概要

学生などの論文を書き慣れていない者は，誤字や脱字，説明不足な表現を含め，論文として不適切な表現を用いがちになる．しかし執筆者本人でその間違いに気づくことは困難である．よって，そのような不適切な表現を検出する支援技術があれば便利である．そこで本研究では，不適切な表現の検出システム支援として，学生論文から論文として不適切な表現の収集と分析を行う．

具体的には，教員による修正の前と後の学生論文の差分を取り，学生論文の不適切な表現の収集と分析を行った．

その結果『語彙表現の修正』『文法誤りの修正』が多く行われていることがわかった．また，差分に含まれる高頻度の 2 単語連続を分析することで，特徴のある傾向を発見した．例えば,「〜ている」という表現が修正前の表現に多く存在していた．この表現は積極性に欠けた他人事のような印象を与える表現であり修正されたものと思われる．

目 次

表 目 次

図 目 次

第1章　はじめに

研究を始めたばかりの論文を書き慣れていない者は，誤字や脱字，説明不足な表現など論文として不適切な表現を用いがちになる．しかし学生自身でその不適切な文や表現に気づくことは困難である．そこで不適切な表現を検出，修正する支援技術があれば便利である．そのためには学生論文の不適切な表現の事例の収集と分析を行うことが重要である．そこで本研究では，論文として不適切な表現の収集と分析を行う．

具体的には，論文を執筆した学生の指導担当教員により修正が行われた学生論文を修正後論文，教員による修正が行われる前の論文を修正前論文とし，この二つで差分を取り，不適切と思われる表現を収集する．またここで得られた差分の分類を行い，どのような不適切な修正があるかも分析する．さらに，単語の頻度に基づく分析も行い，学生に用いられやすい不適切な表現の傾向や偏りを分析する．

本研究の主要な点を以下に示す．

- 差分抽出の技術により修正前と後の学生論文から多くの不適切な表現を抽出した．その不適切な表現を分類することで，どのような種類の不適切な表現が学生論文に存在するかを明らかにした (3.2.4 節の表 3.1)．

- 差分に含まれる高頻度の 2 単語連続を分析することで，種々の特徴のある傾向を発見した．例えば,「ている」という表現が修正前の表現に多く存在していた．この表現は積極性に欠けた他人事のような印象を与える表現であり修正されたものと思われる。

第2章　先行研究

2.1　関連研究

- 古本ら [1] は日本人学生の「論文を書く力」に不満を感じ，特に理系分野ではこの問題が大きな課題となっていたため，日本語専門教員 3 人による調査を行った．具体的には，工学を専門とする日本人学生が書いた文章にみられる基礎的な問題点として分析を行い，学生は内容を客観的に伝える文章を書く力が不足していることを発見した．

- 村田ら [2] は差分を用いた言い換えパターンを抽出する技術を利用して，英語運用における個人的な誤りパターンを抽出するシステムを作成した．英文校閲前のものと英文校閲後のもので差分を取り，この差分を誤りパターンとして抽出し，頻度を計算し，結果を考察した．この研究は差分を用いた分析手段として参考にした．

- 阿辺川ら [3] は下訳から修正訳に至る過程でどのような要因で修正操作が施されるかの解明を行い，翻訳初心者が作成したぎこちない本訳文に対して修正候補が提示できる機能の開発を試みている．

- 藤田ら [4] はコーパスから獲得した大規模な正例に基づいて格構造の適格さを定量化するモデルを構築し，語彙・構文的言い換えにおいて頻繁に生じる動詞格構造の不整合を自動的に検出する方法を提案している．

- 飯田ら [5] は日本語書き言葉を対象とした参照表現の自動生成課題，特に参照表現を省略するか否かを分類する問題を対象に，既存研究で利用されている談話的特徴を考慮した自動分類モデルを提案している．

2.2　先行研究との相違点

古本ら [1] は工学を専門とする日本人学生が書いた文章にみられる基礎的な問題点として，提示したテーマに沿った文章を学生に書かせ，3 人の日本語教員が全て人手で添削を行い，学生の文章にみられる問題点の分析を行った．しかし，論文内の不適切な箇所を検出し修正するために，毎回日本語専門教員によって人手で添削を行うのは多大な労力や手間がかかってしまう．なるべく人手の手間を減らし，不適切な表現を検出し修正する支援システムが自動化に近づいていくことができれば理想的である．そのためには，不適切と思われる表現の収集と分析を行うことが重要であり，これらを行う過程に自動的な技術を用いることで，支援システムの自動化に近づけることができると考えられる．そこで本研究では不適切な表現の収集と分析に工学的な手法を用いてを行い，不適切な表現の検出，修正を行うシステムの自動化に向けた支援を目指す．

