

# 作文指導に向けた表現データベースの構築と 表現評価指標の検討

安藤 秀明      秋吉 信吾      竹内 和広  
大阪電気通信大学 工学研究科 情報工学専攻

高校・大学生の作文の表現指導は評価者の主観が強く影響するため、一定の基準で作文指導を行う上では指導者間で共通する評価指標を設定することが課題となる。我々は現在、標準的な表現で記述された文章データに対して書き換え処理を行い、誤り表現の評価基準を設定するための誤り表現データベースの構築を試みている。今回はそのようなデータベースを用いて高校・大学生の作文表現を客観的に評価する数値指標の比較、検討を報告する。

キーワード：教育支援情報処理，作文自動評価，学習者コーパス

## A Comparison of Statistical Error Detection Methods for Objective Evaluation of Students' Composition

Hideaki Ando      Shingo Akiyoshi      Kazuhiro Takeuchi  
Graduate School of Engineering  
Osaka Electro-Communication University

In this paper, we propose a set of pseudo-error rewriting rules based on an analysis on students' errors in their compositions and introduce a pseudo-error database that is created from regular texts by applying the rules. In order to establish an objective evaluation method which adopts various viewpoints for students' composition, we compare some statistical error detection methods on the database.

Keywords: Computer Aided Education, Automatic Error Detection, Learners' Corpora

### 1. はじめに

人間が書く作文には本質的に多様性がある。そのような多様性は文学作品に作家特有の文体が存在するように、必ずしも悪いものばかりではない。しかし、高校や大学での作文教育の場面では、作文の不適切な部分を指導する基準があいまいであると問題を招くこともある。そこで、学生が書いた作文の不適切さを多様な観点から基準設定することが求められる。しかし、不適切な表現の基準を策定するためには、不適切な事例を大量に収集、蓄積することが最良であるが、そのようなデータ収集は簡単には実現できないことが問題となる。本稿ではこの問題に対して、適切な表現を使っていると考えられる標準的な文章を書き換えて疑似的な不適切表現として利用し、多様な作文評価を可能とする評価指標設計の資源として利用することを検討したい。

### 2. 不適切表現のデータベース化

我々は、既に、作文指導のために学生の作文を教員が添削したデータを収集し、データベ

ス化を行っている[1]。このデータベース化作業を進めていく過程で、学生が作文中に用いる表現の多様性に改めて驚くと同時に、教員が特定の表現を不適切表現ととらえる観点も多様であることに気付いた。このような教員の経験則である観点はそれぞれに妥当性があり、どの観点が優れているとは言い難い。例えば、ある表現が、文体の統一性から見て不適切なのか、表現が含まれる文そのものを見た場合でも不適切なのかは観点の相違であり、その観点は指導の目的によって変化する性質をもつはずである。前者の文体や論理構造に関わる問題は文章レベルの特徴付けが必要であり、その観点に関しても既に我々は検討を進めているが[2]、本稿では後者の文内のレベルの不適切表現をデータベース化することを目的とした研究を報告する。

文脈を文内に限定した表現レベルでも、特定表現を不適切と見る観点には多様性がある。また、このレベルの不適切表現を整理する上での難しさは、不適切さの理由はある程度説明することができるにも関わらず、それが文内レベルのどのような文脈においても成立するわけではなく、不適切となる文脈を限定することが難しい点である。

不適切表現を整理・限定するための文中文脈を統語解析といった言語解析モジュールの解析結果に基づいて行うためには、もう一つの問題が存在する。不適切表現を含んだ学生の作文では、その不適切表現が問題となって係り受け解析や形態素解析が精度よく機能することが期待できないという問題である。そこで、自然言語処理モジュールを使わず、基本的な言語統計量を用いて不適切表現の成立する文脈を特定する方法を検討したい。具体的には、主にOCRの分野で開発されてきた手法を参考に、作文の文を単位に不適切性を議論したい。

