

# 中国人日本語学習者の書いた作文に対する評価行動

— プロトコル分析による日本母語話者教師と中国語母語話者教師の比較 —

方 正

(2017年10月4日受理)

Evaluation Behavior toward Essays by Chinese Learners of Japanese:  
A Comparison between Native Japanese-speaking Teachers and Native Chinese-speaking  
Teachers by Means of Protocol Analysis

Zheng Fang

**Abstract:** This study used think-aloud protocols to compare evaluation behaviors between native Japanese-speaking teachers and Chinese-speaking teachers when they evaluate writing by Chinese learners of Japanese. A group of three native Japanese-speaking teachers and three native Chinese-speaking teachers evaluated four essays using both holistic and analytic scales while providing think-aloud protocols. The findings show that the native Chinese-speaking teachers paid more attention to the grammar used in the essays, while the native Japanese-speaking teachers were more likely to confirm the meaning and also made judgments about the logical aspect of the support for the main idea. Furthermore, it was found that, even for the same essay, the native Japanese-speaking teachers and Chinese-speaking teachers both had different views regarding the coherence of the main idea.

**Key words:** assessment, writing, validity, protocol analysis, evaluation behavior

キーワード：評価, 作文, 妥当性, プロトコル分析, 評価行動

## 1. はじめに

学習者の言語運用を評価するにあたって、評価者は欠かせない存在である。評価研究の分野では、評価者の属性（性別、母語、教師経験の有無等）の違いが評価に影響を与えることが多くの研究によって指摘されており（Lumley, 2005; Shaw & Weir, 2007; Weigle, 2002）、属性の異なる評価者による評価の公平性や構成概念妥当性の確保が課題となっている（Shi, 2001）。

日本語教育分野における評価研究は、これまで評

価者の教師経験の有無が注目され、日本語を母語とする日本語教師と一般日本人による評価を比較する研究が多く行われてきた（田中・坪根・初鹿野, 1998; 崔, 2013; 野原, 2008; 渡部, 2005）。しかし、母語の異なる教師による評価の相違を扱った研究は見当たらない。

海外における日本語教育において、非母語話者教師は欠かせない存在である。国際交流基金（2012）の調査によると、中国の日本語教師の数は16,752名だが、日本語母語話者教師の割合はわずか14.2%である。こうした現状から、日本語母語話者教師（以下、JST）と中国母語話者教師（以下、CST）による評価を比較する必要があると言える。

また、日本語を母語とする日本語教師と一般日本人による評価を比較する研究のうち、学習者の発話を対象とした研究（崔, 2013; 野原, 2008; 渡部, 2005）は

---

本論文は、課程博士候補論文を構成する論文の一部として、以下の審査委員により審査を受けた。

審査委員：松見法男（主任指導教員）、渡部倫子、柳澤浩哉、間瀬茂夫

多いが、学習者の作文に対する評価を取り上げる研究(田中・坪根・初鹿野, 1998)は少ない。「書く」能力は日本語運用能力の重要な一つであり、それを測る作文評価の妥当性、信頼性について検証していく必要があると思われる。CONNOR - LINTON (1995a)は母語話者教師と非母語話者教師の評価基準を比較することによって、両者がお互いの評価を知る必要があると述べている。これにより、教師の属するそれぞれのグループの作文教育の目標を明らかにすることができるためである。加えて、両者の評価における問題点を発見し、その改善策を提案することも可能であろう。

そこで、本研究は中国人日本語学習者の書いた作文に対する JST と CST の評価に違いがないのか、あるとしたらその違いはどこから生まれるのかを明らかにすることを目的とする。

## 2. 先行研究

英語教育において、英語母語話者教師と非英語母語話者教師による作文の評価を比較する研究は多く行われている(CONNOR-LINTON, 1995a; Hamp-Lyons & Zhang, 2001; Kobayashi & Rinnert, 1996; Lee, 2009; Marefat & Heydari, 2016; Shi, 2001等)。Shi(2001)は、中国人英語学習者の書いた説明文10本について、英語母語話者教師(23名)と中国語母語話者教師(23名)に総合的評価(10段階)をさせ、そう評価した理由を自由記述で述べさせた。その結果、総合的評価の点数に違いは見られなかった。一方で、英語母語話者教師は「内容」と「言葉」についてのプラス評価を多く述べているのに対し、中国語母語話者教師は「構成」と「長さ」についてのマイナス評価を多く述べているという違いが明らかになった。

Lee (2009) は、英語母語話者教師(5名)と韓国語母語話者教師(5名)に韓国語母語話者教師の論証文420本を「全体」「内容」「構成」「語彙」「文構造」「文法とメカニクス<sup>1)</sup>」の項目について評価させ、ラッシュ分析を行った。その結果、英語母語話者教師が「全体」と「内容」について、韓国語母語話者教師が「構成」「文構造」「文法とメカニクス」についてより厳しく評価したことが分かった。また、評価者の各項目に対する重視度をアンケートで調べた結果、英語母語話者教師も韓国語母語話者教師も「内容」と「構成」を重視する一方で、韓国語母語話者教師は「文法とメカニクス」をも重視することが分かった。

Marefat & Heydari (2016) は英語母語話者教師(12名)とペルシャ語母語話者教師(12名)にイラン人英語学習者の論述文8本を「内容」「構成」「語彙」「文法」

の項目で評価させ、ラッシュ分析を行った。その結果、英語母語話者教師が「構成」について、ペルシャ語母語話者教師が「文法」についてより厳しく評価したことが分かった。また、評価者の各項目に対する重視度をアンケートで調べた結果、英語母語話者教師は「構成」を一番重視するのにに対し、ペルシャ語母語話者教師は「文法」を一番重視しているとの結果が得られた。

