

## 『児童・生徒作文コーパス』の設計

宮城 信 (富山大学人間発達科学部)<sup>†</sup>  
今田 水穂 (文部科学省初等中等教育局)

### Design of a Written Composition Corpus of Japanese Elementary and Junior High School Students

Shin Miyagi (University of Toyama)

Mizuho Imada (Ministry of Education, Culture, Sports, Science and Technology)

#### 要旨

本研究では、児童・生徒の作文能力の実態を明らかにするため、子どもたちが作成した生の作文を原本とした『児童・生徒作文コーパス』の構築を進めている。本コーパスは、協力校4校(小学校2校、中学校2校)9学年の全クラスを対象に3年間にわたって作文資料を収集・電子化するもので、最終的に300万形態素規模のコーパスになる予定である。同時期に同条件(題や作文時間の指定)で作文を作成させることによって資料の均質性を保証している点、複数年の継続調査により同一児童・生徒の作文能力の変化を追跡できる点の特徴である。本コーパスを利用した研究によって、児童・生徒の学齢別の作文能力の実態や発達を明らかにし、現場の教員の作文指導の手本となる資料の作成を目指す。また、本コーパスの構築と合わせて、独自の検索システムの開発も同時に行っている。現段階の検索システムは、単純な文字列検索が行えるに留まるが、今後システムを更新して、高度な検索処理をできるようにする。本発表では、コーパスの基本的な設計方針、内容の概要、検索システムの紹介を行い、コーパスを活用した研究の展望を述べる。

#### 1. はじめに

近年、コーパスを利用した言語研究が盛んになってきている。国語教育学研究でも子どもたちの書いた作文を資料とした作文能力の実態調査や指導法の開発などが行われている。しかしながら、後者の資料となる児童・生徒の作文でコーパスとして利用可能なものは、資料の収集や公開の難しさから質量ともに十分ではなく、十分な研究環境が整っているとは言いがたい。そのため、本研究では小中学校の児童・生徒の作文を3年間に亘って収集し電子化する大規模な作文コーパスの構築を進めている。本発表では現在構築中の『児童・生徒作文コーパス』(以下、「児童作文コーパス」と略す)の目的と概要を説明し、今後の研究の展望を示す。

#### 2. 児童作文コーパスの必要性

小中学校における現在の作文指導は、多くの場合子どもたちの書いた文章に教員が手を入れて書き改めさせるという方法で行われている。この指導法には次の2点で問題がある。

- (1) 文章の修正(指導)が教師個人の語感によって主観的になされていること。
- (2) 子どもによる作文の推敲が、教師による書き換え例を丸写しすることに留まり、なぜ直すのか、他にどのような表現があるのかなどの検討が行われていないこと。

---

<sup>†</sup> miyagi@edu.u-toyama.ac.jp

(したがって、子ども自身の作文推敲能力が育たない)。

これらの問題は、教師個人のひいては教育現場全体における経験知の不足、またそれを補い補正していく資料の不足によるものと考えられる。

作文指導には特定のマニュアルがあるわけではなく、現場依存的である。また、当然ながら子どもたちの作文能力は個々で異なっている。ベテラン教師は、勘を働かせて上手に子どもたちを誘導し、それなりの文章に推敲させることができるが、経験の浅い教師は、このような技術を持たないため、ベテラン教師の助言や手本となる用例集などの資料が必要になると考えられる。ここでいう手本となる資料は以下の要件を満たす必要がある。

- (3) 子どもたちの発達段階を考慮した、相対的な基準を提示できるものであること。
- (4) 文章を特定の型に揃えることを目標とするものではないこと。(言葉狩りを推奨するものではない。)
- (5) 子どもたちが理解できる理由で説明がなされること。

以上の要件を満たす資料を構築するためには、まず、発達段階に応じた子どもたちの書く作文の実態(語彙や文構造、段落構成など)を知る必要がある。そのため、本研究では、子どもたちの書く作文の実態を明らかにし、それに基づいて指導資料を開発するための基礎的な研究資料として児童作文コーパスを構築する。

### 3. 作文コーパスの設計と基本方針

#### 3. 1 作文コーパスの特徴

本コーパスは、調査協力校4校(小学校2校、中学校2校)9学年(小学1年~中学3年)の全児童・生徒に作文課題を課し(作成時間は小学校40分、中学校45分)、収集して電子化したものである。作文は「夢」などのテーマ(タイトル)のみを提示し、教員は一切の事前指導を行わない。電子化は以下の指針に従って行う。