第3章　差分の抽出と分類

ここでは，指導担当教員による修正が行われる前の学生論文と，修正が行われた後の学生論文で差分をとり，不適切と思われる表現を収集する．差分を得るための手法として mdiff コマンド [6] を用いた．さらに抽出によって得られた差分箇所の分類を行う．分類を行うことで，どのような不適切な表現が多く用いられているかを知ることができる．

図 3.1 は本研究で行う差分抽出の流れである．抽出の処理に用いる mdiff コマンドの説明は 3.1 節にて行う．また，実際に手順に沿った詳しい実験内容の説明については 3.2 節にて行う．

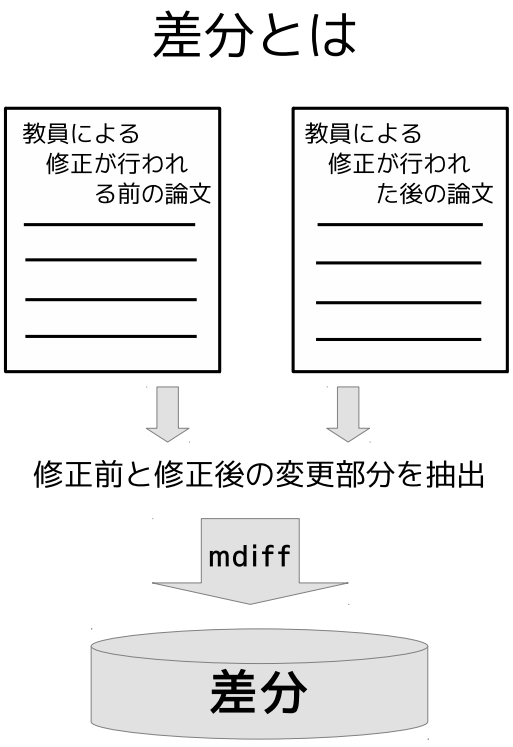

図 3.1: 差分とは

3.1 mdiffとは

UNIXに標準で搭載されているコマンドとしてdiffコマンドがある．このdiffコマンドは一般には差分の検出に利用され，与えられた二つのファイル間の違いを探し，順序情報を保持したまま結果を出力する．例えば，

今日
学校へ
行く

ということが書いてあるファイルを

今日
大学に
行く

という文に修正して，書き換えたファイルがあるとする．これらの2つのファイルに対してdiffコマンドを行うと差分の部分が

< 学校へ
> 大学に

のような形で出力される．

「今日」と「行く」という部分は修正前のファイルにも修正後のファイルにも存在しており，書き換えられていない．これは修正前後での共通部分として扱い，共通部分の間にある書き換えられた「学校」と「大学」が差分として出力される仕組みとなっている．

これに加えて diff コマンドには-D オプションという便利なオプションがあり，これをつけて diff コマンドを使うと差分部分だけでなく共通部分も出力される．つまりファイルのマージが実現される．しかし ifdef という機械的な記号が含まれてくるため人間の目で認識しづらい面もある．そこで差分部分の始まりに▼，二つのデータの境界に●，差分部分の終わりに▲を用いて出力結果を表すことにする．実際の形式としては，

▼▼▼▼▼
(一つめのファイルにだけある部分)
●●●
(二つめのファイルにだけある部分)
▲▲▲▲▲

となる．これをマージを行う diff として mdiff と呼ぶ．そこで先程のデータに対して mdiff を行うと以下のような結果になる．

今日
▼▼▼▼▼
学校へ
●●●
大学に
▲▲▲▲▲
行く

本研究ではこの mdiff コマンドを用いて差分の抽出を行い，学生論文から不適切な表現の収集を行う．

3.2 実験

3.2.1 データ

指導担当教員による修正の行われた 5 人分の学生論文で差分の抽出を行い，そこからさらに適切に差分が抽出されているかを確認するために，差分箇所を含む一文を出力して人手で確認を行った．なお本研究で用いる学生論文データは 2011 年度の言語処理学会年次大会論文である。

