

# 学習指導要領に立脚した 児童作文自動点検システムの実現

研 究 者：藤田 彬

(横浜国立大学大学院環境情報学府博士課程後期学生)

# 研究成果論文

本報告書では、研究プロジェクト『学習指導要領に立脚した児童作文自動点検システムの実現』（以下、本研究）の活動内容について報告する。まず、二種の自動点検システムについてその開発過程と開発成果について述べる。次に、本研究に係わる対外発表の状況について述べる。最後に、本研究の今後の課題について述べる。

## 1. 研究活動の内容と成果について

### 1.1 詳細な言語的要素に及ぶ評価が可能な文章評価システムの開発とその成果

#### 1.1.1 開発目的

従来、作文や小論文などの記述回答型の評価方式では、詳細な言語的要素の特徴を踏まえた評価を実現することが困難とされてきた。全ての言語的要素<sup>1</sup>を参照しながら評価対象文章の総合評価を割り出すという作業に非常に労力がかかることが、その原因の一つとしてあげられる。一度観点（表記法、文法、構成等）毎に評価した後、それらの評点を何らかの規則にしたがってまとめた上で総合評価とするという方法も考えられる。しかしこの方法で多くの文章に対して評価を行う場合、多大な時間を要することになり、評価結果のフィードバックに即時性がなくなってしまう。

特に教育現場においては、限られた時間の中で生徒や受験者の文章を評価して返却する必要がある。また、評価者の人数も限られており、一人の評価者が評価を担当する文章の量も膨大なものになることが多い。そのため、着目する言語的要素と要素毎の基準が、担当する文章毎に変動しやすい。個々の評価者の評価基準が不安定である限り、予め評価作業の前に評価者間で基準を共有していたとしても、結果として教育効果にばらつきが生じたり、試験の公平性が損なわれたりすることが考えられる。

図1、表1、表2に同一の文章群に対する複数人の手による評価結果の例を示す。これは高校二年生による584編の小論文データ<sup>2</sup>とそれを四人の国語教育専門家が評価した結果のデータから、評点当たりの事例数とその分布や一致具合をまとめたものである。評価は特定の観点に沿ったものではなく総合的なもので、10段階の絶対評価により施される。評点は10点が最高点、1点が最低点である。 $\kappa$ 係数は、二人の評価者間の判断の一致度合を表す尺度である。Krippendorffの考察（Krippendorff 1980）によれば、あるデータ間の $\kappa$

---

1 ここでの言語的要素とは、評価の際に判断対象となる個々の項目（仮名遣いの正しさ、文法の正しさ、構成の良さなど）のことを指す。

2 宇佐美により収集されたデータ（宇佐美 2011）を研究利用目的で借用したものである。

係数が0.7未満の場合、両データの関連を示すことは困難であることが多いとされる。表2において評点の $\kappa$ 係数は全評価者間で0.7を下回っている。また、表1と図1より、評点の分布にも異なりが認められる。このことから、本例においては評価者の評価基準が共通のものであったとはいえない。

文章評価の自動化はこれらの評価の公平性を損なう要因となる問題の解消に役立つと考えられる。また、評価者が着目する言語的要素やその配分の定量的な提示を行うことで、正確かつ円滑な評価者間の基準統合が可能になると考えられる。

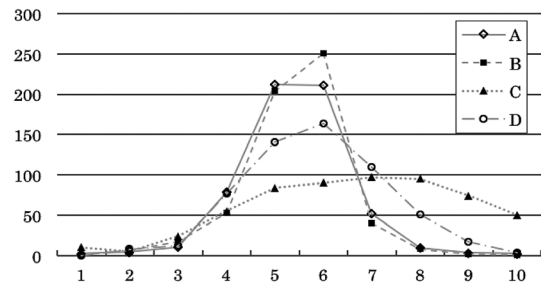


図1 各評価者の評点当たりの事例数

表1 各評価者の評点分布の平均と標準偏差

| 評価者  | A     | B     | C     | D     |
|------|-------|-------|-------|-------|
| 平均   | 5.420 | 5.438 | 6.639 | 5.856 |
| 標準偏差 | 1.081 | 1.037 | 2.092 | 1.436 |

表2 評価者間での評点の $\kappa$ 係数

|             | A - B間 | A - C間 | A - D間 | B - C間 | B - D間 | C - D間 |
|-------------|--------|--------|--------|--------|--------|--------|
| $\kappa$ 係数 | 0.110  | 0.035  | 0.095  | 0.054  | 0.115  | 0.081  |

そこで本研究では、単独の評価者により対象文章に与えられる総合的な評点と、多種の言語的要素についての特徴量<sup>3</sup>から、任意の作文課題設定における個人の評価者の文章評価モデルを自動的に推定する手法を開発した。また、その評価モデルにおける観点毎の配分を他の評価モデルと定量的に比較可能な形で自動的に提示する手法についても開発した。

文章の意味面での適切さの評価に関しては扱わない。ここでいう意味面での適切さとは、文章中の文が示す個々の内容の正しさを指す。例えば、「月は西から昇る」のような文が示す内容が正しいか正しくないかについての判断は、本研究では扱わない。

### 1.1.2 文章評価システムの仕組み

開発したシステムでは、Support Vector Regression (SVR) (Smola and Sch 1998) を用いて、人間の評価者が付けうる評点を予測する。

SVRは、機械学習手法の一種であるSVM (Support Vector Machine) の派生形である。SVMは、ある物事に関する複数の特徴量の集合から、数理的解析によりその物事の

3 言語的要素から測りとれる数値を示す。例えば、「漢字の使用程度」という言語的要素を扱うとすると、「全文字中において漢字が占める割合」が特徴量に該当する。

状態に関する一定のパターンを推し量り、状態が未知である物事の状態を二値で判定する手法の一般的名称である（同様に物事の状態を「実数値」で判定する手法をSVRと呼ぶ）。例えば「明日雨が降るか否か」をSVMにより判定するというアプローチを想定する場合、過去の気象データから降雨日の前日と非降雨日の前日の気温、湿度、風向等の特徴量をそれぞれ多数抽出する。これらのデータから「最も矛盾が少ない翌日の降雨判定ルール（モデル）」を数理計画問題として定式化して導き出し、翌日の天気を予測する<sup>4</sup>。

この判定ルールを人手により導き出す場合、最適なルールの候補が膨大な量になるか、もしくは一切ルールが導き出せない状況に陥り、最適なルールの判定は困難を極める。人工知能分野においてSVMは、このルールの最適化についての効率、正確性において、非常に優れた性能を発揮する手法として知られている。

本システムでは、過去に書かれた文章とそれに対するある一人の評価者による評点をSVRの教師データとする。教師データから表層や使用語彙、構文、文章構造などに関する種々の言語的要素の特徴量を算出し、各文章から抽出された特徴量群とその文章に与えられた評点との間の関係性をSVRにより導き出し、評価モデルとする。一度評価モデルが導出されると、未評価の文章にも、その文章から種々の特徴量を算出して評価モデルに代入することで、その点数を予測することができる。以下では、この評価モデルにおいて評価の根拠として扱われる言語的要素を**素性**、それらの言語的要素から算出される特徴量を**素性値**と呼ぶことにする。

SVRは、全素性について、**回帰係数**とよばれる「素性の重み（重要性）を表す数値」と素性値を乗じたものを足し合わせることで、ラベル（本システムの場合は評点）を推定する。この回帰係数の大きさを参照することで、各素性がラベル推定モデルに寄与する度合を知ることができる。これにより、教師データにラベル（評点）をつけた評価者が、各素性に対して「どの程度の配分で評価していたか」、また「加点要素としたか減点要素としたか」を定量化することができる。

扱う素性には、小学校から高等学校までの国語科教育において扱われる作文の良悪基準に関わるものが多く含まれる。評価者は、様々な観点から文章を捉えて評価を行った上で、その判断結果を評点として数値化するが、本研究ではこの過程を教育的立場からモデル化した。学校教育上考えられる様々な課題や試験を想定して、それらで必要となる観点を合わせて素性として扱い、評価対象文章から自動的にそれぞれの素性値を算出する。

素性値を算出するには、文章を単語レベル（形態素）に分けた上で、文の構造（文節間の係り受け関係）を明らかにする必要がある。本システムでは、文章の単語分割に形

---

4 このときの過去の気象データにあたるデータは、一般的に**教師データ**と呼ばれる。また、翌日が雨天であるか否かにあたる値は、教師データの**ラベル**と呼ばれる。

態素解析器『MeCab』を、文構造の解析に係り受け解析器『CaboCha』を用いた。また、SVRの学習器には、『SVM<sup>light</sup>』を用いた。

以下では、本システムにおいて素性として用いられる言語的要素をカテゴリ毎に列挙した上で、一部の素性についてその詳細を述べる。後掲表4に、全素性の素性名とIDをカテゴリ別に示す。

### ●カテゴリ「表層」

文字数や文数、字種などの表層的特徴に関する素性群である。主に文章の形式面の妥当性についての評価に役立つ。

FV2 (文字数制限の達成度)：文字数制限には、(I)「～字以上」、(II)「～字以内」、(III)「～字程度」の3種類が存在する。制限文字数を  $r$ 、評価対象文章の文字数を  $n$  としたとき、下記のように達成度  $d$  を算出することにする。

$$(I): d = 1 \quad (II): d = \frac{n}{r} \quad (III): d = \begin{cases} \frac{n}{r}, & n \leq r \\ \frac{2r-n}{r}, & n > r \end{cases}$$

FW7 (文の最大文字数)：句点で終わる一続きの文字列を文として扱い、文章中でのその最大文字数を素性値とする。文は長くなるほど、内部の係り受け関係に曖昧さを生じやすいとされている（森岡 1963）。

