



大学教員によるライティング評価の観点を探る

伊集院, 郁子

小森, 和子

奥切, 恵

(Citation)

Learner Corpus Studies in Asia and the World, 3:159-176

(Issue Date)

2018-03-12

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81010125>

(URL)

<https://hdl.handle.net/20.500.14094/81010125>



大学教員によるライティング評価の観点を探る

伊集院 郁子(東京外国語大学)／小森 和子(明治大学)／奥切 恵(聖心女子大学)
 ijuin@tufs.ac.jp／komokazu@meiji.ac.jp／okugiri@u-sacred-heart.ac.jp

An Analysis of Raters' Comments on Japanese Opinion Essays

IJUIN Ikuko (Tokyo University of Foreign Studies)/ KOMORI Kazuko (Meiji University)/ OKUGIRI Megumi (University of the Sacred Heart)

概要

本研究では、同条件で書かれた 30 編の日本語作文のうち、同等の評定結果となった母語話者と学習者 3 組計 6 編に対する評価コメントをデータとし、大学教員による作文評価の観点を探索的に抽出したうえで、その特徴を分析した。評定値の高低(良い作文か悪い作文か)を軸に、書き手の属性(母語話者か学習者か)、評価者の属性(日本語教員か専門教員か)による相違にも着目して評価コメントを比較した結果、評定値の高低、書き手、評価者の属性に関わらず、評価の観点は、「内容」>「言語」>「構成」>「形式」の順に多いという点では共通していることがわかった。一方で、評定値の高低で比べると、上位群に比して中位群では「言語」、下位群では「言語」と「形式」の割合が高くなっていること、「構成」については、母語話者に比して学習者の方がプラス評価されていること、日本語教員の方が専門教員より「言語」と「構成」を重視していることなど、いくつかの相違点も観察され、さまざまな要因によって評価の観点の重みが異なっている可能性も見出された。

キーワード

アカデミック・ライティング、評価、作文コーパス、日本語学習者、初年次教育

1. はじめに

高校から大学への円滑な移行を目的とした初年次教育は、現在 97%の大学で実施されており、中でも「文章作法」に取り組む大学は89%に上っている(文部科学省高等教育局大学振興課大学改革推進室 2017)。日本の大学への進学を目指す留学生に対しても、「日本留学試験」が「記述」試験を課しており、大学入学時の大学生の「書く力」は、入学後の学業に重要不可欠なスキルとして捉えられている。一方、日本の大学は現在、少子化と教育の国際化に伴い、留学生の占める比率が相対的に高くなってきている。そのため、今後、日本人学生と留学生とが同じ科目を履修し、同じ課題でレポートを執筆して評価される機会も少しずつ増えていくと予想されるが、日本人学生と留学生が混在した授業において、大学教員が両者の執筆したものをどのような観点で評価すべきか、明確な示唆を与える研究成果は見当たらない。

良い作文と悪い作文、日本語母語話者によって執筆された作文と学習言語としての日本語を駆使して書かれた作文、日本語教育を専門とする教員が評価した作文と言語教育以外の専門の教員が評価した作文など、さまざまなタイプの作文・作文評価者が想定される中で、評価が一致しや

すい観点と一致しにくい観点とでは、それぞれどのような特徴があるのだろうか。「書く力」の向上が大学の取り組むべき重要な課題となっている現在、大学教員による文章評価の実態を探ることは、留学生を対象とする日本語教育、母語話者を対象とする文章表現教育のみならず、教員を対象とする評価トレーニングの面においても意義があると思われる。本研究では、評価コメントの記述的分析を通し、作文の良し悪しを軸に評価の観点の異同を捉えることによって、大学教員の文章評価の実態を探ることを目的とする。

2. 先行研究

作文評価の際、評価の観点が一致しないことは、複数の研究で明らかにされている。例えば、田中・長阪(2009)は、8名の日本語教師が2種各26編の小論文の評価を行った結果を分析し、評価の内の一貫性はあったが、ずれの大きな小論文も認められたことから、人間によるパフォーマンス評価の限界に言及している。また、田中・坪根(2011)では、日本語教師10名のプロトコル分析と事後アンケート調査の分析から、総合的評価による順位付けの際の主要要素(「課題の達成」、「主張の明確さ」、「内容のオリジナリティ」、「客観的で広い視野からのサポート」、「構成」、「談話展開のテクニック」、「表現力の豊かさ」など)のうち、どの観点が優先されているかについては、日本語教師間に共通の認識が見られなかったと述べている。一方で、作文評価を観点別に捉えると、言語能力的側面については評価が比較的一致しやすいが、内容や構成面に関する評価は一致が難しいとの指摘も見られる(田中他, 1998; 田中・坪根, 2011 など)。

このように、評価者によって作文評価の基本的観点が異なることや、評価の観点によって評価者間の一致しやすさが異なることが既に指摘されているものの、その実態はまだ明らかになっていない。本研究では、評価者が多様な書き手による複数の作文をどのように評価しているかについて、評定結果と評価コメントを照らし合わせて実証的に分析する。帰納的に評価の観点を掘り起こすことによって、書き手の能力や属性の違い、評価者の属性の違いが評価の観点にどのように影響しているかを考察する。

3. リサーチクエスション

上述の研究目的に照らし、本研究では以下の2つの研究課題を設定した。

RQ1 大学教員44名による評価コメントから、どのような評価の観点が抽出できるか。

RQ2 作文の良し悪し(評点の高低)は、評価の観点到に影響を与えるか。

4. 研究の手法

本研究は、伊集院ほか(2016)の予備調査の結果を踏まえ、①一般公開されている作文コーパスなどから30編の意見文を評価対象データとして抽出、②抽出した意見文30編に対し、大学教員44名に5段階での評定(以下、評定値)、および評価を決定づけた「良い点」と「悪い点」の記述(以下、評価コメント)を依頼、③評定値の分析結果に基づき、30編を上位群、中位群、下位群に分け、各群から評定値が最も近い日本語母語話者および日本語学習者をペアにし、3ペア計6

編を抽出, ④全 6 編に付された 44 名の評価コメントの集約, コード化, および分析, という手順で行った。以下に, 各データに関する詳細を述べる。

4.1 評価対象の作文コーパス

大学教員が評価した日本語作文は, 一般公開されている「日本・韓国・台湾の大学生による日本語意見文データベース」(<http://www.tufts.ac.jp/ts/personal/ijuin/terms.html>)と“The Corpus of Multilingual Opinion Essays by College Students”(<http://www.u-sacred-heart.ac.jp/okugiri/links/moecs/moecs.html>)から抽出された 25 編の日本語作文のほか, 作文の多様性を確保するために新規に収集した 5 編の計 30 編(日本語母語話者作文 15 編, 中国語母語話者作文, 韓国語母語話者作文, 英語母語話者作文 5 編ずつ)である。上記の 2 つのコーパスデータおよび新規に収集したデータは, いずれも以下の課題文にしたがって, 辞書などを使わずに 800 字程度で原稿用紙に 1 時間以内で執筆されたものである。