3.2.2 手順

図 3.2 は差分抽出から分類までの一連の流れである．

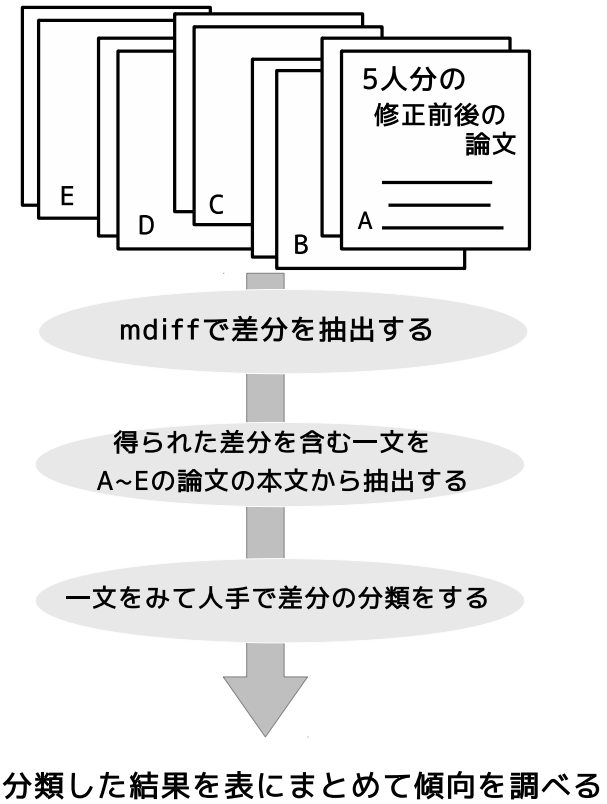

図 3.2: 差分抽出から分類までの流れ

図 3.2 の流れに沿った分類までの具体的な手順を以下に示す.

1. 修正前の学生論文と修正後の学生論文に対して mdiff コマンドを用いて差分を抽出する.
2. 1 で得られた差分箇所を含む一文を原文から抽出する.(例 1, 例 2 に抽出した例を掲載)
3. 差分箇所の前後の共通部分 (例 1 の共通部分 (前)(後)) の文字数を求める.
4. 前後の共通部分の文字数の小さい方の値を調べ, この値の降順にソートする.(共通部分が短いものは断片的に一致しただけであり適切な差分でない場合が多い. そのため前後の共通部分がある程度の長さをもっているものが有用な差分と考える.)
5. ソートの上位のものから順に人手で分類を行う.

例 1

修正前文：どちら が 空白に入れるべきかを推定する
修正後文：どちら を 空白に入れるべきかを推定する
差分部分 が(→を)
共通部分 (前)：どちら
共通部分 (後)：空白に入れるべきかを推定する
共通部分 (前) の文字数：3 文字
共通部分 (後) の文字数：14 文字
長さ:3

例 2

修正前文：結果を さらによく する方法として次の方法が考えられる
修正後文：結果を 改善 する方法として次の方法が考えられる
差分部分 さらによく(→改善)
共通部分 (前)：結果を
共通部分 (後)：する方法として次の方法が考えられる
共通部分 (前) の文字数：3 文字
共通部分 (後) の文字数：17 文字
長さ:3

3.2.3 分類の項目

分類に利用する項目の設定には古本ら [1] の「誤りおよび不適切表現の分類」を参考した．分類には大きく 4 つの項目に分け，さらにそこから詳細な項目へと分類を行った．以下に本実験で使用した分類の項目を示す．

ーーーーーー 4 つの大項目ーーーーーー

1. 表記の修正
2. 語彙表現の修正
3. 文法誤りの修正
4. 文体の修正

ーーーーーー詳細な項目ーーーーーー

(a) 表記の統一
(b) 専門用語の統一
(c) 冗長性
(d) 情報補完・詳細化
(e) 大雑把・安全な表現へ
(f) 適切な表現へ
(g) 助詞・接続詞
(h) 係る語との対応
(i) 時制
(j) 口語
(k) 硬い表現の軟化

実際に抽出した差分箇所から，設定した分類項目ごとの例文をいくつか掲載する．例文の見方はアンダーラインを引いている部分が修正前の表現であり，括弧の中の表現が修正が行われた後の表現である．Φは空表現である．

1. 表記の修正

(a) 表記の統一 (漢字・カナ・ひらがな)