### ●カテゴリ「語」

単語（特に自立語）の用法や品詞、記法に関する素性群である。主に、筆者の語彙の豊かさや、語彙表現の成熟度合の評価に有効である。

FW2 (自立語の多様性)：自立語の異なり数（タイプ数）を延べ数（トークン数）で割った値とする。ただし用言については活用形を計数の考慮に入れず、異なり数、延べ数ともにその用言の原形を数える。

FW4 (オノマトペ使用率)：文章中でのオノマトペの使用数を総単語数で割った値を素性値とする。オノマトペは、運用する者の主観によって際限なく自作することができる。そこで、本素性において扱うオノマトペは小野の日本語オノマトペ辞典（小野 2007）に記載のあるものに限定することとする。オノマトペは、通常の熟語を用いて妥当に表現することが困難な概念を表現する際に用いられることが多い。しかしながら、記述課題の設定によっては、学習した熟語の運用可否を問うものもある。そのような場合、オノマトペの使用は減点対象となり得る。

### ●カテゴリ「文体」

文末の形式や文内で用いられる文体等に関する素性群である。主に、表現の工夫や作文の基本的な技法について、その習熟度の評価に役立つ。

FF4 (名詞述語文の出現率)：述語が“名詞句＋断定の助動詞（「だ」「である」「です」等）”で構成される文の出現数を、文の総数で割った値とする。名詞述語文は、筆者の意見や感想を述べる際に多く用いられる文体である。

FF7 (常体・敬体の混用有無)：文中の助動詞に着目し、文体が一貫している場合は0、混用が認められる場合は1を素性値とする。文章中の文体の一致は、現行の小学校学習指導要領（文部科学省 2008）において指導事項として定められている。

FF8 (文末の単調さ)：式  $-\log(n/N)$  により算出する。ただし、 $n$  は文の最終文節の表記異なり数、 $N$  は文の総数とする。文末の表現が単調な文章は、退屈な印象を与えると考えられる。筆者が作文に不慣れな場合、文体が多様性に乏しく、本素性の素性値が高くなる傾向にあると考えられる。

### ●カテゴリ「係り受け」

同一表層格の多用や文節間の修飾関係の複雑さを捉える素性群である。これらの素性は、主に係り受けの適切さの観点を反映させた素性である。

FD6 (格助詞「ガ」で終わる文節の文内最大出現数)：格助詞「ガ」を付属語に持つ文節が文中に出現する回数の最大値を素性値とする。文中で同じ助詞を繰り返し使用することで、文の表す内容が不明瞭になる場合がある（岩淵 1979）。

例：英語の早期教育がもたらす効果が測れないうちに実行に移すことが問題であることは、言うまでもない。

FD14、FD15 (修飾の詳細さ)：一文中における名詞文節の出現回数を用言の出現回数で割った値を全文について平均した値、またその分散を素性値とする。名詞文節が多いほど、主語や述語が詳しく表現される傾向にある。

FD16 (一文中での中止法の最大使用回数)：一文中で、用言の連用形や接続助詞等によって文が途中で中止される回数の最大値を素性値とする。中止法の多用は、係り受け関係を曖昧にする原因となりうる（岩淵 1979）。

### ●カテゴリ「文章のまとまり」

文章の一貫性と結束性に関する素性群である。主に、文間のつながりの良さや文章全体を通しての記述内容の首尾一貫性の評価に有効である。

一貫性は、概念や事象間の意味的なつながりの良さを指す。隣り合う二文間における一貫性（局所的ー貫性）と、文章全体での話題遷移の一貫性（大域的ー貫性）に区別できる。局所的ー貫性を表すモデルとして entity grid モデル<sup>5</sup>がある（Barzilay and Lapata 2008）。一方、結束性は表層的なつながりの良さを指す。横野らは、結束性に寄与する要

5 文を行、文章中の語を列、文における語の構文役割を成分とする行列を用いて、語の分布パターンを表現するモデル。行列から構文役割の遷移確率を成分とするベクトルを導出し、局所的ー貫性の評価等に用いる。



素<sup>6</sup>をentity grid モデルに組み込み、結束性と局所的一貫性を同時に捉えるモデルを提案している（横野、奥村 2010）。本システムでは、主に横野らのモデルに基づいた素性を用いる。

FC1～FC120：表3において entity grid で表される文章は、文1と2、

文3と4の間がそれぞれ論理的結合関係、文2と3の間が多角的連続関係という接続関係<sup>7</sup>で捉えられている。語1は文1でH、文2でS、文4でX、語2は文1・3でOという構文役割<sup>8</sup>で出現している。

このとき、FC1～36の素性値は以下のようにして求める。ある語の2文間での構文役割の遷移パターンは36通り（「HH」「HS」…「X-」「--」）ある。論理的結合関係で接続する2文間では、合計4カ所の遷移箇所のうち、「HS」「-X」の遷移が1回ずつ、「O-」が2回存在する。このときFC1～FC36の素性のうち、「HS」と「-X」の遷移確率を表す素性の素性値を0.25、「O-」の素性値は0.5とする。その他33素性の素性値は全て0とする。FC37～72の素性値も同様にして求める。「S-」と「-O」の素性値を0.5、その他34素性の素性値を全て0とする。拡充的合成関係の隣接文対は出現しないので、FC73～FC108の素性値を全て0とする。FC109～FC114、FC115～FC120はそれぞれ、文章の始端と終端にダミー要素（始端をB、終端をE）を持つとした時の遷移パターン「BH」…「B-」、「HE」…「-E」の出現確率をFC1～FC108と同様に算出した結果を素性値とする。

FC122（並列接続の有無）：並列的に展開して述べる構造が文章中に存在する場合に1、存在しない場合に0を素性値とする。並列的な接続を明示する特定の表現（人手で設定）の有無により、同定する。

例：…ばくは～に感動しました。一つ目は～です。二つ目は～です。…

### ●カテゴリ「モダリティ」

記述内容に対する筆者の主観的な態度（モダリティ）と判断の強さに関する素性群である。筆者による感想・意見の量的評価に有効である。

FM1～11：松吉らにより体系化された「拡張モダリティ」（松吉、江口、佐尾、村上、乾、

表3 結束性と局所的一貫性を同時に捉える entity grid の例

|    | 前文との接続関係 | 語1 | 語2 |
|----|----------|----|----|
| 文1 | -        | H  | O  |
| 文2 | 論理的結合関係  | S  | -  |
| 文3 | 多角的連続関係  | -  | O  |
| 文4 | 論理的結合関係  | X  | -  |

6 Halliday ら（Halliday and Hasan 1976）は、参照、接続、語彙的結束性、省略を挙げている。

7 接続関係には論理的結合関係（二つの事柄を論理的に結びつけて述べる関係）、多角的連続関係（二つ以上の事柄を別々に述べる関係）、拡充的合成関係（一つの事柄を別々に述べる関係）の3種があるとする。

8 語の文中での構文役割にはH（主題）、S（主語）、O（目的語）、R（述部要素）、X（その他）があるとする。ただし、語が当該文中に出現しない場合は構文役割を-（出現せず）と表記する。

松本 2010) に沿って、態度と真偽判断について文の特徴を捉え、素性として扱う。機能表現辞書つつじ(松吉、佐藤、宇津呂 2007) に助動詞型機能表現として収録される機能表現と、分類語彙表(国立国語研究所 2004) に「精神および行為／心」として収録される用言を手がかりにして、文の態度、真偽判断について分類を行う。これら拡張モダリティ体系に準拠する素性の素性値は全て、当該拡張モダリティカテゴリに分類される文が出現する回数を文の総数で割った値とする。

FM12 (文末思考知覚感覚動詞使用率)：最終文節で思考動詞、知覚・感覚動詞が用いられる文の出現回数を文の総数で割った値を素性値とする。思考動詞、知覚・感覚動詞であるか否かの判断は、分類語彙表を典拠とする。

### ●カテゴリ「内容」

本システムでは、文章の内容(筆者により書かれた行動、出来事、状態)について、その正しさを意味面で判断しない。その代わりに、作文課題(または論題)に対して適合した語彙が使用されていることを捉えるための素性を用意する。

論題に含まれる名詞と文章中の名詞が、EDR 電子化辞書(日本電子化辞書研究所 2010)において同一の概念識別子を持つ、もしくは所属概念が直接の上位または下位関係にある場合、作文課題に適合する語彙として判断する。このように判断される語彙が文章中の全名詞中で占める割合を素性(FS1)とする。

以上の7カテゴリに分けられる186の素性を用いて、評価対象文章の特徴をあらゆる観点から捉えた上で、前述の数理解析的手法により総合的な評価を求める。

#### 1.1.3 文章評価システムが利用する語彙資源の拡張

本システムは、EDR 電子化辞書を参照して評価対象文章中の単語の語義を推定し、カテゴリ「文章のまとめり」「内容」等の素性値抽出に利用する。しかしながら、作文課題の内容によっては、文章中に時事的な単語や一般的でない単語(特に名詞が連なり複合名詞化したもの)が使用されることが考えられる。その単語がEDR 電子化辞書に登録のない単語であった場合、語義を推定することは困難であり、素性値抽出の精度に影響を与えることが考えられる。このことから、広範な分野の最新語彙を随時反映する機能をもった語彙資源が必要となった。

そこで、本研究ではオンライン百科事典 Wikipedia に記載された語彙を抽出して、システムが参照する語彙資源の拡張を行った。Wikipedia には多岐に渡る分野の単語に関する記事が記載されており、それらは不特定多数の人物(編集者)により随時編集されている。また、定期的に全記事を含んだデータ(ダンプデータ)が公開されるため、最新語彙の抽出が自動的に行える。



