

国立国語研究所学術情報リポジトリ

日本語学習者がパソコンを用いて執筆した作文における漢字の使用：中国語母語話者を対象に

メタデータ	言語: Japanese 出版者: 国立国語研究所 公開日: 2025-07-25 キーワード (Ja): 漢字使用率, 執筆過程, キーボード入力, タイピング, W-CoLeJa キーワード (En): rate of kanji, writing process, keyboarding, typing, W-CoLeJa 作成者: 胡, 佳芮 メールアドレス: 所属: 国立国語研究所 非常勤研究員
URL	https://doi.org/10.15084/0002000519

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.



日本語学習者がパソコンを用いて執筆した作文における漢字の使用

——中国語母語話者を対象に——

胡 佳芮

国立国語研究所 非常勤研究員

要旨

本研究は、中国人日本語学習者（CJL）の作文における漢字使用の実態を明らかにすることを目的とし、ベトナム人日本語学習者（VJL）との比較を通じて分析を行った。日本語学習者作文縦断コーパス（W-CoLeJa）に収録された作文と執筆プロセスデータを用い、以下の2点を明らかにした。第一に、テーマ別・学年別の作文における漢字使用率の差異を検証した結果、CJLの作文では初期段階から漢字使用率が高く、学年が進むにつれて安定する一方、テーマによる違いが明確に表れた。体験文では漢字使用率が低く、説明文や意見文では高い傾向が見られた。これにたいし、漢字圏外のVJLでは学年が進むにつれ漢字使用率が徐々に増加し、CJLと異なる傾向を示した。第二に、CJLの漢字入力方法の特徴を分析した結果、日本語ローマ字入力と中国語ピン音入力の2つの方法が併用されており、とくに固有名詞の入力において複雑な試行錯誤が見られた。漢字の日本語の発音を把握していない場合、音読みに頼る傾向が観察された。この結果から、デジタル環境での日本語教育において、日本語の音韻情報の習得やキーボード入力を考慮した指導の必要性が示唆される*。

キーワード：漢字使用率、執筆過程、キーボード入力、タイピング、W-CoLeJa

1. はじめに

現代日本語は漢字仮名交じり文で表記されている。書き言葉の文字種比率¹は、漢字が約3割、仮名が約6割を占めている（胡 2024）。漢字を共有している中国語を母語とする中国人日本語学習者（Chinese Japanese Learner: CJL）は、母語の漢字知識をいかし日本語の読み書きに対応できると考えられるが、CJLの書いた文章において漢字がどの程度使用されているのか、また日本語学習が進むにつれ漢字の使用に変化が見られるのかについてはまだ十分に明らかにされていない。

これまでCJLの作文における漢字の使用にかんする研究は、手書き作文における漢字の書き間違いや日中同形語の使用、簡体字による誤用などに焦点をあててきた。しかし、近年パソコンやスマートフォンといったデジタルデバイスが主な書記媒体となり、手書きで文章を書く機会は

* 本稿はNINJAL国際シンポジウム「中国語話者のための学習者コーパスを用いた作文の研究と教育」（2024年9月15日於北京語言大学）のパネルセッションでの発表内容をもとに作成したものである。本稿の掲載にあたり、貴重なご指摘・ご助言をくださった査読委員の先生に厚く御礼申し上げます。また、本稿は国立国語研究所の共同研究プロジェクト「多様な言語資源に基づく日本語非母語話者の言語運用の応用的研究」のサブプロジェクト「日本語学習者の作文の縦断コーパス研究」（プロジェクトリーダー：石黒圭）、および、JSPS科研費JP21H04417の研究成果である。本研究の遂行にあたり、国立国語研究所の上記プロジェクトのスタッフ各位、調査協力校の教員・学習者のみなさまのご協力を得た。記して感謝申し上げます。

¹『現代日本語書き言葉均衡コーパス（BCCWJ）』の統計結果に基づき、記号や空白を除いた文字のみの比率をさす。

減少しつつある。このような状況をふまえ、日本語学習者の作文にかんする研究でも、石毛(2016)や布施・石黒(2018)などのように、パソコンのキーボード入力を介した作文の執筆過程への注目が高まっている。キーボード入力による作文とその執筆過程を通じて、従来の手書き作文では観察できなかった漢字使用の特徴や潜在的な問題点を分析することが可能となった。国立国語研究所の共同研究プロジェクトで現在構築中の『日本語学習者作文縦断コーパス (W-CoLeJa)』(以下、W-CoLeJa)には、異なる母語をもつ日本語学習者の4年間にわたる作文テキストデータが収録されているだけでなく、その執筆過程を縦断的に追った作文執筆プロセスデータも収集されている。本稿では、W-CoLeJaを用い、CJLがパソコンを用いて作成した作文の産出結果であるテキストデータと、その結果にいたるまでの執筆プロセスデータの両方に焦点をあて、CJLの作文における漢字の使用の実態を報告する。

2. 先行研究と本研究の位置づけ

CJLを対象とした漢字の習得および使用についての研究は、主に漢字で表記されている漢字語彙や漢語に焦点をあてた研究(河住 2005, 小室 2019 など)、漢字の読み書き能力にかんする研究(阿久津 1991, 邱 2003, 山田 2015, 張ほか 2018 など)が中心である。そのうち、本研究とくに関連が深いのは、CJLの作文における漢字や漢字語彙の使用にかんする研究であり、漢字の使用に注目した金澤(2019)、福田(2016)、語彙の使用に焦点をあてた石黒(2004)、費(2020)などが挙げられる。

金澤(2019)は、学習者の手書き作文が収録された『YNU 書き言葉コーパス』を用いてCJLの作文における漢字の使用はレベルが上がるにつれて減少していくと指摘し、「熟語の形になりやすい「音読み」漢字の使用例が多く、その中には日本語では一般的でない語彙も含まれ」(金澤 2019: 197)ることが下位群の作文における漢字の使用が多い要因であると分析している。同じく学習者の手書き作文を対象にした福田(2016)では、CJLの作文における漢字の使用は非漢字圏よりはるかに多いものの、動詞に漢字を使わないなど個人差があることを述べ、CJLであっても必ずしも全般的に漢字を使用しているわけではないことに言及している。一方、石黒(2004)は漢語に焦点をあて、日本に留学しているCJLと中国の大学にいるCJLの作文を用いて分析し、CJLの作文における漢語副詞の多用が文体の違和感や文体的厳密さの欠如につながると指摘している。また、『北京日本語学習者縦断コーパス (B-JAS)』²を用いてCJLの作文における漢字語彙の使用について分析した費(2020)では、漢字語彙の習得過程における母語の影響について以下の3点を指摘していた：①漢字語彙の割合は日本語習得の初期段階で有意に高いが、過剰な使用は作文の流暢さに悪影響を与える、②日本語習得の進行にともない、日中同形同義語、異形語、同形異義語の使用割合はそれぞれ減少する、③初期段階では漢字語彙の割合と作文のレベルに負

² 北京日本語学習者縦断コーパス (B-JAS: Beijing Corpus of Japanese as a Second Language) は日本語を勉強している中国語母語の大学生の発話および作文データを、4年間にわたって収集した縦断コーパスである。1年次は年2回、2～4年次は年3回行い、全11回のデータを収集している。(HP: <https://www2.ninjal.ac.jp/jll/bjas/bjasindex.html>)