### 3. 疑似的な不適切表現の生成

学生の作文を添削したデータベースを参考に、学生が不適切な表現を行う理由を以下の典型的な5種類に大別し、疑似的な不適切表現のデータベースを作成した。

DB1: 1文字単位での誤字脱字

DB2: 格助詞の不適切な使用

DB3: 連体修飾をとともう不適切な名詞句の形成

DB4: 名詞の表記が不適切なもの

DB5: 文節要素の順が不適切なもの

この5つの誤り類型に対して具体的には表1のような11の疑似的な不適切表現への書き換え規則を作成し、Wikipedia<sup>1</sup>のデータに対して表現書き換え規則を適用することで表中の各行ごとに1000文からなる疑似的な不適切表現データベースを作成した。すなわち11ファイルから成る不適切表現データベースが完成した。なお、書き換え処理規則に自然言語処理モジュールの解析結果を利用するが、利用したモジュールは形態素解析器ChaSen[5]と係り受け解析器CaboCha[6]である。

表1のそれぞれのファイルについて、大別類型のDB1から順番に詳細を説明する。

DB1では正データの文字列に対して、文中の適当な文字に対して下記の3種類の書き換え規則を設計した。なお、それぞれの規則は、1文中に1文字程度を書き換えるように設定している。

- ひらがな文字を挿入する (c\_ins)
- 文字を削除する (c\_del)
- 文字をひらがな文字に変換する (c\_henkan)

DB2では書き換え対象文にまず形態素解析を

かけ、その結果に対して、文中の品詞情報が助詞-格助詞-一般となる語に以下の2種類の書き換え規則を用意した。

- 格助詞をランダムに変換する (case\_case)
- 格助詞を削除する (case\_del)

各規則では、1文中の1つの格助詞がランダムに選ばれて書き換えられるように設定している。

表1.疑似不適切表現データベースの内容

	DB	略称	不適切の種類
1	DB1	c_ins	文字挿入
2	DB1	c_del	文字削除
3	DB1	c_henk	文字変換
4	DB2	case_case	格助詞変換
5	DB2	case_del	格助詞削除
6	DB3	rentai_col	連体化部分変換
7	DB3	rentai_del	連体化部分削除
8	DB4	noun_hira	ひらがなに変換
9	DB5	rule_a	格要素の位置交換
10	DB5	rule_b	副詞的文節後方移動
11	DB5	rule_c	副詞的文節前方移動

DB3は書き換え対象文の形態素解析の結果から、連体化が名詞節を同定することを前提とした疑似的な不適切表現生成規則を以下のように作成した。

- 連体修飾節の接続を別の形にランダムに変換する (rentai\_col)
- 連体修飾節中の助詞を1つ削除する (rentai\_del)

以上もDB2と同様に連体修飾節がある文中で1箇所のみ書き換え規則が適用されるようになっている。

DB4では正データの形態素解析で同定された形態素の品詞情報が名詞-一般かつ形態素の字種がカタカナもしくは漢字のみで構成されている場合、形態素の読みをひらがなとして変換する (noun\_hira) という書き換え規則を用意した。

DB5の作成には、係り受け解析で得られた情報を用いた疑似不適切表現の書き換え規則を用いる。具体的には下記の3つの種類がある。

- 同じ係り受け先をもつ格要素を入れ替える (rule\_a, 図1中のa)
- 副詞で終わっている副詞的な節を1つ後の文節と入れ替える (rule\_b, 図1中のb)

<sup>1</sup> <http://ja.wikipedia.org/>

- 副詞で終わっている副詞的な節を1つ前の文節と入れ替える (rule\_c, 図1中のb)