大半の Wikipedia 記事の冒頭部分は、「見出し語（記事で説明される単語）」の後に「定義文（見出し語を簡単に説明する文）」が続くという形で始まる。定義文には、多くの場合見出し語がどのような語の下位語（または上位語）であるかが書かれている。このパターンを活用して、「見出し語」と「見出し語と直接の上位下位関係にある単語」を抽出することができる。本研究では、上位下位関係を is-a 関係と part-of 関係をまとめた関係とする。例えば下記のような二語が上位下位関係にある。

例：(上位) 運動会 ← (下位) 玉入れ、(上位) 雨 ← (下位) ゲリラ豪雨

一つの語に複数の下位語が存在することはもちろん、複数の上位語が存在することもある。例えば、上記の例の「玉入れ」には、別の上位語として「競技」も存在する。このとき、同一下位語の上位語間に類義関係が存在するとは必ずしもいえない。

下記の手順で Wikipedia から見出し語とその上位語を抽出した。抽出の対象となったダンプデータは2011年4月7日時点のものである。

#### 見出し語の抽出手順

1. Wikipedia のダンプデータから page.txt を抽出する。
2. page.txt 中に見出し語を識別するタグで囲まれた文字列を抽出する。
3. 抽出した文字列のうち見出し語として不適当なものを除外する。（Wikipedia では、記事で引用される画像ファイルも一つの記事として扱われており、その画像ファイルのファイル名が見出し語として抽出されることがある。）

#### 見出し語の上位語の抽出手順

1. Wikipedia のダンプデータから text.txt を抽出する。
2. text.txt の各記事の最初の一文が「AはB。」「AとはB。」の二つのパターンのいずれかに合う文章を取り出す。このときAを下位語としBを上位部とする。上位部とは上位語を含む文字列である。
3. 上記の二つのパターンに合った文の上位部のうち "[[" と "]" ]" で囲まれた文字列を上位語とする。例えば「あだち充は [[ 日本 ]] の [[ 漫画家 ]]。」の場合、

上位語：日本、下位語：あだち充

上位語：漫画家、下位語：あだち充

のように抽出する。[[…]] という記法は、…にあたる文字列が他の記事へのリンクであることを示すものである。

このようにして、見出し語を約18.3万語、上位下位関係にある上位語と下位語のペアを約183.5万ペア抽出し、これらの単語と上位下位関係をシステムの既存の語彙資源に付加した。

図2に、開発した文章評価システムのフロー図を示す。

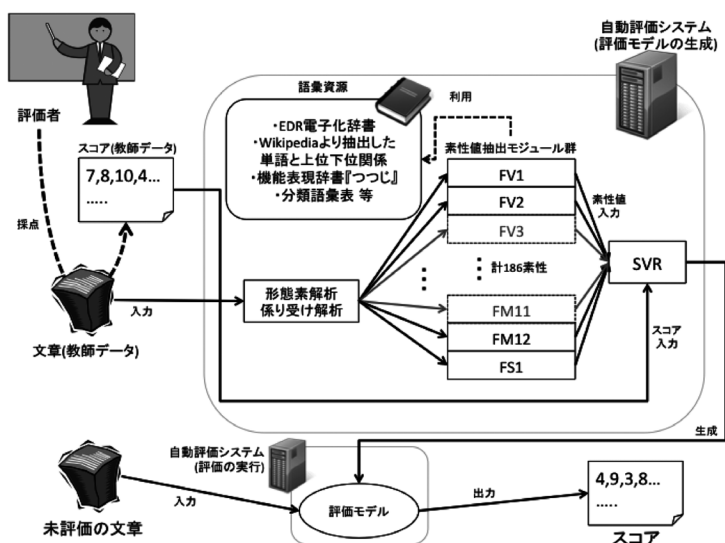


図2 詳細な言語的要素に及ぶ評価が可能な文章評価システムのフロー図

表4 文章評価システムで評価の根拠として扱われる全ての言語的要素（素性）

| カテゴリ      | ID          | 素性名                     | ID          | 素性名                 |
|-----------|-------------|-------------------------|-------------|---------------------|
| 「表層」      | FV1         | 文字数                     | FV9         | カタカナ使用率             |
|           | FV2         | 文字数制限の達成度               | FV10        | 漢字使用率               |
|           | FV3         | 文数                      | FV11        | 記号使用率               |
|           | FV4         | 文の文字数の平均                | FV12        | 英字使用率               |
|           | FV5         | 文の文字数の分散                | FV13        | 算用数字使用率             |
|           | FV6         | 文の最小文字数                 | FV14        | 文当たりの読点数の平均         |
|           | FV7         | 文の最大文字数                 | FV15        | 形式段落当たりの文数の平均       |
|           | FV8         | ひらがな使用率                 |             |                     |
| 「語」       | FW1         | 自立語の最大長                 | FW6         | 複合名詞の使用率            |
|           | FW2         | 自立語の多様性                 | FW7         | カタカナ語の使用率           |
|           | FW3         | 副詞使用率                   | FW8         | 自立語の文字数の平均          |
|           | FW4         | オノマトペ使用率                | FW9         | 自立語内のひらがな使用率        |
|           | FW5         | 感動詞、間投詞の使用率             | FW10        | 数量表現における漢・算用数字の混用有無 |
| 「文体」      | FF1         | 体言止め文の出現率               | FF6         | 括弧（ ）による補足を含む文の出現率  |
|           | FF2         | 体言が存在しない文の出現率           | FF7         | 常体・敬体の混用有無          |
|           | FF3         | 用言が存在しない文の出現率           | FF8         | 文末の単調さ              |
|           | FF4         | 名詞述語文の出現率               | FF9         | 感嘆符が用いられる文の出現率      |
|           | FF5         | 鉤括弧「」による引用を含む文の出現率      | FF10        | 口語的表現の使用有無          |
| 「係り受け」    | FD1         | 格助詞「ノ」で終わる文節の最大連続出現数    | FD9         | 文当たりの連体修飾回数の最大値     |
|           | FD2         | 格助詞「ト」で終わる文節の最大連続出現数    | FD10        | 文当たりの連体修飾回数の最小値     |
|           | FD3         | 格助詞「ニ」で終わる文節の最大連続出現数    | FD11        | 文当たりの連体修飾回数の平均      |
|           | FD4         | 格助詞「ヲ」で終わる文節の最大連続出現数    | FD12        | 文当たりの連体修飾回数の分散      |
|           | FD5         | 係助詞「モ」で終わる文節の最大連続出現数    | FD13        | 全文節中に占める名詞文節の割合     |
|           | FD6         | 格助詞「ガ」で終わる文節の文内最大出現数    | FD14        | 修飾の詳細さ（平均）          |
|           | FD7         | 係助詞「ハ」で終わる文節の文内最大出現数    | FD15        | 修飾の詳細さ（分散）          |
|           | FD8         | 一文中での「ハ・モ・ガ」の最大出現数      | FD16        | 一文中での中止法の最大使用回数     |
| 「文章のまとまり」 | FC1～FC36    | 論理的結合関係で接続する文間の構文役割遷移確率 | FC115～FC120 | 最終文と文章終端の間の構文役割遷移確率 |
|           | FC37～FC72   | 多角的連続関係で接続する文間の構文役割遷移確率 | FC121       | 参照表現が正しく機能している割合    |
|           | FC73～FC108  | 拡充的合成関係で接続する文間の構文役割遷移確率 | FC122       | 並列接続の有無             |
|           | FC109～FC114 | 文章始端と冒頭文の間の構文役割遷移確率     |             |                     |
| 「モダリティ」   | FM1         | 態度「叙述」出現率               | FM7         | 態度「許可」出現率           |
|           | FM2         | 態度「意志」出現率               | FM8         | 態度「問いかけ」出現率         |
|           | FM3         | 態度「欲求」出現率               | FM9         | 真偽判断の程度が「断定」の文の出現率  |
|           | FM4         | 態度「働きかけ－直接」出現率          | FM10        | 真偽判断の程度が「推量」の文の出現率  |
|           | FM5         | 態度「働きかけ－間接」出現率          | FM11        | 真偽判断の程度が不明な文の出現率    |
|           | FM6         | 態度「働きかけ－勧誘」出現率          | FM12        | 文末思考知覚感覚動詞使用率       |
| 「内容」      | FS1         | 論題に適合する名詞の使用率           |             |                     |

#### 1.1.4 文章評価システムの評価精度について

開発した文章評価システムの精度を調査する目的で、実験を行った。

教師データには1.1.1節で用いた高校生による小論文を電子化したものを用いる。これらの小論文は、論題  $\alpha$  「小学校の授業における英語の早期教育は必要であるか否かに対して意見を述べよ」、論題  $\beta$  「グラフと説明文を読み、日本人の子育ての態度に関してどのような特色が読み取れるかに関して述べよ」という2種類の論題に沿って書かれている。また、400字以内と800字以内の2種類の字数制限が存在する。事例は合計で584事例あり、論題と字数制限毎の内訳は表5に示す通りである。

表5 教師データ事例の内訳

|    |          | 字数制限   |        |
|----|----------|--------|--------|
|    |          | 400字以内 | 800字以内 |
| 論題 | $\alpha$ | 153    | 147    |
|    | $\beta$  | 140    | 144    |

これらの584事例に対して4人の評価者が総合的な観点でつけた10段階の評点を、各教師データのラベルとする。

#### 【実験1 評点推定の性能】

教師データを用いてSVRを構築し、評価者がつけた評点とSVRによる評点推定結果の間の差について検討する。SVRによる評点推定の評価指標には平均二乗誤差 (Mean Square Error)<sup>9</sup>を用いる。評点推定性能のベースラインには、人間の評価者による評点間のMSEを設定する。表6に、本システムのMSEとベースラインのMSEを比較して示す。

