概要

一般的に，日本語学習者にとって，助詞の使い分けは難しいとされている．日本語学習者への支援として，助詞の自動推定や，助詞の使い分けのルールがあげられる．助詞の誤り訂正や，誤りの特徴の分析といった研究がなされている．しかし，助詞の推定の研究において,「に・へ」,「に・で」,「に・を」等の二分類を行っている研究はない．また，明確な使い分けのルールを獲得する研究は行われていない．そこで本研究では，日本語学習者の支援を行うため，使い分けが困難な助詞 (「は・が」,「に・へ」,「に・で」,「に・を」) の自動推定を行う．これにより，日本語学習者が助詞の使い分けに迷う場合，どちらを使うべきかを示すシステムを構築可能になる．また，素性の頻度分析，素性の取捨に基づく分析を行うことにより，日本語学習者にとって有用な，使い分けのルール獲得する．推定の結果,「は・が」の分類における提案手法の正解率は 0.760 であり，先行研究の手法と比べて高い値となった.「に・で」,「に・を」の分類において，提案手法が比較手法よりも高い値となった.「に・へ」においては全て「に」でない以外の手法がほぼ同等の正解率であった．各助詞の特徴や使い分けルールを多数獲得した．

目 次

図 目 次

表 目 次

第1章　はじめに

日本語の文法を対象とした様々な研究が行われている [1],[2],[3],[4],[5],[6],[7]．一般的に，ノンネイティブの日本語学習者にとって，助詞の使い分けは難しいとされている．その中でも副助詞「は」と格助詞「が」の使い分けや，格助詞「に・へ・を・で」の使い分けは特に困難である．例えば，副助詞「は」と格助詞「が」の使い分けにおいて,「彼は学生だ」と「彼が学生だ」の二文は文法として誤りでなく，かつニュアンスも近い．田中ら [8] は,「は・が」の使い分けについて「は」は既知情報や説明文,「が」は未知情報や描写文を示すと述べているが，明確な分類法については述べていない．また，Komori[9] は，遷移確率を用いて「は・が」の使い分けの自動推定を行っているが，その他の助詞については推定を行っていない．そこで本研究では，日本語学習者の支援を行うため，使い分けが困難な助詞（「は・が」,「に・へ」,「に・で」,「に・を」）の自動推定を行う．これにより，日本語学習者が助詞の使い分けに迷う場合，どちらを使うべきかを示すシステムを構築可能になる．また，助詞に関わるデータの分析を行うことにより，日本語学習者にとって有用な使い分けのルールを獲得する．

まず，副助詞「は」および格助詞「が」を含む文を京大コーパス 3.0[10] から収集し，これらを教師データとして利用する．次に，同様に別途収集した文から副助詞「は」および格助詞「が」を取り除いた文を収集し，これらをテストデータとして利用する．収集した教師データから Support Vector Machine(以下 SVM) で分類器を学習し，取り除いた助詞の推定実験を行う．最後に，教師データを分析し副助詞，格助詞の使い分けのルールを獲得する．同様の実験を,「に・へ」,「に・を」,「に・で」に対しても行った．

なお，Komori[9] の手法を利用した推定手法を比較手法として用いた．

本研究の主張点は次の 3 つである．

1 機械学習（SVM）を用いて助詞「は・が」,「に・へ」,「に・を」,「に・で」の分類を初めて行った．

2 「は・が」「に・を」「に・で」の使い分けの問題において，比較した手法のなかで機械学習（SVM）が最も高い正解率であった．

3 実験データを用いた素性の分析によって，助詞の使い分けに役立つ表現を多数獲得した．

本論文の構成は以下の通りである．第2章ではこれまでの助詞に対する研究を説明し，分類を行う各助詞の特徴を説明する．第3章では本研究で利用するデータを紹介する．第4章では，助詞を自動推定する手法の説明を行う．第5章では，助詞の自動推定に対する手法ごと，ならびに分類ごとの評価を行う．第6章では，素性分析を行い，各助詞の特徴と使い分けに関するルールを獲得する．第7章では，各助詞の特徴を考察する．第8章ではまとめを行う．

第2章　関連研究

本章では，助詞の自動推定や，助詞の解析に関する研究を紹介する．2.1 節では，助詞の推定に関する先行研究を紹介する．2.2 節では，本研究の助詞の推定で利用する，分類語彙表の変換に関する研究を紹介する．2.3 節では，助詞の日本語学習者に対する助詞の使い分けの誤り等の分析に関する研究を紹介する．

2.1　機械翻訳のための日本語格助詞の予測

2.1.1　概要

鈴木ら [1],[2] は機械翻訳の精度向上のために，文構造と内容語から格助詞の推定を行っている．推定は格助詞同定と分類の二段階で行っており，格助詞同定では，各文節を格助詞が存在するかしないかの 2 分類を行う．分類では，文節に格助詞が存在するもののうち「が，を，の，に，から，と，で，へ，まで，より」に「は，には，からは，とは，では，へは，までは，よりは」を加えた 18 分類を行う．推定手法には，最大エントロピー法を利用している．正解率は，格助詞同定のクローズドテストにおいて 0.984，オープンテストにおいて 0.960 であった．また，分類のクローズドテストにおいて 0.886，オープンテストにおいて 0.724 の正解率であった．鈴木ら [1],[2] と本研究との差異は，機械学習の手法の違いと，機械学習の分類数が違うことが挙げられる．本研究では，素性が増加することが予測されたので，過学習に強い SVM を利用した．また，分類数を 18 分類から 2 分類 (「は・が」「に・へ」「に・を」「に・で」の 2 分類ごと）に減らすことによって，より細かな使い分けの分析を行った．

2.1.2　取立て助詞「は」に関して

鈴木らは，本文中の情報のみからは，予測不可能な「しか・も」のような取り立て助詞（「だけ，ばかり，こそ」など）について述べている.「は」に関しても取立て助詞に分類されることもあり，推定が困難であると述べつつも，次の理由から，推定を行っている．

理由１　「は」の主要な働きは主題を提示することであり，この点，ほかの取り立て助詞と異なり,「提題助詞」と分類されることもある．主題は文の構造上からある程度予測可能である．

理由２　「は」は，日本語の助詞の中で「の」「を」についで頻度が高く，したがって「は」を適切に生成することは日本語文生成にとって重要である．

本研究では，理由 1 から推定が可能であると考え，素性として係り受け解析の情報を利用することで，文の構造を手がかりとして利用している．また，理由 2 から,「は・が」の分析の重要性がわかる．

2.2　分類語彙表の扱い

村田ら [11] は，単語を意味でソートする意味ソートについて述べている．本研究ではその中で行われている，分類語彙表の分類番号の変更を利用することによって，一部の素性を付与した．

2.2.1　分類語彙表

分類語彙表 [21] とはボトムアップ的に単語を意味に基づいて整理した表であり，各単語に対して分類番号という数字が付与されている．電子化された分類語彙表には 10 桁の分類番号が与えられており，7 レベルの階層構造を示している．上位 5 レベルは分類番号の最初の 5 桁，6 レベル目は次の 2 桁，最下層は最後の 3 桁で表現されている．

2.2.2　分類語彙表の分類番号変換表

分類語彙表の分類番号変換表を表 2.1 に示す．表の数字は分類番号の最初の何桁かを変換するものであり，分類番号の分類番号を変換後の分類番号に変換する．表の [] は正規表現で表されており，例えば [1-3] は 1 , 2 , 3 を表す．これにより分類語彙表そのままを付与するよりも粗いレベルでの情報が得られる．例えば，人間が主部に出現した際の傾向などを知ることができる．

表 2.1: 分類番号変換表

意味素性	分類語彙表の分類番号	変換後の分類番号
動物	[1-3]56	511
人間	12[0-4]	52[0-4]
組織・機関	[1-3]2[5-8]	53[5-8]
生産物・道具	[1-3]4[0-9]	61[0-9]
動物の部分	[1-3]57	621
植物	[1-3]55	631
自然物	[1-3]52	641
空間・方角	[1-3]17	657
数量	[1-3]19	711
時間	[1-3]16	811
現象名詞	[1-3]5[01]	91[12]
抽象名詞	[1-3]1[0-58]	aa[0-58]
人間活動	[1-3]58,[1-3]3[0-8]	ab[0-9]
その他	4	d

2.3 先行研究からの知見ー対象の助詞に関してー

助詞の使い分けに関する様々な先行研究 [12],[13],[14],[15],[16] から得られる対象の助詞についての知見を示す.

2.3.1 「は」と「が」の分類について

三上 [12] は「は・が」について,「は」は「が・に・の・を」を代行すると述べている. 例えば

象は鼻が長い

において象をベースにした文にすると

象は鼻が長い

となり，この文のベースを鼻にすると

象の長い鼻

となり，新たに助詞「の」が出現する．このことから,「は」は「の」の代行していると言える.「は」は代行する助詞の中でも,「が」を代行することが多いとしている．しかし,「は」と「が」はいっしょくたにすべきでないと述べており，使い分けの重要性が高いと言える．

2.3.2 「に」と「で」の分類について

安田ら [13] は，中国の大学で学ぶ日本語学習者とタイの大学で学ぶ日本語学習者に対し助詞を選択させるアンケートを行った．この結果から,「に・で」の選択において,「に」を過剰使用している傾向があることがわかった．このことから,「に」の適切な使用を促せるシステムが必要であると言える．若生ら [14] は，韓国人日本語学習者を対象にアンケートを行い，その結果を分析している．その中で,「で・に」の意味について次のように述べている．

で　場所・材料・手段・道具・原因・理由・範囲・まとまり・内容・動作主

に　対象・能力・知覚の主体・存在場所・到着点・受け手・変化結果・移動の方向・出所・時間・割合の分母

2.3.3 「に」と「へ」の分類について

杉村ら [15] は，日本語能力検定試験一級以上の日本語能力を身につけている日本語学習者を対象に助詞（「に・へ」）を選択させるアンケートを行い結果を分析した．分析の結果，上級・超上級日本語学習者はかなりの程度「に」と「へ」の使い分けができているが，日本語母語話者ほどにはその構文パターンの違いや，意味の違いを習得していないことがわかった．また,「に・へ」の使い分けについては次のように述べている．

に　移動や変化の結果を表す傾向がある.（事態の収束）

「A が B に」構文あるいは「A を B に」構文をとりやすい

へ　移動や変化の過程を表す傾向がある.（事態の進化）

「A から B へ」構文あるいは「A を B へ」構文をとりやすい

2.3.4 「に」と「で」と「を」の分類について

蓮池ら [16] は，日本語能力上級レベルの韓国語母語話者（「ＡＫ」），日本語能力上級レベルの中国語母語話者（「ＡＣ」），日本語能力中級レベルの韓国語母語話者（「ＩＫ」），日本語能力中級レベルの中国語母語話者（「ＩＣ」）を対象に助詞（「に・で・を」）を選択させるアンケートを行い結果を分析した．分析の結果を次に示す．

1　場所を示す格助詞のうち「に」と「で」の混乱が顕著であった．しかし，韓国語母語話者では「で」の正答数が多く，中国語母語話者では「に」の正答が多いという違いがあった．

2　韓国語母語話者は，母語からの類推により助詞を選択する傾向があった．特に中級レベルの学習者にこの傾向が顕著である．

3　中国語母語話者には「に」を多用する傾向があり，中級レベルの中国語母語話者には「に」の過剰一般化（自分の中で作られたルールを全て適用してしまうこと．例えば日本語学習者が keep の過去形を keeped のように書いてしまうこと）の現象がみられた．

4　中国語母語話者には，近隣の語（動詞，名詞）をヒントに助詞を選択する傾向が強かった．特に中級レベルの中国語母語話者は,「ある」など特定の動詞を，その意味に関わらずキーワードとして助詞選択に利用する傾向があった．

第3章　実験データ

本章では，実験に用いるデータについて説明を行う．

3.1　京大コーパス

本研究の教師データ，テストデータは京大コーパス [10] から獲得する．京大コーパスとは，新聞記事を自動解析後，人手による修正を加え， 各種言語情報を付与した品詞タグ付きコーパスである． 京大コーパスには，あらかじめ構文解析が行われており，係り先が付与されている．図 3.1 に京大コーパスを示す．ここで#の付いている行は，先頭の行を示しており，京大コーパスの文番号など が付与されている．EOS は文の終わりを表している．＊の付いている行は，左から文節番号，数字部分が係り先の 文節番号，英数字 D，P，A が係り受け関係，並列関係，同格関係を示している．係り受けについて，D の左の番号は係り先の文節の番号を表しており，例では,「ロシア側は」の文節は，入力文の中で 0 番目に出現する文節で，3D は 3 番目の分節に係ることを表している．その他の行は，形態素情報を表しており，左から，表記，読み，原型 (活用しない語の場合は＊)， 品詞，品詞細分類，活用型，活用形を示している．