の相関が見られるが、上級になると正の相関が強まり、母語の影響は習得過程で変化する。

上述の先行研究において、金澤（2019）と費（2020）は公開されている学習者コーパス（『YNU 書き言葉コーパス』, B-JAS）を用いているのに対し、石黒（2004）と福田（2016）は自ら収集した学習者の作文をデータベース化している。なかでも B-JAS のデータは、本研究の研究対象である W-CoLeJa と同様、4 年間にわたる作文調査によって収集されたものである。費（2020）が分析に使用したのは B-JAS の 5 回分の作文データであり、各回の作文テーマは異なっているが、テーマ間の差異には言及せず、5 回分をとおして 4 年間の変化を一括して分析している。しかし、作文テーマが異なれば使用語彙や表記の選択も変化するため、日本語習得過程における漢字や漢字語彙の使用の変化を考察するには、テーマによる影響を考慮する必要があると思われる。そこで本研究では、CJL の漢字使用が日本語の学習とともに変化するか、またどのように変化するかを明らかにするため、CJL が異なる学習時期に書いた同一テーマの作文を用いて分析を行う。

さらに、費（2020）や福田（2016）などでは、学習者の作文における「漢字語彙」を抽出して分析しているが、「漢字語彙」がどのような語をさすのかについては、明確な定義が十分に定まっていない。実際には漢語として一括して扱われる場合もあれば、「食べる」のように 3 文字のうち 2 文字がひらがなで書かれた語も「漢字語彙」として扱われるケースがある。このように、先行研究での「漢字語彙」の定義はゆれており、語単位で集計するか字単位で集計するかについても統一的な基準が確立されていない。

以上の点をふまえると、CJL がパソコンを用いた作文における漢字の使用について、文字種の使い分けという観点からの研究はまだ十分に進んでいないのが現状である。

本研究の目的は、CJL が日本語学習の 4 年間にわたってパソコンで書いた作文における漢字の使用の実態を明らかにすることである。そのため、作文のプロダクトデータとプロセスデータの両方を用い、同一群の学習者が書いた作文について異なるテーマや学年による差に着目し、異なる母語をもつ日本語学習者との比較を通じて CJL の漢字使用の特徴を考察する。さらに、作文執筆過程でのキーボード入力方法を記録したプロセスデータを用いて分析する点で先行研究と大別する。具体的には以下のようなリサーチクエスションを設定した。

RQ1 テーマ別・学年別に、CJL の作文における漢字の使用の多寡に差があるか

RQ2 CJL がパソコンを用いて作文を執筆したさい、漢字をどのように入力しているのか

なお、CJL の漢字使用の特徴を考察するにあたり、中国語が母語ではない学習者との比較が不可欠である。そこで、同じく W-CoLeJa に収録されたベトナム語を母語とする学習者（Vietnamese Japanese Learner: VJL）の作文データを取り上げ、CJL の比較対象とした。VJL を選定した理由は次の 2 点である。第一に、中国語圏の CJL と同様に、2024 年 12 月現在、4 年間の調査が完了しており、学習者の 4 年間にわたる推移を観察できる点でデータの均質性が確保されている。第二に、VJL の母語であるベトナム語には漢字が使用されておらず、非漢字圏学習者として CJL と比較することで、日本語学習者の作文における漢字の使用に母語が与える影響を検討できる点である。とくに、漢字圏学習者と非漢字圏学習者のそれぞれの漢字習得過程や使用傾向を明らかに

することは、異なる母語知識をもつ学習者に向けた漢字教育や教材開発において新たな知見を提供することが可能である。

3. 分析データについて

3.1 分析データにあたる調査の詳細

本プロジェクトでは W-CoLeJa の構築にあたり、海外の複数の大学と連携し、日本語学専攻の大学生を対象に作文調査を実施した。具体的には、1 年に 3 回異なるテーマで学習者に執筆を依頼し、それを 4 年間継続し計 12 回の調査を行っている。本稿ではそのうち中国語圏の 2 大学 (CCA 大学, CCC 大学) とベトナム語圏の 1 大学 (VVA 大学) から収集した作文データを分析対象とし、全 12 回に参加した 90 名の調査協力者による 1,080 本の作文を取り上げる (表 1)。調査は、2020 年春から 2023 年夏にかけて CCA 大学のみがパイロット調査として先行実施され、CCC 大学および VVA 大学は、ほぼ同時期の、2021 年春から 2024 年夏にかけて実施された。作文のテーマは「A1, A2, A3」の 3 種があり、それぞれ「うれしかったプレゼント」(体験文), 「私の好きな有名人」(説明文), 「写真と動画」(意見文) である。

表 1 12 回作文調査の詳細 (実施時期は若干前後する場合がある)

学年	調査回数	実施時期 (開始 - 終了)		作文のテーマ	CJL の人数	VJL の人数
1	01	2020/03	2021/03	A1	70	20
	02	2020/05	2021/05	A2	70	20
	03	2020/06	2021/06	A3	70	20
2	04	2020/12	2021/12	A1	70	20
	05	2021/03	2022/03	A2	70	20
	06	2021/05	2022/05	A3	70	20
3	07	2021/12	2022/11	A1	70	20
	08	2022/03	2023/03	A2	70	20
	09	2022/06	2023/06	A3	70	20
4	10	2022/11	2023/09	A1	70	20
	11	2023/03	2023/11	A2	70	20
	12	2023/06	2024/03	A3	70	20
作文本数の合計					840	240

3.2 分析データの概要

まず、匿名化済みの作文プロダクトデータにたいし、正規表現を用いて文字種別に字数を算出し、分析対象であるデータの概要を把握する³。その結果、CJL の作文には、合計 477,493 字、異なり 2,807 字、延べ 320,800 語、異なり 9,088 語があることがわかった。VJL の作文には、延べ 146,437 字、異なり 1,633 字、延べ 106,618 語、異なり 4,250 語があることがわかった。

次に、テーマ別、学年別に、作文 1 本あたりの字数の平均値を算出する。表 2 に示すように、

³ 文字数の集計は、空白を除いていた。語数の集計については、形態素解析エンジンには MeCab (ver.0.996) を、解析用辞書は unidic-cwj-2023.02 を用い、語数は空白と補助記号も含めている。

CJL と VJL の作文において、テーマ間の差より学年別の差が注目される。表 2 から、CJL と VJL はいずれも学年が上がるにつれて作文における平均字数が増加する傾向があるが、3 年目と 4 年目との差は小さいことがわかった。CJL は 4 年目で 728 字、VJL は 3 年目で 757 字と、それぞれ最大値を示した。テーマ間の差から見てみると、CJL は平均的に A2（説明文）＞ A1（体験文）＞ A3（意見文）の順で作文 1 本あたりの字数が減少し、意見文の字数がもっとも少なかった。それとやや異なる傾向を示した VJL は、A2（説明文）＞ A3（意見文）＞ A1（体験文）の順で体験文の字数がもっとも少ないという結果が見られた。ただし、CJL と VJL の両者において、もっとも多い字数をもつのが説明文である点は共通している。その理由については 2 点が考えられる。1 つは、説明文のテーマである「私の好きな有名人」を執筆するさい、学習者が自身の本当の趣味や興味を自由に表現できるため、内容が充実しやすい点である。もう 1 つは、有名人についての作文では、人名や地名、作品名などの固有名詞が使用され、その固有名詞の意味の説明なども含めた結果として字数が増加したと考えられる。

