

作文コーパスを資料に児童・生徒の漢字使用・選択傾向と発達の実態を明らかにする ——語彙情報つき作文コーパスの構築と学齢別語彙・漢字使用実態調査——

富山大学 人間発達科学部 宮城 信
文部科学省 初等中等教育局 今田 水穂

研究成果要約

1. 研究活動の概要

本研究では現在までに収集し電子化した作文資料を利用して語彙情報を付与したおよそ66万語程度の「児童・生徒作文コーパス」を構築した。このコーパスは調査時の条件を統一した作文コーパスで、児童らの作文の実態を反映している。作文調査は平成26年と27年のそれぞれ春頃（6月）と秋冬頃（12月頃）合わせて計4回実施し、それぞれの題は、春は「ゆめ（夢）」、秋冬は「ぼくの／わたしのがんばったこと」である。この資料を用いて、複数の観点から学齢別の標準的な漢字使用能力の推定を目的とした児童・生徒の語彙および漢字使用の実態調査を実施した。調査結果は、漢字別、語別に一覧表にまとめて教育・研究利用に活用可能な形（報告書にCD-ROMで添付した）で公開する。また、その結果を受けて学齢別に沿った漢字使用の到達目標の検討を行った。

2. 研究成果の概要

本研究では、各種語彙情報を付した「児童・生徒作文コーパス」を基礎資料として、現代の児童らの作文に於ける語彙使用・漢字使用の実態を明らかにしたことが最大の成果である。調査の基礎資料が先行研究と異なり、同一条件下で教師や保護者等大人の手が入っていない、生の作文を収集したことであることにも価値があると考えている。現在利用可能な児童らの作文能力、特に語彙使用・漢字使用の実態を映した同等規模の資料が存在しない状況では、本研究の成果は相応の意味があり、今後、児童らの学齢別の漢字使用の実態を詳説していくための足掛かりとしては、十分に要求を満たすことができたと考える。

本研究では、このコーパスを基礎資料として、次の課題について調査を行った。

（本研究に於ける調査課題）

- ① 頻出漢字と当該漢字を使用した頻出語彙のリストの作成
- ② 主要品詞別頻出語彙のリストおよび当該語彙の漢字使用率の調査
- ③ 和語の頻出語彙のリストおよび当該語彙の漢字使用率の調査
- ④ 学齢別対中学3年比の漢字使用率（飽和率）の調査
- ⑤ 学齢別文章量と漢字使用率、文の特性に関する調査

上記課題の調査結果は、語彙表等にまとめて全データを一覧表として示した。

3. 成果活用について

本研究の成果を掲載した報告書を各教育機関に配布する予定である。また、本研究の成果である語彙表等の資料は、全データを報告書に添付する以外に、HPで公開する予定である。また、成果の現場の教育利用を考えた場合、本研究で示された資料の配布だけでなく、それを具体的に活用するための提案も必要であると考え。現段階では達成できていないが、今後本研究で得られた資料を基に語彙・漢字学習システムを構築し、それを踏まえた学習教材の開発・公開を目指す。

4. 今後の研究課題

本研究の限界は、主たる調査対象の児童・生徒らが国立大学附属学校という特殊な環境にあるという点にある。「児童・生徒作文コーパス」の汎用性を追求するのであれば、公立小・中学校のデータや、公立でも地方部と都市部を比較する観点を踏まえた児童・生徒の作文データの収集も必要であろう。今後は一般の公立小・中学校での継続調査も視野に入れながら資料の拡充を図っていく必要がある。さらに、児童らの文章作成能力に関する評価を相対化するために、他ジャンルの文章との比較、高校生や大学生の文章との比較も視野に入れて、研究対象を拡大していく予定である。

研究成果報告

1. はじめに

1.1 研究の背景

児童や生徒が書いた作文の分析は、国語教育の中でも比較的関心が高い分野の一つである。特に児童を対象とした作文の分析・指導は1955年頃から現場の教員を中心に盛んに行われ、作文の誤りを修正していくという方法で一定の成果を上げてきた。それに対して、より客観的に児童作文の実態に迫ることを目指して、国立国語研究所が中心となり、協力校を募って進められた児童作文の6年間に亘る大規模経年調査が国立国語研究所(1964)である。この調査では、相当数の作文を収集して出現語彙や文章構成力、文章量等を計量的に調査・検証し、児童の作文能力が内容・計量・形式の3方面から分析され、その特徴と問題点が学齢別に指摘された重要な資料である。その後同国立国語研究所(1989)では、1970年から10年間に刊行された文集を資料として語彙の収集と分類に重点を置いた調査を実施している。この調査で収集された作文のレベルに関しては、同書に「文集発行の目的が、コンクールや優秀作品の選抜ではないこと。」(同1989, p.19)と明記されており、文集を資料としたと言っても、特段文章力の高い児童らの作品ばかりではないこと等、一般児童の作文能力に迫ろうという配慮がなされていることが分かる。これ以降、国立国語研究所が実施した規模の調査は現在までほぼ行われておらず、経年変化の可能性に注意すれば、児童らの作文能力を推定する資料として現在でも十分に利用価値が高い資料と言える。比較的規模の大きな調査としては、その後しばらく間を置いてなされた長友和彦編(1997)がある。長友らの調査では1992年にお茶の水女子大学と同附属小中高等で各学年1クラスずつを対象として児童・生徒の作文調査が実施されている。この調査の特徴は、それまでのように公開された雑多な作文を収集するのではなく、作文の題を「手」に限定し、字数(400字)や時期等の調査条件を統一した点にあり、学齢による発達段階の差異を測ることに眼目がある(帰国子女に対しても同様の作文調査を実施している)。大掴みではあるが、これまでこのような形で児童らを対象とした代表的な作文調査が実施されてきた。

これらの調査と並行して、作文や教科書、新聞、一般書籍等の資料の用字・語彙に基づいた教育基本語彙の研究も盛んに行われている(紙幅の都合で詳細は割愛するが、国立国語研究所2009等に詳しい)。近年でも、それぞれアプローチは異なるが、義務教育期間の学習過程で習得に適した語彙はどれかという観点から考察が進められている。例えば、河内昭浩(2015)や田中牧郎(2011)は、「国語政策・国語教育のよりどころとなるような重要語彙リストを作成する」(田中2011, p.86)ことを目的として、子ども達が優先的に学習すべき語彙の選定を念頭として、日常生活における重要語彙を検討している。また、鈴木一史(2011)では、中高一貫の中等教育課程における生徒達の語の使用を調査し、それらの語の重要性を検討している。