3.2　教師の獲得

京大コーパスの 1995 年 1 月 1 日～1995 年 1 月 9 日（休刊日のため 1995 年 1 月 2 日を除く）から教師データを，1995 年 1 月 10 日～1995 年 1 月 17 日のデータからテストデータを生成する．1994 年全日を利用するデータは，1994 年のデータに形態素解析システム JUMAN[17], 構文解析システム KNP[18] を利用し教師を獲得する．まず，対象の助詞が最低 1 つは出現する文を抽出する．次に，対象の助詞を取り除く．対象の助詞を取り除いた文に対して，取り除いた助詞の種類を分類先として与える．文中に対象の助詞が複数存在する文の場合，対象の助詞の出現数分の教師データを獲得する．例えば,「今は鳥

```
# S-ID:950101004-003 KNP:96/10/27 MOD: MEMO:?
0 3D
ロシア ろしあ * 名詞 地名 * *
側 がわ * 接尾辞 名詞性名詞接尾辞 * *
は は * 助詞 副助詞 * *
1 2D
首都 しゅと * 名詞 普通名詞 * *
制圧 せいあつ * 名詞 サ変名詞 * *
の の * 助詞 接続助詞 * *
2 3D
最終 さいしゅう * 名詞 普通名詞 * *
段階 だんかい * 名詞 普通名詞 * *
に に * 助詞 格助詞 * * * 3 4D
入った はいった 入る 動詞 * 子音動詞ラ行 タ形
と と * 助詞 格助詞 * *
4 -1D
み み みる 動詞 * 母音動詞 未然形
られる られる られる 接尾辞 動詞性接尾辞 母音動詞 基本形
。 。 * 特殊 句点 * *
EOS
```

図 3.1: 京大コーパス例

取が熱い」の文からは次のような教師データを獲得する．X は取り除いた「は・が」の位置を表す．

副助詞は　今 X 鳥取が熱い

格助詞が　今は鳥取 X 熱い

の 2 つの教師データを獲得する．教師データの素性の情報は京大コーパスの形態素・構文情報から得た．獲得した教師データ数を表 3.1 に示す．

「に・へ」に関しては「へ」の教師数が少ないため，1994 年の毎日新聞の記事一年分の各分類ごとの教師データ数を揃えたデータ（に：3,339 文，へ：3,339 文）も教師として利用する．教師データ数を表 3.2 に示す．

表 3.1: データ数

助詞	教師データ数	テストデータ数
は	4323	5558
が	4653	6009
に	5529	7045
で	2238	3071
を	6432	8329
へ	85	85

表 3.2: 教師バランスを調整した「に・へ」のデータ数

助詞	教師データ数
に	3,339
へ	3,339

第4章　手法

本章では，提案手法，先行研究手法の説明を行う．

4.1　提案手法

本研究では，日本語学習者が「は」と「が」,「に」と「へ」,「に」と「を」,「に」と「で」の使い分けに迷った場合を想定し，それらの助詞を 1 つ空白にした文を問題とする．その問題に対し，機械学習を利用し空白に入れるべき助詞を推定する．

機械学習には，認識性能が優れている SVM を実装している TinySVM[19] を使用する．カーネル関数には 1 次の多項式カーネルを利用した．

4.1.1　Support Vector Machine

本研究の提案手法で用いるサポートベクトルマシン法は，空間を超平面で分割することにより 2 つの分類からなるデータを分類する手法である．このとき，2 つの分類が正例と負例からなるものとすると，学習データにおける正例と負例のマージン（間隔）を大きくとるほど分類器の誤りが減少するという考えから，このマージンを最大にする超平面を求めそれを用いて分類を行う．一般的に上記の方法の他に,「ソフトマージン」と呼ばれる学習データにおいてマージンの内部領域に少数の事例が含まれてもよいとする手法の拡張や，線形分離が不可能な問題に対応するために，超平面の線形の部分を非線型にする拡張 (カーネル関数の導入) がなされたものが用いられる．この拡張された方法は，以下の識別関数を用いて分類することと等価であり，その識別関数の出力値が正か負かによって二つの分類を判別することが可能である．

$$
\begin{aligned}
f(\mathbf{x}) &= sgn\left(\sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \\
b &= -\frac{max_{i,y_i=-1} b_i + min_{i,y_i=1} b_i}{2}
\end{aligned}
\tag{4.1}
$$

$$b_i = \sum_{j=1}^{l} \alpha_j y_j K(\mathbf{x}_j, \mathbf{x}_i)$$

ただし，$\mathbf{x}$ は識別したい事例の文脈 (素性の集合) を，$\mathbf{x}_i$ と $y_i (i = 1, ..., l, y_i \in \{1, -1\})$ は学習データの文脈と分類先を意味し，関数 sgn は，

$$sgn(x) = \begin{array}{ll} 1 & (x \geq 0) \\ -1 & (otherwise) \end{array} \tag{4.2}$$

であり，また，各 α_i は式 (4.4) と式 (4.5) の制約のもと式 (4.3) の $L(\alpha)$ を最大にする場合のものである．

$$L(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(\mathbf{x_i}, \mathbf{x_j}) \tag{4.3}$$

$$0 \leq \alpha_i \leq C \ (i = 1, ..., l) \tag{4.4}$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{4.5}$$

また，関数 K はカーネル関数と呼ばれ，様々なものが用いられるが本論文では以下の多項式のものを用いる．

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^d \tag{4.6}$$

C, d は実験的に設定される定数である．本論文ではすべての実験を通して C を 1 に d を 1 に固定した．ここで，$\alpha_i > 0$ となる $\mathbf{x}_i$ は，サポートベクトルと呼ばれ，通常，式 (4.1) の和をとっている部分はこの事例のみを用いて計算される．つまり，実際の解析には学習データのうちサポートベクトルと呼ばれる事例のみしか用いられない．

4.1.2 SVM 利用素性

村田ら [11] は，機械学習による助詞の推定を行っている．本研究では機械学習で利用する素性は村田ら [11] の研究を参考にして以下のものを用いる．分類語彙表 [21] を利用

する素性は，村田ら [20] の手法を利用し素性化する．素性 [24〜27] は，本研究で追加した素性である．素性 [24,25] は Komori[9] らの研究を参考にし追加した．素性 [26,27] は一般的に助詞の分析で利用される情報を参考に追加した．N(の文節に相当) は推定すべき助詞を含む文節を表し，V(述部に相当) は N の係り先の文節を表す．

例　私（N）【「が・は」等の対象の助詞】社長（V）だ

１　V における自立語の連続

２　V の最初の自立語の基本形

３　２の単語の品詞

４　２の単語の分類語彙表 [21] の分類番号の 1,2,3,4,5,6,7 桁までの数字．ただし，分類番号に対して文献 [20] の表の変更を行っている．

５　V に出現する付属語

６　N における自立語の連続

７　N の最後の自立語

８　７の単語の品詞

９　７の単語の分類語彙表 [21] の分類番号の 1,2,3,4,5,6,7 桁までの数字．ただし，分類番号に対して文献 [20] の表の変更を行っている．

１０　N に体言が存在するか否か

１１　同一文に共起する語

１２　１１の単語の分類語彙表 [21] の分類番号の 1,2,3,4,5,6,7 桁までの数字．ただし，分類番号に対して文献 [20] の表の変更を行っている．

１３　V の係り先の自立語の連続

１４　V の係り先の文節における最後の自立語の基本形

１５　１４の品詞

１６　１４の単語の分類語彙表 [21] の分類番号の 1,2,3,4,5,6,7 桁までの数字．ただし，分類番号に対して文献 [20] の表の変更を行っている．

１７　V の係り先に体言が存在するか否か

１８　V にかかる N 以外の体言の文節の自立語の連続

１９　V にかかる N 以外の体言の文節の最後の自立語

２０　１９の単語の品詞

２１　１９の単語の分類語彙表 [21] の分類番号の 1,2,3,4,5,6,7 桁までの数字．ただし，分類番号に対して文献 [20] の表の変更を行っている．

２２　V にかかる N 以外の体言が存在するか否か

２３　V にかかる N 以外の体言がとっている格

２４　解析対象の助詞の直前，直後の単語

２５　解析対象の助詞の直前，直後の単語の品詞

２６　解析対象の文内において，解析対象の文節以外にある助詞

２７　解析対象の文節内の名詞がすべて，記事内の前方に存在しているか否か

次に入力文に対して，実際に付与される素性を大まかに示す．

入力　彼【は or が】山に出かけたことがあると兄が言った．

素性１～５が対象とする文節　出かけたことがあると

付与素性例　出かけることある　（述部Ｖの自立語連続）

出かける　（述部Ｖの自立語の原形）

動詞　（述部Ｖの自立語の品詞）

こと，が，ある，と（出現する付属語）

他に分類語彙表 [21] の番号の情報

素性６～１０が対象とする文節　彼

付与素性例　彼　（Ｎの自立語の連続）

彼 （Ｎの最後の自立語）

名詞（Ｎの最後の自立語の品詞）

他に分類語彙表 [21]，体言の有無の情報

素性１１，１２が対象とする文節　全文

付与素性例　動詞：出かける 動詞：言う 動詞：ある

名詞：山 名詞：兄

助詞：が 助詞：と　助詞：に

名詞：こと 特殊：。

素性１３～１７が対象とする文節　言った

付与素性例　言った（V の係り先の自立語の連続）

言う（V の係り先の最後の自立語の基本形）

動詞（V の係り先の最後の自立語の品詞）

他に分類語彙表 [21] の情報，係り先の存在しているかの情報

素性１８～２３が対象とする文節　山に

付与素性例　格に：山 (V に係る N 以外の助詞「に」の文節の自立語の連続)

格に存在 (V に係る N 以外の助詞「に」が存在している)

格に：山 (V に係る N 以外の助詞「に」の文節の最後の自立語)

格に：名詞 (V に係る N 以外の助詞「に」の最後の自立語の品詞)

他に分類語彙表 [21] の情報

素性その他が対象とする文節　全文，同一記事既出文

付与素性例　推定する箇所の前後の単語　既出単語

4.2　先行研究手法

提案手法と比較する Komori[9] の先行研究手法を説明する．

4.2.1　先行研究手法：１

助詞「は・が・に・へ・で・を」（以下対象の助詞）の直前に出現する名詞の統計情報を利用し，X に入る助詞はどちらであるかを推定する．具体的には，現在解いている箇所の直前の単語を A とするとき，学習データにおいて A の後により高い頻度で出現する助詞を求め、それを現在解いている箇所の助詞とする．例えば,「は・が」の分類において，X の直前に「今」という名詞が出現しており，学習データにおいて「今」が出現した場合に副助詞「は」を使う確率が格助詞「が」を使う確率よりも大きいのであれば，X に入る助詞は「は」であると推定する．

4.2.2 先行研究手法：2

助詞「は・が・に・へ・で・を」（以下対象の助詞）の直前直後に出現する名詞の統計情報を利用し，X に入る助詞はどちらであるかを推定する．具体的には，現在解いている箇所の直前の単語の品詞を A，直後の単語の品詞を B とするとき，学習データにおいて A の後に出現する助詞の確率を求め，学習データにおいて B の前に出現する助詞の確率を求め助詞ごとに二つの値を合計する．その確率の合計が高いものを現在解いている箇所の助詞とする．

4.2.3 先行研究手法：3

対象の助詞の直前，直後に出現する品詞を利用し，X に入る助詞はどちらであるかを推定する．具体的には，現在解いている箇所の直前の単語の品詞を A とし，直後の単語の品詞を B とするとき，学習データにおいて A と B に挟まれる個所により高い頻度で出現する助詞を求め、それを現在解いている箇所の助詞とする．例えば,「は・が」の分類において，X の直前に名詞が出現しており，X の直後に動詞が出現している状況で，学習データにおいて直前に名詞が出現し直後に動詞が出現する場合に「は」になる確率が「が」になる確率よりも大きいのであれば，X に入る助詞は「は」であると推定する．

第5章　実験

本章では，助詞の推定実験の評価方法と結果を載せる．実験の前に「は・が」,「に・へ」,「に・で」,「に・を」に使い分けが必要な例がどの程度存在するか調査を行った．

5.1　実験前調査：各助詞は使い分けが必要か否か

助詞の使い分けにおいて，助詞の使い分けが必要でない場合がある．本節では，推定結果を分析することによって，使い分けが必要でない場合の比率を調査する.「は・が」,「に・へ」,「に・で」,「に・を」のテストデータからランダムで文章を抜き出し，各文章に対し使い分けが必要か否かを人手で調査する．この調査により，新聞に出現する文章に，どの程度使い分けが必要な助詞が存在しているかを知ることができる．使い分けが必要な文が多いほど，本研究の手法の有用性が高いといえる．

