

日本語のガ格に対する副助詞「は」の使用の推定

横野 光^{1,a)} 稲邑 哲也^{1,b)}

概要: 本論文では日本語のガ格要素に対して、副助詞「は」を用いて表現するかどうかを推定するモデルを提案する。「は」は主題を表すために用いられることが多く、提案モデルはそれに着目し前文脈において対象要素がどのような使われ方をしているかを考慮する。新聞記事を対象とした実験により、文中要素のみに着目した既存手法よりも提案モデルの方が良い性能を示すことが明らかになった。

キーワード: 「は」と「が」、自然言語生成、教育応用

Estimation of Suitability of Adverbial Particle “ha” for Japanese Ga-case

Abstract: In this paper, we propose a model that estimates whether adverbial particle “ha” is suitable for target ga-case in a Japanese sentence. Adverbial particle “ha” is often used to indicate a topic of the sentence. Our model focuses on this perspective and considers coreferential elements in the context. An experimental evaluation shows that our model outperforms the existing model and that contextual features are effective.

Keywords: “ha” and “ga”, natural language generation, application to education

1. はじめに

書き手が自分の意図を正しく伝えるためには、その読み手にとって分かりやすい文章を作成することが重要である。この、分かりやすい、というのは内容の正確さとは異なる要素であり、例えば、読みやすさや首尾一貫性といったことが影響する。

たとえ文章の内容が正確であっても、その表現方法や展開の仕方によっては読み手に負荷を与えたり、場合によっては誤った解釈を促すことになってしまい、文章の本来の意図が読み手に伝わらないということを引き起こし得る。読み手に負荷を与えないような自然な文章の書き方は統語規則とは異なり、母語話者の直観に頼るところが多く、規則として表現することは困難である。

文章の自然さに関わる文法項目の一つに副助詞「は」の使用がある。副助詞「は」は主題を示すために用いられることが多く [1]、長い文章ではその時点で何が焦点となっ

ているかを明らかにするという役割を持つ。しかし、正しく使用しないと読み手が本来の意図とは異なる解釈をしてしまう可能性が出てくる。例えば、

(1) 太郎が花子にコーヒーを入れた。

(2) 花子 は それを飲んだ。

という文の流れは自然であるが、文 (2) を

(2') 花子 が それを飲んだ。

に変えると、内容としては文 (2) と同じことを表現しているが、文 (2') は読み手によっては違和感を覚えたり、文 (2) とは異なる印象を受けることがある。

副助詞「は」は格助詞「が」や「を」の代わりに用いられったり、格助詞「に」や「から」に接続して用いられるなど、文中の様々な場所に出現し得るが、特に述語のガ格において使用されることが多い。「は」と「が」の使い分けの基準は曖昧であり、上述の例のように母語話者であっても“何故この場合「が」よりも「は」を用いた方が自然であるか”という問いに答えられない場合も少なくない。

このため「は」の使用の習得は比較的困難であると考えられる。特に日本語学習者にとっては大きな壁となり得る。また、日本語母語話者であっても小論文の作成などで同様の状況に遭遇することがある。例えば、中学生の国語教育

¹ 国立情報学研究所
National Institute of Informatics

a) yokono@nii.ac.jp

b) inamura@nii.ac.jp

においてもこの使い分けは課題となっている [2]。教師の指導を受ければある程度の知識を得ることができるが、学習者が自身の作成した文章の全てに対して添削を受けることは困難であるため、自動添削システムや校正支援システムなどが重要となる。

このような背景から本研究では教育への応用を目的とした副助詞「は」の使用モデルの提案を目指す。しかし、前述の通り、「は」は文中の様々な場所で用いられるため、まず最初の段階として、「は」が特に述語のガ格に対して用いられることが多いことから、このガ格を対象を限定した使用モデルを提案する。

以下、2節で関連研究を述べ、3節で副助詞「は」の統語的な性質について述べる。その知見をもとにしたモデルを4節で提案し、5節で提案モデルの評価実験について述べ、6節でまとめと今後の課題を述べる。