表 2 作文 1 本あたりの字数の平均値

学年	CJL				VJL			
	A1	A2	A3	平均	A1	A2	A3	平均
1	327	318	315	320	314	423	397	378
2	496	531	491	506	505	605	545	552
3	721	725	712	719	745	798	729	757
4	747	753	683	728	722	790	748	753
平均	573	582	550	568	571	654	605	610

また、CJL と VJL の作文における文字種別の集計をし、その結果を次の表 3 にまとめた。「粗頻度」とは実際の出現頻度を、「調整頻度」とは百万字あたりの出現頻度をさし、「調整頻度 = 各文字種の延べ数の粗頻度 ÷ 総文字数 × 1,000,000 字」によって算出したものである。なお、表 3 のなかの「漢字」とは、日本語の漢字のみならず、中国語の簡体字と繁体字も含めてカウントしたものである。表 3 から、CJL の作文では、記号も含めて集計した結果、漢字が約 30%、ひらがなが約 55%、カタカナが約 7% を占めていることがわかる。VJL の作文と比較すると、漢字の割合が 6 ポイント高く、ひらがな、カタカナ、アルファベットの割合が 1 ～ 2 ポイント低いことがわかる。

表3 作文における文字種別の集計

文字種	CJL			VJL		
	粗頻度	調整頻度	割合 (%)	粗頻度	調整頻度	割合 (%)
漢字	139,003	295,479	29.55	34,434	235,145	23.51
ひらがな	259,656	551,952	55.20	83,121	567,623	56.76
カタカナ	32,957	70,057	7.01	13,467	91,964	9.20
アルファベット	1,736	3,690	0.37	2,504	17,100	1.71
アラビア数字	1,859	3,952	0.40	721	4,924	0.49
記号	35,221	74,869	7.49	10,321	70,481	7.05
計	470,432	1,000,000	100.00	146,437	1,000,000	100.00

4. テーマごとの作文における漢字使用率の学年別推移

本節では、RQ1「テーマ別・学年別に、CJL の作文における漢字の使用の多寡に差があるか」を明らかにするため、漢字使用率⁴という指標を用い、テーマごとの作文における漢字使用率の学年別の推移を確認する。漢字使用率は「総漢字数 ÷ (総文字数 - 記号の延べ数 - スペースの延べ数)」によって算出した。次頁の表4にはCJLとVJLの作文における漢字使用率の平均値と最大値、最小値をまとめた。

表4から、中国のCJLの作文では漢字使用率の平均値は0.27～0.36の範囲内に収まっており、最小値は0.12～0.25であり、最大値は0.39～0.56と半数を超えているケースも見られた。最大値である0.56は、4年生の11回目の「私の好きな有名人」というテーマの作文で出現しており、この学習者は中国の歴史人物「諸葛亮」を取り上げ、その説明のために「清廉潔白」「晴耕雨読」のような中国語に由来する四字熟語の漢語や、「空城計」「劉備」「漢室」のような中国語にある固有名詞などを多用していたことが、漢字使用率がとくに高い要因であると考えられる。また、学年別の大きな変化は見られなかったが、テーマ別ではA2(説明文) > A3(意見文) > A1(体験文)の順で漢字使用率が低下傾向にあった。

一方、ベトナムのVJLの作文では、漢字使用率の平均値は0.15～0.30とやや低く、学年が上がるにつれて高くなる傾向が見られた。最大値は0.38にとどまり、最小値にかんしては0.004(表4では四捨五入して0.00となっている)と、250字の作文のなかで漢字1字のみ使用されているケースもあった。また、テーマ別による漢字使用率の差については、CJLの作文ではA2(説明文)がもっとも高い漢字使用率を示したのに対し、VJLの作文ではA3(意見文)の漢字使用率がもっとも高かった。ただし、A1(体験文)がもっとも低い漢字使用率を示している点では、CJLとVJLは同様の傾向を示している。このように、テーマや学年、母語の違いにより、漢字使用率に異なる傾向が見られることがわかった。

⁴ 本稿は、日本語学的立場での「漢字含有率」という概念を援用し、漢字の延べ数が総文字数における割合を「漢字使用率」と定義した。『日本語学大辞典』では、「漢字含有率」の定義について、「一定の文章中に用いられる文字(場合によって表記記号も加える)の総数に対する漢字数の比率」(野村 2018: 173)のように記述されている。

表4 CJL と VJL の作文における漢字使用率

テーマ	学年	CJL			VJL		
		最大値	最小値	平均値	最大値	最小値	平均値
A1	1	0.40	0.12	0.27	0.25	0.00	0.15
	2	0.42	0.16	0.27	0.28	0.17	0.22
	3	0.39	0.15	0.28	0.32	0.15	0.24
	4	0.41	0.14	0.28	0.34	0.14	0.26
A2	1	0.54	0.21	0.36	0.29	0.05	0.17
	2	0.55	0.22	0.36	0.38	0.16	0.25
	3	0.51	0.22	0.34	0.35	0.18	0.26
	4	0.56	0.19	0.36	0.36	0.21	0.28
A3	1	0.45	0.25	0.33	0.36	0.14	0.24
	2	0.41	0.23	0.32	0.32	0.23	0.27
	3	0.41	0.23	0.32	0.38	0.13	0.29
	4	0.41	0.24	0.34	0.35	0.24	0.30

