中国語を母語とする学習者の作文における日本語漢語の使用状況

－日本人のブログとの比較－

鄧琪（神戸大学大学院生）

要　旨

本研究は、「日语学习者书面语语料库」（日本語学習者書き言葉コーパス）を用いて、学習段階が進むにつれて、CLJの作文における日本語漢語の使用量と内容がどのように変化するか、また、JNSとのズレがどのように変化していくかを実証的に調査する。その結果、漢語の使用量について、タイプ（異なり語数）は若干の増加が見られたものの、トークン（延べ語数）は上昇が見られず、いずれについてもJNSとの差は大きいまま残されていることがわかった。また、漢語の内容について、テーマに基づいた語彙の内容変化は見られたが、漢語そのものの言語的性質がJNSに接近していないこと、が確認された。

【キーワード】：コーパス　漢語　中国語を母語とする学習者　縦断調査

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、その中で、種類として一番多いのは漢語である（国立国語研究所1962など）。漢語の習得について、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CLJ）は総じて有利であると思われるが、このことはCLJが漢語習得に問題を持たないことを意味するものではない。以下は、中国の日本語専攻大学生の作文を収集した「日语学习者书面语语料库」（日本語学習者書き言葉コーパス）に含まれるCLJの漢語使用の実例である。

（1）先生のおかげで毎日を充実に（⇒充実して）過ごしておりますが、進学のお礼と近況報告とお申し上げます。（作文番号：d0208）

（2）政権も巩固（⇒強固）になるはずなのだ。（作文番号：d0949）

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用することは必ずしも容易ではないことを示す。こうした問題の程度と範囲、また、学習段階による変容の過程を客観的に議論するには、CLJによる産出を時系列的に収集した縦断型の学習者コーパス分析が必要となる。本研究は、大学の第３～５学期にかけてCLJが書いた日本語作文を集めた学習者コーパスと、日本語母語話者によるブログコーパスを比較

し、CLJが使用する漢語の量と内容がどのように変化していくかを分析する。

２．先行研究

横断コーパスを用いた学習者の漢語使用の計量的研究として、胡（2012）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」を利用して学習者の漢語の使用状況を調査した。調査対象は「あなたの国の行事について」というテーマについて、日本語母語話者（Japanese Native Speakers、以下JNS）とCLJ・韓国人（Korean Learners of Japanese：以下KLJ）・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編であった。日本語読解支援システム「リーディングチュウ太」を用いて、作文で使用された日本語の語種別の語数と比率を計算した結果、（1）CLJの漢語使用率は20.8％で6ヵ国の中で一番高いこと、（2）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、（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こと、が明らかになった。

鄧（2020）は「多言語母語の日本語学習者横断コーパス」（I-JAS）を利用し、様々な習熟度段階にあるCLJ、KLJ、英語母語の日本語学習者（L1−English Learners of Japanese、以下ELJ）とJNSの漢語使用を比較した。その結果、（1）漢語使用の総頻度はJNS ＞KLJ ≒ CLJ ≒ ELJとなり、（2）使用する漢語の種別数はJNS ＞CLJ ≒ KLJ ＞ELJとなり、（3）使用する漢語の語彙難度はJNS ＞KLJ ≒ CLJ ＞ELJとなり、（4）上位語におけるJNSとの類似度はKLJ ＞ELJ ＞CLJとなることが確認された。また、CLJは習熟度の上昇に伴い、漢語使用の総頻度や種別数、そして語彙難度の３つの観点においてJNSに接近するが、上位語という観点においてJNSに接近しないことが確認された。

一方、縦断コーパスを用いた漢語研究は少ない。陳（2007）は「Language Acquisition Research Project at Soochow University」（LARP at SCU）を用いて台湾の学習者が使用した漢語の特質分析を行ったが、漢語使用全体の量的概観はなされていない。

このように、先行研究はそれぞれ重要な知見を示したが、各学習段階にあるCLJの漢語使用量や各学習段階を特徴づける漢語の内容、また、学習段階が進むにつれてJNSとのズレがどの程度解消されるかといった点はまだ明らかではない。そこで本研究は、これらの諸点の解明を目指す。