#### ○電子化の指針

- ・できるだけ、正確に紙面を再現するよう心がける。
- ・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。
- ・誤字・脱字、文字種の違いにも注意して、正確に記録する。
- ・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。
- ・個人情報にかかわる部分(個人が特定される可能性のある語句や学校名、氏名・渾名など)は、当該部分を“\*”で置き換える。
- ・1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・性別などが判別できるようにする)

個人情報保護の理由から、収集した作文原本は非公開とし、テキストデータは範囲を限定して利用を認める。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交渉中である。

### 3. 2 作文コーパスの構成

本コーパスは本文テキストとメタデータで構成される。メタデータは本文テキストには含まず、ファイル名と紐付けて別に管理する。メタデータは以下の項目を含む。

作文課題の属性	課題 ID、実施年度、テーマ (タイトル)
執筆者の属性	著者 ID、学校 ID、学年、クラス、性別

作文課題の実施、収集は年 2 回行い、3 年間継続する。2015 年 1 月現在、2014 年度分の課題 2 回について実施済みであり、電子化作業を進めている。

表 1 作文課題の実施計画

年度	2014		2015		2016	
課題	課題 1	課題 2	課題 3	課題 4	課題 5	課題 6
進捗状況	実施済	実施済				

最初の作文課題 (課題 1) について、48 クラス分の作文原稿の収集と、23 クラス分のテキスト入力、11 クラス分のチェック作業が完了している。テキスト入力済みの 23 クラス分のデータについて、文分割と形態素解析処理を行い、文数、形態素数、文字数 (改行文字を除く) を集計した結果を以下に示す。形態素解析処理には MeCab 0.996<sup>1</sup> と UniDic 2.1.2<sup>2</sup> を使用した。学年別集計は 5.1 節を参照されたい。

表 2 課題 1 の概要 (23 クラス分)

学年	クラス数	作文数	文数	形態素数	文字数
小 1～中 3	23	813	11046	237940	378652

23 クラス分のコーパスの形態素数が約 24 万なので、48 クラス分で約 50 万形態素、6 回の作文課題で最終的に 300 万形態素程度の規模のコーパスになる見込みである。

### 3. 3 既存コーパスとの比較

児童・生徒の書き言葉を対象としたコーパスは全国の地域文集 10 年分を収集し約 47 万形態素規模のコーパスを構築した国立国語研究所(1989)などを例外として従来あまり多くなかったが、近年報告が増えている。永田他(2010)は小学 5 年生 81 人の読書ブログを 8 カ月間記録した約 4 万形態素規模のコーパスで、ブログの更新履歴を追跡できる点、一般公開されている点の特徴である。坂本(2010)は全国の小学校 265 校の Web ページで公開されている児童作文を収集した 123 万形態素規模のコーパスである。学校名、県名、学年、性別などの情報が確認できる限り付与されており、一部については著作権処理が完了しているという。鈴木他(2011)は中等教育学校の 1 年から 5 年(中 1～高 2)の冬休みの宿題作文を記録した約 25 万語規模のコーパスである。藤田他(2012)は神奈川県内の小学校 9 校で 2 回に分けて収集した作文 672 編からなるコーパスである。表記や文法の誤りなどの指摘事項と

<sup>1</sup> <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

<sup>2</sup> <http://sourceforge.jp/projects/unidic/>

評価が付与されている点の特徴である。

表3 既存の作文コーパスとの比較

コーパス	国語研(1989)	永田他(2010)	坂本(2010)	鈴木他(2011)	藤田他(2012)	本コーパス
形態	作文	読書ブログ	作文	作文	作文	作文
形態素数	474,243	39,269	1,234,961	249,918	不明	3,000,000
調査対象	小1～小6	小5	小1～小6	中1～高2	小4	小1～中3
調査期間	10年	8カ月	2年	1カ月	1年	3年
収集方法	文集収集	活動記録	Web収集	課題調査	課題調査	課題調査
備考		公開済	著作権処理済 (一部)		誤用、評価情 報つき	