作文教育にとって教育基本語彙の選定は重要な課題である。それと同時に児童らがどのように語彙・漢字を用いているのかに関する実態の解明も喫緊の課題である。従来、児童の漢字能力の実態と到達目標に関して、日本教材文化財団(2010)やベネッセ教育総合研究所(2007)等の漢字テストによる実態調査や、田中(2011)のような大規模コーパスを用いて社会生活に必要な漢字を推定する研究等が行われてきた。しかしながら、これらの調査で得られた結果は、

次2. 1節で述べるような児童・生徒の語彙および漢字使用の実態そのものを表しているわけではないと考えられる。果たして児童らは既習の漢字を総動員して作文しているのだろうか、拙いながら筆者のこれまでの中学・高校での作文指導の経験上その答えは否である。では、児童らが実際作文する場でどのような語を選択し、どのように用字して書き記しているのか、実はその実態は未だ明らかにされてはいない。特に現代の児童らの文章産出能力と用字の傾向についての研究はまだ黎明期にあると言って良い。

1.2 本報告書の構成

以下に本報告書の構成を述べる。まず2節では児童らの漢字使用についての仮説を述べ、調査を行う理由について述べる。次に3節では本研究の概要を述べ全体の流れを捉える。そして4節では本研究に於ける調査の方法を詳説し、得られたデータの概要を述べる。続けて5節では本調査で得られた結果を考察する。最後に6節では本研究の成果を総括して展望を述べる。また本研究で得られた語彙表や頻度表等の基礎データは、本報告書添付のCD-ROMに収録した。

2. 研究の目的

2.1 児童・生徒の作文に於ける漢字使用の実態を明らかにする

本研究では計画的に構築した作文コーパス（以下、「児童・生徒作文コーパス」とする）を用いて児童・生徒の語彙および漢字使用の実態調査を実施し、学齢別の標準的な漢字使用能力の推定と到達目標の検討を行う。既に述べたように、児童らの作文能力の解明には、使用語彙の調査と共に用字（本研究では特に漢字使用に着目した）に関する調査も重要な課題となる。漢字習得に関しては、日本教材文化財団（2010）やベネッセ教育総合研究所（2007）等が採用しているようなテスト形式による調査によって児童のおよその漢字習得状況を測ることは可能であろう。一方で、その結果が必ずしも児童らの漢字使用の実態を反映したものとは言えない。言い換えれば、児童らは書ける漢字を必ず作文時に用いているのかという疑念が残る。また、社会生活に必要な漢字の習得は長期的目標としては妥当であるが、個別の学齢における漢字使用の到達目標についての検討は十分にされていない（学年別配当教育漢字の習得そのものではないことに注意）。そのために、必要とされる調査・研究のためには児童の漢字使用の実態を客観的、計量的に評価するための研究資料が不可欠である。しかしながら、そのような資料を収集するための調査は昨今個人情報保護の観点等から収集・公開が難しく、十分な質・量を備えた利用可能な研究資料はこれまでほとんど公開されていない。また漢字使用の実態を調べるためには文字単位だけではなく語単位でも選択傾向の調査をする必要があるため、例えば単純に全文字数に対する漢字の使用頻度を調べるだけでは、漢字使用の実態を捉えているとは言えない。そのため資料に対する使用語彙調査を併せて実施する必要がある。そこで本研究では、現在構築中の「児童・生徒作文コーパス」（宮城信・今田水穂2015aを参照のこと）を利用し、このコーパスに語彙情報を付与した研究資料を作成して¹児童の語彙および漢字使用の実態を

1 本調査で使用する資料は児童の作文を誤字等も含め忠実に反映した電子化テキストである（4.2節を参照のこと）。この資料に対して辞書形、読み等の語彙情報を付与することにより、学齢別、語別の漢字使用の実態を精密に調査することができるようになる。平文テキストは誤字脱字を含み、低学年では漢字の使用頻度が極端に少ない。また、「体いくかん」（「体くかん」のような誤

調査し、冒頭の課題の達成を図る。

2.2 児童らの漢字使用に関する仮説と成果の応用

本研究では、児童らの作文に於ける漢字使用について、「習得した漢字をすぐに用いるわけではなく、使用状況は字毎や語毎に異なる」という仮説を立てる。即ち、漢字によって学習後すぐに使用される字としばらくしてから使われるようになる字があるといったような差異があるのではないか。または、語によっても漢字の使用傾向（漢字書き率）に異なりがあるのではないかという疑義が生じている（以上は筆者が拙い作文指導の経験から得た実感である）。それを検証するために、本研究では2.1節で述べたように児童・生徒らが書く作文を資料として漢字使用傾向を調査して学齢毎の標準的な漢字運用力を推定し、その発達を明らかにすることを目的とする。作文資料を調査していくと、いくつかの漢字に於いて仮説に即する形で、学習後すぐ漢字書きする語、遅れて漢字書きする語、漢字書きされ難い語等の違いがあることが分かった。このような児童の漢字使用の実態が詳細に記述できるようになれば、現場での漢字学習の復習や作文指導の場に於いて、証左を有する軽重を付ける有効な基準となり得る²。それを受けて、成果の活用法として最終的には小中学校の現場の教師に漢字指導の具体的な指針を提案することを目指す。これまで筆者らが実施したアンケート小調査、講演、免許更新講習会などの場で、現場の教師からも、そのような資料に対する同様な要望があったことも需要の高さを物語っている。

