

論文B

「過去」「未来」を主題とする作文の分析

浅原正幸（人間文化研究機構国立国語研究所）

川崎采香（文教大学人間科学部）

上原泉（お茶の水女子大学人間発達教育科学研究所）

酒井裕（玉川大学脳科学研究所）

須藤百香（お茶の水女子大学基幹研究院自然科学系大学院生）

谷口巴（お茶の水女子大学基幹研究院自然科学系大学院生）

小林一郎（お茶の水女子大学基幹研究院自然科学系）

越智綾子（人間文化研究機構国立国語研究所）

鈴木彩香（人間文化研究機構国立国語研究所）

責任著者：浅原正幸（masayu-a@ninjal.ac.jp）

要旨

本研究では「過去」「未来」を主題とする作文の分析を行う。電子化した作文に対して UniDic 形態論情報を付与したうえで、その品詞の出現傾向に基づく分析を行う。さらに、作文で利用されている語彙を教育基本語彙と突合し、語彙配当の出現傾向について分析を行う。一般化線形混合モデルに基づき、作文種別・年代差・性差の分析を行ったところ、それぞれの属性ごとに傾向の異なることを確認した。体・用・相の出現傾向が異なるほか、語彙配当の出現傾向についても差異が見られることを確認した。

キーワード：作文，品詞，語彙配当，線形混合モデル，過去・未来

1. はじめに

我々は、新学術領域研究「時間生成学 時を生み出すところの仕組み」に参画し、ヒトで特に発達した「過去」「現在」「未来」の時間の意識—ところの時間—はどこからどのように生まれるのかについて研究を進めている。浅原・須藤・谷口・小林・越智・鈴木は同領域の A01 班に属し、言語使用の観点から時間のとらえ方の研究を進めている。川崎・上原・酒井は同領域の D01 班に属し、発達過程の観点から時間のとらえ方の研究を進めている。同領域の連携研究として、ヒトの発達過程において、ヒトの「過去」「未来」のとらえ方が、作文の表現にどのように表出するかについての計量的な分析をおこなう。具体的には、「過去」「未来」を主題とする作文を収集し、その品詞および教育基本語彙の語彙配当の出現傾向を調査する。収集した作文は人手で形態論情報を付与し、樺島（1963）・樺

島・寿岳（1965）が提唱した体比率と相用比などを計量する。計量においては、調査対象者の揺れの影響を統制するために、実験協力者をランダム効果としたうえで、年代差・性差を固定効果として考慮した、線形混合モデルにより統計分析をおこなう。結果、「過去」「未来」を主題とする作文を対比させた場合、年代差・性差を固定効果として考慮したとしても、「過去」を主題とする作文が体言を多用する表現であり、「未来」を主題とする作文が用言を多用する表現であることが確認された。さらに、作文に使用された語彙について『教育基本語彙データベース-増補改訂版-』（国立国語研究所 2009）（以下「教育基本語彙」と略す）の語彙配当と突合し、その出現傾向を確認した。結果、低学年配当語彙については「未来」作文のほうが「過去」作文よりもより多く利用される傾向と、高学年・中学校配当語彙については「過去」作文のほうが「未来」作文よりも多く利用される傾向を確認した。

以下、2 節では、関連研究について述べる。3 節では、評価対象である作文データについて紹介する。4 節では、統計分析手法について述べる。5 節に結果を示し、6 節に考察を示す。最後にまとめを示す。

2. 関連研究

以下では、日本語において整備されている作文データとその分析について示す。作文データの構築時期に基づき時系列で示したうえで、各作文についてどのような分析・検討が進められてきたかについて示す。残念ながら、個々の研究が必ずしも収集方法・分析方法ともに体系的に進められているものではないために、分野としての俯瞰的な体系化は困難である。重要な点として、「過去」「未来」を主題とした作文調査はなされておらず、成人も含めた作文調査は管見の限り確認されなかった新規性を主張する。

国立国語研究所は 1971 年より 3 年計画で「現代児童・生徒の言語能力の動態調査」という題目で調査研究を実施した（国立国語研究所 1978）。小学校 6 年生を対象に大都市部・中都市部・郡部の約 620 名の児童に対して、「ともだち」「私の教室」「私が聞いた音」「最近のできごとについての私の意見」「小さなときの私の思い出」「手紙」の 6 つの課題作文が実施され、漢字使用・オノマトペ・題材・印象（ポジティブ／ネガティブ）・敬語使用についての検討が行われた。

その後、国立国語研究所は 1982 年から 7 年計画で「児童の作文に関する調査研究」という題目で調査研究を実施した（国立国語研究所 1989）。この調査では小学校の国語・作文の授業で副教材として利用される地域文章を調査対象とし、自治体教育委員会等が編纂する文集 10 種 10 年分を収集し、単語分割を行ったうえで語彙表を構築した。分析においては、語彙量・初出学年・共出現語・出現度数・品詞構成・語種構成・阪本ランク（阪本 1984）・分類語彙表による意味分野（国立国語研究所 2004）の検討が行われた（一例として茂呂 1986）。

