

中国人日本語学習者の文法能力は 作文の語彙特性にどう影響するか¹

楊 艶萍²

玉岡 賀津雄³

張 婧禕⁴

DOI: 10.18999/stul.33.147

要約: 中国人日本語学習者の文法能力が、作文の語彙特性にどう影響するかを検討した。日本語学習者 26 名の作文の語彙特性を数量化した。同じ学習者に、形態素変化、局所依存、構造の複雑性の3つの下位尺度を持つ文法テストを実施した。文法テストの下位尺度で作文の語彙特性を予測する重回帰分析を行った。構造の複雑性の理解が、動詞、副詞、助動詞、普通名詞、助詞の頻度を有意に予測した。また、局所依存の理解が、形容詞の頻度を有意に予測した。形態素変化の理解は、作文の品詞の頻度を予測しなかった。複雑な文を理解する能力を持つ日本語学習者は、動詞、助動詞および副詞など述部を構成する品詞群と普通名詞と助詞など主語と目的語を構成する品詞群を頻繁に使って作文を書き、語と語の組み合わせである局所依存の理解がよくできる日本語学習者は修飾的機能の形容詞を頻繁に使って作文を書く傾向がみられた。さらに、26名の日本語学習者の作文の動詞、助動詞、副詞、普通名詞、形容詞、助詞の頻度を使ってクラスタ分析をした。その結果、3つのクラスタが得られ、文法能力の向上と共に品詞の使用が「普通名詞、助詞」から「普通名詞、助詞、動詞、助動詞」と拡張していく傾向がみられた。

キーワード: 作文, 文法能力, 品詞, 使用頻度, リーダビリティ, クラスタ分析

¹ English title: How do grammatical abilities affect readabilities and lexical characteristics of essays written by native Chinese speakers learning Japanese?

² YANG, Yanping, Central South University of Forestry and Technology, China, E-mail: 314651338@qq.com

³ TAMAOKA, Katsuo, Nagoya University, Japan, E-mail: ktamaoka@gc4.so-net.ne.jp

⁴ ZHANG, Jingyi, Nagoya University, Japan and Japan Society for the Promotion of Science (JSPS) fellow, E-mail: jingyizhang@nagoya-u.jp

1. 研究の目的

情報化社会の進展の中で、「書く」ことによる表現力の育成が求められている。村田(2018)は、ドイツ人日本語学習者の日本語の習熟度が、作文の文の長さ、誤り数などに影響することを報告している。しかし、習熟度と作文の内容の関係について詳しく示したわけではない。「書く」能力には、語の配列や結合規則の総体としての文法能力が影響すると思われる。そこで、本研究では、中国人日本語学習者の文法能力を形態素変化、局所依存、文の複雑性の3つの尺度で測定して、それが個々の学習者の作文の言語的な特性にどう影響するかを統計的な手法を用いて検討した。

2. 日本語学習者の作文に関する先行研究

日本語学習者の作文については、誤りのパターンを分類する研究が行われてきた。市川(1993)は、中級学習者に短文作りの課題を課して、得られた 288 の短文について、読み手の理解を妨げる視点から、「全体的誤り」と「局部的誤り」について分析を行った。その結果、接続語そのものが文の意味を理解し難くしている最大の要因であり、次に主語・主題の扱い、文末・ムード、語彙・表現も「全体的誤り」に大きく関わっていると報告している。その後、川口(1995)では、異なる国籍の5名の中上級レベルの学習者の授業時の作文2つと、終了テストの作文2つの計20点、それぞれ400字のテキストを分析し、以下の結果を得た。(1)日本人の平均文長を35~36字とすると、その数字に近づくほど誤用は少なくなり、離れていくほど誤用は多くなり、特に長すぎる文では誤用が多くなる傾向があった。(2)267例の誤用のうち、文法項目の中で最も多く見られる誤用は「助詞」である。その助詞の中でも「格助詞」の誤用が多く、また、「は」と「が」の誤用は上級まで引きずる傾向がある。さらに、原沢(2012)は、初中級の学習者15名の300~600字程度の33点の作文で290例の誤用を分析し、文の構造に関する誤用が全体の60%になり、初・中級レベルの学習者は、ある特定の表現に特化した誤用がみられると報告している。

楊帆(2014)は、中級日本語学習者の作文における誤用を5種類に分類している。それによると、文法的な誤り、意味的な誤り、表記上の誤り、論理的な誤り、語用論的な誤りである。合計2,875例の誤用のうち、文法的な誤りは全誤用の54.6%を占め、意味的な誤りは全誤用の33.8%を占めた。これらの2つの誤用が全誤用の9割弱となり、圧倒的に多か

った。誤用の半数以上を占める文法的な誤りについては、さらに詳細に分類すると、6割以上を占める上位5項目として、助詞の誤り、文構造の呼応関係の誤り、句読点の誤り、修飾の誤り、語句の選択の誤りがみられた。

これまでの先行研究(市川, 1993;大塚・林, 2010;鈴木, 2002)は、語彙などに特化した誤用に関して分析したものが多く、学習者の作文における誤用の全体的な傾向については言及していないことが多かった。川口(1995)と原沢(2012)は学習者の作文における誤用の全体的な傾向を把握したものの、それぞれ川口(1995)が20点、原沢(2012)が33点の作文を分析対象としており、データ数は決して多いとはいえない。そのため、これらが日本語学習者の誤用を代表していると言い切れるかどうか疑問が残る。さらに、川口(1995)は、助詞の誤用が最も多く、原沢(2012)は文の構造に関する誤用が最も多いという異なる結果を得ている。したがって、より多くの作文データを詳細に分析し、学習者の誤用傾向を詳しく検討する必要があると考えられる。

3. 研究の方法

本研究は、中南林業科技大学外国語学院日本語学部生を対象に、文法テストおよび作文タスクを実施した。詳細は以下の通りである。

3.1 本研究に参加した日本語学習者

中南林業科技大学で日本語を専攻する2～4年生の合計26名に、約1,500字の作文を書いてもらった。平均年齢は20歳10カ月で、標準偏差1歳2カ月であった。3カ月以上日本に滞在した日本語学習者はいなかった。

