素情報,

- *W_*の *POS*_助詞-連体化 *CT_CF_*

の意味は，単語の文字列が「の」，品詞が「助詞−連体化」を表す.「助詞−連体化」は名詞

に接続して体言にかかる「の」の品詞である[56]．また，助詞は活用しないため，活用型と活用形には設定を行わない．

助詞「の」の形態素情報に続く，

- (?!*W_*(の|ん))

は，表現式 3.1 と同様に，文字列「の」または「ん」を *NOT* で返すように設定した．

■ **表現式 3.2 のユニット 4**

- *W_*[一-㷀ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*(名詞-|副詞-一般|未知語|動詞-自立|形容詞-自立)

の形態素情報は，表現式 3.1 と同様に設定した．

■ **形容動詞を使って「的な」で後ろの単語を修飾している誤用表現**

「巨大的なビル」のように，形容動詞の単語を使用して「的な」と共起させ，後続する単語を修飾している表現にマッチさせる式を作成した．作成した表現式 3.3 を示す．

*POS_*名詞-形容動詞語幹 *CT_CF_W_*(的|てき)*POS_*名詞-接尾-形容動詞語幹 *CT_CF_W_*な *POS_*助動詞 *CT_*特殊・ダ *CF_*体言接続 *W_*[一-㷀ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*(名詞-|副詞-一般|未知語|動詞-自立|形容詞-自立)　(3.3)

表現式 3.3 は，四つのユニットから構成されている．

■ **表現式 3.3 のユニット 1**

- *POS_*名詞-形容動詞語幹 *CT_CF_*

は，形容動詞である．正しい「的」の名詞修飾表現は形容動詞と接続しない．つまり，「形容動詞＋的」の誤用文にマッチさせるため，設定した．

■ **表現式 3.3 のユニット 2**

- *W_*(的|てき)*POS_*名詞-接尾-形容動詞語幹 *CT_CF_*

は，「的」の形態素情報である．

この形態素情報，

- *W_*(的|てき)

は，単語の文字列が，漢字または平仮名の「的」を表す.

単語の文字列に続く，

- *POS_*名詞-接尾-形容動詞語幹 *CT_CF_*

は，「的」の形態素情報を表す.「的」は，ipadic で名詞に分類している[39]．名詞は活用しないため，活用型と活用形には設定を行わない.

■ **表現式 3.3 のユニット 3**

- *W_*な *POS_*助動詞 *CT_*特殊・ダ *CF_*体言接続

は，「的な」の助動詞「な」の形態素情報である．この形態素情報は表現式 3.1 と同様である.

■ **表現式 3.3 のユニット 4**

- *W_*[一-龠ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*(名詞-|副詞-一般|未知語|動詞-自立|形容詞-自立)

は，助動詞「な」に続く形態素情報であり，表現式 3.1 と同様の設定である.

■ **「的の」を使って後ろの単語を修飾している誤用表現**

「伝統的の文化」のように，「的の」の前に単語を接続して共起させ，共起表現に後続する単語を修飾している表現にマッチさせる式を作成した．作成した表現式 3.4 を示す.

*W_*(的|てき)*POS_*名詞-接尾-形容動詞語幹 *CT_CF_W_*の *POS_*助詞-連体化 *CT_CF_W_*[一-龠ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*(名詞-|副詞-一般|未知語|動詞-自立|形容詞-自立)　(3.4)

表現式 3.4 は三つのユニットから構成されている.

■ **表現式 3.4 のユニット 1**

- *W_*(的|てき)*POS_*名詞-接尾-形容動詞語幹 *CT_CF_*

は，「的」の形態素情報であり，表現式 3.3 と同様である.

■ 表現式 3.4 のユニット 2

- *W_*の *POS_*助詞-連体化 *CT_CF_*

は，助詞「の」の形態素情報であり，表現式 3.2 と同様である．「的」を使った名詞修飾表現は「的＋の＋名詞」とはならない．つまり，この誤用文にマッチさせる設定である．

■ 表現式 3.4 のユニット 3

- *W_*[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*(名詞-|副詞-一般|未知語|動詞-自立|形容詞-自立)

助詞「の」に続く形態素情報であり，表現式 3.1 と同様の設定である．

表 3.3 形容動詞名詞修飾表現の表現式で使用した正規表現

番号	正規表現	働き	番号	正規表現	働き
1	\	エスケープ	9	ぁ-ん	平仮名
2	*	0 回以上の繰り返し	10	ァ-ヶ	片仮名
3	+	1 回以上の繰り返し	11	Ａ-Ｚ	全角大文字アルファベット
4	()	「（」から「）」までの範囲	12	ａ-ｚ	全角小文字アルファベット
5	?!	*NOT*(先読みして否定)	13	０-９	全角アラビア数字
6	|	*OR*(または)	14	A-Z	半角大文字アルファベット
7	[]	[]内の一文字	15	a-z	半角小文字アルファベット
8	一-纊	漢字	16	0-9	半角アラビア数字

3. 2. 7　表現式を使った誤用文の検出結果

使用するデータに形態素情報を付けて，作成した表現式を用いて検索した．その結果，誤用文が検出された．検出された文と表現の件数を表 3.4 に示す．

表 3.4 形容動詞名詞修飾表現の表現式で検出された文と表現の件数

検出した文	誤用文	誤用表現
37	28	30

本実験における表現式は，次のような単語と表現が書かれた文も検出された．

- 形態素解析器の辞書に登録されていなかったために，正しく解析しなかった単語
  例：堅強
- 形態素解析器の辞書に判別とは異なる品詞で登録されていた単語
  例：普通
- 単語が正しく書かれていないために，誤って解析された単語と表現
  単語の例：「コンサートのないよ」の「ないよ」(正しくは「内容」)
  表現の例：行こなできた(正しくは「行うことができた」)
- 正しく書かれているけれども，形態素解析器が誤って解析した表現
  例：たいくつなこと
- 誤用表現であると判別できるけれども，基準にはない表現
  例：同じのように

これらの文も検出されたが，本実験で作成した表現式の検索は，誤用文を検出できた．

3．2．8　目視による判別と表現式による検出の比較

目視によって判別した文と表現式を使って検出した文を比較した．その結果，表現式を使った検索は，いくつかの文が検出されなかった．目視で判別して取り出した誤用表現と表現式で検出した誤用表現の件数を表 3.5 に示す．

表 3.5 目視で判別した件数と表現式で検出した件数

誤用表現	目視で判別	表現式で検出	表現式で検出なし
「な」の誤用表現	31	23	8
「の」の誤用表現	11	6	5
「的な」の誤用表現	2	1	1
「的の」の誤用表現	0	0	0

3.2.9 項において，検出されなかった文についての分析を行う．

3. 2. 9　表現式で検出されなかった誤用文の分析

本実験において，表現式を使った検索で検出されなかった誤用文を目視の結果と比較して分析する．

本実験において，目視では誤りであると判別したが，表現式を用いた検索では検出されなかった誤用表現を表 3.6 に示す．

表 3.6 表現式で検出されなかった文の表現

誤用表現	番号	検出されなかった文の表現	件数
「な」の誤用表現	1	いろいろな友たち	1
	2	大事な友たち	1
	3	特別な表演	1
	4	不知名な声	1
	5	綺麗なかざら物(正しくは「飾り物」)	1
	6	活溌な子供	1
	7	大切な想い出	1
	8	静かな曲	1
「の」の誤用表現	9	幸せいのきもち	1
	10	いろいろの経験	1
	11	暇のとき	2
	12	好きなのファン	1
「的な」の誤用表現	13	熱情的なファン	1

表 3.6 の文 3～5 と文 9 は，日本語として正しく単語が書かれていないため，検出されなかった．これら形態素情報を表 3.7 に示す．

表 3.7 正しく書かれていない単語の形態素情報

正しく書かれていない単語	形態素情報			
	単語	品詞	活用型	活用形
不知名	不	接頭詞-名詞接続		
	知名	名詞-形容動詞語幹		
表演	表	名詞-一般		
	演	未知語		
(綺麗)なかざら物	なら	名詞-一般		
	ざら	名詞-一般		
	物	名詞-接尾-一般		
幸せい	幸せ	名詞-形容動詞語幹		
	い	動詞-自立	一段	連用形

本実験で作成した表現式は，「ラッキな人」のような正しく書かれていない誤用表現をいくつか検出しが，表 3.7 のように単語を分割して形態素情報が付けられると，表現式の作成において，形態素情報を予測して設定が行うことができない．そのため，本研究で提案した手法では，日本語として正しく書かれていない単語の誤用表現は有用な検出を行えない．ただし，このような誤用は，単語の誤りであり文法的な誤りではないと考えることができる．そのため，文法的な誤りでないと考えるならば，検出の必要はない．

文 6，7 は，形態素解析器の辞書に単語が登録されていたために，誤用文を検出するように設定した表現式にマッチせず，検出されなかった．文 6 の「活溌」と文 7 の「想い出」は，一般に「活発」と「思い出」と書かれる．そのため，これらの単語を使って書かれた表現を誤りと判別した．しかし，形態素解析器の辞書には「活溌」が「名詞-形容動詞語幹」，「想い出」が「名詞-一般」と登録されていた．辞書に単語登録されていたことによって，「活溌な子供」と「大切な想い出」は，正しい品詞間の接続である「形容動詞＋助動詞(な)＋名詞」の表現となった．そのため，設定した表現式にマッチせず，検出されなかった．したがって，誤用文の検出において，誤りであると想定した単語が形態素解析器の辞書に登録されていると検出されない．

文 8 は形態素解析器が正しく単語を分割しいないかったため，検出されなかった． 文 8 は常用漢字ではない「靜」と「かな」をそれぞれ一つの単語に分割していた．単語が正しく分割されていなかったために，表現式にマッチせず，検出されなかった．したがって，文に形態素情報を付けるとき，形態素解析器が正しく単語を分割しなかった誤用文は検出されない．

文 10，11 は，形態素解析器が想定した品詞と異なる品詞を付けたために検出されなかった．文 10 の「いろいろ」と文 11 の「暇」は，一般に形容動詞の単語である．実際，これらを正しく書いた「いろいろな経験」と「暇な時」を形態素解析器に入力すると，「いろいろ」と「暇」は，どちらも形容動詞の品詞「名詞-形容動詞語幹」で返された．しかし，文 10 と文 11 においては，「いろいろ」が「副詞-助詞類接続」，「暇」が「名詞-一般」の品詞が付けられていた．これらの品詞と「の」の品詞の接続は，正しい表現になるため，誤用文を検出する表現式にマッチせず，検出されなかった．したがって，形態素解析器が想定した形態素情報と異なる情報を付けた誤用文は検出されない．

さらに，文 1，2 と文 12，13 は検出されなかったが，以下のように表現式を改良することで検出が可能であることが考えられる．本実験において作成した表現式は「単語＋な/の＋単語」を想定した．「いろいろな友たち」のような誤用表現は，「単語＋な＋単語＋単語」に単語が分割されていた．このことから，表現式にさらに単語を加えて設定する．また，「的な」の表現式は，「的」の前の品詞を「名詞-形容動詞語幹」に設定した．「熱情」の品詞は「名詞-一般」が付けられていた．このことから表現式にさらに「名詞-一般」の品詞を加えて設定する．このように表現式の設定を改良することで，検出される文はある程度増えることが予想されるが，マッチする誤用文を増やすことができる．

加えて，本章で作成した表現式は，形容動詞の名詞修飾表現における自動分類器としての側面を持つ．3.2.6 で述べたように，機械が特定の文を捜して検出するという作業は，自動で分類を行っていると捉えることができる．そのため，本章で作成した表現式は，形容動詞の名詞修飾表現における自動分類器である．

3．3　品詞の接続法の誤りがある誤用文検出の検討

実験の結果，目視と比較して日本語として単語が正しく書かれていない誤用文や形態素解析器と辞書の問題などから検出できない誤用文があったが，品詞の接続法に誤りがある形容動詞の名詞修飾表現を検出することができた．

本実験において，表現式検出法を用いた検出は，効率よく誤用文を検出した．実験では，学習者が書いた 2,185 文を使用した．目で判別して誤用文を取り出す方法は，文を一つずつ確認する必要がある．また，表現式検出法を用いた検出は，一般的な検出手法である文字列検索よりも，一度の検索で多様な文を検出したため，効率のよい誤用文の検出が行えた．

表現式検出法による誤用文の検出は，他の品詞の接続法の誤りがある文法的な誤用文においても有効に用いることができると考える．実験で検出した形容動詞の名詞修飾表現は，品詞の接続法が違っている日本語として正しくない文法的な誤用文であった．このような誤用文は，たとえば，1.2.7 項の文献で示されていた「あの背が高いの人」も「形容詞＋の＋名詞」のような表現式を作成して検索することで検出が可能であると考える．そのため，他の文法的な誤用文も有効な検出が行えると予測する．

また，1.2.7 項で述べたように，作成した表現式は，同じ環境であれば，他の日本語学習者コーパスで再利用が可能である．実験では茶筌と ipadic 辞書を使って，台湾人日本語学習者コーパスに形態素情報を付けた．本実験と同じ環境で他の学習者コーパスに形態素情報を付けて，実験で作成した表現式で検索すると，他のコーパスのなかに表現式とマッチする誤用文があれば，有効な検出が行える．特に，学習者は母語る誤りに傾向があることから，同じ母語の学習者のデータにおいて有効な検出ができる．

さらに，表現式における形態素情報の組み合わせは，柔軟に設定できることから，他のユーザーも形態素情報を把握して文にマッチする表現式を正しく作ることができれば，誤用文を検出することができる．また，考察において示したように，表現式を改良することで，検出される文は増えるものの，一つの表現式でさらに多様な誤用文を検出することができることが考えられる．

以上のことから，品詞間に接続の誤りがある誤用文おいて，表現式検出法の有効性が確認できた．

3．4　本章のまとめ

本章では，表現式検出法を用いて品詞法に接続の誤りがある誤用文の検証を行い，有効性を確認した．品詞の接続の誤りがある誤用文の例には，形容動詞の名詞修飾表現に着目して実験を行った．実験の結果，誤用文が検出され，有効性が確認できた．

また実験の結果と分析をふまえて，品詞の接続法に誤りがある誤用文における表現式検

出法について検討した．検討した結果，他の品詞の接続法に誤りがある誤用文においても有効であることが予測でき，また効率的に誤用文を取り出せることが考えられた．

第 4 章

従来の分類において同類である誤用文の再現

4. 1　本章のあらまし

本章では，人が目視で確認した分類で同類とした誤用文を表現式検出法で再現できるか検証する．

3.2.6 項で述べたように，表現式検出法は機械が自動で誤用文を分類して検出すると考えることができる．従来の方法では，誤用を研究する研究者らが，誤用文に対して基準を設けるとともに，目視で確認ししながら分類作業を行っている．この人が目で見て分類したものに，本研究の機械による分類が再現できるかを本章では検証する．

4. 2　表現式検出法の分類と目視による分類

3.2.6 項で述べたように，表現式検出法による誤用文の検出は，誤用文を機械が見つけるとともに，それを分類していると考えることができる．これは，作成した表現式は，いわば分類器のようなものであると捉えることができるからである．また，機械による分類であることから，分類作業は客観的に行われている．

従来，誤用を研究する研究者らが用いている方法は，特定の誤りの基準を定めて，その基準に従って，分類を行う手法が行っていた．この方法は汎用性があるが，分類過程において，他の基準以外の要素が入り込む可能性があり，客観性が保てていない可能性がある．

しかし，機械による分類は客観的ではあるものの，厳密性が高いために，人が目で見て判断するように柔軟に対応できないかもしれない．そこで，表現式検出法を用いた手法が目視で判別して分類した誤用文を再現できるか検証する．検証の例として，テ形と呼ばれる動詞の活用を誤った誤用文で実験を行う．

4. 2. 1　動詞の活用誤りとテ形における学習者の誤用

動詞の活用を誤った誤用文は，「日本へ来った時」のような来るの活用形「た」の誤り，「駅のぼしょ(場所)を聞かられました」の「聞かられました」のような「れる/られる」の誤り，「早く終わせて」のような「せる/させる」などがある[46]．また，4.2 節の「じしょを

買いて千円はらいました」の「買いて」のように，音便形の活用の形が不正確な誤りがある．この「て」は日本語教育ではテ形と呼ばれる動詞の活用の誤りである．

日本語教育では動詞の連用形に接続助詞「て」が付いた形をテ形と呼んでいる[69]．テ形は「食べてください」などの依頼・勧誘，「食べています」，「食べてしまった」といった時間に関する表現などに用いられる[69]．

日本語教育における動詞活用の規則は，日本人向けの学校文法とは異なる．学校文法では，活用のし方によって，動詞を五段，上一段，下一段，カ変，サ変に分けている．日本語教育では，動詞の種類をⅠグループ(五段)，II グループ(上一段・下一段)，III グループ(カ変・サ変)の三つに分け，グループ毎の活用変化を指導する[70]．テ形は音便化であり，Ⅰグループのテ形は語末の音節により変化形が複雑に変わる．そのため，誤用も多い．たとえば，作文において「買って」を「買いて」，「歩いて」を「歩って」のようにする誤用はよく見られる．