例文：『余分な漢字表現を含む言い回しは，冗長で 分かり（→わかり）にくい』

解説：同一論文の中でひらがなの「わかる」を用いているので，ひらがなで統一している．

(b) 専門用語の統一

例文：『教師あり機械学習 手法で(→に) は性能の優れたサポートベクトルマシンを利用する』

解説：「機械学習」か「機械学習手法」どちらかの表現で統一させている．助詞の修正もあり。

2. 語彙表現の修正

(a) 冗長性（文を短く書き換えても大きな意味変化をもたらさないような修正）

例文：『機械学習 を行った場合(→では) あまりよい結果は得られなかった』

解説：文自体に大きな意味の変化をもたらさずに，より短い表現へと修正している。

例文：『要約前の文章から得られる情報を用いて文の順序推定を行う 手法(→の) が主な手法である』

解説：同じ単語や文が二回以上用いられて冗長なため修正.（例文の場合「手法」が二回用いられているため.）

(b) 情報補完・詳細化（読者に意味が伝わりづらい表現に言葉を補いわかりやすくなるようにする）

例文：『また，Φ(→副助詞「は」と格助詞「が」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情報を獲得する』

解説：どのようなデータを分析したのかを明確にするために情報を補完した.

例文：『適合率では 優るものの(→ベースラインより高かったが，) F 値ではベースラインより低かった』

解説：何に優るのかが書かれておらず，わかりにくい表現になっている．

(c) 大雑把・論文として安全な表現へ

例文：『対象語列の出現頻度 と照合(→を利用) して誤り表現の検出を行う』

解説：英語で use の意味をもつ「利用」という語を用いて違和感のない表現に修正している．

例文：『素性を拡充することでさらに性能向上 が期待できる(→を目指したいと考えている)』

解説：論文として指摘を受けにくいような安全な表現へと修正している．

(d) 適切な単語・表現へ (適切な単語や論文の内容に沿った語への書き換えなど)

例文：『素性を拡充することで より良い精度(→さらに性能) 向上を目指したいと考えている』

解説：向上と言う語に係る語として，良い精度向上は日本語としておかしいため修正している．

3. 文法誤りの修正

(a) 助詞・接続詞の修正

例文：『どちら が(→を) 空白に入れるべきかを推定する』

解説：「入れる」に係る語として「が」は不適切なため修正している．

(b) 係る語との対応

例文：『機械学習を用い Φ（→た) 格助詞「が」，副助詞「は」の分類を初めて行った』

解説：「用い」の係り先として『行った』ではなく「分類」に係ってほしいので「用いた」に修正している．

(c) 時制の修正

例文：『ヒューリスティックルールに加え教師あり機械学習法を利用することで性能の改善が可能であることが わかる(→わかった)』

解説：実験などを行った際の結果なので過去形で表している．

4. 文体の修正

(a) 口語の修正

例文：『結果を さらによく(→改善) する方法として次の方法が考えられる』

解説：口語を論文らしい表現に修正している．

(b) 硬い表現の軟化

例文：『近年，パソコンやインターネットの普及により，計算機を使って文字を入力する機会が 増し(→増え) ている』

解説：硬い印象を与える語を柔らかい表現へと修正している．

3.2.4 分類の結果と考察

3.2.2 節の手順に基づき，差分抽出，分類を行った．抽出した差分の上位約 650 個を人手で確認し，分析に有用な差分を 258 個獲得した．この 258 個の差分の分類を行った。分類した不適切な表現の結果を表 3.1 に示す.

表 3.1: 分類の結果

修正項目＼人物	A	B	C	D	E	合計
(表記の修正)						
用語の統一	1					1
その他の表記の統一		1		1	2	4
小計	1	1	0	1	2	5
(語彙表現の修正)						
冗長性	4	5		3	20	32
情報補完・詳細化	9	13	7	13	34	76
大雑把・安全な表現へ	1	5		2		8
適切な表現へ	9	16	5	8	44	82
小計	23	39	12	26	98	198
(文法誤りの修正)						
助詞・接続詞	6	6	3	3	20	38
係る語との対応			2			2
時制	2	1		1	3	7
小計	8	7	5	4	23	47
(文体の修正)						
口語	1			1	3	5
硬い表現の軟化				2	1	3
小計	1	0	0	3	4	8
合計	33	47	17	34	127	258

表 3.1 より，大きな項目の分類で見ると『語彙表現の修正』が圧倒的に多く，次いで『文法誤りの修正』が多いことがわかった．さらに細かい分類項目で見てみると『適切な表現への修正』,『情報補完・詳細化の修正』,『助詞・接続詞の修正』,『冗長性の修正』が多くみられた.『適切な表現への修正』が多かった原因としては，学生のような若いものは基本的な知識不足の面から言葉の表現をあまり知らないということに加えて，論文自体を書き慣れていないということもあり，読み手に伝わりづらい表現を多用しがちに