３．リサーチデザイン

３．１．目的とRQ

本研究は、学習段階が進むにつれて、CLJの作文における日本語漢語の使用量と内容がどのように変化するか、また、JNSとのズレがどのように変化していくかを実証的に調査することを目的とする。上記に従い、以下の２つのRQを設定した。

RQ1　学習段階の変化につれて、CLJの作文における日本語漢語の使用量はどのよう

に変化するか？ また、JNSとのズレはどのように変化するか？
RQ2　学習段階の変化につれて、CLJの作文における日本語漢語の内容はどのように変化するか？ また、JNSとのズレはどのように変化するか？

なお、最近の応用言語学の研究では、学習者の産出を議論する際に、JNSをモデルとすることへの批判も認められる（石川2016、Leech 1998など）。実際、外国人が日本語漢語を使用する際に、すべての面においてJNSと一致しなければならないというのは極論である。本研究では、この点を念頭に置きつつ、CLJの変化を見る際の1つの目安としてJNSの産出データを位置付けたい。JNSと比較することで、CLJが学習目標にどこまで近づいているのか、どのような問題を抱えているのかが明らかになる。

３．２．データ

本研究ではCLJ側のデータとして「日语学习者书面语语料库」を、JNS側のデータとして「現代日本語書き言葉均衡コーパス」（Balanced Corpus of Contemporary Written Japanese、以下BCCWJ）を利用する。以下、それぞれについて概観していく。

「日语学习者书面语语料库」は上海交通大学文科科研基金「学習者書き言葉および日本語教学研究」（代表者：張建華、課題番号：08TS1）の一環として、中国の5つの大学における日本語科の学習者401名が書いた作文を収集したコーパスである。作文のテーマは、学習者が興味を持ちそうな各種の社会問題となっている。なお、同じ学期ではあっても、大学間で、学習の進度やレベルが異なる可能性がある。この点を統制するために、本研究では、学習者数が多く、作文テーマが比較的揃っている「東北師範大学縦断データ」30名分のみを使用した。学習期間は3学期から5学期までで、総語数は201,857語となる。なお、コーパスに収集された学習者の習熟度をテストなどの形で直接調査していなかったため、同じ学習段階の学習者の中にも習熟度の差が存在していた可能性が否定できない。ただし、中国の大学の日本語専門学科で使われているシラバス（『高等院校日语专业基础阶段教学大纲』）に基づくと、学習者は2学期が終わる段階で3000語程度、また、4学期が終わる段階で5600語程度の語彙が習得されると期待されている。これは日本語能力試験JLPTにおける語彙量の認定基準（N3：1600語、N2：6000語、N1：10000語）と照らし合わせると、3学期と4学期の学習者はおおよそＮ３程度、5学期はおおよそＮ２程度であると推定される。

BCCWJは、国立国語研究所が中心となって開発した日本語の大規模均衡コーパスである。BCCWJは書籍・新聞・雑誌・ブログなど様々なレジスターのテキストを集め、日本語の書き言葉の多様な実態をとらえられるデータとなっている。本研究では、このうち、ブログデータのみを利用する。ブログは、学習者が日本語で書く作文と完全に一致するものではないが、自分の意見を自由なスタイルで書き、かつ、通例、編集が加えられていない点で、作文に近い言語的性質を持つと考えられる。

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。データの処理にあたっては、UniDic 辞書が搭載されたWeb 茶まめを利用する。