4．2．2　実験に使用するデータ

1.2.2 項で述べたように， 3 章と同様のデータを使用する．

4．2．3　正しいテ形表現の調査とその結果

3 章と同様の目的で，データにおけて，正しく使用されたテ形を調査した．調査方法も 3 章と同様に，目視で確認した．

調査の結果， 1046 件の正しいテ形の使用が確認できた．うち，「やってる」，「してる」のように口語で使用されたテ形が 17 件あった．

以上の結果から，実験で使用するデータにおいて，テ形が使用されていることが確認できた．

4．2．4　誤用表現とその判別基準

3 章と同様に，一文中にあるテ形の活用の誤りを誤用表現とする．

また，3 章と同様に，誤用表現を判別する基準を設ける．本実験では，次の基準を動詞のテ形における誤用表現とした．

- ルールにそぐわない活用をさせたもの

例：正「歩いて」　→　誤「歩って」

- 常用漢字表[70]と異なる漢字の使用と送り仮名のもの

  漢字の例：正「頼んで」　→　誤「賴んで」

  送り仮名の例：正「暮らして」　→　誤「暮して」

- サ変動詞に関する誤り

  例 1：正「びっくりして」　→　誤「ぴっくりして」

  例 2：正「黙って」　→　誤「默して」

サ変動詞に関する誤りは，例 1 のように，厳密にはテ形の誤りではなく「して」に接続する単語の誤りであると判別することができる．しかし，例 2 の「默して」のような誤りは，教師が添削を加える場合，「黙して(もくして)」とするのではなく，「黙って(だまって)」と修正する．つまり，例 2 のような誤りは，単純に単語の誤りであると判別できない．このような単語の誤りと判別できない例が少なくないと予想されるため，サ変動詞に関する誤りもテ形の誤用として判別することにする．

4．2．5　目視でテ形の誤用文を判別した結果

3 章と同様に，目視でテ形の誤用表現が使用された文を判別した．判別の結果を表 4.1 に示す．

基準にしたがって，目視で文を判別した．

判別の結果，使用したデータにはテ形の誤用表現と判別されるものが文中に書かれていた．誤用文と誤用表現の件数を表 4.1 に示す．

表 4.1 目視で判別したテ形の誤用文と誤用表現の件数

誤用文の件数	誤用表現の件数
88	90

表 4.1 の結果から，使用したデータにはテ形の誤用文が存在していた．また，誤用表現は，一文中に一つ以上存在する文があったしていた．

4. 2. 6　表現式作成のためのテ形活用語尾の参照

4.2.1 項で述べたように，テ形は「て」となる音便形の活用の形の活用である．つまり，これらの正しい音便形の形は「て」に関するものとなる．

目視による確認では 3 種類の基準を設けたが，これらはいずれも「て」に関するものであり，学習者の誤用も「て」に関する誤りを 4.2.5 項で確認された．そこで，表現式においても「て」に関する活用語尾の文字列を参照して作成することにした．表 4.2 に「て」に関する活用語尾の誤用表現を調べた結果を示す．

表 4.2 「て」に関する活用語尾の誤用表現

番号	「て」に関する活用語尾の誤用表現	誤用表現の例	件数
1	特定の文字列と「て」の誤用表現	過て	5
2	「って」の誤用表現	歩って	8
3	「で」の誤用表現	泣いで	11
4	「て」の誤用表現	入いて	66

表 4.2 を参照して，それぞれの表現式を作成する． 表現式は，それぞれの誤用表現と一対一に対応させて作成するのではなく，形態素情報の組み合わせや正規表現を利用して，できるだけまとめたものを作成して目視の分類を再現する．

4. 2. 7　目視の分類を再現するテ形誤用文の表現式

「て」に関する活用語尾の文字列を参照して表現式を作成した．表 4.3 に示す．

表 4.3 テ形の誤用文を検出する表現式の種類

番号	テ形の誤用文を検出する表現式の種類	表現式の件数
1	特定の文字列と「て」の誤用表現を検出する表現式	2
2	「って」の誤用表現を検出する表現式	2
3	「で」の誤用表現を検出する表現式	6
4	「て」の誤用表現を検出する表現式	20

表 4.3 は，「て」に関する活用語尾を参照して作成したテ形の誤用表現を検出する表現式であるが，これが本研究におけるテ形の分類となる．また，検索を行うことで，それぞれのテ形の誤用表現とその他の表現を区別する分類器となる．

以下に本研究の分類である表現式を示す．また，テ形の分類器であるの表現式の作成に使用した正規表現の一覧を表 4.4 に示す．

■ 特定の文字列と「て」の誤用表現を検出する表現式

表 4.3 で示した特定の文字列と「て」を検出する 2 件の表現式を作成した．

- **特定の文字列と「て」の表現式 1**：

表現式 4.1 は，表 4.2 の番号 1 のうち，次の 4 件の誤用表現を対象に検出する式である．

*W*_[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+て *POS*_動詞-自立 *CT*_(五段・タ行|一段)*CF*_(命令ｅ|連用形) (4.1)

- 出してしまて
- 泊まて
- 育てきた
- 休日を過て

表現式 4.1 は，一つのユニットによるものである．

■ 表現式 4.1 のユニット

- *W*_[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+て

は，表現式 3.1 と同様に，正規表現の漢字・平仮名・片仮名・全角大文字アルファベット・全角小文字アルファベット・全角アラビア数字・半角大文字アルファベット・半角小文字アルファベット・半角アラビア数字が 1 回以上を表す．これに文字列「て」と設定して，特定の文字列と「て」の単語にマッチするように設定した．

単語の文字列に続く，

- *POS*_動詞-自立

は，表現式 3.1 と同様に，品詞を自立語動詞に設定した．表 4.2 の 4 件の品詞はいずれも自

立語動詞で返されていたからである.

品詞に続く,

- *CT_*(五段・タ行|一段)

は，自立語動詞の活用型であり,「五段・タ行」と「一段」を「( )」の範囲で指定して *OR* を設定した．表 4.2 の 4 件の活用型は,「五段・タ行」または「一段」で返されていたからである．活用型「五段・タ行」は「持つ」など五段タ行で「助詞 接続助詞」の「て」に接続するときに促音便になるもの,「一段」は「着る」など上一段・下一段活用の活用型を表す[39].

活用型に続く,

- *CF_*(命令ｅ|連用形)

は，自立語動詞の活用形であり,「命令ｅ」と「連用形」を「( )」の範囲で指定して *OR* を設定した．表 4.2 の 4 件の活用型は,「命令ｅ」または「連用形」で返されていたからである．活用形「命令ｅ」は「読め」や「(とは)いえ」など五段動詞の命令形，文語已然形，一段動詞の語幹止め命令用法を表し[39],「連用形」は動詞の語幹を「～ます」と接続したとき，五段動詞と上一段動詞は「い段」，下一段動詞は「え段」となる活用形である．たとえば，五段動詞「読む」は「読みます」になり「み」は「い段」，上一段動詞「着(き)る」は「着(き)ます」になり「着(き)」は「い段」になる．また，下一段動詞「食べる」は「食べます」になり「べ」は「え段」になる．さらに，カ変動詞「来(く)る」は「来(き)ます」の「き」，サ変動詞「する」は「します」の「し」になる.

- **特定の文字列と「て」の表現式 2**:

表現式 4.2 は，表 4.2 の番号 1 のうち，次の 1 件の誤用表現を対象に検出する式である.

*W_*[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+て *POS_*動詞-自立 *CT_*一段 *CF_*未然形 (4.2)

- 金があまり持てない

表現式 4.2 は，一つのユニットによるものである.

■ 表現式 4.2 のユニット

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+て

は，表現式 4.1 と同様に，特定の文字列と「て」の単語にマッチするように設定した.

単語の文字列に続く，

- *POS_*動詞-自立

は，品詞を自立語動詞に設定した.「持て」の品詞は自立語動詞で返されていたからである.

品詞に続く，

- *CT_一段*

は，動詞の活用型「一段」に設定した.「持て」の活用型が「一段」で返されていたからである.

活用型に続く，

- *CF_未然形*

は，動詞の活用形「未然形」に設定した.「持て」の活用形が「未然形」で返されていたからである.「未然形」は五段動詞の場合，語幹を「～ない」と接続したとき「あ段」,「～う」と接続させたとき「お段」となる活用形である．たとえば,「読む」は「読まない／読もう」になり「ま」は「あ段」,「も」は「お段」となる．また，一段動詞の場合，上一段動詞は「～ない」,「～よう」と接続したとき「い段」，下一段動詞と接続は「え段」になる．たとえば，上一段動詞「着(き)る」は「着(き)ない／着(き)よう」になり「着(き)」は「い段」,「食べる」は「食べない／食べよう」になり「べ」は「え段」になる．さらに，カ変動詞「来(く)る」は「来(こ)ない／来(こ)よう」の「こ」，サ変動詞「する」は「しない／しよう」の「し」になる.

■ 「って」の誤用表現を検出する表現式

表 4.3 で示した「って」の誤用表現を検出する 2 件の表現式を作成した.

- 「って」の表現式 1：

表現式 4.3 は，表 4.2 の番号 2 のうち，次の 6 件の誤用表現を対象に検出する式である.

POS_*(名詞[\-一-㍊ァ-ヶ]+([\-一-㍊ァ-ヶ]+)|助詞-[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+|未知語)*CT_CF_W_*って *POS_*助詞-格助詞-連語 (4.3)

- 見って
- 歩って
- 連って
- 戻って
- 黙って
- 年齢をしたかって(正しくは「したがって」)

これらの誤用表現は,「って」の直前で分割されていた単語の品詞が名詞，助詞，未知語のいずれかで返されていた．そこで， 表現式 4.3 は二つのユニットから構成した．

■ **表現式 4.3 のユニット 1**

- *POS_*(名詞[\-一-纊ァ-ヶ]+([\-一-纊ァ-ヶ]+)*|助詞-[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+|未知語)*CT_CF_*

の品詞フラグ *POS_*を「( )」の範囲で指定し，名詞，助詞，未知語を *OR* の「|」とした．また，これらの品詞は活用型・活用形が存在しないため，設定を行わず「*CT_CF_*」とした．品詞の名詞,

- 名詞-[一-纊ァ-ヶ]+(-[一-纊ァ-ヶ]+)*

は, 表現式 3.1 と同様に, 名詞とその階層で使用される文字列にマッチするように設定した．名詞に続く *OR* の品詞,

- 助詞-[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，助詞にマッチさせる設定である．助詞の階層「-」に続く「[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]」は，表現式 3.1 と同様の文字列の意味を表す．これらの文字列を 1 回以上の「+」に設定した．

助詞に続く *OR* の品詞,

- 未知語

は，表現式 3.1 と同様に，辞書にない単語で返された品詞である．

■ **表現式 4.3 のユニット 2**

- *W_*って *POS_*助詞-格助詞-連語

は，単語の文字列「って」とその品詞を表す． 6 件の誤用表現は，いずれも「って」で分

割され，品詞が「助詞-格助詞-連語」で返されていたため，「*W_*って *POS_*助詞-格助詞-連語」に設定した．

先頭の，

- *W_*って

は，単語の文字列を表す．

文字列に続く，

- *POS_*助詞-格助詞-連語

は，「って」の品詞を表す．「助詞-格助詞-連語」は「って」，「にしたがって」，「に関して」など，格助詞と動詞との連語で，主に格助詞に相当するような働きを持つ品詞である[39]．

- **「って」の表現式 2**：

表現式 4.4 は，表 4.2 の番号 2 のうち，次の 2 件の誤用表現を対象に検出する式である．

*POS_*動詞-自立 *CT_*[一-纔ァ-ヶ]+*CF_*命令ｅ *W_*って *POS_*助詞-格助詞-連語 (4.4)

- 彼氏とわかれって
- 電話をかけって

表現式 4.4 は三つのユニットから構成されている．

■ 表現式 4.4 のユニット 1

- *POS_*動詞-自立 *CT_*[一-纔ァ-ヶ]+*CF_*命令ｅ

は，「って」の直前で分割された単語の品詞，活用型，活用形である．

単語の品詞，

- *POS_*動詞-自立

は，2 件ともに自立語動詞で返されていたため，「動詞-自立」に設定した．

品詞に続く，

- *CT_*[一-纔ァ-ヶ]+

は，動詞の活用型における文字列である．動詞の活用型は漢字と片仮名の文字列が使用されているため[39]，「[一-纔ァ-ヶ]」にし，それらの文字列が 1 回以上の「+」を設定した．

活用型に続く，

- *CF*_命令ｅ

は，活用形は動詞の活用形である．2 件の動詞は，いずれも「命令ｅ」で返されてため，「命令ｅ」を設定した．

■ **表現式 4.4 のユニット 2**

- *W*_って *POS*_助詞-格助詞-連語

は，表現式 4.3 と同様に，単語の文字列「って」とその品詞を表す．2 件の誤用表現の「って」は，「*W*_って *POS*_助詞-格助詞-連語」で返されていたため，これらの形態素情報を設定した．

■ 「で」の誤用表現を検出する表現式

表 4.3 で示した「で」の誤用表現を検出する 6 件の表現式を作成した．

- **「で」の表現式 1**：

表現式 4.5 は，表 4.2 の番号 3 のうち，次の 1 件の誤用表現を対象に検出する式である．

*POS*_未知語 *CT_CF_W*_で *POS*_助詞-格助詞-一般　　(4.5)

- 家族と一緒に遊で

表現式 4.5 は，二つのユニットから構成されている．

■ **表現式 4.5 のユニット 1**

- *POS*_未知語 *CT_CF*_

は，「って」の直前で分割された単語「遊」の形態素情報である．「遊」は未知語で返されていた．そこで，表現式 4.5 と同様に，「*POS*_未知語 *CT_CF*_」に設定した．

■ **表現式 4.5 のユニット 2**

- *W_*で *POS_*助詞-格助詞-一般

は,「遊」に接続した「で」とその品詞である．未知語「遊」に接続した「で」は「助詞-格助詞-一般」で返されていたため,「*W_*で *POS_*助詞-格助詞-一般」に設定した．品詞「助詞-格助詞-一般」は,「で」,「が」,「から」,「と」など，一般的な格助詞である[39].

- **「で」の表現式 2**：

表現式 4.6 は，表 4.2 の番号 3 のうち，次の 2 件の誤用表現を対象に検出する式である．

*POS_*未知語 *CT_CF_W_*[一-㍻ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+
*POS_*名詞-非自立-一般 *CT_CF_W_*で *POS_*助動詞 *CT_*特殊・ダ *CF_*連用形　(4.6)

- 病室に送って下さいと頼んで
- 何人も頼んで

こららの誤用表現は，いずれも「頼んで」の誤用文である．表現式 4.6 は，三つのユニットから構成されている．

■ **表現式 4.6 のユニット 1**

- *POS_*未知語 *CT_CF_*

は,「頼」の形態素情報である．「頼」は未知語で返されていた．そこで，表現式 4.5 と同様に,「*POS_*未知語 *CT_CF_*」に設定した．

■ **表現式 4.6 のユニット 2**

- *W_*[一-㍻ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+ *POS_*名詞-非自立-一般 *CT_CF_*

は,「ん」のにマッチさせる形態素情報である．
単語の文字列,

- *W_*[一-㍻ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，表現式 3.1 と同様であり，これらの文字列が 1 回以上の「+」を設定した．誤用表現は「ん」の一文字であるが，表現式 4.6 では，全ての単語の文字列にマッチさせるため,

「*W_*[一-龠ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+」に設定した.

単語の文字列に続く,

- *POS_*名詞-非自立-一般 *CT_CF_*

は,「ん」で返された品詞，活用型，活用形である．品詞「名詞-非自立-一般」は，連体詞,「の（格助詞)」，活用語の基本形に接続して使われるもののうち,「名詞-非自立」の下位階層に当てはまらないものであり，形式名詞「こと」,「もの」,「ところ」や口語の「ん」,「もん」などの品詞である[39]．また，名詞は活用型と活用形がないため，表現式では設定を行わなかった.

■ **表現式 4.6 のユニット 3**

- *W_*で *POS_*助動詞 *CT_*特殊・ダ *CF_*連用形

は,「頼ん」に接続した「で」の形態素情報を表す.

単語の文字列,

- *W_*で

は,「で」を表す.

文字列に続く.

- *POS_*助動詞 *CT_*特殊・ダ

は,「で」の品詞と活用形を表す．品詞「助動詞」と活用型「特殊・ダ」は，表現式 3.1 の「な」と同様の形態素情報である.

品詞と活用型に続く,

- *CF_*連用形

は,「で」の活用形である.「で」は断定を表す助動詞「だ」の連用形であり,「彼は学生で 18 歳だ」,「簡単で便利だ」など,「で」の前後に名詞と形容動詞が接続する.

- **「で」の表現式 3**：

表現式 4.7 は，表 4.2 の番号 3 のうち，次の 3 件の誤用表現を対象に検出する式である.

*POS_*(動詞-自立|形容詞-自立)*CT_*[一-龠ァ-ヶ]+*CF_*(連用タ接続|連用形|基本形)*W_*で *POS_*(助動詞|助詞-格助詞-一般) (4.7)

- 先祖様の前にひざまずいで
- 出で
- 近づいで

表現式 4.7 は，二つのユニットから構成されている．

■ **表現式 4.7 のユニット 1**

- *POS_*(動詞-自立|形容詞-自立)*CT_*[一-纊ァ-ヶ]+*CF_*(連用タ接続|連用形|基本形)

は，「で」の直前で分割された単語の品詞，活用型，活用形である．

単語の品詞，

- *POS_*(動詞-自立|形容詞-自立)

は，表現式 3.1 と同様に，自立語動詞と自立語形容詞を表す．「出」と「近づい」が自立語動詞，「ひざまずいで」の「まずい」が自立語形容動詞で返されていた．そこで，「()」で品詞の範囲を指定し，*OR* で「動詞-自立|形容詞-自立」に設定した．

品詞に続く，

- *CT_*[一-纊ァ-ヶ]+

は，自立語動詞と自立語形容詞の活用型を表す．これらの活用型の文字列は漢字と片仮名が使用されているため，「[一-纊ァ-ヶ]」にし，文字列が 1 回以上の「+」を設定した．

活用型に続く，

- *CF_*(連用タ接続|連用形|基本形)

は，自立語動詞と自立語形容詞の活用形を表す．「まずい」が「基本形」，「出」が連用形，「近づい」が「連用タ接続」で返されていた．そこで，「( )」で範囲を指定し，*OR* で「連用タ接続|連用形|基本形」に設定した．「連用タ接続」は「書いた／書いて」，「読んだ／読んで」など「～た／～て」に接続する品詞の活用形である[39]．

■ **表現式 4.7 のユニット 2**

- *W_*で *POS_*(助動詞|助詞-格助詞-一般)

は，「ひざまずいで」，「出で」，「近づいで」の「で」に返された形態素情報である．

単語の文字列,

- *W_*で

は，表現式 4.6 と同様に，「で」を表す.