以上3つの先行研究のように、評者のつけた点数とアンケート回答に基づいて母語話者教師と非母語話者教師による評価を比較した研究が多い。しかし、CONNOR - LINTON (1995b)は評価の結果(点数)を分析の対象とするだけでは、そうした評価は何を意味するかが解明できないと指摘し、発話思考法<sup>2)</sup>等の方法を用いて、評価者が実際にどのように作文を評価しているのかを調べる必要があると指摘している。

発話思考法を用いて属性の異なる評価者による作文評価を比較する研究には、熟練した評価者と未熟な評価者を扱ったものが多いが(Barkaoui, 2010; Shaw & Weir, 2007等)、母語話者教師と非母語話者教師の評価を比較する研究はHamp-Lyons & Zhang (2001)がある。この研究では、英語母語話者教師(6名)と中国語母語話者教師(6名)が中国人英語学習者の書いた論述文2本を評価する際のプロトコル<sup>3)</sup>を分析した。その結果、評価者の異なる文化的背景が、レトリックパターン、客観性、説得力に関する認識の違いをもたらし、同じ作文内容についても両者の判断が異なることが報告された。ただし、Hamp-Lyons & Zhang (2001)は評価者の個別な発話を取り上げるだけで、評価者の行動全体を分析の対象としなかった。

以上のように、先行研究の多くは、母語話者教師と非母語話者教師の間に評価の違いがあることを示している。しかし、両者の評価行動全体を取り上げて比較する研究はない。本研究では、発話思考法を用いて、母語話者教師と非母語話者教師が実際に作文を評価する際の行動、つまり作文のどの要素に注目し、それから、どのように判断を行っているかを比較する。

## 3. 方法

本研究は、まず評価材料の作文と評価項目を選定し、評価表を作成した。それから、発話思考法を用いて評価者に作文を評価させ、その過程を録音した。録音したものを文字化し、評価者の評価行動を抽出し分類を行った。

### 3.1 評価材料

本研究は中国人日本語学習者の書いた4本の作文を評価材料とした。作文は、インターネット公開されて

いる I-JAS (『多言語母語の日本語学習者の横断コーパス』) に収録されている中国人日本語学習者の書いたものである。作文は「ファーストフードと家庭料理」をテーマとした600字前後の意見文である。作文の書き手は中国の大学で日本語を専攻として学ぶ3年生である。

### 3.2 評価項目

調査で用いる評価項目を選定するのにあたって、まず、日本語教育の研究分野で作文の評価基準を提案する先行研究から、142の評価項目を収集した。それらの評価項目を、作文指導経験のある JST 2名と CST 2名の協力を得て、KJ法<sup>4</sup>を用いて20評価項目のリスト案を作成した。さらに、日本語作文教育の専門家に意見を求め、評価項目リスト案の修正を行った。以上のプロセスを経て、本研究で用いる評価項目リスト(総合的評価項目1項目、分析的評価項目20項目<sup>5</sup>)を作成した。

### 3.3 評価者

JST10名、CST10名に4本の作文について、先述した評価項目リストを用いて評価させた。10名のJSTの日本語教師歴は最短2年、最長22年であり、平均9年1か月であった。彼らの作文指導歴は最短4か月、最長9年であり、平均3年半であった。10名のCSTの日本語教師歴は最短6年5か月、最長27年であり、平均12年8か月であった。彼らの作文指導歴は最短0、最長10年であり、平均3年3か月であった<sup>6</sup>。なお、本論文では、日本語教師歴と作文指導歴による影響を考慮し、日本語教師歴と作文指導歴の比較的近い3ペア(JST 3名、CST 3名)を比較の対象として取り上げる。6名の評価者の属性を表1に示す。

### 3.4 手続き

評価にあたっては、評価者に作文の課題文及び作文の書き手の日本語学習状況を示し、また作文を添削する必要がないように教示を行った。一人の評価者につき、4本の作文を評価してもらった。順序によるバイアスを避けるため、作文はランダムに提示した。

表1 評価者の属性<sup>7</sup>

|       | 日本語教師歴 | 作文指導歴     |
|-------|--------|-----------|
| JST 1 | 4年6か月  | 4か月       |
| JST 2 | 7年     | 4年        |
| JST 3 | 10年    | 4年        |
| CST 1 | 6年5か月  | 0年(4年)    |
| CST 2 | 7年     | 3年(0年)    |
| CST 3 | 11年3か月 | 2年3か月(5年) |

具体的な手続きは以下の通りである。(1) 1本の作文を読んでもらった後、(2) 評価項目リストに載せてある総合的評価項目1項目、分析的評価項目20項目について7段階(1点~7点)で評価してもらった。(3) 4本の作文の評価が全部終わった後、評価者が普段行っている作文の評価活動等についてフォローアップインタビューを行った。

なお、上記(1)(2)の過程において、発話思考法を用いた。評価者に作文を読む際、それから各評価項目について点数をつける際、読んでいること、考えていることをすべて声に出して述べるように指示し、それを録音した。また、発話思考法に慣れてもらうために、評価材料の4本の作文を評価してもらう前に、2本の250字程度の作文を読んで評価する練習を行った。

### 3.5 分析方法

調査で録音した評価者のプロトコルを文字化した。文字化した資料から、(1) 評価者が作文を読む際の評価行動を抽出した。抽出した評価行動は、Cumming, Kantor & Powers (2002)の評価行動の分類システム<sup>8</sup>によって分類し、日本語の作文の評価基準を提案する田中・阿部(2014)を参考に、カテゴリー名を決定した。(2) 各評価項目について点数をつける際の評価行動は、作文を読む際のものとの質の異なるものと見なし、本稿では主にその評価行動を評価項目ごと(分かりやすさ、正確さ等)にまとめる。また、(3) フォローアップインタビューで述べられた評価者の作文評価に対する考え方を評価行動に影響する要素として分析する。