3.2 作文のテーマ

作文のテーマと人数は、「中国の若者が見つけた日本の新しい魅力」が14名、「日本の「心に残る、先生のあの言葉」が10名、「日本の中国語の日に私ができること」が2名であった。

3.3 学習者の文法能力の測定

作文を書くには、日本語学習者の文法能力が影響するのではないかとと思われる。そこで、宮岡・玉岡・酒井(2014)の改訂版である早川・玉岡(2015)の構造分類による日本語

の文法知識テストを使って、26 名の日本語学習者の文法能力を測った。図1に示したように、この文法能力テストは形態素変化 (morphological inflections)、局所依存 (local dependency) および構造の複雑性 (complex structure) の3つの構造分類からなる各 12 問の合計 36 問のテストである。日本語学習者の文法に関するテストなので、難しい語彙は使わず、旧・日本語能力試験の配当級で4～3級にあたる初級段階で学習する語彙のみが使用されている。

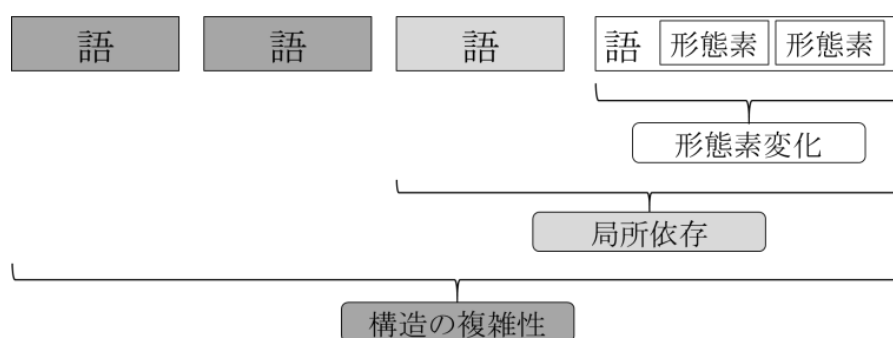


図1 文の構成要素に基づいた文法能力の下位尺度

注: 宮岡・玉岡・酒井(2014)および早川・玉岡(2015)より

早川・玉岡(2015)は、文法能力を、1つのまとまった文を理解するのに必要な語内外の一連の規則に関する知識と定義している。これら3つの関係は、1つの文の中で生じる形態素と形態素の関係(形態素変化)、隣接する語と語の関係(局所依存)、文全体での隣接しない語と語の関係(構造の複雑性)と、小さな規則単位から大きな規則単位へと構造的に拡大している。宮岡・玉岡・酒井(2014)および早川・玉岡(2015)の文法テストでは、これら3つをテストの下位尺度としている。

具体的には、形態素変化は、1つの単語内における形態的变化や助動詞との複合的接続と配列に関する知識であり、「形態素」+「形態素」で一つの単語を作るレベルの規則である。たとえば、「さっきの説明は()。 」という文に対し、「ややこしくありませんでしたか」「ややこしくありませんでしたか」「ややこしいじゃなかったですか」「ややこしいじゃないでしたか」の4つの選択肢から、正しいものを1つ選ばせるものである。つまり、形態素変化は単語の品詞を判別した上で、助動詞における複合的接続形の適切さの知識が問われる。局所依存は、隣接する「単語」+「単語」の適切な共起関係の知識である。たとえば、「私は()人が好きだ。」の「人」という名詞に隣接する形容詞が、「かっこいい」「かっこいい

の「かっこよく」「かっこいいな」のどれがふさわしいかを判断させる問題が挙げられる。つまり、局所依存は、「かっこいい人」のように、隣接する単語との関係によって活用の形の選択に複数の可能性があり、隣接する単語は相互依存的である。最後に、構造の複雑性は、1つの文の中で、離れて位置する単語間の相互関係に関する知識である。たとえば、「彼女がいつ日本に行くと(), 田中さんに聞いてみてください。」のように、共起的制約の知識を問うものなどがある。括弧内に「思う」「思うか」「思うかどうか」「思いますか」のいずれか1つを選ぶ問題である。局所依存が隣接による結合関係であるのに対し、構造的複雑性は、語が呼応を要求し、結合関係にある語同士が文中で離れた位置関係にある。

なお、改訂版の文法テストは、中国人日本語学習者 143 名を対象に実施している。クロンバック信頼度係数は、 $\alpha=0.82$ と非常に高く、日本語学習者の文法知識を簡便かつ安定して測定することができることを確認した(早川・玉岡, 2015)。

表1 26名の学習者の文法テスト得点

文法テスト	平均	標準偏差	最小値	最大値
形態素変化	6.73	2.07	2	11
局所依存	6.96	1.87	2	10
構造の複雑性	6.15	2.07	2	12
文法点数	19.85	4.72	11	29

本研究の 26 名の学習者の文法テストの得点は、表1に示したように、最も得点の低い学習者は 11 点で、最も得点の高い学習者は 29 点であった。26 名の学習者の平均は 19.85 点で、標準偏差が 4.72 点であった。文法知識の下位尺度は、形態素変化、局所依存、構造の複雑性の3つであるが、形態素変化で最も得点が低い学習者は2点で、最も高い学習者は 11 点であった。平均は 6.73 点で、標準偏差は 2.07 点であった。局所依存は、最も得点の低い学習者は2点で、最も高い学習者は 10 点であった。平均は 6.96 点で、標準偏差は 1.87 点であった。構造の複雑性は、最も得点の低い学習者は2点で、最も高い学習者は 12 点であった。26 名の学習者の平均は 6.15 点で、標準偏差は 2.07 点であった。

これらの下位尺度の得点はRによるバイオリンプロット(violin plot)で図2のような分布となる。バイオリンプロットはそれぞれの分布を四分位で視覚的に分かり易く示してくれる。たとえば、構造の複雑性については、第1四分位(Q_1)は5点、第2四分位あるいは中央値

