



中国語を母語とする日本語学習者の発話における漢語使用の実態：「多言語母語の日本語学習者横断コーパス」を用いた調査

鄧, 琪

(Citation)

統計数理研究所共同研究リポート, 435:121-138

(Issue Date)

2020-03

(Resource Type)

journal article

(Version)

Accepted Manuscript

(URL)

<https://hdl.handle.net/20.500.14094/90007761>



中国語を母語とする日本語学習者の発話における漢語使用の実態

—「多言語母語の日本語学習者横断コーパス」を用いた調査—

鄧 琪(神戸大学 大学院生)

dengqixq123@gmail.com

A Survey on the Use of Kanji words by Chinese Learners of Japanese

—Focus on the spoken language—

DENG Qi (Kobe University, Graduate Student)

概要

本研究は、大規模な日本語学習者コーパスを利用し、様々な習熟度段階にある中国語を母語とする日本語学習者の話し言葉における漢語の使用実態を明らかにすることを目指す。具体的には、(1)総語数・(2)総種類数・(3)語彙難度・(4)上位語の内容という点で、中国語を母語とする日本語学習者の漢語使用はどのような特徴を持っているのか、という4つのRQを設定し、また、それぞれのRQについて、(1)中国語を母語とする学習者の漢語使用は日本語母語話者、および、他の母語を持つ学習者と比べてどのような違いがあるか、(2)習熟度はどのように影響しているのか、という2つの観点から考察を行う。その結果、(1')中国語を母語とする学習者の漢語使用は不十分であること、(2')漢語使用が不十分であるという問題は、中国語を母語とする学習者だけではなく、韓国語および英語を母語とする学習者も持っている問題と言えること、(3')習熟度の上昇につれて、中国語を母語とする学習者の漢語使用は日本語母語話者に接近しているが、母語話者との距離はまだ大きいこと、が明らかになった。

キーワード

漢語使用, 学習者コーパス, 話し言葉

1. はじめに

日本語の語彙には和語・漢語・外来語・混種語という4種の語種が存在するが、その中、種類として一番多いのは漢語である(野村(1999), 陳(2003))。たとえば、現代日本語書き言葉均衡コーパス(BCCWJ)を例として挙げると、漢語は日本語の語彙全体の43.5%を占め、和語(33.1%)や外来語(18.6%),そして混種語(4.8%)よりも比率が高い。

漢語の習得についていうと、一般に、漢字の知識を持っている中国語を母語とする日本語学習者(L1-Chinese learners of Japanese, 以下はCLJと称する)は総じて有利な状況にあると思われる。たとえば、玉岡(1997)は、日本語と中国語の漢字の共通性が高いため、視覚呈示課題での漢字語処理においては、CLJが有利であると述べた。また、陳(2003)は、CLJの漢語学習にお

いては、日中で字形が同じである同形語が多いため、学習者にとって大きな助けになっていると述べた。

このように、CLJ は日本語の漢語習得において有利であることは否定できない事実である。しかしながら、このことは CLJ が日本語の漢語習得に何も問題を持たないことを意味するものではない。以下は、「日本語学習者書き言葉コーパス」における CLJ の漢語の誤用例である。

(1) 学生は、真面目に本を勉強して、いい初中(⇒中学校)に入るのため、頑張つてそれから、高中(⇒高校)と大学、最後に、いい仕事を探したいです。(作文番号:j0008)

(2) ですから、国慶節で 1 日の休みは十分で、こんな長い休みは需要(⇒必要)でなはいと思います。(作文番号:j0012)

上記の用例は CLJ であっても、日本語において漢語を正しく使うことは必ずしも容易ではないことを示す。過去の研究で、荒川(1979)は、日本人と中国人が「基本的」に同じ漢字を使っているということは、大きな利点にもなれば大きな「落とし穴」にもなっていると述べた。また、陳(2007)は、CLJ が日本語の文字や語彙を母語の中国語の形・音・義(意味)のスキーマを生かしながら認知し、母語と目標言語の発音、意味、用法などの相似性や相違性に目をむけて学んでいくという点は、他の学習者にはないメリットであるが、教師などが否定的証拠(「○○とは言わない」というようなデータ)を与えなければ、学習者の中間言語における誤りが固定化して矯正されない化石化現象が起きやすいデメリットもあると述べた。

では、どのような指導を行っていけばよいのであろうか。この点を明らかにするには、まず、CLJ が漢語使用に関して抱える問題を特定する必要があるだろう。この時、学習者の漢語使用の媒体としては書き言葉と話し言葉の 2 つがあるが、従来の研究は書き言葉を課題にし、話し言葉の分析は必ずしも十分ではなかった。また、CLJ に特有の問題点を明らかにするためには、日本語母語話者(Japanese native speakers, 以下は JNS と称する)と比較したり、異なる母語を持つ他の学習者と比較したりすることは重要であろう。そこで、本研究は、CLJ の話し言葉に注目し、様々な習熟度段階にある CLJ の漢語使用実態を調査し、JNS との間にどのような差が存在するのか、また、異なる母語を持つ学習者と比べて CLJ に特有のものであるのかを明らかにする。

2. 先行研究

以下、CLJ の漢語習得に関わる研究と、学習者コーパスを利用して CLJ の漢語使用実態を調査した研究、という 2 つの観点について概観する。

まず、CLJ の漢語習得に関わる研究について、たとえば、張(2017)は、CLJ の漢語習得について、いわゆる **Overlap** 語(同形類義語: 日中両言語における意味が一部重なるが、両者の間にずれのある漢語)に焦点を絞って考察を行った。具体的には、**Overlap** 語に属する 2 字漢語を含む正用文と誤用文の正誤を学習者に判断してもらい形式で調査を行った。調査対象語は 60 語となり、協力者は 12 名(修士 6 名と学士 6 名)であった。その結果、正用文と誤用文を判断する場

合、中国語と意味使用一般性が一致しない語(例:始末, 下手)よりは、意味使用一般性が一致する語(例:質問, 資格)のほうが誤りが起きやすく、CLJ は母語とする中国語の漢語知識に依存しすぎる傾向があることが報告された。

許(2011)は、CLJ が日本語を勉強する際に、母語干渉の影響を受けているかを明らかにするため、日本語と中国語の同形語の意味・用法の差異をめぐって使用実態を調査した。調査方法は、文脈の中で日本語の漢字熟語と意味に対応する中国語の熟語を選択させるアンケート調査であった。設問は主に金(1987)の『同じ漢字でもこれだけ違う日本語と中国語』と『日中辞典』から抜粋した 20 問である(なお、許(2011)は設問 10 まで検討した)。具体的な設問は以下の表 1 のようになる。有効回答はJNS837 人と CLJ174 人であった。