本コーパスはコーパスの規模が約 300 万形態素と既存の作文コーパスと比べても最大規模である点、義務教育課程（小1～中3）の全体をカバーしている点、同一の調査対象に対して 3 年間継続して調査を行う点などが特徴である。一方で、特定の学校のみを調査対象としているため、必ずしも全国の児童、生徒作文全体に対する代表性を保証しているわけではない点、構築したコーパスを研究目的で公開し、共有する方法が確定していない点などに課題が残る。

#### 4. 児童作文検索システム

本コーパスの構築に合わせて、「児童作文コーパス」のデータを検索するシステム（以下、「検索システム」とする）を開発する。検索システムを独自に開発する利点は、コーパスの仕様変更（5 節を参照）に合わせて、適切な検索が実行できるように検索システムを改修することができる点である。また、本コーパスは教育現場での利用も視野に入れており、現場の教員が手軽に検索を行えるインターフェイス設計を指向している。以下、検索システムの現在のバージョンにおける概要を示す。

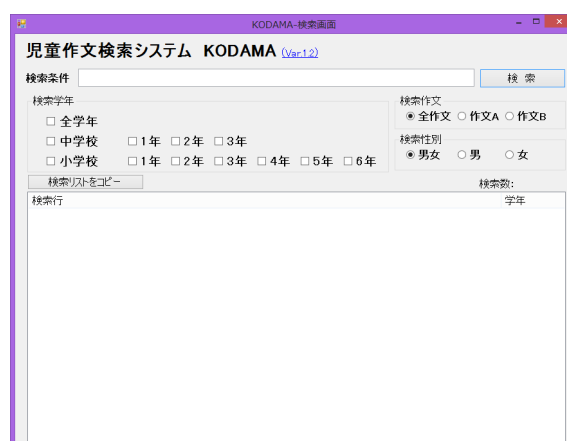


図1 基本操作画面

検索システム（図1）には以下のような検索項目がある。組み合わせて、検索したい作文の条件を設定する。

検索学年 (全学年／中学校／小学校／学年指定)  
 検索作文 (全作文／生活作文／意見文)  
 検索性別 (男／女)

現在のバージョンは単純文字列検索である（正規表現には対応していない）。検索条件に文字列を指定すれば結果が得られるようになっている。

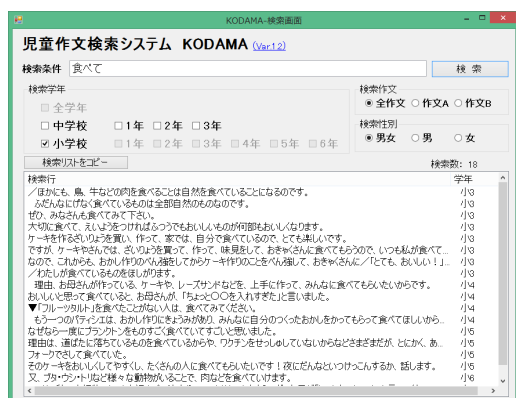


図2 検索結果（一覧表示）

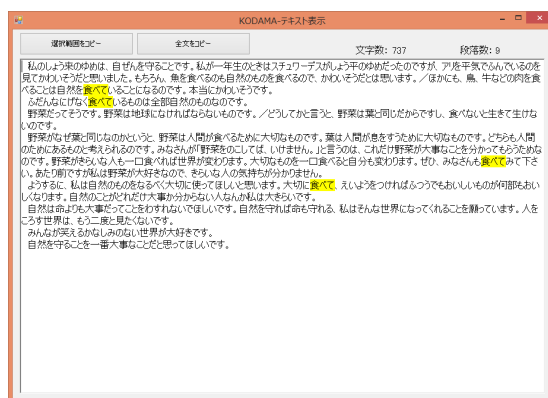


図3 検索結果（個別・全文表示）

検索条件を「食べて」に指定して検索すると、検索された一文（句点から句点までの文字列）が一覧表で表示される（図2）。合わせて、各文の横には作文した児童・生徒の学年も付される。また、一覧左上の「検索リストをコピー」をクリックすると、結果一覧を excel に直接貼り付けることができる。

結果一覧の任意の文をクリックすると、当該の文が検索された作文の全文が表示される（図3）。作文全文の中の検索した文字列は、例の「食べて」のように黄色で反転表示される（複数の候補がある場合、すべての文字列が対象となる）。画面左上の「選択範囲をコピー」や「全文をコピー」をクリックすることによって、excel や word などに、当該の文章を直接貼り付けることができる。また、画面右上に、当該作文の段落数（改行記号の数）や文字数（記号も1字と数える）も表示される。