(Q_2)は6点, 第3四分位(Q_3)は7点であった。これに対して, 局所依存については, 第1四分位は6点, 第2四分位は7点, 第3四分位は8点であった。さらに, 形態素変化については, 第1四分位の値は5点, 第2四分位は7点, 第3四分位の値は8点であった。

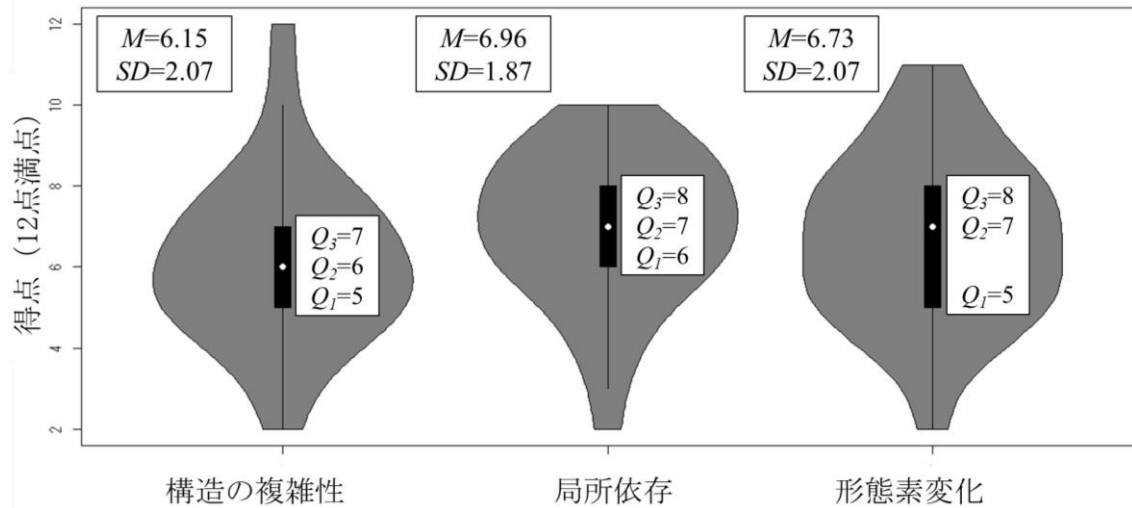


図2 文法得点のバイオリンプロット

注: M は平均, SD は標準偏差を示す。

また, Q_1 は第1四分位, Q_2 は第2四分位, Q_3 は第3四分位を指す。

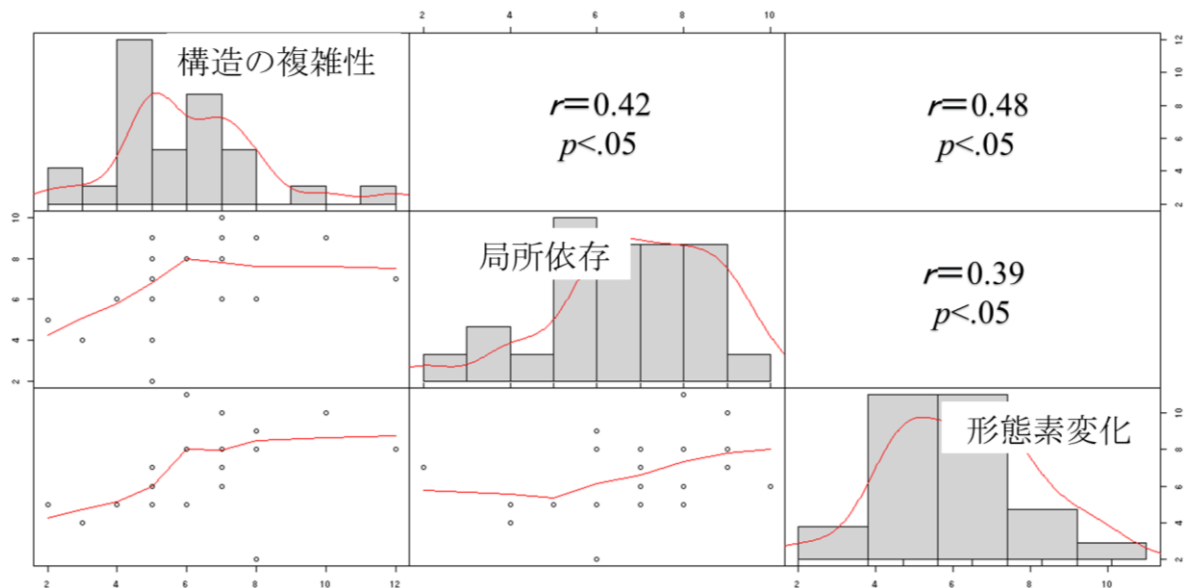


図3 3つの下位尺度の分布と相関

また、文法テストが持つ3つの下位尺度内の得点分布および3つの下位尺度間の相関をもっと分かり易く説明するため、R の `pairs.panels()` 関数を使って、図3の散布および相関行列図を作成した。文法テストが持つ3つの下位尺度間の相関係数は、図3に示したように、構造の複雑性と局所依存の相関係数は 0.42 ($p<.05$) であり、構造の複雑性と形態素変化の相関係数は 0.48 ($p<.05$) であった。また、形態素変化と局所依存の相関係数は 0.39 ($p<.05$) であった。これらの相関係数はすべて5%水準で有意であるが、全体的に中くらいの相関が示された。さらに、これら3つの下位尺度の得点について反復のない分散分析で検討した。その結果、主効果は有意ではなく [$F(2, 75)=1.18, p=.32$]、ほぼ同じ得点であることが分かった。つまり、文法テストの3つの下位尺度の問題は、異なるテスト内容であるものの、3つの下位尺度は同じような平均を示した。