5.1.1　「は・が」使い分けが必要な文の比率の調査

「は・が」のテストデータからランダムで 10 件抜き出し，新聞の文で使い分けが必要な「は・が」がどの程度存在しているのかを調査した．表 5.1 に結果を示す．使い分けが必要な文は，10 文中 9 文であった．このことから，ほとんどの「は・が」を含む文章において使い分けが必要であることがわかった．次にランダムで抽出した 10 文とその分析結果を載せる．

使い分け必要　締まりのない社会党騒動をみていると、だんだん腹【が】立ってくる。

使い分け必要　山花氏はとりあえず、社会党会派からの離脱届を中執委に提出するが、「離脱【が】了承されなければ、同日中に離党届提出だ」との強硬論もあり、山花氏は背水の陣を敷いた形だ。

使い分け必要　久保ら【は】当時、そのことをまったく知らない。

使い分け必要　作業速度【は】三キロで、前後走行が可能。

使い分け必要　「もともと社会党自体が第三極の考え方なのだから、何も党を割らなくてもいいのではないか。次の総選挙で新進党と手を握りたい人【は】、第三極などと言わずに正直にそう言えばいい」

使い分け必要　王国崩壊の危機に立つ西武だが、むしろ選手【は】燃えるのではないか。

使い分け必要　５カ月間の禁漁効果もあるが、秋に台風【が】なかったのと、高水温が続いたせいらしい。

使い分け必要　今回の交渉が決着の運びとなった背景には、公的年金市場についての米国の強い市場開放圧力に加えて、規制緩和が国内経済政策の焦点となるなか、厚生省など国内からも開放要求が出たこと【が】挙げられる。

使い分け必要　同行はこれまで日本の南ア向け輸出企業に融資した前例はあるが、南アへの直接融資【は】これが第一号となる。

使い分け不要　村山富市首相と武村正義さきがけ代表の十六日の会談で、さきがけ【が】政策研究会設置提案という形で両党を軸にした再編に向けた協議にゴーサインを出した。

表 5.1: 「は・が」の使い分けが必要か否か

使い分け	必要	不要
文数	9	1

5.1.2　「に・へ」使い分けが必要な文の比率の調査

「に・へ」のテストデータからランダムで 10 件抜き出し，新聞の文で使い分けが必要な「に・へ」がどの程度存在しているのかを調査した．表 5.2 に結果を示す．使い分けが必要な文は，10 文中 9 文であった．このことから，ほとんどの「に・へ」を含む文章において使い分けが必要であることがわかった．次にランダムで抽出した 10 文とその分析結果を載せる．

使い分け必要　われわれ国民【に】分かろうはずもない。

使い分け必要　自社政権論ともう一つ、第三極論について、次の二人のリーダーの分かりやすい発言【に】、山花らは明快な回答を用意したほうがいい。

使い分け必要　予算決定後【に】、九社にその年度の工事内容や予算を詳細に伝え、メーカー側はこの直後に、「ドラフト会議」を開催、シェア枠内で工事を割り振っていた。

使い分け必要　いつまでもラモス、三浦【に】頼るのでは、次期ワールドカップ出場は難しい。

使い分け必要　連勝したことで期待が予想以上【に】膨らみ、プレッシャーも感じていたに違いない。

使い分け必要　同市職員らでつくる「闘牛クラブ」の国下和男会長は「やはり近鉄に残って優勝に貢献してもらうのが希望。今の投手陣を見ても野茂投手【に】頑張ってもらわないと、優勝への道は険しい」と話していた。

使い分け必要　上田署員が駆けつけたところ、桜井さんの庭先【に】止めてあった軽乗用車が燃えており、車内から大人一人、子供二人の焼死体を発見した。

使い分け必要　同郵便局は午前九時【に】業務を始めたばかりで、客はおらず、局員四人だけだった。

使い分け必要　どのアナウンサーも「さんない」の「さん」【に】アクセントを置いているが、津軽弁では四文字を平ったく発音する。

使い分け不要　そこ【へ】ベルルスコーニ氏が「クリーン政治」「イタリアの回生」を掲げて実業界から政界入りを表明。

表 5.2: 「に・へ」の使い分けが必要か否か

使い分け	必要	不要
文数	9	1

5.1.3 「に・で」使い分けが必要な文の比率の調査

「に・で」のテストデータからランダムで 10 件抜き出し，新聞の文で使い分けが必要な「に・で」がどの程度存在しているのかを調査した．表 5.3 に結果を示す．使い分けが必要な文は，10 文中 9 文であった．このことから，ほとんどの「に・で」を含む文章において使い分けが必要であることがわかった．次にランダムで抽出した 10 文とその分析結果を載せる．

使い分け必要　Ｇ７各国から、米国最大手の電気通信事業者のＡＴ＆Ｔなど、マルチメディア企業のトップ約四十人が出席すること【に】なっている。

使い分け必要　六五一年、ササン朝が新興のイスラム軍に滅ぼされたとき、大勢の王族・貴族がシルクロードを経て唐の長安【に】亡命した。

使い分け必要　先日、テレビ【で】輸入農産物の安全性について放映していました。

使い分け必要　「ＮＡＦＴＡが機能すれば、メキシコ国内に十分な雇用の場が創出され、不法移民の流出にブレーキをかけることができる」と言うセディジョ大統領【に】とって、その長期的ビジョンをくじくようなカリフォルニア州民の審判は「両国関係を緊張させ、経済を後退させかねない不幸な問題」と映ったようだ。

使い分け必要　チェチェン情勢【に】関するＯＩＣの公式声明は初めて。

使い分け必要　政府筋は九日、ブリュッセルで二月二十五、二十六日に開かれる先進七カ国情報通信担当閣僚【に】よる「情報サミット」に合わせ、二月二十五日にＧ７各国の主要電気通信事業者や電機メーカー、マルチメディア関連などの企業トップも「民間情報サミット」を開くと明らかにした。

使い分け必要　国連仲介による外相会談は五回目【で】、東ティモールの各政治勢力との対話など平和解決の道を探る。

使い分け必要　また世界貿易機関に引き継がれる関税貿易一般協定二四条に従えば、農業を含むすべての分野【に】ついて自由化を進めるか、域外国にも最恵国待遇で自由化の成果を適用しなければならない。

使い分け必要　十日【に】大阪府枚方市で開かれる松下グループの今年の経営方針発表会に出席するためだが、十一、十二日には松下幹部との最高経営会議も予定されている。

使い分け不要　通貨切り下げ【に】端を発したメキシコ経済の混乱が広がる中、昨年まで政権の座にあったサリナス前大統領に対する非難が高まっている。

表 5.3: 「に・で」の使い分けが必要か否か

使い分け	必要	不要
文数	9	1

5.1.4 「に・を」使い分けが必要な文の比率の調査

「に・を」のテストデータからランダムで 10 件抜き出し，新聞の文で使い分けが必要な「に・を」がどの程度存在しているのかを調査した．表 5.4 に結果を示す．使い分けが必要な文は，10 文中 10 文であった．このことから，ほとんどの「に・を」を含む文章において使い分けが必要であることがわかった．次にランダムで抽出した 10 文とその分析結果を載せる．

使い分け必要　首相は社会党の党内情勢【を】説明したうえで、さきがけの協力を要請した。

使い分け必要　連合滋賀は参院の議席を死守するため、高田氏　【に】「連合公認候補なら支援する」との条件で十月末に擁立を決めた。

使い分け必要　近畿の各連合は近く政治団体「リベラル近畿」【を】結成し、それぞれ支援してきた国会議員の受け皿づくりとして連合新党の結党を模索している。

使い分け必要　一方、共産党県委員会の国政対策委員長、川内卓氏【を】擁立する同委員会は「なれ合い政治の典型」と批判する。

使い分け必要　子供が父親から認知された途端に児童扶養手当【を】打ち切るよう定めた児童扶養手当法施行令は、非嫡出子に対する差別だとして、関西の「婚外子差別と闘う会」など三団体が十三日、厚生省に施行令の改正を求める要望をした。

使い分け必要　「小笠原料理といえるのは島ずしぐらい。カンパチやオナガダイ、アカバなど小笠原近海のもの【を】使い、八丈の料理を組み合わせました」と宮沢さん。

使い分け必要　こうした追い風【を】受け香港は九五、九六年も実質五％台の安定した経済成長が見込まれている。

使い分け必要　香港がこれほど意欲的な姿勢【を】みせる背景には、中国との経済統合が加速しているという現実がある。

使い分け必要　受益者とは車に乗って高速道路を走る人のこと【を】言うらしい。

使い分け必要　ニュースで、アナウンサーが「三陸はるか沖地震」と言う度に、美しく詩的に耳【に】響いて、二人が亡くなり、いまだに重傷を負って入院治療をしている多くの人、家屋などの被害に遭った人々の気持ちを思うと、何かやりきれない気がする。

表 5.4: 「に・を」の使い分けが必要か否か

使い分け	必要	不要
文数	10	0

5.2 評価値の計算

評価は，正解率と F 値で行う．正解率は (5.1) の式を用いて算出する．

$$\text{正解率} = \frac{\text{正解数}}{\text{出力}} \tag{5.1}$$

F 値とは適合率と再現率の調和平均であり，(5.2) の式を用いて算出する．適合率はシステム出力の正解率，再現率は問題に対する取りこぼしの指標であり，適合率は (5.3)，再現率は (5.4) の式で表される．

$$F\text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \tag{5.2}$$

$$\text{適合率} = \frac{\text{システムの正解数}}{\text{システムの出力数}} \tag{5.3}$$

$$\text{再現率} = \frac{\text{システムの正解数}}{\text{テストデータ中の正解数}} \tag{5.4}$$

5.3 「は・が」推定実験

副助詞「は」，格助詞「が」が取り除かれた文に対し，SVM を利用し取り除かれた助詞の推定を行った．提案手法の正解率の他に，全てを「が」に分類する手法，全てを「は」に分類する手法，先行研究手法の正解率も求めた．これらの正解率を表 5.5 に示す．表のように，SVM の正解率は 0.760 であり，比較手法の中で最も高い値となった．また，SVM の「が」の推定は F 値 0.768(再現率：0.765，適合率：0.772),「は」の推定は F 値 0.751(再現率：0.755，適合率：0.748) であった．

表 5.5: 「は・が」での正解率

手法	正解率
SVM	0.760
先行研究手法 1	0.615
先行研究手法 2	0.522
先行研究手法 3	0.646
全て「が」に分類	0.519
全て「は」に分類	0.480

表 5.6: 「は・が」での F 値・再現率・適合率

手法	分類先	F 値	再現率	適合率
SVM	が	0.768	0.765(4598/6009)	0.772(4598/5956)
	は	0.751	0.755(4200/5558)	0.748(4200/5611)
先行研究手法 1	が	0.672	0.761(4576/6009)	0.602(4576/7590)
	は	0.532	0.457(2544/5558)	0.639(2544/3977)
先行研究手法 2	が	0.661	0.900(5412/6009)	0.523(5412/10739)
	は	0.185	0.113(631/5558)	0.513(631/1228)
先行研究手法 3	が	0.599	0.509(3061/6009)	0.727(3061/4206)
	は	0.683	0.793(4413/5558)	0.599(4413/7361)
全て「が」	が	0.683	1.000(6009/6009)	0.519(6009/11567)
全て「は」	は	0.648	1.000(5558/5558)	0.480(5558/11567)

5.4 「に・へ」,「に・を」,「に・で」推定実験

「に」と「へ」,「に」と「を」,「に」と「で」が取り除かれた文に対しても，同様に助詞の推定を行った．正解率を表 5.7 に示す．表のように,「に・で」「に・を」の分類において，SVM の手法が比較手法の中で最も高い値となった．また,「に・へ」においては全て「に」でない以外の全ての手法がほぼ同等の正解率であった．

表 5.7: 「に・へ」「に・を」「に・で」での正解率

手法	正解率		
	にへ	にで	にを
SVM	0.987	0.812	0.889
先行研究手法 1	0.985	0.736	0.617
先行研究手法 2	0.988	0.696	0.582
先行研究手法 3	0.988	0.700	0.585
全て「に」	0.988	0.696	0.458
全て「に」でない	0.011	0.303	0.541