表 1 許(2011)におけるアンケート調査の設問

設問 1	公共の(A 場合・B 場面)では秩序を守るべきだ。
設問 2	公開の(A 場・B 場合)では話に気をつけないといけない。
設問 3	文章の内容を十分(A 把握・B 理解)する。
設問 4	実態を(A 把握・B 理解)する。
設問 5	会議は(A 円満・B 円滑)にとり行われた。
設問 6	旅行に行けない(A 事情・B 理由)を話す。
設問 7	僕はずっと前から中国の(A 事情・B 状況)を知るべきだと思った。
設問 8	(A 検討・B 討論)の結果, 二つの意見が出た。
設問 9	(A 検討・B 反省)の色があまり見えない。
設問 10	この論文の中味は一体どうか(A 評判・B 検討)して欲しい。

アンケート調査を行った結果、10 問のうち、JNSと CLJ の回答が一致したと言えるのは設問 3、設問 4、設問 8、設問 9 の 4 問であり、日本語の漢字熟語の選択率と中国語の訳文の熟語の選択率が逆になった設問は設問 1、設問 2、設問 5、設問 7、設問 10 であった。10 問のうち 5 問が逆となり、CLJ は日本語の漢字熟語を母語のルールそのまま理解しようとする傾向が見られたと述べた。

次に、学習者コーパスを利用して CLJ の漢語使用実態を調査した研究について、胡(2013)は、「日本語学習者による日本語作文と、その母語訳との対訳データベース(作文対訳 DB)オンライン版」を利用して日本語学習者の漢語の使用状況について調査を行った。調査対象は、上記のデータベースにおいて「あなたの国の行事について」というテーマで、JNS と CLJ・韓国人・シンガポール人・アメリカ人・インド人学習者が書いた作文各 10 編を収集したものであった。調査方法については、「日本語読解支援システム リーディング チュウ太」を使い、調査対象を「語彙チェッカー」で判定し、判定された漢語の語数と比率を計算する。その結果、(1)CLJ の漢語使用率は 20.8% であり、6 カ国の中で一番高いこと、(2)日中の同形同義漢語が多いため、CLJ は日本語がわからないときに、中国語の漢字や漢語を試しに使うこと(例:「春節を祝うために、人々が商店やスーパ

一マーケットに行って、いろいろな野菜や、果物や、肉や、それからピーナッツや瓜子(「ひまわりの種」のこと)などを買う、とても忙しい。」), (3) CLJ は使用頻度の低い漢字や常用漢字表外字を多用する(例: 辺鄙, 諺)こと, などが明らかになった。

過去の研究はそれぞれ重要な知見を示しているものの, 大規模な学習者コーパスを利用した研究, 話し言葉を分析した研究, 学習者の習熟段階を区別した研究, 他の母語話者と比較した研究, などについてはほとんど行われていない。本研究はこれらの制約の解決を目指す。

3. リサーチデザインと手法

3.1 研究目的と RQ

本研究では, 大規模な日本語学習者コーパスを利用し, CLJ の話し言葉における漢語使用の状況と問題点の解明を目指す。なお, 「漢語使用」という概念は極めて広範であるため, 本研究では語彙使用の基礎的指標として, 使用された語彙の総語数・総種類数・語彙難度, また, 高頻度で使用された上位語の内容の 4 点に着目する。以上をふまえ, 本研究は以下のような 4 つの RQ を設定した。

RQ1 総語数という点で, CLJ の漢語使用はどのような特徴を持っているのか?

RQ2 総種類数という点で, CLJ の漢語使用はどのような特徴を持っているのか?

RQ3 語彙難度という点で, CLJ の漢語使用はどのような特徴を持っているのか?

RQ4 上位語の内容という点で, CLJ の漢語使用はどのような特徴を持っているのか?

これらの RQ を論じる際に, 具体的な分析観点を設定する必要がある。本研究は, これまでの先行研究の分析の枠組みを精緻化し, (1) JNS および他の母語を持つ学習者(韓国語を母語とする学習者(以下, KLJ: L1-Korean learners of Japanese と称する)と英語を母語とする学習者(以下, ELJ: L1-English learners of Japanese と称する))との比較, (2) 習熟度別比較, という 2 つの分析観点を導入する。

なお, 以下はこれらの RQ を実例データのもとに検証していくわけであるが, あとで行う検証をより体系化するため, 事前にそれぞれの課題に対する仮説を考えておきたい。

まず, CLJ と JNS および KLJ・ELJ の比較について, 母語においてすべて漢字で表記する CLJ は日本語においても JNS より多くの漢語を使用するのではないかと考えられる。また, かつて漢字使用圏に属していた KLJ が CLJ に次ぎ, 母語において漢語を使用しない ELJ がもっとも低い値になるのではないかと予想される。すなわち, 全体として $CLJ > JNS > KLJ > ELJ$ という順序関係になると推定される。同様の観点から, CLJ が使用する漢語の語彙難度も高いものと推定され, 続いて $JNS > KLJ > ELJ$ というようになると予想される。最後に, 上位語の内容についていえば, 上記の順序関係にふまえると, JNS に隣接した CLJ と KLJ との一致度が相対的に高く, ELJ との類似度がもっとも低くなるのではないかと予想される。

次に, 習熟度の影響に関していえば, どの母語を持つ学習者においても習熟度が上昇につれ

て、次第に JNS の漢語使用パターンに近接していくと考えられる。

3.2 データ

本研究は「多言語母語の日本語学習者横断コーパス」(I-JAS)を使用する。従来の学習者コーパスと比べ、I-JAS の構築上の特徴としては、(1)12 種の母語を持つ幅広い学習者のデータを集めていること、(2)CLJ については 200 人のデータを集めており、従来のコーパスに比べて大量のデータを集めていること、(3)学習者情報が詳細であること、(4)日本語能力の客観テストを実施したこと、(5)タスクのバリエーションが豊富であること、という 5 つが挙げられる。そこで、本研究は、学習者の話し言葉における漢語使用を観察するため I-JAS を利用する。

なお、I-JAS において話し言葉を観察できるタスクとしては、絵描写タスク、ロールプレイタスク、対話タスクの 3 種が存在する。そのうち、絵描写タスクはモノログであり、言語の産出場面としてはかなり人工的なものとなっている。また、ロールプレイタスクはダイアログではあるが、会話の目的が狭く定義されており、自然な対話データとは言いにくい。これに対し、対話タスクはダイアログであり、また、幅広い内容を打ち解けた雰囲気ですることから、学習者の持つ言語能力がもっとも自然な形で表れていると考えられる。以上の理由により、本研究は対話タスクを分析対象とする。