島村（1987）は、1983 年に千葉県松戸市の公立小学校 2 年生・4 年生・6 年生約 120 名を対象とし、「わたしの学校」「先生」「ともだち」という 3 つの課題の作文を収集し、漢字含有率・文字数などの性差を検討した。

坂本（2010）は 2004 年から 2005 年にかけて、全国の小学校 265 校のウェブページで公開されている児童作文を収集し、123 万語規模のデータを構築した。一部については

著作権処理を行った。

永田ほか（2010）は2008年から2009年にかけて、兵庫県内の公立小学校5年生を対象に読書ブログコーパスを構築し、38,269語規模のデータを公開した。

鈴木ほか（2011）は2009年に東京大学教育学部附属中等教育学校の中学1年生から5年生（高校2年生）452名を対象に「年末・年始の行事をふまえて、文化について述べよ」という課題を冬休みの宿題として実施した。形態素解析を行ったうえで約25万語規模のデータを構築し、語種・「学校・社会対照語彙表」に基づくレベル・初出学年・教科特徴語などの語彙調査を行った（国立国語研究所2011）。

藤田・田村（2012）は神奈川県内の小学校9校で小学4年生を対象に672編からなる作文コーパスを構築した。表記や文法誤りを付与しているほか、評価が付与されている。

『児童・生徒作文コーパス』（宮城・今田2015, 富士原ほか2015）は2014年から2016年に国立小中学校を対象に収集した作文コーパスである。「ゆめ（夢）」「ぼくの／わたしのがんばったこと」を課題とした作文を電子化したうえで形態素解析されている。漢字・語彙使用（富士原ほか2015）、程度修飾表現（宮城2016）、接続詞（富士原ほか2016）、修辭ユニット分析（田中2018）、脱文脈化観点（田中ほか2021）などの検討が行われた。

阿部ほか（2017）は1992年に実施された「手」を主題にした作文（320編）を2016年に再度収集（979編）し、対照コーパスとして電子化したうえで経年変化の計量的分析を行った。この作文について、現職教員により評価を行い、評価基準の検討が進められた（宮城ほか2018）。さらに同データを用いた語彙多様性の評価も進められた（今田2021）。

笹島（2017）は都内公立小学校3年生から5年生145名を対象に「朝起きてから学校にくるまで」を課題とした作文を収集し、羅列型・類集型など6種類の文章様式（土部・宝示1963a,b, 1966a,b）の検討を行った。

3. 作文データと教育基本語彙データ

本節では利用した成人および児童の作文データについて示す。

3.1 作文課題

調査対象者に、作文課題2問と記号問題を記した回答用紙1枚、作文を書くための罫線入り用紙2枚の計3枚を配布し回答をもとめた。作文課題はできる限り、用紙の半分以上を書くようにもとめた。記号問題は、本発表では分析対象とはしていないため説明を割愛する。本発表で分析対象としたのは、作文課題の2問への回答（作文）であった。作文課題は(1)過去経験した一番楽しかった出来事を書くようもとめる課題（以下「過去作文」）、(2)将来経験すると予想する最も楽しい出来事を書くようもとめる課題（以下「未来作文」）の2問であった。成人と児童で教示文は同じであった。

3.2 手続き

成人のデータは調査会社を通じて募集をかけ、参加を希望された方々に、依頼の手紙、作文課題と記号問題を記した回答用紙1枚と作文を記入するための回答用紙2枚（過去作文用と未来作文用）、返信用封筒をお送りし、回答用紙3枚を返信用封筒に入れて投函するようにもとめ、収集した。募集から回収までの作業を調査会社に委託した。

児童のデータは、学校長の承諾のもと、学校を通じて協力いただける児童に配布し収集した。首都圏の公立小学校 2 校に依頼し、1 校から小学 4 年生 (70 名) に参加いただき 66 名の分析対象データを得た。もう 1 校から、小学 3 年生から 6 年生 (185 名) に参加いただき 181 名の分析対象データを得た。なお本調査は、お茶の水女子大学人文社会科学研究所の倫理委員会の承認を経て実施された。

3.3 調査対象者

成人 499 名 (20-30 代 158 名 ; 40-50 代 171 名 ; 60 代以上 170 名) が、本調査に参加したが、記入漏れにより 5 名が排除され、分析対象となったデータは 494 名分 (20-30 代 158 名 [男性 75 名, 女性 83 名], 平均年齢 30.4 歳 ; 40-50 代 168 名 [男性 79 名, 女性 89 名], 平均年齢 50.0 歳 ; 60 代以上 168 名 [男性 78 名, 女性 90 名], 平均年齢 68.5 歳) であった。