RQ1 については、トークン（延べ語数）とタイプ（異なり語数）の両観点から考察する。学期間を比較する際には、段階ごとの作文の長さ（総語数）が異なるため、トークンとタイプともに粗頻度を作文総語数で割って1000語あたりに変換した調整値を算出した。たとえば、学習者d1 が３学期に書いた作文の総語数は2,784 語であり、使用された漢語の延べ語数は491 語、異なり語数は211 種であった場合、調整トークンは491/2784*1000 = 176.36 語、調整タイプは211/2784*1000 = 75.79 となる。参照基準となるJNSのタイプの算定にあたっては、CLJの作文の分量（学期あたり平均2268.06 語）と揃えるため、BCCWJからランダムに50 本分のブログを抽出して調整タイプを計算する。なお、サンプル数が少なかったため、平均値の差の検定については、クラスカル・ウォリス検定を使用する。RQ2 については、各学習段階において特徴的に使用される漢語を調べるため、学習段階を第１アイテム（CLJの３～５学期とJNS)、上位語を第２アイテム（上記４グループが使用した高頻度漢語上位50 語（合計200 語、テーマ語は除く）から重複を除いた全120 種の語）として対応分析を行う。ただし、こうした分析は通例学期別の高頻度語を調べるという形で行うが、学期というのは学習時間によって便宜的に定められたもので、必ずしも相互独立の関係にあるわけではない。実際に複数の異なる学期において漢語使用が変化していないことは十分に考えられる。この場合、学期ではなく、漢語使用の実態に即した学習段階別に見ていく方がいいと考えられるため、本研究は学習段階を再構成できる対応分析を実施する。

なお、上記の分析を実施する場合、学習者個人の発達状況と、学習者集団の発達状況とは必ずしも一致しない。本研究は、特に学習時期による集団としての振る舞いの違いに焦点を当てるため、個々の学習者から得られた頻度を平均して議論を行った。また、本研究はCLJの漢語使用における基礎的な事実を客観的に抽出するという主目的の下、計量的・統計的なデータ分析を主眼とするが、分析結果の考察の際には、関連する学習者の用例の質的な検討と解釈を行い、総合的な判断を行うことを目指す。

４．結果と考察

４．１．RQ１　漢語の使用量

まず、CLJの漢語使用量の段階的変化を見ていく。CLJの各学期の漢語使用の調整

トークンと調整タイプを調査したところ、以下のような図が得られた。

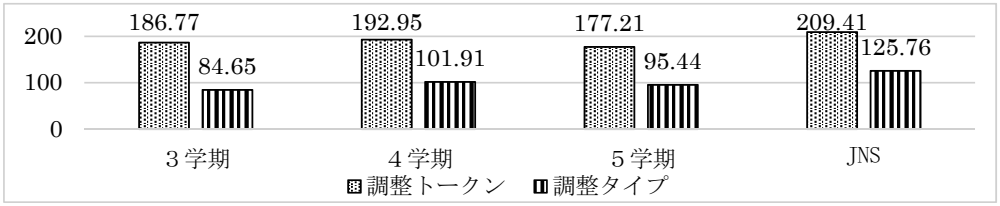

図1　各学習段階にあるCLJおよびJNSの調整トークンと調整タイプ

図1より、3学期と5学期を比較するとタイプは増えているが、トークンは減っており、いずれの場合もJNSには届かないという傾向が観察された。この傾向の有意性を確認するため、クラスカル・ウォリス検定を行った結果、トークン（K = 56.33、df = 3、p<.001）、タイプ（K = 76.08、df = 3、p<.001）ともに群間に差があることが確認された。しかし、ボンフェローニ下位検定を行ったところ、JNSとの差については両観点ともにCLJの3～5学期のすべてにおいて差が確認されたが（p<.001）、学期間の差については、トークンでは4学期と5学期の間のみ（p<.05）、タイプでは3学期と4学期の間に（p<.001）、また、3学期と5学期との間に（p<.01）差が確認された。学習期間の増加により、トークン・タイプともに増加し、JNSに近接すると予測したが、今回のデータについて言うと、タイプは若干の増加が見られたが、トークンは上昇が見られず、いずれもJNSとの差は大きいまま残されているという結果となった。

胡（2012）は、CLJの漢語使用率は20.8%で、調査した6ヵ国の中で一番高く、JNSよりも多用すると述べたが、今回は異なる結果となった。これは2つの調査におけるデータや漢語の認定基準の違いの影響が考えられる。では、なぜ今回の調査において、CLJが使用した漢語はJNSより少なかったのだろうか。実際のデータを観察したところ、胡（2012）が指摘したように、日本語がわからない時にCLJが中国語の漢字や漢語（「非日漢語」）を使用する傾向は今回も認められた（例（3））。しかし、それらを除くと、全体的に漢語の使用度は低く、特に、文体やフォーマリティへの配慮が欠如しているため、本来は改まったニュアンスを持つ漢語を使うべきところで、くだけた和語を使う事例が散見された（例（4、5）はCLJの、例（6）はJNSの用例である）。