これらの規則も適用されるのは1文中、1箇所限定するように調整した。

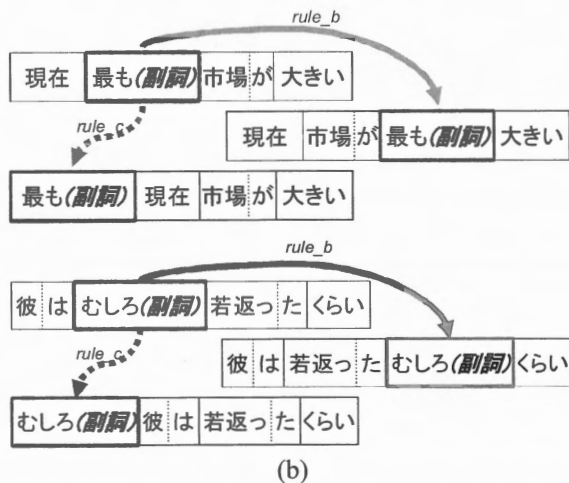
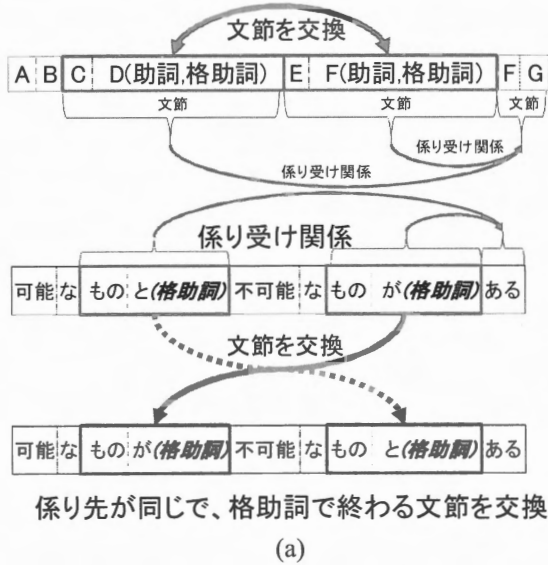


図1.係り受け情報を用いた不適切表現への書き換え規則

#### 4. 不適切表現の検出指標

本稿では、適切な表現が使われていることが想定される標準的な文書集合(正の例)を用意しておき、正の例から採集した統計情報を用いて前節で作成した疑似的な不適切表現のデータベースから本当に不適切表現となる例を検出することを検討する。

このように文中の表現の不適切性を評価する手法としては、図2に示したように文中の表現(灰色で示す)に対して基本的な2つの方向性がある。1つは文内に検査用の区切りを設け、

区切りの文頭側と文末側の文字列の結合性(つながりやすさ)を基準にする方向性である。他方は、文中の連続文字列が基準となる正の例においてどの程度存在するか(正の例での出現度合)を基準とする方向性である(図2中の手法2)。

手法1: 区切りの左右での文字列結合性で評価

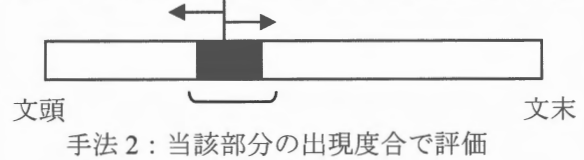


図2.文中の表現の不適切性を検査する方法

手法1の方向性に着目して表記誤りを検出する研究には、村田手法[3]がある。村田手法では、「表現xが実際の正の例には出現していない場合に表現xが負の例である度合P(x)を定義し、この値を使って不適切表現の検出を行う。具体的にはP(x)は式(1)によって求める。

$$P(x) = 1 - \left(1 - \left(\frac{O(a)}{N} \times \frac{O(b)}{N}\right)^N\right) \quad (1)$$

式(1)で用いる統計量はO(a), O(b), Nであるが、ここで、O(a)とO(b)は表現x中に分割点を用意し、分割点の文頭側の文字列a, 文末側の文字列をbとした場合の正の例中の出現数である。Nは正の例中に出現する全ての文字の出現数である。

実際の不適切表現の抽出には、作文中の文の文頭から文末まで1文字ずつ順に分割点を移動させ、分割点を基準に文頭側, 文末側の文字列の生起数を比較する。

それぞれの値の具体例を表2に示す([3]中の表を説明のために簡略化して引用)。この例は「負の事令を検出する。」という文の「事令」の間に分割点を置き、左右の文字列a, bをそれぞれ3文字から1文字まで考えた場合の統計値を示した表である。

表2. P(ab)を計算した実例

A	O(a)	B	O(b)	P(ab)
負の事	0	令の検	2	0
...		...		
の事	20837	令の	648	0.0324
...		...		
事	384718	令	16227	0.9999

表2は分割点を基準に文頭側文末側3文字以内のパターンを整理しているため、実際の表の行

数は 9 パターンにわたる。分割点を基準にどれだけの文字数を表現の単位と考えるかは指標の使用者側で決定する必要がある。例えば、分割点の前後 5 文字までを見るならば、25 パターンの統計値を計算することとなる。また、 $P(x)$ を認定する際に不適切表現である度合い  $P(x)$ にどのような閾値を用意すればよいかも決める必要がある。以上の村田手法は、特定の表現を二項関係(a,b)からなるものと仮定し、二項関係(a,b)の独立性を調べることを発展させた手法と考えることができる。本稿では村田論文を参考にして実装した手法を以後手法 1 と呼ぶ。