児童 255 名 (3 年生 46 名 ; 4 年生 125 名 ; 5 年生 35 名 ; 6 年生 49 名) が参加したが、記入漏れにより 8 名が排除され、分析対象となったデータは、247 名分 (3 年生 46 名 [男子 20 名, 女子 26 名] ; 4 年生 118 名 [男子 55 名, 女子 63 名] ; 5 年生 35 名 [男子 17 名, 女子 18 名] ; 6 年生 48 名 [男子 28 名, 女子 20 名]) であった。

結果 741 人分の作文データを収集した。表 1 に調査対象者の属性の基礎統計を示す。

表 1 : 調査対象者の属性

	小学生	20-30 代	40-50 代	60-80 代	計
女性	127 名	83 名	89 名	90 名	389 名
男性	120 名	75 名	79 名	78 名	352 名
計	247 名	158 名	168 名	168 名	741 名

3.4 教育基本語彙

「教育基本語彙の基本的研究-増補改訂版-」(国立国語研究所 2009) の付属 CD-ROM に採録されている『教育基本語彙データベース-増補改訂版-』(国立国語研究所 2009) は、「教育基本語彙の基本的研究-教育基本語彙データベースの作成」(国立国語研究所 2001) の増補版として刊行されたものである。『岩波国語辞典』により見出し語の正規化を行ったほか、分類語彙表番号も付与されている。阪本の新旧の基本語彙(阪本 1958, 阪本 1984) を含む基本語彙 6 種のデータベースなどに基づき、小学校低学年配当語彙・小学校高学年配当語彙・中学校配当語彙の 3 つの区分の語彙配当を定義した。語彙配当の区分別語彙数(異なり語数) は小学校低学年 6,865 語・小学校高学年 8,620 語・中学校 11,749 語の計 27,234 語である。本研究はこの語彙配当の分布についても検討をおこなう。

4. 統計分析手法

4.1 品詞

以下では、表現の分析として、UniDic 形態論情報に基づき、品詞情報の統計分析をおこなう。得られた 741 人分の過去作文・未来作文を電子化したうえで、国語短単位で分

から書きし、UniDic 品詞体系に基づく形態論情報を人手により付与した。樺島（1963）や樺島・寿岳（1965）が提唱した体比率（体言：名詞・代名詞）・用比率（用言：動詞）・相比率（形容詞・形状詞・副詞・連体詞）・相用比（MVR: Modifier Verb Ratio）について検討を行い、表現の性差・年代差および作文種別（過去作文か未来作文か）の差を確認する。但し、絶対評価ではなく、研究協力者（subj）をランダム効果とした lme4（Bates et al. 2015）による一般化線形混合モデルに基づき分析を行う。具体的には、1 作文中の体・用・相の単語数について、1 作文の延べ語数（word_token）（体比率・用比率・相比率の推定）もしくは用の単語数（相用比の推定）と年代（age={elementary, 20_30, 40_50, 60_80} の離散値）と性別（gender={m:男性, f:女性}）と作文種別（future_past={future:未来作文, past:過去作文}）を固定効果とした線形式を回帰により求め、その傾きを評価した。具体的には次のような式を用いて回帰分析をおこなった。

$$\begin{aligned} \text{体比率：体（の単語数）} &= \text{word_token} + \text{age} * \text{gender} * \text{future_past} + (1 | \text{subj}) \\ \text{用比率：用（の単語数）} &= \text{word_token} + \text{age} * \text{gender} * \text{future_past} + (1 | \text{subj}) \\ \text{相比率：相（の単語数）} &= \text{word_token} + \text{age} * \text{gender} * \text{future_past} + (1 | \text{subj}) \\ \text{相用比：相（の単語数）} &= \text{用（の単語数）} + \text{age} * \text{gender} * \text{future_past} + (1 | \text{subj}) \end{aligned}$$

統計モデルは正規分布を用い、最尤推定法によりモデルを求めた。式中 * は交互作用を表し、(1|subj) が研究協力者のランダム効果である。当該式を最大モデルとし、対数尤度・AIC（赤池情報量：Akaike Information Criterion）・BIC（Bayesian Information Criterion）により anova によりモデル選択を行ったところ、最大モデルが最良モデルであったためにその結果を示す。分析時には、いったん全データで回帰を行ったうえで得られた回帰式から±3 標準偏差離れたものを外れ値として排除したうえで、最終的な回帰式を得た。すべてのモデルは収束した。