2. 関連研究

近年、非母語話者に対する言語学習支援や文章校正支援など教育応用に焦点を当てた自然言語処理研究が多くなされるようになった。

日本語学習者に多い誤りには助詞の使用に関するものがあり、そのため日本語学習者支援に関する研究においては、助詞の誤り訂正に関する研究が多くなされている。例えば、笠原らは誤り傾向を考慮した格助詞の訂正手法 [3] を、また、今村らは誤り訂正を元文章から修正文章への翻訳として捉え、中国語母語話者の日本語作文の助詞誤りに対する助詞誤り訂正手法 [4] をそれぞれ提案している。

統計的手法に基づいた誤り訂正では一般的に非母語話者によって作成されたテキストとそれを修正したテキストが必要となる。このようなデータを集めた学習者コーパスは存在する (e.g. Konan-JIEM Learner Corpus) が、コーパスの作成にはコストがかかる。Mizumoto らは言語学習者向けの SNS に着目し、そのサービスのユーザが作成したデータを用いた誤り訂正手法を提案している [5]。

「は」と「が」の使い分けは日本語母語話者にとっては比較的容易であるが非母語話者にとっては理解が困難なものである。英語における同様の文法項目としては、定冠詞と冠詞の使い分けや前置詞の使用などが挙げられ、これらの誤り訂正に関しても研究が行われている。例えば、Felice らは第二言語学習者の冠詞誤り訂正手法を提案している [6]。また、Boyd らは単語のクラスターを考慮した前置詞の誤り訂正手法を提案している [7]。

一方、母語話者に対する応用としては、作文の自動採点などがあり [8]、例えば、藤田らは SVR による自動評価モデルを提案している [9]。

助詞「は」と「が」の使い分けの推定に関しては、三浦らは同文中の要素に着目したモデルを提案している [10]。これに対して、本研究で提案する手法は同文中の要素に加

えて前文脈を考慮したモデルである。

3. 「は」の用法

副助詞「は」には大きく分けて、主題を示す「は」と対比を示す「は」の2通りの用法があるため、「は」と「が」の使い分けに関してはガ格の要素を主題として扱うかどうか、主語を取り立てて表すかどうかという点が影響する [1]。即ち、文中の要素が主題として扱われているのであれば、その要素は「は」を用いて表されうることである。

文が主題を取るかどうか、という点は叙述内容と関連する。例えば、

(3) コーヒー は 苦い。

のように、属性を述べる文では主語は主題となりやすい。逆に、話し手が発話時点で知覚した内容を表す文や、過去の事態を客観的に報告するような文では、主題を取るとは少なくなる。

談話においては、話し手が、聞き手が指示対象を特定できると考えているものについてはそれを主題として示すことができる。例えば、前文脈において既に主題として取り上げられている、主題ではないが既出である場合、或いは前文脈には直接現れていないが、前文脈に出現している要素と何らかの関係がある場合には主題として表すことができる。そのような要素に対しては「は」を用いることができる。また、聞き手と話し手の共有の知識であるようなものに関しては、前文脈に出現していなくても、主題として取り上げることができる。

節に関しては、従属節が主題を取り得るかどうかは主節との従属度によって決まり [11]、その度合いは節の種類によって異なる。

「～ながら」や「～つつ」などで表される従属節は従属度が高く、主節に対する独立度が低いため、主節とは別に節内で主語を取ることができない。従って、この節内の述語のガ格は主節と共有されているため、「は」と「が」の使い分けは問題にならない。

「～ので」や「～など」のような従属度が中程度の従属節に関しては、その節内の述語は主節とは異なる主語を持つことができるが、その主語を主題として扱うことはできない。

従属度が低い等位節や「～から」節では、節内の述語のガ格を主題として表すことができる。

これらのことから、どの節のガ格であるかによって「は」を使用できるか否かがある程度は推測できると考えられる。

一方、対比の用法に関しては、典型的には対比される要素が明示されるが、片方しか明示されず、比較対象は文脈や語彙知識から推測される場合もある。対比を表す「は」は平叙文でよく用いられるが、疑問文や、意志、勧誘、行為要求を表す文で用いられることもある。