2.3 児童らの漢字使用の実態調査の必要性

1節で概説したように、児童らの作文能力を推定する資料は1970年代（長友編1997を含めても1990年代）を最後にほぼ存在していない。既出の国立国語研究所の調査（1964,1989）に於いては、調査が行われたのが50年近く前で、他人の推敲があり児童ら自身の作文との乖離がある疑いが拭えないことや、とりわけ高度に発達した情報メディアの影響を多分に受けた現在の児童の拙い作文技術の実態を示す資料としては時代にそぐわないものとなっている可能性がある。近年では宮城・今田（2015a）で整理されるように、利用可能な児童作文のデータベースとして、いくつかのコーパスが存在している。しかしながら、それぞれ資料の等質性という点で問題を含んでいる（詳しくは宮城・今田2015aを参照のこと）。現在の児童らの作文能力を反映した客観的な資料が存在していない以上、彼らの作文能力の発達段階を適切に掴むことは難しい。この問題は、作文指導にも直接的に波及する。先のような状況から時宜を得た適切な指導方法が確立できず、また、現在具体的な指導の指針等が示されていないため、教師が児童らの作文の評価を適切に行うのも困難という状況にある。さらに前述の先行研究の蓄積の活用も容易ではないため、現段階では多くの教師が自らの経験に基づいて指導を行っており、教師間の指導法のゆれが存在していることも想像に難くない。そしておそらく作文の善し悪しを

表記も散見される)のような一般的ではない漢字の書き分けや誤表記を含むデータも散見される。これをそのまま形態素解析器に掛けても適正な語彙情報付与は困難であることが予想され、本研究では、まず人手で語彙情報を付与したコアデータを作成することから始めることにした。

2 一方で、安易な基準の提示は言葉狩りを誘発しかねない。本研究の成果が適切に運用され、文章作成時に「よく書かれる漢字」や「書けるようになって欲しい漢字」の選定、必要に応じた重点指導の実施等学習の改善が期待される。

経験的に判断するしかない綱渡りの状況が続いていると考えられる（特に経験の少ない若手の教師にとっては切実な問題である）。以上の点から、児童らの作文能力を適切に反映し、また調査条件等に等質性のある基礎資料の整備が喫緊の課題となっている。

3. 研究の概要

3.1 研究方法の構成

本研究は先に挙げた課題を解決するために、およそ図1のような方法で進められた。



図1 研究方法

図1で示すように、本研究は、すでに基礎資料となる「児童・生徒作文コーパス」の構築を進めつつある段階から開始されている³ので、当該資料の特性や限界等もある程度把握できている。また、富士原紀絵他（2016）で示されるように、当該資料を用いた研究もいくつか試行しており、信頼性が担保されている。その結果を受けて、本調査は先の研究の注意点を踏まえつつ周到に進められた。まず、[試行]では「児童・生徒作文コーパス」の構築と共に進められてきた予備調査等を検討し、当該コーパスが研究資料として妥当なものかどうかという評価を実施した。次に、[実態調査]では本調査でどのような調査課題を扱えば良いのか、現場の教師にインタビュー小調査を実施し、要望が多かった項目について、実際に調査が可能か検討を行った（その中で検証の実効性があった課題を後述の(1)のように採用した）。さらに[資料の整備]では構築中の「児童・生徒作文コーパス」に対して新たな語彙情報を人手で付与する等、本調査に合わせて追加のアノテーションを実施した。そして[分析]では分析後のデータの妥当性の検証を実施すると共に、CD-ROMに収録した語彙表等の資料を作成した。最後に[教育利用]では、本研究の成果を現場での教育利用で活かすための教材等の開発についての展望を示した（ただし、現時点に於いてはこの目的を十分に達せられていない）。なお、基本的にはこれらの作業に筆者らが分担して当たったが、情報を交換しながら必要に応じて共同で調整作業を実施している。

本調査で使用する原作文資料は児童らの作文を誤字等も含め忠実に反映した電子化テキストである（詳細は4.2節を参照のこと）。これに辞書形や読み等の語彙情報を付与したコーパスを資料として、学齢別、語別の漢字使用の実態の調査を実施する。研究の精度を高めるために、人手で修正を行った語彙情報を付与したデータを用意する必要がある。その際主立った情報付与に関する作業を専門の業者に依頼し、一部を研究補助者に展開して、およそ66万語程度の

3 第9回 博報財団 児童教育実践についての研究助成（代表：富士原紀絵）および科研費基盤研究（B）（代表者：矢澤真人、研究課題番号：26285196）の研究で収集・電子化された平文テキストに対して、本助成を受け、専門技術者が人手で語彙情報を付与した基礎データが「児童・生徒作文コーパス」である。

語彙情報つきの基礎データである「児童・生徒作文コーパス」を構築した（基礎データの概要については4. 3節を参照のこと）。

本研究では、次のような調査課題を立てて、同コーパスを資料として分析を進めることにする。なお、次の①～⑤の調査課題の具体的な分析は後の5節に詳説した。

(1) 本研究に於ける調査課題

- ①頻出漢字と当該漢字を使用した頻出語彙のリストの作成 (5. 1節)
- ②主要品詞別頻出語彙のリストおよび当該語彙の漢字使用率の調査 (5. 2節)
- ③和語の頻出語彙のリストおよび当該語彙の漢字使用率の調査 (5. 3節)
- ④学齢別対中学3年比の漢字使用率（飽和率）の調査 (5. 4節)
- ⑤学齢別文章量と漢字使用率、文の特性の調査 (5. 5節)

3.2 研究体制

本研究に於いて、次のような研究体制を組織し作業の分担を行った。それぞれの業務の責任者を配置した上で、綿密に連絡を取りながら実質的には共同で研究を進めた部分もある。

(2) 研究体制

- ・研究の統括、調査協力校との連絡調整、作文データの下処理、「児童・生徒作文コーパス」の企画：宮城 信（研究代表者）
- ・「児童・生徒作文コーパス」アノテーションの設計、コーパス作成の統轄、語彙表等資料の作成：今田水穂（共同研究者）

3.3 調査概要

作文調査の研究協力校として、国立大学附属小学校と中学校2校、9学年に依頼し、全児童・生徒を対象とする悉皆調査を実施した。調査は春期と秋冬期の年2回、全3年計画で実施している途中である。本研究の資料は平成26年と27年に実施した計4回分の調査によって得られたものである。附属中学校には附属小学校からの連絡進学があることから、本研究では追跡調査まではできていないものの、今後小学校1年生から中学校3年生までの義務教育期間9年間の長期に亘る縦断調査を実施する可能性も想定している。調査に際しては、国語科の担当教員が監督・指示し、国語の授業時間を使って作文を書かせた。調査条件等の等質性を担保するため、時間や作文用紙、記述上の諸注意等を統一したマニュアルを作成してそれに従って調査を実施した。作文を書かせる際に、担当教師から成績には一切関係が無いこと、今後の作文指導に活かされること等を口頭で伝え、児童らの心的負担を軽減する配慮を行った（調査の詳細については4. 1節を参照のこと）。