3.4 リーダビリティの測定と作文の特徴

リーダビリティは、文章の読み易さのことであり、1文あたりの文字数や語数といったコーパス情報をもとに、文章の難しさを数値化したものである(英語の研究は、Flesch, 1948; Smith & Kinkaid, 1970; 日本語の研究は、Hasebe & Lee, 2015; 酒井, 2011; 佐藤, 2011; 柴崎・原, 2010; 建石・小野・山田, 1988)。詳細の26名の作文の特性は表2に示した。まず、26名の作文のリーダビリティ(読解可能指標)を『日本語文章難易度判別システム』(<https://jreadability.net/sys>)で調べた。なお、リーダビリティ(=Y)の算出式は、

$$Y = -0.056 \times \text{平均文長} - 0.126 \times \text{漢語率} - 0.042 \times \text{和語率} - 0.145 \times \text{動詞率} - 0.044 \times \text{助詞率} + 11.724$$

である。この指標は、11.724から、文長、漢字、和語、動詞、助詞の5つの変数の割合が大きくなるにつれて、リーダビリティの得点が低くなるという計算である。

リーダビリティは0.50~6.40の連続変数で示される。数値が小さいほうがより読み難い文章であり、大きい方が読み易い文章とされる。26名の日本語学習者の作文のうち、最も読み易かったのは3.87で、読み難かったのは2.86であった。また、平均は3.38であり、標準偏差は0.28であったので、26名の作文のリーダビリティ得点は2.86~3.87の狭い範囲に収まっており、26名の学習者がほぼ同じような難しさ(即ち、中級レベル)の作文を書いたといえよう。

表2 26名の作文の特性および相関係数

作文の特性	1	2	3	4
1 リーダビリティ	—			
2 総文数	0.66 ***	—		
3 総文字数	0.34	0.16	—	
4 一文の平均語数	-0.68 ***	-0.92 ***	0.10	—
平均	3.38	44	1,552	23
標準偏差	0.28	6	64	3
最小値	2.86	35	1,326	17
最大値	3.87	61	1,686	30

注: $N=26$. *** $p<.001$.

表2に示したように、総文数は、最も少ない作文が 35 文で、最も多い作文が 61 文であった。26名の平均の総文数は 44 文で、標準偏差が6文であった。総延べ文字数は、最も少ない作文が 1,326 字で、最も多い作文が 1,686 字であった。26名の平均は 1,552 字で、標準偏差は 64 字であった。1文の平均語数は、最も少ない作文は 17 語で、最も多い作文は 30 語であった。26名の平均は 23 語で、標準偏差は 3 語であった。また、作文の総文数と1文の平均語数 ($r=-.92$, $p<.001$) の相関係数は極めて高く、逆相関であった。つまり、文数が多い作文は、1文あたりの語数が少ないことを示している。文数の多い作文は、語彙数の少ないシンプルな文で書かれているのではないかとと思われる。リーダビリティと総文数 ($r=.66$, $p<.001$) は高い正の相関であるが、リーダビリティと 1 文あたりの語数 ($r=-.68$, $p<.001$) は高い負の相関である。

この関係を視覚的に把握するために、26名の作文について、リーダビリティと総文数および1文あたりの平均語数の散布図を図4に描いた。散布図に描かれた近似直線は、お互いに逆の方向を示している。相関係数で因果関係を語ることはできないが、これにより、文数が多い作文はリーダビリティの得点が高く、難易度の低い作文であり、1文当たりの語数が多い作文はリーダビリティの得点が低く、難易度の高い作文である傾向があることが窺える。総文数が多く、文字数の少ないシンプルな文で書かれている場合のほうがリーダビリティが高く、分かり易い作文になっていることが分かる。つまり、文の数が多い作文は、短い文の繋がりて内容を説明しようとする傾向があるからではないかと思われる。平均文長がリーダビリティの重回帰式に含まれていることもあるが、この傾向はかなり普遍性がある可能性がある。また、1 文あたりの平均語数が少ないほうが、リーダビリティの得点が高く