【作文の課題文】

今, 世界中で, インターネットが自由に使えるようになりました。ある人は「インターネットでニュースを見ることができるから, もう新聞や雑誌はいらぬ」と言います。一方, 「これからも, 新聞や雑誌は必要だ」という人もいます。あなたはどのように思いますか。あなたの意見を書いてください。

なお, 執筆者は全員が国内外の高等教育機関に所属する学生で, 学習者作文は, 日本語能力試験 2 級(学習時間 600 時間相当)以上または N2(2010 年以降の新試験の名称で, 旧試験の 2 級とはほぼ同じレベル)以上という条件で募集し, 応募してきた学習者によって執筆されたものである。中国語母語話者は台湾, 韓国語母語話者は韓国の大学で日本語を専門に学ぶ学生であり, 英語母語話者は豪州の大学で日本語を学んでいる学生で, 専門は多岐にわたる。実際の日本語習得レベルを知る目安として, 筑波大学留学生センターで開発された日本語能力簡易試験 (SPOT ver.2) も実施し, 得点についてもコーパスで公開している。

4.2 評価者と評価方法

評価者は, 日本の大学で 3 年以上の教育歴のある日本語教育の教員 22 名および日本語教育以外の人文社会系(言語学, 社会学, 教育学など)の専門の教員 22 名, 計 44 名である。ここでは, 日本の高等教育機関で留学生への日本語教育の経験が 3 年以上ある大学教員を日本語教員, 留学生に語学として「日本語」を教えた経験がない大学教員(大学での教育歴 3 年以上)を専門教員と呼ぶ。この定義に当てはまらない評価者は, 事前に分析データから除いてある。

依頼に際しては, 大学 1 年生のアカデミック・ライティングとして自身の評価基準で総合的に 5 段階(「非常によい」「よい」「ふつう」「あまりよくない」「悪い」)で判断するよう依頼した。本研究の目的の 1 つとして, 評価コメントから評価の観点を掘り起こすことが挙げられるため, ルーブリックなどに基づく分析的評価ではなく, 包括的評価がふさわしいと判断した。5 段階で評定値を付与した後に

「評価を決定付けたポイント」(良い点・悪い点)を記述すること、評価を決定付けたポイントが良い点または悪い点の一方しかない場合は、一方に「特になし」と記述すること、半日(5~6 時間程度)の時間を確保して一気に 30 編の評価を行うことも依頼した。

4.3 本研究の分析対象データ

4.2 で収集したデータのうち、本分析に用いるのは、「悪い」から「非常に良い」までの 5 段階で付与された評定値を間隔尺度と見なし、1~5 点に換算した値の 44 名分(すなわち、5 点×44 名=220 点満点)を各作文の得点として計算し、その得点に基づいて分類、抽出した上位群、中位群、下位群の各 2 編ずつに対して付された評価コメントである。

44 編の作文の得点は、最低が 79 点(220 点満点における百分率得点は 35.91 点、以下同)、最高が 202 点(91.82 点)で、平均が 138.13 点(62.79 点)、標準偏差が 37.30 点(16.95 点)であった。群分けは得点の高い順から 10 編ずつとし、上位群は 202 点(91.82 点)から 166 点(75.45 点)まで、中位群は 148 点(67.27 点)から 119 点(54.09 点)まで、下位群は 117 点(53.18 点)から 79 点(35.91 点)までとした。なお、この 3 群の平均値の差は統計的に有意であった [$F(2,27)=112.88, p<.01$]。また、多重比較の結果、いずれの群間においてもその差は 1%水準で有意であった。そのうち、本分析の対象とした 3 組計 6 編の作文とその得点は、表 1 のとおりである。表中の ID は、「NS」が母語話者、「NNS」が学習者を表し、さらに「上」が上位群、「中」が中位群、「下」が下位群を表す。以下、本文中でも、執筆者は NS 上、NNS 上といった ID で示す。

表 1 分析対象作文の内訳

	ID (NS)	得点 (百分率得点)	ID (NNS)	得点 (百分率得点)
上位群	NS 上	182 点 (82.73 点)	NNS 上	176 点 (80.00 点)
中位群	NS 中	123 点 (55.91 点)	NNS 中	125 点 (56.82 点)
下位群	NS 下	109 点 (49.55 点)	NNS 下	106 点 (48.18 点)

なお、NS 上は「日本・韓国・台湾の大学生による日本語意見文データベース」に格納されている JP004、NNS 上は TM032、NNS 中は TM024、NNS 下は TM002 と同一であり、ウェブ上で作文の原文や執筆者の年齢、出身地等を確認することが可能である。日本語能力簡易試験 (SPOT ver.2) の得点は、65 点満点中 NNS 上が 64 点、NNS 中が 62 点で日本語能力試験 N1 相当と判断できるレベルであり、NNS 下が 29 点である。NNS 下は SPOT の得点が低いが、2 年 5 か月の日本語学習歴があり、日本語能力試験 2 級は取得している学生である。(NS 中と NS 下は新規に収集したデータであり、未だ一般公開されていないデータである。)

5. 結果と考察

5.1 RQ1 大学教員 44 名による評価コメントから、どのような評価の観点抽出できるか。

表 1 の 6 編の作文に対して与えられた評価コメントから評価の観点を抽出するため、M-GTA (修

正版グラウンデッド・セオリー・アプローチ(木下, 1999)の手法を用いて、意味的なまとまりを分析単位として評価コメントの切片化を行った。ここで用いる「切片化」という用語は、評価コメントを一語一句ばらばらにしていくことではなく、文脈を理解したうえで、意味をもったまとまりにコメントを分けていく作業を意味する。以下に示した「切片化基準」の中で、「例」で示すのは原文のままの評価コメント、「→」で示すのは意味的なまとまり(分析単位)に切片化した後の評価コメントである。