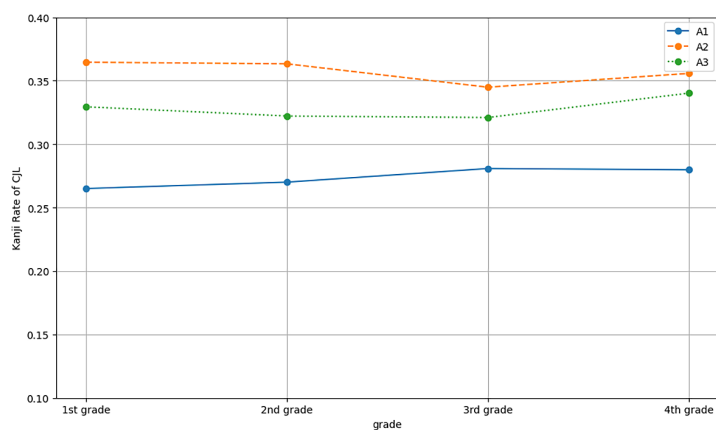


図1 CJL の漢字使用率の平均値の推移

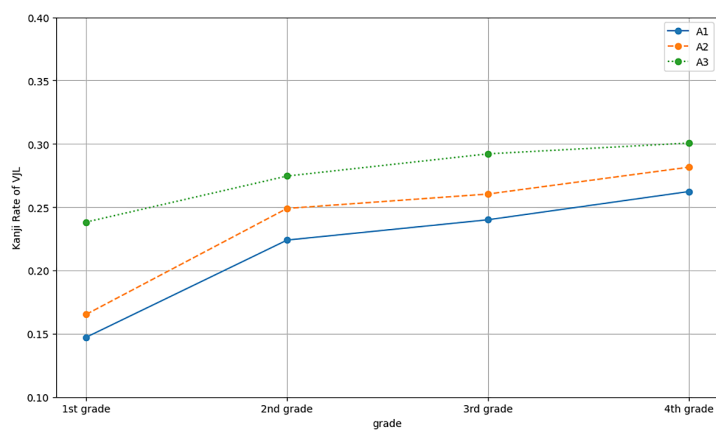


図2 VJL の漢字使用率の平均値の推移

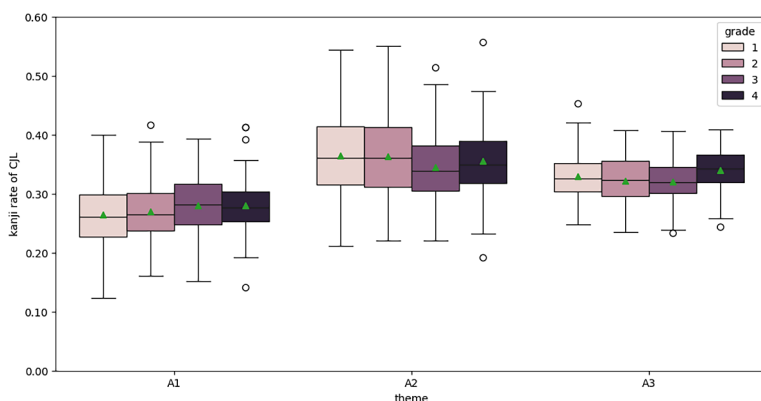


図3 CJLの漢字使用率の分布

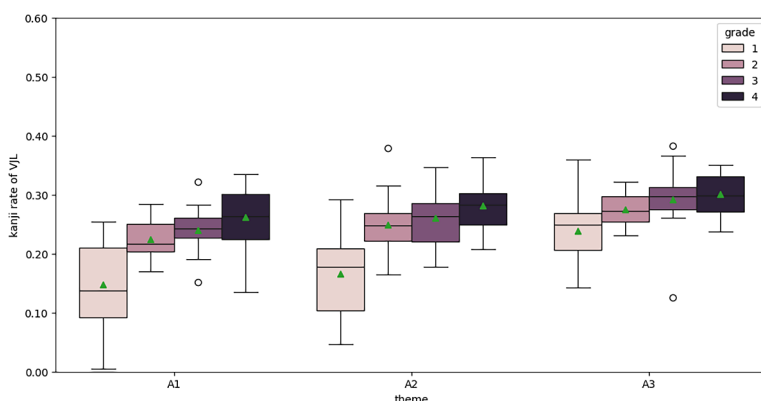


図4 VJLの漢字使用率の分布

図1と図2は作文における漢字使用率の1年目から4年目までの推移を3つのテーマごとに示したもので、図3と図4は作文における漢字使用率の学年別の分布をテーマごとに示している。VJLと比べ、CJLの作文における漢字使用率は学年別で明確な増減パターンが見られず、0.10以内でゆれていることがわかった。また、CJLはテーマごとの漢字使用率のばらつきが小さく、テーマA2においてはややばらつきが広がったものの、中央値は比較的安定していることが見られた。それにたいし、VJLの漢字使用率は、とくに1年目では漢字使用率をもっとも低く、学習者間のばらつきも顕著であった。

そこで、テーマ別および学年別の漢字使用率の平均値が統計的に有意に異なるかを検証するため、CJLとVJLそれぞれにたいしてテーマと学年の主効果および交互作用効果を評価する対応ありの2要因分散分析を行った。それぞれの結果は以下のとおりである。

CJLについて、テーマの主効果は有意であった（テーマ、 $F(1.78, 122.77) = 141.8486, p < .0001, \eta^2 = 0.3066$ ）が、学年の主効果は有意ではなかった（学年、 $F(3, 207) = 1.5730, p = .1970, \eta^2 = 0.0032$ ）。

また、テーマと学年の交互作用は有意であった（テーマ × 学年, $F(5.1, 352.08) = 5.86, p < .05, \eta^2 = 0.0109$ ）。VJL について、テーマの主効果は有意であった（テーマ, $F(2, 38) = 31.9826, p < .0001, \eta^2 = 0.1278$ ）が、学年の主効果も有意であった（学年, $F(1.95, 36.97) = 29.0018, p < .0001, \eta^2 = 0.3045$ ）。また、テーマと学年の交互作用は有意ではなかった（テーマ × 学年, $F(4.05, 77.04) = 2.0850, p = 0.09, \eta^2 = 0.0207$ ）。

この結果から、CJL について、異なるテーマによっては作文における漢字使用率には有意な差があるが、異なる学年によっては有意な差はないことが示された。一方、VJL は学年とテーマの両方が作文における漢字使用率に有意な影響を与えていることが確認され、とくに学年の主効果が大きかった。また CJL のように明確な有意差ではないものの、テーマと学年の交互作用に有意傾向（ $p < .10$ ）が見られた。

本節では、RQ1「テーマ別・学年別に、CJL の作文における漢字の使用の多寡に差があるか」について、VJL と比較した CJL の漢字使用率の特徴には次の 2 点があると確認できた。

- ① 同じテーマで 4 年間の推移を見ると、CJL は日本語学習の初期段階から作文における漢字使用率が高く、学年が上がっても大きな変動が見られなかった。それにたいし、VJL は学習開始時の漢字使用率が低く、学習時間とともに徐々に上がっていく傾向が見られた。また、学習者間でのばらつきについても、CJL は学年別の変化が小さく、安定した分布を示していた。一方、VJL は日本語学習の初期段階では学習者間でのばらつきが大きく、学年が上がるにつれてそのばらつきが徐々に収まる様子が観察された。
- ② 同じ学年で異なるテーマ間の差異を見ると、作文における漢字使用率はテーマによって異なることがわかった。CJL と VJL の両者とも体験文（A1）では比較的低い漢字使用率を示しているが、説明文（A2）と意見文（A3）ではやや高い傾向がある。このなかで、CJL は説明文における漢字使用率をもっとも高かったが、その理由として、説明文の作文 1 本あたりの字数が多いこと、また漢字表記である漢語や中国語の固有名詞が多用されていたことが挙げられる。一方、VJL は「好きな有名人」という説明文の執筆において、カタカナ表記やアルファベット表記の人名が多く使用され、20 人中 3、4 人程度しか漢字表記の人名を記述していなかった。そのため、CJL のように説明文における漢字使用率をもっとも高いという結果にはならなかった。

5. パソコンを用いて執筆した作文における漢字の入力

本節では、RQ2「CJL がパソコンを用いて作文を執筆したさい、漢字をどのように入力しているのか」を明らかにするため、作文の執筆過程を記録しているプロセスデータを参照して、CJL と VJL が漢字を入力する過程を観察する。5.1 では、CJL の作文に多く出現している漢字表記の固有表現を代表例として取り上げ、CJL がパソコンで作文を執筆したさいの漢字の入力方法について考察する。5.2 では、CJL と VJL の双方がパソコンを用いて執筆した作文における漢字の入力過程を分析し、そこで現れる問題点や困難点を考察する。

5.1 CJL の漢字の入力方法

CJL の作文には、中国語の語句の引用、中国の人物やスポットの紹介などが多いため、中国語の固有表現が頻繁に登場している。とくに「私の好きな有名人」というテーマの作文では、中国の人名や地名が多く出現している。たとえば、「周恩来」や「袁隆平」といった歴史上の人物や、「迪丽热巴」といった現在の芸能人の名前、「爱奇艺」のような動画サイト名、さらには「胡同」「四合院」のような中国伝統の概念をさす普通名詞も見られた。