なってしまうということが考えられる．

『情報補完・詳細化の修正』が多かった原因としては，学生は文そのものを書く機会をあまり経験してこなかったために論文を書くことに慣れていないため情報の欠落した，読み手に対して不親切な文章を書きがちになってしまうということが考えられる．

また『冗長性の修正』は都藤ら [7] の研究でも行われている通り，冗長な表現は論文などに用いるのはふさわしくない表現であり，修正を行う必要がある箇所である．この冗長な表現も差分抽出の結果より，文章を書き慣れていない学生は多用しがちな傾向があるように思われた.『助詞・接続詞の修正』も学生の国語力や知識不足などの問題から不適切な使用やうっかりミスのような用いられ方がされていた．この分析結果から項目別にみた学生の用いやすい不適切な表現の傾向を知ることができた．不適切な表現としてこれらの要因を検出できるようなれば，論文作成支援への貢献が期待できる．

第4章　差分の頻度による分析

前章では差分箇所の抽出を行い，多くの不適切と思われる表現を得ることができた．これより抽出された差分から特に偏りや傾向のみられる表現があるのではないかと考えた．そこで抽出した差分箇所から，頻度に着目して分析を行う．

4.1　差分の頻度分析

4.1.1　頻度の集計

3 章で抽出した差分の頻度の集計を行った．ここでは抽出した差分をそのまま集計し、頻度を数えた．

その結果頻度 2 以上となった差分を表 4.1 に示す．表中の「修正前」とは，学生が論文で用いていた表現であり,「修正後」とは，教員によって書き換えられた修正前表現である．

表中のΦは空表現を意味し，修正前の文に矢印の後の文を追加した箇所に相当する．例えば表 4.1 の番号 1 の修正前と後の表現は以下のとおりである．

例 3

修正前文：要約前文章から得られる情報
修正後文：要約前 の 文章から得られる情報
差分部分 Φ(→の)

例 4

修正前文：簡潔な 文へと修正を行う
修正後文：冗長でない 文へと修正を行う
差分部分 簡潔な(→冗長でない)

抽出した差分の頻度をそのまま集計した結果を以下に示す.

表 4.1: 頻度 2 以上の差分

番号	修正前		修正後	頻度
1	Φ	→	の	11
2	簡潔な	→	冗長でない	7
3	Φ	→	は	5
4	単語	→	自立語	5
5	前後	→	文の順序を	4
6	した	→	する	4
7	のとっている	→	につく	3
8	もの	→	表現	3
9	Φ	→	法	2
10	している	→	する	2
11	Φ	→	で	2
12	付く	→	存在する	2
13	付く	→	出現する	2
14	や	→	と	2
15	に	→	で	2
16	Φ	→	すべての	2
17	Φ	→	本研究の	2

4.1.2 考察

表 4.1 より，表現の挿入が多くなされていることがわかった．そのなかでも特に助詞の挿入が多く行われていることがわかった．しかし抽出した差分を，そのままの形で頻度の集計を行っても，これ以上の発見はみられず，あまり特徴的な偏りや傾向は見られず，抽出した差分の 9 割以上が頻度 1 となる抽出結果であった．そのため次節では，抽出した差分箇所を単語単位に分解してから頻度を用いた分析を行う.

4.2　差分に含まれる単語単位での頻度分析

抽出した差分をそのままの形で頻度を数えると，頻度 2 以上のものは 17 箇所しか見つからなかった．よって，3 章で抽出した差分の表現から単語単位で表現を取り出し頻度の集計を行い，修正前に用いられていた不適切な表現と修正後に書き換えられた表現の出現傾向を調べる．

ここでは実験を行うにあたって抽出した差分を 1 単語，2 単語連続，3 単語連続で分解を行ってみた．1 単語，2 単語連続，3 単語連続の結果の例を以下に示す．

例 5

抽出した原文が『機械学習を用いて行っている』という文だった場合 … 1 単語

機械
学習
を
用い
て
行っ
て
いる

例 6

抽出した原文が『機械学習を用いて行っている』という文だった場合 … 2 単語

機械 学習
学習 を
を 用い
用い て
て 行っ
行っ て
て いる

例 7

抽出した原文が『機械学習を用いて行っている』という文だった場合 … 3 単語

機械 学習 を
学習 を 用い
を 用い て
用い て 行っ
て 行っ て
行っ て いる

このように分解して，単語単位での頻度を集計する．

その結果，1 単語だと抽出される量は多いが，そこから得られる情報量が少なく特徴が発見しづらいため分析に用いるにはあまり向かないものと思われる．3 単語連続だと抽出される量が少なくデータ不足であり，頻度自体が 1 となるものが多くなってしまい特徴も発見しづらいためこれも分析に用いるには不向きと判断した．