なり、読みやすくなっていることが分かる。

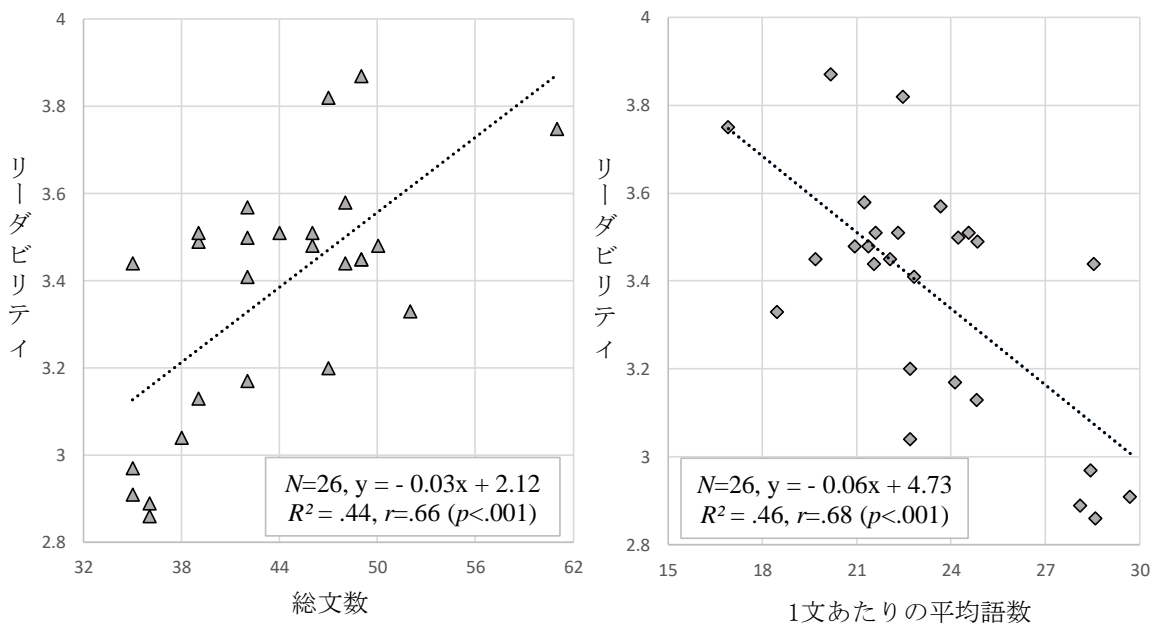


図4 リーダビリティと総文数および1文あたりの平均語数の散布図

3.5 作文の語彙特性

26名の作文の語彙的な難易度を判定するために、『日本語教育語彙表』(科学研究費補助金基盤研究(A)「汎用的日本語学習辞書開発データベース構築とその基盤形成のための研究」研究代表者:砂川有里子)を利用した。『日本語教育語彙表』は、「現代日本語書き言葉均衡コーパス(BCCWJ)」および日本語教科書のコーパスの語彙調査によって17,920語を選定して、経験豊富な日本語教師が初級・中級・上級の3段階で前半・後半にそれぞれ分かれた6段階で難易度を主観的に判定した結果をデータ化したものである。旧・日本語能力試験の級、品詞・語種・表記・アクセントなどの情報および語義と用例の検索エンジン(jreadability.net/jev)が公開されている。表3に示したように、26名の作文で使われた語彙は、初級レベルが最も多く、初級前半の平均語数が116語で、標準偏差が35語であった。初級後半は、平均語数が106語で、標準偏差が23語であった。中級レベルでは、中級前半の平均語数が96語で、標準偏差が18語であった。中級後半は、平均語数が84語で、標準偏差が23語であった。上級レベルになると、語彙数が著しく少なくなり、上級前半の平均語数が19語で、標準偏差が7語であった。上級後半の語彙はきわめて少なく、26名の上級後半の平均語数は1語となった。

また、品詞分類に従って分類すると、助詞、普通名詞、助動詞および動詞という4種類の品詞が多く使われていることが分かる。これらの品詞の平均語数は表3に示した通りである。具体的には、「の」「は」などのような助詞の平均語数が 293 語で、標準偏差が 27 語であった。「伝統」「観光」などのような普通名詞の平均語数が 230 語であり、標準偏差が 23 語であった。時制の「た」(過去を表す)またはムードである「たい」(願望を表す)などのような助動詞では、平均語数が 84 語で、標準偏差が 21 語であった。「いる」「する」などのような動詞では、平均語数が 57 語であり、標準偏差が 12 語であった。

表3 26名の作文で使用された語彙特性別の平均頻度

	語彙特性	平均	標準偏差	最小値	最大値
学 習 レ ベ ル	初級前半	116	35	65	196
	初級後半	106	23	57	144
	中級前半	96	18	57	126
	中級後半	84	23	41	132
	上級前半	19	7	7	35
	上級後半	1	1	0	6
品 詞	助詞	293	27	219	337
	普通名詞	230	23	192	282
	助動詞	84	21	44	137
	動詞	57	12	35	77
	固有名詞	25	19	0	61
	代名詞	23	16	5	57
	形容詞	21	8	10	42
	副詞	19	6	12	32
	形状詞	16	6	7	26
	連体詞	10	6	2	22
	接続詞	4	2	0	10
	その他	219	23	151	261
	語 種	和語	659	79	493
漢語		172	34	115	268
外来語		17	16	1	52
混種語		9	5	3	19

一方、26名の学習者による作文においては、品詞の使用頻度が低い順に並べると、接続詞、連体詞、形状詞、副詞、形容詞、代名詞、固有名詞であった。「しかし」「そして」などのような接続詞は、平均語数が4語であり、標準偏差が2語であった。「この」「大きな」な

どのような連体詞は、平均語数が 10 語であり、標準偏差が6語であった。「好き」「いろいろ」などのような形状詞は、平均語数が 16 語であり、標準偏差が6語であった。「ずっと」「たくさん」などのような副詞は、平均語数が 19 語であり、標準偏差が6語であった。そして、「長い」「面白い」などのような形容詞では、平均語数が 21 語であり、標準偏差が8語であった。「それ」「私」などのような代名詞では、平均語数が 23 語であり、標準偏差が 16 語であった。また、「日本」「中国」などのような固有名詞は、平均語数が 25 語であり、標準偏差が 19 語であった。つまり、簡単な文を構成するための一般的な項目である助詞、普通名詞、助動詞および動詞より、複雑な文を構成するための接続詞、連体詞、形状詞及び副詞はあまり使われない傾向がみられる。

最後に、語種による平均語数からみると、「言葉」「分かる」などのような和語は圧倒的によく使われ、平均語数が 659 語であり、標準偏差が 79 語となった。その次に、「料理」「交流」などのような漢語は平均語数 172 語であり、標準偏差が 34 語であった。それに対し、「アニメ」「カメラ」などのような外来語は、平均語数が17語であり、標準偏差が16語となった。「素晴らしい」「気持ち」などのような混種語は、4つの語種のうち一番少なく、平均語数が9語であり、標準偏差が5語であった。

4. 文法能力による作文の品詞使用の予測—重回帰分析

本研究の目的である、日本語学習者の文法能力が作文のどの言語特性に影響するかを、重回帰分析のステップワイズ法(以下、同じ分析法)で検討した。26名の文法テストの3つの下位尺度で、作文の語彙の特性を予測した。以下、どの重回帰式でも有意な予測変数は1つしかなかったため、 β 値と t 値は、唯一の有意な変数についての数値である。また、ステップワイズ法では、有意でない変数の結果は算出されないため「—」で示した。

リーダビリティを従属変数とし、形態素変化、局所依存、構造の複雑性を独立変数とした重回帰分析を行ったが、どの下位尺度も有意な予測変数にはならなかった。文法能力はリーダビリティとは関係がなかった。この26名の日本語学習者の作文のリーダビリティ得点は、2.86から3.87の狭い範囲に収まっており、ほぼ同じような難しさの作文を書いているので、有意な予測変数が取れ難いと考えられる。