3.4 成果発表

現在までの本研究に関する成果発表は以下の通り実施された。

[口頭発表]

- ・宮城信・今田水穂「『児童・生徒作文コーパス』を用いた漢字使用能力の推定」
第7回コーパス日本語学ワークショップ（2015年9月1日：於 国立国語研究所、発表資料：
宮城・今田2015b）

[依頼講演]

・宮城信「作文指導について一言語活動の充実を図った指導を目指して」

(2015年11月9日：於 富山県射水市立射北中学校)

なお、今後本報告書の内容を踏まえて論文化を進め、学会誌等に論文を投稿する予定である。

4. 作文による実態調査

4.1 調査対象・方法

既出の先行研究等を検討した結果、本研究では作文の収集から、調査条件等を整備した適切な方法で始めることによって研究全体の質の保証を目指すこととし、次の条件を満たすことを企図した。

(3) 作文調査に関する必要条件

- ①教師や保護者等大人の推敲が入っていないもの(児童・生徒自身の力だけで書いたもの)
- ②同一の条件で作成されたもの(年齢・課題・指示・時間・作文用紙等)
- ③分析に耐えうる量の確保(学年全体、全校で取り組んだ作文)

作文調査は、平成26年と27年のそれぞれ春頃(6月)と秋冬頃(12月頃)合わせて計4回実施し、それぞれの題は、春は「ゆめ(夢)」、秋冬は「ぼくの／わたしのがんばったこと」とした。題は先行研究を踏まえて筆者らが検討した結果、抽象的なものと生活文的なものを与えてみることにした。「ゆめ(夢)」については、将来の夢を書くことと、就寝時に見た夢を書くことのいずれの可能性も想定した(どのような「ゆめ」を期待しているのかと言った指示は行っていない)。実際「ゆめ」作文については、「昨日見た夢」「将来の夢」「夢を見ることの意義」等様々な内容の作文が収集された。

作成時間は、小学校40分、中学校45分を目安とした(小学校低学年では生徒の実態を踏まえて時間を調整したクラスもあった)。原稿用紙も、発達段階に合わせてマスの数や大きさを変え、専用の用紙を用意した。小学校1年用が240字(B4版、15字×16行)、小学校2年用300字(B4版、15字×20行)、中学年以上が400字(B4版、20字×20行)の3種類を用意した。担当教員には作文を書かせた後、回収用封筒に実施日時、開始時間と終了時間を書き込んでもらい厳封後、筆者らが速やかに各学校に出向いて回収した。

本調査は富山大学の研究倫理審査を受け承認されている。その際、児童・生徒の個人情報の保護、保護者への研究協力の同意を得ることが必須条件だったことから、各担任を通じて事前に保護者へ依頼文を配布し、調査への非協力も許容している。なお本調査の範囲では調査拒否はなかった。

4.2 調査資料の電子化作業

収集した児童・生徒の作文を電子化する際に、以下の指針を立てた。

(4) 電子化の指針

- ①できるだけ、正確に紙面を再現するよう心がける。
- ②段落初めの一字下げや空欄(意味不明なものも含めて)も正確に記録する。
- ③誤字・脱字、送り仮名の誤り、文字種の違いにも注意して、正確に記録する。
- ④入力後に入力者以外の者が原本と照合し、入力ミスを修正する。
- ⑤個人情報に関わる部分(個人が特定される可能性のある語句や学校名、氏名・渾名等)

は、当該部分を“*”で置き換える。

- ⑥1作文1ファイルで記録し、整理番号を付す。(整理番号から、課題・学年・クラス・性別等が判別できるようにする)

以下のようなテキストデータが電子化された作文の一例である。(誤字脱字等の誤りもできるだけ正確に記録するようにした。枠内の文章の下線は、誤字脱字や文の乱れを示す。)

わたしはしょうらいしょうがっこうのせんせいになーたいです。なぜかというと、まずし
いいくにをまはっておやがいないひとたちやまずしくておかねがなくてしょうがっこうにい
けないこどもたちにおかねも、もらずべんきょうをおしえ、みんなによろこんでおべんきょ
うをおしえたいし、おやにきにってもらえずべんきょうをしたいっていつもやらせてくれ
ずないてるこどもちをやさしくこえをかけて、いっしょにべんきょうをやるおよといって
いっしょにべんきょうをやりたいです。

(「ゆめ」小学校1年 女子)

現在、本テキストデータは範囲を限定して利用を認めている(詳細は巻末の関連HPを参照のこと)。本コーパスの現在の公開範囲は限定的であるが、児童・生徒の個人情報に関する処理を施した後、学術的研究、特に学校現場への還元を目的とした研究に利用する場合での一般公開が可能になるよう協力校に依頼済みである。

4.3 語彙情報の付与とデータの規模

(語彙情報の付与)

本研究の基礎資料となる「児童・生徒作文コーパス」本体は平文テキストとメタデータの形式で構成される。今後研究利用可能な言語学的情報の付与を進めたい。現時点では、自動処理による形態論情報(短単位、長単位)、文節境界情報、構文情報(係り受け情報)を付与している。形態素解析器はMeCab 0.996⁴、辞書はUniDic 2.1.2⁵を使用し、その他長単位と文節はComainu 0.70⁶、構文解析にはCaboCha 0.68⁷を使用した。今後は、本研究で作成した人手で修正を実施したコアデータを基に解析精度の向上を目指す。

(データの規模)

まず、データの基本情報を以下に示す。資料とした作文には「ゆめ(夢)」(夢)と「ぼくの／わたしががんばったこと」(がんばったこと)とがあり、それぞれ別に集計した(資料1)。