そこで 2 単語連続では，抽出される表現の量も多く，頻度もばらけており特徴のある表現を発見しやすいと考えられたため，4.2 節の分析では 2 単語連続を用いて行うこととした．

なおここで取り出す 2 単語連続は 5 人中 2 人以上の論文に出現している表現に限定して行った．

4.2.1 手順

図 4.1 は 2 単語連続の頻度を集計するまでの一連の流れである.

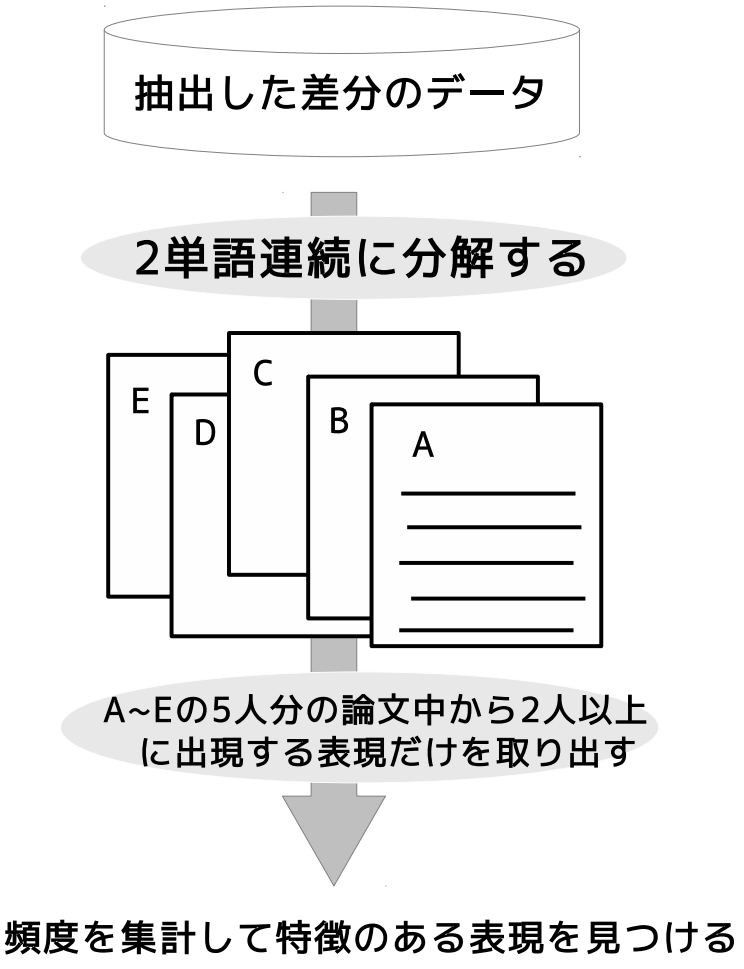

図 4.1: 頻度集計までの流れ

2 単語連続の頻度分析の具体的な手順を以下に示す.

1. 差分抽出によって得られた差分表現を利用する.
2. 1 の差分表現の修正前表現と修正後表現に対して形態素解析 (ChaSen[8]) を行い，それらを品詞単位の単語に分解する.
3. 2 で分解された単語を 2 単語連続にして取り出す.
4. さらにそこから 5 人中 2 人以上の論文に出現している 2 単語連続のみを抽出する.
5. 修正前表現で得られた 2 単語連続の頻度を a とし，修正後表現で得られた 2 単語連続の頻度を b とする．a/(a+b) という式を利用して 0〜1 までの数値で得る.
6. 上記の式で得られた値が 1 に近いものが，より修正後表現で利用される可能性が高いと考えられる．逆に 0 に近いものは修正前表現で利用される可能性が高いと考えられる。