## 4. 結果と考察

ここでは、評価者が(1) 作文を読む際の評価行動を「読む」段階のものとし、(2) 各評価項目について点数をつける際の評価行動を評定段階のものとして分けて分析し、それぞれの段階におけるJSTとCSTの評価行動の類似点と相違点をまとめる。両者の評価に違いが生じる原因については、先行研究と(3) インタビューの結果を用いて検討する。

### 4.1 「読む」段階について

3.で述べた評価行動の分類に基づき、「読む」段階において、JSTとCSTによる評価行動を比較する。まず、上位カテゴリー(「自己モニタリング」「内容」「構成」「日本語」)に基づいて分析を行う。次に、評価行動をカテゴリーごとにさらに下位分類し、各カテゴリー内におけるJSTとCSTの評価行動を比較する。

#### 4.1.1 上位カテゴリーに基づく分析の結果

本項では、上位カテゴリーに基づいて集計した評価

表2 上位カテゴリーに基づく評価行動の集計<sup>9</sup>

|       | 自己モニタリング |      | 内容 |      | 構成 |      | 日本語 |      | その他 |      | 総数 |
|-------|----------|------|----|------|----|------|-----|------|-----|------|----|
|       | 回数       | 出現率  | 回数 | 出現率  | 回数 | 出現率  | 回数  | 出現率  | 回数  | 出現率  |    |
| JST 1 | 13       | 0.17 | 15 | 0.20 | 7  | 0.09 | 39  | 0.51 | 2   | 0.03 | 76 |
| JST 2 | 12       | 0.15 | 11 | 0.13 | 6  | 0.07 | 45  | 0.55 | 8   | 0.10 | 82 |
| JST 3 | 7        | 0.12 | 21 | 0.35 | 8  | 0.13 | 18  | 0.30 | 4   | 0.07 | 58 |
| JST平均 | 11       | 0.16 | 16 | 0.23 | 7  | 0.10 | 34  | 0.45 | 5   | 0.06 | 73 |
|       |          |      |    |      |    |      |     |      |     |      |    |
| CST 1 | 14       | 0.21 | 4  | 0.06 | 1  | 0.01 | 41  | 0.60 | 7   | 0.10 | 67 |
| CST 2 | 13       | 0.32 | 1  | 0.02 | 0  | 0.00 | 27  | 0.66 | 1   | 0.02 | 42 |
| CST 3 | 13       | 0.22 | 12 | 0.20 | 3  | 0.05 | 31  | 0.53 | 0   | 0.00 | 59 |
| CST平均 | 13       | 0.24 | 6  | 0.10 | 1  | 0.02 | 33  | 0.60 | 3   | 0.04 | 56 |

自己モニタリング=自分の評価行動に対するコントロール、或はメタ評価 内容=作文の内容に関する評価行動  
 構成=作文の構成に関する評価行動 日本語=作文の日本語に関する評価行動 その他=判断できない評価行動

表3 「自己モニタリング」(下位カテゴリー)に基づく評価行動の集計

|       | 進行 |      | 再読 |      | 課題文 |      | 印象 |      | 比較 |      | 反省 |      | 総数 |
|-------|----|------|----|------|-----|------|----|------|----|------|----|------|----|
|       | 回数 | 出現率  | 回数 | 出現率  | 回数  | 出現率  | 回数 | 出現率  | 回数 | 出現率  | 回数 | 出現率  |    |
| JST 1 | 0  | 0.00 | 5  | 0.45 | 0   | 0.00 | 6  | 0.55 | 1  | 0.00 | 0  | 0.00 | 13 |
| JST 2 | 0  | 0.00 | 6  | 0.50 | 1   | 0.08 | 5  | 0.42 | 0  | 0.00 | 0  | 0.00 | 12 |
| JST 3 | 0  | 0.00 | 3  | 0.43 | 0   | 0.00 | 1  | 0.14 | 2  | 0.29 | 1  | 0.14 | 7  |
| JST平均 | 0  | 0.00 | 5  | 0.46 | 0   | 0.03 | 4  | 0.37 | 1  | 0.10 | 0  | 0.05 | 10 |
|       |    |      |    |      |     |      |    |      |    |      |    |      |    |
| CST 1 | 0  | 0.00 | 12 | 0.86 | 0   | 0.00 | 1  | 0.07 | 0  | 0.00 | 1  | 0.07 | 14 |
| CST 2 | 0  | 0.00 | 13 | 1.00 | 0   | 0.00 | 0  | 0.00 | 0  | 0.00 | 0  | 0.00 | 13 |
| CST 3 | 1  | 0.08 | 12 | 0.92 | 0   | 0.00 | 0  | 0.00 | 0  | 0.00 | 0  | 0.00 | 13 |
| CST平均 | 0  | 0.03 | 12 | 0.93 | 0   | 0.00 | 0  | 0.02 | 0  | 0.00 | 0  | 0.02 | 13 |

進行=評価行動を進める。或は止める 再読=既に読んだ文、表現、語彙を読み返す 課題文=課題文を確認する  
 印象=作文全体の印象を述べる 比較=既に評価した作文、或は評価者自身の評価経験を言及し、評価対象となる作文と比較する  
 反省=評価者自身の評価について反省する

行動について述べる。各カテゴリーに属する評価行動が出現した回数と、それが行動全数の中で占める割合(出現率)を表2に示す。なお、各評価者による評価行動の全数が異なるため、カテゴリーの出現率を用いて比較を行う。