表 5.8: 「に・で」での F 値・再現率・適合率

手法	分類先	F 値	再現率	適合率
SVM	に	0.868	0.894(6300/7045)	0.845(6300/7451)
	で	0.669	0.625(1920/3071)	0.720(1920/2665)
先行研究手法 1	に	0.827	0.915(6448/7045)	0.756(6448/8519)
	で	0.427	0.325(1000/3071)	0.626(1000/1597)
先行研究手法 2	に	0.820	1.000(7045/7045)	0.696(7045/10114)
	で	0.001	0.0006(2/3071)	1.000(2/2)
先行研究手法 3	に	0.821	0.990(6978/7045)	0.701(6978/9943)
	で	0.065	0.034(106/3071)	0.612(106/173)
全て「に」	に	0.820	1.000(7045/7045)	0.696(7045/10116)
全て「で」	で	0.465	1.000(3071/3071)	0.303(3071/10116)

表 5.9: 「に・を」での F 値・再現率・適合率

手法	分類先	F 値	再現率	適合率
SVM	に	0.876	0.858(6050/7045)	0.895(6050/6755)
	を	0.899	0.915(7624/8329)	0.884(7624/8619)
先行研究手法 1	に	0.635	0.729(5139/7045)	0.564(5139/9106)
	を	0.596	0.523(4362/8329)	0.695(4362/6268)
先行研究手法 2	に	0.267	0.166(1174/7045)	0.682(1174/1720)
	を	0.707	0.934(7783/8329)	0.570(7783/13654)
先行研究手法 3	に	0.281	0.177(1247/7045)	0.683(247/1824)
	で	0.708	0.930(7752/8329)	0.572(7752/13550)
全て「に」	に	0.628	1.000(7045/7045)	0.458(7045/15374)
全て「を」	を	0.702	1.000(8329/8329)	0.541(8329/15374)

5.5 「に・へ」教師の調整

「に・へ」に関しては「へ」の教師数が少ないため，1994 年の毎日新聞の記事一年分の各分類ごとの教師データ数を揃えたデータ（に：3,339 文，へ：3,339 文）を教師として利用し，各手法で推定を行う実験を追加で行った．正解率を表 5.10 に，SVM の教師の違いによる F 値の差を表 5.11 に示す．教師を増やすことによって，正解率は下がったが,「へ」の F 値は上昇した．

表 5.10: 94 年データを利用した「に・へ」での正解率

手法	正解率
SVM	0.862
先行研究手法 1	0.867
先行研究手法 2	0.180
先行研究手法 3	0.216
全て「に」	0.988
全て「に」でない	0.011

表 5.11: 教師の違いによる「に・へ」での F 値・再現率・適合率

手法	教師	分類先	F 値	再現率	適合率
SVM	95 年 1 月 1〜9 日	に	0.993	0.998	0.988
		へ	0.042	0.023	0.200
	94 年 全日	に	0.924	0.863	0.996
		へ	0.109	0.717	0.059
先行研究手法 1	95 年 1 月 1〜9 日	に	0.992	0.997	0.988
		へ	*	*	*
	94 年 全日	に	0.928	0.872	0.992
		へ	0.073	0.447	0.040
先行研究手法 2	95 年 1 月 1〜9 日	に	0.993	1.000	0.988
		へ	*	*	*
	94 年 全日	に	0.291	0.171	0.997
		へ	0.025	0.964	0.013
先行研究手法 3	95 年 1 月 1〜9 日	に	0.994	1.000	0.988
		へ	*	*	*
	94 年 全日	に	0.343	0.207	0.997
		へ	0.028	0.952	0.014
全て「に」	なし	に	0.993	1.000	0.988
		へ	*	*	*

第6章　分析

本章では，素性の取捨に基づく分析による素性の分析と頻度分析による素性の分析を行った．また,「は・が」の推定結果の分析を行った．

6.1　素性の取捨に基づく分析による効果的な素性の俯瞰

素性の取捨に基づく分析によって，SVM による推定実験において有用な素性にはどういったものがあるかを調査した．教師データは毎日新聞 1995 年 1 月 1 日～1995 年 1 月 9 日から獲得し，推定実験は交差検定によって行う．

6.1.1　交差検定

交差検定（cross-validation）は，教師データを n 分割し 1 個をテストデータ，残りの n-1 個を教師データとして学習を行い，分割された教師データが全て 1 回テストデータとして用いられるのように繰り返して精度を求める方法である．本実験では，手作業で作成された教師データの SVM に対して，10 分割の交差検定によって精度を求める．

6.1.2　分析：手順

教師データを利用し，10 分割交差検定による学習を行う．次に，教師データから，特定の素性を取り除き，10 分割交差検定による学習を行う．

二つの結果の正解率を比べることにより，有用な素性か否かを判断する．特定の素性を取り除いた教師データの交差検定の正解率が，取り除く前の教師の交差検定の正解率よりも低い場合は，その素性は有効であるといえる．また，特定の素性を取り除いた教師データの交差検定の正解率が，取り除く前の教師の交差検定の正解率よりも高いあるいは同等の場合は，その素性は有効でないといえる．取り除く素性群は，述部 V に関係する素性群「述部 V」（素性番号 1～5），体言の文節 N に関係する素性群「体言の文節 N」（素性番号 6～10），共起単語に関する素性群「共起」（素性番号 11,12），述部 V の係り先に関する素性群「係り先」（素性番号 13～17），述部 V に係る N 以外の体言の文節に関係する素性群「N 以外体言」（素性番号 18～23），品詞に関する素性群「品詞」（素性番号 3,8,15,20），分類語彙表に関する素性群「分類」（素性番号 4,9,12,16,21），推定す

る助詞の直前直後に出現する素性群「直」（素性番号 24,25），解析対象の文内において，解析対象の文節以外にある助詞を表す素性「他助詞」（素性番号 26），解析対象の文節内の名詞がすべて，記事内の前方に存在しているか否かを表す素性「文脈情報」（素性番号 27）であり，これらを増減させることで，素性の取捨に基づく分析を行う．

6.1.3 分析１：「は・が」使い分け

素性の取捨に基づく分析により「は・が」の分析を行った．表 6.1 に取り除いた素性ごとの正解率を示す．分析の結果，「は・が」の分類の際に素性群「述部 V，体言の文節 N，N 以外体言，品詞，直，文脈情報」が有効であるとわかった．最も効果の高い述部 V に関する例文を次に示す．

述部 V がなくても判断可能　党首選のしこり【が or は】[述部 V] とすれば、羽田氏人気は新進党にとってかえって波乱要因となるかもしれない。

述部 V ＝のこっている；【】＝が

述部 V がないと判断困難　穏健派【が or は】、統一地方選前を [述部 V]。

述部 V ＝主張する；【】＝は

述部 V がなくても判断可能な文は，文章の工夫で「は」をいれることも可能だが，「とすれば」が続く場合，一般的には「が」を使うことが多いだろうと判断できる．この問題に関しては，述部 V に付く付属語からも人は予測している可能性がある．判断困難な文は，使い分けを推定することは難しい．一方の文は助詞がある程度予測可能であるものの，他の情報（「とすれば」）から読み取る必要がある．どちらの例文も，述部 V の情報がないために判断が難しくなっていることから，述部 V が「は・が」の使い分けの判断に重要であることがうかがえる．

6.1.4 分析２：「に・へ」使い分け

素性の取捨に基づく分析により「に・へ」の分析を行った．表 6.2 に取り除いた素性ごとの正解率を示す．分析の結果，「に・へ」の分類の際に素性群「述部 V，体言の文節 N，共起，係り先 N，N 以外体言，品詞，直，文脈情報」が有効であるとわかった．「に・へ」の分類において突出して有用な素性は見つからなかった．次に，「に・へ」が使われる例文を示す．

分類：に　「ＮＡＴＯ【に】加わる四カ国と加わらない東欧諸国との間に新たなカーテンが下りる」と警告し、「ＮＡＴＯ拡大に戦略的・政治的な合理性はなく、ロシアが近隣諸国に侵略的態度を取り始めた場合に拡大すると警告するだけで十分」と反論した。

表 6.1: 素性の取捨に基づく分析：「は・が」

取り除いた素性	正解率
なし	0.7467
述部 V	0.7091
体言の文節 N	0.7359
共起	0.7610
係り先	0.7468
N 以外体言	0.7298
品詞	0.7392
分類	0.7521
直	0.7301
他助詞	0.7477
文脈情報	0.7430

分類：へ　さらに同党はここ【へ】きて「ロシアがＮＡＴＯの将来や欧州における米国の役割について拒否権を持つのは許せない」、「米国は冷戦の勝利者として行動すべきだ」などと強硬姿勢を鮮明にし始めた。

6.1.5　分析３：「に・で」使い分け

素性の取捨に基づく分析により「に・で」の分析を行った．表 6.3 に取り除いた素性ごとの正解率を示す．分析の結果,「に・で」の分類の際に素性群「述部 V，体言の文節 N，N 以外体言，品詞，分類，直，他助詞」が有効であるとわかった．最も効果の高い述部 V に関する例文を次に示す．

述部 V がなくても判断可能　イスラエルからの報道によると、原因は入植者【に or で】[述部 V] 居住区域の拡張工事。

述部 V ＝よる ;【】＝に

述部 V がないと判断困難　田弘子はひとり【に or で】[述部 V].

述部 V ＝歩いてきた ;【】＝で

述部 V がなくても判断可能な文は，一般的には「で」を使うべきではないと判断できる．この問題に関しては，体言の文節 N から人は予測している可能性がある．体言の文節 N も効果のある素性としてあげられているため，これらの情報を総合的に判断していると考えられる．判断困難な文は，使い分けを推定することは難しい．結果から，述部 V が「に・で」の使い分けの判断に重要であることがうかがえる．

表 6.2: 素性の取捨に基づく分析：「に・へ」

取り除いた素性	正解率
なし	0.9838
述部 V	0.9836
体言の文節 N	0.9834
共起	0.9786
係り先	0.9822
N 以外体言	0.9829
品詞	0.9836
分類	0.9843
直	0.9831
他助詞	0.9838
文脈情報	0.9836

表 6.3: 素性の取捨に基づく分析：「に・で」

取り除いた素性	正解率
なし	0.8066
述部 V	0.7801
体言の文節 N	0.7900
共起	0.8338
係り先	0.8076
N 以外体言	0.7818
品詞	0.8025
分類	0.8037
直	0.7824
他助詞	0.8064
文脈情報	0.8074

6.1.6　分析４：「に・を」使い分け

素性の取捨に基づく分析により「に・を」の分析を行った．表 6.4 に取り除いた素性ごとの正解率を示す．分析の結果,「に・を」の分類の際に素性群「述部 V，体言の文節 N，係り先 N，N 以外体言，品詞，分類，直，他助詞，文脈情報」が有効であるとわかった．最も効果の高い述部 V に関する例文を次に示す．

述部 V がなくても判断可能　散文の根っこ【に or を】[述部 V]、水分がたっぷりあった、ということだと思う。

述部 V ＝ほりかえしてみたら；【】＝を

述部 V がないと判断困難　一葉の作品をえらんだのは、文学【に or を】[述部 V] ことで、かえって自由な世界がくると信じたからであろう。

述部 V ＝しばられる；【】＝に

述部 V がなくても判断可能な文は，体言の文節 N から判断している可能性がある．判断困難な文は，受身かどうかわからないと使い分けを推定することは難しい．一方の文は助詞がある程度予測可能であるものの，どちらの例文からも，述部 V が「に・を」の使い分けの判断に重要であることがうかがえる．

表 6.4: 素性の取捨に基づく分析：「に・を」

取り除いた素性	正解率
なし	0.8835
述部 V	0.8226
体言の文節 N	0.8744
共起	0.9011
係り先	0.8799
N 以外体言	0.8695
品詞	0.8795
分類	0.8741
直	0.8722
他助詞	0.8834
文脈情報	0.8831

6.2 素性の頻度分析によるルールの獲得

どういった素性が出現すると「は」,「が」,「に」,「へ」,「で」,「を」が使われやすいのかを明らかにするために，素性の頻度分析を行った.「は・が」,「に・を」は本研究で用いた教師データを利用して分析を行う.「に・で」は教師データ数が偏っているため，データ数を揃えたデータ（に：2,238 文，で：2,238 文）を利用し分析を行う.「に・へ」においては「へ」の教師が少ないため，1994 年の毎日新聞の記事一年分のデータ数を揃えたデータ（に：3,339 文，へ：3,339 文）を利用する.

素性の出現頻度が 50 回以上であり，テストデータにおいてその素性が出現した場合にその分類先が出現する確率が 0.75 以上の素性を使い分けに有用な素性として獲得する.獲得された使い分けに有用な素性の数を表 6.5 に示す.