26名の書いた作文の語彙の難易度別の出現比率について、文法テストの3つの下位尺度で予測する重回帰分析を行った。結果は表4に示した。初級前半の語彙 $\beta=0.43$,

$t=2.36, p<.05$]は、構造の複雑性が有意な予測変数となった。初級後半、中級前半の語彙は多いものの、有意な予測変数はなかった。中級後半の語彙[$\beta=-0.50, t=-2.81, p<.01$]は、構造の複雑性が有意な予測変数となった。上級前半の語彙[$\beta=-0.46, t=-2.52, p<.05$]は、形態素変化が有意な予測変数となった。上級後半の語彙の出現比率 ($M=0.19\%$) は非常に少ないために、予測する下位尺度がなかったのであろう。

表4 文法能力による作文の語彙の難易度の予測

語彙 難易度	比率(%)		回帰分析の結果		重回帰分析の結果			R^2
	M	SD	β 値	t 値	複雑性	局所依存	形態素変化	
初級前半	27.31	7.16	0.43	2.36	*	—	—	0.19
初級後半	24.92	4.53	—	—	—	—	—	—
中級前半	22.77	4.42	—	—	—	—	—	—
中級後半	20.23	5.98	-0.50	-2.81	**	<i>ns</i>	<i>ns</i>	0.25
上級前半	4.58	1.77	-0.46	-2.52	<i>ns</i>	<i>ns</i>	*	0.21
上級後半	0.19	0.40	-0.41	-2.19	—	—	—	—

注: * $p<.05$. ** $p<.01$. *** $p<.001$.

初級前半の語彙がもっとも頻繁に使われていた。これは基本語彙であり、文のさまざまな位置で使用される。そのため、構造の複雑性が有意な予測変数となったと考えられる。また、中級後半の語彙の出現率 ($M=20.23\%$) も高く、これらはレベルの高い語彙であり、こうした語彙の使用は、文法力の複雑性の下位尺度が有意な予測変数となった。基本的に、文の構造の複雑性とやや弱いものの形態素変化の知識が、作文の語彙使用における重要な要因になっているようである。

表5 作文における語種の使用頻度に影響する文法能力

語 種	比率(%)		回帰分析の結果		重回帰分析の結果			R^2
	M	SD	β 値	t 値	複雑性	局所依存	形態素変化	
和 語	76.73	5.65	0.58	3.47	**	<i>ns</i>	<i>ns</i>	0.33
漢 語	20.38	4.64	-0.52	-2.97	**	<i>ns</i>	<i>ns</i>	0.27
外来語	1.92	1.94	—	—	—	—	—	—
混種語	1.00	0.63	—	—	—	—	—	—

注: * $p<.05$. ** $p<.01$. *** $p<.001$.

4つの語種について、文法テストの3つの下位尺度で予測する重回帰分析を行った。表5に示したように、外来語と混種語を有意に予測する下位尺度はなかった。和語 [$\beta=0.58, t=3.47, p<.01$] および漢語 [$\beta=-0.52, t=-2.97, p<.01$] については、共に構造の複雑性の能力のみが有意な予測変数となった。

また、品詞別の頻度を文法テストの3つの下位尺度で予測する重回帰分析を行った。結果は、表6に示したように、連体詞、形状詞、接続詞、代名詞および固有名詞については、有意に予測する下位尺度はなかった。動詞 [$\beta=0.48, t=2.68, p<.05$]、副詞 [$\beta=0.45, t=2.43, p<.05$]、助動詞 [$\beta=0.54, t=3.12, p<.01$]、普通名詞 [$\beta=-0.50, t=-2.84, p<.01$]、助詞 [$\beta=-0.42, t=-2.26, p<.05$] のすべてにおいて、構造の複雑性が有意な予測変数となった。形容詞 [$\beta=0.41, t=2.22, p<.05$] については、局所依存が有意な予測変数となった。形容詞は名詞を修飾する品詞であり、語の並び方の規則である局所依存が、こうした品詞の使用を促したのであろう。しかし、動詞や形容詞の活用に関係した形態素変化は、作文のどの品詞の頻度も予測しなかった。

表6 作文における品詞の使用頻度に影響する文法能力

品 詞	比率(%)		回帰分析の結果		重回帰分析の結果			
	<i>M</i>	<i>SD</i>	β 値	<i>t</i> 値	複雑性	局所依存	形態素変化	<i>R</i> ²
動 詞	7.23	1.53	0.48	2.68	*	<i>ns</i>	<i>ns</i>	0.23
副 詞	2.42	0.58	0.45	2.43	*	<i>ns</i>	<i>ns</i>	0.20
助 動 詞	10.77	2.44	0.54	3.12	**	<i>ns</i>	<i>ns</i>	0.29
普通名詞	29.50	3.47	-0.50	-2.84	**	<i>ns</i>	<i>ns</i>	0.25
助 詞	37.35	2.13	-0.42	-2.26	*	<i>ns</i>	<i>ns</i>	0.18
形 容 詞	2.69	0.97	0.41	2.22	<i>ns</i>	*	<i>ns</i>	0.17
連 体 詞	1.23	0.82	—	—	—	—	—	—
形 状 詞	2.00	0.85	—	—	—	—	—	—
接 続 詞	0.42	0.50	—	—	—	—	—	—
代 名 詞	2.88	1.90	—	—	—	—	—	—
固有名詞	3.31	2.54	—	—	—	—	—	—

注: * $p<.05$. ** $p<.01$. *** $p<.001$.