【切片化基準】

- ①原則として、「節」を1単位として切片化する。なお、コーディングの正確性を期すために、切片化の際に、節の主語や補語などを補うこともある。
例:「主張がはっきりしており、論拠も客観的、論理的である。」
→「主張がはっきりしている。」「論拠が客観的である。」「論拠が論理的である。」
- ②評価者の着目点が複数含まれる場合は、着目点ごとに切片化する。
例:「文法も語彙も誤りがない。」
→「文法は誤りがない。」「語彙は誤りがない。」
- ③複数の節からなる場合でも、評価者の着目点が1つのみの場合は、切片化せず、全体で1単位とする。
例:「新聞で得られる情報はネットで得られると述べているが、ネットは新聞・雑誌の情報に頼っていることが多いため、理由として納得しがたい。」
→「新聞で得られる情報はネットで得られると述べているが、ネットは新聞・雑誌の情報に頼っていることが多いため、理由として納得しがたい。」
- ④複数の節からなる場合でも、分かりやすく言い換えたり具体例を示したりするために、評価者が同じことに繰り返し言及していることが明らかな場合は、主要でない節は削除し、全体で1単位とする。
例:「文末表現が「思う・思っている」ばかりである点。意味的に、2段落目の最後の文は、「思われる・考えられる」にしないといけないのではないか？」
→「文末表現が「思う・思っている」ばかりである点。意味的に、2段落目の最後の文は、「思われる・考えられる」にしないといけないのではないか？」
- ⑤複数の節からなる場合でも、原因・理由が、結果の詳述や具体例となっている場合は、主要でない節は削除し、全体で1単位とする。
例:「利点の1つ1つはインターネット記事にも当てはまることなので、効果的な論拠になっていないように見受けられた。」
→「利点の1つ1つはインターネット記事にも当てはまることなので、効果的な論拠になっていないように見受けられた。」

なお、評価コメントに評定結果に関する言及が含まれる場合(例:「(評点は)2か3か迷う」や、「良い点」に否定的なコメントまたは「悪い点」に好意的なコメントが含まれる場合は、それらの記述

は無視することとした。

続いて、切片化後の評価コメント 1 つにつき、評価の観点のコードを 1 つずつ付与する作業を行った。手順としては、本研究対象データ 6 編を含む全 10 編分の評価コメントを切片化し、切片化後のコメントに対し意味を解釈してコードを生成し、同一のコードで解釈できない場合は新たにコードを生成していく、という方法をとった。作業過程の中で、複数のコードを統括する上位概念が生成された場合はコードのグループ化を行った。作文の原文に当たらないとコメントの文意が明らかにならない場合には、作文を読んだうえで評価者の意図を解釈し、コードを付した。

切片化基準およびコーディング表の原案は、研究代表者がこれらの作業を行う中で作成し、その後、3 名の言語教育に携わる研究者が原案にしたがって同作業を行った。作業結果を持ち寄り、コードが一致しない箇所を中心に計 4 名で検討作業を重ね、基準の見直しおよび精緻化作業を繰り返す作業を経て、最終的に全評価コメントを説明するコード基準表を完成させた。評価コメントから抽出されたコードは、次ページの表 2 に示すとおり、「言語(表現力、文体、文法、語彙、接続詞、表記、その他)」、「内容(論旨、立場、根拠、状況把握、譲歩、反論、問題提起)」、「構成(段落間、段落内、バランス、立場)」、「形式(文字数、原稿用紙、書字)」、「印象」、「題」、「分類不能」、「なし」の 25 項目、8 カテゴリーに集約された。

なお、これらのカテゴリーに含まれるコメントの説明を簡潔にまとめると、「内容」は、論旨や根拠などの内容に関わるコメントの上位概念で、その下位分類の「内容(論旨)」は論旨や論点の明確さや一貫性に関するコメント、「内容(立場)」は意見文の課題に対する書き手の立場や主張に関わるコメント、「内容(根拠)」は立場や主張を支える根拠に関するコメント、「内容(状況把握)」はテーマの背景の情報に関するコメント、「内容(譲歩)」は執筆者と相反する立場に有利な情報や執筆者の立場に不利な情報の記述に関するコメント、「内容(反論)」は譲歩に対する反論や反駁に関するコメント、「内容(問題提起)」は問題提起、問いの立て方に関するコメントである。

「言語」は、日本語の使用に関するコメントの上位概念で、その下位概念として、「言語(表現力)」は表現の豊かさ、洗練さへの言及や、読みやすさのための表現上の工夫に関するコメント、「言語(レジスター)」は意見文の文体やフォーマリティーに関するコメント、「言語(文法)」、「言語(語彙)」、「言語(接続詞)」、「言語(表記)」は、それぞれ、文法、語彙、接続詞、表記に関するコメント、「言語(その他)」はこれらに分類されない日本語の誤用や自然さに関するコメントである。

「構成」は、文章全体の構造や段落間の関係、段落内のまとめなどに関するコメントの上位概念で、その下位概念の「構成(段落間)」は序論、本論、結論などの全体的な構成や論の展開などに関するコメント、「構成(段落内)」は 1 つの段落に 1 つの概念が記されているか否かや段落内のまとめに関するコメント、「構成(バランス)」は分量的なバランスに関するコメント、「構成(立場)」は書き手の立場や主張の出現位置に関するコメントである。

「形式」は、「形式(文字数)」、「形式(原稿用紙)」、「形式(書字)」があり、それぞれ文字数の過不足、原稿用紙の使い方、書字の読みやすさに関するものである。

「印象」は、特定の評価の観点に言及せずに、「面白い」、「丁寧でない」、「完璧」など、漠然とした全体的な印象を記したコメントで、「題」は題の適切性などに関するコメントである。このほかに、コ

メント欄が空欄,あるいは、「特になし」と書かれていたものは「なし」と分類した。最後に,これらのいずれにも分類することができないコメント,同じ観点のコメントが複数見られなかったものは「分類不能」とした。

表2 コード基準表(抜粋)