「日本語」を注目点とする評価行動は、JST 1、JST 2、及びCST全員において、最も高い出現率を示している。JST 3においては、「日本語」は「内容」に次ぎ、二番目に出現率の高い評価行動のカテゴリーとなる。JSTもCSTも正確で適切な言語運用(文法の正しさ、語彙選択の適切さ等)に多く注目して評価対象の作文を読んでいることが分かる。「構成」を注

目点とする評価行動は、全体的にCSTよりJSTの出現率のほうが高かった。JSTはCSTより構成に多く注目して評価対象の作文を読んでいることがうかがえる。しかし、他のカテゴリーと比べると、「構成」についてはCSTはほとんど注目しておらず、JSTも出現率が低かった<sup>10</sup>。

「内容」を注目点とする評価行動は、CSTよりもJSTの出現率のほうが高い傾向であった。しかし、CSTグループの中でCST 3はJSTと同程度の出現率で「内容」について言及している。フォローアップインタビューでは、他のCSTと違って、CST 3は普段作文を評価する際に重視する項目として、「筋道が

通っている」ことを挙げている。CST グループ内でも、重視する項目に違いがあり、その違いによって評価行動が影響されることが想定される。

「自己モニタリング」を注目点とする評価行動は、全体的に JST より CST の出現率のほうが高かった。日本語話者ではない CST はより自分の評価活動に注意を払いながら、評価を行っていることが推測できる。

#### 4.1.2 下位カテゴリーに基づく分類の結果

本項では、各上位カテゴリー（「自己モニタリング」「内容」「構成」「日本語」）の評価行動を、さらに下位分類し、集計した評価行動について述べる。

まず、「自己モニタリング」について述べる。表 3 の示しているように、「自己モニタリング」の下位カテゴリーにおいて、JST による「再読」行動は 5 割、或は 5 割未満の割合を占めているのに対し、CST による「再読」行動は 8 割以上の割合を占めている。

「再読」の具体的な発話例 (1) (2) を見てみると、JST も CST も作文の中の文、表現、語彙を読み返し、文法の正確さ、語彙選択の適切さ等を吟味していることが分かる。CST は「再読」をより多く行っていることは、4.1.1 の分析で CST が「日本語」により多く注目している結果と一致したものだと言えよう。

##### 発話例<sup>11</sup> (1) JST 2

(読み) ファーストフードが短時間で作れる、あるいは注文してからすぐ食べられる手軽な食品や食事のことだ。

(自一再) ファーストフードが短時間で作れる

(日一修) はですね これはは

##### 発話例 (2) CST 2

(読み) それでも、私は家庭料理が最高だと思っています。朝早く起きたら、料理を作る時間があります。

(自一再) 速く起きたら

(日一修) 起きれば

(日一正) このたらの使い方がおかしい

2. で紹介した評価点数を分析の対象とする Lee (2009) 及び Marefat & Heydari (2016) は母語話者教師より非母語話者教師のほうが文法についての評価を重視し、厳しく評価すると報告している。評価者の評価行動を分析の対象とする本研究でも、そのような傾向が見られた。

次に、「内容」について述べる。表 4 が示しているように、「内容」の下位カテゴリーにおいて、JST の評価行動は「内容の確認」と「サポートの質の判断」に

集中している傾向がうかがえる。それに対し、CST による「内容」に関する評価行動が少なく、本稿で取り上げた 3 名の評価者では一致した傾向が見られなかった。以下の発話例 (3) ~ (6) が示しているように、「内容の確認」では、評価者は直前に読んだ文、或は段落の意味をまとめている。また、「サポートの質の判断」では、評価者はサポートとして提示されている材料の真正性、論理性等について判断を行っている。

##### 発話例 (3) JST 3

(読み) 手軽さ、安さなどの面から、一人で暮らしをはじめる際に、好んで食べられる場合がある。現在では、コンビニエンスストアや外食産業が盛んなため、根強い人気がある。

(内一内) ここがいいところね ファーストフードの  
発話例 (4) CST 3

(読み) アメリカでは、マクドナルドやペプシコなど 11 の主要なメーカーが、12 歳以下の子どもにファーストフードなどの広告をやめることで合意している。

(内一内) この段落は ファーストフードの長所と欠点

##### 発話例 (5) JST 3

(読み) しかし、現代人にとって、家庭料理を作るのは時間がかかることもさることながら、料理を作る現代若い人も少ない。

(内一サ) ほんとう？ これはなんか自分の勝手な思い込みでは

##### 発話例 (6) JST 2

(読み) でも、いつも外食に頼るのは、体にとってあまり良くありません。なぜなら、どうしても油ものや炭水化物が多くなり、野菜の量が不足するからです。

(内一サ) その通りです

JST による「内容の確認」と「サポートの質の判断」の回数と出現率が高いことから、JST はより作文の「内容」に注目し、特に「内容の確認」と「サポートの質の判断」をより頻繁に行っていることが分かる。それに対し、CST による「内容の確認」は少ない。

非母語話者教師は、目標言語の学習経験を作文の書き手と共有しており、書き手に多く見られる誤用、特に母語の移転に起因する誤用をよく理解している。したがって、作文の言葉遣いが分かりにくくても、非母語話者教師が書き手の表現しようとする内容を推測できると言われている (Lee, 2009; Hyland & Anan, 2006)。本研究の CST は中国語母語話者であり、作文