プロセスデータを確認した結果、CJL が中国語の固有表現の漢字を入力するさいには、日本語のローマ字入力と中国語のピン音入力の2つの方法を使用していることが明らかになった。ここで、具体例として「袁隆平」の入力過程を取り上げ、実際の執筆過程を記録したプロセスデータの一部を提示する。プロセスデータは、作文を執筆すると同時に自動生成されている。本調査は、本プロジェクトが開発した Essaylogger という作文執筆ソフトを用い、学習者が必ずパソコンを使って執筆するように依頼した。Essaylogger の機能は、文字入力時のパソコンのキーストロークを入力時間も含めてすべて記録する（石黒 2025）ことであり、本稿に示すプロセスデータの例は次の図5のように2列構成となっている。1列目は学習者のキーボード操作を示し、どのキーを押したか、どのタイミングで変換操作を行ったかなどの過程である。2列目はキーボード入力後にパソコンの画面上で表示された結果を示し、カーソル前後の文脈を確認できる内容となっている。

操作内容2	カーソル前後の文字列
KeyE	生まれました。彼はえ ^\
KeyN	生まれました。彼はえ n ^\
変換	生まれました。彼は円 ^\
変換	生まれました。彼は縁 ^\
変換	生まれました。彼は園 ^\
変換	生まれました。彼は延 ^\
変換	生まれました。彼は宴 ^\
変換	生まれました。彼は塩 ^\
変換	生まれました。彼はえん ^\
変換	生まれました。彼は袁 ^\
変換確定	生まれました。彼は袁 ^\

【袁の入力過程：EN→袁】

操作内容2	カーソル前後の文字列
変換	ました。彼は袁law ^\
変換	ました。彼は袁low ^\
変換	ました。彼は袁row ^\
BackSpace	ました。彼は袁row ^\
BackSpace	れました。彼は袁ro ^\
変換	れました。彼は袁ro ^\
変換	れました。彼は袁 r o ^\
BackSpace	れました。彼は袁 r o ^\
BackSpace	まれました。彼は袁 r ^\
BackSpace	生まれました。彼は袁 ^\
KeyR	まれました。彼は袁 r ^\
KeyO	まれました。彼は袁ろ ^\
変換	まれました。彼は袁ロ ^\
変換	まれました。彼は袁炉 ^\
変換	まれました。彼は袁露 ^\
変換	まれました。彼は袁ろ ^\
変換	まれました。彼は袁魯 ^\
変換	まれました。彼は袁櫓 ^\
変換	まれました。彼は袁紹 ^\
変換	まれました。彼は袁路 ^\
変換	まれました。彼は袁路 ^\
変換	まれました。彼は袁櫓 ^\
変換	まれました。彼は袁芦 ^\
変換	まれました。彼は袁蘆 ^\
変換	まれました。彼は袁路 ^\
変換	まれました。彼は袁瀘 ^\
変換	まれました。彼は袁戸 ^\
変換	まれました。彼は袁ロ ^\
変換	まれました。彼は袁櫓 ^\
変換	まれました。彼は袁路 ^\
変換	まれました。彼は袁鷺 ^\

【隆の入力過程②：ROU→RO→変換失敗】

操作内容2	カーソル前後の文字列
KeyR	まれました。彼は袁 r ^\
KeyO	まれました。彼は袁ろ ^\
KeyU	れました。彼は袁ろう ^\
変換	まれました。彼は袁朗 ^\
変換	まれました。彼は袁郎 ^\
変換	まれました。彼は袁老 ^\
変換	まれました。彼は袁勞 ^\
変換	れました。彼は袁ろう ^\
変換	まれました。彼は袁楼 ^\
変換	れました。彼は袁ロウ ^\
変換	まれました。彼は袁蠅 ^\
変換	まれました。彼は袁蠅 ^\
変換	まれました。彼は袁蠅 ^\
変換	まれました。彼は袁蠅 ^\
変換	まれました。彼は袁廊 ^\
変換	まれました。彼は袁浪 ^\
変換	まれました。彼は袁漏 ^\
変換	まれました。彼は袁弄 ^\
変換	まれました。彼は袁瑯 ^\
変換	まれました。彼は袁蠅 ^\
変換	まれました。彼は袁妻 ^\
変換	まれました。彼は袁榔 ^\
変換	まれました。彼は袁淹 ^\
変換	まれました。彼は袁狼 ^\
変換	まれました。彼は袁竜 ^\
変換	まれました。彼は袁龍 ^\
変換	まれました。彼は袁耀 ^\
変換	まれました。彼は袁露 ^\
変換	まれました。彼は袁隴 ^\

【隆の入力過程①：ROU→変換失敗】

操作内容2	カーソル前後の文字列
変換	まれました。彼は袁盧 ^\
BackSpace	まれました。彼は袁盧 ^\
BackSpace	生まれました。彼は袁 ^\
KeyR	まれました。彼は袁 r ^\
KeyY	れました。彼は袁 r y ^\
KeyU	れました。彼は袁りゅ ^\
KeyU	ました。彼は袁りゅう ^\
変換	まれました。彼は袁流 ^\
変換	まれました。彼は袁劉 ^\
変換	まれました。彼は袁竜 ^\
変換	まれました。彼は袁龍 ^\
変換	まれました。彼は袁柳 ^\
変換	まれました。彼は袁留 ^\
変換	まれました。彼は袁立 ^\
変換	まれました。彼は袁隆 ^\
Digit8	まれました。彼は袁隆 ^\
変換確定	まれました。彼は袁隆 ^\

【隆の入力過程③：RO→RYUU→隆】

操作内容2	カーソル前後の文字列	操作内容2	カーソル前後の文字列
KeyH	れました。彼は袁隆 h ^	KeyH	れました。彼は袁隆 h ^
KeyE	れました。彼は袁隆へ ^	KeyE	れました。彼は袁隆へ ^
KeyE	ました。彼は袁隆へえ ^	KeyY	ました。彼は袁隆へ y ^
変換	ました。彼は袁隆へえ ^	KeyI	ました。彼は袁隆へい ^
変換	ました。彼は袁隆経え ^	変換	れました。彼は袁隆兵 ^
変換	ました。彼は袁隆経得 ^	変換	れました。彼は袁隆堀 ^
変換	ました。彼は袁隆へ得 ^	変換	ました。彼は袁隆へイ ^
変換	ました。彼は袁隆兵衛 ^	変換	ました。彼は袁隆へい ^
変換	ました。彼は袁隆へエ ^	変換	れました。彼は袁隆幣 ^
変換	ました。彼は袁隆へえ ^	変換	れました。彼は袁隆弊 ^
変換	ました。彼は袁隆経え ^	変換	れました。彼は袁隆閉 ^
変換	ました。彼は袁隆経得 ^	BackSpace	れました。彼は袁隆閉 ^
変換	ました。彼は袁隆へ得 ^	BackSpace	まれました。彼は袁隆 ^
変換	ました。彼は袁隆兵衛 ^	KeyH	れました。彼は袁隆 h ^
変換	ました。彼は袁隆へエ ^	KeyE	れました。彼は袁隆へ ^
BackSpace	ました。彼は袁隆へエ ^	KeyY	ました。彼は袁隆へ y ^
BackSpace	れました。彼は袁隆へ ^	KeyI	ました。彼は袁隆へい ^
BackSpace	まれました。彼は袁隆 ^	変換確定	れました。彼は袁隆平 ^