対話タスクというのは、学習者と調査実施者が自然な会話を 30 分程度行うものである。データ間での比較を行えるよう、このタスクではある程度統一された話題(全 15 項目)が設定されており、自然な会話の流れで学習者の言語運用が引き出されている(迫田他, 2016)。具体的な流れは以下の表 2 ようになる。

表 2 I-JAS における対話タスクの流れ

ステップ	内訳
(1)ウォーミングアップ	
(2)現在のことを聞く	日本語学習の動機, 好きな本・ドラマ, 出身地の産物, 観光スポットなど
(3)過去の体験を聞く	誕生日の祝い方, 幼少期の体験, 恩師の話, 怖かった・辛かった体験など
(4)未来のことを聞く, 意見陳述	将来の夢, 議論「都会に住むか田舎に住むか」, 議論「お金と時間とどちらが大事か」など
(5)クールダウン	

基本的には、表 2 のように前半は学習者本人に関する過去、現在、未来について話せるものをバランスよく配置し、後半には意見陳述や反論ができるような話題を設定した。

なお、I-JAS を使用して異なる母語を持つ学習者グループを比較するとき、学習者グループ間で習熟度が統制されていないことに注意する必要がある。以下、それらの学習者グループにおける J-CAT の成績分布を見てみよう。

表 3 I-JAS の母語別学習者の各習熟度にあたる人数

習熟度	CLJ	KLJ	ELJ	合計
初級(0 点～99 点)	0	1	7	8
中級前半(100 点～149 点)	2	5	13	20
中級(150 点～199 点)	22	6	16	44
中級後半(200 点～249 点)	41	13	12	66
上級前半(250 点～299 点)	31	14	2	47
上級(300 点～349 点)	3	10	0	13
日本語母語話者相当(350 点～)	1	1	0	2

表 3 を見るとわかるように、CLJ については初級のデータが不足しており、また、ELJ については上級および日本語母語話者相当の 2 つのレベルのデータが不足している。そこで本研究は、3 つの学習者グループのデータがすべて揃っている中級前半～上級前半のデータに絞って考察を行っていきたい。

3.3 手法

個々の RQ の分析手法を説明する前に、本研究における漢語の語数の数え方について述べておきたい。日本語の語数を考えるときに様々な立場があるが、本研究は BCCWJ の品詞タグづけの単位である短単位を採用する。短単位というのは、言語の形態的側面に着目して規定した言語単位であり、現代日本語において意味を持つ最小の単位のことである。たとえば、短単位の基準によると、「図書館」は「図書」と「館」という 2 語に分割される。一般的な漢語の概念と一致していない場合もあるが、ここでは、処理の一貫性を優先し、筆者による調整は行わず、BCCWJ に付された短単位情報を漢語の認定基準とする。

次に、本研究における「漢語使用」を考察する 4 観点の調査方法を以下でまとめておきたい。RQ1(総語数)については、各学習者が対話タスクにおいて使用した漢語の延べ語数を数え、(延べ語数/総語数)*1000 という調整式により 1000 語あたり頻度に統一する。RQ2(総種類数)については、各学習者が対話タスクにおいて使用した漢語の異なり語数を数える。RQ3(語彙難度)については、旧日本語能力試験で示されていた語彙基準(級外, 1 級, 2 級・3 級, 4 級, 5 級という 5 つの段階)に基づく。具体的には、旧日本語能力試験で示されていた語彙レベルを自動判定する日本語読解学習支援システム「リーディング・チュウ太」を用いる。なお、値が大きいほうが高い難度を示すように、難度は(6－級数)という式で計算する。

なお、上記の 2 観点について、まず、(1)CLJ と KLJ・ELJ・JNS の間には差があるかを確認するため検定を行う。通例、異なる文の平均値を比較するには、 t 検定を使用することが多いが、 t 検定は分布の正規性を前提としている。今回のデータでは、正規性を必ずしも前提視することはできなかったため、ノンパラメット検定のクラスカル・ウォリス検定で検証する。次に、(2)異なる習熟度段階の CLJ の漢語使用を比較する際に、上記の 4 つの習熟度段階ごとに比較するが、サンプル数

が少ない場合が多いため、統計的検定を行わない。

RQ4(上位語の内容)については、CLJ の上位語の特徴を見るためコレスポンデンス分析を行う。コレスポンデンス分析とは、頻度表における行・列の関係を組み替え、頻度表に含まれる情報を少数の次元(dimension)にまとめることで、行・列を整理する解析法である(石川他, 2010)。コレスポンデンス分析によって、データ表の行や列の関係を散布図上に布置することができ、視覚的なデータの俯瞰が可能になる。本研究は、JNS・CLJ 中級前半・CLJ 中級・CLJ 中級後半・CLJ 上級前半・CLJ 全体・KLJ 全体・ELJ 全体という 8 種の話者グループを第 1 アイテムとする。また、上記の 8 種の話者グループに対してそれぞれの上位 30 語を取って重複の単語を削除し、残る 60 語を第 2 アイテムとする。