4 <http://taku910.github.io/mecab/>

5 <https://osdn.jp/projects/unidic/>

6 <http://sourceforge.jp/projects/comainu/>

7 <https://code.google.com/p/cabochoa/>

資料1 データの基本情報

作文	学年	作文数	文数	語数 (異なり)	語数 (延べ)	字数 (異なり)	字数 (延べ)	修飾語数	述語数	漢字数
夢	小1	91	593	1057	10498	147	18811	737	1424	0
夢	小2	85	692	1385	14895	352	25881	956	1964	1237
夢	小3	83	967	1649	20244	494	34159	1200	2719	3253
夢	小4	68	967	1653	21687	609	34929	1326	2975	4578
夢	小5	74	1206	2052	26151	791	41699	1571	3584	6912
夢	小6	46	715	1566	16470	792	25617	1014	2336	5224
夢	中1	198	3149	3540	71485	1264	110329	4484	10358	26126
夢	中2	189	3125	3620	65576	1357	100416	4187	9630	23910
夢	中3	184	3055	3654	68914	1372	105667	4245	10248	26109
小計①		1018	14469	20176	315920	7178	497508	19720	45238	97349
がんばったこと	小1	62	473	972	8690	203	15445	569	1189	258
がんばったこと	小2	73	860	1281	15768	370	26679	1006	2071	2097
がんばったこと	小3	76	1092	1540	21322	529	35217	1328	2705	4484
がんばったこと	小4	98	1594	2154	34392	736	55610	2056	4754	8209
がんばったこと	小5	94	1494	2288	31767	822	50667	1811	4316	9402
がんばったこと	小6	105	1576	2289	32835	926	51843	1897	4482	11154
がんばったこと	中1	186	3187	3101	67682	1142	105722	3890	9602	25827
がんばったこと	中2	186	3231	3730	66968	1316	104006	4067	9439	26760
がんばったこと	中3	166	3063	3214	65926	1260	102523	3982	9308	27574
小計②		1046	16570	20569	345350	7304	547712	20606	47866	115765
総計 (①+②)		2064	31039	40745	661270	14482	1045220	40326	93104	213114

次に、以下に1作文当たり、1文当たりの語数や字数等の文体指標を示す(資料2)。

資料2 データの文体指標(1作文当たり、1文当たり)

作文	学年	語数/作文数	字数/作文数	語数/文数	字数/文数
夢	小1	115.4	206.7	17.7	31.7
夢	小2	175.2	304.5	21.5	37.4
夢	小3	243.9	411.6	20.9	35.3
夢	小4	318.9	513.7	22.4	36.1
夢	小5	353.4	563.5	21.7	34.6
夢	小6	358.0	556.9	23.0	35.8
夢	中1	361.0	557.2	22.7	35.0
夢	中2	347.0	531.3	21.0	32.1
夢	中3	374.5	574.3	22.6	34.6
平均①		294.2	468.8	21.5	34.7
がんばったこと	小1	140.2	249.1	18.4	32.7
がんばったこと	小2	216.0	365.5	18.3	31.0
がんばったこと	小3	280.6	463.4	19.5	32.3
がんばったこと	小4	350.9	567.4	21.6	34.9
がんばったこと	小5	337.9	539.0	21.3	33.9
がんばったこと	小6	312.7	493.7	20.8	32.9
がんばったこと	中1	363.9	568.4	21.2	33.2
がんばったこと	中2	360.0	559.2	20.7	32.2
がんばったこと	中3	397.1	617.6	21.5	33.5
平均②		306.6	491.5	20.4	32.9
平均 (①, ②)		300.4	480.2	20.9	33.8

資料1から、本研究で使用了資料（「生徒・児童作文コーパス」）の規模は、66万語（短単位）規模であることが分かった（詳細な分析は5節を参照のこと）。資料2から、文章量（1作文当たりの語数）を知ることができる。両作文共に学齢の進行に従って文章量が増加していくことが分かる。概括的な捉え方になるが、学齢毎に多少の揺れがあり、題の違いがあることを考慮に入れたとしても、小1～小3辺りで文章量は爆発的に増加し、その後は緩やかに増加していくようである。便宜的に学齢別に分析したが、本資料は年2回の調査を課しているので、6カ月毎の変化を追ってより詳細な記述を行うことも可能である。次に1文当たりの語数を見てみると、6月頃に作成された「夢」で小1と小6の差が5.3語、12月頃に作成された「がんばったこと」で2.4語とその差は縮まってきており（全体の平均は20.9語）、作文全体の字数（文章量）を比較すると、1文当たりの字数は学齢によってさほど差がないことが分かる。

なお、本コーパスの各種データは、基本的に機械解析に基づいている。小学校各学年1クラス分の作文を人手で修正し、それを正解データとしてMeCab / UniDicによる形態素解析の精度を評価したところ、F値が0.89～0.98だった。

5. 調査の目的とデータの概要

本節では、3. 1節（1）で挙げた各調査課題について調査方法と調査結果を示し、考察を行った。なお本節での各調査課題の結果を示した資料に関しては添付のCD-ROMを参照されたい。

5.1 頻出漢字と当該漢字を使用した語彙リスト

（目的）

既に述べたように、児童らの漢字使用の実態を調査し、作文能力を測るには漢字毎の使用頻度を調査するだけでは不十分である。学年別配当の教育漢字の使用状況や語彙の使用頻度を知ることが必要である。よって次の調査を実施した。

（5）学年別配当漢字使用頻度表（資料3）

（6）学年別配当漢字別頻度表と語例（資料4）

（資料）

資料3：学年別配当漢字に関して学齢別の使用頻度を示した。

資料4：学年別配当漢字の使用頻度を示し、併せて当該漢字を含む語の異なり語数、延べ語数も示した。また、文中で当該漢字を使用して書かれた語例を挙げた（上位10語まで）。

（資料の概要と活用）

資料3は児童生徒の書く文章中で、どのような漢字が、どの学齢で、どの程度使用されているのかを示した調査結果である。当然ながら、児童らは学年別配当に沿って漢字を習得し使用するわけではなく、字毎、語毎にその使用傾向は異なるはずであり、その実態を映し取ることを目的とした資料である（2. 2節も参照のこと）。次に、資料4では当該漢字を含む語の異なり語数、延べ語数を示し、さらに具体的な語例も挙げた。漢字の使用頻度だけからは見えない使用実態を示す資料であり、漢字指導や教材作りで活用されることが望まれる。