4. 副助詞「は」の使用推定モデル

本研究では、「は」と「が」の使い分けの問題を、述語のガ格に対して副助詞「は」を用いて表現するか否かを推定する二値分類問題としてモデル化する。従って、モデルが「は」を使用しないと分類した場合は格助詞「が」が用いられると解釈する。

3節で述べたように、副助詞「は」には主題を表す用法があり、また、どの節の述語の要素であるかということが関係する。このことを考慮して、本研究で提案するモデルは推定の対象となる要素が文章中でどのように用いられたのかという情報と、推定対象の要素を含む文において他の要素がどのように用いられているのかという情報に着目する。

与えられた文章のどのような情報を推定に利用するかは、どういう応用を想定するかによって異なってくる。一般的なタスクでの入力としてのテキストは書き手から離れた状態で与えられることが多く、そのような場合、表層的な情報以外は何らかの手段で推定する必要がある。これに対して本研究は日本語学習支援や文章作成支援などを考えており、書き手がテキストを書いている段階でのモデルの適用を想定している。従って、共参照関係などの情報などは書き手が直接その場で入力できると考え、共参照情報や述語の格などは既知であるとする。

具体的には本研究では前提として以下の条件を設定する。

- 推定対象の要素が文中に出現する
- テキスト中の要素の共参照関係は既知である
- 述語の格要素は既知である

自然な文章の作成支援ということを考えると、ガ格に対しては「は」と「が」の使い分け以外に、その要素を省略するか否かという選択があるが、本研究では扱わない。

以降、提案モデルにおいて副助詞「は」の使用を推定するガ格要素のことを**推定対象要素**と呼び、推定対象要素をガ格とする同文中の述語を**対象述語**と呼ぶ。直接係り受けの関係にない場合であっても、推定対象要素をガ格とする述語が同文中に出現していれば、その述語は対象述語とみなす。従って、一つの推定対象要素に対して複数の対象述語が存在する場合もある。

一文中に複数の推定対象要素が存在する場合があるが、本研究ではある要素の推定時には他の推定対象要素には正解が与えられているものとする。例えば、

(4) 太郎は花子が入れたコーヒーを飲んだ。

という文では「太郎」と「花子」が推定対象要素であるが、「太郎」の推定時には「花子」は「が」を取っていることが分かっていることにする。

以降、提案モデルで考慮する素性について述べる。

推定対象要素についての素性

推定対象要素については以下の4種類の素性を考慮する。

- 推定対象要素が与えられたテキストにおいて初出であるか否か
- 代名詞か否か
- 推定対象要素から最も近い対象述語までの間に他の要素をガ格とする述語があるか否か
- 推定対象要素から最も遠い対象述語までの間に推定対象要素をガ格とする述語があるか否か

下2つの素性は、格助詞「が」を用いて表現されると直近の述語に対するガ格であると解釈しやすい傾向があるという直観に基づいている。

対象述語についての素性

- 述語の原形
- 述語の品詞
- 述語に後接する機能表現
- 述語のカテゴリ

述語のカテゴリは竹内らの動詞項構造シソーラス [12] の大分類1を利用する。

一文中に複数の対象述語が存在することがあるため、対象述語が文中のどの節にあるかによってこれらを区別する。例えば、

(5) 太郎は昨日買った小説を読んだ。

という文では、「太郎」を推定対象要素としたとき、主節にある述語の原形は「読む」であり、連体修飾節の述語の原形は「買う」である、というように素性に加える。複数の同じ種類の節が文中に存在する場合は、主節に近い節のみを対象とする。