表 6.5: 有用な素性の数

分類問題	分類先	獲得ルール数
は・が	は	40
	が	49
に・へ	に	63
	へ	127
に・で	に	28
	で	34
に・を	に	74
	を	114

6.2.1 有用な素性の内訳

有用な素性の数の内訳を表 6.6，表 6.8，表 6.7，表 6.9 に示す．表の素性群は，述部 V に関係する素性群「述部 V」（素性番号 1〜5），体言の文節 N に関係する素性群「体言の文節 N」（素性番号 6〜10），共起単語に関する素性群「共起」（素性番号 11,12），述部 V の係り先に関する素性群「係り先」（素性番号 13〜17），述部 V に係る N 以外の体言の文節に関係する素性群「N 以外体言」（素性番号 18〜23），品詞に関する素性群「品詞」（素性番号 3,8,15,20），分類語彙表に関する素性群「分類」（素性番号 4,9,12,16,21），推定する助詞の直前直後に出現する素性群「直」（素性番号 24,25）である．出現数は獲得した有用な素性の数であり，比率はその分類において獲得された有用な素性の総数で出現数を割ったものである．なお，定義した各素性群は重複しているものもあるため，比率の合計は 1 にならず，出現数の合計は獲得した教師に必ずしも一致しない.

表 6.6: 獲得した「は・が」のルールの割合

分類先	素性群	比率	出現数
は	述部 V	0.112	10
が	述部 V	0.089	8
は	直	0.033	3
が	直	0.067	6
が	共起	0.022	2
が	係り先	0.303	27
は	体言の文節 N	0.067	6
が	体言の文節 N	0.033	3
は	N 以外体言	0.224	20
が	N 以外体言	0.033	3
は	品詞	0.056	5
が	品詞	0.022	2
は	分類	0.280	25
が	分類	0.348	31

表 6.7: 獲得した「に・へ」のルールの割合

分類先	素性群	比率	出現数
に	述部 V	0.142	27
へ	述部 V	0.205	39
に	直	0.036	7
へ	直	0.073	14
に	共起	0.047	9
へ	共起	0.084	16
に	体言の文節 N	0.105	20
へ	体言の文節 N	0.268	51
へ	N 以外体言	0.036	7
に	品詞	0.015	3
へ	品詞	0.015	3
に	分類	0.178	34
へ	分類	0.336	64

表 6.8: 獲得した「に・で」のルールの割合

分類先	素性群	比率	出現数
に	述部 V	0.338	21
で	述部 V	0.193	12
に	直	0.080	5
で	直	0.032	2
に	体言の文節 N	0.032	2
で	体言の文節 N	0.129	8
で	N 以外体言	0.193	12
で	品詞	0.016	1
に	分類	0.306	19
で	分類	0.387	24

表 6.9: 獲得した「に・を」のルールの割合

分類先	素性群	比率	出現数
に	述部 V	0.207	39
を	述部 V	0.319	60
に	直	0.047	9
を	直	0.047	9
に	共起	0.005	1
を	共起	0.021	4
に	体言の文節 N	0.047	9
を	体言の文節 N	0.069	13
に	N 以外体言	0.085	16
を	N 以外体言	0.148	28
に	品詞	0.005	1
を	品詞	0.005	1
に	分類	0.250	47
を	分類	0.430	81

6.2.2　分析１：「は・が」使い分け

分析の結果,「は」および「が」の使い分けに，述部Vに存在する判定詞「だ」が関係することが確認できた．この結果は我々の先行研究 [22] と同様の結果である．有効な素性の例を表 6.10 に示す．確率はテストデータにおいてその素性が出現した場合にその分類先が出現する確率であり，頻度はテストデータでのその素性の頻度である．以下に表 6.10 の素性を含む例文を示す.

述部Vに存在する判定詞「だ」　二番目は、この制度は大きな政党同士に政権交代を可能ならしめるもの だ。

述部Vのかかり先に体言が存在　八五年四月に「電電公社」が民営化された 際、電気通信事業法が成立。

また，この他にも，推定する助詞の直後の単語に「記号：「」が出現する場合，直後に「首相」が出現する場合，述部Vに出現する最初の自立語が「話す」の場合に「は」が使われやすく，推定する助詞の直後の単語に「あう・ある・できる・出る」が出現する場合に「が」が使われやすい等のルールも獲得できた.

表 6.10:「は・が」での素性分析

分類先	素性	確率	頻度
は	述部Vに判定詞「だ」が存在	0.754	809
が	述部Vの係り先に体言が存在	0.873	1255

6.2.3　分析２：「に・へ」での使い分け

分析の結果,「に」および「へ」の使い分けに，述部Vにおける最初の自立語が関係することがわかった．有効な素性の例を表 6.11 に示す．以下に表 6.11 の素性を含む例文を示す.

述部Vにおける最初の自立語「つく」　生命保険について 思うこと

述部Vにおける最初の自立語「行く」　東京・二子玉川園の「ナムコ・ワンダーエッグ」へ行って みた。

また，この他にも，推定する助詞の直前の単語に「こと」が出現する場合，直後に「よる」が出現する場合に「に」が使われやすく，同一文中に格助詞「から」が存在する場合，推定する助詞の直前の単語に「そこ・ところ・外」が出現する場合,「へ」が使われやすい等のルールも獲得できた.

表 6.11: 「に・へ」での素性分析

分類先	素性	確率	頻度
に	述部 V の最初の自立語「つく」	1.000	189
へ	述部 V の最初の自立語「行く」	0.948	466

6.2.4 分析３:「に・で」使い分け

分析の結果,「に」および「で」の使い分けに，助詞の直後の単語，述部 V における最初の自立語が関係することがわかった．有効な素性の例を表 6.12 に示す．以下に表 6.12 の素性を含む例文を示す.

助詞の直後の単語「対する」犯罪者に対して「ムチ打ち」の刑を導入する法案が四日までに米ミシシッピ州議会に提出された。

述部 V における最初の自立語「行われる」ルワンダからの報道によると、交換は各州中心都市の銀行で三日と四日に 行われ、それ以降旧紙幣は無価値となる。

この他にも，推定する助詞の直後の単語に「つく・なる・よる」が出現する場合「に」が使われやすく，同一文中に格助詞「に」が存在する場合，推定する助詞の直前の単語に「中」が出現する場合「で」が使われやすい等のルールも獲得できた.

表 6.12: 「に・で」での素性分析

分類先	素性	確率	頻度
に	対象の助詞の直後の単語が「対する」	1.000	52
で	述部 V の最初の自立語が「行われる」	0.903	62

6.2.5 分析４:「に・を」使い分け

分析の結果,「に」および「を」の使い分けに，述部 V における最初の自立語が関係することがわかった．有効な素性の例を表 6.13 に示す．以下に表 6.13 の素性を含む例文を示す.

述部 V における最初の自立語「よる」わが国にふさわしい国際貢献による 世界平和の創造

述部 V における最初の自立語「持つ」アンケートでは「好感を持つ 各政党の首脳、幹部名」も挙げてもらった。

この他にも，同一文中に格助詞「を」が出現する場合，推定する助詞の直前の単語に「前・日」が出現する場合,「に」が使われやすく，推定する助詞の直前の単語に「開く・求める・見る」が出現する場合「を」が使われやすい等のルールも獲得できた．

表 6.13: 「に・を」での素性分析

分類先	素性	確率	頻度
に	述部 V の最初の自立語が「よる」	1.000	371
を	述部 V の最初の自立語が「持つ」	0.844	123

6.3 推定結果の分析

6.3.1 「は・が」推定結果分析

「は・が」の推定の結果から，SVM が,「は」を「が」と誤って推定,「が」を「は」と誤って推定したものを各一文示す．

次の一文は，SVM が「は」を「が」と誤って推定したものである．素性の頻度分析の結果から，述部 V の係り先に「こと」が存在していると,「が」が出現しやすいことがわかっている（頻度:286，確率:0.898）．このルールから，誤った推定をした可能性がある．

「は」を「が」と誤って推定　しかし、実際【は】分かっていても、口にだせないことが多いのだろう。

次の一文は，SVM が「が」を「は」と誤って推定したものである．素性の頻度分析の結果から，述部 V に係る体言の文節 N 以外の主部に格助詞「が」が存在している場合,「は」が出現しやすいことがわかっている（頻度:396，確率:0.909）．このルールから，誤った推定をした可能性がある．

「が」を「は」と誤って推定　グロズヌイについては捕虜の大半【が】ほとんど予備知識がなかったという。

6.3.2 「に・へ」推定結果分析

「に・へ」の推定の結果から，SVM が,「に」を「へ」と誤って推定,「へ」を「に」と誤って推定したものを各一文示す．次の二文は，SVM が「に」を「へ」と誤って推定したもの,「へ」を「に」と誤って推定したものである．素性の頻度分析の結果からは，この例に合致する素性は見つからなかった．素性の取捨に基づく分析の結果では，有効的な素性も存在しなかったことから，他の分類問題よりも分類が困難であると考えられる．

また，他の素性分析の結果に比べて，獲得したルールの数も少ない．獲得ルールを増やすためには，分析データを増加させる必要がある．

「に」を「へ」と誤って推定　都立へ行っても塾へ通えばかえって教育費がかかるからと思ったものの、下の二人の子供も私立高校【に】進み、教育費は年々かさむ一方である。

「へ」を「に」と誤って推定　都立へ行っても塾【へ】通えばかえって教育費がかかるからと思ったものの、下の二人の子供も私立高校に進み、教育費は年々かさむ一方である。

6.3.3　「に・で」推定結果分析

「に・で」の推定の結果から，SVM が,「に」を「で」と誤って推定,「で」を「に」と誤って推定したものを各一文示す．素性の頻度分析の結果からは，この例に合致する素性は見つからなかった．素性の取捨に基づく分析の結果では，有効的な素性も存在しなかったことから，他の分類問題よりも分類が困難であると考えられる．

「に」を「で」と誤って推定　同省は通常国会【に】改正法案を提出し、今年十一月に実施する考え。

「で」を「に」と誤って推定　首相は社会党の党内情勢を説明したうえ【で】、さきがけの協力を要請した。

6.3.4　「に・を」推定結果分析

「に・を」の推定の結果から，SVM が,「に」を「を」と誤って推定,「を」を「に」と誤って推定したものを各一文示す.「に」を「を」と誤って推定した文に関しては，素性分析において，主部に組織名が出現すると,「を」が出現しやすいことがわかっている（頻度:55，確率:0.781）．また,「を」を「に」と誤って推定した文に関しては，素性分析において，述部に抽象名詞を含むと「に」が出現しやすいことがわかっている（頻度:541，確率:0.787）．これらルールから，誤った推定をした可能性がある．

「に」を「を」と誤って推定　山花貞夫新民主連合会長に@に@言い切っている。@と山花貞夫新民主連合会長【に】言い切っている。

「を」を「に」と誤って推定　有効期間を@を@延長する@同省はパスポートの有効期間【を】現行の二倍の十年に延長することをすでに決めているが、延長の対象となるのは二十歳以上が申請した場合で、取得・更新手数料は一万五千円とする方針。

第7章　考察

本章では，実験と分析の結果から考察を行う．各結果の考察は分析で行っているので，ここでの考察は総括的なものとなる．

7.1　助詞の推定実験について

推定の結果,「は・が」の分類における SVM の正解率は 0.760 であり,「に・へ」「に・で」「に・を」の正解率は 8〜9 割程度の結果であった．この数値は実際に，日本語学習者の文章の誤り訂正を行う際に十分な正解率であるといえる．ただし,「に・へ」に関しては，他の分類問題に比べて「へ」の F 値が低い．これは「へ」の出現数が低いためであると考えられる．また，94 年全てを用い教師数を増加させ,「に・へ」の教師のバランスを整えた SVM の実験では，F 値が増加した．このことから，さらに教師を増加させることによる，F 値の向上が期待される．

7.2　機械学習を用いた分析の利点と欠点

先行研究で助詞を分析する場合，アンケートを実施して分析を行っているのが主である．それらと機械学習や頻度分析を利用した分析を比較する．表 7.1 にアンケートと機械学習や頻度分析を利用した分析の違いを載せる．

表 7.1: 各分析の利点と欠点

アンケートの特徴	機械学習や頻度分析の特徴
大量のデータを用意することは困難	大量のデータから統計的に分析を行える
アンケートの内容を精査することにより，目的にあったデータを獲得可能	「へ」など，出現頻度がすくないものは，獲得が困難
分析は分析者の能力に左右される	客観的な分析を行える