4. 結果と考察

4.1 RQ1 総語数

まず、各言語を母語とする日本語学習者および母語話者が使用した漢語の総語数について見ていきたい。これらについて調査を行った結果、以下の図 1 のようになる。

図 1 CLJ・KLJ・ELJ および JNS の漢語総語数

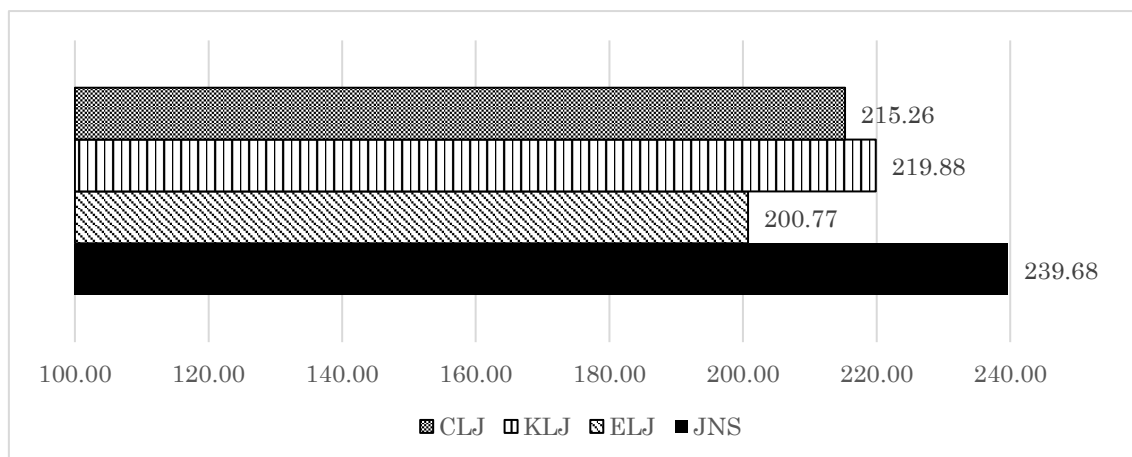


図 1 からわかるように、CLJ と JNS・KLJ・ELJ が使用した漢語の総語数はそれぞれ 215.26 語・239.68 語・219.88 語・200.77 語となった。CLJ の漢語使用量は JNS と KLJ より少なく、JNS の約 88.48%であり、KLJ の約 97, 90%であった。母語という要因は漢語使用に影響を及ぼすかについて、クラスカル・ウォリス検定を行った結果、話者間の間には 0.1%水準で差があることが確認された($K=17.32, df=3, p<.001$)。また、ボンフェローニ下位検定を行った結果、CLJ と JNS との間には 5%水準で、ELJ と JNS との間には 0.1%水準で差があることが確認されたが、他のグループ間には差が見られなかった。仮説では、漢語使用の総語数について $CLJ > JNS > KLJ > ELJ$ という順序関係になると推定されたが、今回のデータは $JNS > KLJ \approx CLJ \approx ELJ$ という順序関係となり、仮説が再現されなかった結果となった。

すでに述べたように、胡(2013)では CLJ の作文において JNS の作文より漢語率が高いと述べたが、今回の結果はこれを支持しなかった。その原因としては、今回使用したデータが話し言葉であることが考えられる。たとえば、以下の学習者と JNS の実例を見てみよう。

表 4 CLJ と JNS の実例(1)

発話番号	実例
CCM17-I-01350～ CCM17-I-01380 (CLJ)	調査者: あーでもラーメンみたいな感じ? <u><うん></u> ラーメンとどう違うんですか? 学習者: あー、んー、多分あー、すんー、あー、ラーメンでは <u><うん></u> 水があります <u><うんうん></u> 、と、とあーこ、こ、あーこれは肉、んー水がありません
CCM17-I-00930～ CCM17-I-00980 (CLJ)	調査者: へーでも『大奥』を、見ても、日本語の勉強にならないでしょう 学習者: あー、はい、んー、あー、んー 調査者: でもやっぱ面白いですか? 学習者: <u>んー面白い</u> 調査者: はい他(ほか)にはどんな、印象に残った、映画とか、ドラマとかありますか? 学習者: <u>んー、あー、お、覚えません</u>

表 4 のように、学習者は整った句の単位で発話ができず、あいづちなどを何回も反復して使用する例が多く見られる。学習者の日本語能力が不足しており、相手の話に乗って次の会話に行くことはできず、あいづちなどを何回も使用して会話のポーズを埋めようとするため、全体として和語の使用率が多くなるのではないかと考えられる。

次に、習熟度の観点について見てみよう。

図 2 CLJ の習熟度段階別の漢語使用総語数

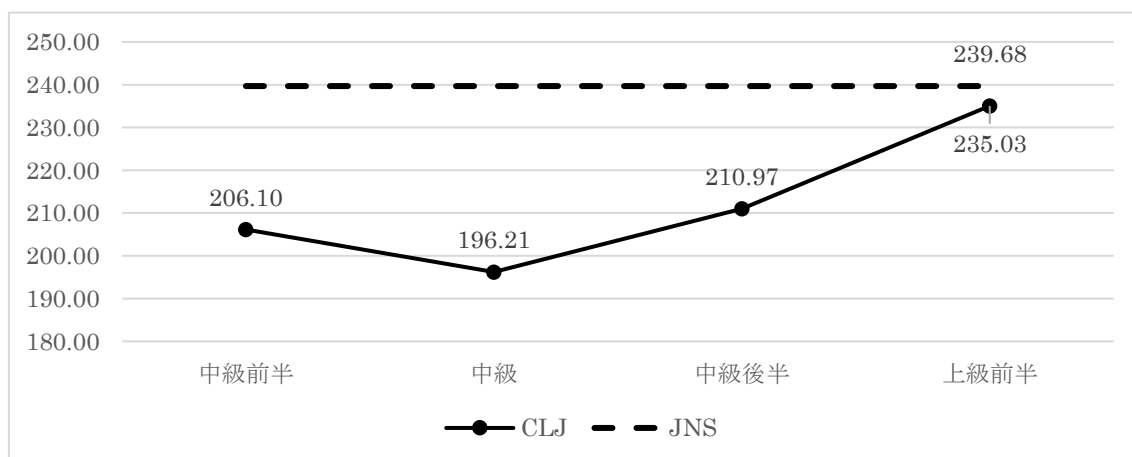


図 2 を見るとわかるように、CLJ の漢語使用の総語数は中級前半から中級にかけて低下しているが、中級から上級前半になるにつれて増加している傾向が見られた。

では、なぜ中級前半から中級になるにつれて漢語使用の総語数は大幅に減少しているのでしょうか。実際の学習者データを確認した結果、この点については、学習者の全般的習熟度の上昇と関わる可能性が示された。すなわち、学習者の習熟度の上がるにつれて流暢性が高くなり、単語の重複が少なくなるため、結果として使用総数が減少しているように見えるのではないかと考えられる。

以下、「誕生日の祝い方」というテーマについて学習者の実例である。

表 5 CLJ の実例 (2)

習熟度	発話番号	実例
中級前半	CCH49-I-02590 ~ CCH49-I-02620 (CLJ)	調査者: {笑} そうですね、友達とかーあの、に、こう例えば何(なん)かプレゼントーをーもらってすごく嬉しかったこととかー、なんか、思い出に残っていることとかありますか？ 学習者: はいはい、うーん、あー、あーんあー、い、あー、吉林よいにはー、あ、た、あ、大学の時、〈うん〉あー、あんー、文けい部(文芸部)、〈うん〉私の一、友達ーあー、いる一の一でー私の一誕生日の時、あー、ちけんー(試験)、あー、試験をしにー、あーわた、私は全部私の誕生日を忘れー、ちゃった 調査者: うん 学習者: ああ、ああ、彼はーおー覚えてー、ああー〈うん〉私ーのああ、寮、寮からーああ、ああ、呼びー出す〈うんうんうんうん〉、ああ、ああーああー嘘をついてー〈うん〉、ああー、ああー、ああー会議をーは、んー、会議をー、ああー、するー
中級前半	CCH26-I-02170 ~ CCH26-I-02260 (CLJ)	調査者: ケーキを食べたり、しません？ 学習者: こんな習慣がありません 調査者: そうなんですかー 学習者: だから〈うん〉、あー今まで、あー誕生日にはケーキをもらう、おーもらったの(過剰使用)、あーことは、あー 調査者: ありませんか 学習者: ありません 調査者: そうですね、この(連体詞) 大学に来てからはどうですかー？ 学習者: うーん、来てから、あー誕生日ーの、〈うん〉日に