4.2 教育基本語彙・機能語・補助記号

次に、語彙の難易度を評定するために教育基本語彙（国立国語研究所 2009）の語彙と突合し、出現する語彙の配当学年の傾向を検討する。教育基本語彙の「見出し」と「表記」を UniDic 形態論情報の語彙素と突合したところ、延べ語数で 185,741 語中 144,945 語（78.0%）、異なり語数で 8,047 語中 5,078 語（63.1%）が適合した。適合しなかった語の多くは、補助記号（読点 9,734 件・句点 7,037 件・括弧 990 件・補助記号一般 345 件）・名詞-数詞（4,060 件）・助動詞（1,461 件）・名詞-固有名詞-地名一般（1,047 件）であった。品詞比率と同様に、低学年配当比率・高学年配当比率・中学校配当比率について、各種属性を固定効果とし、研究協力者をランダム効果とした線形式を回帰により求め、その傾きを評価した。さらに、作文を読んでいて体言止めの生起傾向や文の長さなど、機能語や補助記号の使い方についても性差・年代差があることが認められたため、機能語の比率・補助記号の比率についても同様に回帰分析をおこなった。具体的には次のような式を用いて回帰分析をおこなった。

$$\text{低学年配当比率：低学年配当単語数} = \text{word_token} + \text{age} * \text{gender} * \text{future_past} + (1 | \text{subj})$$

高学年配当比率：高学年配当単語数= $\text{word_token} + \text{age} * \text{gender} * \text{future_past} + (1 | \text{subj})$
 中学校配当比率：中学校配当単語数= $\text{word_token} + \text{age} * \text{gender} * \text{future_past} + (1 | \text{subj})$
 機能語の比率：機能語の単語数= $\text{word_token} + \text{age} * \text{gender} * \text{future_past} + (1 | \text{subj})$
 補助記号の比率：補助記号の単語数= $\text{word_token} + \text{age} * \text{gender} * \text{future_past} + (1 | \text{subj})$

品詞分析時と同様に、いったん全データで回帰を行ったうえで得られた回帰式から±3標準偏差離れたものを外れ値として排除したうえで、最終的な回帰式を得た。すべてのモデルは収束した。また、最大モデルが最良モデルであった。

5. 結果

5.1 品詞

品詞分布の統計分析結果を表2に示す。表の作成には `stargazer` (Hlavac 2018) を用いた。各係数の推定値の右のアスタリスクは有意水準(* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$)を表し、下のカッコ書きは標準誤差を表す。列は体比率・用比率・相比率・相用比の分析を示す。`word_token` は1作文あたりの延べ語数に対する係数を表す。用は相用比における用の単語数を表す。年代は `age=20_30` (20-30代) に対する係数の差分を示す。性別は `gender=f` (女性) に対する係数の差分を示す。作文種別は `future_past=future` (未来作文) に対する係数の差分を示す。固定効果の：による組み合わせは、交互作用を意味する。`Constant` は線形式の切片に相当する。

表2：品詞分布に基づく分析

	従属変数			
	体	用	相	
	体比率	用比率	相比率	相用比
<code>word_token</code> 1 作文中の延べ語数	0.278*** (0.003)	0.128*** (0.002)	0.052*** (0.001)	
用 1 作文中の延べ語数				0.303*** (0.011)
<code>age=40_50</code> 40-50代 (⇔20-30代)	1.948** (0.939)	-1.468*** (0.563)	-1.004** (0.475)	-0.548 (0.504)
<code>age=60_80</code> 60-80代 (⇔20-30代)	3.834*** (0.937)	-1.775*** (0.562)	-2.026*** (0.474)	-1.381*** (0.503)
<code>age=elementary</code> 小学生 (⇔20-30代)	1.319 (0.916)	-2.914*** (0.550)	-2.290*** (0.464)	-2.770*** (0.498)
<code>gender=m</code> 男性 (⇔女性)	1.844* (0.981)	0.558 (0.589)	-0.858* (0.498)	-1.188** (0.528)
<code>future_past=past</code> 過去作文 (⇔未来作文)	3.076*** (0.778)	-3.740*** (0.519)	-0.164 (0.434)	1.300*** (0.457)
<code>age=40_50:gender=m</code>	-0.513	-0.704	0.612	0.852