次に図 2 中の手法 2 の方向性の例として、荒木手法[4]について述べる。これは文中の連続文字列が正の例においてどれだけ出現するかを基準とする方法であるが、手法 1 と同様に 1 から無限長の長さの文字列についての出現検査をするわけではない。具体的には、 $m$  重マルコフモデルを使い、前方  $m$  文字に続く文字  $x_i$  の連鎖確率  $P(x_i|x_{i-m},x_{i-m+1},\dots,x_{i-1})$ を評価する。

一般的に文字  $x_i$  の連鎖確率は式(2)のように表され、文字  $x_i$  の連鎖確率  $P(x_i|x_{i-m},x_{i-m+1},\dots,x_{i-1})$ は、文字  $x_i$  を含む長さ  $m+1$  の文字列  $x_{i-m}x_{i-m+1}\dots x_{i-1}x_i$  の出現確率と文字  $x_i$  の前方の長さ  $m$  の文字列  $x_{i-m}x_{i-m+1}\dots x_{i-1}$  の出現確率の比として求められる。

$$P(x_i | x_{i-m}, x_{i-m+1}, \dots, x_{i-1}) \\ \equiv \frac{P(x_{i-m}, x_{i-m+1}, \dots, x_{i-1}, x_i)}{P(x_{i-m}, x_{i-m+1}, \dots, x_{i-1})} \quad (2)$$

実際にはこの  $m$  として大きな値を使うのではなく、 $m=2$  といった値を用いる。すなわち、図 3 のように検査対象文の文頭から特定文字数（図の場合は 3 文字、丸が 1 つの文字を示す）ずつ文字列をとり、その正の例での出現数を(2)の確率値として計算する。図 3 ではそれぞれの確率値を縦軸として示してある。

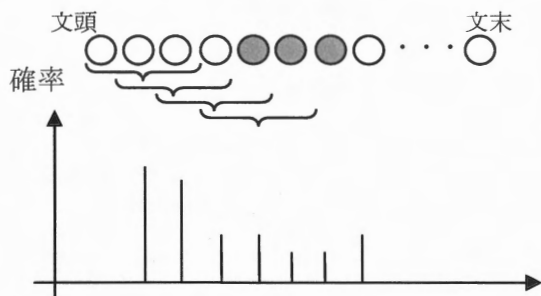


図 3. 文中の 3 文字単位ごとの確率値の計算

例えば図 3 中の不適切な表現が灰色の丸とすならば、その確率値は通常表現よりも低くなるのが考えられるので、この値が閾値より小

さい（落ち込む）部分を不適切表現と判定する手法である。

さらに荒木はこの連鎖確率の落ち込み回数と位置によって、

- (1) 誤りの開始位置が検出できる
- (2) (i)誤字誤りかまたは誤挿入誤り、(ii)脱落誤りの 2 種類が区別できる
- (3) また(i)の場合、誤り文字数が決定できるとしている。

本稿では荒木論文を参考に実装した手法を手法 2 と呼ぶ。

## 5. 実験

### 5.1 作文および未知の適切文に対する実験

前節で紹介した指標（手法 1,2）を使って、擬似的な不適切表現から、確度高い不適切表現を抽出・蓄積するために、いくつかの文章集合に対し適応し、それぞれの誤り傾向をみた。対象とした文章集合は、以下の 3 つの文章集合である。

- 大学生が書いた不適切性の高い文（以下「作文」と呼ぶ、390 文）
- インターネット上の掲示板 2ちゃんねる<sup>1</sup>（以下「2ch」と呼ぶ、610 文）
- 青空文庫<sup>2</sup>より収集した文学作品（以下「青空文庫」と呼ぶ、1000 文）