## 5. 作文コーパスの展望

### 5. 1 作文コーパスの今後の展開

本コーパスは平文テキストとメタデータの形式で構築しているが、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）の付与を試行している。課題1の23クラス分のデータについて、各種情報を学年別に集計した結果を表4に示す。前述の MeCab 0.996、UniDic 2.1.2 の他、長単位と文節は Comainu 0.70<sup>3</sup>、構文解析は CaboCha 0.68<sup>4</sup>を使用している。

<sup>3</sup> <http://sourceforge.jp/projects/comainu/>

<sup>4</sup> <https://code.google.com/p/cabocha/>

表4 課題1の学年別集計(23クラス分)

	小1	2	3	4	5	6	中1	2	3	計	
作文数	104	100	100	80	68	110	74	68	109	813	
段落数	227	281	396	365	365	646	362	313	536	3491	
文数	649	791	1263	997	1030	1919	1214	1334	1849	11046	
文節数	4206	5342	9573	7577	7629	15988	10340	10360	16562	87577	
長単位数	10111	13541	23051	18658	18636	38562	24487	24579	39098	210723	
短単位数	11271	15149	26070	20953	21123	43680	27600	27643	44389	237878	
品詞	名詞	2024	2687	5437	3880	4132	8836	5641	5412	9188	47237
	代名詞	220	393	638	537	528	1055	682	714	1136	5903
	形状詞	149	204	260	235	286	518	359	398	629	3038
	連体詞	70	76	205	195	214	452	339	331	544	2426
	副詞	284	384	557	451	479	924	554	593	813	5039
	接続詞	30	55	82	91	103	197	136	139	236	1069
	感動詞	130	63	58	90	53	75	26	31	29	555
	動詞	1400	1893	3402	2993	2888	6205	4021	4131	6559	33492
	形容詞	261	255	475	383	395	719	501	520	785	4294
	助動詞	1643	1932	3129	2494	2412	4809	3097	3154	5009	27679
	助詞	3126	4199	7450	5964	5857	12654	8078	8217	13186	68731
その他	1934	3008	4377	3640	3776	7236	4166	4003	6275	38415	
語種	和語	8545	10579	18756	15212	14740	30956	19950	20438	32392	171568
	漢語	982	1449	2645	2012	2418	5189	3352	3178	5779	27004
	外来語	167	455	833	460	472	854	423	294	545	4503
	混種語	71	106	154	109	135	238	164	180	282	1439
	固有名詞	74	93	217	91	148	224	149	84	195	1275
	記号	1428	2432	3415	3065	3164	6189	3556	3457	5188	31894
	その他	4	35	50	4	46	30	6	12	8	195
文字数	20070	26089	43987	33767	33701	67827	42470	42432	68309	378652	
文字種	ひらがな	18008	20415	32518	24279	22094	43550	26523	27400	43310	258097
	カタカナ	497	1765	3313	1768	2115	3485	1980	1266	2353	18542
	漢字	128	1237	4153	4315	5924	14018	10121	10081	17144	67121
	その他	1437	2672	4003	3405	3568	6774	3846	3685	5502	34892

これらの情報を用いると、言語単位の比、品詞や語種の比、文字種の比などについて、学年別に調べることができる。例として、作文あたりの平均文数、文あたりの平均短単位数(平均文長)、MVR<sup>5</sup>、漢語比率、漢字比率を表5に示す。学年が上がるにつれて平均文数、平均文長、漢語比率、漢字比率などが増加すること、MVRが減少することなどが観察できる。

表5 学年別の言語単位、品詞、語種、文字種比率

	小1	2	3	4	5	6	中1	2	3	平均
文/作文	6.24	7.91	12.63	12.46	15.15	17.45	16.41	19.62	16.96	13.59
短単位/文	17.37	19.15	20.64	21.02	20.51	22.76	22.73	20.72	24.01	21.54
MVR	0.55	0.49	0.44	0.42	0.48	0.42	0.44	0.45	0.42	0.44
漢語/短単位	0.09	0.10	0.10	0.10	0.11	0.12	0.12	0.11	0.13	0.11
漢字/文字	0.01	0.05	0.09	0.13	0.18	0.21	0.24	0.24	0.25	0.18