文字列に続く.

- *POS_*(助動詞|助詞-格助詞-一般)

は,「で」の品詞を表す．品詞は「ひざまずいで」と「近づいで」の「で」が「助動詞」,「出で」の「で」が「助詞-格助詞-一般」で返されていた．そこで,「()」で範囲を指定し，*OR* で「助動詞|助詞-格助詞-一般」に設定した.

- **「で」の表現式 4**：

表現式 4.8 は，表 4.2 の番号 3 のうち，次の 1 件の誤用表現を対象に検出する式である.

*POS_*動詞-自立 *CT_*五段・[一-纔ァ-ヶ]+音便 *CF_*連用タ接続 *W_*で *POS_*助詞-接続助詞 (4.8)

- 泣いで

表現式 4.8 は，二つのユニットから構成されている.

■ **表現式 4.8 のユニット 1**

- *POS_*動詞-自立 *CT_*五段・[一-纔ァ-ヶ]+音便 *CF_*連用タ接続

は,「で」の直前で分割された「泣い」の品詞，活用型，活用形である.

単語の品詞,

- *POS_動詞*-自立

は,「泣い」が自立語動詞で返されていたため,「動詞-自立」に設定した.

品詞に続く,

- *CT_*五段・[一-纔ァ-ヶ]+音便

は，動詞の活用型である.「泣い」は「五段・カ行イ音便」で返されていた.「五段」は五段活用の動詞を表し,「・」は活用型の下位階層の区切りを表す．たとえば,「泣い」で返されていた「五段・カ行イ音便」は,「解く」,「聞く」など,「助詞-接続助詞」の「て」に接続するときにイ音便になるカ行の五段活用動詞を表す[39].「・」以下の下位階層では,

漢字と片仮名の文字列が使用されているため,「[一-龥ァ-ヶ]」とし，これらの文字列が 1 回以上の「+」を設定した．誤用表現は「泣い」の単語のみであるが，表現式 4.8 では全ての五段活用の音便にマッチさせるため,「五段・[一-龥ァ-ヶ]+音便」に設定した．

活用型に続く,

- *CF*_連用タ接続

は,「泣い」で返された活用形にマッチさせる設定である．「連用タ接続」は，表現式 4.7 と同様の活用形であり,「～た／～て」に接続する．

■ **表現式 4.8 のユニット 2**

- *W*_で *POS*_助詞-接続助詞

は,「泣いで」の「で」に返された形態素情報である．

単語の文字列,

- *W*_で

は，表現式 4.6 と同様に,「で」を表す．

文字列に続く．

- *POS*_助詞-接続助詞

は,「で」の品詞を表す．「助詞-接続助詞」は,「で」,「と」,「が」などの助詞で，主に用言に接続して前後の文の意味的な関係を表す．たとえば,「見た目は悪いが，うまい」の「が」は，逆接の意味を表す．

- **「で」の表現式 5**：

表現式 4.9 は，表 4.2 の番号 3 のうち，次の 3 件の誤用表現を対象に検出する式である．

*POS*_(名詞-一般|助詞-格助詞-一般)*CT_CF_W*_[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]{3}*POS*_名詞-一般 *CT_CF_W*_で *POS*_助詞-格助詞-一般　(4.9)

- 歩いで
- 外で住でいました
- 癌で死でしまった

表現式 4.9 は，三つのユニットから構成されている．

■ **表現式 4.9 のユニット 1**

- *POS_*(名詞-一般|助詞-格助詞-一般)*CT_CF_*

は，「歩いで」の「歩」，「外で」の「で」，「癌で」の「で」の品詞・活用型・活用形を表す．「歩」が「名詞-一般」，「で」が「助詞-格助詞-一般」で返されていた．そこで，「( )」で品詞の範囲を指定し，*OR* で「名詞-一般|助詞-格助詞-一般」に設定した．また，「名詞-一般」と「助詞-格助詞-一般」は活用しないため，活用型と活用形は設定をしない．

■ **表現式 4.9 のユニット 2**

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]{3}*POS_*名詞-一般 *CT_CF_*

は，「で」の直前で分割された「歩いで」の「い」，「住で」の「住」，「死で」の「死」の形態素情報にマッチさせる．

先頭の

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]{3}

は，単語の文字列を表す．表現式 4.9 においては，まず，「[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]」ですべての文字列を正規表現で指定し，次に「{3}」で文字列が 3 回以上の繰り返しに設定した．

単語の文字列に続く，

- *POS_*名詞-一般 *CT_CF_*

は，品詞，活用形．活用型を表す．「歩いで」の「い」，「住」，「死」は，いずれも「名詞-一般」で返されていた．そこで，品詞に「名詞-一般」を設定した．また，は活用しないため，活用型と活用形は設定をしない．

■ **表現式 4.9 のユニット 3**

- *W_*で *POS_*助詞-格助詞-一般

は，「歩いで」，「住で」，「死で」の「で」の形態素情報である．

単語の文字列，

- *W_*で

は，表現式 4.6 と同様に，「で」を表す．

文字列に続く.

- *POS*_助詞-格助詞-一般

は,「で」の品詞を表す. 品詞「助詞-格助詞-一般」は,「が」,「で」,「を」などの格助詞である. 格助詞は主に体言に接続して文中の関係を示す助詞である. たとえば,「本を読む」の「を」は, 動詞「読む」の対象を表す.

- **「で」の表現式 6**:

表現式 4.10 は, 表 4.2 の番号 3 のうち, 次の 1 件の誤用表現を対象に検出する式である.

*POS*_助詞-格助詞-一般 *CT_CF_W*_[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_名詞-一般 *CT_CF_W*_で *POS*_助詞-格助詞-一般 *CT_CF_W*_[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_助詞-係助詞 (4.10)

- 年を流れでも

表現式 4.10 は, 四つのユニットから構成されている.

■ **表現式 4.10 のユニット 1**

- *POS*_助詞-格助詞-一般 *CT_CF_*

は,「年を」の「を」の品詞・活用型・活用形にマッチさせる. 品詞「助詞-格助詞-一般」は, 表現式 4.9 と同様に, 格助詞を表す. 格助詞は活用しないため, 活用型と活用形の設定をしない.

■ **表現式 4.10 のユニット 2**

- *W*_[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_名詞-一般 *CT_CF_*

は,「で」の直前で分割された「流れ」の形態素情報にマッチさせる.
先頭の,

- *W*_[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は, 表現式 3.1 と同様に, 文字列を表す. また, 表現式 4.6 と同様に, 字列が 1 回以上の「+」を設定した.

単語の文字列に続く，

- *POS_*名詞-一般 *CT_CF_*

は，「流れ」の品詞・活用型・活用形である．「流れ」は「名詞-一般」で返され，「名詞-一般」は活用しないため，「*POS_*名詞-一般 *CT_CF_*」と設定した．

■ **表現式 4.10 のユニット 3**

- *W_*で *POS_*助詞-格助詞-一般 *CT_CF_*

は，表現式 4.9 と同様に，「で」の形態素情報にマッチさせる設定である．

■ **表現式 4.10 のユニット 4**

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*助詞-係助詞

は，「流れでも」の「も」にマッチさせる設定である．

先頭の，

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，表現式 4.6 と同様であり，全ての単語の文字列にマッチさせる設定をした．

単語の文字列に続く，

- *POS_*助詞-係助詞

は，「も」の品詞にマッチさせる設定である．ipadic では，「は」，「も」，「しか」などは「助詞-係助詞」としている[39]．係助詞は述語に関わる語に付いて意味を添える助詞である．たとえば，「彼は学生だ」の「は」は主題を表し，「本は買ったが，ノートは買わなかった」の「は」は対比を表す．

■ **「て」の誤用表現を検出する表現式**

表 4.3 で示した「て」の誤用表現を検出する 21 件の表現式を作成した．

- **「て」の表現式 1**：

表現式 4.11 は, 表 4.2 の番号 4 のうち, 次の 1 件の誤用表現を対象に検出する式である.

*POS*_未知語 *CT_CF_W*_[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+
*POS*_動詞-非自立 *CT*_五段・[一-纊ァ-ヶ]+音便 *CF*_連用タ接続
*W*_[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_動詞-自立 *CT*_五段・ラ行 *CF*_連用形 *W*_し *POS*_動詞-自立 *CT*_サ変・スル *CF*_連用形 *W*_て *POS*_助詞-接続助詞 (4.11)

- ぴっくりして

この誤用表現は,「ぴ」,「っ」,「くり」,「し」「て」の五つの単語に分割されていた．そこで，表現式 4.11 は，五つのユニットから構成した.

■ **表現式 4.11 のユニット 1**

- *POS*_未知語 *CT_CF*_

は,「ぴっくりして」の「ぴ」にマッチさせる形態素情報である．「ぴ」は未知語で返されていた．そこで，表現式 4.5 と同様に,「*POS*_未知語 *CT_CF*_」に設定した.

■ **表現式 4.11 のユニット 2**

- *W*_[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_動詞-非自立 *CT*_五段・[一-纊ァ-ヶ]+音便 *CF*_連用タ接続

は,「っ」にマッチさせる形態素情報である.
単語の文字列にマッチさせる,

- *W*_[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，表現式 4.6 と同様の設定を行った.
単語の文字列に続く,

- *POS*_動詞-非自立

は,「っ」の品詞にマッチさせる設定である．「動詞-非自立」は，補助動詞など，動詞に接続して文法的な機能を果たす動詞である．たとえば,「ついて行く」の「行く」が補助動詞であり，ipadic では「動詞-非自立」の品詞とされている[39].

品詞に続く，

- *CT_*五段・[一-龥ァ-ヶ]+音便

は，表現式 4.8 と同様の活用型にマッチさせる設定である．

活用型に続く，

- *CF_*連用タ接続

は，表現式 4.8 と同様の活用形にマッチさせる設定である．

■ **表現式 4.11 のユニット 3**

- *W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*五段・ラ行 *CF_*連用形

は，「くり」にマッチさせる形態素情報である．

単語の文字列にマッチさせる，

- *W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，表現式 4.6 と同様の設定を行った．

単語の文字列に続く，

- *POS_*動詞-自立

は，品詞にマッチさせる設定である．自立語動詞にマッチさせる．

品詞に続く，

- *CT_*五段・ラ行

は，活用型にマッチさせる設定である．「五段・ラ行」は，「切る」，「なる」など，「助詞-接続助詞」の「て」に接続するときに促音便する五段動詞ラ行の活用型である[39]．

活用型に続く，

- *CT_*連用形

は，活用形にマッチさせる設定である．連用形にマッチさせる．

■ **表現式 4.11 のユニット 4**

- *W_*し *POS_*動詞-自立 *CT_*サ変・スル *CF_*連用形

は，「し」にマッチさせる形態素情報である．

単語の文字列,

- *W_*し

は,「し」にマッチさせる.

単語の文字列に続く,

- *POS_*動詞-自立

は，品詞にマッチさせる設定である．自立語動詞にマッチさせる.

品詞に続く,

- *CT_*サ変・スル

は，活用型にマッチさせる設定である.「サ変・スル」は，自立語動詞「する」とサ変名詞「名詞-サ変接続(たとえば,「インプット」などの単語で,「する」,「できる」などと後接できる名詞)」に接続する「する」の活用型である[39].

活用型に続く,

- *CT_*連用形

は，活用形にマッチさせる設定である．連用形にマッチさせる.

■ **表現式 4.11 のユニット 5**

- *W_*て *POS_*助詞-接続助詞

は,「て」にマッチさせる形態素情報である.

単語の文字列,

- *W_*て

は,「て」にマッチさせる.

単語の文字列に続く,

- *POS_*助詞-接続助詞

は，表現式 4.8 と同様に，品詞「助詞-接続助詞」にマッチさせる設定である.

- 「て」の表現式 2：

表現式 4.12 は，表 4.2 の番号 4 のうち，次の 10 件の誤用表現を対象に検出する式である．

POS_*(名詞-一般|名詞-固有名詞-[一-纊]+(-[一-纊]+)|未知語)*CT_CF_W_*し *POS_*動詞-自立 *CT_*サ変・スル *CF_*連用形 *W_*て *POS_*(助詞-接続助詞|動詞-非自立)*CT_*(.*|一段)*CF_*(.*|連用形)　(4.12)

- 充實して
- 默して
- 联絡して
- 迷路して
- 尋して
- 教して(2 件あり)
- 滿意して
- 毎日トキトキして(正しくは「ドキドキして」)
- バービーキュして(正しくは「バーベキューして」)

表現式 4.12 は，三つのユニットから構成されている．

■ 表現式 4.12 のユニット 1

- *POS_*(名詞-一般|名詞-固有名詞-[一-纊]+(-[一-纊]+)*|未知語)*CT_CF_*

は，「して」の「し」の直前で分割された単語の品詞，活用型，活用形にマッチさせる設定である．「し」の直前で分割された単語の品詞は，「名詞-一般」，「名詞-固有名詞」，「未知語」のいずれかで返されていた．そこで，「( )」で範囲を指定し，*OR* の設定を行った．また，これらの品詞は活用しないため，活用型と活用形の設定はしない．

「名詞-固有名詞」おける，

- 名詞-固有名詞-[一-纊]+(-[一-纊]+)*

の設定は，固有名詞の下位階層にマッチさせる設定である．「名詞-固有名詞」は，人名，地名，組織名などが三つ以上四つ以内に下位階層されおり，それらは漢字の文字列が使用されている [39]．そこで，まず，「名詞-固有名詞-[一-纊]+」とし，三つ目の下位階層にマッ

チさせ，次に，「()*」のなかに「-[一-纊]+」を入れて，0 回以上の繰り返しにマッチするように設定した．

■ 表現式 4.12 のユニット 2

- *W_*し *POS_*動詞-自立 *CT_*サ変・スル *CF_*連用形

の形態素情報は，表現式 4.11 の単語「し」と同様の設定である．

■ 表現式 4.12 のユニット 3

- *W_*て *POS_*(助詞-接続助詞|動詞-非自立)*CT_*(.*|一段)*CF_*(.*|連用形)

は，「て」にマッチさせる設定である．

単語の文字列，

- *W_*て

は，「て」にマッチさせる．

単語の文字列に続く，

- *POS_*(助詞-接続助詞|動詞-非自立)

は，「て」の品詞にマッチさせる．「て」の品詞は「助詞-接続助詞」または「動詞-非自立」で返されていた．そこで，「()」で範囲を指定し，*OR* で「助詞-接続助詞|動詞-非自立」に設定した．

品詞に続く，

- *CT_*(.*|一段)

は，「て」の活用型を表す．品詞「助詞-接続助詞」は活用を持たない．また，「動詞-非自立」で返された誤用表現の活用型は「一段」だった．そこで，まず，「()」で範囲を指定した．次に，活用型を持たないものは「.」で任意の一文字をしてしてそれを「*」で 0 回以上の繰り返しと，活用型「一段」を入れた．そして，*OR* の設定を行った．

活用型に続く，

- *CF_*(.*|連用形)

は，「て」の活用形を表す．品詞「動詞-非自立」は「連用形」で返されていた．そこで，活用型と同じように，任意の文字を 0 回以上の繰り返いと「連用形」を「( )」に入れ，これらに *OR* を設定した．

- **「て」の表現式 3**：

表現式 4.13 は, 表 4.2 の番号 4 のうち, 次の 8 件の誤用表現を対象に検出する式である.

*POS_*(動詞-(.*自立|接尾)|名詞-固有名詞-[一-纞]+(-[一-纞]+)*|助動詞|未知語)*CT_*.**CF_*.**W_*て *POS_*(動詞-非自立|助詞-格助詞-連語) (4.13)

- つかれくて(正しくは「つかれて(疲れて)」)
- 感心くて(正しくは「感心して」)
- かわがって(正しくは「かわいがって」)
- 並んて
- 手をつぐて
- 歩くて
- 眠て
- 思て

表現式 4.13 は，二つのユニットから構成されている.

■ **表現式 4.13 のユニット 1**

- *POS_*(動詞-(.*自立|接尾)|名詞-固有名詞-[一-纞]+(-[一-纞]+)*|助動詞|未知語)*CT_*.**CF_*.*

は,「て」の直前で分割された単語の品詞，活用型，活用形にマッチさせ設定である.

「て」の直前で分割された品詞,

- *POS_*(動詞-(.*自立|接尾)|名詞-固有名詞-[一-纞]+(-[一-纞]+)*|助動詞|未知語)

は, 動詞,「名詞-固有名詞」,「助動詞」,「未知語」のいずれかで返されていた. そこで,「()」で範囲を指定し，*OR* の設定を行った．また,「名詞-固有名詞」は表現式 4.12 と同様の設定を行った．動詞で返された品詞は「動詞-自立」,「動詞-非自立」,「動詞-接尾」のいずれかであった．そこで，まず，文字列「動詞-」の後ろに「( )」で範囲を指定した．次に，表現式 4.12 の活用型・活用形と同様に，任意の一文字が 0 回以上にマッチする「.*」を文字列「自立」の直前に設定した．この設定により「自立」と「非自立」をマッチさせる．そして,「.* 自立」と「接尾」を *OR* で設定した.「動詞-接尾」は,「せる／させる」,「れる

／られる」,「がる」など, 学校文法では助動詞とされている動詞の活用語尾の品詞である[39]. 品詞に続く,

- *CT_.*CF_.**

は,「て」の直前で分割の活用型と活用形にマッチさせる設定である．動詞と「助動詞」は活用する品詞であるが,「名詞-固有名詞」,「助動詞」は活用しない．そこで,「.*」を設定した．この設定により，動詞と「助動詞」の活用型と活用形にマッチさせる．

■ **表現式 4.13 のユニット 2**

- *W_*て *POS_*(動詞-非自立|助詞-格助詞-連語)

は,「て」にマッチさせる設定である．

単語の文字列,

- *W_*て

は,「て」にマッチさせる．

単語の文字列に続く,

- *POS_*(動詞-非自立|助詞-格助詞-連語)

は,「て」の品詞にマッチさせる．「て」の品詞は「動詞-非自立」または「助詞-格助詞-連語」で返されていた．そこで,「( )」で範囲を指定し，*OR* で「動詞-非自立|助詞-格助詞-連語」に設定した．

- **「て」の表現式 4**：

表現式 4.14 は, 表 4.2 の番号 4 のうち, 次の 4 件の誤用表現を対象に検出する式である．

*POS_*名詞-一般 *CT_CF_W_*て *POS_*助詞-接続助詞　　(4.14)

- かわがれて(正しくは「かわいがられて」)
- しまて(正しくは「しまって」)
- 起こて(正しくは「起こって」)
- 大学へ行て

表現式 4.14 は，二つのユニットから構成されている．

■ **表現式 4.14 のユニット 1**

- *POS_*名詞-一般 *CT_CF_*

は,「て」の直前で分割された単語の品詞，活用型，活用形にマッチさせる設定である．これらの品詞は，いずれも「名詞-一般」で返されていた．また,「名詞-一般」は活用しない．そこで,「*POS_*名詞-一般 *CT_CF_*」と設定した.