#### 実験1の考察

表6より、全ての結果において、設定したベースラインよりも低いMSEが得られている。すなわち、人間の評価者間の評価の差異よりも、人間の評価者とその評価者を手本として生成した評価モデルとの差の方が小さかった。このことから、評価者ごとに再現性の高い評価モデルの導出が可能であるといえる。

教師データのラベル系列（評価者）別にMSEを比較すると、差がみられる。これは、評価者が評価の依拠とした言語的要素の中に、本実験で設定した素性群に含まれないものが存在したことに因るものと考えられる。

#### 【実験2 評点推定の性能】

SVRを構築し、回帰係数について検討する。表7に、回帰係数の絶対値が大きいもの上位5素性を正負別に示す。ただし、FC1からFC120までの素性については、素性が示す意味自体をIDとして表記する。これらの素性IDはそれぞれ、1文字目が前文、2文字目が後文の構文役割を、3文字目が2文間の接続関係を表す。構文役割「B」は文章始

9  $\frac{1}{n} \sum_{i=1}^n (y_i - f_i)^2$   $y_i$ を評価者による評点、 $f_i$ をSVRにより推定される評点、 $n$ を事例数とする。

端、「E」は文章終端のダミー要素に、「N」は前述の「-: 出現せず」に対応する。また接続関係は、1 が論理的結合関係、2 が多角的連続関係、3 が拡充的合成関係に対応する。

### 実験2の考察

表7より、「文字数制限の達成度 (FV2)」「複合名詞の使用率 (FW6)」「真偽判断の程度が『断定』の文の出現率 (FM9)」は加点要素、「文末思考知覚感覚動詞使用率 (FM12)」「オノマトペ使用率 (FW4)」「文末の単調度 (FF8)」「形式段落当たりの分数の平均 (FV15)」は減点要素として、それぞれ評価者に依らない傾向があることがわかる。「HN-」「NH-」の素性は「ある文で主題として出現する語が隣接する前後の文で出現しない」という文章のつながりの悪さを示す素性である。これらの素性についても、減点要素とされる傾向にあることがわかる。

一方、ほとんど影響を持たない素性も存在する。カテゴリ「文章のまとまり」に関する素性群には「OE3」「HN3」など、回帰係数の絶対値が大きくかつ評価者間で極性が一致する傾向にある素性もあるが、大半の素性は回帰係数が0に近い。カテゴリ「内容」に関する素性も、全評価者において回帰係数の絶対値が小さく、あまり評点に影響を与えていないことがわかる。

表6 文章評価システムの評点推定性能 (MSE) (ベースラインは評価者による評点間の MSE)

| 提案手法  |       |       |       | ベースライン |       |       |       |       |       |
|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|
| A     | B     | C     | D     | A-B    | A-C   | A-D   | B-C   | B-D   | C-D   |
| 0.443 | 0.517 | 0.528 | 0.514 | 1.098  | 1.070 | 1.086 | 1.140 | 1.048 | 0.863 |

表7 評価者別の評価モデルの回帰係数 (数値は素性の回帰係数、{+, -} は回帰係数の正負)

|   | 順位 | 評価者 A |        | 評価者 B |        | 評価者 C |        | 評価者 D |        |
|---|----|-------|--------|-------|--------|-------|--------|-------|--------|
| + | 1  | FV2   | 0.091  | FV2   | 0.117  | FV2   | 0.247  | FV2   | 0.225  |
|   | 2  | NE3   | 0.026  | ON3   | 0.071  | FW6   | 0.090  | FW6   | 0.137  |
|   | 3  | FV5   | 0.023  | FW1   | 0.068  | FV1   | 0.084  | NO1   | 0.112  |
|   | 4  | FV10  | 0.022  | RN2   | 0.067  | OX3   | 0.079  | FM2   | 0.107  |
|   | 5  | FD6   | 0.022  | FD8   | 0.050  | BS3   | 0.077  | FV7   | 0.082  |
| - | 5  | FW9   | -0.016 | RN3   | -0.050 | BR3   | -0.073 | FF1   | -0.054 |
|   | 4  | FW7   | -0.016 | FF3   | -0.050 | FD8   | -0.075 | FM12  | -0.055 |
|   | 3  | OE3   | -0.020 | FM6   | -0.052 | FW9   | -0.082 | FF8   | -0.071 |
|   | 2  | FV8   | -0.022 | FF1   | -0.063 | FD1   | -0.102 | FW4   | -0.076 |
|   | 1  | FM12  | -0.024 | OE3   | -0.066 | FV15  | -0.106 | FW9   | -0.099 |

## 1.2 即時的に指導コメントを示す機能を持った作文支援システムの開発とその成果

### 1.2.1 開発目的

小学校の国語科の授業における「書くこと」の指導の中では、身の回りで起きた出来事や自己の考えを他者に的確に伝達するための基礎的な技術の学習活動が行われる。この学

習活動は主として、児童が作文を記述し、教師による点検（個別の言語的要素別の添削と作文に対する総括的な評価）の結果を参照するという枠組を通して行われる。教師による個別の文章表現技術の指導によって、児童は自身の作文に含まれる具体的な問題点を認識し、修正を行うことができる。このような点で、点検は、「書くこと」の能力の育成に重要な働きを持つものと考えられる。可能な限り作文中の全ての問題点に言及し、かつ適切な助言を与える必要があるといえる。

しかしながら、表現能力が発達段階にある児童の作文は多様であり、改善を必要とする事項や誤りには、多くのパターンが存在する。これらを俯瞰的に捉えて、全てを把握しながら点検を行うことは困難であるといえる。日頃より点検を行う教師によれば、一回の点検で留意できる観点は少なく、点検したい全ての観点について点検を行うにはその都度作文を読み直す必要がある。結果として多くの時間がかかり、点検結果が返却される時期には、児童自身が作文を作成した際に思いついた工夫や自信を持てなかった点等を忘れていることも考えられる。このような場合は特に、作文授業自体の教育効果が低減すると考えられる。

これに対して、学習指導要領上で指導事項として定められた種々の観点について、作文の問題点や賞賛すべき点を包括的かつ即時的に把握することができるシステムは、教師にとって効率的な点検の一助となると考えられる。また、点検事項の即時的な提示は、児童にとっても有用なものであると考えられる。作文が完成してから時間が経過した後でコメントを受ける場合に対して、作文の編集集中にリアルタイムでコメントが提示される場合、児童は自身が加えた工夫や見落とししていた間違いを記憶の新しいうちに知ることができる。その結果、より効率的な作文技術の習得が可能になると考えられる。

このような機能を持ったシステムの実現には、ユーザである児童が電子的に日本語を入力する技術を有することが求められる。昨今の教育現場では、多くの場合、小学3年生までに日本語のローマ字表記とローマ字を用いたキーボードの入力方法を学習する。したがって、一般的には小学4年生以降の児童であればシステムの使用が可能であると思われる。ただし、通常のテキストエディタを用いて作文を編集する場合、入力欄に明確な区切りがなく、児童は分量や構成方法の適切な目安を知ることができない。作文の入力フォームは少なくとも、「書くこと」についての指導事項で基礎的とされる構成に沿った作文を作成できるアフォーダンスを備えたものである必要がある。

また、電子的に文章を入力すること自体にも利点がある。まず、作文の作成後に段落を入れ替えるなどの大がかりな編集を平易に行うことができる。さらに、入力文章からの単語切り出しと語彙資源への照合が即座にできるため、児童が使用した単語に類似する言葉や表現を提示して言葉選びの補助を行うことができる。



以上より本研究では、問題点と良い点に対するコメントを即時的に出力する機能と、平易に文章の構成が行える入力フォームを有した作文支援システムを構築する。

### 1.2.2 作文支援システム構築のためのコーパス作成

システムは、実際の児童の言語活動の特徴を踏まえた上で設計される必要がある。そこで本研究では、まず児童の作文を網羅的に収録したコーパスを用意することとした。多様な言語活動を把握することが目的である以上、コーパスには、改善すべき事項の多い作文から優れた作文まで、幅広い能力水準の児童により書かれた作文が収録されることが求められる。また、それぞれの作文の問題点や優れた点に対する指摘事項が収録されることも求められる。

管見の限り、日本の児童の言語活動をまとめた従来のコーパスは、その多くが話し言葉を収録したコーパスである。その中で、永田は児童の書き言葉を収録した「こどもコーパス」を構築し、公開している（永田 2010）。小学校 5 年生 3 学級 81 人が書いた、図書をテーマとするブログ形式の文章を 1,256 編（39,629 形態素）収録している。さらに、児童が文章を編集した履歴も同時に収録しており、記述活動の「トレース」が可能という特徴を有している。規模、内容ともに有用性の高いコーパスであるといえる。しかしながら、作文の収集対象が一校の小学校に限られており、本システムの構築に必要な作文の水準の多様性からすると不十分であると思われる。坂本は、国内の小学校のホームページ上に公開された作文を収集し、コーパスを構築した（坂本 2010）。このコーパスには、265 校から収集した 10,006 編（1,234,961 形態素）の作文が収録されている。この収録編数は、それまでに存在した児童の言語活動に関するコーパスと比較して規模が大きい。また、作文の筆者（児童）の学年と性別等の情報も収録されており、有用性の高いものであるといえる。しかしながら、坂本によるコーパスの収集源となった「小学校の Web ページに掲載された作文」は、掲載までの過程において何らかの基準で選抜された作文である可能性が高い。少なくとも、指導目標に達しない作文が掲載されることは少ないと考えられる。

このように、日本語の児童書き言葉コーパスに関する先行研究には、幅広い能力水準の児童により書かれた作文が収録されており、なおかつ個別の箇所に対する添削や指導事項を基準とする総括的な評価をコーパスに含めたものは、存在しないと考えられる。

そこで、複数の小学校に在籍する様々な文章能力水準にある児童により実際のカリキュラムの中で書かれた作文を収集した。さらに、それらの作文に教師による点検を再現した情報を付与し、コーパスを構築した。