推定対象要素を含む文中の他の要素についての素性

推定対象要素を含む文において推定対象要素以外の、文中の述語の格となる要素に対して、以下の素性を考慮する。

- 副助詞「は」を用いて表現されているか否か
- 省略されているか否か

ここで考慮する要素は、その要素が属している節と、述語の何格かによって区別する。例えば、

(6) 太郎は花子が入れたコーヒーを飲んだ。

という文の「花子」が考慮する要素であるとする、連体修飾節のガ格がどのような要素であるかという観点から前述の素性を作成する。

推定対象要素と共参照関係にある要素についての素性

推定対象要素を含む文から m 文前までを前文脈とし、この文脈に含まれる推定対象要素と共参照関係にある要素について以下の素性を考慮する。 m の値はあらかじめ決定しておく。

- 副助詞「は」を用いて表現されているか否か

● その要素と推定対象要素との表層の違い (完全に一致, 主辞のみ一致, 代名詞, 省略, その他, のいずれか)
各共参照関係にある要素は前述の推定対象要素を含む文中の他の要素についての素性と同様に, その要素が属している節と述語の何格かによって区別する。

また, この素性に関しては, 何文前に共参照要素が存在するかによって素性を区別する。即ち, 1文前に存在する共参照要素と2文前に存在する共参照要素はそれぞれ別の素性とみなす。

その他の素性

上記以外の素性として以下のものを考慮する。

- 推定対象要素を含む文の文頭に接続詞がある場合, その表層形

5. 評価実験

提案モデルの有効性を検証するために新聞記事を用いた実験を行った。

5.1 実験設定

本研究では表層に現れているガ格を副助詞「は」を用いて表現するかどうかを推定する。従って, 必要な格の情報は述語の原形に対する格 (深層格) ではなく, 表層格である。そのため, 本実験には NAIST テキストコーパスに表層格の情報がアノテーションされたものを利用した。

モデルの学習には 1995 年 1 月 1 日から 11 日までの記事と 1995 年 1 月から 8 月までの社説を, テストに 1995 年 1 月 14 日から 17 日までの記事と 1995 年 10 月から 12 月までの社説を利用した。

前節で述べたように提案モデルでは節の種類を素性に取り入れている。節境界の同定と節の種類推定には丸山らの節境界検出プログラム CBAP を利用した [13]。CBAP では節の分類として益岡, 田窪 [14] の分類を参考に 4 種類の大分類と 10 種類の小分類を定義し, さらにその小分類を形態的特徴などで細かく分類したものを利用している。本研究では一番細かい分類のラベルを節の種類として利用した。

分類モデルには Support Vector Machine(SVM) を採用し, 実装には TinySVM^{*1} を利用した。カーネルは線形カーネル, モデルパラメータの調整は行わず, デフォルトの値である。

提案モデルでは何文前までを前文脈として考慮するかを決定する必要がある。そこで予備実験として, コーパスの 1995 年 1 月 12 日, 13 日の記事と 1995 年 9 月の社説を用いて, 文脈の数を変えながら 10 分割交差検定を行い, もっとも良い正解率を示した値を実験に利用する。

考慮する文脈の数 $m(0 \leq m \leq 6)$ に対する予備実験の結果を表 1 に示す。

表 1 予備実験結果

| m | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|-----|--------|--------|--------|--------|--------|--------|--------|
| 正解率 | 0.7501 | 0.7501 | 0.7515 | 0.7519 | 0.7519 | 0.7510 | 0.7490 |

この結果から本実験では対象要素を含む文から 3 文前までを前文脈として考慮する。また, この結果から対象要素を含む文のみのモデル ($m = 0$) に比べて, 前文脈の共参照関係にある要素を考慮した方が良い結果になることが明らかになった。

比較手法として三浦らのモデル [10] を利用した。

5.2 結果と考察

実験結果を表 2 に示す。majority は全ての事例に対して訓練データにおいて最も多かった方 (本実験の場合は「が」) を割り当てた場合を示す。

表 2 実験結果

| | majority | 比較手法 | 提案手法 |
|-----|----------|-------|--------------|
| 正解率 | 0.518 | 0.734 | 0.760 |