コードの名称	「良い点」の記述例		「悪い点」の記述例		
内容	論旨	要点が明確で、一貫している	論点が整理されている	説明不足	重複情報がある
	立場	主張が明確	設問に簡潔に答えている	主張が不明	作文課題に対して明確な答えがない
	根拠	根拠の理由が述べられている	メリットを明確に述べている	根拠が弱い	理由が外面的で浅い
	伏況把握	現状を踏まえている	まず最初に論題の置かれた伏況を確認している		
	譲歩	他人の意見に譲歩している	反対意見を想定している	新聞や雑誌を読む上での欠点に言及していない ※「新聞必要論」の意見文に対するコメントの場合	ネットの利便性に触れていない
	反論	論じ返している	反対意見を踏まえた上でそれに対する反論をしている	反駁が反駁になっていない	反論部分の二番目と三番目の論拠が薄弱である
	問題提起	独自の問いを立てている	問題提起はわかりやすい	問題提起の文がない	先に問いかけを述べていないのは論説文としては弱い
言語	表現力	表現が豊かである	文にメリハリがあり単語でない	同じ表現の多用	意見の文末表現としては単語
	レジスター	文体が統一されている	丁寧な文体で書かれている	話し言葉	十分にかたい表現になっていない
	文法	文法的破たんがない	複雑な構文を使っている	文法ミスが見受けられる	文法の誤りが多い
	語彙	語彙が豊富	「〇〇性」で語彙を統一	語彙の選択に問題がある	語彙に不正確なところがある
	接続詞	まず、次に、最後になど	「第一に」「第二に」等接続詞をうまく利用している	接続詞がない	接続表現が足りない
	表記	漢字が使ふべきところできちんと使われている	表記ミスがない	表記に誤りがある	漢字の間違ひ
	その他	日本語のまちがひが少ない	誤りがほとんどない	3カ所ほど間違ひがある	不自然な表現が目立つ
構成	段落間	文章全体の構造が明確	序論・本論・結論のまとめ方が簡易でわかりやすい	全体の構造を考慮してから書き出していない	まとめで新しいことを述べている
	段落内	各パラグラフで、1つの話題を詳しく説明している	1つの段落に大体1点の内容が書かれている	1段落—1主文が守られていない	ほとんど重要な部分が段落の最後にかけている
	バランス	構成のバランスがよい		段落の分け方(構成)のバランスがよいかない	結論に関わる記述が少なく、逆の論点に多くを費やしている
	立場	冒頭で立場を明らかにしている	最初に意見が述べられている	最後まで読まないと主張が分らない	前置きが長く、著者の立場をなかなか明らかにしない
形式	文字数	文字数が守られている	字数が適当である	字数が少ない	短い
	原稿用紙			句読点が行頭にくるなど、原稿用紙の正しい使い方に沿っていない	基本的な原稿用紙の使用法がわかっているかない
	書字	字がきれい	読み易い字である	字に癖があつてきつめて読みにくい	文字が雑である
印象	ほほ完璧	感性は面白い	読みにくい	丁寧でない印象がある	
題	内容と関連性がある題	題名がおもしろい	タイトルの意図が不明	題名との不一致	
分類不能	伝えようとする意欲は買いたい	推察がなされている	読解力が低い学生		
なし	特になし		特になし		

分類に最も困難を伴ったのは、「内容」と「構成」が密接に関連している場合であるが、本研究では、「論点」の「一貫性」、「明確さ」、「正確さ」、「説得力」などのロジックを焦点としたコメントは「内容」に、「段落」や「構造」に着目しているコメント、および「展開」、「順序立て」、「話の流れ」を焦点としたコメントは「構成」に分類した。

このような基準で、6編に対する評価コメント(総文字数 21,212 文字)の切片化およびコーディングを行った結果、抽出された総コード数、すなわち切片化後のコメント数は1,136(表3)、各コードの出現数は表4のとおりとなった。表の中で、評価コメントの「良い点」は「+」、「悪い点」は「-」で示す。

表3 評価コメント(原文)の文字数および切片化後のコメント数

作文ID	上位群						中位群						下位群						合計
	NS上		小計	NNS上		小計	NS中		小計	NNS中		小計	NS下		小計	NNS下		小計	
	+	-		+	-		+	-		+	-		+	-		+	-		
文字数	1,739	1,877	3,616	1,969	1,609	3,578	1,188	2,408	3,596	1,290	2,135	3,425	957	2,496	3,453	876	2,668	3,544	21,212
コード数	116	69	185	114	67	181	73	122	195	78	110	188	61	137	198	59	130	189	1,136

表4 評価コメント出現総数詳細

コード分類	上位群				中位群				下位群				合計	
	NS上		NNS上		NS中		NNS中		NS下		NNS下			
	+	-	+	-	+	-	+	-	+	-	+	-		
内容	論旨	24	13	22	4	6	13	12	14	0	22	6	15	151
	立場	12	1	11	1	6	15	8	2	14	5	9	3	87
	構換	22	18	27	12	17	14	17	27	19	22	10	31	236
	状況把握	0	0	1	0	2	0	1	0	0	0	0	0	4
	譲歩	2	3	1	3	0	0	0	2	0	0	2	2	15
	反論	2	1	0	0	0	0	0	0	0	0	2	0	5
	問題提起	0	0	0	0	1	0	1	0	0	1	0	0	3
	小計	62	36	62	20	32	42	39	45	33	50	29	51	501
言語	表現力	3	3	4	14	4	10	4	8	0	7	1	6	64
	レジスター	2	2	2	3	1	35	1	2	0	8	0	14	70
	文法	2	0	1	4	0	2	0	5	1	5	0	12	32
	語彙	2	0	1	5	0	3	2	17	1	5	2	13	51
	接続詞	6	0	2	0	1	0	2	0	0	5	0	0	16
	表記	1	0	2	0	1	1	1	5	0	6	0	3	20
	その他	6	0	2	7	2	6	1	13	5	5	1	14	62
	小計	22	5	14	33	9	57	11	50	7	41	4	62	315
構成	段落間	17	6	24	0	12	4	14	1	1	14	9	6	108
	段落内	6	2	2	0	2	3	2	1	2	5	1	0	26
	バランス	0	0	1	0	0	1	0	1	0	2	0	0	5
	立場	3	0	3	0	1	5	1	1	1	4	3	0	22
	小計	26	8	30	0	15	13	17	4	4	25	13	6	161
形式	文字数	1	0	2	0	1	0	1	0	1	16	1	7	30
	原稿用紙	0	0	0	7	0	5	0	1	0	0	0	0	13
	書字	1	0	1	0	0	0	0	1	0	0	0	0	3
	小計	2	0	3	7	1	5	1	2	1	16	1	7	46
	印象	2	0	4	0	3	2	4	5	2	2	1	1	26
	題	2	8	0	0	0	1	2	4	0	2	0	2	21
	分類不能	0	0	0	0	0	0	0	0	0	1	1	0	2
なし	0	12	1	7	13	2	4	0	14	0	10	1	64	
合計	116	69	114	67	73	122	78	110	61	137	59	130	1136	

NS と NNS, さらに「良い点(+)」と「悪い点(-)」のすべてのコメント数を合算すると, 上位群, 中位群, 下位群のいずれにおいても, コメント量は「内容」>「言語」>「構成」>「形式」の順であり, 評価の際に内容的側面が最も重視されていることがわかった。また, 下位分類(合計)を見ると, 「内容(根拠)」>「内容(論旨)」>「構成(段落間)」>「内容(立場)」の順に多く, 意見文においては適切な「根拠」によって「立場」(主張)が明確に伝わること, 全体の「論旨」が「構成」(展開)に支えられて理解しやすいことが重視されていると考えられる。