表4 「内容」(下位カテゴリー)に基づく評価行動の集計

|       | 主張 |      | 内容 |      | 精緻化 |      | 明確さ |      | サポート |      | オリジナリティ |      | 余剰 |      | 総数 |
|-------|----|------|----|------|-----|------|-----|------|------|------|---------|------|----|------|----|
|       | 回数 | 出現率  | 回数 | 出現率  | 回数  | 出現率  | 回数  | 出現率  | 回数   | 出現率  | 回数      | 出現率  | 回数 | 出現率  |    |
| JST 1 | 3  | 0.20 | 5  | 0.33 | 0   | 0.00 | 0   | 0.00 | 5    | 0.33 | 0       | 0.00 | 2  | 0.13 | 15 |
| JST 2 | 0  | 0.00 | 2  | 0.18 | 3   | 0.27 | 0   | 0.00 | 4    | 0.36 | 0       | 0.00 | 2  | 0.18 | 11 |
| JST 3 | 3  | 0.10 | 7  | 0.52 | 1   | 0.03 | 3   | 0.10 | 6    | 0.21 | 1       | 0.03 | 0  | 0.00 | 21 |
| JST平均 | 2  | 0.10 | 5  | 0.34 | 1   | 0.10 | 1   | 0.03 | 5    | 0.30 | 0       | 0.01 | 1  | 0.11 | 16 |
| CST 1 | 0  | 0.00 | 1  | 0.25 | 0   | 0.00 | 0   | 0.00 | 3    | 0.75 | 0       | 0.00 | 0  | 0.00 | 4  |
| CST 2 | 0  | 0.00 | 1  | 1.00 | 0   | 0.00 | 0   | 0.00 | 0    | 0.00 | 0       | 0.00 | 0  | 0.00 | 1  |
| CST 3 | 2  | 0.17 | 7  | 0.58 | 0   | 0.00 | 1   | 0.08 | 0    | 0.00 | 2       | 0.17 | 0  | 0.00 | 12 |
| CST平均 | 1  | 0.06 | 3  | 0.61 | 0   | 0.00 | 0   | 0.03 | 1    | 0.25 | 1       | 0.06 | 0  | 0.00 | 6  |

主張=主張の確認 内容=内容の確認 精緻化=主張の精緻化の判断 明確さ=主張の明確さの判断  
サポート=サポートの質の判断 オリジナリティ=内容の独創性の判断 余剰=内容の余剰の判断

の書き手と日本語の学習経験を共有しているため、書き手の表現意図をより簡単に読み取ることが考えられる。その結果、作文の内容について確認することも少ない。それに対し、JSTは頻繁に書き手の意図を確認しながら、作文を読み進んでいくことが想定される。

また、「サポートの質の判断」についての評価の出現率の違いは、JSTとCSTの文化的背景に影響されると考えられる。作文の書き手が主張のサポートとして用いている材料は、書き手と文化的背景を共有しているCSTにとって、このトピックについて論述する際に一般的に考えられるものが多い。そのため、サポートの質に関しては、CSTによる判断が少ない。それに対し、JSTは作文の書き手とは文化的背景が異なっている。サポートとして用いている材料は、JSTにとって予期しなかったものが多いため、サポートの質についての判断も多く行われたと考えられる<sup>12</sup>。

次に、「構成」について述べる。表5が示しているように、「構成」の下位カテゴリーにおいて、数少ないJSTの評価行動は「構成要素の判断」に集中している傾向がうかがえる。CSTの出現数はさらに少ないが、「結束性の判断」に集中している。

発話例(7)～(11)から分かるように、JSTは段落の配分、文章の終わり方等、作文全体に関わるマクロ構成に注目するのに対し、CSTは主に文と文のつながりといった段落内部のミクロ構成の部分を問題視している。CONNOR-LINTON(1995a)も「構成」に関する評価理由については、母語話者教師は作文全体レベルで評価するのに対し、非母語話者教師は段落レベルで評価すると報告している。しかし、本研究で確認できた「構成」を注目する評価行動が少ないため、今後回顧法を取り入れてさらに調査する必要がある。

発話例(7) JST 3

頭でっかちなんだよね ファーストフードの話がたくさん書いてあった なんて家庭料理のほうがいいのかっていう説明がすごく少ないから

発話例(8) JST 2

作文として そんな終わり方でいいのでしょうか

発話例(9) JST 1

短っ こっち多いほうがいい 量が

発話例(10) CST 2

文のつながりがよくないですね 何か理由 原因を表すつながりの言葉を入れたらいいと思いますけど

発話例(11) CST 2

この文の接続は問題がある はっきり表していない これはたぶん 何々のでにしたほうが この二つの部分を(そのまま)繋げてはおかしい

最後に、「日本語」について述べる。表6で示しているように、「日本語」の下位カテゴリーにおいて、JSTもCSTも「日本語の正確さの判断」と「日本語の修正」を多く行っている傾向がうかがえる。さらに、教師歴の違いによる評価行動の変化も見られた。教師歴が長くなるにつれ、JSTもCSTも「日本語の修正」の出現率が減り、また、JSTは教師歴が長くなるにつれ、「日本語の正確さの判断」の出現率が高くなっていくことが分かった。

作文の誤用の修正は、評価者が書き手の表現意図を理解するのに必要であり、作文を評価するにあたって欠かせない行動だと考えられる。一方で、教師歴の長い評価者は、誤用の修正が自動化され、発話思考法では確認しにくいと推測できる。ただし、本稿で取り上げる評価者の数が少ないため、今後評価者を増やして