提案モデルは比較手法と比べて良い結果を得ることができた。

提案モデルの正解と出力の対応を表 3 に示す。

表 3 分類表

| 正解 \ 出力 | は | が |
|---------|------|------|
| は | 3168 | 916 |
| が | 1114 | 3269 |

実験に使用したデータでは「は」と「が」の分布に偏りはなく, システムの出力もそれほど偏りはない。

正しく推定できた事例を図 1 に, 誤って推定した事例を図 2 に示す。太字は推定対象要素であり, 下線部が提案モデルによる推定結果である。

図 1 の 1 番目の事例では推定対象要素の“SMA”は前文脈において出現した要素であるため主題となる可能性があり, このことから正しく推定できたと考えられる。2 番目の事例においても対象要素は前文脈に出現しているが, 同文で既に「は」が使われていること, 従属節内の述語のガ格であるということから正しく推定できたと考えられる。

一方, 図 2 の 1 番目の事例は文章の 1 文目であり, “近畿”という単語は初出であるため「が」の使用が考えられるが, 正解は「は」である。これは, ここでの「は」は主題を表すために用いられているというわけではなく, その前にある“北陸, 関東, 九州”との対比を示すために使われていると考えられる。

提案モデルは, 前文脈での推定対象要素の現れ方に注目

*1 <http://chasen.org/~taku/software/TinySVM/>

- 遺伝性の重い神経疾患である脊髄性筋萎縮症の原因と思われる遺伝子を日本とカナダの国際共同研究グループが発見した。SMAの診断や治療に結びつく成果で、十三日付の米国の論文誌「セル」に発表した。
SMA は常染色体劣性の遺伝性疾患で、脊髄中の神経細胞が障害を受けて筋肉が委縮し、手足のまひや呼吸障害が起きる。
- 社会党の山花貞夫・新民主連合会長らは十七日に衆参両院で新たな国会内会派を届け出ることを党に通告したが、新会派結成のためには、まず、現在の社会党会派からの離脱手続きが必要。
山花氏らにとっては、**社会党** が離脱を認めるかどうか、最初の関門となる。

図 1 正しく推定できた事例

- 高成長を続けるサービス業の中堅企業は、北陸、関東、九州などに多く、**近畿** が低調なことが、ニッセイ基礎研究所のまとめたレポート「一九九五年度主要産業の展望」でわかった。
- 同教育庁は「拒否の理由があいまい。受け入れの余裕があれば拒否できない」として、調査のうえ強制命令を出す見通しだ。南アでは五年前に集団地域法が廃止されるまでは、法的にも白人と黒人の居住地域が分離され、学校も人種別だった。同法廃止後も、概して人種ごとに収入が違ってくるから人種別の居住地域は続いており、白人地域の学校の方が施設が整備され、**教育水準** は高い。

図 2 誤って推定した事例

している。これは間接的に前文脈で主題として使われているかという点を考慮しているとみなすことができるため、提案モデルは暗黙的に主題の「は」に特化したモデルとなっており、これがこの誤りの原因であると考えられる。

また、図 2 の 2 番目の事例に関しては、出力は正解とは異なっているものの、この部分を見る限りはそれほど不自然には見えない。1 節で述べたように、副助詞「は」の使用には明確な基準は存在せず、人によって異なることがある。つまり、場合によっては「が」と「は」のどちらも使用することができるということであり、正解と異なるからといって必ずしもその使用が誤っていると判断することはできない。

そこで、「は」と「が」の使い分けを実際に人が判断したデータを用いて評価実験を行った。実験に使用したデータは飯田ら [15] によって作成されたものである。

飯田らは新聞記事のテキストを用いて、文中の述語のガ格、ヲ格、ニ格を格助詞で表現するか、副助詞「は」を用いるか、省略するかを 3 人の作業者がそれぞれ判定したデータを作成している。