	(1.362)	(0.820)	(0.692)	(0.734)
age=60_80:gender=m	-0.698 (1.363)	-1.057 (0.820)	0.920 (0.695)	1.234* (0.736)
age=elementary:gender=m	-0.256 (1.252)	-1.131 (0.753)	0.573 (0.636)	0.860 (0.675)
age40_50:future_past=p	0.246 (1.072)	0.263 (0.718)	0.579 (0.601)	0.335 (0.637)
age60_80:future_past=p	1.721 (1.076)	-0.381 (0.718)	-0.159 (0.602)	0.329 (0.637)
age=elementary: future_past=p	-2.317** (0.990)	1.535** (0.664)	0.483 (0.555)	-0.056 (0.587)
gender=m:future_past=p	0.904 (1.116)	-0.728 (0.749)	0.047 (0.627)	0.237 (0.663)
age=40_50:gender=m: future_past=p	1.704 (1.553)	0.215 (1.044)	-0.869 (0.872)	-0.700 (0.923)
age=60_80:gender=m: future_past=p	-0.005 (1.560)	0.986 (1.044)	0.249 (0.875)	-0.147 (0.925)
age=elementary:gender=m: future_past=p	-1.223 (1.426)	1.100 (0.959)	0.031 (0.801)	-0.243 (0.847)
Constant 切片	-2.840*** (0.797)	4.058*** (0.477)	2.426*** (0.404)	3.153*** (0.441)
データポイント数	1,474/1,482	1,482/1,482	1,473/1,482	1,470/1,482
対数尤度	-4,713.996	-4,025.165	-3,746.154	-3,824.278
Akaike Inf. Crit.	9,465.992	8,088.330	7,530.307	7,656.556
Bayesian inf. Crit.	9,566.611	8,189.052	7,630.913	7,787.123

* p<0.1; ** p<0.05; ***p<0.01

まず、年代別の傾向について確認する。体比率においては、年代が高くなるほど体言が多くなる傾向がみられた。用比率においては、20-30代が顕著に用言が多く、それ以外の年代では少ない傾向がみられた。相比率においては、20-30代が顕著に修飾表現が多く、それ以外の年代では少ない傾向がみられた。相用比においては、20-30代が高く、60-80代と小学生が低い傾向がみられた。

性別においては、相比率の差異から男性のほうが相用比が小さい傾向がみられた。

作文種別においては、過去作文は未来作文と比べて顕著に体言が多く用言が少ない傾向がみられた。相比率には有意差がみられなかったが、用比率の差異から過去作文のほうが相用比が大きい傾向がみられた。

交互作用効果として、小学生の過去作文が体比率が低く、用比率が高くなる傾向が見られた。

5.2 教育基本語彙・機能語・補助記号

教育基本語彙に基づく語彙配当および機能語・補助記号の統計分析結果を表3に示す。

表3：教育基本語彙に基づく語彙配当および機能語・補助記号の分析

	従属変数				
	低学年	高学年	中学校	機能語	補助記号
word_token	0.652*** (0.004)	0.084*** (0.002)	0.031*** (0.001)	0.397*** (0.003)	0.096*** (0.002)
1 作文中の延べ 語数					
age=40_50 40-50代 (⇔20- 30代)	-2.823*** (1.179)	0.707 (0.601)	-0.067 (0.388)	-0.666 (0.828)	1.259** (0.613)
age=60_80 60-80代 (⇔20- 30代)	-3.934*** (1.176)	1.335** (0.598)	0.297 (0.386)	-1.793** (0.828)	1.748*** (0.614)
age=elementary 小学生 (⇔20-30 代)	-7.360*** (1.151)	1.430** (0.582)	-0.013 (0.380)	-1.997** (0.809)	2.632*** (0.596)
gender=m 男性 (⇔女性)	-2.640** (1.232)	2.068*** (0.626)	0.865** (0.404)	-1.831** (0.868)	0.073 (0.640)
future_past=pas t 過去作文 (⇔未 来作文)	-4.060*** (1.078)	0.987* (0.565)	1.177*** (0.364)	0.957 (0.712)	-0.409 (0.411)
age=40_50: gender=m	1.019 (1.715)	-1.247 (0.873)	0.057 (0.563)	0.085 (1.207)	-0.124 (0.892)
age=60_80: gender=m	0.715 (1.716)	-1.095 (0.872)	0.175 (0.562)	-0.020 (1.211)	-0.720 (0.894)
age=elementary: gender=m	1.337 (1.576)	-1.399* (0.799)	-0.403 (0.515)	1.570 (1.109)	-0.724 (0.820)
age=40_50: future_past=p	-0.907 (1.492)	0.464 (0.783)	0.057 (0.504)	-0.817 (0.985)	-0.712 (0.567)
age=60_80: future_past=p	-0.192 (1.499)	0.273 (0.781)	-1.061** (0.502)	-1.452 (0.987)	-1.344** (0.573)
age=elementary: future_past=p	3.450** (1.379)	-1.934*** (0.720)	-1.718*** (0.462)	-0.749 (0.911)	0.761 (0.524)
gender=m: future_past=p	-0.646 (1.557)	-0.161 (0.817)	-0.536 (0.522)	-0.739 (1.030)	0.452 (0.589)
age=40_50:	0.025	0.871	-0.077	0.056	-0.114