なお資料3表末の「漢字以外」は、本コーパスに含まれる漢字以外の全ての文字（仮名、英数字、句読点、括弧、記号、空白等）の合計を示す。

5.2 主要品詞別頻出語彙リストおよび当該語彙の漢字使用率

(目的)

作文の使用語彙の品詞構成を知ることによって、児童らの書く文章の分析を精緻化することができる。例えば、児童らの書く文章では品詞毎に漢字使用率に差異がある可能性が予測される。また、従来の教育基本語彙との比較や他の文章ジャンルの品詞構成比との比較も今後の課題として興味深い。よって次の調査を実施した。

(7) 品詞別語彙頻度表 (資料5)

(8) 品詞別語彙漢字使用率表 (資料6)

(資料)

資料5：品詞別に語彙の使用頻度を示した。併せて学齢毎の別も示した。

資料6：当該語彙毎の漢字使用率を示した。併せて学齢毎の別も示した。

(資料の概要と活用)

資料5では学齢別の使用頻度も示した。今後、学齢毎の品詞構成変化や他の文章ジャンルとの比較を実施する可能性があり、その点にも配慮している。次に、資料6では品詞別語彙の漢字使用率を調査した(当該語彙使用数に占める漢字が使用された語の使用数の比率)。児童らの漢字使用の基準として品詞の別が関わっている可能性がある。おそらく接続詞や副詞、一部の代名詞等では漢字の使用率が低い。本資料では、動詞、名詞、形容詞、形状詞(いわゆる形容動詞語幹)の主要4品詞を抽出して、品詞別の漢字使用率の調査ができるようにした。

なお漢字使用の有無に関しては、当該語彙の中に漢字が含まれていれば、漢字が使用されている(漢字書きがされている)と判定した。例えば、「上り詰める」という語の場合、「上り詰める」「上りつめる」「のほり詰める」のいずれも漢字が使用されていると判定した(以下の資料でも同様に処理した)。

5.3 和語のリストおよび当該語彙の漢字使用率

(目的)

当然ながら漢字使用され易い語種として漢語や漢語を含む混種語が挙げられる。一方で外来語や記号等は基本的に漢字で書かれることはない。問題となるのが和語の漢字使用率である。これまで述べてきたように、語別に漢字使用率の傾向が異なるのだとしたら、和語を一括りにして安易な分析ができるものではない。資料6で示したように品詞の別によってもその差異があることが予測される。よって次の調査を実施した。

(9) 和語漢字使用率表 (資料7)

(資料)

資料7：和語(動詞、名詞、形容詞、形状詞⁸)に着目して、学齢別に語彙の使用頻度と、漢字使用率を示した。

(資料の概要と活用)

最も漢字使用の傾向に幅があると考えられる和語の状況を検討することによって、児童らが書く作文の平均的な漢字使用の傾向を明らかにすることができる。学校現場では、「習った漢字は全て使って作文を書きなさい」といった類の指導がよく見られるが、詳細に検討していけ

8 「へろへろだ」「ぐちゃぐちゃだ」や「軽やかだ」「賑やかだ」等が和語として分類される。

ば「ある（有る）」や「なる（成る）」のようにむしろ仮名書きの方が自然に感じられる語も少なくない。本調査によってこの語を抽出し、リスト化することができれば、今後の漢字指導や作文指導（特に推敲の場面⁹）等で有効な指針を示すことが可能になる。

また、資料7では漢字使用率を20%未満、20%以上～80%未満、80%以上に分け、3段階でランク付けを行った。基本的な捉え方は、漢字を使用して書かれるか、書かれないかであるので、20%以上～80%未満の語は揺れのある書き分け意識の低い語であり、今後の動向に注意すべき語ということになる¹⁰。

5.4 学齢別対中学3年比の漢字使用率（飽和率）

（目的）

既に述べたように、児童らの漢字使用の実態を調査し、作文能力を測るには漢字毎の使用頻度を調査するだけでは不十分である。漢字学習の最終的な到達状態に対する、発達段階ごとの到達状況を知る必要がある。そこで、飽和頻度と飽和率という二つの尺度を設定した。飽和頻度とは、漢字学習の最終的な到達状態を示す指標である。本調査では、中3年次までには小学校配当漢字の学習が一通り定着するものと仮定し、中3年次における100万字当たり文字頻度（CPM: Character per million）を飽和頻度とした。「飽和率」とは、発達段階ごとの到達状況を示す指標であり、各学齢におけるCPMを飽和頻度（中3年次におけるCPM）で割った数値である。これを踏まえて次の調査を実施した。なお、他学年で使用があっても中3年次の使用がない場合、飽和率は「値なし」（空欄）となっている。

(10) 学齢別漢字飽和率表（資料8）

(11) 配当学年前後の漢字飽和率表（資料9）

（資料）

資料8：学齢別CPMを中3年次の飽和頻度で割った値（飽和率）を示した。

資料9：配当学年前後の当該漢字の飽和率とその差を示した。

（資料の概要と活用）

既に5.3節で一部述べたが、語別に漢字使用率が異なっており、資料8を利用して、例えば習得済みなのになかなか漢字書きされない語を調査することもできる。一方で、総体的な漢字指導・語彙指導という観点から見た場合、個別的な事例に囚われずに児童らの漢字使用の傾向を全体的に捉える必要もある。宮城・今田（2015b）は、全体的な傾向として、配当学年より1年遅れて60%程度の飽和率が得られると指摘している（2年遅れてほぼ70%に達する）。また、資料9では漢字使用率の定着の変異を推定する調査を実施した。学年別配当の漢字の飽和率（または使用率）は学習直後に暫時的に上昇して下降する語（暫時使用型）や、また学習時期から少し遅れて漸進的に上昇する語（漸進使用型）がある。もちろん、学習直後に安定的に使用されるようになる語も存在する（即時使用型）。その場合、習得時期と使用（活用）開始時期が

9 筆者が作文指導に関する講演を行った際、推敲についての質問が少なくない。よく聞かれるのが、『です』や『ます』の乱れを直し、習った漢字はできるだけ使うように…』という指導法であるが、おそらく教師も文言通りに既習漢字は全て漢字書きに直しているわけではないと思われる。