上記の文集合はすべて正の例には含まれない文から成り、かつ、作文、2ch は青空文庫よりも不適切な表現が多いことが期待できる。このことを利用して、手法 1,2 を使い、下記の適切率を用いてそれぞれの文集合の特質を区別できるかを実験した。

$$\text{適切率} = \frac{\text{指標が適切と判定した文数}}{\text{対象とした文数}}$$

まず、作文と 2ch の検出率を比較したグラフを図 4 に示す。次に、学生の作文と 2ch の書き込みを一つの文集合として、青空文庫との不適切率を比較し、その結果を図 5 として示す。

<sup>1</sup> <http://www.2ch.net/>

<sup>2</sup> <http://www.aozora.gr.jp/>

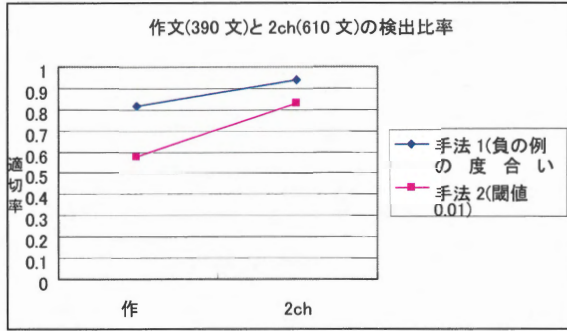


図 4. 作文(390 文)と 2ch(610 文)の比較

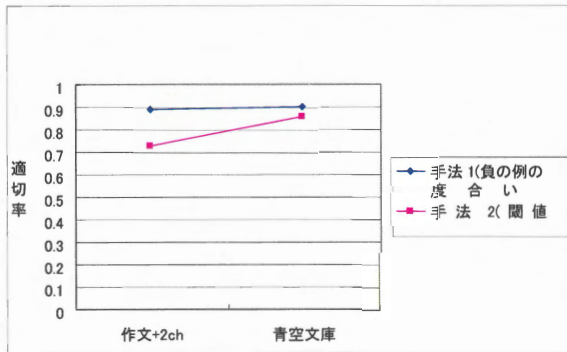


図 5.(作文+2ch)と(青空文庫)の比較

以上の図 4, 図 5 から, 手法 2 によって, 特に文集合の不適切な違いを分割することが可能となると考えられる. また, 我々が収集した学生の作文 360 文が 2ch の例と比較しても手法 1, 2 どどちらを用いても適切率が低いと判定された.

### 5.2 表 1 のデータベースを用いた実験

5.1 節で確認した手法 1, 2 の不適切表現検出の傾向をさらに調査する. 検討手法として 3 節で作成した各不適切表現ファイルに対する適切率をみることで検討を行う. 適切率は 3 節で作成した各ファイルの文数が各 1000 文であることから以下のように定義する. 5.1 節の文集合とは異なり, これらのファイルに含まれる文は表 1 の書き換え規則を用いて 1 文に必ず何らかの擬似的な不適切表現があるように書き換えられている.

$$\text{適切率} = \frac{\text{手法が適切とした文数}}{1000}$$

手法の閾値や比較文字列長を 5.1 節の実験と同様の設定で, 各ファイルに対する適切率を出したグラフが図 6 となる.

図 6 から不適切性の原因によって手法 1, 2 が不適切性を検出する度合いが変化することがわかる. すなわち, 特定の表現が不適切となる文

脈を区別する評価指標が設計可能ではないかと考える.

例えば, 5.1 節の実験結果である図 5 の作文+2ch と青空文庫に比較では, 手法 1 による適切率にはあまり差がなかったものの, 手法 2 では明確な差があった理由は, 作文+2ch の文集合に手法 2 が感度良く検出するタイプの不適切表現が多く含まれている可能性が考えられる.