より高度な言語学的情報としては、文の成分(主語、述語、修飾語など)、係り受けの

<sup>5</sup> (形状詞+連体詞+副詞+形容詞)/動詞で計算した。

種類（並列など）、節の種類などの文法情報や、誤用情報などの付与がある。文法情報は、文の複雑さを評価するために必要となる。誤用情報は、発達段階別の誤用実態の分析や指導資料の開発のために必要となる。こうした研究の展望については、次節を参照されたい。

## 5. 2 作文コーパスを用いた研究の展望

現時点での児童作文コーパスおよび検索システムの概要は以上である。児童作文コーパスによって明らかにされる子どもたちの作文活動の実態と研究の展望について言及する。

### ① 学習漢字の使用の実態

子どもたちが作文で使用する漢字は、多くの場合授業で学習済みのものであると推測される。表 6 は学年別の使用漢字を集計し、1 万文字あたりで示したものである。

表 6 学年別使用漢字（1 万文字あたり）

	小1	2	3	4	5	6	中1	2	3	平均
1年配当漢字	12.5	319.3	376.5	379.1	434.7	472.2	521.3	497.7	515.2	430.7
2年配当漢字	15.4	154.5	389.2	478.3	565.3	605.2	668.2	683.7	699.3	535.8
3年配当漢字	9.0	0.4	147.1	258.5	382.2	419.4	461.5	409.6	435.5	326.0
4年配当漢字	9.0	0.0	8.4	83.8	158.5	221.4	251.7	240.1	248.4	162.7
5年配当漢字	5.0	0.0	2.3	19.0	103.3	160.6	202.3	239.7	262.9	137.1
6年配当漢字	3.5	0.0	15.7	44.4	73.6	125.0	159.9	167.8	157.2	100.0
非配当漢字	9.5	0.0	5.0	14.8	40.4	62.8	118.2	137.2	191.2	80.4
合計	63.8	474.1	944.1	1277.9	1757.8	2066.7	2383.1	2375.8	2509.8	1772.6

表 6 を見ると、学年が上がるに連れて漢字の使用頻度が増加すること、低学年では未習漢字の使用は稀だが学年が上がるにつれて未習漢字の使用頻度が増加することなどが観察できる。児童の作文の中には、様々な外的要因によって、学習前の漢字で書くことが多い語句や、学習後でも仮名書きのままを書くことが多い語句が混在している可能性があり、非配当漢字の使用状況も併せて、詳しく調査する必要がある。また、各学年の使用漢字の比率を見ると、高学年であっても 3 年生までの配当漢字の使用比率が高い。これは使用頻度が高い語彙に使われる漢字が 3 年生までに配当されていることの帰結である可能性があり、語彙の分布と合わせて調査する必要がある。児童作文コーパスのデータと学習漢字の学年配当表を照らし合わせることによって、多くの児童に共通して観察される学習漢字の配当と使用実態のずれを明らかにすることができる。

### ② 接続詞の使用や文の展開の傾向性

低学年の児童の書く作文では、ある段階から「それで」や「あと」などの接続詞の使用が多くみられる（小学校中学年頃から論理的な文章を書けるようになるため「しかし」などの使用が増えるとの指摘もある）。その後、子どもたちは段階的に接続詞の種類と使用頻度を増やしていくが、ある段階から不要な接続詞の使用を控えるようになる。児童作文コーパスを使用することで、その変化を追跡する調査をすることができる。表 7 は、接続詞の学年別出現頻度を集計し、上位 10 語を 1 万形態素あたりで示したものである。接続詞は短単位では複数の語に分割されるものも多いため（表 7 の「でも」「だから」「すると」「ですが」「それから」など）、長単位で集計している。

表7 接続詞の学年別出現頻度 (長単位 1 万形態素あたり・上位 10 語)