文全体の複雑な構造の理解は、述部を構成する動詞、助動詞および副詞、さらに主語と目的語を構成する普通名詞と助詞の使用頻度に影響していた。動詞、副詞、助動詞は述部を、普通名詞と助詞は名詞句を作る。いずれも、文を作る主要な品詞である。これは、

文全体をつかさどる文法規則の理解が、この種の品詞の使用を促進したのであろう。一方、形容詞は局所依存が有意な予測変数であった。これは、形容詞が名詞を修飾することを考えれば、隣り合う語の局所依存の関係の理解が、形容詞の使用頻度に影響するのは理解できよう。動詞や形容詞に関係している形態素変化の文法能力は、どの品詞の使用頻度にも影響がなかった。語自体の活用などの知識は、作文の語彙の頻度特性には影響しないようである。総じて、文全体をつかさどる文法規則(複雑な文の理解)が作文の文全体を作るのに貢献しているようである。

5. 26 名の作文特性についての検討—クラスタ分析と判別分析

重回帰分析の結果で有意であった動詞、副詞、助動詞、普通名詞、助詞、形容詞の6つの使用頻度の比率を変数として、26名の日本語学習者のクラスタ分析を行った。クラスタ間の距離はワード法、日本語学習者間の距離は、平方ユークリッド距離を使って、階層的クラスタ分析を行った。その結果、25ポイント中の7ポイントで3つのクラスタが得られた。これらのクラスタの分類の精度を判別分析で調べた結果、本データのクラスタ化について100%が正しく判別された。また、交差妥当化の結果、96.2%の判別的中率であった。2つの判別関数が、有意に3つのクラスタを予測した。第1正準判別関数(#1)は、69.6%を説明し、有意であった[Wilk's $\lambda=0.05$, $\chi^2(12)=61.85$, $p<.001$]。第2正準判別関数(#2)は、30.4%(累積は100%)を説明し、やはり有意であった[Wilk's $\lambda=0.31$, $\chi^2(5)=24.35$, $p<.001$]。構造行列から各変数と判別関数間の最大絶対相関が有意であった変数をみると、3つのクラスタ化について助詞が最も強い負の方向の原因帰属(#1=-0.79)を示し、次に形容詞(#1=-0.22)、普通名詞(#2=0.82)、助動詞(#2=-0.57)、副詞(#2=-0.22)、動詞(#2=-0.21)であった。

2つの正準判別関数の係数を使って、3つのクラスタに分類した26名の日本語学習者による3つの下位分類を含む文法得点および作文における品詞の使用状況について、表7のようにまとめた。3つのクラスタにおける6種類の品詞(助詞、形容詞、普通名詞、助動詞、副詞、動詞)の使用率をみると、全体的に助詞と普通名詞がほかの4種類より遥かに多く使われて、24~40%の使用率をもっていることに対して、形容詞と副詞の使用率は少なく、1~5%の間の値をとる。これは3つのクラスタにおける共通特徴である。

表7 各クラスタにおける学習者による品詞使用

ID	助詞 (%)	形容詞 (%)	普通名詞 (%)	助動詞 (%)	副詞 (%)	動詞 (%)	文法得点M	複雑性M	局所依存M	形態素変化M	文法点数	複雑性	局所依存	形態素変化
ク	38	3	33	8	2	7					15	4	6	5
ラ	36	3	38	9	2	5					19	5	8	6
スタ	38	3	34	8	3	7					21	5	9	7
1	37	2	32	9	2	7	17	5	6	6	19	6	8	5
	38	3	33	10	2	6					12	2	5	5
	38	2	33	8	2	7					21	7	6	8
	36	2	34	9	2	7					14	5	2	7
ク	35	4	25	13	4	9					27	12	7	8
ラ	37	3	31	12	2	6					16	5	6	5
スタ	34	2	33	13	2	6					14	5	4	5
2	35	2	30	13	2	9					23	8	6	9
	33	2	27	13	3	10	22	8	7	7	23	7	9	7
	34	2	28	16	2	9					16	8	6	2
	35	1	28	12	3	9					21	7	6	8
	36	2	24	14	3	10					25	8	9	8
	37	2	28	13	3	6					29	10	9	10
ク	40	3	29	10	2	6					17	5	7	5
ラ	39	4	30	6	3	5					19	5	7	7
スタ	40	2	28	8	3	5					18	5	7	6
3	40	3	29	8	2	6					23	7	8	8
	40	2	29	11	2	8	20	6	8	7	17	5	7	5
	40	3	26	12	2	8					25	6	8	11
	39	5	27	11	2	9					26	7	9	10
	38	5	26	12	3	6					23	7	10	6
	38	3	24	13	2	8					22	6	8	8
	40	2	28	9	3	7					11	3	4	4

注：IDは学習者ID, Mは平均得点を指す。

しかし、3 クラスタのうち、もっとも文法能力が高かったクラスタ 2 の学習者 ($M=22$ 点) は、動詞と助動詞を多く使っているのが顕著である。こうした学習者は文法能力が高いので、文を完成するための構文要素としての助詞、普通名詞、助動詞、動詞をよく使う特徴を持ち、文法能力が高い学習者とされる。つまり、クラスタ 2 は文を完成するために多様な品詞をうまく使用できる文法能力の高い学習者と考えられる。

それに対して、3 つのクラスタのうち、もっとも文法能力が低かったクラスタ 1 ($M=17$ 点) では、動詞と助動詞が比較的になく使われ、両者の使用率がほぼ 10% に留まっている。つまり、クラスタ 1 の学習者は文法能力が低いので、助詞と普通名詞はある程度使えるが、基本的な構文要素であるほかの品詞使用はまだ養成されていないと思われる。こうした学習者は助詞と普通名詞のみを使い、文法能力はやや低い学習者といえよう。

さらに、表 7 にまとめた結果を参照すると、クラスタ 3 ($M=20$ 点) の学習者はクラスタ 1 と比べれば、文法能力がやや高かった。特に、文法能力における「局所依存」の得点に関しては、3 つのクラスタのうち、もっとも高かった学習者である ($M=8$ 点)。このような学習者による助詞の使用率は 40% となり、もっとも高かったのが一番の特徴である。また、動詞と助動詞の使用はクラスタ 1 とクラスタ 2 の間に位置しているが、使用傾向についてクラスタ 3 はクラスタ 2 の学習者に類似するのが使用率から分かる。クラスタ 3 の学習者は、文法能力が向上すれば、品詞を多様に使用できるようになると期待される学習者と考えられる。