7.3 「は・が」の使い分け

素性の取捨に基づく分析による素性分析の結果,「は・が」の分類の際に素性群「述部 V，体言の文節 N，N 以外体言，品詞，直，文脈情報」が有効であるとわかった．また，頻度分析による素性分析の結果,「述部 V，係り先，体言の文節 N，N 以外体言，分類」の各ルールは 10 以上獲得できた．これらは素性として有効であると予測される．両分析の共通する素性は「述部 V，体言の文節 N，N 以外体言」であることがわかった．田中ら [8] は,「は・が」の使い分けについて「は」は既知情報,「が」は未知情報を表すと述べている．本研究では，素性の取捨に基づく分析による素性分析の結果から有効な素性として「文脈情報」が得られており，先行研究の結果と一致する.「は・が」の使い分けが必要な文の比率の調査から，10 文に 9 文は使い分けが必要であることがわかった．

7.4 「に・へ」の使い分け

素性の取捨に基づく分析の結果,「に・へ」の分類の際に素性群「述部 V，体言の文節 N，共起，係り先 N，N 以外体言，品詞，直，文脈情報」が有効であるとわかった．また，頻度分析による素性分析の結果,「述部 V，直，共起，体言の文節 N，分類」の各ルールは 10 以上獲得できた．これらは素性として有効であると予測される．2.3.2 節で説明した先行研究では,「へ」の使い分けのルールに「A から B へ」が挙げられている．本研究の頻度分析による素性分析の結果から,「から」が文に出現した場合「へ」なりやすいことがわかっている（確率：355，頻度：0.802）．この結果は先行研究と一致している．また，両分析の共通する素性は「述部 V，体言の文節 N，共起，直」であることがわかった.「に・へ」の使い分けが必要な文の比率の調査から，10 文に 9 文は使い分けが必要であることがわかった.「に・へ」の使い分けの課題としては，分析時のデータを増加させることによってさらに有用な素性の分析を行う必要がある．

7.5 「に・で」の使い分け

素性の取捨に基づく分析の結果,「に・で」の分類の際に素性群「述部 V，体言の文節 N，N 以外体言，品詞，分類，直，他助詞」が有効であるとわかった．また，頻度分析による素性分析の結果,「述部 V，体言の文節 N，N 以外体言，分類」の各ルールは 10 以上獲得できた．これらは素性として有効であると予測される．両分析の共通する素性は「述部 V，体言の文節 N，N 以外体言，分類」であることがわかった.「に・で」の使い分けが必要な文の比率の調査から，10 文に 9 文は使い分けが必要であることがわかった.「に・で」の使い分けの課題としては，分析時のデータを増加させることによってさらに有用な素性の分析を行う必要がある．SVM の推定の正解率は 8 割程度あるため，問題のモデル化はある程度上手くいっているといえる．そのため，素性の分析方法に工夫が必要な可能性がある．

7.6 「に・を」の使い分け

素性の取捨に基づく分析の結果,「に・を」の分類の際に素性群「述部 V，体言の文節 N，係り先 N，N 以外体言，品詞，分類，直，他助詞，文脈情報」が有効であるとわかった．また，頻度分析による素性分析の結果,「述部 V，直，体言の文節 N，N 以外体言，分類」が有効であるとわかった．両分析の共通する素性は「述部 V，体言の文節 N，N 以外体言，分類，直」であることがわかった.「に・を」の使い分けが必要な文の比率の調査から，10 文に 10 文は使い分けが必要であることがわかった．

7.7 使い分けが必要でない文の扱い

助詞の使い分けにおいて，文法的に助詞の使い分けが必要でない場合がある．本研究で利用した新聞記事の文にも，文法的に使い分けが必要でない助詞が存在している．使い分けが必要でない助詞はデータから取り除き実験する方法が考えられる．しかし，文法的に正しい場合でも一般的に使われない表現は存在している．また，文章を作成する際に,「は・が」どちらでも意味は通るが,「は」を使った場合の方が，より良い文になる場合がある．そういった情報は，使い分けが必要でない助詞を取り除いたデータからは得ることができない．そのため本研究では，使い分けが必要でない助詞を含む文を省く等の処理を行っていない．

第8章　おわりに

日本語学習者の支援のために，本研究では，機械学習を利用した助詞の推定を行った．実験を行う前に，助詞の使い分けが必要か否かを，対象の助詞に対して調査した．結果から，使い分けが必要な助詞は9〜10割程度であり，本研究の必要性が確認できた．実験の結果,「は・が」の分類におけるSVMの正解率は0.760,「が」の推定はF値0.768(再現率：0.765，適合率：0.772),「は」の推定はF値0.751(再現率：0.755，適合率：0.748)であった．また,「に・で」「に・を」の分類において，SVMの手法が比較手法の中で最も高い値となった．また,「に・へ」においては全て「に」でない以外の手法がほぼ同等の正解率であった．実験データにおいて素性の分析（素性の取捨に基づく分析，素性の頻度分析）を行い,「が・は」の使い分けに役立つ表現を獲得した．副助詞「は」になりやすい表現として，述部に存在する助詞「だ」があった．また,「に・へ・で・を」の使い分けに役立つ表現を多数獲得した．これらの知見は，今後の助詞に関わる研究に役立つと思われる．

また，推定結果の誤り解析を行い，誤った原因を調査した．結果から,「に・へ」「に・で」において素性が不足している可能性があることがわかった．素性を拡充することで，さらなる正解率の向上が期待される．

謝辞

本研究を進めるにあたり，種々の御助言を頂きました鳥取大学工学部知能情報工学科計算機工学講座Ｃの村田真樹教授に心から御礼申し上げます．本研究を進めるにあたり，御指導を頂きました徳久雅人講師に心から御礼申し上げます．また，村上仁一准教授には，終始に渡り研究の進め方や本論文の書き方など，細部にわたる御指導を頂きました．ここに深く感謝いたします．本論文をご精読頂き有用なコメントを頂きました木村周平准教授に深く感謝いたします．その他様々な場面で御助言を頂いた計算機工学講座Ｃ研究室の皆様に感謝の意を表します．

参考文献

[1] H. Suzuki and K. Toutanova: 2006, “Learning to Predict Case Markers in Japanese,” ACL-COLING.

[2] H. Suzuki and K. Toutanova: 2006, “機械学習による日本語格助詞の予測” 言語処理学会年次大会発表論文集, pp.1119-1122.

[3] 久野 ススム, 染谷 方良: 1989, 日本語学の新展開, くろしお出版.

[4] 森田 良行: 1995, 日本語の視点　ことばを創る日本人の発想, 創拓社.

[5] 益岡 隆志, 田窪 行則: 1992, 基礎日本語文法—改訂版—, くろしお出版.

[6] 玉村 文郎: 1992, 日本語を学ぶ人のために, 世界思想社.

[7] 内元 清貴，村田 真樹，馬 青，関根 聡，井佐原 均 : 2000, “コーパスからの語順の獲得”, 自然言語処理，pp.163-180.

[8] 田中 稔子: 1990, 田中　稔子の日本語の文法—教師の疑問に答えます—, 近代文藝社.

[9] Y. Komori: 2003, “Disambiguating between ‘wa’ and ‘ga’ in Japanese”, CPSC, pp.30-34.

[10] 京大コーパス: http://nlp.ist.i.kyoto-u.ac.jp

[11] 村田 真樹: 2001, “機械学習手法を用いた日本語格解析—教師信号借用型と非借用型，さらには併用型—”, 情報処理学会自然言語処理研究会, 2001-NL-144, pp.113-120.

[12] 三上 章: 1960, 象は鼻が長い, くろしお出版.

[13] 安田 晴子, 森まどか, 劉 玉琴, 許 清平, 小野由美子: 2008, “格助詞「に」「で」の誤用研究 : タイ・中国の日本語学習者を対象に”, 鳴門教育大学実技教育研究, pp.19-25.

[14] 若生 正和: 2012, “韓国人日本語学習者による場所の格助詞「に」と「で」の選択に関する研究”, 大阪教育大学紀要　第 I 部門人文科学, pp.91-99.

[15] 杉村 泰: 2005, “上級・超上級日本語学習者に見る格助詞「に」と「へ」の使い分け”, 2005, 言語文化論集, pp.91-102.

[16] 蓮池 いずみ: 2004, “場所を示す格助詞選択のストラテジー　—韓国語母語話者と中国語母語話者の比較—”, 言葉と文化, pp.105-117.

[17] juman: http://nlp.ist.i.kyoto-u.ac.jp/index.php

[18] knp: http://nlp.ist.i.kyoto-u.ac.jp/index.php

[19] TinySVM: http://chasen.org/˜taku/software/Tiny SVM/

[20] 村田 真樹, 神崎 享子, 内元 清貴, 馬青, 井佐原 均: 2000, “意味ソート msort —意味的並びかえ手法による辞書の構築例とタグつきコーパスの作成例と情報提示システム例—”, 自然言語処理, 7 巻, 1 号, 89-96.

[21] 分類語彙表: http://www.ninjal.ac.jp/products-k/kanko/goihyo/

[22] S. Miura, L. Fan, M. Murata and M. Tokuhisa: 2012, “Automatic Selection and Contextual Analysis of the Japanese Particles ‘Wa’ and ‘Ga’ Using Machine Learning”, SCIS-ISIS, PT-4.

付 録A　追加実験：先行研究手法３の改良

先行研究３について単語を利用した方が正解率が向上すると指摘を受けた．ここに追加実験として，先行研究手法３を単語で行った結果を紹介する．

A.1　先行研究手法３＋

対象の助詞の直前，直後に出現する単語を利用し，X に入る助詞はどちらであるかを推定する．具体的には，現在解いている箇所の直前の単語の品詞を A とし，直後の単語の品詞を B とするとき，学習データにおいて A と B に挟まれる個所により高い頻度で出現する助詞を求め、それを現在解いている箇所の助詞とする．例えば,「は・が」の分類において，X の直前に名詞が出現しており，X の直後に動詞が出現している状況で，学習データにおいて直前に名詞が出現し直後に動詞が出現する場合に「は」になる確率が「が」になる確率よりも大きいのであれば，X に入る助詞は「は」であると推定する．

A.2　結果

先行研究手法３＋を利用し推定を行った．結果を表 A.1 に示す．結果から，先行研究手法３＋は,「に・で」において，先行研究手法３よりも高い正解率となった．しかし，それらよりも提案手法（SVM）の正解率のほうが高かった．

表 A.1: 先行研究手法３＋の正解率

分類問題	正解率
「は・が」	0.556
「に・へ」	0.987
「に・で」	0.716
「に・を」	0.551

付 録B　獲得ルール一覧

頻度分析によって獲得されたルールを載せる．素性は以下の記号で表現されている．Bの後に続く数値は分類語彙表の番号である．

用全単 述部Vにおける名詞、動詞、形容詞、指示詞の単語の連続（述部Vにおける自立語の連続）

用単 述部Vにおける最初の名詞、動詞、形容詞、指示詞（述部Vにおける最初の自立語）

用単品 「用単」の単語の品詞

用Ｂ７など 「用単」の分類語彙表の分類番号分類語彙表 [12] の分類番号の 1,2,3,4,5,6,7 桁までの数字．ただし，分類番号に対して文献 [9] の表の変更を行っている．

用形列 述部Vの文節内の「用単」より右側（後続）の単語

連体__ 述部Vの係り先の主部の文節の情報（この文節は体言の文節Nではない場合)

自格__ 体言の文節Nの文節の情報

格 [助詞 __] 述部Vにかかる体言の文節N以外の主部の文節の情報体言の文節Nとは [助詞] の助詞で結ばれる。これら 3 つはそれぞれ下記の素性を持つ

__全単 その文節における名詞、動詞、形容詞、指示詞の単語の連続（自立語の連続）

__存在 その文節の存在

__単 その文節における最後の名詞、動詞、形容詞、指示詞の単語（最後の自立語）

__単品 「__単」の品詞

__Ｂなど 「__単」の分類語彙表の分類番号分類語彙表 [12] の分類番号の 1,2,3,4,5,6,7 桁までの数字．ただし，分類番号に対して文献 [9] の表の変更を行っている．

直前単語 解析対象の助詞の直前の単語

直前品詞 「直前単語」の品詞

直後単語 解析対象の助詞の直後の単語

直後品詞 「直後単語」の品詞

共単 文内の単語（動詞、名詞、形容詞、副詞、指示詞のみ）

共Bなど 「共単」の分類語彙表の分類番号分類語彙表 [12] の分類番号の 1,2,3,4,5,6,7 桁までの数字．ただし，分類番号に対して文献 [9] の表の変更を行っている．