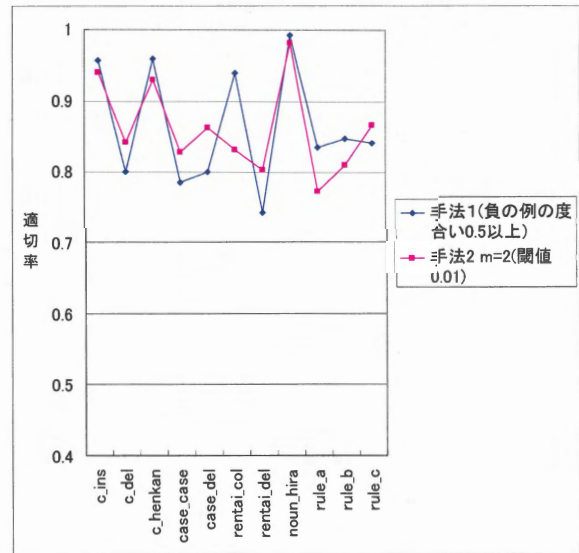


図 6. 擬似不適切表現への手法 1, 2 の適用

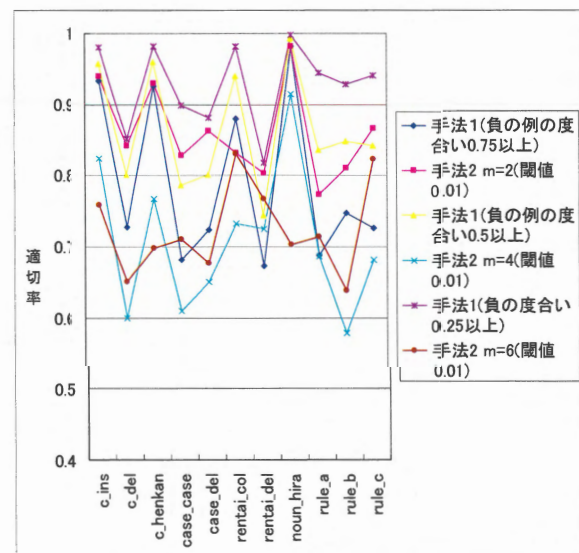


図 7. 手法 1 の負の度合いと手法 2 の m の次数を変えた状態の再現率

さらに, 手法 1 と手法 2 の設定を変え, 細かく傾向をみる実験も行った. 主に手法 1 では不適切性を検出する際の不適切性と適切性を分ける閾値, 手法 2 では誤りを特定する上での走査文字列長が感度調整に役立つ. 実験結果を図 7

に示す。図 7 が示すように、これらの設定が、図 6 よりもさらに、指標間の不適切性検知の感度が調整可能であることが期待できる。

## 6. まとめ

本稿では、人間が書いた不適切な表現を大量に収集、蓄積する代わりに、適切な表現を使っていると考えられる標準的な文章を書き換えて疑似的な不適切表現を作成する研究を報告した。疑似的な不適切表現はそのままでは工学的応用には用いることができず、評価指標開発と組みになってはじめて機能する。本稿では、この点に関して既存の代表的な文表現適切性を評価する指標を用いて、このような目的に使用可能かを評価した。本稿の内容は、それら手法の疑似的な不適切表現に対する総合的な評価のみにとどまった。今後はさらに評価指標の精緻な分析をすすめると同時に、評価指標と人間の不適切性を検知する観点との相関を調査していきたい。

## 謝辞

本研究は科学研究費補助金（基盤研究(C)課題番号 21500113）の補助を受けて遂行されている研究の一部である。ここに記して謝意を示す。

## 参考文献

- [1] 竹内和広：作文指導のための作文添削データベースの構築, 第 12 回「人文科学とデータベース」シンポジウム予稿, pp.55-60, 2006.
- [2] 橋本喜代太 他：複数コーパスからの特徴表現抽出を利用した文体統一性の評価とその活用について, 言語処理学会第 14 回年次大会, pp.384-387, 2008.
- [3] Masaki Murata, Hitoshi Isahara : Automatic Detection of Mis-Spelled Japanese Expressions Using a New Method for Automatic Extraction of Negative Examples Based on Positive Examples, IEICE Transactions on Information and Systems, VOL.E85-D, No.9, pp.1416-1424, 2002.
- [4] 荒木哲郎,池原悟,塚原信幸,小松康則,田川崇史,橋本憲久：m 重マルコフ連鎖モデルを用いた日本語の誤字・脱落・誤挿入誤り文字列の検出と訂正法,信学論(D-II), vol.J83-D-II, no.6, pp.1516-1528, 2000.
- [5] 形態素解析器 ChaSen:  
<http://chasen-legacy.sourceforge.jp/>
- [6] 係り受け解析器 CaboCha:  
<http://chasen.org/~taku/software/cabocha/>