【平の入力過程①：HEE→変換失敗】

【平の入力過程②：HEYI→平】

図 5 「袁隆平」のローマ字入力過程 (CCC178-A2-01-02)

たとえば、図 5 は CJL の 1 年生が 2 回目の作文で「袁隆平」の 3 文字を入力する過程を示している。まず、【EN→袁】の部分は、【EN→袁】という日本語ローマ字入力から漢字変換にいたる過程を示した。キー「E」とキー「N」を入力したあと、変換候補として「円」「縁」「園」「延」など 6 つの漢字が順に提示されてきたが、7 つ目に目標とする漢字「袁」が表示され、変換を成功していた。

「袁」は「EN」と入力することで即座に変換が成功したが、一方で、「隆」は「ROU→RO→RYUU」, 「平」は「HEE→HEYI」と複数回の変換を経た試行錯誤の末に目標の漢字が表示されていた。これら漢字 3 文字の入力過程を確認したところ、合計で 1 分以上もかかっていた。このように、手書きであればすぐ書ける漢字でも、パソコンで書くには日本語での読み方を把握していなければスムーズに入力できないことが明らかになった。

また、それとは違ってピン音入力によるケースも観察された。図 6 は同じく CJL の 1 年生が 2 回目の作文で「袁隆平」の 3 文字を入力する過程を示しているが、5 秒以内でピン音入力によって成功している様子がうかがえた。

操作内容2	▼ カースル前後の文字列 ▼
KeyY	^
KeyU	^
KeyA	^
KeyN	^
KeyL	^
KeyO	^
KeyN	^
KeyG	^
KeyP	^
KeyI	^
KeyN	^
KeyG	^
変換確定	袁隆平 ^

【YUANLONGPING→袁隆平】

図6 「袁隆平」のピン音入力過程
(CCC009-A2-01-02)

a	ilibili や 「 ^
i	ilibili や 「 ^
q	ilibili や 「 ^
i	ilibili や 「 ^
y	ilibili や 「 ^
i	ilibili や 「 ^
	bili や 「爱奇艺 ^

【AIQIYI→爱奇艺】

図7 簡体字「爱奇艺」のピン音入力過程
(CCA009-A2-01-02)

以上をまとめると、CJL が中国語の固有表現の漢字を入力するには、ローマ字入力とピン音入力の2つのパターンを使用し、それぞれ以下の傾向が見られた。

まず、日本語のローマ字入力によって生成されている例として、中国の人名や地名などが挙げられるが、これらの漢字は日本語と同じ形をもち、授業で扱ったことがある、もしくは日本語で聞き覚えがある可能性が高いため、日本語のローマ字入力が第一選択となると考えられる。

一方、中国語のピン音入力によって生成されている例には、図6のみならず、次の図7に示す「爱奇艺」や「迪丽热巴」「山川异域风月同天」などのような日本語にはない簡体字も含まれている。これらの漢字を入力するときに、日本語のローマ字入力ではなく、中国語のピン音入力を選択した理由として、日本語での聞き覚えがない、または日本語に対応する日本語の漢字を把握していないのではないかと推測される。

5.2 CJL と VJL の漢字の入力にかかわる問題点

5.1 では、CJL がローマ字入力で漢字を入力することを述べているが、とくに漢字の読み方を知らない場合、CJL は漢字の音読みのローマ字つづりを入力する傾向が見られた。たとえば、次の図8に示すように「ビールの都」と書きたい場合には、訓読みの「都 (miyako)」が用いられるべきであるが、その訓読みの発音を把握していないためか、既習単語である「都市 (toshi)」を先に入力し、後ろの「市」を削除することで目標漢字「都」を残すという方略が使用されている。また、図9は中国の都市名である「青島」を書く例であるが、ここでも「青島」の2漢字それぞれの音読みを入力しようとする様子が見られた。その候補には目標の漢字が出てこなかったため、「青島」の中国語音のローマ字つづりを入力した結果となった。

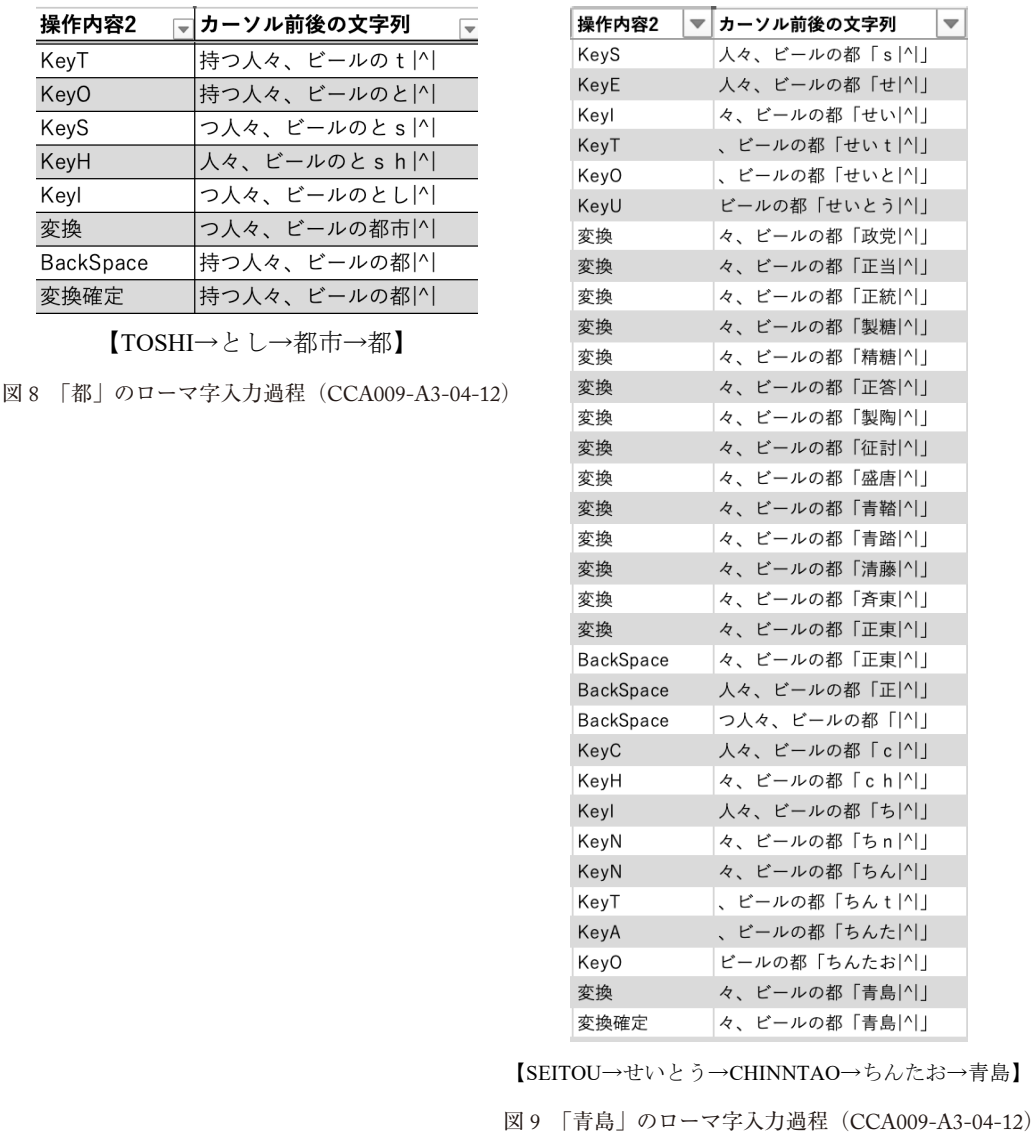


図 9 「青島」のローマ字入力過程 (CCA009-A3-04-12)