以上のように、クラスタ化した結果から、異なる文法能力を持つ学習者の品詞使用の特徴がみえてくる。3 つのクラスタにおける使用特徴をまとめてみると、26 名の学習者による文法能力の向上と共に、品詞の使用が「普通名詞、助詞」から、「普通名詞、助詞、動詞、助動詞」へと拡張していく傾向があるようである。

6. まとめ

本研究は、26 名の日本語学習者が作成した作文における文法能力を考察した。文法と品詞との因果関係について、文法的な理解を構成する複雑な構造、局所依存、形態素変化から 12 種類の品詞の使用を重回帰分析で検討した。3 つの文法的な理解力のうち、作文における動詞、助動詞、副詞の使用頻度には複雑な構造の理解力が強く

影響した。局所依存は形容詞の頻度のみで、形態素変化はどの品詞にも影響しなかったとみられた。つまり、文法能力、特に、複雑な構造と局所依存のような文法規則に従って作られたものが、作文における品詞の使用頻度に影響することが示された。

作文における品詞の使用頻度（比率）に基づいて、26名の日本語学習者に対して、クラスタ分析を行った。その結果、26名の学習者は「助詞と普通名詞を頻繁に使い、文法能力はやや低い学習者」「文を完成するために品詞を多様に使える文法能力の高い学習者」「文法能力の向上とともに、品詞の使用も豊富になると期待される学習者」という3つのクラスタに分けられた。どのクラスタでも助詞、普通名詞の使用頻度が高かった。それぞれの特徴をみると、クラスタ1は文法能力の低い学習者であり、ほかの品詞と比べ、助詞と普通名詞だけをもっともよく使っていた。クラスタ2は文法能力の高い学習者であり、特に動詞、助動詞をよく使用していた。このクラスタの学習者は3つの下位分類を持つ文法能力が高いので、文を構成する要素としての品詞を多様に使用できる傾向がみられた。クラスタ3は文法能力が、ほかの2つのグループの間に位置する学習者であるが、「局所依存」という文法能力が3つのクラスタのうちもっとも高かった。そのため、クラスタ1より、助詞と普通名詞を多く使っているが、動詞、助動詞もよく使い始めているようになったと示された。

以上のように、本研究で考察した26名の学習者の作文における品詞の使用分布から、文法能力の向上と共に、品詞の使用が普通名詞、助詞から、普通名詞、助詞、動詞、助動詞へと拡張していく傾向があるようである。つまり、文法の習得によって、より繊細で微妙な表現が可能となり、作文の内容に多様性をもたらすという効果があるといえよう。

[引用文献]

- 市川保子（1993）「中級レベル学習者の誤用とその分析—複文構造習得過程を中心に—」『日本語教育』81号，55-66.
- 大塚薫・林翠芳（2010）「中上級レベルの日本語学習者の作文教育—意見文にみる語彙・漢字使用及び誤用の分析結果を踏まえて—」『高知大学総合教育センター修学・留学生支援部門紀要』第4号，47-66.
- 川口良（1995）「中上級日本語学習者の作文にみる誤用の一例」『言語文化と日本語

- 教育』9, 178-188.
- 酒井由紀子 (2011) 「健康医学情報を伝える日本語テキストのリーダビリティの改善とその評価——一般市民向け疾病説明テキストの読みやすさと内容理解のしやすさの改善実験——」『Library and Information Science』65, 1-35.
- 佐藤理史 (2011) 「均衡コーパスを規範とするテキスト難易度測定」『情報処理学会論文誌』52-4, 1777-1789.
- 柴崎秀子・原信一郎 (2010) 「12 学年を難易尺度とする日本語リーダビリティ判定式」『計量国語学』27-6, 215-232.
- 鈴木智美 (2002) 「2000 年度中級作文に見られる語彙・意味に関わる誤用——初中級レベルにおける語彙・意味教育の充実を目指して——」『東京外国語大学留学生日本語教育センター論集』28, 27-42.
- 建石由佳・小野芳彦・山田尚勇 (1988) 「日本文の読みやすさの評価式」『文書処理とヒューマンインターフェース』18-1, 1-8.
- 村田裕美子 (2018) 「ドイツ語を母語とする日本語学習者の作文に見られる言語的特徴——習熟度の差は産出にどう現れるのか——」『第二言語としての日本語の習得研究』21, 61-76.
- 原沢伊都夫 (2012) 「日本語初中級学習者の作文指導——学習者の誤用分析をもとに——」『静岡大学国際交流センター紀要』6, 79-92.
- 早川杏子・玉岡賀津雄 (2015) 「改訂版・構造分類による日本語文法知識テストの開発——中国人日本語学習者のデータによるテスト評価——」『ことばの科学』29, 5-24.
- 宮岡弥生・玉岡賀津雄・酒井弘 (2014) 「日本語の文法能力テストの開発と信頼性——日本語学習者のデータによるテスト評価——」『広島経済大学研究論集』36(4), 33-46.
- 楊帆 (2014) 「中級日本語学習者の作文における困難点——文構造の呼応関係について——」『秋田大学国際交流センター紀要』3, 15-28.
- Flesch, Rudolf. (1948). A new readability yardstick. *Journal of Applied Psychology* 32, 221-33.
- Hasebe, Yoichiro and Lee, Jae-Ho. (2015). Introducing a Readability Evaluation System for Japanese Language Education. *Proceedings of the 6th International Conference on Computer Assisted Systems for Teaching & Learning Japanese (CASTEL/J)*, 19-22.
- Smith, Edgar A. and J. Peter Kincaid. (1970). Derivation and validation of the automated readability index for use with technical materials. *Human Factors* 12(5), 457-464.

日本語文章難易度判別システム<<https://jreadability.net/>> (2019年5月4日, アクセス)

楊 艶萍 - 中南林業科技大学外国語学院 講師

玉岡 賀津雄 - 名古屋大学大学院人文学研究科 教授

張 婧禕 - 名古屋大学・日本学術振興会 外国人特別研究員