文内助詞存在 解析対象の文内の解析対象の文節以外の文節にある助詞

全ての名詞が前方に存在：1 解析対象の文節内の名詞がすべて前方に存在している。名詞がわかれて前方に存在していてもよい。

表 B.1: 「は・が」の使い分け「は」のルール

分類先	素性	確率	頻度
は	格がＢ１：5	0.912	57
は	格がＢ１：a	0.901	203
は	格がＢ２：aa	0.908	120
は	格がＢ２：ab	0.891	83
は	格が存在	0.909	396
は	格が単品：名詞	0.907	391
は	格とＢ３：aa3	0.764	212
は	格とＢ３：ab1	0.759	108
は	格とＢ４：aa31	0.776	170
は	格とＢ５：aa311	0.777	144
は	格とＢ７：aa31104	0.779	136
は	格と単：する	0.784	130
は	格と単品：形容詞	0.763	93
は	格をＢ４：aa10	0.818	99
は	格をＢ４：ab16	0.792	101
は	格をＢ５：aa100	0.818	99
は	格をＢ７：aa10003	0.903	83
は	格を全単：こと	0.913	81
は	格を単：こと	0.913	81
は	格推定：の単品：名詞	0.758	170
は	自格Ｂ４：8114	0.904	125
は	自格Ｂ５：81141	0.972	72
は	自格Ｂ７：5201001	0.760	75
は	自格Ｂ７：5241106	0.758	91
は	自格Ｂ７：5371003	0.790	62
は	自格単：首相	0.769	65
は	自品：判定詞	0.754	806
は	直後単語：「	0.827	436
は	直後品詞：特殊	0.812	464
は	直前単語：首相	0.769	65
は	用Ｂ１：5	0.756	82
は	用Ｂ３：aa1	0.863	95
は	用Ｂ４：7116	0.764	51
は	用Ｂ４：aa10	0.971	71
は	用Ｂ５：aa100	0.971	71
は	用Ｂ５：ab231	0.861	72
は	用Ｂ７：ab23101	0.880	67
は	用形列：だ	0.754	809
は	用全単：話す	0.962	54
は	用単：話す	0.962	54

表 B.2: 「は・が」の使い分け「が」のルール

分類先	素性	確率	頻度
が	格Ｂ１：8	0.789	76
が	格推定：はＢ２：81	0.789	76
が	格推定：はＢ３：811	0.789	76
が	共単：接尾辞：ごろ	0.801	121
が	共単：動詞：決まる	0.769	52
が	自格Ｂ１：9	0.803	117
が	自格Ｂ２：91	0.803	117
が	自格Ｂ３：912	0.836	98
が	直後単語：あう	0.888	117
が	直後単語：ある	0.914	222
が	直後単語：できる	0.853	75
が	直後単語：出る	0.923	52
が	直後単語：多い	0.808	73
が	直後品詞：動詞	0.823	2061
が	用Ｂ５：aa652	0.761	67
が	用Ｂ７：aa22001	0.769	152
が	用全単：あう	0.763	148
が	用全単：出る	0.845	71
が	用全単：必要だ	0.867	53
が	用単：あう	0.769	152
が	用単：出る	0.845	71
が	用単：必要だ	0.867	53
が	連体Ｂ１：5	0.863	95
が	連体Ｂ１：6	0.940	67
が	連体Ｂ１：7	0.811	53
が	連体Ｂ１：8	0.934	153
が	連体Ｂ１：a	0.870	770
が	連体Ｂ２：53	0.883	60
が	連体Ｂ２：71	0.811	53
が	連体Ｂ２：81	0.934	153
が	連体Ｂ２：aa	0.868	579
が	連体Ｂ２：ab	0.874	191
が	連体Ｂ３：711	0.811	53
が	連体Ｂ３：811	0.934	153
が	連体Ｂ３：aa1	0.878	379
が	連体Ｂ３：aa2	0.873	103
が	連体Ｂ３：aa4	0.886	53
が	連体Ｂ３：ab1	0.860	93
が	連体Ｂ４：8119	0.941	51
が	連体Ｂ４：aa10	0.875	354
が	連体Ｂ４：aa21	0.824	74
が	連体Ｂ５：81190	0.941	51
が	連体Ｂ５：aa100	0.875	354
が	連体Ｂ５：aa212	0.779	59
が	連体Ｂ７：aa10003	0.899	288
が	連体全単：こと	0.898	286
が	連体存在	0.873	1255
が	連体単：こと	0.898	286
が	連体単品：名詞	0.873	1253

表 B.3: 「に・へ」の使い分け「に」のルール

分類先	素性	確率	頻度
に	共Ｂ５：ab820	0.754	57
に	共単：助動詞：べきだ	0.761	67
に	共単：接尾辞：%	0.757	107
に	共単：動詞：つく	0.772	465
に	共単：動詞：関する	0.839	87
に	共単：動詞：述べる	0.761	126
に	共単：動詞：対する	0.821	235
に	共単：名詞：記者	0.756	82
に	共単：名詞：発表	0.753	73
に	自格Ｂ３：aa4	0.829	117
に	自格Ｂ３：ab1	0.838	383
に	自格Ｂ３：ab2	0.753	146
に	自格Ｂ３：ab4	0.764	102
に	自格Ｂ４：aa10	0.892	112
に	自格Ｂ４：aa21	0.790	143
に	自格Ｂ４：aa40	0.881	59
に	自格Ｂ４：ab16	0.830	130
に	自格Ｂ４：ab17	0.848	86
に	自格Ｂ４：ab18	0.870	77
に	自格Ｂ５：71160	0.893	159
に	自格Ｂ５：aa100	0.892	112
に	自格Ｂ５：aa212	0.942	52
に	自格Ｂ５：ab170	0.981	55
に	自格Ｂ７：7116001	0.895	124
に	自格Ｂ７：aa10003	0.898	59
に	自格全単：	0.916	60
に	自格全単：こと	0.894	57
に	自格単：	0.916	60
に	自格単：こと	0.894	57
に	直後単語：する	0.966	118
に	直後単語：なる	0.995	208
に	直後単語：よる	1.000	171
に	直後単語：対する	0.991	119
に	直前単語：こと	0.894	57
に	直前品詞：接尾辞	0.750	621
に	直前品詞：副詞	0.910	56
に	用Ｂ３：aa2	0.849	765
に	用Ｂ４：8113	0.988	173
に	用Ｂ４：aa21	0.829	451
に	用Ｂ４：aa22	0.960	227
に	用Ｂ４：aa30	0.795	137
に	用Ｂ４：aa32	0.919	62
に	用Ｂ５：81135	0.988	173
に	用Ｂ５：aa212	0.887	276
に	用Ｂ５：aa220	0.960	227

表 B.4: 「に・へ」の使い分け「へ」のルール

分類先	素性	確率	頻度
へ	格からＢ１：5	0.823	102
へ	格からＢ１：a	0.784	93
へ	格からＢ２：53	0.923	65
へ	格から存在	0.802	355
へ	格から単品：名詞	0.797	336
へ	格にＢ１：a	0.772	246
へ	格にＢ２：ab	0.821	129
へ	共Ｂ５：aa626	0.768	329
へ	共Ｂ５：aa627	0.806	992
へ	共単：助詞：ね	0.775	58
へ	共単：接尾辞：る	0.759	54
へ	共単：動詞：運ぶ	0.904	63
へ	共単：動詞：帰る	0.935	108
へ	共単：動詞：向かう	0.838	149
へ	共単：動詞：向ける	0.788	208
へ	共単：動詞：行く	0.912	570
へ	共単：動詞：進む	0.797	79
へ	共単：動詞：走る	0.766	60
へ	共単：動詞：送る	0.800	100
へ	共単：動詞：飛ぶ	0.864	59
へ	共単：動詞：戻る	0.797	79
へ	共単：動詞：来る	0.809	121
へ	共単：名詞：電話	0.771	105
へ	自格Ｂ２：53	0.783	1290
へ	自格Ｂ２：65	0.763	715
へ	自格Ｂ３：535	0.836	552
へ	自格Ｂ３：536	0.820	462
へ	自格Ｂ３：614	0.754	118
へ	自格Ｂ３：657	0.763	715
へ	自格Ｂ４：5352	0.792	53
へ	自格Ｂ４：5359	0.844	347
へ	自格Ｂ４：5360	0.757	95
へ	自格Ｂ４：5363	0.887	142
へ	自格Ｂ４：5364	0.831	83
へ	自格Ｂ４：5365	0.831	113
へ	自格Ｂ４：6140	0.840	69
へ	自格Ｂ４：6570	0.895	230
へ	自格Ｂ４：6573	0.870	131
へ	自格Ｂ４：8119	0.758	178
へ	自格Ｂ５：53520	0.792	53
へ	自格Ｂ５：53590	0.844	347
へ	自格Ｂ５：53600	0.757	95
へ	自格Ｂ５：53630	0.887	142
へ	自格Ｂ５：53640	0.831	83
へ	自格Ｂ５：53650	0.831	113

表 B.5: 「に・へ」の使い分け「に」のルール 2

分類先	素性	確率	頻度
に	用Ｂ７：8113512	1.000	171
に	用Ｂ７：aa21001	0.829	82
に	用Ｂ７：aa21201	0.923	236
に	用Ｂ７：aa22001	0.959	148
に	用Ｂ７：aa30001	0.811	101
に	用Ｂ７：aa31102	0.990	218
に	用全単：ある	0.820	100
に	用全単：つく	1.000	189
に	用全単：なる	0.928	223
に	用全単：対する	0.991	119
に	用単：ある	0.820	100
に	用単：つく	1.000	190
に	用単：なる	0.927	233
に	用単：よる	1.000	171
に	用単：対する	0.991	119
に	用単品：形容詞	0.770	144

表 B.6: 「に・へ」の使い分け「へ」のルール 2

分類先	素性	確率	頻度
へ	自格Ｂ５：61400	0.840	69
へ	自格Ｂ５：65700	0.895	230
へ	自格Ｂ５：65720	0.754	118
へ	自格Ｂ５：65730	0.870	131
へ	自格Ｂ５：65740	0.788	52
へ	自格Ｂ５：81190	0.758	178
へ	自格Ｂ７：5359001	0.794	73
へ	自格Ｂ７：5359006	0.830	130
へ	自格Ｂ７：5359007	0.884	78
へ	自格Ｂ７：5363011	0.912	57
へ	自格Ｂ７：6570002	0.960	153
へ	自格Ｂ７：6573001	0.784	51
へ	自格Ｂ７：6573025	0.915	59
へ	自格Ｂ７：8119001	0.818	154
へ	自格Ｂ７：aa14002	0.815	76
へ	自格全単：そこ	0.924	66
へ	自格全単：ところ	0.970	68
へ	自格全単：どこ	0.987	81
へ	自格全単：外	0.923	52
へ	自格全単：日本	0.814	70
へ	自格単：そこ	0.924	66
へ	自格単：ところ	0.970	68
へ	自格単：どこ	0.987	81
へ	自格単：外	0.919	62
へ	自格単：学校	0.927	55
へ	自格単：所	0.882	51
へ	自格単：日本	0.791	72
へ	自格単：方向	0.784	51
へ	自格単品：指示詞	0.761	252
へ	直後単語：帰る	0.986	73
へ	直後単語：向かう	0.875	129
へ	直後単語：向ける	0.852	170
へ	直後単語：行く	0.940	404
へ	直後単語：出る	0.823	85
へ	直後単語：来る	0.841	63
へ	直前単語：そこ	0.924	66
へ	直前単語：ところ	0.970	68
へ	直前単語：どこ	0.987	80
へ	直前単語：外	0.918	61
へ	直前単語：学校	0.925	54
へ	直前単語：日本	0.802	71
へ	直前単語：方向	0.784	51
へ	直前品詞：指示詞	0.782	244

表 B.7: 「に・へ」の使い分け「へ」のルール 3

分類先	素性	確率	頻度
へ	用Ｂ４：6573	0.824	324
へ	用Ｂ４：8110	0.830	100
へ	用Ｂ４：aa62	0.831	1474
へ	用Ｂ４：ab24	0.765	64
へ	用Ｂ５：65730	0.824	324
へ	用Ｂ５：81100	0.830	100
へ	用Ｂ５：aa620	0.844	77
へ	用Ｂ５：aa622	0.842	76
へ	用Ｂ５：aa626	0.913	186
へ	用Ｂ５：aa627	0.911	646
へ	用Ｂ５：ab241	0.836	55
へ	用Ｂ７：6573002	0.851	309
へ	用Ｂ７：8110001	0.913	69
へ	用Ｂ７：aa31112	0.885	61
へ	用Ｂ７：aa62001	0.870	54
へ	用Ｂ７：aa62101	0.777	72
へ	用Ｂ７：aa62603	0.901	61
へ	用Ｂ７：aa62605	0.926	95
へ	用Ｂ７：aa62707	0.839	112
へ	用Ｂ７：aa62713	0.944	507
へ	用形列：いく	0.883	86
へ	用形列：込む	0.788	85
へ	用全単：帰る	0.950	80
へ	用全単：向かう	0.875	128
へ	用全単：向ける	0.835	176
へ	用全単：行く	0.948	443
へ	用全単：出る	0.824	91
へ	用全単：送る	0.901	51
へ	用全単：戻る	0.865	52
へ	用全単：来る	0.828	70
へ	用単：運ぶ	0.905	53
へ	用単：帰る	0.952	85
へ	用単：向かう	0.877	131
へ	用単：向ける	0.836	177
へ	用単：行く	0.948	466
へ	用単：出る	0.805	103
へ	用単：送る	0.910	67
へ	用単：戻る	0.870	54
へ	用単：来る	0.821	73