4.2.2　2 単語連続の抽出例

図 4.2 は実際に例文を用いて 2 単語連続の抽出を行った例である．

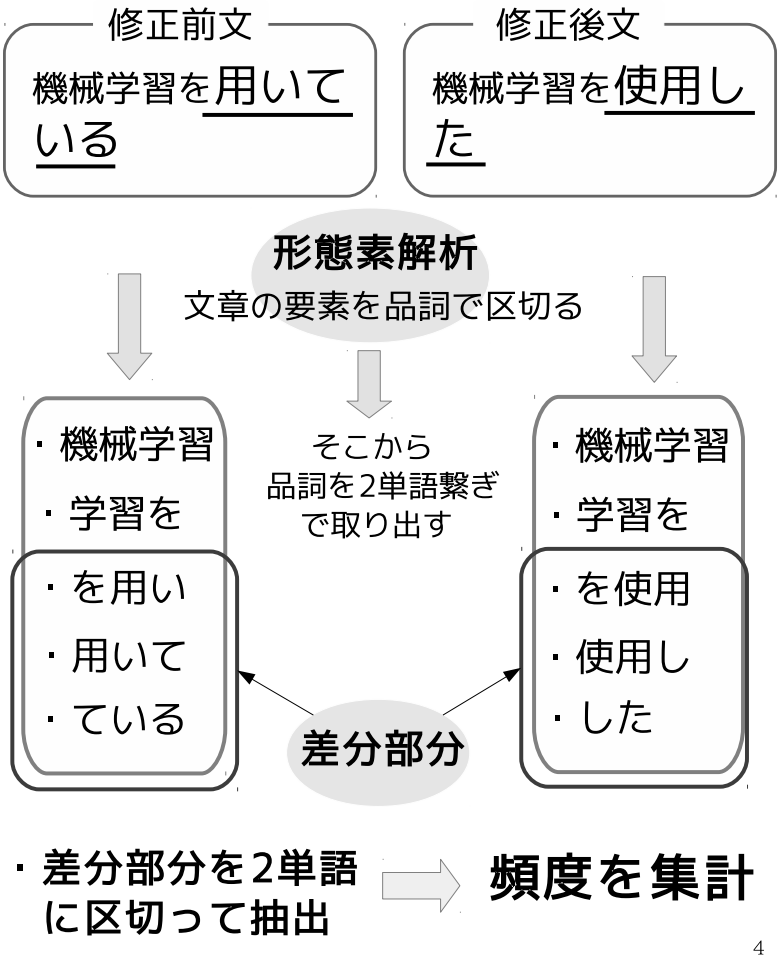

図 4.2: 2 単語連続の抽出例

4.2.3 結果

分析の結果，数値が 0 または 1 に近く，なおかつ特徴的な傾向が見受けられたものを表 4.2，表 4.3，表 4.4 に示す．表では修正前表現での頻度 (修正前頻度) と修正後表現での頻度 (修正後頻度) を表示している．

表 4.2: 修正後に数が減った 2 単語連続

2 単語	修正前頻度		修正後頻度	数値
ている	38	→	7	0.84
用いて	21	→	5	0.81
を行う	10	→	3	0.77
とし	16	→	5	0.76

表 4.3: 修正後に数が増えた 2 単語連続

2 単語	修正前頻度		修正後頻度	数値
ように	4	→	11	0.27
である	14	→	22	0.39

表 4.4: 修正前または修正後にしか出現しなかった 2 単語連続

2 単語	修正前頻度		修正後頻度	数値
され	16	→	0	1.00
行って	8	→	0	1.00
を利用	0	→	9	1.00
それを	0	→	8	1.00

4.2.4 考察

表 4.2 より「〜ている」という表現が修正後には大幅に数を減らしていることがわかった．この原因としては,「〜ている」という表現は執筆者が論文を書くにあたって積極性に欠けた他人事のような文を書いている印象を読者に与えてしまうため，修正がなされたと考えられる．実際に抽出した文からは「用いている」といった表現を「用いた」などに書き換えられている場合が多く見受けられた．2 番目の「とし」が修正されているものとして,「結果としては」という表現が「結果は」のように書き換えられている場合が多く見受けられた．これは冗長性による書き換えに該当し，表 3.1 でも冗長性が多く修正されていることがわかるため，これはあまり論文に用いるにはふさわしくない表現の可能性がある.「〜を行う」も同様に冗長性の問題から「〜する」に修正されている場合が多く存在した.

実際の修正の例文を示す.