5.2 RQ2 作文の良し悪し(評点の高低)は, 評価の観点に影響を与えるか。

ここからは, コメント数上位 4 種の「内容」, 「言語」, 「構成」, 「形式」に絞って, その特徴を詳細に見ていくこととする。上述のとおり, 上位群, 中位群, 下位群のいずれにおいても, コメント量は「内容」>「言語」>「構成」>「形式」の順であるが, 表 5 を見ると, 上位群では「内容」が 49.18%, 「構成」を合わせると 66.67%を占めるのに対し, 中位群と下位群ではその比率が下がり, その分, 中位群では「言語」, 下位群では「言語」と「形式」に比重が置かれており, 作文の良し悪しによって, 評価の観点の比重が異なっていることがわかる。

表 5 群別 上位 4 種のコメント数

	上位群	中位群	下位群
内容	180(49.18)	158(41.25)	163(42.12)
言語	74(20.22)	127(33.16)	114(29.46)
構成	64(17.49)	49(12.79)	48(12.40)
形式	12(3.28)	9(2.35)	25(6.46)
合計	330(90.16)	342(89.56)	350(90.44)

注:()は各群の総コメント数に占める比率

さらに, 評価コメントのうち, 「良い点」として言及された観点と「悪い点」として言及された観点に分けてみると, 「内容」, 「言語」, 「構成」, 「形式」のいずれにおいても, 上位群と下位群とでは, 「良い点」は 0.30 倍から 0.50 倍に減り, 「悪い点」は 1.80 倍から 3.88 倍に増えている(表 6)。中位群は, 「良い点」では「形式」が下位群と同数, 「悪い点」では「形式」が上位群と同数, 「言語」が下位群を若干上回る数となっているが, 中位群の「悪い点」で「言語」のコメント数が多くなっているのは, NS 中へのコメントで, 「レジスター」に関する言及が多く出現しているためである。

表 6 群別, 観点(良い点・悪い点)別 上位 4 種のコメント数

	上位群		中位群		下位群	
	+	-	+	-	+	-
内容	124	56	71	87	62	101
言語	36	38	20	107	11	103
構成	56	8	32	17	17	31
形式	5	7	2	7	2	23
合計	221	109	125	218	92	258

内訳の詳細(表 7)をみると、「内容」は「根拠」、「論旨」、「立場」、「構成」は「段落間」、「段落内」、「立場」の順で、「良い点」でも「悪い点」でも着目されているのに対し、「言語」は、「良い点」では「その他」、「表現力」、「接続詞」、「悪い点」では「レジスター」、「表現力」、「その他」の順となっており、「接続詞」を除くすべての下位項目が「良い点」での言及数よりも多くなっている。「良い点」の特徴である一方、下位群の「良い点」には見られなかった「接続詞」への言及は、「『第一に』『第二に』」など接続詞をうまく利用している、「段落冒頭に接続詞を使用している」などのコメントで、文章構成にも関わる観点である。接続詞を効果的に用いることで読み手の文章理解を助けることができれば、プラス評価につながる可能性があることが窺える。一方、「言語(レジスター)」に関する言及(例:「話し言葉が混在している」、「文体が統一されていない」)、および「形式(原稿用紙)」(例:「改行後の1文字下げが出来ていない」、「原稿の修正方法が多少難あり」)、「形式(文字数)」(例:「最終的には5行も表現の余地を残してしまっていることも大きな問題である」)に関する言及は、「悪い点」で多く出現しており、アカデミック・ライティングの文体から逸脱している場合や原稿用紙の使い方、文字数のルールから逸脱する場合は、マイナス評価につながりやすいと考えられる。「構成」に関する特徴は、次節で取り上げる。

表 7 群別、観点(良い点・悪い点)別 評価コメント出現総数詳細

コード	良い点(+)						合計	悪い点(-)						合計	
	上位群		中位群		下位群			上位群		中位群		下位群			
	NS上	NNS上	NS中	NNS中	NS下	NNS下		NS上	NNS上	NS中	NNS中	NS下	NNS下		
内容	論旨	24	22	6	12	0	6	70	13	4	13	14	22	15	81
	立場	12	11	6	8	14	9	60	1	1	15	2	5	3	27
	根拠	22	27	17	17	19	10	112	18	12	14	27	22	31	124
	状況把握	0	1	2	1	0	0	4	0	0	0	0	0	0	0
	論歩	2	1	0	0	0	2	5	3	3	0	2	0	2	10
	反論	2	0	0	0	0	2	4	1	0	0	0	0	0	1
	問題提起	0	0	1	1	0	0	2	0	0	0	0	1	0	1
小計	62	62	32	39	33	29	257	36	20	42	45	50	51	244	
言語	表現力	3	4	4	4	0	1	16	3	14	10	8	7	6	48
	レジスター	2	2	1	1	0	0	6	2	3	35	2	8	14	64
	文法	2	1	0	0	1	0	4	0	4	2	5	5	12	28
	語彙	2	1	0	2	1	2	8	0	5	3	17	5	13	43
	接続詞	6	2	1	2	0	0	11	0	0	0	0	5	0	5
	表記	1	2	1	1	0	0	5	0	0	1	5	6	3	15
その他	6	2	2	1	5	1	17	0	6	13	5	14	4	45	
小計	22	14	9	11	7	4	67	5	33	57	50	41	62	247	
構成	段落間	17	24	12	14	1	9	77	6	0	4	1	14	6	31
	段落内	6	2	2	2	2	1	15	2	0	3	1	5	0	11
	バランス	0	1	0	0	0	0	1	0	0	1	1	2	0	4
	立場	3	3	1	1	1	3	12	0	0	5	1	4	0	10
	小計	26	30	15	17	4	13	105	8	0	13	4	25	6	56
形式	文字数	1	2	1	1	1	1	7	0	0	0	0	16	7	23
	原稿用紙	0	0	0	0	0	0	0	0	7	5	1	0	0	13
	書字	1	1	0	0	0	0	2	0	0	0	1	0	0	1
小計	2	3	1	1	1	1	9	0	7	5	2	16	7	37	
印象	2	4	3	4	2	1	16	0	0	2	5	2	1	10	
題	2	0	0	2	0	0	4	8	0	1	4	2	2	17	
分類不能	0	0	0	0	0	1	1	0	0	0	0	1	0	1	
なし	0	1	13	4	14	10	42	12	7	2	0	0	1	22	
合計	116	114	73	78	61	59	501	69	67	122	110	137	130	635	

5.2.1 書き手の属性による比較

次に、書き手がNSかNNSかによる違いを考察するために、NSとNNSのそれぞれにおいて、コメント数が多かったコードを確認したところ、前節で示したように、NSもNNSも、「内容」>「言語」>「構成」>「形式」の順であった(表8)。書き手がNSかNNSかによって、評価の観点が異なるということであろう。この上位4つのコメント数を合計すると、NSが512、NNSが511で、近似している。ただし、表9に示したように、総コメント数が異なるため、この4つのコメント数が全体に占める比率はNNSの方がやや多い(NS:88.58%, NNS:91.58%)。また、「内容」と「形式」に関しては、NSとNNSとでコメントの比率はそれぞれ44%、4%程度でほぼ同じだが、「言語」に関するコメントはNNSの方が多く(NS:24.39%, NNS:31.18%)、「構成」に関するコメントはNSの方が多い(NS:15.74%, NNS:12.54%)。