表5 「構成」(下位カテゴリー)に基づく評価行動の集計

|       | 構成要素の確認 |      | 構成要素の判断 |      | 論の展開 |      | 結束性 |      | 首尾一貫性 |      | 総数 |
|-------|---------|------|---------|------|------|------|-----|------|-------|------|----|
|       | 回数      | 出現率  | 回数      | 出現率  | 回数   | 出現率  | 回数  | 出現率  | 回数    | 出現率  |    |
| JST 1 | 2       | 0.29 | 4       | 0.57 | 0    | 0.00 | 1   | 0.14 | 0     | 0.00 | 7  |
| JST 2 | 0       | 0.00 | 3       | 0.50 | 0    | 0.00 | 2   | 0.33 | 1     | 0.17 | 6  |
| JST 3 | 1       | 0.13 | 3       | 0.38 | 4    | 0.50 | 0   | 0.00 | 0     | 0.00 | 8  |
| JST平均 | 1       | 0.14 | 3       | 0.48 | 1    | 0.17 | 1   | 0.16 | 0     | 0.06 | 7  |
| CST 1 | 0       | 0.00 | 0       | 0.00 | 0    | 0.00 | 1   | 1.00 | 0     | 0.00 | 1  |
| CST 2 | 0       | 0.00 | 0       | 0.00 | 0    | 0.00 | 0   | 0.00 | 0     | 0.00 | 0  |
| CST 3 | 0       | 0.00 | 1       | 0.33 | 0    | 0.00 | 2   | 0.67 | 0     | 0.00 | 3  |
| CST平均 | 0       | 0.00 | 0       | 0.11 | 0    | 0.00 | 1   | 0.56 | 0     | 0.00 | 1  |

構成要素の確認＝段落分け、段落の配分(バランス)、導入、結びについての確認  
 構成要素の判断＝段落分け、段落の配分(バランス)、導入、結びの適切性の判断 論の展開＝意見と理由の配列の判断  
 結束性＝結束性(文と文のつながり)の判断 首尾一貫性＝作文全体を通しての主張の一貫性の判断

表6 「日本語」(下位カテゴリー)に基づく評価行動の集計

|       | 原因 |      | 分かりやすさ |      | 正確さ |      | 修正 |      | 語彙選択 |      | 置き換え |      | 豊富さ |      | 総数 |
|-------|----|------|--------|------|-----|------|----|------|------|------|------|------|-----|------|----|
|       | 回数 | 出現率  | 回数     | 出現率  | 回数  | 出現率  | 回数 | 出現率  | 回数   | 出現率  | 回数   | 出現率  | 回数  | 出現率  |    |
| JST 1 | 2  | 0.05 | 3      | 0.08 | 6   | 0.15 | 27 | 0.69 | 0    | 0.00 | 0    | 0.00 | 1   | 0.03 | 39 |
| JST 2 | 0  | 0.00 | 11     | 0.24 | 14  | 0.31 | 13 | 0.29 | 5    | 0.11 | 0    | 0.00 | 2   | 0.04 | 45 |
| JST 3 | 0  | 0.00 | 4      | 0.22 | 9   | 0.50 | 3  | 0.17 | 1    | 0.06 | 0    | 0.00 | 1   | 0.06 | 18 |
| JST平均 | 1  | 0.02 | 6      | 0.18 | 10  | 0.32 | 14 | 0.38 | 2    | 0.06 | 0    | 0.00 | 1   | 0.04 | 34 |
| CST 1 | 3  | 0.07 | 4      | 0.10 | 9   | 0.22 | 21 | 0.51 | 2    | 0.05 | 1    | 0.02 | 1   | 0.02 | 41 |
| CST 2 | 0  | 0.00 | 2      | 0.07 | 17  | 0.63 | 8  | 0.30 | 0    | 0.00 | 0    | 0.00 | 0   | 0.00 | 27 |
| CST 3 | 0  | 0.00 | 7      | 0.23 | 14  | 0.45 | 2  | 0.06 | 8    | 0.26 | 0    | 0.00 | 0   | 0.00 | 31 |
| CST平均 | 1  | 0.02 | 4      | 0.13 | 13  | 0.43 | 10 | 0.29 | 3    | 0.10 | 0    | 0.01 | 0   | 0.01 | 33 |

原因＝誤用の原因の分析 分かりやすさ＝文の分かりやすさの判断 正確さ＝日本語の正確さの判断 修正＝日本語の修正  
 語彙選択＝語彙選択の判断 置き換え＝語彙の置き換え 豊富さ＝語彙・表現の豊富さの判断

その傾向を確認していきたい。

#### 4.2 評定段階について

ここでは、JSTとCSTが作文を読んだ後、与えられた評価項目について点数をつける際の発話プロトコルの分析の結果について述べる。各評価項目についての両者の評価、及びそう評価した理由を比較した結果、「内容の分かりやすさ」と「主張の一貫性」の二つの評価項目において、JSTとCSTによる評価の違いが見られた。

まず、JSTとCSTによる「内容の分かりやすさ」についての評価を比較する。以下の発話例(12)～(17)が示しているように、「内容の分かりやすさ」を評価

する際、CSTは作文の言語面の問題点を指摘している。それに対し、JSTは、言語面の問題点よりも書き手の意見の不明確について多く言及している。先述したように、CSTは作文の書き手とは文化的背景と日本語の学習経験を共有していることから、文の意味、或は書き手の意見を読み取ることが難しくないと思われる。一方で、言語面の誤りは意味の理解に阻害する重要な要因の一つであり、同時に、その誤りについては、比較的客観的に判断することができると考えられる。CSTは作文の「分かりやすさ」を判断するのにあたって、言語運用を重要な指標として用いることが推測できる。

発話例 (12) JST 1  
 分かりやすいはわかりやすいんだけど 5点 6点はいかないかな 立場が分からない 立場もないかあるか

発話例 (13) JST 3  
 なんか 食生活についての意見っていうところを読み手が探さないといけないつらさがあるから 私はこの内容は分かりにくいと思う 2

発話例 (14) JST 2  
 分かりにくいです なんてだろう 助詞を間違えてるせいでか なんか どう読んでいいのか分からない文章が多いですね

発話例 (15) CST 2  
 あんまり 言葉づかいが間違ってることもありま  
 すし あんまり分かりやすいとは思いません

発話例 (16) CST 3  
 内容は 分かりやすいは分かりやすい しかしこの文章の言語の構造ははっきりしていないので 4点

発話例 (17) CST 2  
 それは ちょっと自分の主張が分かりやすく表現してないので 5点にします

次に、JST と CST による「主張の一貫性」についての評価を比較する。以下の表7が示しているように、4本の作文の「主張の一貫性」については、CST はプラスに評価するものが多いのに対し、JST はマイナス評価するものが多い。特に、作文Ⅲに関しては、JST 全員がマイナスに評価するのに対し、CST 全員がプラスに評価している。