#### 【作文事例の収集】

神奈川県または東京都内の 10 校の小学校から 4 年生の作文事例を収集した。収集に協力

をいただいた小学校は下記の通りである。

- ・厚木市立戸田小学校 ・鎌倉市立深沢小学校 ・東京学芸大学附属小金井小学校
- ・茅ヶ崎市立室田小学校・大和市立渋谷小学校 ・大和市立引地台小学校
- ・横浜市立港北小学校 ・横浜市立並木中央小学校 ・横浜市立本町小学校
- ・横浜市立六つ川小学校

以下では、上記10校を50音順に a, b, …, i, j 校とそれぞれ記すことにする。

収集期間は、2011年7月～10月（第Ⅰ期）と2011年12月～2012年1月（第Ⅱ期）の2回に分かれる。第Ⅰ期には9校（a, b, d, e, f, g, h, i, j）から652編、第Ⅱ期には2校（c, f）から628編、合計1,280編の作文を収集した。

収集対象とした作文は、本研究のために特別に書かれたものではなく、各校のカリキュラムに沿って、作文活動を行う特定の単位の中で書かれたものである。表8に、作文の課題設定と学校別の作文収集事例数の内訳を示す。

表8 作文の課題設定と学校別の作文収集事例数の内訳

|      |   | 学 校 |     |     |    |    |     |    |    |    |    | 計   |
|------|---|-----|-----|-----|----|----|-----|----|----|----|----|-----|
|      |   | a   | b   | c   | d  | e  | f   | g  | h  | i  | j  |     |
| 課題設定 | ① | 63  | 131 | 0   | 55 | 48 | 51  | 0  | 52 | 80 | 84 | 564 |
|      | ② | 0   | 0   | 0   | 0  | 0  | 52  | 0  | 0  | 0  | 0  | 52  |
|      | ③ | 0   | 0   | 0   | 0  | 0  | 0   | 56 | 0  | 0  | 0  | 56  |
|      | ④ | 0   | 0   | 0   | 0  | 0  | 0   | 32 | 0  | 0  | 0  | 32  |
|      | ⑤ | 0   | 0   | 462 | 0  | 0  | 0   | 0  | 0  | 0  | 0  | 462 |
|      | ⑥ | 0   | 0   | 38  | 0  | 0  | 0   | 0  | 0  | 0  | 0  | 38  |
|      | ⑦ | 0   | 0   | 37  | 0  | 0  | 0   | 0  | 0  | 0  | 0  | 37  |
|      | ⑧ | 0   | 0   | 39  | 0  | 0  | 0   | 0  | 0  | 0  | 0  | 39  |
| 計    |   | 63  | 131 | 576 | 55 | 48 | 103 | 88 | 52 | 80 | 84 |     |

作文の課題設定は下記の通りである。

- ①：光村図書教科書4年生上編に収録の物語『一つの花』についての紹介文

『一つの花』は戦時中に生まれた少女とその家族を中心にして戦争体験を描写した作品である。『一つの花』を扱う単位では、感想を交えて作品の内容を他者に紹介する文章（紹介文）が書かれる。

- ②：光村図書教科書4年生下編に収録の物語『三つのお願い』についての感想文

『三つのお願い』は、三つの願いがかなうという一セント硬貨をめぐって、少女と少年の人間関係が描かれた作品である。『三つのお願い』を扱う単位では、作品に対する感想を記述する文章（感想文）が書かれる。

- ③：体験学習で体験したことについての感想文（4年生）

- ④：身の回りにある記号を探し出して紹介する作文（3年生）

- ⑤：4 学年の学校行事で体験したことについての感想文（4 年生）
- ⑥：運動会で体験したことについての感想文（5 年生）
- ⑦：教育出版教科書 6 年生収録『ぼくの世界、きみの世界』についての感想文
- ⑧：運動会で体験したことについての感想文（6 年生）

作文は、一部を除いて、400字詰め原稿用紙に記入されている。これらを複写した後、各校に原本を返却した。個人情報保護の観点から、児童の氏名や本文中の個人を特定する記述（児童の友人の名前、教諭の名前、学校名等）を手でマスキングした上で複写を行った。児童の氏名には筆者の異なりを表現するための ID を、本文中の個人を特定する記述には任意の名称を、それぞれ代替的に記入した。また、個々の児童が所属する学級についての情報は作文に含まれておらず、学級を区別することはできない。

これらの作文事例をコーパスとして公開するための著作権処理は行っていない。本研究での利用を目的とすることのみについて承諾を得ている。今後、各校と相談した上で公開を検討したい。

#### 【作文の点検】

収集した作文事例に対して、15名の人物（以下、点検者）の手で点検を行った。点検者全員が国語教育に関係する人物である。過去に国語教師の経験を持つ者、国語教育系の研究室に所属する大学院生と学部生、塾の講師が含まれる。点検者は、現役の横浜市立小学校教諭 1 名により、横浜版学習指導要領（横浜市教育委員会事務局 2009）に沿った点検方法の研修を事前に受けた。特に、小学校第 4 学年の「書くこと」の構成と記述に関する指導事項の解釈について詳細な解説を行った。

点検は、点検者が作文を通読した上で、原稿用紙に指摘事項を赤ボールペンで付記する形で行われた。ここでいう指摘事項とは、児童らが既に指導を受けたと考えられる指導事項と照らし合わせて、不十分あるいは優れていると判断される言語活動に言及する注釈を指す。個別の作文に対する指摘事項の形式には、指摘事項が指し示す範囲について二種の形式がある。

一つは、作文に対する総括的なコメントである。作文全体を通して不十分な点や優れている点、改善につながるアドバイス等を作文の欄外に付記する。このうち不十分な点と優れている点に関するコメントについては、それぞれ個別の記号を添えることで判別できるようにした。

もう一つは、作文中の個別の箇所に対する注釈である。主に以下に示す事項に該当する箇所について、それを修正する目的のコメントや記号が付記される。

- ・表記上の誤り（判読が難しい文字や綴りに誤りのある文字も含む。ただし、原稿用紙の使い方に関する誤りは含まない。）

例：放課後たども → 放課後なども、思ました → 思いました

- ・句読法上の難点

例：…でした私は… → …でした。私は…

- ・習得済の漢字の不使用

例：仲なおり → 仲直り

- ・文法 / 語用上の誤り

例：私が思ったのは、…だと私は思いました。 → 私は…だと思いました。

例：…だということをゆみ子の知っていました。

→ …だということをゆみ子は知っていました。

- ・主述関係の曖昧性
- ・段落分けに関する不適切さ

また、誤りを修正するだけでなく、優れている箇所についてのコメントや記号も付記する。

上記の指摘事項とは別に、点検者の総合的な判断による評価を付記した。この総合評価は絶対的な評価で A（良い）、B（標準的）、C（指導の必要がある）の 3 段階とした。

複数人の観点から点検を行う目的で、1 編の作文事例につき、異なる 2 名の点検者がそれぞれ個別に点検を行った。f 校の設定②の作文事例については、点検者の点検基準の経時変化について調査する目的で、同一点検者が一週間の間隔を置いて二度点検を行った。また、c 校の設定⑤の作文事例のうち 280 編については、2 名により付記される点検事項の言及範囲の十分性を確認する目的で、2 名の点検者に加えてさらに 1 名の点検者が点検を行った。なお、g 校の設定④の作文事例については、点検者が受けた研修の内容を鑑みて不十分な点検となることが考えられたため、現時点では点検を実施していない。

### 【作文と点検事項の電子コーパス化】

作文とそれに対する点検事項をコンピュータによる処理が可能な形に電子化した。基本方針として、収集した児童の作文を可能な限り改変せずにコーパスに収録することを目標とした。点検者による点検事項については、本文中に挿入されるタグ（以下、点検タグ）としてデータ化した。

作文本文の電子化に際して、判読が難しい文字や対応する活字が存在しない文字については、個々の電子化担当者の判断により、推測される文字を充当した。その他の誤りを含む表現や文字列については、改変を加えずにそのまま収録した。また、改行箇所については、行の途中で次行に移るか字下げが見られる箇所を改行箇所として、改行文字を挿入した。段落冒頭への空白文字挿入の有無は、本文中での字下げの有無に従うこととした。

作文の題名も電子化の対象とした。題名は下記のタグを用いて本文と区別する。

<title>…</title>

作文の総合評価についても下記のタグを用いてデータ中に記載する。

<score>…</score>

作文に対する総括的なコメントについての点検タグには、下記の3種類がある。

- ・優れている点についての講評：<praise>…</praise>
- ・不十分な点についての講評：<pointout>…</pointout>
- ・その他の講評：<comment>…</comment>

本文中の個別の箇所に対する点検タグは、主に注釈が本文に対して及ぼす編集操作の種類に基づいて体系化した。

- ・文字列挿入 <ADD seq="", type="">：文字列 seq を当該箇所に挿入する。
- ・文字列削除 <RM type="">…</RM>：タグで囲まれた文字列を削除する。
- ・文字列変更 <MOD seq="", type="">…</MOD>：タグで囲まれた文字列を文字列 seq に変更する。
- ・コメントの付記 <ANT seq="">…</ANT>：タグで囲まれた文字列にコメント (seq) を付記する。
- ・優れた箇所の明示 <GOOD>…</GOOD>：タグで囲まれた箇所が優れていることを明示する。

<ADD>、<RM>、<MOD> タグは type 属性を持つ。この type 属性は、注釈の内容的な分類を表す属性である。表記上の誤り、句読法上の難点、習得済みの漢字の不使用、文法 / 語用上の誤りに関する指摘事項の type 属性を他の事項と区別して、"e" という値で示す。現時点では、e 以外の値は設定しておらず、その他の内容の指摘事項については type 属性を null としている。