（3）でも、<u>失主</u>に対して、私は責任がある事を忘れてしまいました。（作文番号：d0104）

（4）赤信号の場合、<u>車</u>と<u>人</u>は止めらなければならない。もし、とめらなったら、交通事故があるかねない。（作文番号：d0331）

（5）赤信号がつける時、<u>止まな</u>かったら、交通事故が起こりやすい。（作文番号：d0334）

（6）信号無視といっても進行方向に交差する側の<u>歩行者</u>信号が赤になった途端に進行

方向の歩行者信号も自動車用信号も赤なのに横断を始めていった…(OY03_08150)

上記のように、JNSは「自動車」「歩行者」「横断」などの漢語を使うことで、自身の主張にフォーマリティと客観性を持たせている（例（6））。一方、CLJは同様の内容に言及する際に、「車」や「人」（例（4））、そして「止まらない」（例（5））のような和語を多用する。結果として、CLJの作文は、フォーマリティが低く、主観的な印象を与えている。漢語の持つこうした文体やフォーマリティの表出機能に習熟することがCLJにとっての課題の1つであると言えるだろう。

４．２．RQ２　漢語の内容

ここでは、各学期にCLJが特徴的に使用する漢語はどのように変化しているかを見ていく。対応分析を行った結果、図2のような散布図が得られた。

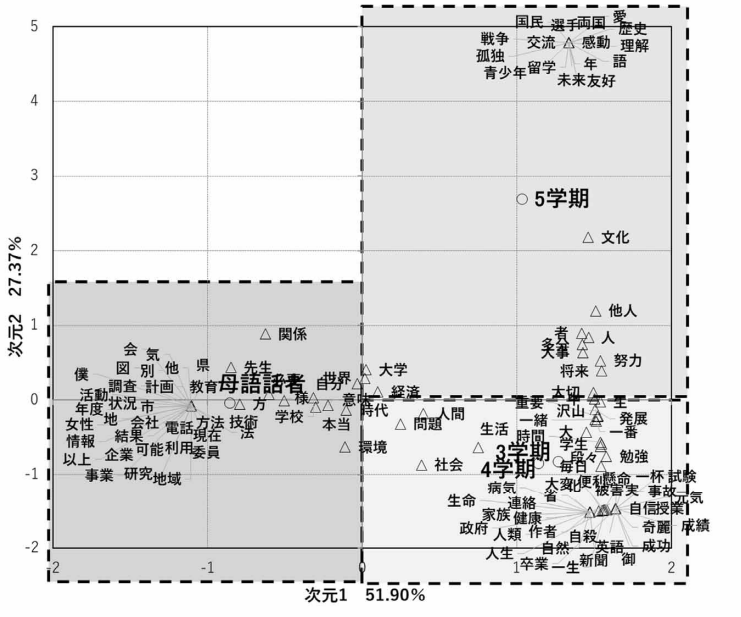

図２　各学習段階にあるCLJおよびJNSの漢語使用の特徴語

ここでは、（1）書き手分類、（2）漢語分類の2種に分けて結果を概観する。まず、（1）に関して、次元1（横軸）を見ると、左側にJNSが、右側にCLJがまとまって布置されていることから、次元1はJNSとCLJの区分軸になっていることがわかった。学習の最終段階である5学期の学生であっても散布図上の配置は右側であってJNSの側に入っていないことは、漢語使用についてJNSとCLJの間に大きな差が存在することを示した。次に、次元2（縦軸）を見ると、学習段階の後期（5学期）の学生が上部に、初期（3・4学期）の学生が下部に布置されていることから、次元2はCLJの習熟度の区分軸になっていることがわかった。また、CLJの漢語使用能力は線形的に上昇