このように、CJL は知らない漢字（または漢字の読み方）を使用したい場合、既習の音である音読みのローマ字つづりを入力する傾向がある。一方、VJL の「私の好きな有名人」というテーマの作文のプロセスデータを確認すると、CJL とは異なる方略を使っていることがわかった。未習の漢字音についてどう処理しているのかという点について、VJL は知らない漢字（または漢字の読み方）や、知っているつもりで実際には習得していない漢字（または漢字の読み方）を使用したい場合、漢字の音読みと訓読みを併用する傾向が見受けられた。

操作内容2	カーソル前後の文字列	操作内容2	カーソル前後の文字列
KeyT	技と紬の短時間で心の ^\	変換確定	紬の短時間で心の戸枠 ^\
KeyO	と紬の短時間で心のと ^\	BackSpace	と紬の短時間で心の戸 ^\
KeyW	紬の短時間で心のとw ^\	KeyW	紬の短時間で心の戸w ^\
KeyA	紬の短時間で心のとわ ^\	KeyA	紬の短時間で心の戸わ ^\
KeyK	の短時間で心のとわk ^\	KeyK	の短時間で心の戸わk ^\
KeyU	の短時間で心のとわく ^\	KeyU	の短時間で心の戸わく ^\
BackSpace	紬の短時間で心のとわ ^\	変換	紬の短時間で心の戸枠 ^\
BackSpace	と紬の短時間で心のと ^\	変換	の短時間で心の戸わく ^\
KeyU	紬の短時間で心のとう ^\	変換	の短時間で心の戸沸く ^\
KeyW	の短時間で心のとうw ^\	変換	の短時間で心の戸ワク ^\
KeyA	の短時間で心のとうわ ^\	変換	の短時間で心の戸和久 ^\
KeyK	短時間で心のとうわk ^\	変換	の短時間で心の戸湧く ^\
KeyU	短時間で心のとうわく ^\	変換	紬の短時間で心の戸惑 ^\
BackSpace	の短時間で心のとうわ ^\	変換確定	紬の短時間で心の戸惑 ^\
BackSpace	紬の短時間で心のとう ^\	KeyI	の短時間で心の戸惑い ^\
BackSpace	と紬の短時間で心のと ^\	変換確定	の短時間で心の戸惑い ^\
KeyW	紬の短時間で心のとw ^\		
KeyA	紬の短時間で心のとわ ^\		
KeyK	の短時間で心のとわk ^\		
KeyU	の短時間で心のとわく ^\		
変換	紬の短時間で心のと枠 ^\		
変換	の短時間で心の間わく ^\		
変換	紬の短時間で心の都枠 ^\		
変換	紬の短時間で心のト枠 ^\		
変換	紬の短時間で心の十枠 ^\		
変換	紬の短時間で心の戸枠 ^\		
変換	紬の短時間で心の途枠 ^\		
変換	紬の短時間で心の戸枠 ^\		
変換確定	紬の短時間で心の戸枠 ^\		

【戸枠→戸わく→戸惑, I→い】

【TOWAKU→とわく→とうわく→とわく→戸枠】

図 10 「戸惑い」のローマ字入力過程 (VVA029-A2-03-08)

図 10 は VJL の 3 年生が 8 回目の作文で「戸惑い」という語を入力しようとしたさいの過程を示している。この学習者は和語である「戸惑い」の読み方「tomadoi」を「to」まで記憶しているようであり、まず「towaku」と入力した。結果として「to」で「戸」と変換が成功したあと、「惑い」を 2 つに分割して入力している。訓読みであるべき「惑」を音読みの「waku」と入力して変換し、その後「い」を「i」と入力して「戸惑い」を完成させた。この過程から、この学習者が「戸惑い」を 1 語としての読み方を十分に習得していなかったことがわかる。また、漢字「惑」の訓読みも理解していないため、「戸惑い」1 語全体としての入力ではなく、1 字ずつ分割して入力した結果となった。一方で、音読みであるべき箇所を訓読みのローマ字つづりを入力するケースもあった。次の図 11 に示すように、VJL の 1 年生が「有名人」という語を入力しようとしたさいに、「有名」と「人」の 2 つに分割し、音読みであるべき「人」を入力するときに訓読みのローマ字つづりを入力した。

操作内容2	カーソル前後の文字列
KeyY	私の好きな y ^
KeyU	私の好きなゆ ^
KeyM	私の好きなゆ m ^
KeyE	私の好きなゆめ ^
KeyI	私の好きなゆめい ^
BackSpace	私の好きなゆめ ^
BackSpace	私の好きなゆ ^
KeyU	私の好きなゆう ^
KeyM	私の好きなゆう m ^
KeyE	私の好きなゆうめ ^
KeyI	私の好きなゆうめい ^
変換	私の好きな有名 ^
KeyA	私の好きな有名 ^
BackSpace	私の好きな有名 ^
KeyH	私の好きな有名 h ^
KeyI	私の好きな有名ひ ^
KeyT	私の好きな有名ひ t ^
KeyO	私の好きな有名ひと ^

【YUUMEI→ゆうめい→有名，HITO→ひと→人】

図 11 「有名人」のローマ字入力過程（VVA033-A2-01-02）

操作内容2	カーソル前後の文字列
KeyT	私の t ^
KeyA	私のた ^
KeyN	私のた n ^
KeyN	私のたん ^
KeyY	私のたん y ^
KeyO	私のたんよう ^
KeyU	私のたんよう ^
KeyB	私のたんよう b ^
KeyI	私のたんようび ^
変換確定	私の単曜日 ^
変換確定	私の単曜日 ^
KeyN	私の単曜日 n ^
KeyO	私の単曜日の ^
Digit1	私の単曜日の 1 ^
Digit8	私の単曜日の 1 8 ^
KeyS	私の単曜日の 1 8 s ^
KeyA	私の単曜日の 1 8 さ ^
KeyI	私の単曜日の 1 8 さい ^

【TANNYOUBI→たんようび→単曜日】

図 12 「単曜日」のローマ字入力過程（VVA161-A1-01-01）

そのほか、図 12 に示すような漢字の誤字に気づけない問題が VJL の学習段階の各時期に見られた。図 12 は VJL の 1 年生が「うれしかったプレゼント」というテーマで 1 回目の調査で書かれた作文の冒頭を示している。この例では、「誕生日」の読み方を「tannyoubi」と誤認し、そのまま変換した結果が「単曜日」となっている。それにもかかわらず、その誤りに気づかず使用しており、誤字であることを認識していない様子がうかがえる。