表 B.8: 「に・で」の使い分け「に」のルール

分類先	素性	確率	頻度
に	自格Ｂ２：52	0.770	227
に	自格Ｂ３：524	0.846	104
に	直後単語：する	0.983	62
に	直後単語：つく	1.000	96
に	直後単語：なる	1.000	172
に	直後単語：よる	1.000	129
に	直後単語：対する	1.000	52
に	用Ｂ１：8	0.826	196
に	用Ｂ２：65	0.836	61
に	用Ｂ２：81	0.826	196
に	用Ｂ３：657	0.836	61
に	用Ｂ３：811	0.826	196
に	用Ｂ３：aa2	0.787	541
に	用Ｂ４：8113	0.984	132
に	用Ｂ４：aa21	0.794	350
に	用Ｂ４：aa22	0.804	138
に	用Ｂ５：81135	0.992	131
に	用Ｂ５：aa210	0.851	94
に	用Ｂ５：aa212	0.779	254
に	用Ｂ５：aa220	0.804	138
に	用Ｂ７：8113512	1.000	129
に	用Ｂ７：aa21201	0.788	232
に	用Ｂ７：aa22001	0.760	96
に	用Ｂ７：aa31102	0.945	128
に	用形列：は	0.826	69
に	用全単：なる	0.786	230
に	用単：なる	0.786	230
に	用単：対する	1.000	52

表 B.9: 「に・で」の使い分け「で」のルール

分類先	素性	確率	頻度
で	格にＢ１：5	0.882	85
で	格にＢ１：6	0.803	51
で	格にＢ１：7	0.811	69
で	格にＢ１：a	0.775	138
で	格にＢ２：71	0.811	69
で	格にＢ２：ab	0.803	66
で	格にＢ３：711	0.811	69
で	格にＢ４：7116	0.818	55
で	格にＢ５：71160	0.803	51
で	格に存在	0.784	479
で	格に単品：名詞	0.788	477
で	格推定：はＢ３：537	0.835	85
で	自格Ｂ４：5365	0.781	55
で	自格Ｂ４：aa20	0.759	158
で	自格Ｂ５：53650	0.781	55
で	自格Ｂ５：aa201	0.758	124
で	自格Ｂ５：ab161	0.767	56
で	自格Ｂ７：aa20102	0.810	100
で	自格全単：中	0.846	78
で	自格単：中	0.846	78
で	直後単語：開く	0.826	52
で	直前単語：中	0.767	86
で	用Ｂ３：ab5	0.816	136
で	用Ｂ４：ab23	0.807	52
で	用Ｂ４：ab53	0.851	108
で	用Ｂ４：ab71	0.796	54
で	用Ｂ５：aa602	0.751	137
で	用Ｂ５：ab530	0.851	108
で	用Ｂ７：aa60204	0.842	76
で	用Ｂ７：ab53002	0.861	94
で	用全単：開く	0.833	72
で	用全単：行われる	0.903	62
で	用単：開く	0.833	72
で	用単：行われる	0.903	62

表 B.10: 「に・を」の使い分け「に」のルール

分類先	素性	確率	頻度
に	格がＢ２：61	0.763	55
に	格がＢ３：ab1	0.800	90
に	格がＢ３：ab2	0.792	53
に	格をＢ１：5	1.000	121
に	格をＢ１：7	1.000	51
に	格をＢ１：a	0.992	562
に	格をＢ２：52	1.000	58
に	格をＢ２：61	1.000	66
に	格をＢ２：71	1.000	51
に	格をＢ２：aa	0.995	218
に	格をＢ２：ab	0.991	344
に	格をＢ３：711	1.000	51
に	格をＢ３：aa6	0.981	53
に	格をＢ３：ab1	0.992	140
に	格を存在	0.994	1002
に	格を単品：名詞	0.993	991
に	共単：動詞：あたる	0.781	55
に	自格Ｂ４：5371	0.967	62
に	自格Ｂ４：6570	0.900	60
に	自格Ｂ４：6574	0.841	126
に	自格Ｂ５：53710	0.967	62
に	自格Ｂ５：65700	0.900	60
に	自格Ｂ５：65742	0.857	77
に	自格Ｂ５：81150	0.814	70
に	自格Ｂ７：6574203	0.960	51
に	自格Ｂ７：7116001	0.806	238
に	直後単語：ある	0.986	73
に	直後単語：つく	0.986	215
に	直後単語：よる	1.000	371
に	直後単語：向ける	0.941	51
に	直後単語：対する	1.000	138
に	直後単語：入る	1.000	79
に	直前単語：月	1.000	83
に	直前単語：前	0.961	78
に	直前単語：日	0.935	78
に	用Ｂ１：6	0.765	235
に	用Ｂ２：65	0.886	159
に	用Ｂ３：657	0.886	159
に	用Ｂ３：aa2	0.874	1272
に	用Ｂ４：6573	0.893	94
に	用Ｂ４：8113	0.997	373
に	用Ｂ４：aa21	0.906	779
に	用Ｂ４：aa22	0.986	305
に	用Ｂ４：aa30	0.922	193

表 B.11: 「に・を」の使い分け「を」のルール

分類先	素性	確率	頻度
を	格でＢ１：7	0.776	67
を	格でＢ２：71	0.776	67
を	格でＢ３：711	0.776	67
を	格にＢ１：5	0.874	239
を	格にＢ１：6	0.845	175
を	格にＢ１：7	0.757	140
を	格にＢ１：a	0.824	466
を	格にＢ２：52	0.896	125
を	格にＢ２：53	0.845	110
を	格にＢ２：61	0.825	63
を	格にＢ２：65	0.864	103
を	格にＢ２：71	0.757	140
を	格にＢ２：aa	0.811	271
を	格にＢ２：ab	0.841	195
を	格にＢ３：524	0.929	71
を	格にＢ３：657	0.864	103
を	格にＢ３：711	0.757	140
を	格にＢ３：aa1	0.809	63
を	格にＢ３：aa2	0.822	107
を	格にＢ３：aa6	0.769	52
を	格にＢ３：ab1	0.910	67
を	格にＢ４：7116	0.752	117
を	格にＢ４：aa21	0.773	75
を	格に存在	0.808	1368
を	格に単品：名詞	0.807	1346
を	格推定：はＢ３：537	0.765	183
を	格推定：はＢ４：5371	0.873	63
を	格推定：はＢ５：53710	0.873	63
を	共単：動詞：つくる	0.785	56
を	共単：動詞：果たす	0.815	65
を	共単：動詞：変える	0.773	53
を	共単：動詞：目指す	0.751	145
を	自格Ｂ３：aa5	0.766	103
を	自格Ｂ４：aa50	0.766	103
を	自格Ｂ４：ab10	0.769	104
を	自格Ｂ４：ab13	0.763	55
を	自格Ｂ４：ab14	0.881	144
を	自格Ｂ４：ab20	0.770	87
を	自格Ｂ４：ab25	0.793	63
を	自格Ｂ４：ab67	0.761	63
を	自格Ｂ５：aa500	0.843	51

表 B.12: 「に・を」の使い分け「に」のルール 2

分類先	素性	確率	頻度
に	用Ｂ５：65730	0.893	94
に	用Ｂ５：81135	0.997	373
に	用Ｂ５：aa210	0.799	249
に	用Ｂ５：aa212	0.958	526
に	用Ｂ５：aa220	0.986	305
に	用Ｂ５：aa300	0.922	193
に	用Ｂ５：aa627	0.835	73
に	用Ｂ７：6573002	0.950	81
に	用Ｂ７：8113512	1.000	371
に	用Ｂ７：aa21001	0.949	99
に	用Ｂ７：aa21201	0.991	462
に	用Ｂ７：aa22001	1.000	184
に	用Ｂ７：aa22002	0.984	65
に	用Ｂ７：aa31102	0.848	303
に	用Ｂ７：aa63201	1.000	90
に	用形列：れる	0.837	424
に	用全単：ある	1.000	126
に	用全単：つく	0.986	217
に	用全単：なる	0.991	448
に	用全単：よる	1.000	371
に	用全単：向ける	0.943	53
に	用全単：出る	0.843	51
に	用全単：対する	1.000	138
に	用全単：入る	1.000	85
に	用単：つく	0.986	220
に	用単：なる	0.991	452
に	用単：よる	1.000	371
に	用単：向ける	0.927	55
に	用単：出る	0.843	51
に	用単：入る	1.000	87

表 B.13: 「に・を」の使い分け「を」のルール 2

分類先	素性	確率	頻度
を	自格Ｂ５：ab161	0.795	98
を	自格Ｂ５：ab181	0.773	53
を	自格Ｂ５：ab200	0.796	54
を	自格Ｂ５：ab670	0.761	63
を	用Ｂ３：ab1	0.770	1526
を	用Ｂ３：ab3	0.802	71
を	用Ｂ４：8110	0.788	180
を	用Ｂ４：ab14	0.765	298
を	用Ｂ４：ab15	0.803	66
を	用Ｂ４：ab16	0.786	657
を	用Ｂ４：ab19	0.819	222
を	用Ｂ４：ab20	0.788	123
を	用Ｂ４：ab49	0.897	127
を	用Ｂ４：ab62	0.831	107
を	用Ｂ４：ab80	0.859	135
を	用Ｂ４：ab95	0.807	57
を	用Ｂ５：81100	0.788	180
を	用Ｂ５：aa604	0.834	205
を	用Ｂ５：aa620	0.822	90
を	用Ｂ５：aa663	0.780	73
を	用Ｂ５：ab101	0.761	63
を	用Ｂ５：ab142	0.864	229
を	用Ｂ５：ab150	0.803	66
を	用Ｂ５：ab161	0.910	67
を	用Ｂ５：ab165	0.903	83
を	用Ｂ５：ab166	0.853	82
を	用Ｂ５：ab167	0.764	123
を	用Ｂ５：ab191	0.816	71
を	用Ｂ５：ab192	0.831	107
を	用Ｂ５：ab203	0.873	63
を	用Ｂ５：ab492	0.948	116
を	用Ｂ５：ab620	0.822	96
を	用Ｂ５：ab800	0.876	97

表 B.14: 「に・を」の使い分け「を」のルール 3

分類先	素性	確率	頻度
を	用Ｂ７：8110002	0.829	129
を	用Ｂ７：aa31008	0.807	166
を	用Ｂ７：aa32005	0.862	80
を	用Ｂ７：aa60201	0.792	82
を	用Ｂ７：aa60407	0.924	93
を	用Ｂ７：aa62001	0.775	58
を	用Ｂ７：ab14204	0.857	77
を	用Ｂ７：ab14205	0.972	74
を	用Ｂ７：ab16206	0.775	116
を	用Ｂ７：ab16706	0.863	66
を	用Ｂ７：ab19201	0.779	68
を	用Ｂ７：ab73014	0.761	63
を	用Ｂ７：ab80001	0.923	65
を	用形列：せる	0.772	193
を	用形列：ながら	0.791	72
を	用全単：求める	0.857	77
を	用全単：決める	0.931	58
を	用全単：見る	0.890	73
を	用全単：使う	0.762	59
を	用全単：持つ	0.844	122
を	用全単：示す	0.904	73
を	用全単：受ける	0.951	83
を	用全単：目指す	0.966	60
を	用単：求める	0.857	77
を	用単：決める	0.931	58
を	用単：見る	0.894	76
を	用単：使う	0.770	61
を	用単：持つ	0.840	125
を	用単：示す	0.906	75
を	用単：受ける	0.952	84
を	用単：目指す	0.967	61

付 録C　追加実験：先行研究手法の改良ーバイグラムー

追加実験として，バイグラムで推定を行った結果を載せる．

C.1　結果

先行研究手法を利用し推定を行った．結果を表 C.1 に示す．結果から，バイグラムの手法は,「に・で」「に・を」において，SVM を除く他の全ての手法よりも高い正解率となった．しかし，それらよりも提案手法（SVM）の正解率のほうが高かった．

表 C.1: 先行研究手法 (n-gram)

分類問題	正解率
「は・が」	0.618
「に・へ」	0.972
「に・で」	0.769
「に・を」	0.744