NSとNNSのそれぞれにおいて、各評価の観点が、「良い点」と「悪い点」のいずれでより多く言及されているかを比較してみると、表10に示したように、コードによってNSとNNSの「良い点」と「悪い点」の比率の差が異なることがわかった。その傾向が特に顕著なのは、「構成」である。そこで、ここでは、NSとNNSにおける顕著な相違点として、「構成」に着目して、考察を行いたい。

表8 NS・NNS別 評価コメント出現総数

コード	NS									NNS										
	NS上			NS中			NS下			合計	NNS上			NNS中			NNS下			合計
	+	-	小	+	-	小	+	-	小		+	-	小	+	-	小	+	-	小	
内容	62	36	98	32	42	74	33	50	83	255	62	20	82	39	45	84	29	51	80	246
言語	22	5	27	9	57	66	7	41	48	141	14	33	47	11	50	61	4	62	66	174
構成	26	8	34	15	13	28	4	25	29	91	30	0	30	17	4	21	13	6	19	70
形式	2	0	2	1	5	6	1	16	17	25	3	7	10	1	2	3	1	7	8	21
印象	2	0	2	3	2	5	2	2	4	11	4	0	4	4	5	9	1	1	2	15
題	2	8	10	0	1	1	0	2	2	13	0	0	0	2	4	6	0	2	2	8
分類不能	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	1	1
なし	0	12	12	13	2	15	14	0	14	41	1	7	8	4	0	4	10	1	11	23
合計	116	69	185	73	122	195	61	137	198	578	114	67	181	78	110	188	59	130	189	558

表9 NS・NNS別 上位4種のコメント数

コード	NS	NNS
内容	255(44.11)	246(44.09)
言語	141(24.39)	174(31.18)
構成	91(15.74)	70(12.54)
形式	25(4.33)	21(3.76)
合計	512(88.58)	511(91.58)

表10 NS・NNS別、観点(良い点・悪い点)別
上位4種のコメント数

コード	NS+	NS-	NNS+	NNS-
内容	127(49.80)	128(50.20)	130(52.85)	116(47.15)
言語	38(26.95)	103(73.05)	29(16.67)	145(83.33)
構成	45(49.45)	46(50.55)	60(85.71)	10(14.29)
形式	4(16.00)	21(84.00)	5(23.81)	16(76.19)
合計	214(41.80)	298(58.20)	224(43.84)	287(56.16)

注:()はNSとNNSの各々の総コメント数に占める比率

注:()はNSとNNS各々における各観点総数を100とした場合の+と-の比率

まず、NSは、「構成」が「良い点」として評価される比率は49.45%、「悪い点」として評価される比率は50.55%で、ほぼ等しい(表10)。しかし、NNSでは、「良い点」として評価される比率が圧倒的に高く、85.71%であった。上に示した表8の「構成」部分だけを表11として切り出して、コメント数を比較してみると、NSでは、上位群で「良い点」>「悪い点」、中位群で「良い点」≒「悪い

点」, 下位群で「良い点」<「悪い点」となっている。このことから, 評価の高低が, 構成の良し悪しと関連があると言えよう。一方, NNS はいずれの群でも「良い点」>「悪い点」である。具体的には, 上位群ではすべてのコメントが「良い点」であり, 「悪い点」としてのコメント数はゼロ, 下位群でも「良い点」の 13 コメントに対して, 「悪い点」は 6 コメントである。これらの数値だけを見ると, NS よりも NNS の方が「構成」が整った意見文を書いている可能性が示唆される。

表 11 NS・NNS 別, 観点(良い点・悪い点)別「構成」のコメント数

構成	NS 上			NS 中			NS 下			合計	NNS 上			NNS 中			NNS 下			合計
	+	-	小計	+	-	小計	+	-	小計		+	-	小計	+	-	小計	+	-	小計	
	26	8	34	15	13	28	4	25	29	91	30	0	30	17	4	21	13	6	19	70

ここで, 具体的に, どのようなコメントが付されているか, 「構成」に関するコメントの記述を, NS と NNS とで比較するために, 4 つの下位項目のそれぞれについて, コメント数を計算した(表 12)。

表 12 NS・NNS 別, 観点(良い点・悪い点)別「構成」下位項目のコメント数

コード	NS								NNS							
	NS 上		NS 中		NS 下		合計		NNS 上		NNS 中		NNS 下		合計	
	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
段落間	17	6	12	4	1	14	30	24	24	0	14	1	9	6	47	7
段落内	6	2	2	3	2	5	10	10	2	0	2	1	1	0	5	1
立場	3	0	1	5	1	4	5	9	3	0	1	1	3	0	7	1
バランス	0	0	0	1	0	2	0	3	1	0	0	1	0	0	1	1
合計	26	8	15	13	4	25	45	46	30	0	17	4	13	6	60	10

まず, 「構成」の中で最も多かったコメントは, 「段落間」である。表 12 を見ると, 上位群と中位群では, NS と NNS の両方で, 「段落間」は「良い点」として言及されることが多いことがわかる。実際のコメントを見てみると, 上位群では, 「構成が良く練られている」, 「構成がしっかりしている」, 「議論の流れがわかりやすい」, 「順序立てて説明している」などのコメントが, NS にも NNS にも同様に付されている。中位群に対するコメントは, 主に, 「構成ができていない」, 「流れは悪くない」, 「段落に分けて書かれている」などで, 極めて高い評価というより, 段落間の構成は悪くない, という程度の良さとして言及されている。

また, 上位群と中位群の「段落間」のコメントのうち, NS では 2 割強が「悪い点」として言及されている。しかし, NNS では「悪い点」として付されたコメントは中位群の 1 つしかない。上位群の NS に対する「悪い点」として付された 6 つのコメントを見てみると, 5 つが「最終段落が蛇足で, 不要である」という趣旨のコメントであった。結論部分となる最終段落にそれまでに記述されていた内容とは異なる記述があったことが, 文章全体の一貫性を損なうものとしてマイナス評価されたようだ。また, 中位群の NNS に付された 4 つのコメントは, いずれも「段落分けに問題がある」というコメントであった。