本文中の個別の箇所に対する点検タグは適用範囲が重なり合うことがあるが、このような場合はタグを入れ子にして示す。

以下に作文と点検事項のデータ例を示す。

<title>『一つの花』を読んで</title>

<score>B</score>

<pointout> 接続詞のあやまりが多い。</pointout>

<praise> 感想をたくさん書いている。</praise>

<GOOD><ANT seq="呼びかけの言葉"> みなさんは <ADD seq=", ", type="">  
『一つの花』という本を <RM type="">、</RM><MOD seq="知", type="e"> し  
</MOD> っていますか。</ANT></GOOD><GOOD><ANT seq="ゆみ子の紹介をしている"> 主人公はゆみ子という女の子です。</ANT></GOOD> (以下略)

表9に作文事例（本文と題名）の合計文字数、1編あたりの平均文字数、形態素数、1



編あたりの平均形態素数を作文の課題設定毎に示す。なお、形態素数の計数には形態素解析器 MeCab を用いた。

表 9 収集した作文事例の規模

|      |   | 合計形態素数  | 平均形態素数 | 合計文字数   | 平均文字数 |
|------|---|---------|--------|---------|-------|
| 課題設定 | ① | 152,297 | 270.0  | 261,621 | 463.9 |
|      | ② | 18,749  | 360.6  | 31,944  | 614.3 |
|      | ③ | 16,594  | 296.3  | 28,627  | 511.2 |
|      | ④ | 7,988   | 249.6  | 12,870  | 402.2 |
|      | ⑤ | 198,200 | 429.0  | 324,383 | 702.1 |
|      | ⑥ | 16,584  | 436.4  | 26,479  | 696.8 |
|      | ⑦ | 13,831  | 373.8  | 22,341  | 603.8 |
|      | ⑧ | 19,216  | 492.7  | 30,587  | 784.3 |
| 計    |   | 443,459 | 346.5  | 738,852 | 577.2 |

表10に点検タグの付与事例数の合計（各作文事例につきそれぞれ2名以上の点検者が付与した点検タグ事例の合計数）をタグの種類別に示す。type 属性がeのタグは全付与事例のうち、9,586事例存在する。

表10 点検タグ別の付与事例数の総計

| 点検タグの種類    | 付与事例数の合計 |
|------------|----------|
| <praise>   | 1,071    |
| <pointout> | 873      |
| <comment>  | 253      |
| <ADD>      | 6,089    |
| <RM>       | 3,268    |
| <MOD>      | 12,021   |
| <ANT>      | 3,633    |
| <GOOD>     | 2,915    |

### 1.2.3 構築したコーパスに基づく児童の学習傾向分析

前述の通り、構築したコーパスに含まれる点検事項は横浜版学習指導要領に沿ったものである。横浜版学習指導要領の内容は、文部科学省が策定した学習指導要領に沿っているが、指導事項の設定範囲に異なりがある。一般の学習指導要領が指導事項の設定範囲を低学年、中学年、高学年と区別するのに対して、横浜版学習指導要領は指導事項を1学年毎に詳細化して設定している。本研究では、横浜版学習指導要領に沿って、主に小学4年生の作文の点検事項から、児童の学習傾向の分析を行った。

<pointout> タグは、児童の作文が既習の事柄（カリキュラム上既に指導を受けたと考えられる作文技術）に対して不十分と判断される際に、個別の言語要素に言及した講評を付記するタグである。ここでは、コーパスに付与された<pointout> タグによって指摘される内容を、指導事項に関わる言語要素別に分類して、その付与頻度を分析する。表11に、<pointout> タグのうち付与頻度の高い指摘内容を、タグが付与された作文の編数と共に示す。ここでは、2名の点検者のうちのどちらかがタグを付与した場合に、その作文にタグが付与されたとして、編数を計数した。

「段落分け」と「常体・敬体の使い分け」については、小学校3年生で明確に指導事項として設定される内容である。本コーパスの中には、これらの事項を満たしていない作文が高い頻度で見受けられる。

段落分けに関して <pointout> タグが付与された作文は、一切段落分けが行われない作文が大半である。また、段落分けはできていても中心となる段落が明確でない作文が存在する。段落分けには書き手の述べようとすることを明確化する効果があり、「接続詞の使用法」と共に、小学校5年生以降で扱われる「事実と意見や感想を分けて書く」技術の基礎となる重要な技法である（国語教育研究所 1996）。

常体・敬体の使い分けの指導は、それぞれの文体が持つ特性を知り、その表現効果を状況別に活かす技術を習得する目的で行われる（国語教育研究所 1996）。文体の使い分けについての通常の指導では、作文が完成した後で児童が混用に気付くことになる。作文の作成中に即時的に混用に対する意識を与えることができれば、より効果的な文体の指導が実現されることが考えられる。

「自身の考えを記述すること」について指摘のある作文も多く存在する。これらの作文の大半が、読んだ物語のあらすじや登場人物、時代背景についての客観的な事柄のみを記述したものである。一方で、<praise> タグにより、自身の考えや感想を記述したことに対して優れている旨の講評を付記された作文は200編以上存在する。このことから、自身の考えを記述することを意識する児童とそうでない児童の人数が二極化していると推測される。

「句読点の使用法」については、規範が曖昧であり（比毛 1992）、適切な使用法を理解できない児童に対して指導を行うことが困難であると考えられる。本コーパスには、<pointout> タグが付与された作文以外にも、句読点の位置や有無を修正する点検事項が多く含まれる。これらの点検事項の持つ情報は、句読法の規範の体系化に役立つものと考えられる。

「テーマの一貫性」について <pointout> タグが付された作文には、記述のされ方に一定の傾向がみられた。このような作文の大半は、隣接する文間では連続性のある内容を記述している。しかしながら、作文全体での構成を意識していないためか徐々に主題がそれていき、結果として書き出しとむすびで整合性があるとは言い難い作文となっている。田中はテーマを明確化する方法として、作文の題材を児童が考え出す過程で、扱う対象となる事物についてどう考えているかを意識させる必要があるとしている（田中 2010）。作文の記述を始める前に、児童が題材について思考した内容を自発的に整理して作文の構成を検討できるような環境が設定されるべきであると思われる。

表11 &lt;pointout&gt; タグによる主な指摘内容

| タグによって指摘される内容    | タグが付与された作文の編数 |
|------------------|---------------|
| 適切な段落分けができていない   | 80            |
| 自身の考えについての記述が少ない | 80            |
| 常体と敬体の混用が見られる    | 34            |
| 接続詞の使用法に問題がある    | 18            |
| 句読点の使用法に問題がある    | 23            |
| テーマがしぼられていない     | 14            |

#### 1.2.4 作文支援システムの構築

構築したコーパスを対象とした小学4年生の学習傾向分析の結果や、4～6年生の作文のうち複数の教師により共通の指摘があった事項、有識者会議<sup>10</sup>での議論によって得られた知見に基づいて、作文支援システムを構築した。

本システムは、Webアプリケーションとして開発されており、一般的なWebブラウザをインストールしたPCやモバイルデバイスから使用することができる。

児童は、「クラス」という単位のアカウントに、「ニックネーム」という個人を認識するユーザ名を用いてログインしてシステムを使用する。アカウントは、事前にシステムの管理者より発行される。アカウントは、実際の授業でのシステム使用を想定して学級単位で発行する設計としたが、自習向けに個人的なアカウントを発行することも可能である。ユーザ名は、半角・全角に関わらず任意の文字列を入力することが可能である。ただし、同じアカウント内に同じユーザ名を持つユーザがログインしている場合は、入力がキャンセルされる。

教師は、専用の画面を通して、点検項目の設定等の環境設定、自身が担任するクラス内の作文記述状況の閲覧、完成した作文の外部出力を行うことができる。作文を行う間は点検に関する機能を無効にし、作文が完了したのちに有効化することで、即時的ではない事後の点検を行うことも可能である。

図3に本システムの主要画面と主な機能について示す。

本システムは、主に以下に示す5つの機能を備える。

<推敲促進機能、コメント機能、構成補助機能、類義語表示機能、交流機能>

##### ●推敲促進機能

作文中の問題点や記述の状況を即時的に提示する。作文中の問題点については、小学校第4学年の「書くこと」の構成と記述に関する指導事項に沿って、主に下記の項目に関係する内容の指摘を提示する。

10 有識者会議は、東京学芸大学附属小金井小学校において同校の細川太輔教諭にご臨席いただき行った。(2012年1月12日・2月20日・2月25日・2月27日)

- ・自分の考えを明確に書くこと
- ・理由を挙げて書くこと
- ・指示語を適切に使用すること
- ・各文を簡潔な分量にまとめること
- ・はじめ・なか・おわり部の分量をバランス良く配分すること
- ・中心的段落の設定
- ・敬体と常体との違いに注意して書くこと
- ・適度な分量で段落に分けること

記述の状況については、下記の項目について提示する。

- ・はじめ・なか・おわり各部の文字数・文数・段落数
- ・はじめ・なか・おわり各部の文字数の目安

### ●コメント機能

基礎的な作文技法の運用について賞賛するコメントを即時的に出力する。主に下記の項目に関係する内容についてコメントを出力する。

- ・自分の考えを明確に書くこと
- ・理由を挙げて書くこと
- ・文のつながりを明確にする接続詞使用
- ・並列的に展開して述べる構成

### ●構成補助機能

作文の入力フォームが「はじめ」「なか」「おわり」に分かれている。また、作文の記述を始める前に、題材について思考した内容を整理して入力する必要がある。

「なか」については、作文の中心的内容を、展開する順に箇条書きにして入力するアフォーダンスが存在する。この中心的内容を入力しないうちは、作文本文の入力フォームが表れない。「おわり」については、作文のむすびを、「報告」「紹介」「意見」「提案」のうちのいずれの形にするかを事前に明示するアフォーダンスが存在する。