作文Ⅲは、導入の部分でファーストフードの人気を述べ、その次にファーストフードのメリット、デメリットを述べ、さらに、家庭料理のデメリット、メリットを述べる構成になっている。文章量の配分を文字数から見ると、ファーストフードのメリットとデメリット

表7 「主張の一貫性」に対する JST と CST の評価

|       | 作文Ⅰ | 作文Ⅱ | 作文Ⅲ | 作文Ⅳ |
|-------|-----|-----|-----|-----|
| JST 1 | ×   | ×   | ×   | ×   |
| JST 2 | ○   | ○   | ×   | ×   |
| JST 3 | ○   | ×   | ×   | ×   |
| CST 1 | △   | ○   | ○   | ○   |
| CST 2 | ○   | ×   | ○   | ○   |
| CST 3 | ○   | ×   | ○   | ×   |

○=プラス, ×=マイナス, △=中立的

トはほぼ同じ文章量が割り当てられ、家庭料理のデメリットはやや少なく、家庭料理のメリットはやや多い文章量が割り当てられている。

以下の作文Ⅲの「主張の一貫性」について評価する際の発話例 (18) ~ (21) から分かるように、JST は作文Ⅲが内容のバランスが悪く、主張が一貫していないと判断するのに対し、CST はそれを特に問題としていない。JST と CST の作文Ⅲの「主張の一貫性」に対する評価が異なっていることが分かる。JST と CST が持っている「主張の一貫性」の判断基準、つまり、主張の一貫性のある作文とはどのようなものかという認識は異なると考えられる。

発話例 (18) JST 2  
 ないですね ファーストフードのことが なんていうか ファーストフードも これを段落を二つに分けてるんだけど メリット デメリットっていうので 結局どちらのほうの視点に立っているのかが分からないですね

発話例 (19) JST 1  
 ないのもったいないな なんか 結局何が はじめは ファーストフードは簡単だけど 不健康 家庭料理は 健康 時間かかる まあ 単純に違う お互いのメリット デメリットとか 相反するから それを言ってるだけか 最後は 主な違い点だったら なんか もっと こう バランスよく言ってほしいから

発話例 (20) JST 3  
 この人は主張が見えない 最後のところで ただ家庭料理がいいよって言うだけなのかな ないと思う 2

発話例 (21) CST 2  
 一貫性があると思います これは … それはファーストフードがよくないと 家庭料理がその良さ 家庭料理のその良さについて 書かれているものですね まあ 主張に一貫性がある 6点にします

Kobayashi & Rinnert (1996) は、日本語のレトリックパターンで書かれた英語の作文の導入部分について、英語母語話者教師より日本語母語話者教師のほうが高く評価することを報告している。作文の構成の良し悪しに関する評価基準そのものは、評価者の母語のレトリックパターンに影響され、母語の異なる評価者の判断が異なる可能性がうかがえる。

## 5. まとめと今後の課題

本研究の目的は中国人日本語学習者の書いた作文に対する JST と CST の評価の違いを明らかにすることであった。発話思考法を用いて、評価者の実際の評価行動を収集し、分析した結果、JST よりも CST のほうが「日本語」に注目し、再読を多く行いながら、評価対象となる作文を読んでいることが分かった。それに対し、CST より JST のほうが「内容」に注目し、「内容の確認」と「サポートの質の判断」多く行っていることが分かった。また、JST と違って、CST は作文の言語面の適切さを用いて作文の分かりやすさを判断していることが明らかになった。さらに、母語話者教師と非母語話者教師の下した評価の傾向（プラスかマイナスか）にも違いが見られた。本研究では、作文の「主張の一貫性」についての JST と CST の判断が異なることが分かった。

本研究の結果から、これからは CST が学習者の作文を評価する際、日本語の正確さだけでなく、内容及び文章レベルの構成について多く注目する必要があることが示唆された。また、日本語作文の「主張の一貫性」に関する判断基準を明確にし、JST と CST の間で共有させるべきだろう。

今後は、作文を読む際的评价行動は点数をつける際の判断にどう影響するか、「読む」段階のプロトコルと判定段階のプロトコルを関連付けて分析していきたい。さらに、JST と CST の評価になぜ違いが生じたかについて、フォローアップインタビューを分析し、その原因を明らかにしたい。

また、本稿の分析で明らかになったように、評価者の作文指導経験の長さ、評価者個人の作文評価に対する考え方も作文評価に影響すると考えられるため、そういった要素を考慮に入れ、JST と CST による評価を比較していきたい。

### 【注】

- 1) メカニックとは、英語の綴り、句読法に関する項目である。
- 2) 発話思考法とは、ある作業をしている間、思ったこと、考えたことを可能な限り口に出してもらい、それを録音して分析するという手法である（海保・原田, 1993）。
- 3) プロトコルとは、発話思考法を用いて産出した発話である。
- 4) 本研究は143の項目を一枚ずつカードに書き、それらのカードを近い感じのするもの同士で分類をし