		は、あー私の、おー、おー、んー、くら、んールームメイトと 〈うーん〉一緒にあー、<u>食事</u>に、<u>食事</u>をしま、します 調査者:あそうですかー〈はい〉でもそれは友達と一緒に 食事をするのは、楽しいですねー 学習者:はい
--	--	--

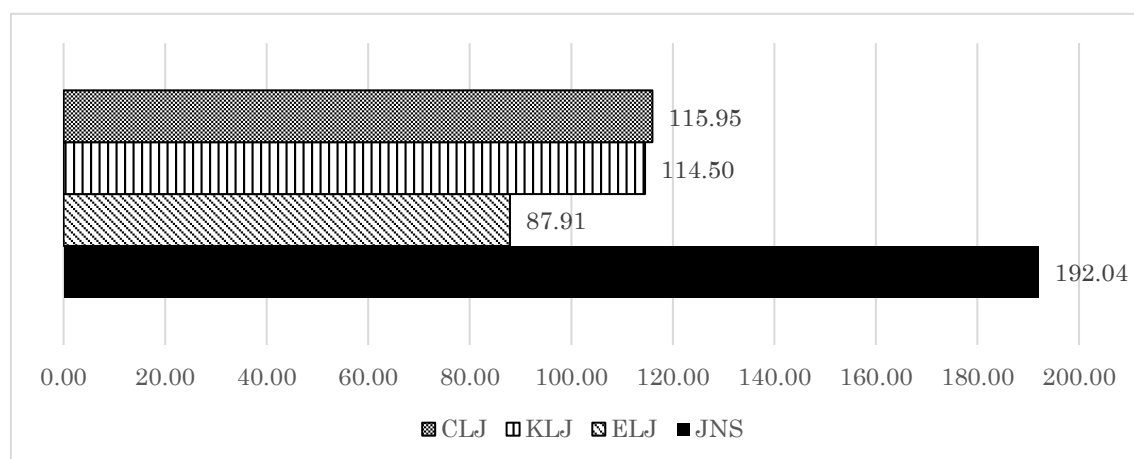
表 5 を見るとわかるように、中級前半の学習者の発話はほとんど句にならず、主要の語彙(例えば、「試験」や「会議」など)を反復して使うことで自分の意思を相手に伝達する。それに対して、中級から上級前半のレベルの学習者は、まだ流暢性が足りないため主語となる名詞(例えば、「食事」など)を重複して使用する現象が見られるが、全体として自分の意思がある程度整えている句で表現できるようになり、結果として漢語の総語数は減少したのではないかと考えられる。

以上、RQ1 では、「漢語使用」の総語数について考察を行った。その結果、(1)全ての学習者グループは JNS より少なく、全体として $JNS > CLJ \div KLJ \div ELJ$ という順序関係となること、(2)習熟度の上昇につれて、CLJ の漢語使用総語数は中級前半から中級にかけて低下しているが、中級から上級前半になるにつれて増加していること、が確認された。

4.2 RQ2 総種類数

次に、漢語の総種類数について見ていく。

図 3 CLJ・KLJ・ELJ および JNS の漢語総種類数

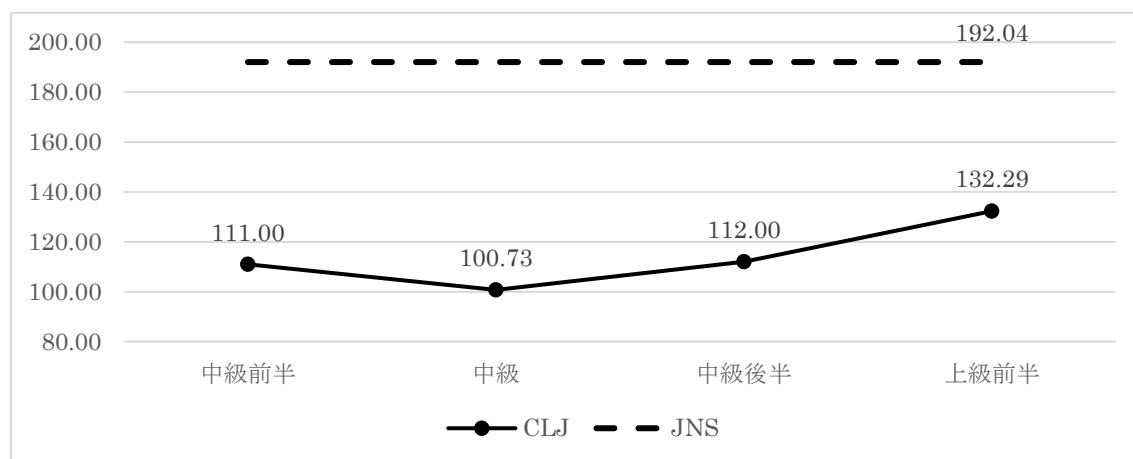


まず、母語別比較について、CLJ・KLJ・ELJ が使用した漢語の総種類数はそれぞれ 115.95 語・114.50 語・87.91 語となり、JNS より 39.62%・40.38%・54.22%少ないことが確認された。クラスカル・ウォリス検定を行った結果、話者間の間には 0.1%水準で差があることが確認された ($K=106.61, df=3, p<.001$)。また、ボンフェローニ下位検定を行った結果、CLJとKLJ以外に、すべてのグループ間において 0.1%水準で差があることが確認された。すなわち、仮説とは異なり、

全体として $JNS > CLJ \approx KLJ > ELJ$ という順序関係となり、総語数と同じように使用する漢語種類が少ないという問題は、CLJ 特有のものではなく、他の言語を母語とする学習者共通の特徴と言える。