■ **表現式 4.14 のユニット 2**

- *W_*て *POS_*助詞-接続助詞

は,「て」にマッチさせる設定である．これは表現式 4.11 の「て」にマッチさせる設定と同様である.

- **「て」の表現式 5**：

表現式 4.15 は,表 4.2 の番号 4 のうち,次の 3 件の誤用表現を対象に検出する式である.

*POS_*(形容詞-自立|未知語)*CT_*.**CF_*.**W_*がっ *POS_*動詞-接尾 *CT_*五段・ラ行 *CF_*連用タ接続 *W_*て *POS_*助詞-接続助詞 (4.15)

- かいがって(正しくは「かわいがって」)
- 可愛いがって(正しくは「可愛がって」)
- 違がって(正しくは「違って」)

表現式 4.15 は，三つのユニットから構成されている.

■ **表現式 4.15 のユニット 1**

- *POS_*(形容詞-自立|未知語)*CT_*.**CF_*.*

は，分割された単語である「かいがって」の「かい」,「可愛いがって」の「可愛い」,「違がって」の「違」にマッチさせる設定である.

単語の品詞

- *POS_*(形容詞-自立|未知語)

は,「かい」と「可愛い」が「形容詞-自立」,「違」が「未知語」で返されていた．そこで,

「()」で範囲を指定し，*OR* の設定を行った．

品詞に続く，

- *CT_.*CF_.**

は，活用型と活用形の設定である．「形容詞-自立」は活用するが，「未知語」は活用しない．そこで，表現式 4.13 の活用型と活用形と同じ設定を行った．

■ **表現式 4.15 のユニット 2**

- *W_*がっ *POS_*動詞-接尾 *CT_*五段・ラ行 *CF_*連用タ接続

は，「かいがって」，「可愛いがって」，「違がって」の「がっ」にマッチさせる設定である．これらは，いずれも文字列「がっ」で分割され，品詞は「動詞-接尾」，活用型は「五段・ラ行」，活用形は「連用タ接続」で返されていた．そこで，「*W_*がっ *POS_*動詞-接尾 *CT_*五段・ラ行 *CF_*連用タ接続」と設定した．

■ **表現式 4.15 のユニット 3**

- *W_*て *POS_*助詞-接続助詞

は，「て」にマッチさせる設定である．これは表現式 4.11 の「て」にマッチさせる設定と同様である．

- **「て」の表現式 6**：

表現式 4.16 は，表 4.2 の番号 4 のうち，次の 1 件の誤用表現を対象に検出する式である．

*CF_*体言接続特殊２ *W_*[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*助詞-格助詞-一般 *CT_CF_W_*し *POS_*動詞-自立 *CT_*サ変・スル *CF_*連用形 *W_*て *POS_*助詞-接続助詞　(4.16)

- 探がして(正しくは「探して」)

この誤用文は「探」，「が」，「し」，「て」の四つの単語に分割されていた．そこで，表現式 4.16 は，四つのユニットから構成した．

■ **表現式 4.16 のユニット 1**

- *CF_*体言接続特殊２

は,「探」の活用形にマッチさせる設定である．活用形「体言接続特殊」は，口語の活用形であり,「る」で終わる動詞が「の」などに接続する場合の音便化である[39]．たとえば,「何するの？」は「何すんの？」と「ん」に音便化する．「体言接続特殊２」は,「体言接続特殊」の語末の「ん」が欠落した形とされている[39].

■ **表現式 4.16 のユニット 2**

- *W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*助詞-格助詞-一般 *CT_CF_*

は,「が」にマッチさせる設定である.

単語の文字列,

- *W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，表現式 4.6 と同様の設定を行った.

単語の文字列に続く,

- *POS_*助詞-格助詞-一般 *CT_CF_*

は，品詞，活用形，活用型の形態素情報である．これは表現式 4.10 の「を」と同様の設定である.

■ **表現式 4.16 のユニット 3 とユニット 4**

- *W_*し *POS_*動*詞-自立 CT_サ変・スル CF_連用形*
- *W_*て *POS_助詞-*接続助詞

は，表現式 4.11 の「し」と「て」にマッチさせる設定と同様である.

- **「て」の表現式 7**：

表現式 4.17 は, 表 4.2 の番号 4 のうち, 次の 1 件の誤用表現を対象に検出する式である.

*POS_*助詞-格助詞-一般 *CT_CF_W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*助詞-接続助詞 *CT_CF_W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-非自立 *CT_*[一-龥ァ-ヶ]+*CF_*連用形 *W_*て *POS_*助詞-接続助詞 (4.17)

- 運動がてきて

表現式 4.17 は，四つのユニットから構成されている．

■ **表現式 4.17 のユニット 1**

- *POS*_助詞-格助詞-一般 *CT_CF_*

は，「が」にマッチするように設定した．これは表現式 4.10 の「を」と同様の設定である．

■ **表現式 4.17 のユニット 2**

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_助詞-接続助詞 *CT_CF_*

は，「が」の直後の「て」にマッチさせる設定である．

単語の文字列，

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，表現式 4.6 と同様の設定を行った．

単語の文字列に続く，

- *POS*_助詞-接続助詞 *CT_CF_*

は，品詞，活用形，活用型の形態素情報である．「が」の直後の「て」は，品詞が「助詞-接続助詞」で返されていた．そこで，品詞を「助詞-接続助詞」に設定した．また，「助詞-接続助詞」は活用しないため，活用型と活用形は設定しない．

■ **表現式 4.17 のユニット 3**

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_動詞-非自立 *CT_*[一-㍊ァ-ヶ]+*CF_*連用形

は，「てきて」の「き」にマッチさせる設定である．

単語の文字列，

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，表現式 4.6 と同様の設定を行った．

単語の文字列に続く，

- *POS*_動詞-非自立

は，「き」が返された品詞「動詞-非自立」を設定した．

品詞に続く，

- *CT_*[一-纔ァ-ヶ]+

は，表現式 4.4 と同じ活用型の設定を行った．

活用型に続く，

- *CF_*連用形

は，「き」が返された活用形「連用形」を設定した．

■ **表現式 4.17 のユニット 4**

- *W_*て *POS_*助詞-接続助詞

は，「き」の直後の「て」にマッチさせる設定である．これは表現式 4.11 の「て」にマッチさせる設定と同様である．

- **「て」の表現式 8**：

表現式 4.18 は，表 4.2 の番号 4 のうち，次の 1 件の誤用表現を対象に検出する式である．

*POS_*名詞-一般 *CT_CF_W_*[一-纔ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*助詞-格助詞-一般 *CT_CF_W_*[一-纔ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*名詞-一般 *CT_CF_W_*[一-纔ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*助詞-格助詞-一般 *CT_CF_W_*[一-纔ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*一段 *CF_*連用形 *W_*て *POS_*助詞-接続助詞 (4.18)

- 最初から気に入れて(正しくは「気に入って」)

この誤用文は「最初」，「から」，「気」，「に」，「入れ」，「て」の六つの単語に分割されていた．そこで，表現式 4.18 は，六つのユニットから構成した．

■ **表現式 4.18 のユニット 1**

- *POS_*名詞-一般 *CT_CF_*

は，「最初」の形態素情報にマッチさせる設定である．これは表現式 4.14 における先頭の形態素情報と同様の設定である．

■ 表現式 4.18 のユニット 2

- *W_*[一-纞ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*助詞-格助詞-一般 *CT_CF_*

は,「から」にマッチさせる設定である．この設定は，表現式 4.16 の「が」にマッチさせる設定と同様である．

■ 表現式 4.18 のユニット 3

- *W_*[一-纞ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*名詞-一般 *CT_CF_*

は,「気」にマッチさせる設定である．この設定は，表現式 4.10 の「流れ」にマッチさせる設定と同様である．

■ 表現式 4.18 のユニット 4

- *W_*[一-纞ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*助詞-格助詞-一般 *CT_CF_*

は,「に」にマッチさせる設定である．この設定は，表現式 4.16 の「が」にマッチさせる設定と同様である．

■ 表現式 4.18 のユニット 5

- *W_*[一-纞ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*一段 *CF_*連用形

は,「入れ」にマッチさせる設定である．

単語の文字列,

- *W_*[一-纞ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，表現式 4.6 と同様の設定を行った．

単語の文字列に続く,

- *POS_*動詞-自立

は,「入れ」が返された品詞「動詞-自立」を設定した．

品詞に続く,

- *CT_*一段

は,「入れ」が返された活用型「一段」を設定した．

活用型に続く，

- *CF*_連用形

は，「入れ」が返された活用形「連用形」を設定した．

■ **表現式 4.18 のユニット 6**

- *W*_て *POS*_助詞-接続助詞

は，「入れて」の「て」にマッチさせる設定である．これは表現式 4.11 の「て」にマッチさせる設定と同様である．

- **「て」の表現式 9**：

表現式 4.19 は，表 4.2 の番号 4 のうち，次の 1 件の誤用表現を対象に検出する式である．

*POS*_未知語 *CT_CF_W*_[一-㷀ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+ *POS*_名詞-サ変接続 *CT_CF_W*_し *POS*_動詞-自立 *CT*_サ変・スル *CF*_連用形 *W*_て *POS*_助詞-接続助詞 (4.19)

- 體験して

この誤用文は「體」，「験」，「し」，「て」の四つの単語に分割されていた．そこで，表現式 4.19 は，四つのユニットから構成した．

■ **表現式 4.19 のユニット 1**

- *POS*_未知語 *CT_CF_*

は，「體」の形態素情報である．「體」は未知語で返されていた．そこで，表現式 4.5 と同様に，「*POS*_未知語 *CT_CF_*」に設定した．

■ **表現式 4.19 のユニット 2**

- *W*_[一-㷀ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_名詞-サ変接続 *CT_CF_*

は，「験」の形態素情報である．

単語の文字列,

- *W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，表現式 4.6 と同様の設定を行った.

単語の文字列に続く,

- *POS_*名詞-サ変接続 *CT_CF_*

は,「験」が返された品詞「名詞-サ変接続」を設定した．また,「名詞-サ変接続」は活用しないため，活用型と活用形は設定しない.

■ **表現式 4.19 のユニット 3 とユニット 4**

- *W_*し *POS_*動詞-自立 *CT_サ変・スル CF_連用形*
- *W_*て *POS_助詞-接*続助詞

は，表現式 4.11 の「し」と「て」にマッチさせる設定と同様である.

- **「て」の表現式 10**：

表現式 4.20 は, 表 4.2 の番号 4 のうち, 次の 1 件の誤用表現を対象に検出する式である.

*POS_*名詞-一般 *CT_CF_W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+ *POS_*副詞-一般 *CT_CF_*(?!*W_*.+POS_助詞-格助詞-一般 *CT_CF_)W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-龥ァ-ヶ]+*CF_*連用タ接続 *W_*て *POS_*助詞-接続助詞 (4.20)

- 母はてつかって(正しくは「母はてつだって(手伝って)」)

この誤用表現は「母」,「はて」,「つかっ」,「て」の四つの単語に分割されていた．また，NOT のユニットを一つ組み込んだ．そのため，表現式 4.20 は，五つのユニットから構成されている.

■ **表現式 4.20 のユニット 1**

- *POS_名詞-*一般 *CT_CF_*

は,「母」の形態素情報にマッチさせる設定である．これは表現式 4.14 における先頭の形態素情報と同様の設定である.

■ 表現式 4.20 のユニット 2

- *W_*[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+ *POS_*副詞-一般 *CT_CF_*

は，「はて」の形態素情報である．

単語の文字列，

- *W_*[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は，表現式 4.6 と同様の設定を行った．

単語の文字列に続く，

- *POS_*副詞-一般 *CT_CF_*

は，「はて」が返された品詞「副詞-一般」を設定した．また，「副詞-一般」は活用しないため，活用型と活用形は設定しない．

■ 表現式 4.20 のユニット 3

- (?! *W_*.+POS_助詞-格助詞-一般 *CT_CF_)*

は，「?!」で先読みして返す *NOT* の設定を行った．*NOT* の設定を行ったのは，格助詞の形態素情報を持つ単語に対してである．この設定を行ったのは，後に続く単語「つかって」が「～をつかって」の「を」や「～につかって」の「に」など，正しい文にもマッチするからである．そのため，「を」や「に」などの格助詞の形態素情報を持つ単語を *NOT* にする設定を行った．

まず，*NOT* の範囲を「()」で指定した．次に，*NOT* の正規表現「?!」を「()」のなかに入れた．次に単語の文字列を

- *W_*.+

とした．この「.+」は，任意の文字が一つ以上を表す正規表現である．

単語の文字列に続く，

- *POS_*助詞-格助詞-一般 *CT_CF_*

は，格助詞の品詞「助詞-格助詞-一般」，活用型，活用形である．格助詞は活用しないため，品詞以外は設定しない．

■ **表現式 4.20 のユニット 4**

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-㍊ァ-ヶ]+*CF_*連用タ接続

は,「つかっ」の形態素情報である.

単語の文字列,

- *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+

は, 表現式 4.6 と同様の設定を行った.

単語の文字列に続く,

- *POS_*動詞-自立

は,「つかっ」が返された自立語動詞を設定した.

品詞に続く,

- *CT_*[一-㍊ァ-ヶ]+

は, 表現式 4.4 と同じ活用型の設定を行った.

活用型に続く,

- *CF_*連用タ接続

は,「つかっ」が返された活用形「連用タ接続」を設定した.

■ **表現式 4.20 のユニット 5**

- *W_*て *POS_*助詞-接続助詞

は, 表現式 4.11 の「て」にマッチさせる設定と同様である.

- **「て」の表現式 11**:

表現式 4.21 は, 表 4.2 の番号 4 のうち, 次の 2 件の誤用表現を対象に検出する式である.

*POS_*(動詞-接尾 | 助動詞)CT_(一段 | 特殊・ナイ)*CF_*(連用形 | ガル接続)*W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-㍊ァ-ヶ]+*CF_*(連用形 | 連用タ接続)*W_*て *POS_*助詞-接続助詞 (4.21)

- 言われいて(正しくは「言われていて」)
- なちゃって(正しくは「なっちゃって」)

表現式 4.21 は，三つのユニットから構成されている．

■ **表現式 4.21 のユニット 1**

- *POS_*(動詞-接尾|助動詞)CT_(一段|特殊・ナイ)*CF_*(連用形|ガル接続)

は，「言われいて」の「れ」，「なちゃって」の「な」にマッチさせる設定である．

- *POS_*(動詞-接尾|助動詞)

は，「れ」が「動詞-接尾」，「な」が「助動詞」の品詞で返されていた．そこで，「( )」で範囲を指定し，*OR* を設定した．

品詞に続く，

- *CT_*(一段|特殊・ナイ)

は，「れ」が「一段」，「な」が「特殊・ナイ」の活用型で返されていたため，品詞と同様に「( )」で範囲を指定し，*OR* を設定した．「特殊・ナイ」は否定の助動詞「ない」の活用型である[39]．

活用型に続く，

- *CF_*(連用形|ガル接続)

は，「れ」が「連用形」，「な」が「ガル接続」の活用形で返されていたため，品詞・活用型と同様に「( )」で範囲を指定し，*OR* を設定した．「ガル接続」は「～がる」，「～げ」，「～そう」に接続する活用形である[39]．たとえば，「寒い」は「寒がる」，「寒げ」，「寒そう」になる．

■ **表現式 4.21 のユニット 2**

- *W_*[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-纊ァ-ヶ]+*CF_*(連用形|連用タ接続)

は，「言われいて」の「い」，「なちゃって」の「ちゃっ」にマッチさせる設定である．

単語の文字列，品詞，活用型，

- *W_*[一-纊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+ *POS_*動詞-自立 *CT_*[一-纊ァ-ヶ]+

は，表現式 4.20 の「つかっ」と同様の設定を行った．「い」と「ちゃっ」は表現式 4.20「つかっ」と同じ形態素情報になるからである．

活用型に続く，

- *CF_*(連用形|連用タ接続)

は,「い」が「連用形」,「ちゃっ」が「連用タ接続」で返されていたため,「()」で範囲を指定し，*OR*を設定した.

■ **表現式 4.21 のユニット 3**

- *W_*て *POS_*助詞-接続助詞

は，表現式 4.11 の「て」にマッチさせる設定と同様である.

- **「て」の表現式 12**：

表現式 4.22 は,表 4.2 の番号 4 のうち,次の 2 件の誤用表現を対象に検出する式である.

*POS_*(名詞-副詞可能|名詞-一般)*CT_CF_*(?!*W_*.+*POS_*助詞-格助詞-一般 *CT_CF_*)*W_*[一-㐀ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-㐀ァ-ヶ]+*CF_*(連用タ接続|体言接続特殊２)*W_*て *POS_*助詞-接続助詞 (4.22)

- かんばって(正しくは「がんばって」)
- 晩くて(正しくは「遅くて」)

表現式 4.22 は，四つのユニットから構成されている.