gender=m: future_past=p	(2.170)	(1.138)	(0.728)	(1.435)	(0.822)
age=60_80: gender=m: future_past=p	0.125 (2.177)	0.008 (1.140)	0.842 (0.728)	0.970 (1.438)	0.582 (0.825)
age=elementary: gender=m: future_past=p	1.288 (1.993)	0.203 (1.042)	0.424 (0.666)	1.086 (1.319)	-0.750 (0.756)
Constant 切片	8.497*** (1.000)	-2.809*** (0.505)	-0.274 (0.336)	1.378** (0.702)	-1.019** (0.514)
データポイント 数	1,478/1,48 2	1,466/1,48 2	1,461/1,48 2	1,474/1,48 2	1,467/1,48 2
対数尤度	-5,104.094	-4,067.857	-3,407.256	-4,511.460	-3,972.992
Akaike Inf. Crit.	10,248.190	8,173.714	6,852.511	9,140.919	7,983.985
Bayesian inf. Crit.	10,346.860	8,274.229	6,952.962	9,241.538	8,084.513

* p<0.1; ** p<0.05; ***p<0.01

まず、年代別の傾向について確認する。成人においては、20-30代に顕著に記号が少なく低学年語彙が多い傾向が見られた。60-80代に顕著に機能語は少なく高学年語彙が多い傾向が見られた。小学生においては、顕著に低学年語彙が少なく高学年が多い傾向が見られた。また機能語が少なく、補助記号が多い傾向が見られた。中学校配当語彙については、世代差が見られなかった。

性別においては、男性について、低学年語彙が少なく、高学年語彙・中学校語彙が多い傾向が見られた。また男性のほうが機能語が少なかった。

作文種別においては、過去作文は未来作文と比べて低学年語彙が少なく、中学校語彙が多い傾向が見られた。

交互作用効果として、60-80代の過去作文は中学校語彙が少なく、補助記号が少ない傾向が見られた。小学生の過去作文は低学年語彙が多く、高学年語彙・中学校語彙が少ない傾向が見られた。

6. 考察

調査対象者の属性や作文種別（「過去」「未来」）の体比率と相用比の傾向から、樺島（1963）、樺島・寿岳（1965）に倣い、「要約的」⇔「描写的」もしくは「有様描写的」⇔「動き描写的」などの特性を検証する。樺嶋らは体比率が高く相用比の低いものを「要約的」とした。60-80代の作文が典型的な「要約的」な表現であった。体比率が低いもののうち相用比が高いものが「有様描写的」とされ、20-30代の作文が典型的な「有様描写的」な表現であった。

作文の主題の対比では、未来作文が、体比率が低く相用比も低い「動き描写的」な表現

であった。過去作文は用比率が低いために、体比率と相用比が高い傾向が見られた。しかしながら、小学生の過去作文の交互作用の結果から、成人に見られる「過去作文の体比率が高く用比率が低い傾向」が打ち消されていることがわかった。小学生は過去作文であっても用言を用いる傾向が見られた。

表 4：結果に関する考察（語彙配当）

	少ない←→多い
低学年配当語彙	小学生<<60-80代<40-50代<20-30代 男性<女性 過去<<未来
高学年配当語彙	20-30代<40-50代<60-80代≒小学生 女性<<男性 未来<過去
中学校配当語彙	世代差は見られない 女性<男性 未来<<過去
機能語	小学生≒60-80代<40-50代<20-30代 男性<女性 作文種別差は見られない
補助記号	20-30代<<40-50代<60-80代<小学生 性差・作文種別差は見られない

さらに、教育基本語彙による語彙配当の検討においては、表 4 のような傾向が確認された。

まず世代差について確認する。小学生は 4 年生を中心に 3 年生以上を対象として調査を行っており、高学年配当語彙を多用する傾向が見られた。成人においては若い世代が高学年配当語彙よりも低学年配当語彙を用いる傾向が見られた。中学校配当語彙においては、世代差が見られなかった。機能語については、小学生や 60-80 代が少なく、20-30 代が多い傾向が見られた。20-30 代のほうが、助詞・助動詞を多用する傾向が見られた。補助記号（文や節を区切る読点・句点の数）は、小学生や 60-80 代が多く、20-30 代が少ない傾向が見られた。これは、前者が文や節が短く、後者が文や節が長いことを表し、節を長くする要因として 20-30 代が修飾語（相の類）を多用していることとも合致している。

性差を見ると男性のほうが、より高学年の配当語彙を用いる傾向が見られた。

表 5：教育基本語彙における「.1642 過去」・「.1643 未来語彙」の分布

	低学年配当	高学年配当	中学校配当
過去 .1642	21	27	21
未来 .1643	16	33	12

作文種別を確認すると過去作文のほうがより高学年の配当語彙を用いる傾向が見られた。教育基本語彙には分類語彙表番号が付与されており、時間を直接表す語彙である、分類語彙表番号が「.1642 過去」と「.1643 未来」の単語分布を調べると表 5 のようになった。中学校配当語彙においては、過去の語彙が未来の語彙より多く、表 4 の傾向と合致するが、低学年配当・高学年配当語彙においては必ずしも表 4 の傾向と合致しない。過去・未来を表す語彙としてさまざまな習得時期の語があったとしても、既に経験した既知事象を表現するにあたってはより高学年の語彙を使うが、まだ経験していない未知事象を表現するにあたってはより低学年の語彙を使う傾向が確認された。