次に、CLJ の習熟度段階の比較について見ていきたい。

図 4 CLJ の習熟度段階別の漢語使用総種類数



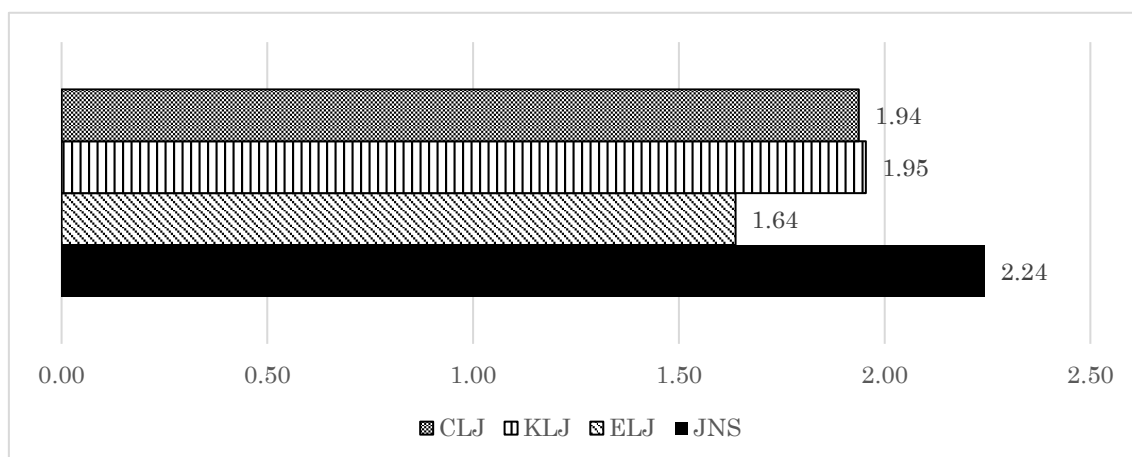
習熟度段階の比較について、CLJ の総種類数は中級前半から中級にかけて少し低下したが、以後は上級になるにつれて増加していることがわかった。すなわち、全体として CLJ は習熟度の上昇につれ、漢語総種類数が増加し、JNS に接近している傾向を示した。なお、上級前半になっても JNS との距離はまだ大きいこともわかった。

以上、RQ2 では、「漢語使用」の総種類数について考察を行った。(1) 総語数と同じように、すべての学習者グループは JNS より少なく、全体として $JNS > CLJ \approx KLJ > ELJ$ という順序関係となり、使用する漢語種類が少ないという問題は、CLJ 特有のものではなく、他の言語を母語とする学習者共通の特徴と言えること、(2) 全体としては習熟度の上昇につれて、CLJ が使用した漢語の種類数は増加し、JNS に接近していること、が確認された。

4.3 RQ3 語彙難度

以上、RQ1 と RQ2 においては、各言語を母語とする日本語学習者および JNS が使用する漢語の総語数と総種類数について見てきた。以下、それらが使用した漢語の語彙難度について見ていく。

図 5 CLJ・KLJ・ELJ および JNS の漢語語彙難度



まず、母語比較について、CLJ・KLJ・ELJが使用した漢語の語彙難度はそれぞれ1.94・1.95・1.64となり、CLJの語彙難度はJNSとKLJより少なかったことがわかった。クラスカル・ウォリス検定を行った結果、話者間の間には0.1%水準で差があることが確認され($K=133.13, df=3, p<.001$), また、ボンフェローニ下位検定を行った結果、CLJとKLJ以外に、すべてのグループ間において0.1%水準で差があることが確認された。この点については、総種類数と同じように仮説を棄却し、全体として $JNS > KLJ \approx CLJ > ELJ$ という順序関係となり、3つの学習者グループともにJNSより簡易な漢語を使っていることが確認された。以下は、CLJとJNSの実例を見ていく。

表6 CLJとJNSの実例(3)

発話番号	実例
CCM50-I-00810～ CCM50-I-00870 (CLJ)	調査者:あー、その(連体詞)小説は日本語で読むんですか？ 学習者:うんはい 調査者:すごいですね 学習者:まあ辞書(5級) 調査者:難しくないですか 学習者:いや辞書(5級)、辞書(5級)えっと辞書(5級)がないと困るんです
JJJ05-I-00600-K (JNS)	本、古本で〈はい〉一万円(5級)ぐらいなんですそれを購入(2級)して〈はい〉ようやくそれ一昨日届いたので〈はい〉開封(1級)してみたんですが、〈はい〉古本、屋さんのその情報(3級)とゆうのが、ちょっとあの、掲載(2級)されている、ネットに掲載(2級)されてるものから買ったんですが、〈はい〉ただ情報(3級)があまりよくなくて〈はい〉中にはその、もうぼろぼろでその〈はい〉日焼け〈はい〉、カバーが日焼けしているってのあったんですが、一冊(5級)、そうですね、手の平サイズの、まあ超えるようなまあ大きさの本(5級)が三十一冊(5級)入っていたんですが、〈はい〉まあカ

	ビはない、まてまあ、なない、無かったんですが、ただ中を開けるとその虫の死骸(1 級)がたくさん詰まっていたとゆうのがありまして
--	---

表 6 を見ると、同じく「好きな本・ドラマ」という話題について、JNS は語彙レベルの高い語を使ったが、それに対して学習者は使った語彙が非常に限定されており、そして比較的簡単なものしか使わないことがわかった。

次に、CLJ が習熟度段階別に使用した漢語の語彙難度変化について見てみよう。

図 6 CLJ の習熟度段階別の漢語語彙難度

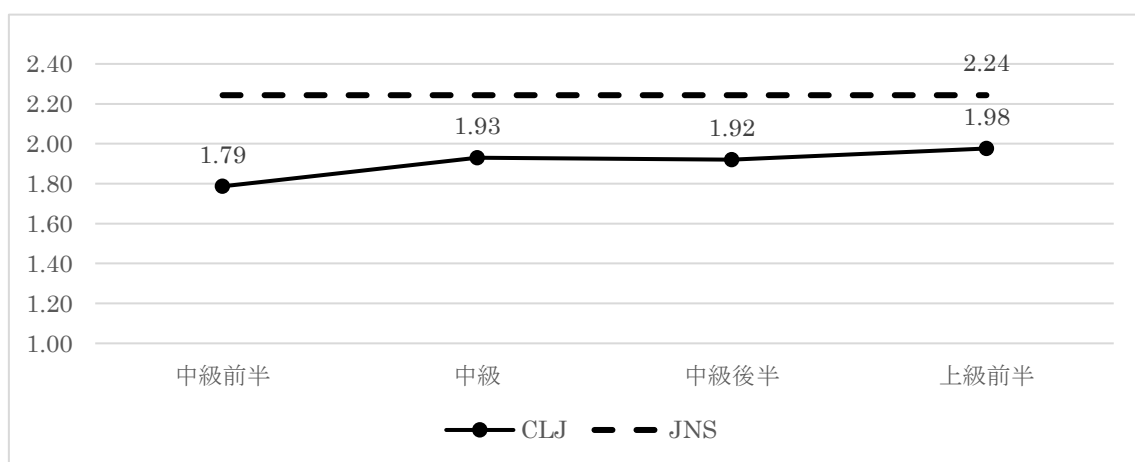


図 6 からわかるように、中級前半から上級まで CLJ が使用する漢語の難度は徐々に増加している。以下、学習者の実例を見ていきたい。

表 7 CLJ の実例(4)