■ **表現式 4.22 のユニット 1**

- *POS_*(名詞-副詞可能|名詞-一般)*CT_CF_*

は,「かんばって」の「かん」,「晩くて」の「晩」にマッチさせる設定である.

- *POS_*(名詞-副詞可能|名詞-一般)

は,「かん」が「名詞-副詞可能」,「晩」が「名詞-一般」の品詞で返されていた. そこで,「()」で範囲を指定し，*OR*を設定した.「名詞-副詞可能」は曜日や月など時間を表す副詞的な用法を持つ名詞, または量や割合などを表し副詞的に使うことのできる名詞である[39]. また,「名詞-副詞可能」と「名詞-一般」は活用しないため，活用型と活用形の設定はしない.

■ **表現式 4.22 のユニット 2**

- (?!*W_*.+*POS_*助詞-格助詞-一般 *CT_CF_*)

は，表現式 4.20 と同様に，*NOT* の設定を行った．*NOT* の設定を行ったのは，後に続く単語との関係から，正しい文にマッチさせないようにするためである．たとえば，「月曜日に行って」は，「月曜日(名詞-副詞可能)」＋「に(助詞-格助詞-一般)」＋「行っ(動詞-自立，五段・カ行促音便，連用タ接続)」＋「て(助詞-接続助詞)」の形態素情報の並びになる．表現式 4.22 では，このような正しい文にマッチしないように格助詞に *NOT* の設定を行った．

■ **表現式 4.22 のユニット 3**

- *W_*[一-纞ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-纞ァ-ヶ]+*CF_*(連用タ接続|体言接続特殊２)

は，「かんばって」の「ばっ」，「晩くて」の「く」にマッチさせる設定である．

単語の文字列，品詞，活用型，

- *W_*[一-纞ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+ *POS_*動詞-自立 *CT_*[一-纞ァ-ヶ]+

は，表現式 4.20 の「つかっ」と同様の設定を行った．「ばっ」と「く」は表現式 4.20「つかっ」と同じ形態素情報になるからである．

活用型に続く，

- *CF_*(連用タ接続|体言接続特殊２)

は，「ばっ」が「連用タ接続」，「く」が「体言接続特殊２」で返されていたため，「()」で範囲を指定し，*OR* を設定した．

■ **表現式 4.22 のユニット 4**

- *W_*て *POS_*助詞-接続助詞

は，表現式 4.11 の「て」にマッチさせる設定と同様である．

- **「て」の表現式 13**：

表現式 4.23 は，表 4.2 の番号 4 のうち，次の 6 件の誤用表現を対象に検出する式である．

*POS_*(接頭詞-名詞接続|未知語)*CT_CF_*(?!*W_*し *POS_*動詞-自立 *CT_*サ変・スル CF_連用形)*W_*[一-龠ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-龠ァ-ヶ]+*CF_*(連用形|連用タ接続)*W_*て *POS_*助詞-接続助詞 (4.23)

- 入いて
- 脱いて
- おかけて(正しくは「おっかけて(追っかけて)」)
- おしゃべって(正しくは「しゃべって」または「おしゃべりして」)
- 少なって(正しくは「少なくなって」)
- 犬に吠かれて

表現式 4.23 は，四つのユニットから構成されている．

■ **表現式 4.23 のユニット 1**

- *POS_*(接頭詞-名詞接続|未知語)*CT_CF_*

は，「て」の二つ前に分割された単語の品詞にマッチさせる設定である．二つ前に分割された単語は，「入いて」の「入」「脱いて」の「脱」，「おかけて」と「おしゃべって」の「お」，「少なって」の「少」，「吠かれて」の「吠」である．これらの単語は「吠」が「未知語」，それ以外は「接頭詞-名詞接続」で返されていた．そこで，「()」で範囲を指定し，*OR* を設定した．「接頭詞-名詞接続」は，数に接続するもの以外の名詞の接頭語である(「お水」の「お」，「ご立派」の「ご」，「高品質」の「高」など)[39]． また，「接頭詞-名詞接続」と「未知語」は活用しないため，活用型と活用形の設定は行わない．

■ **表現式 4.23 のユニット 2**

- (?!*W_*し *POS_*動詞-自立 *CT_*サ変・スル CF_連用形)

は，「～して」となる表現に対して，*NOT* の設定を行った．これは表現式 4.12 とマッチする誤用表現が重複するからである．たとえば，表現式 4.12 で検出する誤用表現には「默し

て」があった.「默」は「未知語」,「し」は「動詞-自立，サ変・スル，連用形」,「て」は「助詞-接続助詞」は返されていた. *NOT*の前では「未知語」にマッチする単語を設定している. また，*NOT* の後ろの単語の設定は「し」の「動詞-自立，サ変・スル，連用形」と，「て」の「助詞-接続助詞」にマッチする. したがって，重複を避けるため，*NOT*の設定を行った. *NOT*の設定は，表現式 4.20 と同様に，「()」で範囲を指定し，「()」のなかに正規表現「?!」を入れた. 次に，表現式 4.11 の「し」と同じ，

- *W*_し *POS*_動詞-自立 *CT*_サ変・スル CF_連用形

の形態素情報を「?!」後ろに入れた.

■ **表現式 4.23 のユニット 3**

- *W*_[一-㍾ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_動詞-自立 *CT*_[一-㍾ァ-ヶ]+*CF*_(連用形|連用タ接続)

は,「て」の直前に接続する単語の形態素情報である. この形態素情報は, 表現式 4.12 の「い」と「ちゃっ」と同じであるため，同様の設定を行った.

■ **表現式 4.23 のユニット 4**

- *W*_て *POS*_助詞-接続助詞

は，表現式 4.11 の「て」にマッチさせる設定と同様である.

- **「て」の表現式 14**:

表現式 4.24 は, 表 4.2 の番号 4 のうち, 次の 2 件の誤用表現を対象に検出する式である.

*POS*_名詞-サ変接続 *CT_CF*_(?!*W*_.+*POS*_助詞-格助詞-(一般|連語)*CT_CF*_)(?!*W*_し *POS*_動詞-自立 *CT*_サ変・スル *CF*_(未然形|連用形))(?!*W*_.+*POS*_動詞-接尾 *CT*_一段 *CF*_連用形)*W*_[一-㍾ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_動詞-自立 *CT*_[一-㍾ァ-ヶ]+*CF*_連用形 *W*_て *POS*_助詞-接続助詞 (4.24)

- 呈現して(正しくは「現れて」)
- 心の中に刻かれて(正しくは「刻まれて」)

表現式 4.24 は，六つのユニットから構成されている．

■ **表現式 4.24 のユニット 1**

- *POS_名詞*-サ変接続 *CT_CF_*

は，「呈現して」の「呈」，「刻かれて」の「刻」にマッチさせる設定である．「呈」と「刻」は，どちらも品詞が「名詞-サ変接続」で返されていた．そのため，サ変名詞の設定を行った．また，活用に対する設定は，表現式 4.19 と同様に，設定を行わない．

表現式 4.24 のユニット 1 に続いて，次の *NOT* の設定を三つ行った．これらの *NOT* を設定したは，正しい文を返さないようにするためである．

■ **表現式 4.24 のユニット 2**

- (?! *W_*.+*POS_*助詞-格助詞-(一般|連語)*CT_CF_)*

は，格助詞に対して行った．*NOT* の前にはサ変名詞を設定している．サ変名詞は「インプット」などの単語である．この単語は，「インプットを」の「を」などの「助詞-格助詞-一般」と，「インプットに関して」の「に関して」などの「助詞-格助詞-連語」と接続する．これららは正しい文であるため，*NOT* を設定した．*NOT* の設定は，正規表現と範囲，単語の文字列，活用形，活用型は，表現式 4.20 と同様である．
品詞は，

- *POS_*助詞-格助詞-(一般|連語)

「助詞-格助詞-」に続く下位階層の範囲を「()」で指定し，そのなかに「()」*OR* で「一般」と「連語」を設定した．

■ **表現式 4.24 のユニット 3**

- (?! *W_*し *POS_*動詞-自立 *CT_*サ変・スル *CF_*(未然形|連用形))

は，自立語動詞「する」に対して行った．*NOT* の前にはサ変名詞を設定しているため，正しい文のテ形にマッチする(たとえば，「インプットして」など)．そのため，自立語動詞「する」に対して *NOT* を設定した．*NOT* の設定は，正規表現と範囲，単語の文字列，品詞，活用型は，表現式 4.23 と同様である．

活用形は,

- *CF_*(未然形|連用形)

「()」で範囲を指定し，そのなかに「()」*OR* で「未然形」と「連用形」を設定した．自立語動詞「する」は「未然形」のとき，「インプットしない」のように，語幹が「し」に変化して活用する．また，「連用形」のとき，「インプットして」のように，「し」となる．そのため，これら二つの活用形に対して *OR* の設定を行った.

■ **表現式 4.24 のユニット 4**

- (?! *W_*.+*POS_*動詞-接尾 *CT_*一段 *CF_*連用形)

は，学校文法では助動詞とされている「せる／させる」，「れる／られる」などに対して行った．*NOT* の前にはサ変名詞を設定しているため，「インプットされて」，「インプットさせて」などの正しい文にマッチする．そのため，これらの正しい助動詞の文にマッチしないように設定を行った．*NOT* の設定は，正規表現と範囲，単語の文字列，表現式 4.23 表現式 4.20 と同様である.

品詞，活用型，活用形は,

- *POS_*動詞-接尾 *CT_*一段 *CF_*連用形

を設定した.「されて」の「れ」,「させて」の「せ」の形態素情報は，品詞が「動詞-接尾」，活用型が「一段」，活用形が「連用形」だからである.

以上の三つの *NOT* を設定し，正しい文を返さないようにした.

■ **表現式 4.24 のユニット 5**

- *W_*[一-㍾ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-㍾ァ-ヶ]+*CF_*連用形

は,「呈現して」の「現し」,「刻かれて」の「かれ」にマッチさせる設定である.

- *W_*[一-㍾ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+ *POS_*動詞-自立 *CT_*[一-㍾ァ-ヶ]+

は，表現式 4.20 の「つかっ」と同じ形態素情報であるため，同様の設定を行った.

続く活用形,

- *CF_*連用形

は，どちらも「連用形」で返されていた．したがって，活用形を「連用形」に設定した．

■ **表現式 4.24 のユニット 6**

- *W_*て *POS_*助詞-接続助詞

についても，表現式 4.11 の「て」にマッチさせる設定と同様である．

- **「て」の表現式 15**：

表現式 4.25 は, 表 4.2 の番号 4 のうち, 次の 2 件の誤用表現を対象に検出する式である．

*POS_*動詞-自立 *CT_*(一段|.+促音便)*CF_*(未然形|連用形) *W_*[一-㍊ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-非自立 *CT_*[一-㍊ァ-ヶ]+*CF_*(連用タ接続|連用形) *W_*て *POS_*助詞-接続助詞 (4.25)

- 想い出して(正しくは「思い出して」，常用漢字表によると「想(おも)」の読み方はない)
- 似って

表現式 4.25 は，三つのユニットから構成されている．

■ **表現式 4.25 のユニット 1**

- *POS_*動詞-自立 *CT_*(一段|.+促音便)*CF_*(未然形|連用形)

は，「想い出して」の「想い」，「似って」の「似」にマッチさせる設定である．「想い」と「似」は，どちらも品詞が「動詞-自立」で返されていた．そのため，自立語動詞の設定を行った．

続く活用型,

- *CT_*(一段|.+促音便)

は，「似」が「一段」，「想い」が「五段・ワ行促音便」で返されていた．表現式 4.25 においては，全ての促音便にマッチさせるため，一文字以上の繰り返しの正規表現「.+」とした．そして，「一段」と「.+促音便」を「( )」の範囲で指定し，*OR* を設定した．自立語動詞の

また,「五段・ワ行促音便」は,「言う」,「食う」など，本来，五段ワ行で「助詞-接続助詞」の「て」に接続するときに促音便になる動詞の活用型である[39].

続く活用形,

- *CT_*(未然形|連用形)

は,「似」が「未然形」,「想い」が「連用形」で返されていたため,「()」の範囲で指定して*OR*を設定した.

■ **表現式 4.25 のユニット 2**

- *W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-非自立 *CT_*[一-龥ァ-ヶ]+*CF_*(連用タ接続|連用形)

は,「想い出して」の「出し」,「似って」の「っ」にマッチさせる設定である．二つの単語の品詞は「動詞-非自立」で返されていた．そこで，表現式 4.17 の「運動がてきて」の「き」と同様の単語の文字列，品詞，活用型の設定をした.

活用型に続く,

- *CF_*(連用タ接続|連用形)

は,「っ」が「連用タ接続」,「出し」が「連用形」の活用形で返されていたため,「()」で範囲を指定した *OR* を設定した.

■ **表現式 4.25 のユニット 3**

- *W_*て *POS_*助詞-接続助詞

は，表現式 4.11 の「て」にマッチさせる設定と同様である.

- **「て」の表現式 16**：

表現式 4.26 は,表 4.2 の番号 4 のうち,次の 3 件の誤用表現を対象に検出する式である.

*POS_*動詞-自立 *CT_*[一-龥ァ-ヶ]+*CF_*連用形(?!*W_*し *POS_*動詞-自立 *CT_*サ変・スル *CF_*連用形)*W_*[一-龥ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-龥ァ-ヶ]+*CF_*(連用形|連用タ接続)*W_*て *POS_*助詞-接続助詞 (4.26)

- 思い出て(正しくは「思い出して」)
- 思い浮いて(正しくは「思い浮かべて」)
- 失しなって(正しくは「思い浮かべて」)

表現式 4.26 は，四つのユニットから構成されている．

■ **表現式 4.26 のユニット 1**

- *POS*_動詞-自立 *CT*_[一-纔ァ-ヶ]+*CF*_連用形

は，「思い出て」，「思い浮いて」の「思い」，「失しなって」の「失し」にマッチさせる設定である． これらの単語は，いずれも品詞が自立語動詞，活用形が「連用形」で返されていた．そのため，品詞を「動詞-自立」，活用形を「連用形」に設定した．
活用型は，それぞれ異なる形態素情報が返されていたため，

- *CT*_[一-纔ァ-ヶ]+

を設定した．「思い出て」の「思い」が「一段」，「思い浮いて」の「思い」が「五段・カ行イ音便」，「失しなって」の「失し」が「五段・ラ行」で返されていた．表現式 4.4 で示したように，自立語動詞の活用型は，文字列が使用されているため「[一-纔ァ-ヶ]」にし，それらの文字列が 1 回以上の「+」とすることで，これらの活用型にマッチさせる．

■ **表現式 4.26 のユニット 2**

- (?!*W*_し *POS*_動詞-自立 *CT*_サ変・スル CF_連用形)

先頭の形態素情報に続いて，次に *NOT* の設定を行った．この *NOT* は，「～して」となる表現を返さない設定である．また，*NOT* の形態素情報は表現式 4.23 と同様である．*NOT* に後続する形態素情報の設定は「して」の「し」にあたる「動詞-自立，サ変・スル，連用形」と，「て」にあたる「助詞-接続助詞」にマッチする．この「～して」を返さないようにするために，*NOT* の設定を行った．

■ **表現式 4.26 のユニット 3**

- *W*_[一-纔ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS*_動詞-自立 *CT*_[一-纔ァ-ヶ]+*CF*_(連用形|連用タ接続)

は，「思い出て」の「出」，「思い浮いて」の「浮い」，「失しなって」の「なっ」にマッチさ

せる設定である．また，この形態素情報は，表現式 4.21 における「言われいて」の「い」と「て」,「なちゃって」の「ちゃっ」と「て」と同様の設定である．単語の文字列をすべての文字列に設定した. 品詞は, いずれも自立語動詞で返されていたことから,「動詞-自立」を設定した．活用形は,「出」が「一段」,「五段・カ行イ音便」,「五段・ラ行」と異なるものが返されていた. そこで, すべての活用型にマッチさせる設定を行った. 活用形は,「出」が「連用形」,「浮い」と「なっ」が「連用タ接続」で返されていたため，範囲を指定して *OR* を設定した.

■ **表現式 4.26 のユニット 4**

- *W_*て *POS_*助詞-接続助詞

は，表現式 4.11 の「て」にマッチさせる設定と同様である.

- **「て」の表現式 17** :

表現式 4.27 は, 表 4.2 の番号 4 のうち, 次の 1 件の誤用表現を対象に検出する式である.

*POS_*接頭詞-名詞接続 *CT_CF_W_*[一-龠ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-龠ァ-ヶ]+*CF_*基本形 *W_*[一-龠ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-龠ァ-ヶ]+*CF_*連用タ接続 *W_*て *POS_*助詞-接続助詞 (4.27)

- 近つくなって(正しくは「近づいて」)

この誤用表現は「近」,「つく」,「なっ」,「て」の四つの単語に分割されていた．そこで, 表現式 4.27 は，四つのユニットから構成した.

■ **表現式 4.27 のユニット 1**

- *POS_*接頭詞-名詞接続 *CT_CF_*

は,「近」にマッチさせる設定である.「近」は「接頭詞-名詞接続」で返され,「接頭詞-名詞接続」は活用しないため,「*POS_*接頭詞-名詞接続 *CT_CF_*」に設定した.

■ **表現式 4.27 のユニット 2 とユニット 3**

- *W_*[一-㍾ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-㍾ァ-ヶ]+*CF_*基本形
- *W_*[一-㍾ぁ-んァ-ヶＡ-Ｚａ-ｚ０-９A-Za-z0-9]+*POS_*動詞-自立 *CT_*[一-㍾ァ-ヶ]+*CF_*連用タ接続

は,「つく」と「なっ」にマッチさせる設定である．単語の文字列，品詞，活用形は，表現式 4.20 の「つかっ」と同様の設定である．また，この二つの単語は活用形のみ異なる．文字列はすべての文字列に設定し，品詞はどちらも自立語動詞で返されていた．活用型は「つく」が「五段・カ行イ音便」,「なっ」が「五段・ラ行」で返されていた．これらの形態素情報は，表現式 4.20 の「つかっ」と同じ設定にマッチする．そこで,「つかっ」と同じ設定にした．活用形は,「つく」が「基本形」,「なっ」が「連用タ接続」で返されていた．そのため，それぞれの活用形に対し,「つく」を「*CF_*基本形」,「なっ」を「*CF_*連用タ接続」に設定した.