10 全体の19.5%の語がこのカテゴリに含まれる。「揺れ」が生じる原因は、語毎に異なり様々であるが、漢字使用が無批判に先行したため、学年が進むと揺り戻しが起こり、逆に漢字使用率が低下する語もあると考えられ、今後詳細に検討しなければならない。

ほぼ一致しているということになる。このような漢字別の違いを捉えた指標が「飽和率差A」(次年度－配当年度)、および「飽和率差B」(次年度－前年度)という指標である。飽和率差Aが大きな正の値を取れば「漸進使用型」の可能性があり、値が微小であれば「即時使用型」の可能性もある。飽和率差Bは大きな正の値を取るものが多いが、この差が微小な語は配当学年より早く習得・使用され、学習時期の前倒しが起こっている語の可能性もある。このような必ずしも学年別配当の時期に従わない児童らの漢字習得の傾向を把握しておくことにより、児童らの漢字使用能力に即した適切な指導・支援ができるようになる。

5.5 学齢別文章量と漢字使用率

(目的)

平均的な児童らがどのような内容の文章を書くことができるのかは、現在のところ教師の経験に頼るしかない。作文の出来に関する評価¹¹は、主観的な基準でなされることが少なくない。これを完全に否定するものではないが、例えば平均的な作文の特徴を数量的に示すことができれば、それを併用して、ある種の作文の文体(型)を評価することは可能であろう。よって次の調査を実施した。

- (12) データの基本情報(文数や語数、延べ語数、異なり語数等)(資料1)
- (13) 学齢・作文別の文体指標表(文長)(資料10)
- (14) 学齢・作文別の文体指標表(文の特性)(資料10)
- (15) 学齢・作文別の文体指標表(漢字使用率)(資料10)

(資料)

資料1: コーパス全体の文数や語数、延べ語数、異なり語数等を示した。

資料10: 学齢別・作文別に1文・1作文当たりの文章量や文の特性、漢字使用頻度等を示した。

(資料の概要と活用)

資料1では「児童・生徒作文コーパス」の基本情報を示した。コーパスの規模やデータ量による妥当性の検証等に利用されたい。また、資料10では、文章の長さ、文の長さ、MVR (Verb Modifier Ratio¹²)、KPM (Kanji Per Million) 等、本コーパスがどのような文体的特徴を有しているかを示す各種の指標を示した。MVRは「(形容詞+形状詞+副詞+連体詞)/動詞」の値を表し、文がどれくらい描写的に偏っているのかを推し量る指標である。KPMは「100万×漢字数/字数」を表し、文章中の漢字使用頻度を示す指標である。それぞれのデータは、本コーパスがどのような文体的特徴を有しているのかを推定するための指標である。

実際の作文指導の場面に於いては、例えば類題の作文を書かせる場合、本調査の結果と比較して、1作文当たりの文章量の多寡や文章中の漢字使用率の高低を測るといった評価方法が可能になる。今後、これらの文体指標を利用して対象となる学習者の作文と発達段階に応じた平均的な作文とを簡便に比較し、評価するシステムの構築を目指している。

11 作文コンクールの評価基準に於いてさえ「気持ちをはっきりと書いているので良い」等とする場合があると聞く。主観的な評価に偏らないように、今後客観的な文体指標に基づく基準も取り入れていく必要があると考えられる。

12 樺島・寿岳(1965)で示された指標。

5.6 資料の評価

既存の漢字テストが理想的な学習者を想定した漢字習得修得の達成目標であるとするれば、本調査で明らかになった指標はある児童らの標準的な実態であり、既習漢字活用の達成目標を示すことになると考えられる。また、テストによる漢字習得調査からは、児童らが習得済みの漢字を実際に使用するかどうかまでは測ることは難しいが、本調査で用いた「児童・生徒作文コーパス」によって、使用実態にかなりの程度迫ることができたと考えられる。

資料1～資料10の語彙表が本研究の最も大きな成果であり、児童らの語彙使用・漢字使用の実態を捉えていると思われる。本研究の成果をどう使うか、どう読み取るか、何を知りたいか等の探求すべき課題は個々に異なるので、本報告書では、各語彙表を作成した目的とその概要を述べるに留めておく。本資料は、現場での教材開発や漢字指導の方針、児童らの作文を対象とした様々な調査を実施する時の基礎資料としての活用が望まれる。なお、本研究の成果である語彙表類の資料は、付録のCD-ROMに全データをExcel形式で収録してある。

6. 本研究の成果と展望

6.1 本研究の成果

本研究では、各種語彙情報を付した「児童・生徒作文コーパス」を基礎資料として、現代の児童らの作文に於ける語彙使用・漢字使用の実態を明らかにしたことが最大の成果である。調査の基礎資料が先行研究と異なり、同一条件下で教師や保護者等大人の手が入っていない、生の作文を収集したことであることにも価値があると考えている。現在利用可能な児童らの作文能力、特に語彙使用・漢字使用の実態を映した資料が存在しない状況では、本研究の成果は相応の意味があり、今後、児童らの学齢別の漢字使用の実態を詳説していくための足掛かりとしては、十分に要求を満たすことができたと考える。

また、本研究の調査によって、漢字使用の実態だけではなく、使用語彙の実態、作文の表現力や児童らの言語の発達の実態に関する複数の新たな知見も見いだされた。今後、言語資料として本コーパスを活用することで、広く文章作成に関わる能力を解明していきたい。研究の範囲は、語彙使用・漢字使用に留まらない。例えば、「子どもたちの記号や表記ルール適応度についての研究」、「学齢別にどの程度の分量を書くのか、また標準的な1文の長さはどの程度なのか、だたら文の修正はどうすれば良いのか等についての研究」、「学齢に応じた慣用句、感情・感覚表現の使用についての研究」、「学齢別の書き出しやまとめの表現についての研究」、「はじめ-なか-おわりの型の作文の論理的適切性についての研究」、「接続表現や文末表現の多様性についての研究」等に発展させることができる。これらは現場の教師達が切望しているにもかかわらず、1950年代の先行研究が先鞭を付けながら、これまでほぼ客観的な手法での研究がなされてこなかった。今後本研究で構築した資料や調査結果が児童・生徒の作文能力に関する実証的研究の突破口、即ち児童らの作文能力の実態を解き明かす手掛かりとなることが期待される。