7. おわりに

本研究では、「過去」と「未来」を主題とする作文を収集し、品詞の分布傾向を確認し、表現の違いについて検証した。「過去作文」は体言を多く使う傾向にあり、用言に比べて修飾表現が多い傾向が確認された。「未来作文」は体言を少なく使う傾向にあり、用言に比べて修飾表現が少ない傾向が確認され、動き描写的な表現であった。さらに語彙配当の観点からも分析を行い、既に経験した事象を述べる「過去作文」のほうがまだ経験していない事象を述べる「未来作文」よりも習得時期が遅い配当語彙を利用する傾向が見られた。時間生成学の Tang ら (2020) では、作例に基づく「過去」「現在」「未来」の表現を呈示した際の脳活動データの分析において、3つの表現の差異が脳活動でも差分があることを確認している、A01 班の別の研究 (未発表) では、テキストの読み時間を評価した際に「過去」の表現よりも「未来」の表現のほうが読み時間が有意に短いことが確認されている。本研究の言語の生成過程の分析と、読み時間に基づく言語の受容過程の分析から、「未来」と比べて「過去」を主題にした言語表現に負荷がかかることが示唆される。

年代別・性別の語彙配当の分析においては、小学生の作文において補助記号を多用する傾向が見られた。データの電子化・形態論情報付与の過程において、小学生の文が短い傾向が確認されている。句点の比率が大きくなったために、補助記号の比率に差異が見られたと考える。20-30代が低学年配当語彙を多く使い、60-80代が高学年配当語彙を多く使う傾向が見られた。男性のほうが女性よりも習得時期が遅い配当語彙を利用する傾向が確認された。これらの傾向も、電子化・形態論情報付与の作業時に定性的に確認できたことであったが、定量的にも確認できた。

付記

本稿は、日本語学会第 2021 年度秋季大会 (於オンライン) での口頭発表に加筆・修正をしたものである。

謝辞

本研究は科研費 JP18H05521, JP18H05524 および国立国語研究所コーパス開発センター共同研究プロジェクトによるものです。

文献

- Douglas Bates, Martin Machler, Ben Bolker, and Steve Walker (2015). “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, 67:1, pp. 1–48.
- Marek Hlavac (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables.*, Central European Labour Studies Institute (CELSI) Bratislava, Slovakia. R package version 5.2.2
- Long Tang, Toshimitsu Takahashi, Tamami Shimada, Masayuki Komachi, Noriko Imanishi, Yuji Nishiyama, Takashi Iida, Yukio Otsu, and Shigeru Kitazawa (2020). “Neural correlates of temporal presentness in the precuneus: a crosslinguistic fMRI study based on speech stimuli”, *Cerebral Cortex* 31 (3), pp.1538-1552.
- 阿部藤子・今田水穂・宗我部義則・富士原紀絵・松崎史周・宮城信(2017). 「児童生徒の「手」作文に於ける経年変化の計量的分析：1992年と2016年の作文を比較して」 *言語資源活用ワークショップ発表論文集 1 巻*, pp. 234-247.
- 今田水穂 (2021). 「児童作文における語彙多様性の評価」 *計量国語学*, 33:3, pp.145-161.
- 樺島忠夫 (1963). 『表現論-ことばと言語行動』 綜芸舎.
- 樺島忠夫・寿岳章子 (1965). 『文体の科学』 綜芸舎.
- 国立国語研究所(1978). 『児童の表現力と作文』 国立国語研究所報告；63 東京書籍.
- 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.
- 国立国語研究所 (2001). 『教育基本語彙の基本的研究：教育基本語彙データベースの作成』 国立国語研究所報告；117 明治書院.
- 国立国語研究所 (2009). 『教育基本語彙の基本的研究 -増補改訂版-』 国立国語研究所報告；127 明治書院.
- 国立国語研究所 (2011). 『特定領域研究『日本語コーパス』言語政策班最終成果 CD-ROM』, application/zip.
- 阪本一郎 (1958). 『教育基本語彙』 牧書店.
- 阪本一郎 (1984). 『新教育基本語彙』 学芸図書.
- 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.
- 笹島眞実(2017). 「言語形式に基づく児童作文の類型化」 *言語資源活用ワークショップ発表論文集 2 巻*, pp. 290-296.
- 島村直己(1987). 「児童の漢字使用：課題作文の漢字含有率から」 国立国語研究所報告；90, pp. 77-94.
- 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」 「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.
- 田中弥生(2018). 「児童・生徒作文の日本語修辭ユニット分析と教員評価の検討」 *言語資源活用ワークショップ発表論文集 3 巻*, pp. 91-104.
- 田中弥生・佐尾ちとせ・宮城信(2021). 「児童作文の評価に向けた脱文脈化観点からの検討」 *言語処理学会第 27 回年次大会発表論文集*, pp. 750-755.
- 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.