〜ている

- 修正前：本実験は機械学習を 用いて行っている
- 修正後：本実験は機械学習を 使用した

〜として

- 修正前；冗長な表現になりやすいもの として 考えられる
- 修正後：冗長な表現になりやすいもの であると 考えられる

表 4.3 から読み取れた特徴としては「ように」は，文の情報を明確にするために書き足された表現ということがある．実際に「ように」が用いられた例として「読みやすく修正した」という文を「読みやすくなるように修正した」というように書き換えが行われている.「である」という表現は，表 4.2 の考察の例文にも示している他に，以下のような修正がみられ，論文らしい表現へとするために修正したと考えられる．

実際の修正の例文を示す.

〜ように

- 修正前：冗長な文は読みにくいため，読みやすく _ 修正する
- 修正後：冗長な文は読みにくいため，読みやすく なるように 修正する

〜である

- 修正前：不適切 な 可能性が高い
- 修正後：不適切 である 可能性が高い

表 4.4 の考察として,「され」という表現は「する」「した」「でき」という表現に修正されている場合が多く見受けられた．これは「〜ている」の考察と同様に，受け身文ということで執筆者自身に積極性が欠ける表現なために修正が多くなされたと考えられる.「行って」は表 4.2 の「を行う」と同様に冗長性の問題から修正が行われていた.「を利用」という表現は,「使う」や「用いる」といった表現を書き換えている場合が多く見受けられた．これは論文修正者の指導担当教員の好みの問題と考えられるが，論文として安全な表現への修正とも考えられる.「それを」の修正は，文にあえて「それを」を用いなくても意味は通じるが，直前の文からの情報が欠落してしまうので，不親切な文になってしまう．そのため，文の情報補完，詳細化の点から修正がなされていると考えられる.

実際の修正の例文を示す.

〜され

- 修正前：提案手法により有用性 が確認された
- 修正後：提案手法により有用性 を確認した

〜それを

- 修正前：データベースに含まれる冗長な文と，_修正した文を比較し
- 修正後：データベースに含まれる冗長な文と，それを 修正した文を比較し

第5章　おわりに

本研究では差分を用いた学生論文の不適切な表現の収集と分析を行った．その結果『語彙表現の修正』『文法誤りの修正』が多く行われていることがわかった．詳細な項目で見てみると,『適切な表現への修正』,『情報補完・詳細化の修正』,『助詞の修正』が多くみられた．また，差分に含まれる高頻度の 2 単語連続を分析することで，特徴のある傾向を発見した．例えば,「〜ている」という表現が「〜た」,「〜した」といった表現に書き換えられている場合が多くあることを発見した．これは「〜ている」といった表現が，積極性に欠けた他人事な印象を読者に与えてしまう表現なため修正されていると考えられる．これらの結果を得る過程で，差分の抽出や 2 単語連続の抽出に工学的な手法を用いたことから将来的に不適切な表現の検出，修正を行うシステムが開発される際の自動化の支援になれば幸いである．今後は本研究で得られた知見を活かして，学生論文の作成支援に役立つ技術の開発をしたいと考えている．

謝辞

本研究を進めるに当たり，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，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．また，本研究を進めるにあたり，御指導，御助言を頂きました，村上仁一准教授，徳久雅人講師に心から御礼申し上げます．ここ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] 古本裕子, 苗田敏美, 八重澤美知子, 川西琢也. 工学を専門とする日本人学生が書いた文章に見られる基礎的な問題点. 専門日本語教育研究, Vol. 7, pp. 47–52, 2005.

[2] 村田真樹, 井佐原均. 自動言い換え技術を利用した三つの英語学習支援システム. 情報科学レターズ, Vol. 3, pp. 85–88, 2004.

[3] 阿辺川武, 影浦峡. 下訳と修正訳を用いた訳文修正パターンの発見. 言語処理学会第13回年次大会発表論文集, pp. 919–922, 2007.

[4] 藤田篤, 乾健太郎, 松本裕治. 自動生成された言い換え文における不適格な動詞格構造の検出. 言語処理学会誌, Vol. 45, No. 4, pp. 1176–1187, 2004.

[5] 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, Vol. 2012-NL-206, No. 15, pp. 1–10, 2012.

[6] 村田真樹. diff を用いた言語処理-便利な差分検出ツール mdiff の利用-. 言語処理学会誌, Vol. 9, No. 2, pp. 91–110, 2002.

[7] 都藤俊輔, 村田真樹, 徳久雅人, 馬青. 冗長な文の機械的分析と機械的検出. 言語処理学会第18回年次大会発表論文集, pp. 1114–1117, 2002.

[8] ChaSen:. http://chasen.naist.jp/hiki/ChaSen/.