なお本機能は、本研究において行った有識者会議においてシステムが備えるべきインタフェースについての議論の結果、実装されたものである。

### ●類義語表示機能

作文入力画面下部のボタンをクリックすることで、作文本文中で使用されている単語に似た語義を持つ語彙が提示される。ユーザは、これらの語彙を参照して、推敲に役立てることができる。また、多くの候補を表示することで、言葉を吟味する習慣をつけることができると思われる。

類義語の候補として表示される語彙は、分類語彙表（国立国語研究所 2004）に記載のある語彙のうち、光村図書出版国語教科書（全10冊）中で用いられる語彙に限って提示される。なお、教科書中で用いられる語彙は、光村図書出版の小学校教科書語彙表（光村図書出版 2011）にまとめられている。

### ●交流機能

同一クラスにログインしている他ユーザの作文編集状況をリアルタイムに参照できる。

同一クラスにログインしたユーザは「広場」という仮想的な空間に存在しており、他のユーザが文字を入力する様子と削除する様子を閲覧することができる。なお、他ユーザの作文は、直接コピーできないように設計した。

本機能は、学習指導要領の「交流に関する指導事項」にある、「書いたものを発表し合い、書き手の考えの明確さなどについて意見を述べ合うこと」を編集段階から活発に行うことを可能にする効果があると考えられる。

### 1.2.5 作文支援システムの臨床テスト

2012年2月25日・2月27日に、東京学芸大学附属小金井小学校5年3組（細川級）において、作文支援システムの運用臨床テストを行った。

臨床テストは、タブレット型コンピュータ<sup>11</sup>（以下、タブレット）で運用した場合と、通常のデスクトップ型コンピュータ（以下、PC）で運用した場合の使用感を比較できる形で実施した。2月25日にタブレットでの臨床テストを、2月27日にPCでの臨床テストをそれぞれ行った。タブレットのWebブラウザには、『Safari』を、PCには『Internet Explorer』を用いた。また、タブレットの入力にはタッチペン（スタイラスペン）を用いた。

タブレットでの臨床テストには15名の児童が、PCでの臨床テストには39名の児童がそれぞれ参加した。タブレットでの臨床テストに参加した児童は、全員がPCでの臨床テストにも参加した。児童は、テストの実施時期に社会科で学んだ地理の知識に基づいて、二者択一型の作文課題に答える。作文中に資料（社会科資料集）を参照することは、可能とした。

テストの後には、被験者である児童に使用感についてのアンケートを実施した。アンケートには、下記のような問いを設定した。

- ・「最初に目次を作るのは便利でしたか。」（1～4の整数での評価）
- ・「字数の目安があるのは便利でしたか。」（1～4の整数での評価）
- ・「切り取り・はりつけは便利でしたか。」（1～4の整数での評価）
- ・「敬体のアドバイスは便利でしたか。」（1～4の整数での評価）
- ・「言葉のヒントは便利でしたか。」（1～4の整数での評価）
- ・良かった点、悪かった点についての感想（自由記述式）

段階型の評価項目は、数字が小さいほど高評価と設定した。以下ではテスト別に、アンケートの結果とそれに対する考察を加える。表9に、段階型の評価項目の回答結果の平均値を示す。

11 本研究では、タブレット型コンピュータにApple社製の『iPad2』を用いた。



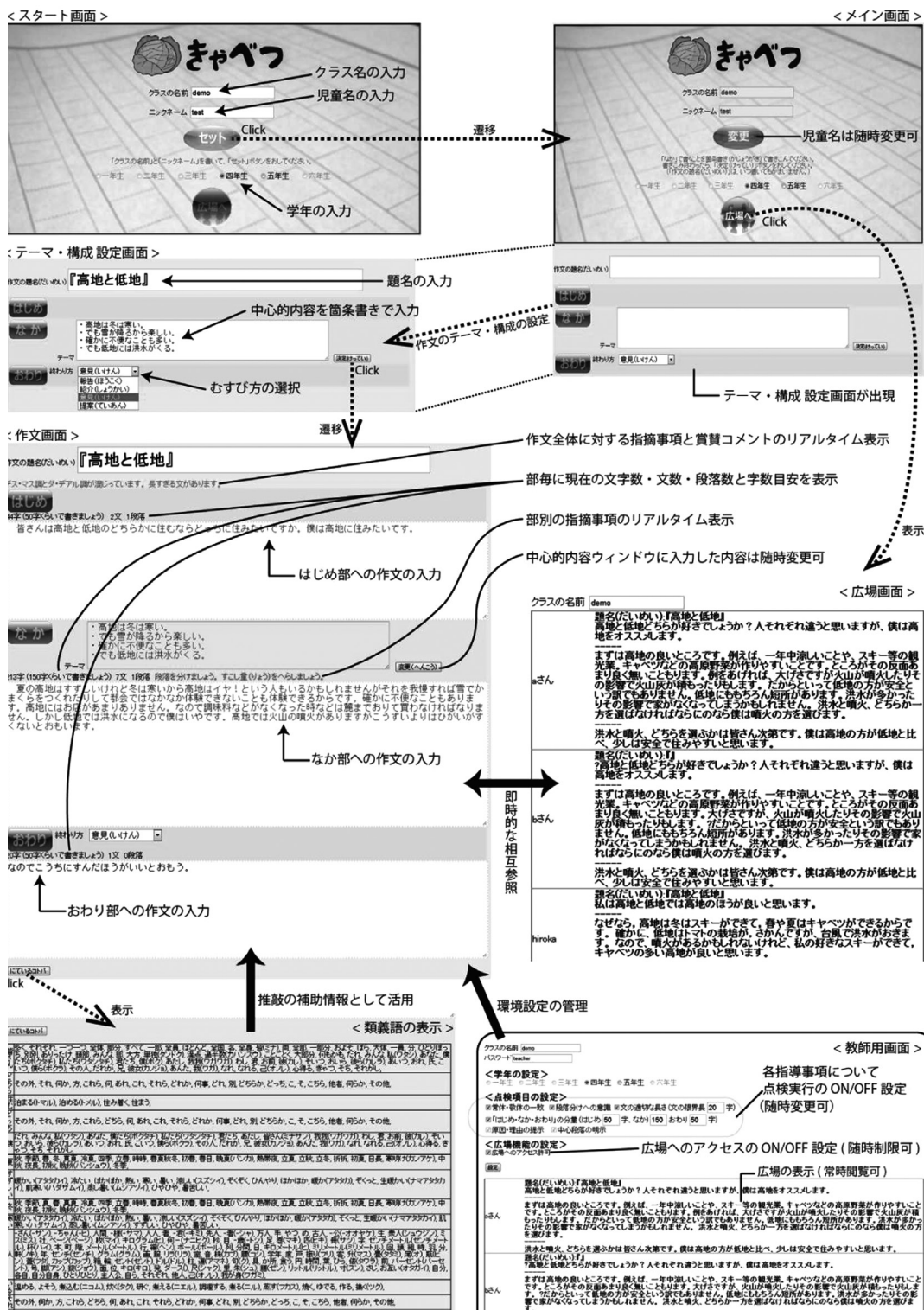


図3 作文支援システム『きゃべつ』の表示画面と主な機能

なお、両テストの実施段階において、コメント機能は未実装であった。

### 【タブレットでの臨床テスト】

与えた作文課題：「北国と南国ではどちらに住みたいですか？」

一人当たりの作文制限時間：30分

アンケートの回答者：15名（参加者15名のうち）

はじめ・なか・おわりの字数目安設定：50字・150字・50字

#### ＜感想の要約＞

利点・設定したテーマを参照しながら記述することで、文章がまとまりやすくなった。

- ・類義語が提示されることで新しい言葉を使うことができた。
- ・簡単に画面を拡大・縮小することができて視認性が良かった。
- ・現在の字数と字数の目安が表示されることによって作文の冗長性に注意することができた。
- ・かな漢字変換機能が使えることによって、文章構成に集中できた。

欠点・編集（切り取りとはりつけ）の方法がわかりにくかった。

- ・表示される類義語が過剰に多く、語彙の検討が難しかった。
- ・字数の指摘に意識が向いて、書きたいことが曖昧になった。

### 【P Cでの臨床テスト】

与えた作文課題：「高地と低地ではどちらに住みたいですか？」

一人当たりの作文制限時間：40分

アンケートの回答者：32名（参加者39名のうち）

はじめ・なか・おわりの字数目安設定：50字・150字・50字

#### ＜感想の要約＞

利点・設定したテーマを参照しながら記述することで、文章がまとまりやすくなった。

- ・文字数の指摘によって、一文を簡潔にすることを意識できた。
- ・段落分けの指摘によって、構成を工夫する意識ができた。
- ・類義語が提示されることで新しい言葉を使うことができた。
- ・交流機能によって、様々な書き方を考えながら記述できた。
- ・キーボードが押しやすく入力ミスが少なかった。
- ・編集（切り取りとはりつけ）が簡単に行えた。

欠点・字数の指摘に意識が向いて、書きたいことが曖昧になった。

- ・ローマ字入力が苦手であるため、入力に時間がかかった。

段落分けや文の簡潔さについての指摘、構成の事前検討、類義語の提示、交流機能など、本システムが独自に実装した機能については、効果的であったと考えられる。また、表12からも常体・敬体のアドバイスも有効であったと考えられる。

表12 使用デバイス別の段階型評価項目の回答結果（全回答の平均値）

|        |       | 設 問               |                   |                   |                  |                |
|--------|-------|-------------------|-------------------|-------------------|------------------|----------------|
|        |       | 最初に目次を作るのは便利でしたか。 | 字数の目安があるのは便利でしたか。 | 切り取り・はりつけは便利でしたか。 | 敬体のアドバイスは便利でしたか。 | 言葉のヒントは便利でしたか。 |
| 使用デバイス | タブレット | 1.47              | 1.60              | 1.78              | 1.64             | 1.60           |
|        | P C   | 1.56              | 1.69              | 1.53              | 1.63             | 1.66           |