習熟度	発話番号	実例
中級前半	CCM53-I-01180 ~ CCM53-I-01200 (CLJ)	学習者: <u>青島の海は旅行(5 級)、旅行、旅行、旅行する、旅行するの所です</u> 調査者: あーみんなが、旅行するとこ(所)なのね、あーそう、そうですかじゃああの青島のその、なんていうかな、えっと一町は一、あの上海と比べて一、あのどこが違いますか 学習者: あーはいはい、えー、てんきよ(天気)(5 級)、てんきよ(天気)が、<u>天気が違います</u>
中級後半	CCH38-I-01950 ~ CCH38-I-02000 (CLJ)	調査者: わかりました、その(連体詞)岳陽市ってゆうのは、どんな町ですか? 学習者: えっと、<u>〈うん〉それは、そんなに一繁栄(1 級)で</u>

		はないに、ではなくて、〈はい〉でも一、そのの、あ一、人たちは一〈うん〉とても親切(4 級)で、〈へー〉えーと、あ一、あ一、食べ物に一、情熱(1 級)がある{笑} 調査者: えどうゆう意味、食べ物に情熱があるってー？ 学習者: えっと、あ一、あ一、魚とか一〈はい〉さまざまな魚とか一〈うん〉えーと、何(なん) (5 級)ですか、あーいろいろな食べ物に一〈はい〉興味(4 級)が一、あって一、〈はい〉えー、う、ち、違一一、作り方、で、〈うんうん〉んー料理(5 級)を一、食べる{笑} 調査者: あーそうなんですか一、〈はい〉すごいですね一、じゃあいろんな料理を、あの、作ってみる 学習者: はい
上級	CCM50-I-01810 ~ CCM50-I-01900 (CLJ)	調査者: ん一、えっと、CMさんが上海で一番好きな場所ってどこですか？ 学習者: えっと一上海博物館(3 級) 調査者: うん、上海はつく、ぶつかん(博物館)〈はい〉はどんなものが置いてあるんですか？ 学習者: えっと一、陶磁器(級外)、えっと一しよつと(書道)(3 級)のものはえと一 調査者: 書道ですか？ 学習者: かなえ、あはい 調査者: はい 学習者: んーびんやかんや{笑}えっと一、あ一、えと、ん一貨幣(1 級)、かへん(貨幣)、昔のかへ(貨幣) 調査者: あーお金 学習者: うんはい

表 7 からわかるように、同じ「出身地の産物、観光スポット」というテーマについて、中級前半の学習者は簡単な語彙(旅行や天気など)を反復して使うのに対し、習熟度の高い中級後半・上級学習者は比較的複雑な語彙(繁栄や陶磁器など)を使用する。

以上 RQ3 では、各言語を母語とする日本語学習者および JNS が使用した漢語の語彙難度について見てきた。その結果、(1)全体として JNS > KLJ ≧ CLJ > ELJ という順序関係となり、3 つの学習者グループともに JNS より簡易な漢語を使っていること、(2)習熟度の上昇につれて、CLJ が使用した漢語の難度は徐々に増加していること、が確認された

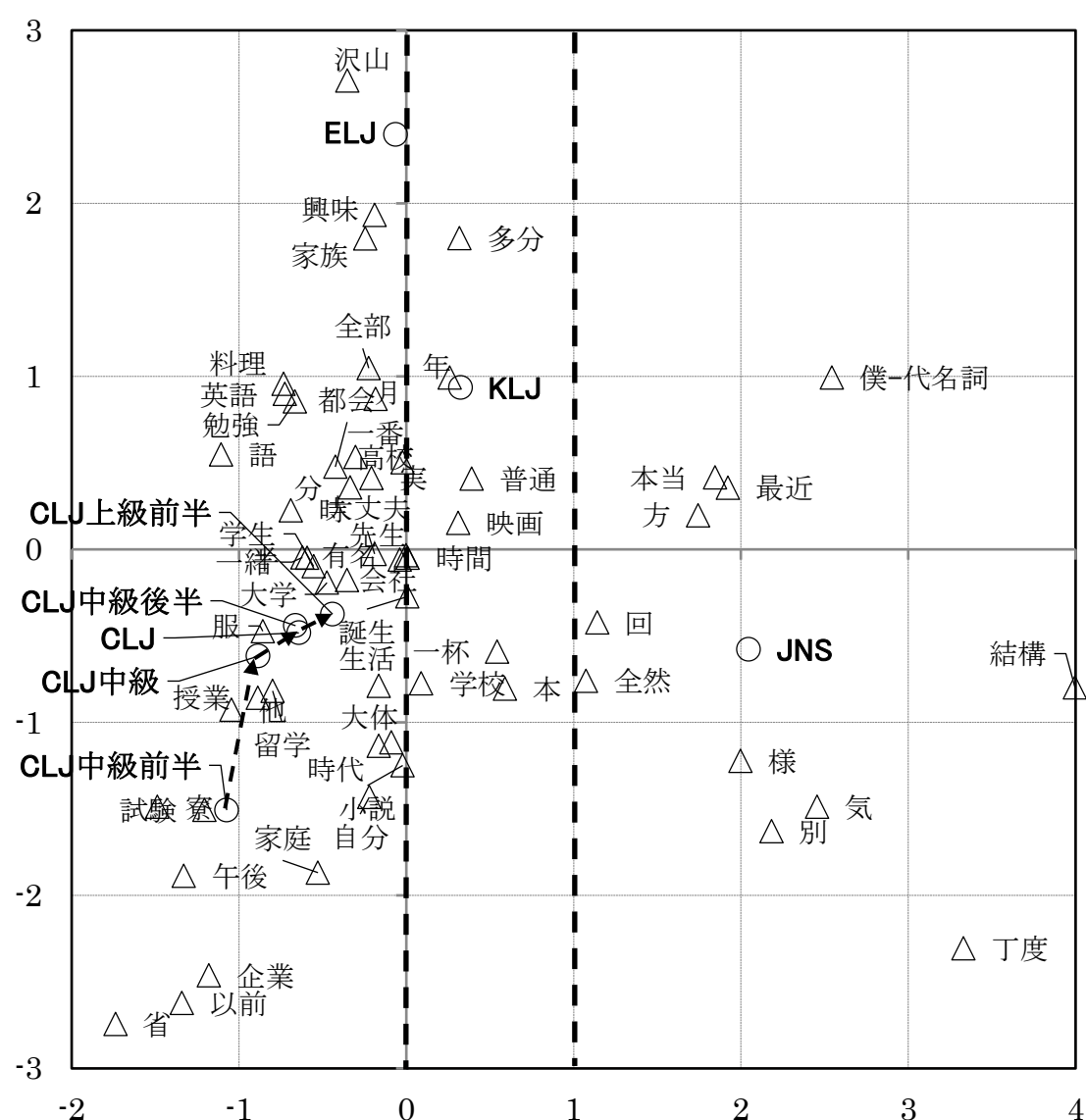
以下 RQ4 では、CLJ 中級前半・CLJ 中級・CLJ 中級後半・CLJ 上級前半・CLJ 全体・KLJ 全体・ELJ 全体・JNS という 8 種の話者グループが使用した漢語の上位語の内容はどのような特徴