このデータにおいてアノテーションされたガ格のうち、提案モデルの前提に合致するもの、すなわち、同文中に出現している要素のみを対象とした。

3 人の作業者全員が同じ判定をしたものを正解としたデータを strict、2 人以上が同じ判定をした物を正解としたデータを majority と呼ぶ。各設定において、その値が“省略”になった場合、その要素は対象外とした。これは本提案システムが対象要素が出現することを前提としているためである。

また、作業者の判定が揺れていても全員がどれか 1 つの選択を行わなかった事例は、正解は複数考えられるが選ば

なかったその 1 つは誤りであるとみなすことができる。例えば、ある事例に対して 3 人の判定が「省略、省略、「は」」であった場合、少なくともそこに「が」を用いるのは適切ではない、と作業者が判断したと考えられる。そこで、作業者全員が「は」かまたは「が」を選ばなかった事例に対して、提案モデルがそれを選ばなければ正解と見なした評価も行った。この設定を negative と呼ぶ。

各設定において、アノテーション結果を正解としたときの評価を human、評価対象となった事例に対して元テキストを正解としたときの評価を original とする。それぞれの設定におけるシステムの正解率を表 4 に示す。

表 4 人手データとの比較

| 設定 | human | original |
|----------|-----------------|-----------------|
| strict | 0.723 (94/130) | 0.808 (105/130) |
| majority | 0.658 (123/187) | 0.840 (157/187) |
| negative | 0.866 (142/164) | 0.884 (145/164) |

どの設定においてもコーパスに現れた事例を正解と見なした場合と比べて結果は良くない。これは新聞記事での「は」と「が」の使い分けには、普通の文章での「は」と「が」の使い分けの仕方と異なる特有の特徴があり、これに対して、提案モデルが新聞記事に表れている「は」と「が」の使い分けを学習に用いているため、このような結果になったと考えられる。

作業者の判断が揺れなかった事例 (strict) と多少揺れた事例 (majority) では、前者の方が正解率が高く、判断が揺れるような事例はシステムにとっても判断が難しい事例であることが分かる。また、negative の結果は他の設定に比べて良く、複数の可能性がある場合でも、少なくとも不適切な使用を検出するという目的でこのモデルを適用するこ

- たとえば、高倉健と倍賞千恵子のような俳優を頭に置いて、一杯飲み屋での“告白”シーンを想像してみましょう。
暖簾を下ろした小さな呑み屋。
倍賞がおでんの火なんか見ながら、「明日は雪かもね」とつぶやきます。
高倉(は | が) 生返事をしただけで、ぐっと日本酒をあおります。
- しかし、金泳三大統領は「混乱を避けるためにも、歴史の判断にまかせる問題である」との立場を表明し、韓国の検察当局も「起訴猶予処分」にしてきた。
ところが、秘密資金口座問題で盧前大統領が収賄容疑で逮捕されたことをきっかけに、金大統領(は | が) 旧政権時代の問題清算に踏み切った。

図 3 作業者の判断が揺れた事例

とは可能であると考えられる。

どのような事例に対しては判断が揺れるのかを分析するため、実験に使用したデータにおいて、実際にガ格に対して作業者の判断が「が」と「は」で揺れた事例を図 3 に示す。

最初の事例では、2 文目から事象を述べる文が続いており、対象の「高倉」に対して、「が」を使うと事実を淡々と述べるという印象を受け、「は」を使うと、この文で場面の主題が「高倉」に固定され、以降これを中心に話が展開するという予測ができる。2 番目の事例は、前文脈に共参照関係にある要素が主題として出現しているために、引き続き主題として「は」を用いたということが考えられるが、一方で、「が」を用いた理由としては、接続詞“ところが”によってこれまでの文脈と対象の文との間に転換が生じ、改めて述べる必要があると見なしたからと考えられる。