■ **表現式 4.27 のユニット 4**

- *W_*て *POS_*助詞-接続助詞

は，表現式 4.11 の「て」にマッチさせる設定と同様である.

- **「て」の表現式 18**：

表現式 4.28 は, 表 4.2 の番号 4 のうち, 次の 1 件の誤用表現を対象に検出する式である.

(?! *W_*(な | 無 | 亡)くなっ *POS_*動詞-自立 *CT_*五段・ラ行 *CF_*連用タ接続) *W_*[一-㍾ぁ-ん]+なっ *POS_*動詞-自立 *CT_*[一-㍾ァ-ヶ]+*CF_*連用タ接続 *W_*て *POS_*助詞-接続助詞 (4.28)

- 失くなって(正しくは「亡くなって」)

この誤用表現は「失くなっ」,「て」の二つの単語に分割されていた．また *NOT* のユニットを背呈した．そのため，表現式 4.28 は，三つのユニットを構成した.

■ 表現式 4.28 のユニット 1

- (?!*W_*(な|無|亡)くなっ *POS_*動詞-自立 *CT_*五段・ラ行 *CF_*連用タ接続)

先頭の文字列では，正しい文を返さないように設定した．この形態素情報は，正しい文である三つの「なくなって」,「無くなって」,「亡くなって」を *NOT* にしている．正しい文は，それぞれ「なくなっ」,「無くなっ」,「亡くなっ」と「て」に単語が分割される．単語の文字列は「な」,「無」,「亡」が異なり，後ろが共通している．そこで，単語の文字列に対して，

- *W_*(な|無|亡)くなっ

と，設定した．

続く品詞，活用型，活用形も共通していため，

- *POS_*動詞-自立 *CT_*五段・ラ行 *CF_*連用タ接続

を設定した．そして，これらの形態素情報の範囲を「( )」で指定し，形態素情報の先頭に *NOT* の正規表現「!?」を設定した．

■ 表現式 4.28 のユニット 2

- *W_*[一-㐀ぁ-ん]+なっ *POS_*動詞-自立 *CT_*[一-㐀ァ-ヶ]+*CF_*連用タ接続

は，「失くなっ」にマッチさせる設定である．

単語の文字列，

- *W_*[一-㐀ぁ-ん]+なっ

は，漢字と平仮名の文字列を一つ以上「[一-㐀ぁ-ん]+」，続く文字列「なっ」を指定した．

単語の文字列に続く，

- *POS_*動詞-自立 *CT_*[一-㐀ァ-ヶ]+*CF_*連用タ接続

は，返された品詞，活用型，活用形を指定した．活用型は「五段・ラ行」が返されていたが，すべての活用形の文字列を返すように「*CT_*[一-㐀ァ-ヶ]+」を設定した．

■ 表現式 4.28 のユニット 3

- *W_*て *POS_*助詞-接続助詞

は，表現式 4.11 の「て」にマッチさせる設定と同様である．

- **「て」の表現式 19**：

表現式 4.29 は,表 4.2 の番号 4 のうち,次の 1 件の誤用表現を対象に検出する式である.

*POS*_動詞-自立 *CT*_五段・マ行 *CF*_連用タ接続 *W*_て *POS*_助詞-接続助詞　(4.29)

- 寝込んて

表現式 4. 29 は，二つのユニットから構成されている.

■ 表現式 4. 29 のユニット 1

- *POS*_動詞-自立 *CT*_五段・マ行 *CF*_連用タ接続

は,「寝込ん」にマッチさせる設定である．この単語は，品詞が自立語動詞の「動詞-自立」，活用形は正しいテ形に接続する「連用タ接続」で返されていた.

また,「五段・マ行」で返されていた.

- *CT*_五段・マ行

「五段・マ行」は「進む」など，テ形の品詞「助詞-接続助詞」に接続するとき，撥音「ん」になるマ行の五段動詞である[39].「五段・マ行」は，正しいテ形に変化するとき,「で」となり，文字列「て」には接続しない．そこで，活用型を「五段・マ行」に設定した.

■ 表現式 4. 29 のユニット 2

- *W*_て *POS*_助詞-接続助詞

は，表現式 4.11 の「て」にマッチさせる設定と同様である.

- **「て」の表現式 20**：

表現式 4.30 は,表 4.2 の番号 4 のうち,次の 2 件の誤用表現を対象に検出する式である.

*POS*_動詞-自立 *CT*_五段・(ワ行促音便|バ行)CF_連用形 W_て POS_助詞-接続助詞　(4.30)

- 習いて(正しくは「習って」)
- 選びて(正しくは「選んで」)

表現式 4.30 は，二つのユニットから構成されている．

■ **表現式 4.30 のユニット 1**

- *POS*_動詞-自立 *CT*_五段・(ワ行促音便|バ行)CF_連用形

は，「習い」と「選び」にマッチさせる設定である．これらの単語は，品詞が自立語動詞の「動詞-自立」で返されていた．

また，活用型は「習い」が「五段・ワ行促音便」，「選び」が「五段・バ行」で返されていた．そこで，「五段・」以下の下位階層の範囲を指定し，OR を設定した．

- *CT*_五段・(ワ行促音便|バ行)

なお，「五段・バ行」は「呼ぶ」など，テ形の品詞「助詞-接続助詞」に接続するとき，撥音「ん」になるバ行の五段動詞である[39]．

活用形は，どちらも「連用形」で返されていた．

- *C*F_連用形

五段動詞の「ワ行促音便」と「五段・バ行」は，テ形の形態素情報「*W*_て *POS*_助詞-接続助詞」に接続するとき，活用形が「連用タ接続」になり，「連用形」にはならない．そこで，活用形を「連用形」に設定した．

■ **表現式 4.30 のユニット 2**

- *W*_て *POS*_助詞-接続助詞

は，表現式 4.11 の「て」にマッチさせる設定と同様である．「習いて」，「遊びて」の二つの「て」は，どちらも文字列が「て」，品詞が「助詞-接続助詞」で返されていたため，「*W*_て *POS*_助詞-接続助詞」と設定した．

表 4.4 テ形の表現式で使用した正規表現

番号	正規表現	働き	番号	正規表現	働き
1	.	任意の一文字	9	一-纞	漢字
2	*	0 回以上の繰り返し	10	ぁ-ん	平仮名
3	+	1 回以上の繰り返し	11	ァ-ヶ	片仮名
4	()	「（」から「）」までの範囲	12	A-Z	全角大文字アルファベット
5	?!	*NOT*(先読みして否定)	13	ａ-ｚ	全角小文字アルファベット
6	|	*OR*(または)	14	０-９	全角アラビア数字
7	[]	[]内の一文字	15	A-Z	半角大文字アルファベット
8	{*m*}	{*m*}回以上の繰り返し(m は 0 以上の整数)	16	a-z	半角小文字アルファベット
			17	0-9	半角アラビア数字

4. 2. 8　テ形の誤用文を検出する表現式を用いた検出結果

テ形の誤用文を検出する作成した表現式を使って検索した．結果を表 4.5 に示す.

表 4.5 テ形の表現式で検出された文と表現の件数

検出した文	誤用文	誤用表現
138	75	77

表 4.5 の結果と目視で分類した表 4.1 を比べると，目視の誤用文は 88 件であったが，表現式では 75 件が検出できた．また，誤用表現は目視では 90 件であったが，77 件が検出できた．加えて，表 4.5 の表現式で検出した文 138 件を見ると，テ形の誤用文ではないものも検出された.

以上の結果，テ形の誤用文とは異なる文がいくつか検出され，また検出されない文と表現がいくつかあったが，テ形の誤用において目視での判別を再現できた.

4. 2. 9　表現式で検出されなかった誤用文の分析

本実験において，目視で誤用表現と判別したが，表現式を用いた検索では検出されなかった誤用文を分析する.

実験において，目視では誤りであると判別したが，テ形の誤用文を検出する表現式を用いた検索では検出されなかった誤用表現を表 4.6 に示す.

表 4.6 表現式で検出されなかったテ形の誤用表現

番号	テ形の誤用表現	件数
1	終って	3
2	決って	2
3	泊って	1
4	分って	1
5	向って	1
6	暮して	2
7	来って	1
8	(制服を)着いて	1
9	(獣医生を)見て	1

表 4.6 の番号 1～6 のテ形は，形態素解析器の辞書に単語の登録がされていたために，誤用文を検出するように設定した表現式にマッチせず，検出されなかった．常用漢字表の送り仮名によると，番号 1～6 の送り仮名と異なる．常用漢字表の送り仮名と正しいテ形を表 4.7 に示す.

表 4.7 常用漢字表の送り仮名を基にした正しいテ形と辞書の登録

番号	常用漢字表の送り仮名	正しいテ形	見出し語として辞書に登録されている単語
1	終わる	終わって	終わる／終る
2	決まる	決まって	決まる／決る
3	泊まる	泊まって	泊まる／泊る
4	分かる	分かって	分かる／分る
5	向かう	向かって	向かう／向う
6	暮らす	暮らして	暮らす／暮す

表 4.7 の番語 1〜6 のテ形の正しい形は表中の見出しにある「正しいテ形」のようになる．しかし，形態素解析器の辞書を確認すると，番号 1〜6 の動詞は，表 4.7 の辞書の登録のように，送り仮名の異なる 2 種類の単語が見出し語に登録されていた．形態素解析器は，辞書に登録があることから，正しいテ形の形態素情報を付けたと考えられる．そのため，誤用文を検出する表現式にマッチせず，検出されなかった．これらの漢字の送り仮名は揺れていると考えられるため，2 種の単語が登録されている可能性がある．常用漢字表にしたがうならば，表現式を使った検出においては，有用な誤用文の検出は行えない．ただし，送り仮名が揺れているということは，正しいテ形であるとも考えられる．これらのテ形が正しいテ形であると判別するならば，目視の分類を再現する必要はない．

番号 7 も形態素解析器の辞書に単語の登録がされていたために，表現式にマッチせず，検出されなかった．辞書には「来る」の形で，二つの単語が登録されていた．形態素解析器は登録されているどちらか一方の単語の情報を付け，付けた情報が正しく書かれたテ形と同じであったため，表現式にマッチせず検出されなかった．「来る」が同じ形で二つ辞書に登録されているのは，「カ変」と呼ばれる特殊な活用をする動詞であるからだと考えられる．このような特殊な活用をする単語の再現は難しいと考える．ただし，「カ変」のような特殊な単語は，日本語において数は多くない．したがって，このような特殊な単語が再現できない場合は，いわゆる決め打ちなどの方法で表現式を設定することで対応できる．

番号 8 は漢字の読み方が二つあるために検出されなかった．番号 8 の誤用文は，正しくは「制服を着て」と書く．漢字「着」は，「着る」と書くと「きる」と読み，「着く」と書くと「つく」と読む．つまり，テ形の誤用表現である「着いて」は，「ついて」と読むことができる．形態素解析器は，「着いて」の形態素情報を正しく付けたため，表現式にマッチせず，検出されなかった．表現式を使って目視を再現する際，漢字の読み方が一つ以上ある場合は再現が難しい．ただし，このような誤りも数は多くないと予想される．実際に，本実験においても，目視での確認で誤用表現は 90 件あったが，読み方が一つ以上ある誤りは，この 1 件しかなかった．そのため，読み方が一つ以上ある単語についても，予測ができるならば決め打ちで対応することができる．

番号 9 は漢字の使い方は誤っているけれども，使用された漢字は正しく書かれているため，検出されなかった．番号 9 の文全体を見て判別したところ，正しくは「獣医が診て」と書く．学習者は「診て」を「見て」と誤っているが，単語「見る」のテ形は「見て」で

あるため，正しいテ形が書かれている．したがって，形態素解析器は正しく形態素情報を付けたため，表現式にマッチせず，検出されなかった．表現式を使った目視の再現において，正しく書かれおり，意味・使い方に誤りは再現できない．ただし，このような間違いは，文法的な誤りではなく，単語の選択の問題である．そのため，単語の選択の問題と考えるならば，再現の対象ではない.

4．3　誤用文の分類としての表現式検出法の検討

テ形の誤用文とは異なる文がいくつか検出され，また検出されない文と表現がいくつかあったが，テ形の誤用において目視での判別を再現できた.

4.2 節で述べたように，表現式検出法による誤用文の検出は，誤用文を機械が分類していると捉えることができる．そのため，作成した表現式は，ある特定の誤用の分類基準と考えられる．その基準に基づいて分類(検出)されて文は，従来の人が目で見て行った分類よりも客観性を保ったものである．そのため，表現式検出法で検出された誤用文は，人が分類した誤用文よりも正しく分類されている可能性がある.

実験においては，形態素解析辞書の問題により再現できない誤用もあったが，それらは，日本語そのものの揺れの問題であるとも考えられた．したがって，4.2.8 項で述べたように，再現の必要はない．加えて，文法的な誤りではなく，単語の選択の誤りについても，再現の対象ではない.

また，再現できない誤りのには，特殊な単語があったが，そのような単語は日本語として数が少ないため，いわゆる決め打ちとして対応できる．同様に，読み方が二つあるものについても，数がすくない．そして，そのような誤りの予測ができるならば，決め打ちで対応できる.

以上のことから，動詞テ形の活用誤りにおいて，表現式検出法を用いて目視の分類を再現でき，有効性が確認できた.

4．4　本章のまとめ

本章では，人が目視で確認した分類した誤用文を表現式検出法で再現できるか検証し，有効性を確認した．検証の実験例として，日本語教育において動詞の活用誤りの一つとされているテ形を例に行った．実験の結果，目視で分類した誤用を再現でき，有効性が確認できた.

また，誤用文を分類する機械としての表現式検出法を検討したところ，従来の方法である人が目で見て行った分類よりも客観性があり，表現式検出法で検出された誤用文は，人が分類した誤用文よりも正しく分類されている可能性があった．

第 5 章

誤用研究における手法の有効性

5. 1　本章のあらまし

本章では，第 3 章で作成した表現式を再利用して，日本語学習者が誤りを生じさせる原因の調査と分析して，表現式検出法の有効性を確認する．

本研究では，1.2.7 項で述べたように，表現式検出法で取り出したデータを誤用研究に有効に活用できるか考えている．つまり，表現式検出法で取り出したデータを使って学習者が誤用を生じさせる原因が分析することができるかということを考えている．また，127 項および 2.4 章で述べたように，表現式検出法が誤用研究を効率的に行えると考えている．そこで，本章では，表現式検出法を用いて効率的に誤用文を検出し，検出したデータを活用して学習者が誤りを原因の調査と分析を行い，提案した手法が誤用研究に有効であるか確認する．ここでの有効とは，誤用文の検出が目視や文字列検索よりも効率的であったか，さらに表現式検出法で取り出したデータが学習者の誤用を分析することができたかである．

5. 2　日本語学習者の誤用分析と表現式検出法によるデータの活用

日本語学習者の誤用文は，誤った原因を分析することで，日本語をどのように習得していくかということや教材・テスト・教授法の応用などに利用できる [31]．そのため，誤用が生じる原因を明らかにした研究や，研究を通じて行われた教授法への提言などが少なくない [1〜4，67]．さらに，誤用文の分析は，日本語教育も含め，外国語教育にたずさわる者にとって，大いに興味をひかれる分野であるとされる[48]．

本章おいては，表現式検出法によって検出した誤用文を活用して，学習者が誤用を生じさせる原因を分析する．表現式検出法による検出例として，中国語を母語とする学習者による品詞の接続に誤りがある接尾辞「的」の誤用を日本語学習者コーパスから検出する．また，検出には第 3 章で作成した「的」に関する表現式を再利用しする．

5. 2. 1　接尾辞「的」と日本語学習者の「的」関する誤用

日本語の接尾辞「的」は，もとは中国語の助辞の用法であり，明治初期の翻訳文のなか

で英語の「-tic」などの形容詞的な語の訳語として，二字の漢語に付けて用いられるようになった[71]．調査によると，「的」の前に接続する語基は，和語，混種語，外来語よりも，漢語を使用する頻度が高い[72]．名詞を修飾するとき，「的」には二つの用法がある．たとえば，「環境」という単語が「問題」という単語を修飾するとき，「環境的問題」とする用法と「環境的な問題」とする用法がある．二つの用法において，学習者は日本語母語話者とくらべて，「的な」を使用する傾向がある．日本語母語話者と多くが中国語を母語とする学習者(中国語母語話者 27 名，韓国母語話者 7 名)に対して調査を行ったところ，日本語母語話者は「環境的問題」のような文を作る頻度が高く，学習者は「環境的な問題」のような文を作る頻度が高かった[73].

中国語を母語とする学習者において，「的な」を誤って使用した用法には，「危険的なゲーム」のような「的」を過剰に使用した誤用，「人力的な物」のような創作した誤用，「屈辱な歴史」のような「的」が欠落した誤用などが見られたと報告されている[3，67]．また，語基には二字の漢語を使用した誤りが多いことも報告されている[3].

「的」の過剰な使用など，学習者が使い方を誤った原因として，日本語で誤りとなる語基は，中国語では名詞の修飾語になりうるものが多いからだとされる[3]．中国語「的(de)」は，修飾表現は前後に並ぶ 2 つの成分が修飾と被修飾の関係にあるとき，その表現は修飾連語と呼ばれて修飾語の後ろに「的(de)」をともなう場合がある[74]．「的(de)」は助詞と呼ばれ，その意味用法は多様であるが，定語と呼ばれる用法があり連体修飾を標示する機能がある[72]．一般に，修飾連語であるとき，名詞などの体詞(副詞「不(bu)」などを直接修飾に受けられない語)が被修飾語となり，その前の修飾語が定語とされる[74].