一方、下位群の「段落間」のコメントを見ると、NS では「良い点」として言及されたのは 1 つのみで、残りの 14 コメントはすべて「悪い点」であった。そのコメントを見てみると、「構成が不明確」、「流れが見えにくい」、「論の流れや論理展開が良くない」、「構成に工夫が見られない」などで、上位群や中位群では「良い点」として着目された側面が、下位群では「悪い点」として言及されていることがわかる。また、NNS の「段落間」の「悪い点」としては、6 つのコメントが付されており、「段落分けがうまくできていない」、「段落のつながりが明確でない」、「流れが悪い」など、コメントの内容は NS と似ている。ただし、下位群の NNS には「良い点」としてのコメントが 9 つあり、それらの中には、「序論、本論、結論の構成は作っている」、「おおむね構成ができています」、「基本的な構成はできている」など、書き手の NNS が構成を作ろうとしている努力や意図を汲みとって、「良い点」として評価していることが見て取れる。

以上のように、「段落間」に関するコメント数やコメントの内容から、大学教員は、論文全体の構成を評価の重要な点として見ており、序論、本論、結論の基本的な全体構成が整っているかどうか、段落間の関係性が明確かどうか、一貫性のある論の展開ができていいるかどうかなどに着目して、評価していると言えよう。

次に、「立場」について見ておきたい。「立場」は、NS では「悪い点」として言及される比率が高いのに対して、NNS ではほとんどが「良い点」として言及されている。まず、上位群では、NS も NNS もすべて「良い点」として言及されている。コメントを見てみると、いずれも、「冒頭で立場を明確に示している」という趣旨のコメントであった。文章の最初に立場を明示し、それ以降で、根拠や事例を論述するという構成が、意見文としては高く評価されるということであろう。また、表 12 から明らかなように、中位群と下位群の NS では、「立場」は「悪い点」として評価される方が多い。コメントを見ると、いずれも、「最後まで読まない主張がわからない」、「先に自分の主張を述べていないのは論説文としては弱い」、「主張は最初にも書くべきである」、「前置きが長い」など、書き手の立場が冒頭に記されていないことを「悪い点」として評価していることがわかる。さらに、NNS では「悪い点」として付されたコメントは中位群に対しての 1 つのみである。このコメントは「はじめのほうで主張文が明確に書かれていない」であった。NS と類似のコメントであるにも関わらず NNS では 1 つしか付されていないのはなぜか、疑問に思い、当該意見文に当たってみた。NNS の中位群の意見文は、タイトルが「古さの良さ」となっており、書き手の主張がタイトルから読み取れるため、ほとんどの評価者は冒頭に主張が明示されていると捉えたものと考えられる。また、当該意見文の第一段落には、「あらゆるものは必ずそのよさがあり、かけがえのない価値があると思う。新聞や雑誌もそうである」と書かれており、ここが書き手の主張部分であり、立場は示されていると言えよう。ただし、上位群の書き方とは大きく異なっている。上位群では、NS の場合は、最初の一文目に「私は『これからも新聞や雑誌は必要だ』という意見を支持します」とあり、NNS も第一段落に「私はこれからも新聞や雑誌は必要だと思っている」と書かれており、いずれも立場表明が明確である。こうした書き方と比較すると、NNS の中位群の書き方に対して、主張が「明確に」は書かれていないと評価されてしまったのだら

う。

以上の結果から、いずれかの立場に立って論じるような意見文の場合、伊集院（2017：48）にも指摘があるとおおり、大学教員は、書き手が冒頭で立場を示しているかどうか、明確な言語表現で主張を述べているかどうか、に着目していると言えよう。

5.2.2 評価者の属性による比較

最後に、評価者が日本語教育を専門とする日本語教員（以下、J）かそれ以外の人文社会系の専門教員（以下、S）かによって、評価の観点に相違が見られるか否かを検討したい。

表 13 は、J と S の「良い点」と「悪い点」の両者を合算したコメント数を示している。この中から上位 4 コードの「内容」、「言語」、「構成」、「形式」に着目すると、J は、上位群と下位群、S は全ての群において「内容」>「言語」>「構成」>「形式」の順でコメントが多くなされており、いずれにおいても「内容」が評価の焦点となっていることがわかる。また、「内容」以外のコードで J と S を比較してみると、表 14 に示されているとおおり、J は「言語」のコメント割合が全体の 31.81%（S:22.13%）、また「構成」は 17.66%（S:9.39%）と、S に比して「言語」と「構成」に焦点が置かれていることがわかる。一方 S は、「形式」が全体の 6.89%で J の 1.98%よりも多く、S は J より「形式」に着目していることがわかる。

表 13 日本語教員・専門教員別 評価コメント出現総数

	日本語教員(J)				専門教員(S)			
	上位群	中位群	下位群	合計	上位群	中位群	下位群	合計
内容	100	79	86	265	80	79	77	236
言語	48	84	77	209	26	43	37	106
構成	41	34	41	116	23	15	7	45
形式	2	1	10	13	10	8	15	33
印象	3	7	4	14	3	7	2	12
題	3	3	2	8	7	4	2	13
分類不能	0	0	0	0	0	0	2	2
なし	10	8	14	32	10	11	11	32
合計	207	216	234	657	159	167	153	479

表 14 日本語教員・専門教員別 上位 4 種のコメント数

	日本語教員(J)				専門教員(S)			
	上位群	中位群	下位群	合計	上位群	中位群	下位群	合計
内容	100 (48.31)	79 (36.57)	86 (36.75)	265 (40.33)	80 (50.31)	79 (47.31)	77 (50.33)	236 (49.27)
言語	48 (23.19)	84 (38.89)	77 (32.91)	209 (31.81)	26 (16.35)	43 (25.75)	37 (24.18)	106 (22.13)
構成	41 (19.81)	34 (15.74)	41 (17.52)	116 (17.66)	23 (14.47)	15 (8.98)	7 (4.58)	45 (9.39)
形式	2 (0.97)	1 (0.46)	10 (4.27)	13 (1.98)	10 (6.29)	8 (4.79)	15 (9.80)	33 (6.89)
小計	191 (92.27)	198 (91.67)	214 (91.45)	603 (91.78)	139 (87.42)	145 (86.83)	136 (88.89)	420 (87.68)

注：() は J と S の各々の総コメント数に占める比率

さらに、JとSのコメントを観点(「良い点」と「悪い点」)別に分析してみると、「良い点」のコメント数はJ・Sともに「内容」>「構成」>「言語」>「形式」であるが、JはSに比して「言語」(例:「表現が豊かである」、「意見文にふさわしい語彙が使われている」、「段落冒頭に接続詞が使われている」と)と「構成」(例:「序論—本論—結論の流れが比較的わかりやすい」)、SはJに比して「形式」(例:「文字数が制限いっぱいである」、「読み易い字である」)を重視していることがわかる(表15)。また、「悪い点」については、Jは「言語」>「内容」>「構成」>「形式」であるのに対し、Sは「内容」>「言語」>「形式」>「構成」となっており(表16)、着目している観点の比重が異なっていることがより鮮明に表れている。