いずれも、受け取るニュアンスの違いであり、どちらを選択しても文章全体の基本的な構造には変わりはないと考えられる。

6. おわりに

本論文では、文中の述語のガ格要素に対して、それが文の表層に現れる場合、格助詞「が」を用いて表すか、副助詞「は」を用いて表すかを推定するモデルを提案した。

提案モデルは推定対象の要素が前文脈においてどのような使われ方をしているかを考慮しており、文中要素の情報のみを利用したモデルよりも良い性能であることが新聞記事を用いた実験によって明らかになった。

提案モデルは既存モデルと同様に、対象をガ格のみに限定し、推定対象以外の要素の用いられ方は正解を利用している。しかし、実際には「は」は他の格に対しても用いられ、また、一文中に複数のガ格が現れることもあるため、少なくとも文を単位として、その中に含まれる全ての推定対象の用いられ方を決定する必要がある。

また、「は」には一般的に主題を表す用法と対比を表す用法がある。提案モデルは前文脈からの遷移を注目しており、これは暗黙的には主題の「は」のみに限定している。対比の「は」の考慮に関しては、要素間の類似度や構造間

の類似度を考慮する必要がある。さらに、より正確な推定のためには、そもそも文中の「は」がどちらの用法で用いられているかを推定する必要があると考えられる。

謝辞 本論文での実験に用いたコーパスは東京工業大学飯田龍氏に提供していただいた。記して謝意を表す。

参考文献

- [1] 日本語記述文法研究会 (編) : 現代日本語文法 5, くろしお出版 (2009).
- [2] 伊坂淳一: 中学生の日本語表現における文法的不適格性の分析, 千葉大学教育学部研究紀要第 60 巻, pp. 63–71 (2012).
- [3] 笠原誠司, 藤野拓也, 小町 守, 永田昌明, 松本裕治: 日本語学習者の誤り傾向を反映した格助詞訂正, 言語処理学会第 18 回年次大会, pp. 14–17 (2012).
- [4] 今村賢治, 齋藤邦子, 貞光九月, 西川 仁: 識別的系列変換を用いた日本語助詞誤りの訂正, 言語処理学会第 18 回年次大会, pp. 18–21 (2012).
- [5] Mizumoto, T., Komachi, M., Nagata, M. and Matsumoto, Y.: Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners, *Proceedings of 5th International Joint Conference on Natural Language Processing*, pp. 147–155 (2011).
- [6] Felice, R. D. and Pulman, S. G.: A Classifier-Based Approach to Preposition and Determiner Error Correction in L2 English, *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pp. 169–176 (2008).
- [7] Boyd, A., Zepf, M. and Meurers, D.: Informing Determiner and Preposition Error Correction with Hierarchical Word Clustering, *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pp. 208–215 (2012).
- [8] 石岡恒憲: 小論文自動採点, 電子情報通信学会誌, Vol. 92, No. 12, pp. 1036–1040 (2009).
- [9] 藤田 彬, 藤田 央, 田村直良: 多様な教育的観点を考慮した機械学習による日本語文章の評価と評価モデルの顕在化, 情報処理学会研究報告 NL-202 (2011).
- [10] 三浦 智, 村田真樹, 徳久雅人: 機械学習による「が」と「は」の使い分け, 言語処理学会第 18 回年次大会, pp. 1118–1121 (2012).
- [11] 南不二男: 現代日本語の構造, 大修館書店 (1974).
- [12] 竹内孔一, 乾健太郎, 竹内奈央, 藤田 篤: 意味の包含関係に基づく動詞項構造の細分類, 言語処理学会第 14 回年次大会 (2008).
- [13] 丸山岳彦, 柏岡秀樹, 熊野 正, 田中英輝: 日本語節境

界検出プログラム CBAP の開発と評価, 自然言語処理,
Vol. 11, No. 3, pp. 39-68 (2004).

- [14] 益岡隆志, 田窪行則: 基礎日本語文法—改訂版—, くろしお出版 (1992).
- [15] 飯田 龍, 徳永健伸: 日本語書き言葉を対象とした参照表現の自動省略-人間と機械処理の省略傾向の比較-, 情報処理学会研究報告 NL-206 (2012).