	小1	2	3	4	5	6	中1	2	3	平均
ソシテ	10.9	17.7	20.8	27.9	24.1	23.1	23.7	14.2	17.6	20.5
デモ	19.8	24.4	24.7	22.5	17.2	17.6	16.7	16.7	8.4	17.4
シカシ	0.0	0.0	3.9	2.7	5.4	7.3	11.8	24.0	25.3	11.3
ダカラ	4.9	7.4	11.7	8.0	12.3	12.2	10.2	9.4	10.5	10.3
マタ	4.0	3.0	3.0	6.4	14.0	11.7	12.3	8.5	10.5	9.0
スルト	0.0	0.0	2.2	3.8	4.3	2.3	6.1	1.2	1.0	2.4
ケレド	0.0	3.0	4.3	3.8	5.4	1.6	2.0	2.0	0.8	2.4
デスガ	0.0	0.7	0.9	2.1	1.6	3.6	1.6	1.2	1.5	1.8
ソレカラ	2.0	0.0	0.9	4.3	1.1	1.8	2.5	0.8	1.3	1.6
タダ	0.0	0.7	0.0	1.1	1.1	1.0	2.5	2.0	3.1	1.5

「あと」は自動解析では接続詞ではなく名詞として解析されるため、個別に名詞用法、副詞用法、接続詞用法などの区別を判断し、集計する必要がある。参考として、それらの区別をせずに「あと」の出現頻度を集計したものを表8に示す。

表8 「あと」の学年別出現頻度 (長単位 1 万形態素あたり)

	小1	2	3	4	5	6	中1	2	3	平均
アト	44.5	10.3	9.5	4.8	3.2	4.4	2.9	2.0	1.5	6.2

表7と表8を見ると、学年が上がるにつれて「しかし」の使用頻度が増加すること、「でも」「あと」が減少すること、「そして」「だから」「また」が一度増加したのち減少することなどが確認できる。一方で、「すると」「けれど」などのように習得後もあまり定着しない(使用されない)接続詞もあり、文の展開や類似する接続詞との棲み分け意識などにも注目して分析を進める必要がある。現在のコーパスの規模では用例数が少なく、十分な分析をすることができないが、今後、コーパスの規模を拡充することによって、より詳細な分析を進めることができる。また、接続詞に限らず接続表現全体を視野に入れた(接続助詞を含む)節の複雑化に関する作文能力の変化についても実態を明らかにすることができる。

### ③ 文構造の複雑化に関する発達

子どもたちは発達段階に応じてどの段階でどのような複雑さの文を作文することができるのか、またどの順で文の構造を複雑化させていくのか(修飾・接続関係の習得順序)などの実態を明らかにすることができる。例えば、連体修飾と連用修飾ではどちらの方が、より早く複雑化する傾向にあるのか、また最終的にはどちらの修飾関係の文が作文されやすいかなど、子どもたちの作文表現の実態を明らかにすることができる。この研究は、いわゆる「だらだら文」(長すぎる文やくどく感じる過修飾文、主述の不对応やねじれがある文)の認定や原因の究明に寄与することも期待される。

この研究のためには、既存の構文解析器で付与可能な係り受け情報に加えて、連体、連用など係り受けの種類に関する情報や、主語、述語など文の成分に関する情報の付与が必要になる。現在、これらの情報を自動付与するスクリプトの作成を進めている。図4は文法情報の自動付与し、結果を可視化したものである。

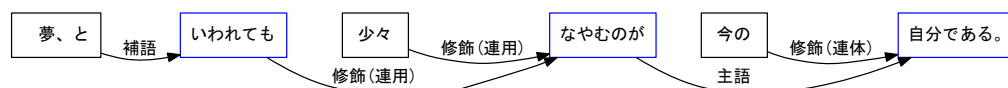


図4 係り受けの種類と文の成分の付与



表9は、このスクリプトにより付与した係り受けの種類を集計し、1万文節あたりで示したものである。

表9 学年別の係り受け分類（1万文節あたり）

	小1	2	3	4	5	6	中1	2	3	平均
主語	485.0	481.1	509.8	595.2	550.5	542.9	589.9	551.2	563.3	548.3
修飾(連体)	1738.0	1576.2	1840.6	1893.9	1964.9	2056.5	2295.9	2305.0	2469.5	2102.0
修飾(連用)	2758.0	2708.7	2704.5	2636.9	2654.3	2714.5	2412.0	2643.8	2450.2	2609.1
接続	1150.7	1471.4	1166.8	1293.4	1293.7	1220.3	1184.7	1044.4	1057.2	1183.3
独立	195.0	69.3	38.7	68.6	48.5	25.0	18.4	21.2	10.9	39.3
補語	2099.4	2154.6	2309.6	2168.4	2116.9	2189.1	2298.8	2143.8	2296.8	2215.8