本節の最後に、本研究での成果を総括すると以下のように集約することができる。

(16) 本研究の成果

- ①各種語彙情報を付した真正性の高い「児童・生徒作文コーパス」を構築した。
- ②学齢別に児童らの標準的な漢字使用能力の推定を行った。

- ③当該コーパスを資料として、児童らの語彙および漢字使用の実態を調査した。
- ④品詞別や語種別（和語）に注目して、語彙の漢字使用率を調査した。
- ⑤必要な観点から自由に分析が行えるよう語彙表を作成した。
- ⑥上記の全データを教育・研究用に公開した。

6.2 今後の課題

本研究の限界は、主たる調査対象の児童・生徒らが国立大学附属学校という特殊な環境にあるという点にある。「児童・生徒作文コーパス」の汎用性を追求するのであれば、公立小・中学校のデータや、公立でも地方部と都市部を比較する観点を踏まえた児童・生徒の作文データの収集も必要であろう。今後は一般の公立小・中学校での継続調査も視野に入れながら資料の拡充を図っていくことになる。もう一点、本報告書によって、これまで捉えられていなかった児童・生徒の漢字使用の実態を周知することができた。一方で現場の教育利用を考えた場合、本報告書で示された資料だけではなく、それを具体的に活用するための提案も必要であると考える。現段階では達成できていないが、本報告書に続いて、本研究で得られた資料を基に語彙・漢字学習システムを構築し、それを踏まえた学習教材の開発を企画している。これら一連の研究が本報告書と合わせて現場で活用して漢字指導・作文指導に資するところがあれば筆者らにとって望外の喜びである。

謝辞

本研究では、「平成27年度 漢字・日本語教育研究助成制度」のご助力によって研究が順当に進められ、十分な成果を得て目標を達成することができた。本研究に助成いただいた公益財団法人 日本漢字能力検定協会の関係者各位に改めて感謝申し上げます。また、本研究の基礎資料の基となる作文調査にご協力いただいた国立大学附属小・中学校の児童・生徒の皆さん、同教職員各位に記して感謝申し上げます（個人情報保護の観点から学校名・個人名を挙げることは控えさせていただきます。）。

引用文献・参考文献

- 阿辻哲次・棚橋尚子（監修）（2015）『小学校国語 漢字指導の方法：確かな定着と活用のために』、光村図書。
- 石井健介（1996）「児童生徒作文の計量的分析の試み」『学芸国語国文学』28、東京学芸大学国語国文学会、pp.82-90。
- 市川孝（1977）「小中学生の作文の表現」『松村明教授還暦記念 国語学と国語史』、明治書院、pp.1039-1058。
- 樺島忠夫（1953）「文の長さについて一条件との相関の分析一」『国語学』15、国語学会、pp.21-30。
- 樺島忠夫・寿岳章子（1965）『文体の科学』、綜芸社。
- 河内昭浩（2015）「国語教育のための「常用漢字表」語例の検討」『第7回コーパス日本語学ワークショップ予稿集』、pp.113-122。（https://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop_no7_papers/JCLWorkshop_No7_web.pdfよりダウンロード可能）

- 国立国語研究所（編）（1964）『小学生の言語能力の発達』、明治図書。
- 国立国語研究所（編）（1989）『児童の作文使用語彙』、東京書籍。
- 国立国語研究所（編）（2009）『教育基本語彙の基本的研究—増補改訂版—』、明治書院。
- 阪本一郎（1995）『読みと作文の心理』、牧書店。
- 鈴木一史（2011）「作文コーパスからみる生徒の使用語彙」『特定領域研究「日本語コーパス」平成22年度公開ワークショップ（研究成果報告会）予稿集』、pp.343-350. (http://www.ninjal.ac.jp/corpus_center/bccwj/doc/workshop/JC-G-10-02.pdfよりダウンロード可能)
- 田中牧郎（2011）「語彙レベルに基づく重要語彙リストの作成—国語政策・国語教育での活用のために—」田中牧郎他『言語政策に役立つ、コーパスを用いた語彙表・漢字表等の作成と活用』、pp.77-87. (http://pj.ninjal.ac.jp/corpus_center/bccwj/doc/report/JC-P-10-01.pdfよりダウンロード可能)
- 田中牧郎（編）（2014）『コーパスと国語教育』、朝倉書店。
- 長友和彦（編）（1997）『児童・生徒・学生及び日本語学習者の作文能力の発達過程に関する研究』平成8年度文部省科学研究費補助金基盤研究（B）（2）研究成果報告書。
- 日本教材文化財団（2010）「生きる力が育つ漢字の学習」(<http://www.jfecr.or.jp/publication/pub-data/kanji/index.html>よりダウンロード可能)
- 富士原紀絵・宮城 信・松崎史周（2016）「児童生徒作文の基礎的研究—児童生徒作文コーパスの構築と活用—」『子ども学研究紀要』4、お茶の水女子大学、pp.9-20.
- ベネッセ教育総合研究所（2007）「小学生の漢字力に関する実態調査」(<http://berd.benesse.jp/shotouchutou/research/detail1.php?id=3198>よりダウンロード可能)
- 宮城 信・今田水穂（2015a）「『児童・生徒作文コーパス』の設計」『第7回コーパス日本語学ワークショップ予稿集』、pp.223-232. (https://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop_no7_papers/JCLWorkshop_No7_web.pdfよりダウンロード可能)
- 宮城 信・今田水穂（2015b）「『児童・生徒作文コーパス』を用いた漢字使用能力の推定」『第8回コーパス日本語学ワークショップ予稿集』、pp.47-56. (http://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop_no8_papers/JCLWorkshop_No.8_web.pdfよりダウンロード可能)
- 森岡健二（1954）「教科書文章の難易調査」、『言語生活』30、筑摩書房、pp.35-39.

関連HP

「作文を支援する語彙文法的事項に関する研究プロジェクト」（代表：矢澤真人）

<https://sites.google.com/site/sakubunshienproject/>

資料

添付のCD-ROMに収録した。同資料は以下の日本漢字能力検定協会webサイトからもダウンロード可能である。

「漢検 漢字・日本語教育研究助成制度（平成27年度）」