- 土部弘・宝示重美(1963a). 「文章意識の発達 (第1報)」 大阪学芸大学紀要 C 教育科学, 4, pp. 92-104.
- 土部弘・宝示重美(1963b). 「文章意識の発達 (第2報)」 大阪学芸大学紀要 C 教育科学, 5, pp. 40-50.
- 土部弘・宝示重美(1966a). 「文章意識の発達 (第3報)」 大阪学芸大学紀要 C 教育科学, 6, pp. 78-89.
- 土部弘・宝示重美(1966b). 「文章意識の発達 (補注)」 大阪学芸大学紀要 C 教育科学, 7, pp. 112-128.
- 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.
- 富士原紀絵・宮城信・松崎史周(2015). 「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-」 漢字・日本語教育研究, 5, pp. 4-20.
- 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」 お茶の水女子大学子ども学研究紀要, 4, pp. 9-20.
- 宮城信 (2016). 「児童作文に見る程度修飾表現の発達」 富山大学人間発達科学部紀要, 10:2, pp. 291-297.
- 宮城信・浅原正幸・今田水穂(2018). 「現職員による児童・生徒作文の評価基準の分析」 言語資源活用ワークショップ発表論文集 3 巻, pp. 421-435.
- 宮城信・今田水穂(2015). 「『児童・生徒コーパス』の設計」 第7回コーパス日本語学ワークショップ予稿集 7 巻, pp. 223-232.
- 茂呂雄二(1986). 「作文の使用語彙から」 児童・生徒の常用漢字の読み書き能力, pp. 17-26.

資料

ツール [ゴシック体, 11pt]

R ver. 4.0.2 <https://www.r-project.org/> (2022年8月19日確認).

MeCab-0.996 <https://taku910.github.io/mecab/> (2022年8月19日確認).

UniDic-2.3.0 <https://clrd.ninjal.ac.jp/unidic/> (2022年8月19日確認).

Title:

Stylistic Analysis of Essays on the themes of 'Past' and 'Future'

Authors:

ASAHARA Masayuki, National Institute for Japanese Language and Linguistics,
National Institutes for Humanities

KAWASAKI Ayaka, Faculty of Human Sciences, Bunkyo University

UEHARA Izumi, Institute for Education and Human Development, Ochanomizu

University

SAKAI Yutaka, Brain Science Institute, Tamagawa University

SUDO Momoka, Graduate School of Humanities and Sciences, Ochanomizu University

TANIGUCHI Tomoe, Graduate School of Humanities and Sciences, Ochanomizu University

KOBAYASHI Ichiro, Graduate School of Humanities and Sciences, Ochanomizu University

OCHI Ayako, National Institute for Japanese Language and Linguistics, National Institutes for Humanities

SUZUKI Ayaka, National Institute for Japanese Language and Linguistics, National Institutes for Humanities

Abstract:

The present research reports on an analysis of essays on the themes of ‘past’ and ‘future’. We assign the UniDic morphological information to the digitized essays and analyse them in terms of part-of-speech realization frequencies. Then we match the identified vocabulary with the Fundamental Vocabulary of Educational Purpose and analyze them in terms of the occurrence trend of vocabulary allocation. We also analyzed the data based on the Generalized Linear Mixed Model and compared them by the essay themes and the author attributes such as age groups and gender types. The results have revealed how each attribute characterizes the essay style of each group. The differences are found in the frequency of the “*tai* (nouns)”, “*yoo* (verbs)”, and “*soo* (modifiers)” as well as in the occurrence trend of vocabulary allocation.

Keywords:

Essays, POS Distribution, Vocabulary Distribution, Linear Mixed Model, ‘past’ and ‘future’

利益相反の開示：

本研究は、文部科学省科学研究費新学術領域研究「時間生成学 時を生み出すところの仕組み」における A01 班と D01 班の共同研究として実施したものである。企業などとの共同研究および企業などからの助成は受けていない。

著者の貢献：

- 浅原正幸：統計的分析・論文執筆
- 川崎采香：作文の収集・発達に関する分析
- 上原泉：作文の収集・発達に関する分析
- 酒井裕：統計的分析手法の検討
- 須藤百香：統計的分析手法の検討

- 谷口巴：統計的分析手法の検討
- 小林一郎：統計的分析手法の検討
- 越智綾子：言語情報の検討
- 鈴木彩香：言語情報の検討