5．2．2　本研究で着目する「的」の誤用と母語の影響を調査する方法

中国語を母語とする学習者は「的な」を使用する頻度が高い[73]ことから，本研究では，中国語を母語とする学習者が誤って書いた「的な」の誤りに着目する．また，「的」に関連する誤りとして「的の」と書かれる可能性があることから「的の」の誤りにも着目する．これらの着目した誤用を，1.2.5 項で述べたデータを使用し，5.1 章で述べたように第 3 章の表現式を使って検出する． そして，検出した文において，日本語では誤りとなる「語基＋的な＋名詞」と「語基＋的の＋名詞」が，母語における中国語では「単語＋的(de)＋名詞」として使用されやすい検証する．検証の方法として，学習者が誤って書いた「語基＋的な＋名詞」と「語基＋的の＋名詞」の表現を自分でどのように中国語に翻訳したか確認する．

また，中国語における「単語＋的」の組み合わせを分析して，単語の組み合わせが共起する強さを測る．さらに，検証結果をふまえて，中国語を母語とする学習者の「的な」と「的の」における誤用の原因を考察する．

5．2．3　使用するデータと文の解析

1.2.5 項で述べたように，日本語学習者コーパスは，次のものを使用し，これらから構造化テキストを確認し，第 3 章で作成した表現式を再利用して検索を行う．

- 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[13](台湾人学習者の 57 作文，1,046 文，以下，翻訳付きデータと呼ぶ)
- 台湾人日本語学習者コーパス[22](581 作文，9,358 文，以下，台湾人学習者データと呼ぶ)

中国語のデータは，総語数が約 500 万語(総語数 4,892,324)とされる次の中国語のコーパスを使用する．

- 中央研究院現代漢語平衡語料庫[75] (以下，現代中国語均衡コーパスと呼ぶ)

使用したデータの解析は，次の環境と形態素解析器を使用する．

- 台湾人学習者データと翻訳付きデータ：本研究の実験環境(第 2 章)
- 中国語のデータ：中文斷詞系統[76] (以下，中国語形態素解析器と呼ぶ)

5．2．4　誤用表現とその判別の手順

第 3 章と第 4 章と同様に，一文中にある「語基＋的な＋名詞」と「語基＋的の＋名詞」の誤りを誤用表現とする．

誤用表現は，表現式を使って検出し，検出された文を判別する．判別の手順を次に示す．

- 検出された文の「語基＋的な＋名詞」と「語基＋的の＋名詞」の誤用表現を目視で判別

- 目視で判別した表現の「語基＋的な」と「語基＋的の」を日本語の大規模コーパスの少納言(BCCWJ) [77]で検索(以下，大規模コーパスと呼ぶ)
- 検索結果が 0 件で返されたものを誤用表現と判定

少納言(BCCWJ)の総語数は，約 1 億語とされる[78]．目視で判別した誤用表現を大規模コーパスで検索するのは，実際に使用されている日本語と誤用表現と判別したものを対照させることがきるからである．つまり，実際の日本語において，存在しない表現は誤用表現であると判定することができる．

大規模コーパスの検索方法は，たとえば，学習者が「巨大的なビル」と書いたものを目視で誤用表現と判別したとする．この表現の「巨大的な」を文字列検索する．

検索の結果については，たとえば，「巨大的な」を検索し，その結果が「コーパスにその文字列はない」として 0 件で返されたものを誤用表現と判定する．

5．2．5　誤用文を検索する表現式

3.1 節で述べたように，表現式は第 3 章の実験で使用したものを再利用する．再利用するのは表現式 3.3 と表現式 3.4 である．

5．2．6　単語と単語が共起する強さ

本研究においては，t スコアを使って中国語の「単語＋的」の強さを測る．t スコア(TScore)は，コーパスの総語数を考慮して全体における個々の単語の出現比率を比較することから，頻繁に用いられる一般性の高い単語と単語が共起する強さを評価するのに適しているとされる[79]．コーパスで共起する単語 x と単語 y の強さは，x を中心語 y を共起語とすると，コーパスの総語数 Nにおける共起頻度 m と中心語頻度 $n1$，共起語頻度 $n2$ から，以下のように計算する[80]．

$$TScore(x,y) = \frac{m - n_1 n_2 / N}{\sqrt{m}} \tag{5.1}$$

一般に，t スコアは，2 以上の強さが測定された場合，組み合わせに意味があるとされる[81]．

5. 2. 7　再利用した表現式による日本語学習者コーパスの検索

5.2.3 項で述べた二つの日本語学習者のテキストコーパスデータを解析して形態素情報を付け，表現式 3.3 と表現式 3.4 で検索した．その結果，文が検出された．検出された文の件数と表現の件数を表 5.1 に示す.

表 5.1 翻訳付きデータと台湾人学習者データの検出件数

データ	文の件数	表現の件数
翻訳付きデータ	2	2
台湾人学習者データ	14	13

5. 2. 8　学習者が翻訳した誤用文の検証

翻訳付きデータから検出した文の表現を目視で判別し，誤用と判別される表現を大規模コーパスで検索した．その結果，0 件で返された表現があった．0 件で返された誤用表現を表 5.2 に示す.

表 5.2 翻訳付きデータの誤用表現

番号	誤用表現
1	適当的な
2	伝統的の

誤用表現の中国語翻訳を確認した．その結果，誤用表現には「的」を使った表現が使用されていた．以下に日本語で書かれた文と学習者が自ら翻訳した文を示す(下線は誤用表現と中国語の表現).

- 日本語:パソーコンとか PDA とか、適当的なことがあれば、資料の検索はすぐできる.
  翻　訳：有電腦或 PDA 等適當的工具，就可以立刻做資料的檢索
- 日本語：伝統的の新聞やテレビはいろいろな限制があります.
  翻　訳：而且傳統的報章雜誌及電視收音機在使用上有種種限制。

使用した翻訳付きデータの中国語翻訳において，日本語では誤りとなる「語基＋的な＋

名詞」と「語基＋的の＋名詞」の表現は，中国語で「単語＋的＋名詞」の表現で書かれていた．

5. 2. 9　中国語における共起の強さの検証

台湾人学習者データから検出した文の表現を目視で判別し，誤用表現と判別されるものを大規模コーパスで検索した．その結果，0 件で返された表現があった．0 件で返された誤用表現を表 5.3 に示す．

表 5.3 台湾人学習者データの誤用表現

番号	誤用表現	番号	誤用表現
1	正確的な	5	公正的な
2	不可欠的な	6	無責任的な
3	健全的な	7	巨大的な
4	有利的な	8	公平的な

表 5.3 の単語を繁体字に変換し，現代中国語均衡コーパスから「単語＋的」となる文を取り出して中国語形態素解析器で解析した．中国語形態素解析器で解析した理由は，「単語＋的」の「単語」が表現である可能性があり，そのような「単語」を取り除くためである．たとえば，日本語の誤り「不可欠」の繁体字「不可缺」は，「不可」と「缺」に単語を分けることができる．したがって，単語ではなく表現として使用されている可能性があるため，中国語の形態素解析器で解析した．また，中国語の「的」が名詞修飾以外の用法で使用されている可能性があるからである．たとえば，中国語の「環境的問題」は名詞を修飾しているけれども，「問題是環境的。」のような「的」は名詞を修飾していない．そのため，名詞を修飾していない形態素情報が付いた表現を取り除くため，中国語の形態素解析器で解析した．

中国語で表現になる可能性がある誤用表現と名詞を修飾していない形態素情報が付いた「的」を取り除いて t スコアを計算して共起の強さを測った．その結果，いずれの「単語＋的」の組み合わせも，共起の割合が 2 以上あった．表 5.4 に共起の強さを測った表現と t スコアの値を示す．

表 5.4 共起の強さを測った表現と t スコアの値

中国語の表現	t スコア	中国語の表現	t スコア
正確的	16.7596	公正的	3.6621
健全的	6.6697	巨大的	9.3176
有利的	5.5092	公平的	7.1261

使用した台湾人学習者データにおいて，日本語では誤りとなる「語基＋的な＋名詞」と「語基＋的の＋名詞」の表現における「語基＋的」は，中国語においては強く共起することがわかった．つまり，日本語では誤りとなる「語基＋的」は，中国語においては，もっともらしい単語と単語の組み合わせであることがわかった．

5. 2. 10　学習者の母語の影響の分析

中国語を母語とする学習者が誤って書いた「的な」,「的の」に着目して表現式検出法で検出を行い，中国語では名詞の修飾語になるかどうか検証した．検証の方法として，検出した文において，学習者が誤って書いた「語基＋的な＋名詞」と「語基＋的の＋名詞」の表現を自らがどのように翻訳したか確認した．また，中国語における「単語＋的」の組み合わせを分析して，単語の組み合わせが共起する強さを測った．翻訳を確認した結果，日本語では誤りとなる「語基＋的な＋名詞」と「語基＋的の＋名詞」の表現は，中国語で「単語＋的(de)＋名詞」の表現で書かれていた．また，日本語の「語基＋的」は，中国語においてはもっともらしい単語と単語の組み合わせであることがわかった．

二つの検証結果から，学習者が「的な」と「的の」の誤った表現を書く原因の一つに母語が干渉していることが考えられる．つまり，中国語を母語とする学習者において，接尾辞「的」に関する誤用の原因の一つに母語の影響があると考える．

さらに,「母語では単語 A と単語 B は共によく使う」という母語におけるコロケーション知識[1]が影響しているのではないかと考える．母語のおけるコロケーション知識とは，たとえば，英語では「good health」のコロケーションは許容されるが，日本語では「ii kenkou (いい健康)」の表現は許容されず，学習者にこのような誤用がみられたため，日本語でも英語のコロケーション知識を使用する傾向があることが指摘しされた[1]．そのため，中国語を母語とする学習者の「的な」と「的の」の誤った使用においても，母語の影響があると考える．

5. 3　誤用研究における手法の有効性について

表現式検出法で検出した誤用文は，学習者が誤りを生じる原因の分析できた．中国語を母語とする学習者が書いた接尾辞「的」は，母語が影響していることが調査からわかり，調査結果を基に分析することができた．

表現式検出法を用いて誤用文を検出したことから効率的に学習者が誤りを生じる原因を分析ることがきた．「的な」と「的の」の誤用表現は，第 3 章でも有効性が確認されたことから，本章でも有効に誤用文を検出することができたが，仮に，分析を行う際に表現式を用いずに，一般的な手法である目視の確認あるいは文字列検索により誤用文を取り出していたならば，効率的に研究が進められなかったかもしれない．使用した日本語学習者コーパスは，一つが約 1,000 文，もう一つが約 10,000 文あったからである．そのため，表現式検出法を用いて，効率的に学習者の誤りの原因を分析できた．

使用した表現式は第 3 章で作成したものを再利用できた．表現式が再利用できたことから，効率的に分析を進めることができた．第 3 章で作成した表現式は，本章でも再利用することができたことから，効率的に分析を進めることができた．

さらに，翻訳付きデータは，3 章とは異なる日本語学習者コーパスであり，台湾人学習者データについても，データ量を増加したが，表現式検出法と再利用した表現式は，誤用文を取り出すことができた．こられの点においても，分析が効率化された．

以上のことから，誤用研究による本研究の手法の有効性が確認できた．

5. 4　本章のまとめ

本章では，本研究では，1.2.7 項で述べたように，表現式検出法で取り出したデータを誤用研究に有効に活用できるか有効性を確認した．その結果，データを活用して学習者の誤用が分析でき，有効性が確認できた．

また，効率的に分析が進められたか検討したところ，誤用研究に際しても誤用文を検出できたこと，表現式が再利用できたこと，二つの異なる日本語学習者のテキストコーパスから誤用文が検出できた．これら三つの点からも効率的に検出を行うことができ，有効性が確認できた．

第 6 章

結論と今後の課題

6. 1 結論

本論文は，電子化処理がされた日本語学習者のテキストコーパスデータから，文中に日本語として誤りがある誤用文を検出する手法を提案した.

近年，日本語学習者の誤用に関する研究においては，日本語学習者コーパスのデータを活用したものが少なくない．日本語学習者コーパスは Web 上などに公開されている．本研究において調査したところ，公開されている日本語学習者コーパスは，作文など学習者が産出した日本語をデータとし，多くがテキストファイルなどの電子化処理のみを施したものであり，またそれらファイルのデータ中に存在する学習者の誤りは訂正されずに電化処理が行われていた．日本語教育の研究者らが，これらテキストコーパスのデータを活用して誤用を分析する際には，誤用を直接目視で確認するか，あるいは，できるだけ多くの誤用文を想定して文字列検索をするといった方法で誤用文の取り出しを行っている．しかし，このような方法は，効率的ではない．そこで，本研究においては，日本語学習者のテキストコーパスデータから多様な文法的誤用文をできるだけ簡便に取り出す手法を提案する．本研究における提案は，日本語学習者のテキストコーパスの文を基にして独自に定義した構造化テキストを作成し，その定義に基づて作成した表現式を用いて構造化テキストに検索をかけ，検索にマッチした文を取り出す手法である．本研究は，提案した手法を用いて日本語学習者のテキストコーパスから誤用文の取り出しを行い，誤用に関する研究を効率的に進めることを目指している．以下に各章を総括する.

第 1 章では,本研究の背景と目的および本研究における提案と関連研究について述べた.

第 2 章においては，日本語学習者のテキスト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について述べた．本研究では，形態素解析器を利用して文を単語に分割し，単語に形態素の情報を付けていく．まず，日本語学習者のテキス

トコーパスの文を形態素解析器にかけ，文を単語に分割し，単語の表層に関する属性(単語文字列・読み方・基本形)と文法的属性(品詞・活用型・活用形)を付与する．単語に付与されたこれらの属性のなかから単語文字列，品詞，活用型，活用形を抽出する．抽出したこれらの属性に対して，本研究で定義したラベルを付けていく．ラベルは，文中に存在しない記号と文字列「 _ 」を用いたもので，単語文字列には「*W_*」，品詞には「*POS_*」，活用型には「*CT_*」，活用形には「*CF_*」と，それぞれ異なる記号に文字列「 _ 」を結合したものを定義した．定義したラベルを抽出した属性に付けていき，ラベルと属性が組み合わさった形の単語を，元の文の単語順序を保持しながら文に埋め込んでいく．本研究では，この一連の処理によって作成したラベル・属性付き単語の文から成るテキストを構造化テキストと定義した．構造化テキストは元の文の単語順序を保持していることから，各ラベルと単語文字列以外の属性を取り除けば，元の文に戻すことができる．また，文法的属性は単語の抽象化した要素であることから，特定の文法的属性に着目して検索すれば，要素の集合にマッチして多様な文法表現をした文を取り出すことができる．つまり，構造化テキストは多角的な視点で文を見ることができるように作られている．構造化テキストを検索する表現式は，文字列のパターンを記述できる正規表現を用いるとともに，構造化テキストの定義に基づいて作成する．定義に基づいて作られた表現式は，構造化テキストに対して検索をかけると，ラベル・属性付き単語が埋め込まれている文の特定の属性にマッチする．これを利用して，誤用にマッチするように属性の組み換えやラベル・属性付き単語の組み合わせの設定を行い，誤用文が検出される表現式を作成する．表現式を用いた検索によって日本語学習者コーパスから取り出される文は，一般的な文字列検索とは異なり，多様な単語と表現による文の集合である．さらに，本研究の簡便性について述べた．

第 3 章では，提案した手法を用いて品詞の接続法誤りがある誤用文を検出する実験を行った．品詞の接続法に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着目した．実験の結果，誤用文を検出することができた．また，実験で検出できなかった誤用文を分析した．実験の結果と考察をふまえて，本研究の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，日本語として単語が正しく書かれていない誤りや形態素解析器と辞書の問題などから検出できない誤用文あったが，品詞の接続法誤りがある誤用文の有効性が確認できた．

第 4 章では，人が目視で確認した分類で同類とした誤用文を提案手法で再現できるか検証実験を行った．従来の分類の方法は，人が誤用の基準をめてその基準に従って誤用文を

分類している．本研究で提案した手法は，検索して見つけたものに取り出しを行っているが，これは機械が自動で誤用文を分類していると考えることができる．そこで，目視で同類であると分類した誤用文を提案手法で再現する検証実験を行った．再現の例には，動詞のテ形活用誤りとし，その誤用と表現式を一対一に対応させるのではなく作成して再現を行った．実験の結果，提案手法を用いて目視で同類である誤用を再現することができ，有効性が確認できた．

第 5 章では，提案手法を用いて効率的に誤用文を検出し，検出したデータを活用して学習者が誤りを原因の調査と分析を行い，提案した手法が誤用研究に有効であるか確認した．ここでの有効とは，誤用文の検出が目視や文字列検索よりも効率的であったか，さらに表現式検出法で取り出したデータが学習者の誤用を分析することができたかである．提案手法を用いて二つの日本語学習者コーパスから誤用文を検出した．誤用文を検出する表現式は，第 3 章で作成したものを再利用した．その結果，提案手法で取り出したデータは学習者が誤用を生み出す原因を分析することができ，また表現式が再利用できたこと，さらに二つの異なる日本語学習者のテキストコーパスから誤用文が検出できたことから，目視や文字列検索よりも効率的に誤用文を取り出すことができ，有効性が確認できた．

6．2　今後の課題

本研究において，提案した文に形態素情報を付けて表現式で検索して誤用文を検出する手法は，これまで提案された誤用文の検出手法ではなく，ユーザーが着目するような学習者の文法的な誤用文を検出する新たな手法である．形態素情報はツールを利用して簡便に付けることができ，表現式も形態素情報を把握して文にマッチする表現式を正しく作ることができれば，他のユーザーも誤用文を検出することができる手法である．また，表現式は柔軟な設定が可能である．したがって，ユーザー同士が同じ環境で表現式を持ち寄って，誤用文を検出することもできるし，持ち寄った表現式をツール化する展望も期待される．

本研究の実験においては，ユーザーが着目するようないくつかの誤用文の例に検出を行った．検出の結果と考察をふまえた検討から，他の誤用文の検出にも有効であると判断した．したがって，表現式検出法を用いた検出をさらに行い，それらの検証を行いたいと考えている．以上，他の誤用文の検出と検証，他のユーザーとの表現式の共有，表現式を集めたツール化の実現を中心とした研究を進めていきたい．

参考文献

[1] Komori, Saeko. (2003) A study of L2 lexical collocations of English-speaking learners of Japanese, 『第二言語としての日本語教育研究』6, pp.33-51.

