

「過去」「未来」を主題にした作文の文体分析

あさはら まさゆき かわさき あや か うえはら いずみ さかい ゆたか
 浅原 正幸* 川崎 采香 上原 泉 酒井 裕
 国立国語研究所 お茶の水女子大学 お茶の水女子大学 玉川大学

すどう ももか こばやし いちろう おち あやこ
 須藤 百香 小林 一郎 越智 綾子
 お茶の水女子大学 お茶の水女子大学 国立国語研究所

1. はじめに

本研究では、ヒトの「過去」「未来」のとらえ方が、作文の文体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」の表現文体について世代間・性別間でどのように差異があるかを調査するために、「過去」「未来」を主題とした作文を収集し、その品詞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一般化線形混合モデルにより統計分析をおこなう。結果、「過去」を主題とした作文が要約的な文体であり、「未来」を主題とした作文が有様描写的であることが確認された。

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

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」という三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。

* masayu-a@ninjal.ac.jp

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／わたしががんばったこと」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)、修辞ユニット分析 (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)。

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。

このように多様な方法で作文データが過去に収集されてきたが、「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。

3. 作文データ

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 20 名、女子 28 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	382 名
計	247 名	158 名	168 名	168 名	741 名

4. 統計分析手法

以下では、文体分析として、UniDic 形態論情報に基づき、品詞情報の統計分析を行う。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、国語研短単位で分かち書きし、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)・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、文体の性差・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。

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った。具体的には、1 作文中の体・用・相の単語数を、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。

体比率：体 (の単語数) = $word.token + age + gender + future_past + (1|subj)$

用比率：用 (の単語数) = $word.token + age + gender + future_past + (1|subj)$

相比率：相 (の単語数) = $word.token + age + gender + future_past + (1|subj)$

相用比：相 (の単語数) = $用 (の単語数) + age + gender + future_past + (1|subj)$

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。

5. 結果

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。Constant は線形式の切片に相当する。

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

表 2 統計分析結果

	従属変数:			
	体	用	相	
	体比率	用比率	相比率	相用比
word_token	0.280***	0.125***	0.050***	
1 作文の異なり語数	(0.003)	(0.002)	(0.001)	
用				0.304***
1 作文中の用の異なり語数				(0.011)
age=40_50	2.187***	-1.512***	-0.685**	-0.112
40-50 代 (↔ 20-30 代)	(0.553)	(0.304)	(0.270)	(0.285)
age=60_80	4.177***	-2.135***	-1.570***	-0.638**
60-80 代 (↔ 20-30 代)	(0.557)	(0.306)	(0.272)	(0.285)
age=elementary	-0.049	-2.554***	-2.000***	-2.391***
小学生 (↔ 20-30 代)	(0.590)	(0.328)	(0.293)	(0.313)
gender=m	2.034***	-0.322	-0.376**	-0.427**
男性 (↔ 女性)	(0.371)	(0.204)	(0.181)	(0.191)
future_past=past	3.009***	-3.227***	0.101	1.400***
過去作文 (↔ 未来作文)	(0.286)	(0.178)	(0.151)	(0.153)
Constant	-2.985***	4.323***	2.415***	2.735***
切片	(0.617)	(0.347)	(0.311)	(0.341)
観測数	1,472/1,482	1,471/1,482	1,471/1,482	1,469/1,482
対数尤度	-4,726.988	-3,941.038	-3,732.384	-3,819.429
Akaike Inf. Crit.	9,471.976	7,900.076	7,482.769	7,656.858
Bayesian Inf. Crit.	9,519.626	7,947.719	7,530.412	7,704.489

Note:

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

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・40-50 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、男性のほうが体言が多く修飾表現が少ない傾向がみられた。用比率には有意差は見られなかったが、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。

最後に作文種別においては、過去作文は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。

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。まとめると表 3 のようになる。

表 3 結果に関する考察

	相用比：高	相用比：低
体比率：高		【要約的】 60-80 代・男性・過去作文
体比率：低	【有様描写的】 20-30 代・40-50 代・女性・未来作文	【動き描写的】 小学生

年代別では、60-80 代の体比率が高く、相用比が低い「要約的」な傾向があった。一方、20-30 代・40-50 代の体比率が低く相用比が高い「ありさま描写的」な傾向と、小学生作文の体比率・相用比がともに低い「動き描写的」な傾向があった。性別では、女性の作文の体比率が低く相用比が高い「ありさま描写的」な傾向が確認され、男性の作文の体比率が高く相用比が低い「要約的」な傾向に差がみられた。「過去」を題材にした作文は体比率が高く相用比が低い「要約的」な傾向にあった。これに対し「未来」を題材にした作文は体比率が低く相用比が高い、「ありさま描写的」なものであった。総合的に評価すると「未来作文」は 20-50 代・女性の文体に近く、「過去作文」は 60-80 代・男性の文体に近い傾向が確認された。

7. おわりに

本研究では、「過去」と「未来」を主題とした作文を収集し、品詞の分布傾向を確認し、文体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少ない傾向が確認され、要約的な文体であった。「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多い傾向が確認され、有様描写的な文体であった。これらの差異は $p < 0.01$ 有意水準で差が確認された。

また、年代別・性別においては、60-80 代・男性といった属性が「過去作文」と同様に要約的な文体であり、20-30 代・40-50 代・女性といった属性が「未来作文」と同様に有様描写的な文体であった。最後に、小学生の文体は体比率・相用比ともに低い、動き描写的な文体であることが確認された。

今後、文体以外の検討を進めていきたい。具体的には教育基本語彙 (国立国語研究所 2001)・単語親密度 (浅原 2020)・分類語彙表による意味分野 (国立国語研究所 2004) についての検討を進めたい。

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

- 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.
 阪本一郎 (1984). 『新教育基本語彙』 学芸図書.
 国立国語研究所 (2004). 『分類語彙表増補改訂版』 国立国語研究所資料集 ; 14 大日本図書.
 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.
 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.
 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.

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 田中弥生 (2018). 「児童・生徒作文の日本語修辞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.
- 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.
- 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.
- 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.
- 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.
- Douglas Bates, Martin Mächler, 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
- 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』 明治書院.
- 浅原正幸 (2020). 「Bayesian Linear Mixed Model による単語親密度推定と位相情報付与」 自然言語処理, 27:1, pp. 133-150.