するのではなく、4学期が修了した段階で大きく伸びるものと推定される。

次に（2）に関して、次元1の左側に存在するJNSゾーンに注目すると、単独では使用できず、他の表現の一部として初めて意味機能を果たす語が多いことがわかった。たとえば、(a)「様」「気」「別」のような1文字語（用例：「気に入った」「別に」「～の様に」)、(b)「本当」や「以上」のような副詞的用語、(c)「市」や「県」のように接尾辞として使用する語などである。一般的な漢語の概念とは若干ずれるものではあるが、こうした特殊な漢語の使用が母語話者らしさを形作っていることが明らかになった。一方、次元1の右側に属するCLJゾーンを見ると、3・4学期学生の使用する「時間」「学生」や、5学期学生の使用する「未来」「交流」など、2文字熟語が大半であった。次元2上の差について、3・4学期では学生や生活など、5学期では日中関係や文化交流などに関わる語彙が多く、テーマに基づいた語彙の内容変化はみられるが、漢語そのものの言語的性質がJNSに接近していることは確認できなかった。ただし、性質的には同様であっても、使用する語彙のレベルが異なる可能性はある。そこで、各ゾーンに含まれる漢語の語彙レベルを調査したところ、表1が得られた。

表1　各ゾーンに属する漢語の語彙レベル

語彙レベル	3・4学期	5学期	JNS
旧5級	40%	24%	21%
旧3級・4級	60%	69%	74%
旧1級・2級	0%	7%	5%

表からわかるように、3～4学期と比べると、5学期では初級（旧5級）語彙の比率が16%減少し、代わりに上級語彙（旧1・2級）の比率が7%上昇し、JNSの比率に近づいている。このことから、3～4学期と5学期の間では内容を除けば使用する漢語の言語的な性質は変化しないが、語彙レベルについては一定の上昇が起こっていることが示された。ただし、このことは漢語使用能力に限った上昇というより、学習段階が進んだことによる日本語能力の全般的上昇を反映した結果にすぎない可能性も残る。

5．まとめ

本研究は「日语学习者书面语语料库」を用いてCLJの日本語漢語使用の段階的変化について調査を行った。その結果、漢語の使用量について、タイプは若干の増加が見られたものの、トークンは上昇が見られず、いずれについてもJNSとの差は大きいまま残されていること、漢語の内容について、テーマに基づいた語彙の内容変化は見られたが、漢語そのものの言語的性質がJNSに接近していないこと、が確認された。

従来、CLJを対象とする日本語教育についていえば、CLJは漢語や漢字の知識をすでに持っているため、指導する必要性が低いと言われてきたが、今回の結果から見れ

ば、CLJの日本語漢語使用の使用量がまだ不足し、漢語の使い方についても母語話者とまだ距離がある。また、学習段階が進むにつれて母語話者の漢語使用に接近する傾向が弱く、特に指導がなければその問題が継続していく可能性が示唆された。

ただし、本研究には残された課題も多い。1点目は学習段階という概念を分析対象としたが、使用したコーパスデータの制約もあり、これが実際の習熟度の上昇とどう関係していたかが議論できていないことである。2点目は学習時期ごとに比較を行ったが、個々の学習者の振る舞いの差が観察できていないことである。3点目は使用された漢語の難度を旧日本語能力試験の語彙リストに依拠した点である。本来、語の難度は使用環境や文脈などによって変化し、語だけを取り出して難度を議論することは適切ではない。これらの点については今後の継続研究において対応を図っていきたい。

参考文献

石川慎一郎（2016）「日本語学習者コーパスの教育応用における留意点―『多言語母語の日本語学習者横断コーパス』に見る母語話者L1産出データの安定性検証を中心に―」『言語資源活用ワークショップ発表論文集』1、190-200

胡環（2012）「日本語学習者の作文における漢語の使用について」『第九回国際日本語教育・日本研究シンポジウム予稿集』香港日本語教育研究会

国立国語研究所（1962）『現代雑誌九十種の用語用字第一分冊：総記および語彙表』東京：秀英出版

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

鄧琪（2020）「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―」『統計数理研究所共同研究リポート』435、121-138

Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on computer* (pp. xiv-xx). Harlow, England: Addison Wesley Longman.

高等院校日语专业基础阶段教学大纲

〈https://wenku.baidu.com/view/40a2e310866fb84ae45c8ddb.html ＞（2021年1月19日アクセス）