推敲促進機能を利用して逐次的に推敲を行いながら、作文を記述する様子が多く見受けられた。特に、段落分けに関する指摘があった場合は、原稿用紙上での作文に比べて編集作業のコストが小さいため、積極的に段落分けを行う傾向が見られた。このように作文中に編集を促進することで却って児童が記述しようとする主題が曖昧になりやすくなることも考えられたが、構成補助機能を用いて作文の骨子を確認する児童が多く見られた。類義語機能は、推敲の活性化に限らず新語彙の学習にも有効であることがわかった。提示された語彙のうち、知らない語彙を試みに作文に使用する児童も見受けられた。また、交流機能については、自作文の編集画面と広場画面を同時に開き、友達の作文の編集過程を観察しながら参考にする様子が見られた。一方、被験者の作文能力の水準が高いためか、理由づけ等の指摘事項が反応することは少なかった。本テストとは別の設定で追試を行う必要がある。

しかしながら、文や段落の分量に関する指摘の機能のように、却って作業の支障と捉えられる場合もあった。これらの機能については、極端な例のみに指摘を加えるように調整を行う必要がある。また、表示される類義語についても、作文での使用が想定されない語彙を除外して表示するなどの制限を行う必要がある。

電子的な入力方法による利点についての感想も見受けられた。中でも、原稿用紙に筆記用具を用いて記入する場合に比べて、大がかりな推敲が容易になることへの感想は多い。編集作業にかかるコストが低減する分、より作文の構成に集中することができるようになったと考えられる。一方、学習した漢字を思い出して筆記する必要があるという点は、漢字学習の観点では欠点となりうる。しかしながら、作文の学習という観点では、編集作業のコスト低減と同様に、構成に集中できる点で利点があると考えられる。

タブレットでの臨床テストとP Cでの臨床テストでは概ね感想が共通しているが、デバイス特有の感想もある。タブレットは画面操作が容易である反面、キーボード入力や文字

列操作に難点があるという感想が多い。一方 P C は、タブレットに比べてキーボード入力  
が容易であるという感想以外に、難点についての感想は特に見受けられなかった。概し  
て、児童にとって、エディタとしてはタブレットに比べて P C の方が使いやすいデバイス  
であったと考えられる。しかしながらタブレットにはビューアとしての利点がある。タブ  
レットを介してシステムが使用される際には、デバイスの特徴が効果を発揮するようなイ  
ンタフェースに改める必要がある。

## 2. 对外発表等の状況について

下記の通り、諸学会に論文の投稿を行った。

- ・ 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  
自動評価と評価モデルの顕在化”, 情報処理学会研究報告, 自然言語処理研究会, 2011-  
NL-202, 平成23年7月発表.
- ・ 藤田央, 藤田彬, 田村直良. “Wikipedia から抽出した語彙関係リソースの小論文自動評価  
タスクへの適用”, 第10回情報科学技術フォーラム (FIT2011), 平成23年9月発表.
- ・ 藤田彬. “自然言語処理手法に基づく作文自動点検の試み：学習指導要領「書くこと」に  
立脚した文章評価補助のための技術”, 全国大学国語教育学会第121回高知大会, 平成  
23年10月発表.
- ・ 藤田彬, 田村直良. “作文事例に基づいた児童の「書くこと」に関する学習傾向について  
の分析－小学四年生による紹介文・感想文を中心に－”, 言語処理学会第18回年次大会  
(NLP2012), 平成24年3月発表.
- ・ 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  
自動評価と評価モデルの構築”, 言語処理学会論文誌『自然言語処理』, (投稿中).
- ・ Akira Fujita, Naoyoshi Tamura. “Automated Evaluation of Japanese Texts Based on  
Educational Points of View and Construction of the Individual Evaluation Model”,  
ICEL2012, 平成24年6月発表予定 (フルペーパー査読あり).

また、作文支援システム『きゃべつ』については、人工知能学会が主催する第64回人工  
知能セミナー「教育と人間科学～理論、実践、社会での取り組み」において、講演「小学  
校教育における I C T 活用」の中で紹介された。

作文支援システム『きゃべつ』は、下記 URL における公開を予定している。

<http://yamato.tamlab.ynu.ac.jp/cabbage/index.html> (2012年6月公開予定)



### 3. 今後の課題について

#### 【文章評価システム開発についての課題】

文章評価システムで扱われる素性のうち「文章のまとまり」「内容」といったカテゴリの素性群は、ほかの素性群に比べ影響が小さいことがわかった。今後、特にカテゴリ「文章のまとまり」を評価の観点に加えるには、修辞構造等の文章構造を大域的にとらえたものを素性として加える必要があると考えられる。また、個々のカテゴリに関して素性の適性の検討を行う上では、カテゴリ別の評点をラベルとして学習を行う必要がある。

本研究では複数の評価者の評価モデルによる評価から最終的な評価判断を導き出すことについて扱っていない。しかしながら、試験等での使用を考慮すると、受験者の最終的な評価を決定する機能が必要となる。このような機能は、自動評価手法一般に期待されるものであるといえる。今後これらの機能について検討を行う必要がある。

本システムは、小学生の「書くこと」についての指導事項を参考にした素性が設定されているが、小学生作文を入力した場合のロバスト性について実証を行っていない。今後、小学生作文を入力した場合の精度評価を行う必要がある。

#### 【作文支援システム開発についての課題】

前述のように推敲促進機能について、一部指摘事項が却って作文記述の支障となる場合があった。指摘事項を表示する条件に調整を加える必要がある。類義語表示機能についても、単語の表示条件を成人の作文での使用頻度に基づいて制限するなどの調整を加える必要がある。また、コメント機能については、本研究において実施した臨床テストにおいて有効性を確認できていないため、再度設定を変えてテストを行う必要がある。

本研究で構築したコーパスについて、今後収集対象とする学年を広げることで、教育分野に限らず、多くの分野（社会学や発達心理学等）での応用が期待される。また収集地域の局所性を解消することで、自治体毎の指導目標の異なりに対応したコーパスとする展開も考えられる。さらに、他の課題設定、文種の作文を収集することで、より有用性の高いコーパスとなると考えられる。今後は、以上のようにコーパスの拡張を進めるとともに、収録作文の著作権処理について検討を行う予定である。

また、コーパス中の点検タグの内容をクラスリング等の統計的手法に従って分類することで、さらに知見を得ることが可能であると考えられる。

#### 【謝辞】

本研究にあたり多大なる御支援を下さった公益財団法人博報児童教育振興会、有益な御助言を戴いた横浜市立港北小学校鳥形昌子先生と東京学芸大学附属小金井小学校細川太輔先生、作文収集や実験に御協力戴いた小学校の皆様、に、深謝の意を表する。



## 【引用文献】

- Barzilay, R. and Lapata, M. (2008). Modeling Local Coherence: An Entity-based Approach, *Computational Linguistics*, 34(1), pp.1-34.
- Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English*, Longman, London.
- Krippendorff, K. (1980). *Content Analysis: An introduction to its methodology*, Beverly Hills, Sage Publications.
- Smola, A. J. and Schoelkopf, B. (1998). A tutorial on Support Vector Regression, Technical report, NeuroCOLT2 Technical Report NC2-TR-1998-030.
- 宇佐美慧 (2011). 小論文評価データの統計解析—制限字数を考慮した測定論的課題の検討—, 行動計量学, 第38巻, 第1号, pp.33-50.
- 横野光, 奥村学 (2010). テキスト結束性を考慮した entity grid に基づく局所の一貫性モデル, 自然言語処理, Vol.17, No.1, pp.161-182.
- 横浜市教育委員会事務局 (2009). 横浜版学習指導要領 国語科編, ぎょうせい.
- 永田亮, 河合綾子, 須田幸次, 掛川淳一, 森広浩一郎 (2010). 作文履歴をトレース可能な子供コーパスの構築, 自然言語処理, Vol.17, No.2, pp.51-65.
- 岩淵悦太郎 (1979). 悪文 第三版, 日本評論社.
- 光村図書出版株式会社 (2011). 語彙に着目した授業をつくる 小学校国語 語彙表編, 光村図書出版株式会社.
- 国語教育研究所 (1996). 「作文技術」指導大事典, 明治図書.
- 国立国語研究所 (2004). 分類語彙表増補改訂版データベース.
- 坂本真樹 (2010). 小学生の作文コーパスの収集とその応用の可能性. 自然言語処理, Vol.17, No.5, pp.75-98.
- 松吉俊, 佐藤理史, 宇津呂武仁 (2007). 日本語機能表現辞書の編纂, 自然言語処理, Vol.14, No.5, pp.123-146.
- 松吉俊, 江口萌, 佐尾ちとせ, 村上浩司, 乾健太郎, 松本裕治 (2010). テキスト情報分析のための判断情報アノテーション, 電子情報通信学会論文誌 D Vol.J93-D No.6.
- 小野正弘 (2007). 擬音語・擬態語4500日本語オノマトペ辞典, 小学館.
- 森岡健二 (1963). 文章構成法 文章の診断と治療, 至文堂.
- 田中定幸 (2010). 作文指導のコツ②中学年, 子どもの未来社.
- 日本電子化辞書研究所 (2010). EDR 電子化辞書 V4.0, 独立行政法人情報通信研究機構.
- 比毛博 (1992). 点のうち方, ことばの科学 5, pp.221-237.
- 文部科学省 (著作権所有) (2008). 小学校学習指導要領解説 国語編, 東洋館出版社.