[2] 大曾惠美子・滝沢直宏(2003)「コーパスによる日本語教育の研究―コロケーションおよびその誤用を中心に―」,『日本語学』22 巻 5 号, pp.234-244.

[3] 曹紅荃・仁科喜久子(2006)「中国人学習者の作文誤用例から見る共起表現の習得及び教育への提言―名詞と形容詞及び形容動詞の共起表現について―」,『日本語教育』130, pp.70-79.

[4] 陳曦(2007)「学習者と母語話者における日本語複合動詞の使用状況の比較―コーパスによるアプローチ―」,『日本語科学』22, pp.79-99.

[5] 大山浩美・小町守・松本裕治(2012)「日本語学習者の作文における誤用タグつきコーパスの構築について―NAIST 誤用コーパスの開発―」, テキストアノテーションワークショップ予稿集 w02, pp.1-8.

[6] 黄淑妙(2009)『日本語習得の達成度分析―「台湾人日本語学習者コーパス」(CTLJ)の構築と分析を中心に―』, 致良出版社.

[7] 金澤裕之(編) (2014) 『日本語教育のためのタスク別書き言葉コーパス』, ひつじ書房.

[8] 鎌田修(2006)「KY コーパスと日本語教育」,『日本語教育』130, pp.42-51.

[9] 杉村泰(2010)「中国語話者のための日本語教育文法の開発と学習者中間言語コーパスの構築」, 科学研究費助成金, 基礎研究(C), 成果報告書(平成 19～21 年度), pp.129-139.

[10] 陳淑娟(2008)「LARP at SCU についてのご案内」, 東吳大學 LARP at SCU 研究工作坊(一) 會議手册, pp.4-13.

[11] 八木豊・鈴木泰山(2012)「学習者作文コーパスの構築と誤用の分析」, 仁科喜久子(監)『日本語学習支援の構築―言語教育・コーパス・システム開発―』, 凡人社, pp.249-273.

[12] 林炫情・李在鎬・宮岡弥生・柴崎秀子・趙堈熙(2012)「言語処理技術を利用した日本語学習者作文コーパスの開発」,『日本文化學報』第 56 輯, pp.129-142.

[13] BTS による多言語話し言葉コーパスー日本語会話 2(日本人と学習者の会話) (2015 年 6 月閲覧) : http://www.tufs.ac.jp/ts/personal/usamiken/corpora.htm

[14] CbLLE 品詞検索エンジン(日本語学習者言語コーパスの品詞検索版) (2012 年 12 月閲

覧)：http://cblle.tufs. ac.jp/tag/ja/search.php?menulang=ja

[15] JLPTUFS 作文コーパス(2015 年 6 月閲覧)：http://www.tufs.ac.jp/ts/personal/SUZUKI_Tomomi/paper/JLPTUFS_Corpus_readme.pdf

[16] KY コーパス(2012 年 6 月閲覧): http://opi.jp/shiryo/ky_corp.html

[17] LARP at SUC(2012 年 12 月閲覧) : http://163.14.2.167/builder/web_page.php?web=156&pid=9346

[18] オンライン日本語誤用辞典(公開版 Ver.1.1) (2012 年 6 月閲覧) :http://cblle.tufs.ac.jp/llc/ja_wrong/index.php?m=default

[19] 学習者作文コーパス「なたね」(2013 年 12 月閲覧) : https://hinoki-project.org/natane

[20] 華東政法大学作文コーパス(2015 年 6 月閲覧)：http://www.lang.nagoya-u.ac.jp/~sugimura/class/corpus/zhengfa.html

[21] 生活場面で必要となる日本語書きことばデータ（生活作文 DB) (2013 年 12 月閲覧) : http://jpforlife.jp/seikatsu-sakubun.html

[22] 台湾人日本語学習者コーパス(2012 年 6 月閲覧)：http://corpora.flld.ncku.edu.tw/index.pl

[23] 寺村秀夫(1990)『外国人学習者の日本語誤用例集』(大阪大学 ; データベース版，国立国語研究所，2011 年) (2012 年 6 月閲覧)：http://teramuradb.ninjal.ac.jp/db/

[24] 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(2012 年 6 月閲覧)：http://www.tufs.ac.jp/ts/personal/ijuin/koukai_data1.html

[25] 日本語学習者言語コーパス(2012 年 6 月閲覧)：http://cblle.tufs.ac.jp/llc/ja/search.php?menulang=ja

[26] 日本語学習者作文コーパス(2015 年 6 月閲覧)：http://sakubun.jpn.org/

[27] 日本語学習者による日本語／母語発話の対照言語データベース（発話対照 DB) (2013 年 12 月閲覧)：http://jpforlife.jp/hatsuwadb.html

[28] 日本語学習者による日本語作文と，その母語訳との対訳データベース(作文対訳 DB) (2013 年 12 月閲覧)：http://jpforlife.jp/taiyakudb. html

[29] オンライン日本語誤用辞典「分類」「タイプ」「キーワード(KW)」「記号」(公開版 ver.1.1：2012-03-26 現在) (2012 年 12 月閲覧)：http://cblle.tufs.ac.jp/llc/ja_wrong/goyo_bunrui-type-kigo.pdf

[30] 寺村誤用集データベース 操作説明書(2015 年 5 月閲覧)：http://teramuradb.ninjal.

ac.jp/db/manual/teramuradb.manual.pdf

[31] 市川保子(2001)「日本語の誤用研究」『日本語教育通信』40 号，pp.14-15.

[32] Corder, S. P. (1967) The Significance of Learners' Errors. Robinett, B.W., and Schachter, J. (Eds.), In *Second Language Learning: Contrastive Analysis, Error Analysis, and Related Aspects.* Ann Arbor: University of Michigan Press, pp.163-172.

[33] Selinker, L. (1972) Interlanguage. Robinett, B.W., and Schachter, J. (Eds.), In *Second Language Learning: Contrastive Analysis, Error Analysis, and Related Aspects.* Ann Arbor: University of Michigan Press, pp.173-196.

[34] 水本智也・小町守(2012)「なんで日本語はこんなに難しいなの？―リアルな日本語学習者コーパスの分析と言語処理の課題―」，『情報処理』Vol.53，No.3，pp.217-223.

[35] 浅原正幸・小野創・狩野芳伸(2012)「コーパスアノテーションと心理言語学」(2014 年 1 月閲覧): http:// www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop_no1_papers/JCLWorkshop2012_54.pdf

[36] 高見澤孟・ハント蔭山裕子・池田悠子・伊藤博文・宇佐美まゆみ・西川寿美(2004)『新・はじめての日本語教育 1 日本語教育の基礎知識 新・はじめての日本語教育』，アスク.

[37] 西尾寅弥(1990)「9. 形容動詞」，日本語教育学会(編)『日本語教育ハンドブック』，大修館書店，pp.452-455.

[38] 茶筌(2014 年 1 月閲覧)：http://chasen-legacy.sourceforge.jp/

[39] ipadic version 2.7.0(2015 年 6 月閲覧)：http://chasen.naist.jp/snapshot/ipadic/ipadic/doc/ipadic-ja.pdf

[40] 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，日本語作文データ入力マニュアル(2013 年 11 月閲覧)：http://www.tufs.ac.jp/ts/personal/ijuin/data_nyuryoku_manual.pdf

[41] 黄淑妙・山本卓司・関口要(2009)「「台湾人日本語学習者コーパス」(CTLJ)試行版の公開」『台灣日本語文學報』25，pp.269-292.

[42] 文化庁(2013)「平成 25 年度 国内の日本語教育の概要」，文化庁文化部国語課(2015 年 6 月閲覧)：http://www.bunka.go.jp/tokei_hakusho_shuppan/tokeichosa/nihongokyoiku_jittai/h25/pdf/h25_zenbun.pdf

[43] 大名力(2012)『言語研究のための正規表現によるコーパスの検索』，ひつじ書房.

[44] Strawberry Perl 5.14.2.1(2014 年 9 月閲覧)：http://strawberryperl.com/

[45] Padre, the Perl IDE(2014 年 9 月閲覧)：http://padre.perlide.org/

[46] 市川保子(編著)(2010)『日本語誤用辞典 外国人学習者の誤用から学ぶ日本語の意味用法と指導のポイント』，スリーエーネットワーク.

[47] 張麟声(2001)『日本語教育のための誤用分析―中国語話者の母語干渉 20 例―』，スリーエーネットワーク.

[48] 水谷信子(1994)『日本語の教え方・実践マニュアル 実例で学ぶ誤用分析の方法』，アルク.

[49] ChaKi(2015 年 6 月閲覧)：https://osdn.jp/projects/chaki/

[50] MeCab(2015 年 6 月閲覧)：http://taku910.github.io/mecab/

[51] KH Coder(2015 年 6 月閲覧)：http://khc.sourceforge.net/

[52] 池原悟・小原 永・高木 伸一郎(1993)「2. 書校正支援システムにおける自然言語処理(<特集>自然言語処理技術の応用)」，『情報処理』34(10)，pp.1249-1258.

[53] 笹野遼平・黒橋禎夫(2007)「形態素解析における連濁および反復形オノマトペの自動認識」，言語処理学会 第 13 回年次大会，pp.819-822.

[54] 今枝恒治・河合敦夫・石川裕司・永田亮・桝井文人(2003)「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(CE), 13, pp.39-46.

[55] 三浦雅則・横山昌一(2009)「留学生の日本語助詞修正システム」，情報処理学会 第 71 回全国大会，2，pp.279-280.

[56] 橋本利典・島田静雄(1997)「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 自然言語処理研究会報告(NL)，4，pp.9-14.

[57] 南保亮太・乙武北斗・荒木健治(2007)「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 自然言語処理(NL)，94，pp.107-112.

[58] 荒木健治(2004)『自然言語処理ことはじめ―言葉を覚え会話のできるコンピュータ―』，森北出版.

[59] 奥村学(2010a) 『言語処理のための機械学習入門』，コロナ社.

[60] 奥村学(2010b) 『自然言語処理の基礎』，コロナ社.

[61] Oyama, Hiromi. (2010) Automatic Error Detection Method for Japanese Particles.

Polyglossia, 18, pp.55-63.

[62] 大山浩美(2010)「日本語学習者支援のための機械学習による正誤判定について」,『Language issues: the international journal of the Academic Information and Media Center』16(1), pp.87-99.

[63] Mizumoto, Tomoya. Komachi, Mamoru. Nagata, Masaaki. Matsumoto, Yuji.(2011) Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners. International Joint Conference on Natural Language Processing (IJCNLP2011), pp.147-155.

[64] 赤野一郎(2006)「英語コーパス言語学と英語教育」『日本語教育』130, pp.11-21.

[65] 石川慎一郎(2012)『ベーシックコーパス言語学』, ひつじ書房.

[66] 広瀬和佳子(2010)「学習者の作文に対する解釈の多様性と「添削」の限界―日本語教師の添削過程の分析を中心に―」『早稲田日本語教育学』8, pp.29-43.

[67] 曹紅荃・仁科喜久子(2006)「自由産出調査から見る形容詞および形容動詞と名詞の共起表現―学習者と母語話者の対照を通して―」, 電子情報通信学会技術研究報告 TL, Vol. 106, No. 363, pp.31-36.

[68] 庵功雄・高梨信乃・中西久実子・山田敏弘, 松岡弘(監) (2000)『初級を教える人のための日本語文法ハンドブック』, スリーエーネットワーク.

[69] スリーエーネットワーク(編) (2000)『みんなの日本語初級 I 教え方の手引き』, スリーエーネットワーク.

[70] 文化庁(2010)「常用漢字表」(2014 年 9 月閲覧) : http://www.bunka.go.jp/kokugo_nihongo/pdf/jouyoukanjihyou_h22.pdf

[71] 小学館(編)(2003)『日本国語大辞典』第 2 版, JapanKnowledge+(2013 年 11 月閲覧) : http://www.japanknowledge.com/top/freedisplay

[72] 北澤尚・李金蓮(2009)「日本語の接尾辞「的」に対応する中国語表現について」『東京学芸大学紀要 人文社会学系 I 』60, pp.161-176.

[73] 望月通子(2010)「接尾辞「～的」の使用と日本語教育への示唆―日本人大学生と日本学習者の調査に基づいて―」『関西大学外外国語学部紀要』2, pp.1-12.

[74] 輿水優(1985)『中国語研究学習双書 8 中国語の語法の話―中国語文法概論―』光生館

[75] 中央研究院現代漢語平衡語料庫(2013 年 11 月閲覧) : http://app.sinica.edu.tw/

kiwi/mkiwi/

[76] 中文斷詞系統(2013 年 11 月閲覧)：http://ckipsvr.iis.sinica.edu.tw/

[77] 少納言(BCCWJ)（2013 年 11 月閲覧)：http://www.kotonoha.gr.jp/shonagon/search_form

[78] 山崎誠(2009)「代表性を有する現代日本語書籍コーパスの構築」『人工知能学会誌』24(5)，pp.623-631.

[79] 石川慎一郎(2008)「コロケーションの強度をどう測るか―ダイス係数，t スコア，相互情報量を中心として―」, 言語処理学会第 14 回大会チュートリアル資料, pp.40-50.

[80] Church, Kenneth W.,William Gale, Patrick Hanks and Donald Hindle. (1991) Using statistics in lexicalanalysis. Zernik, Uri. (ed.), In *Lexical Acquisition: Using On-lineResources to Build a Lexicon*. Hillsdale, N.J.: Lawrence Erlbaum Associates, pp.115-164.

[81] Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

業績リスト

ポスター発表(審査あり)

1. 山本卓司・福川美沙・田中順子・森下淳也・中島和子(2013)「継承語文献データベース(HLDB)―公開に向けて―」，2013 年度母語・継承語・バイリンガル教育(MHB)研究会 10 周年記念大会 予稿集，pp.50-51.

学会・研究会発表(審査あり)

1. 山本卓司・中尾桂子(2012)「日本語学習者の添削済作文に対する違和感とその要因について」，第 21 回小出記念日本語教育研究会 予稿集，pp.43-46.
2. 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.
3. 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.

国際会議(審査あり)

1. 関口要・山本卓司(2012)「場面描写に見るテンスとアスペクト―日本人と台湾人日本語学習者による作文の比較―」，2012 年日語教學國際會議論文集，pp.253-267.
2. 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 年國際学術研討會―應用日語教育的理論與實践―大會論文集，pp.41-48.
3. 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.

原著論文(審査あり)

1. 文健・山本卓司・孫一・清光英成・大月一弘(2015)「学内教員に対する連携本部の情報発信における ICT 利用の効果判定方法の試案―大学組織文化の特徴と産学連携の実態を主な考慮点に―」，『産学連携学』，Vol.11(1)，pp.17-24.

謝辞

本研究の機会を賜り，神戸大学大学院国際文化学研究科グローバル文化専攻情報コミュニケーションコースにおいてご指導を賜りました神戸大学大学院国際文化学研究科 森下 淳也 教授 に深く感謝致します．自身も研究協力者として参加し公開した日本語学習者コーパスが，多くの方に利用されないのは誤用が原因だと気づき，先生にご相談申し上げた折に快く指導教官になってくださいましたこと，今でも本当に感謝しております．また，私の研究の進みが遅いために，先生にはいつもご心配とご迷惑をおかけしました．しかし，先生は，いつも丁寧なご指導と激励をしてくださいました．本当に感謝しております．ありがとうございました．

本研究をまとめるにあたり，多くの有益なご助言を賜りました神戸大学大学院国際文化学研究科 大月 一弘 教授 に深く感謝致します．先生から賜ったお言葉がどれだけ研究に役立ち，どれだけ励みとなったか，計り知れません．先生には深く感謝の意を表します．

本研究に多くのご協力を賜りました神戸大学大学院国際文化学研究科 清光 英成 准教授に深く感謝致します．発表原稿がぎりぎりになったとき，先生はいつも夜遅くまでお付き合いくださいました．本当に感謝しております．

いつも丁寧なご指導を賜りました神戸大学大学院国際文化学研究科 康 敏 教授，村尾 元 教授，西田 健志 准教授 に深く感謝致します．研究室の垣根を越えた自由で明るい雰囲気のなか，先生方から賜ったご指導により，充実した研究生活を過ごすことができました．本当にありがとうございました．

学位審査の依頼を快くお引き受けくださいました 神戸大学 水野 マリ子 名誉教授，神戸大学大学院国際文化学研究科 田中 順子 教授 に深く感謝致します．先生方の TA をさせて頂いた際には，大変勉強になりました．また，遅筆なために，ご迷惑をおかけしたにも関わらず，先生方は審査委員を快く引き受けてくださいました．大変感謝しております．

研究に関して，ご助言を賜りました 神戸情報大学大学院 孫 一 助教 に深く感謝致します．研究を進めるうえで，頂いたお言葉は大変励みになりました．厚く御礼申し上げます．

日々の研究を支えてくださいました 神戸大学大学院国際文化学研究科グローバル文化専攻情報コミュニケーションコース 文 健 さん に深く感謝致します．研究を進める上で，文さんはいつも刺激を与えてくださいました．本当にありがとうございました．

お互いに研究を励まし合った情報コミュニケーションコース博士課程後期 桑野 徹也さん，Dick Martínez Calderón さん，張 帆 さん，同コース博士課程前期の皆さん，OB・OG の皆さんに御礼申し上げます.

最後に，研究生活を天国から支えてくれた 父 へ，いつも体を気遣ってくれた 母 へ，陰ながら支えてくれた 兄 へ，惜しみない支援と励ましをくれた 久保田 佐和子 へ 心からの感謝と御礼を申しげます.