を持っているかを確認していく。

4.4 RQ4 上位語

以下, CLJ 中級前半・CLJ 中級・CLJ 中級後半・CLJ 上級前半・CLJ 全体・KLJ 全体・ELJ 全体・JNS という 8 種の話者グループを第 1 アイテムとし, また, これらが使用した漢語上位語 60 語を第 2 アイテムとしてコレスポネンシ分析を行った結果, 以下のような図 7 が得られた。

図 7 コレスポネンシ分析の結果



次元1(横軸)と次元2(縦軸)の寄与率はそれぞれ69.01%と15.43%となり、全体の約84.45%を説明することがわかった。また、2つの軸について言えば、次元1の寄与率が高く、横軸上の位置がもっとも重要な分類観点になっていると言える。この点をふまえてデータを観察すると、もっとも右側にJNSが位置され、原点付近にKLJとELJが位置され、また、左側にCLJがまとまって布置されている。JNSとの距離という点についていえば、 $KLJ < ELJ < CLJ$ という関係になり、当初の予測とは異なる結果となった。

横軸に従って分類すると、1以上はJNSゾーンに、0~1はKLJゾーンに、また、0以下はELJとCLJゾーンに分類できる。

まず、JNSゾーンに注目すると、JNSに特徴づけられるのは、一人称代名詞「僕」のほか、「結構」や「丁度」「本当」「全然」など副詞的に使用して、物事の程度や確からしさを表す語であることがわかった。また、「様」や「気」「別」について用例分析を行ったところ、「気」については「気に入った」や「気がする」など、「別」については「別に」「別の～」など、「様」については「～のように」などの用例が大半を占めていた。一般的な漢語の概念とは若干ずれるものではあるが、こうした語が母語話者らしい漢字使用を決定していることが明らかになった。

次に、KLJゾーンに注目すると、KLJゾーンは「映画」「学校」「本」など身近な生活に関わる一般名詞が多く含まれているが、「多分」「普通」「一杯」などのような副詞的に使用される語もいくつ含まれている。結果的にみれば、このことは、3種類の学習者グループでKLJがもっともJNSに近い位置に布置されていることの理由となっていると考えられる。

最後に、ELJゾーンとCLJゾーンについていえば、「興味」「家族」「料理」「高校」「大学」「留学」など同じく身近な生活に関わる一般名詞が多い。「沢山」のように副詞的に使われる語も存在するが、その数はKLJより少ない。

また、CLJの習熟度段階の変化について、第2軸に注目すると、下から上に向けてCLJのレベルの上昇に合致していることはわかった。中級前半では「試験」「寮」など、中級では「授業」「留学」など、中級後半では「服」など、上級では「大学」「会社」などの語に特徴づけられ、日常語の中で、より学習に直結したものから、個人生活と関わる語となり、少し内容的な変化はあるものの、その変化は漢語を副詞的に使用するといった母語話者的な漢語使用に近接していることとは言えない。

5. まとめ

本研究は、大規模な日本語学習者コーパス I-JAS を利用し、CLJ の話し言葉における漢語使用の状況を解明した。以下、本研究で明らかになった知見を整理して示しておく。

まず、漢語使用の総語数について、(1)全ての学習者グループはJNSより少なく、全体として $JNS > KLJ \approx CLJ \approx ELJ$ という順序関係となること、(2)習熟度の上昇につれて、CLJの漢語使用総語数は中級前半から中級にかけて低下しているが、中級から上級前半になるにつれて増加していることが確認された。

次に、漢語使用の種類数について、(1)総語数と同じように、すべての学習者グループはJNS

より少なく、全体として $JNS > CLJ \approx KLJ > ELJ$ という順序関係となり、使用する漢語種類が少ないという問題は、CLJ 特有のものではなく、他の言語を母語とする学習者共通の特徴と言えること、(2) 全体としては習熟度の上昇につれて、CLJ が使用した漢語の種類数は増加し、JNS に接近していること、が確認された。

続いて、語彙難度について、(1) 全体として $JNS > KLJ \approx CLJ > ELJ$ という順序関係となり、3 つの学習者グループともに JNS より簡易な漢語を使っていること、(2) 習熟度の上昇につれて、CLJ が使用した漢語の難度は徐々に増加していること、が確認された。

最後に、上位語の内容について、(1) JNS との類似度という点についていえば、 $KLJ > ELJ > CLJ$ という関係になり、KLJ の上位語使用は JNS ともっとも似ていること、(2) CLJ の習熟度段階の変化について、内容的な変化はあるものの、その変化は漢語を副詞的に使用するといった母語話者的な漢語使用に近接していることとは言えないこと、が確認された。

特に CLJ を対象とした日本語教育についていえば、CLJ は漢語や漢字の知識を既に持っているため、日本語漢語について特に取り出して指導する必要が薄いと言われてきたが、今回の結果から見れば、CLJ の漢語使用はまだ多くの問題が存在し、また、一定のレベルの学習者であってもその問題が継続している可能性が示唆された。このことは、中上級 CLJ を対象とした日本語の漢語指導プログラムの開発の必要性が示されたのではないかと考えられる。

引用文献

- 荒川清秀 (1979) 「中国語と漢語：文化庁『中国語と対応する漢語』の評を兼ねて」『愛知大学文学論叢』, 62, 361-388.
- 許恵玉 (2011) 「日本語と中国語の同形語の意味・用法の差異について：漢字熟語の使用に関する調査結果の分析」『山口国文』, 34, 108-120.
- 胡環 (2013) 「中国語を母語とする初級日本語学習者における日本語漢字語彙の学習戦略に関する調査」『アカデミック・ジャパニーズ・ジャーナル』, 8, 37-44.
- 迫田久美子・小西円・佐々木藍子・須賀和香子・細井陽子 (2016) 「多言語母語の日本語学習者横断コーパス」『国語研プロジェクトレビュー』, 6(3), 93-110.
- 張婧禕 (2017) 「中国人日本語学習者の漢語同形語習得：同形類義語 (Overlap 語) を中心に」『愛知工業大学研究報告』, 52, 6-13.
- 陳毓敏 (2003) 「中国語を母語とする日本語学習者における漢語習得研究の概観：意味と用法を中心に」『言語文化と日本語教育』, 増刊特集号, 96-113.
- 陳淑娟 (2007) 「作文における漢語語彙の習得についての考察：LARP at SCU のデータに基づく事例研究」『台湾日本語文学報』, 22, 379-403.
- 野村雅昭 (1999) 「語彙調査データによる基本漢語の抽出」『早稲田大学日本語研究教育センター紀要』, 12, 21-54.