表9を見ると、学年が上がるにつれて主語や連体修飾語が増加すること、連用修飾語が減少することなどが分かる。ただし、このスクリプトはまだ試験的な段階であり、上記のデータは十分に信頼できるものではない。今後、文法情報の付与作業と検証、修正を進め、文の複雑さの評価やねじれ文の自動検出の研究へと繋げたい。

#### ④ 誤用の実態と作文の傾向性

児童作文コーパスのデータは、多くの表記や仮名遣いの誤り（例えば低学年の児童であれば「ごはんおたべた」のような誤りがある）、語句や文法の誤用が、原本に忠実に記録されている。コーパスを使用することによって、これらの誤りが学齢の進行に伴って、質的または量的にどのように変化していくのか、具体例の提示に加えて数量的な傾向性も明らかにすることができる。また、これまでの研究は、語句レベルでの誤用の指摘が中心であり、それ以外では文の主述のねじれの提示など文レベルでの誤用に留まるものが多かった。今後は、文同士の連続の自然さや段落のつながり方、すなわち文の結束性の研究など、比較的大きなレベルでの誤用や不自然さの研究も進めていく必要がある。

この研究のためには、コーパスへの誤用情報の付与と数値化が必要である。今後、誤用情報付与の設計と計画を進めていきたい。

本研究で構築する児童作文コーパスは、以上のような研究課題の究明に寄与する資料として活用が期待できる。

## 6. まとめ

本発表では、児童・生徒の作文能力の実態を映した『児童・生徒作文コーパス』と検索システムの構築について計画と現在の状況を説明し、児童・生徒の作文能力の発達過程の数値化・視覚化など、コーパスを用いた言語研究の展望を示した。本コーパスは義務教育課程9学年の作文活動を3年間に渡って継続的に調査する300万形態素規模（予定）の作文コーパスであり、児童・生徒の作文を収集したコーパスとしては、データの均質性と規模において従来例のない画期的な資料である。また、本コーパスと併せて平易なインターフェイスを備えた検索システムの開発を進めている。今後は、コーパスの構築と並行して、研究利用のために必要な言語学的情報の付与と、検索システムの改良を進めたい。

本研究の最終的な目標の一つは、教育現場における作文教育の改善と適正化を図ることにある。言語研究の立場から現場の教師が手軽に利用できる作文指導の指針を提案し、有効に活用されれば、昨今二者の乖離が叫ばれて久しい研究と教育の現場の協働の一つの形として位置づけることができる。

### 謝 辞

本研究は、博報財団第9回児童教育実践についての研究助成「学校現場との協働による児童作文指導の基礎的研究」(2014年度、研究代表者：富士原紀絵、助成番号：2014042)、および日本学術振興会科学研究費補助金基盤研究(B)「作文を支援する語彙・文法的事項に関する研究」(平成26～30年度、研究代表者：矢澤真人、研究課題番号：26285196)による補助を得ています。

### 文 献

- 国立国語研究所(1989)『児童の作文使用語彙(国立国語研究所報告98)』東京書籍. ([http://www.ninjal.ac.jp/s\\_data/drep/report\\_nijla/R0098.PDF](http://www.ninjal.ac.jp/s_data/drep/report_nijla/R0098.PDF) よりダウンロード可能)
- 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』17:5、pp.75-93. ([https://www.jstage.jst.go.jp/article/jnlp/17/5/17\\_5\\_5\\_75/\\_pdf](https://www.jstage.jst.go.jp/article/jnlp/17/5/17_5_5_75/_pdf) よりダウンロード可能)
- 鈴木一史、棚橋尚子、河内昭浩(2011)「作文コーパスからみる生徒の使用語彙」『特定領域「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](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)
- 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』17:2、pp.51-65. ([https://www.jstage.jst.go.jp/article/jnlp/17/2/17\\_2\\_2\\_51/\\_pdf](https://www.jstage.jst.go.jp/article/jnlp/17/2/17_2_2_51/_pdf) よりダウンロード可能)
- 藤田彬、田村直良(2012)「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回年次大会発表論文集』pp.987-990. ([http://www.anlp.jp/proceedings/annual\\_meeting/2012/pdf\\_dir/D4-3.pdf](http://www.anlp.jp/proceedings/annual_meeting/2012/pdf_dir/D4-3.pdf) よりダウンロード可能)

### 関連 URL

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