6. まとめ

本稿では、CJL の作文における漢字の使用の実態について、VJL との比較や執筆プロセスデータとの参照を通じて分析を行った。

その結果、まず、RQ1「テーマ別・学年別に、CJL の作文における漢字の使用の多寡に差があるか」について、VJL と比較した CJL の漢字使用率の特徴は次の 3 点であると確認できた。①日本語学習の初期段階から漢字使用率が高く、学年が上がっても大きな変動が見られなかった。②漢字使用率はテーマによって異なり、体験文（A1）では比較的 low、説明文（A2）と意見文（A3）ではやや高い傾向がある。③学習者間でのばらつきについても学年別の変化が小さく、安定した分布が見られた。一方、VJL のような母語に漢字がない学習者は、学習開始時の漢字使用率が低く、学習時間とともに徐々に上がっていく傾向が見られ、CJL と対照的な結果となった。

次に、RQ2「CJL がパソコンを用いて作文を執筆したさい、漢字をどのように入力しているのか」については、作文の執筆過程の記録に注目したところ、説明文を書くさいの固有表現の入力において興味深い発見が得られた。具体的には、漢字表記をもつ固有表現を例に取り上げた執筆過程の分析から、CJL の漢字入力には日本語のローマ字入力と中国語のピン音入力の 2 つのバ

ターンを用いることが明らかになった。手書きと違って、キーボード入力の場合は、入力したい漢字の日本語の読み方がわからなかったり聞き覚えがなかったりすると、中国語の漢字知識があっても正確に入力することが難しい可能性がある。また、漢字の未習である日本語音を把握していないときには、既習の読み方である音読みを使って入力する傾向が見られた。一方、VJLの漢字入力については、CJLと異なる特徴が見られた。第一に、音読みであるべき箇所を訓読みで入力するなど、作文のプロダクトでは正しい入力結果となっているものの、実際には漢字の日本語音を正確に把握できていないことが背景にある。第二に、学習の初期段階から漢字表記にかかるミスが多く、高学年になっても誤字が散見されるケースがあった。

日本語教育における漢字教育では、「タイピングを組み合わせた教育はほとんど実施されていない」(中原・岩下 2019) 現状をふまえると、今後、漢字の音韻情報、とくに訓読みや音読みの使い分けに注意を払い、キーボード入力を前提とした漢字の指導が求められると考えられる。たとえば、同じ漢字がどの語で音読みとして使われ、どの語で訓読みとして使われるのかななどを明確にし、学習者の漢字学習における課題に対応した指導方法の調整が必要である。

実際、本稿で観察された漢字入力の方法や問題点について量的な分析が行えなかったため、それが学習者個人による特徴にすぎないのか、全体の傾向を表しているのかが不明であり、それを明らかにするための量的研究が求められる。今後の課題としては、日本語レベル別の違いや、日本語母語話者との比較、ほかの母語をもつ学習者との比較、特定の語における漢字の選択傾向などを含め、CJLの漢字使用の全体像をより包括的に検討する必要がある。また、本稿は中国語圏のうち、簡体字圏の学習者の作文データを用いて分析したが、繁体字圏の学習者の作文においても同様の使用傾向や入力方法が見られるかについては、さらなる研究が必要である。

参考文献

- 阿久津智 (1991) 「漢字圏の学生に対する漢字教育について」『筑波大学留学生教育センター日本語教育論集』6: 129-144.
- 石毛順子 (2016) 「中国語を母語とする日本語学習者におけるパーソナルコンピューターでの作文過程：手書きによる作文過程との比較から」『アカデミック・ジャパニーズ・ジャーナル』8: 19-27.
- 石黒圭 (2004) 「中国語母語話者の作文に見られる漢語副詞の使い方の特徴」『一橋大学留学生センター紀要』7: 3-13.
- 石黒圭 (2025) 「学習者の成長が追体験できる作文コーパスの構築」『中国語話者のための学習者コーパスを用いた作文の研究と教育』国立国語研究所機関拠点型基幹研究プロジェクト「多様な言語資源に基づく日本語非母語話者の言語運用の応用的研究」2024年度研究成果報告書：15-30.
- 金澤裕之 (2019) 「漢字使用の面から見た学習者の表現能力の向上」『日本語／日本語教育研究』10: 197-217.
- 河住有希子 (2005) 「中国人学習者の漢字語彙使用に見られる問題点」『早稲田大学日本語教育研究』7: 53-65.
- 胡佳芮 (2024) 「流行歌の歌詞における文字・表記の特徴：高頻度用例における文字種の選択に着目して」『一橋日本語教育研究』12: 59-72.
- 小室リー郁子 (2019) 「中国語話者の母語知識に配慮した漢字語彙教育：初級漢字名詞の表記に関する側面から」『JSL 漢字学習研究会誌』11: 60-69.
- 邱學瑾 (2003) 「台湾人日本語学習者の日本語漢字熟語の音韻処理について：単語タイプ・単語の習熟年齢・習熟度の観点からの検討」『日本語教育』116: 89-98.
- 張婧樟・玉岡賀津雄・初相娟 (2018) 「中国人日本語学習者は日本語の漢字の書き取りが正しくできるのか？」『中国語話者のための日本語教育研究』9: 52-68.

- 中原郷子・岩下真澄（2019）「日本語学習者における日本語漢字名詞産出時の出力形式の影響：漢字圏学習者と非漢字圏学習者の比較」『長崎外大論叢』23: 43–52.
- 野村雅昭（2018）「日本の漢字・II」日本語学会（編）『日本語学大辞典』172–174. 東京：東京堂出版.
- 福田佳与（2016）「初級作文における中国人学習者の漢字使用」『言語と交流』19: 93–99.
- 布施悠子・石黒圭（2018）「日本語学習者の作文執筆過程における自己修正理由：上級中国人学習者，上級韓国人学習者，日本語母語話者の作文の比較から」『国立国語研究所論集』15: 17–42.
- 山田祐也（2015）「中国人日本語学習者に対する効果的な漢字読みの呈示方法の検証：文字入力と音声入力の観点から」『JSL 漢字学習研究会誌』7: 29–36.
- 费晓东（費曉東）（2020）「基于历时语料库的中国学习者日语汉字词汇习得研究」『高等日语教育』2020(2): 30–42.

Use of Kanji in Essays by Learners of Japanese using Personal Computers: Targeting Native Chinese Speakers

HU Jiarui

Adjunct Researcher, NINJAL

Abstract

This study elucidates the use of kanji in essays by Chinese learners of Japanese (CJL) through a comparative analysis with Vietnamese learners of Japanese (VJL). Using essay data and writing process records from the Written Corpus of Learner Japanese (W-CoLeJa), two key findings emerged. First, the rates of kanji used among CJL remained consistently high from the early stages of learning, stabilizing over time. However, thematic differences were observed, with lower rates in narrative essays and higher rates in expository and editorial essays. In contrast, non-kanji-background learners like VJL showed a gradual increase in the use of kanji usage, following a different trajectory from CJL. Second, an analysis of kanji input methods by CJL revealed a combination of Japanese romaji input and Chinese pinyin input. Complex trial-and-error patterns were particularly evident when entering proper nouns. When unfamiliar with Japanese pronunciation of kanji, learners often relied on the Chinese readings. These findings highlight the need for Japanese language education in digital environments to emphasize phonological acquisition and provide guidance on keyboard input methods.

Keywords: rate of kanji, writing process, keyboarding, typing, W-CoLeJa