てもらった。

- 5) 20項目の内訳は内容に関する8項目、構成に関する6項目、日本語に関する6項目である。また、項目の内容は、作文に対する評価者の行動を調べる Cumming et al (2002) で明らかになった評価者が注目する作文の要素と多く共通している。
- 6) 日本語教育の現場では、日本語母語話者教師が作文の授業を担当する傾向がある。本研究の協力者の場合、JST は平均日本語教師歴が CST より短いが、平均作文指導歴が CST より長い。また、今回の調査では、CST に「作文を取り入れる日本語の総合クラスを担当する年数」についても記述してもらった。協力者の CST 2 は、「作文授業の担当する年数」は0であるが、「作文を取り入れる日本語の総合クラスを担当する年数」は4年であることから、作文指導経験のある教師とする。
- 7) CST の「作文を取り入れる日本語の総合クラスを担当する年数」を作文指導歴のカッコ内で示す。
- 8) Cumming et al (2002) は評価の注目点の違いによって、評価行動をまず「自己モニタリング」「レトリックとアイデア」「言語」の3つの上位カテゴリーに分け、さらにカテゴリーごとに下位分類を行い、27の下位カテゴリーに分けている。Cumming et al (2002) で提案された作文に対する評価行動の分類システムは後に多くの研究で応用され、改善されている（Barkaoui, 2010; Li & He, 2015等）。
- 9) 本論文で示している評価行動の回数の平均値は小数点以下を四捨五入して整数で表示し、出現率及びその平均値は小数第2位以下を四捨五入して表示する。
- 10) 評価者に作文を評価してもらった後に、そう評価した理由を述べてもらうような回顧法を用いる研究では、作文の構成の良し悪しは評価理由として多く挙げられている（Shi, 2001; CONNOR - LINTON, 1995a）。一方で、本研究と同じ発話思考法を用いる Li & He (2015) でも、構成に関する作文の要素に注目する評価行動は低い出現率を示している。作文の構成について判断するには、作文全体の構造を確かめることや、前後の複数の段落の関係を考える等、スキミングを行うことが多いと思われる。そのような作業は、すべて読んでいる、考えていることを声に出して述べる発話思考法では生じにくいと考えられ、本研究の「構成」を注目点とする評価行動の出現率が低い結果につながったと考えられる。
- 11) 発話例では、発話が分類されたカテゴリーを、(上位カテゴリー—下位カテゴリー) の形式で示す。なお、各カテゴリーはそれぞれのカテゴリー名の一文

字目で表す。また、読まれる作文の原文の部分を下線で示し、具体的な表現は「囲み線」で示す。また、JSTのプロトコルは中国語で発話されるものもあり、そのよう場合、その日本語訳で提示する。

12) Hamp-Lyons & Zhang (2001) は書き手のイデオロギー或は社会的／政治的スタンスが表れている作文に対し、評価者と書き手の文化的背景の関係によって評価が異なると報告しており、本研究の結果からサポートの質を判断する際でも類似した傾向が得られたといえよう。

## 【参考文献】

- 海保博之・原田悦子 (1993). 『プロトコル分析入門－発話データから何を讀むか－』新曜社
- 国際交流基金 (2012). 『海外の日本語教育の現状－海外日本語教育機関調査－』国際交流基金日本語国際センター編
- 田中真理・阿部新 (2014). 『Good Writing へのサポート－読み手と構成を意識した日本語ライティング－』くろしお出版
- 田中真理・坪根由香里・初鹿野阿れ (1998). 「第二言語としての日本語における作文評価基準－日本語教師と一般日本人の比較－」『日本語教育』96, 1-12.
- 崔文姫 (2013). 「日本語学習者の発話に対する日本語母語話者の評価－日本語教師と非日本語教師の因果モデルを中心に－」『国立国語研究所論集』5, 1-26.
- 野原ゆかり (2008). 「発話の「分かりやすさ」を判断する要因－一般日本人と母語話者日本語教師の比較を通して－」『人間文化創成科学論叢』11, 165-174.
- 渡部倫子 (2005). 「日本語学習者の発話に対する日本語母語話者の評価－共分散構造分析による評価基準の解明－」『日本語教育』125, 67-75.
- Barkaoui, K. (2010). Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly*, 7(1), 54-74.
- CONNOR - LINTON, J. E. F. F. (1995a). Crosscultural comparison of writing standards: American ESL and Japanese EFL. *World Englishes*, 14(1), 99-115.
- CONNOR - LINTON, J. E. F. F. (1995b). Looking behind the curtain: what do L2 composition ratings really mean. *TESOL Quarterly*, 29(4), 762-765.
- Cumming, A. H., Kantor, R., & Powers, D. E. (2001). *Scoring TOEFL essays and TOEFL 2000 prototype writing tasks: An investigation into raters' decision making and development of a preliminary analytic framework*. Educational Testing Service.
- Hamp-Lyons, L., & Zhang, B. W. (2001). World Englishes: Issues in and from academic writing assessment. *Research perspectives on English for academic purposes*, 101-116.
- Hyland, K., & Anan, E. (2006). Teachers' perceptions of error: The effects of first language and experience. *System*, 34(4), 509-519.
- Kobayashi, H., & Rinnert, C. (1996). Factors affecting composition evaluation in an EFL context: Cultural rhetorical pattern and readers' background. *Language learning*, 46(3), 397-433.
- Lee, H. K. (2009). Native and nonnative rater behavior in grading Korean students' English essays. *Asia Pacific Education Review*, 10(3), 387-397.
- Li, H., & He, L. (2015). A comparison of EFL raters' essay-rating processes across two types of rating scales. *Language Assessment Quarterly*, 12(2), 178-212.
- Lumley, T. (2005). *Assessing second language writing: The rater's perspective*. P. Lang.
- Marefat, F., & Heydari, M. (2016). Native and Iranian teachers' perceptions and evaluation of Iranian students' English essays. *Assessing Writing*, 27, 24-36.
- Shaw, S. D., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing* (Vol. 26). Cambridge University Press.
- Shi, L. (2001). Native-and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing*, 18(3), 303-325.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.