表15 日本語教員・専門教員別 上位4種のコメント数(良い点)

	日本語教員(J)				専門教員(S)			
	上位群	中位群	下位群	合計	上位群	中位群	下位群	合計
内容	72	32	38	142	52	39	24	115
言語	24	15	6	45	12	5	5	22
構成	36	24	12	72	20	8	5	33
形式	0	0	0	0	5	2	2	9
小計	132	71	56	259	89	54	36	179

表16 日本語教員・専門教員別 上位4種のコメント数(悪い点)

	日本語教員(J)				専門教員(S)			
	上位群	中位群	下位群	合計	上位群	中位群	下位群	合計
内容	28	47	48	123	28	40	53	121
言語	24	69	71	164	14	38	32	84
構成	5	10	29	44	3	7	2	12
形式	2	1	10	13	5	6	13	24
合計	59	127	158	344	50	91	100	241

さらに詳細を見るために、「悪い点」のうち、下位コードのコメント数が多いものに注目してみると(表17)、まず「内容」についての評価においては「根拠が信頼性に欠ける」、「根拠が主観的である」といったコメントに見られるように、J・Sともに、「根拠」に適切性がない場合に悪い評価につながっていることがわかる。

また、「言語」に関するコメントで顕著なのは、中位群の「レジスター」である。コメントの代表例は、「『だって・見れて』など会話体を使用している」、「口語表現や体言止めが見られる」というものであった。中位群でのJのコメント数は26で、Sのコメント数10の2.5倍となっており、レジスターに関してはJの方が、マイナスの評価につながりやすいことが見て取れる。

同様に、「構成」についてもJの方に言及が多く、例えば下位群の「段落間」に着目してみると、Sは1コメントのみのところ、Jには、「構成を考えていない」、「論理展開が明確でないところがある」、「段落構成が適切でない」といったコメントが19見られる。このことから、「言語」と同様、「構成」についても、JはSよりその不備を問題視する傾向があると言える。

一方、「形式」については、「文字数が少ない」といった「文字数」に関する指摘、「文頭に二文字空白がある」、「原稿用紙の使い方(行頭に句読点がある)が気になる」といった「原稿用紙」に関する指摘が見られ、JよりSの方が着目していることが明らかとなった。

表 17 日本語教員・専門教員別 下位項目のコメント数(悪い点)

	日本語教員(J)				専門教員(S)				
	上位群	中位群	下位群	合計	上位群	中位群	下位群	合計	
内容	論旨	12	14	17	43	5	13	20	38
	立場	1	12	4	17	1	5	4	10
	根拠	13	20	26	59	17	21	27	65
	状況把握	0	0	0	0	0	0	0	0
	譲歩	1	1	0	2	5	1	2	8
	反論	1	0	0	1	0	0	0	0
	問題提起	0	0	1	1	0	0	0	0
言語	表現力	10	11	8	29	7	7	5	19
	レジスター	4	26	18	48	1	10	4	15
	文法	3	5	14	22	1	2	3	6
	語彙	4	14	14	32	1	6	4	11
	接続詞	0	0	5	5	0	0	0	0
	表記	0	3	5	8	0	3	4	7
	その他	3	10	7	20	4	9	12	25
構成	段落間	5	0	19	24	1	5	1	7
	段落内	0	3	5	8	2	1	0	3
	パランス	0	2	2	4	0	0	0	0
	立場	0	5	3	8	0	1	1	2
形式	文字数	0	0	10	10	0	0	13	13
	原稿用紙	2	1	0	3	5	5	0	10
	書字	0	0	0	0	0	1	0	1
印象	2	10	16	28	0	6	1	7	

6. まとめ

以上、大学教員による作文の評価の観点を洗い出してコメント数を分析し、共通して見出される特徴を抽出すると同時に、作文の良し悪しや書き手、評価者の属性の違いに応じて、評価される観点の重みが異なっている可能性を指摘した。本研究データから見出された主な分析結果は、以下のとおりである。

- ・大学教員の評価の観点を網羅的に抽出することを目的に、評価コメントの切片化およびコード化を行った結果、評価の観点は「言語(表現力、文体、文法、語彙、接続詞、表記、その他)」、「内容(論旨、立場、根拠、状況把握、譲歩、反論、問題提起)」、「構成(段落間、段落内、パランス、立場)」、「形式(文字数、原稿用紙、書字)」、「印象」、「題」、「分類不能」、「なし」の25項目、8カテゴリーに集約された。
- ・上位群、中位群、下位群のいずれにおいても、評価コメント量は「内容」>「言語」>「構成」>「形式」の順であり、評価の際に内容的側面が最も重視されていることがわかった。また、下位分類まで見ると、「内容(根拠)」に関するコメントが最も多く、「内容(論旨)」

「構成（段落間）」、「内容（立場）」と続くことから、適切な「根拠」によって「立場」（主張）が明確に伝わり、全体の「論旨」が「構成」（展開）に支えられて理解しやすいことが重要であることが確認できた。

一方で、以下のような相違点も観察された。

- ・上位群では「内容」と「構成」を合わせると 66.67% を占めるのに対し、中位群と下位群ではその比率が下がり、その分、中位群では「言語」、下位群では「言語」と「形式」に比重が置かれていた。
- ・「構成」については、NS より NNS の方が良い点として評価される割合が高かった。
- ・日本語教員は専門教員に比して「言語」と「構成」、専門教員は日本語教員に比して「形式」に、より多く着目していた。

これらの結果から、作文の良し悪しや執筆者、評価者の属性によって、評価の観点の比重が異なっている可能性があると言える。ただし、本分析結果は、抽出された一部のデータに対する探索的分析であり、分析した作文独自の影響を大きく受けている可能性も否めない。残りの 24 編の作文に対する評価コメントについても同様の枠組みで分類可能か否かを精査しうえて、分析の精緻化を行う必要がある。また、書き手が日本語母語話者の場合と学習者の場合、評価者が専門教員の場合と日本語教員の場合での相違も見られたことから、これらの観点についてもさらに詳細な分析を進める必要がある。

文章評価は、通常、同一の採点基準に基づき、複数の教員が評点を付与することが多いが、採点基準の文言の解釈や評価の観点の重み付けは各教員に任されており、同一の作文に対して異なる評点が付けられても、その要因を把握することは難しい。今後さらに評価コメントの分析を進めることによって、大学教員による文章評価の実態だけでなく、大学生のアカデミック・ライティングに必要とされるスキルの解明につながることを期待できる。また、大学教員の作文評価に関する基礎的データの構築は、評価者が自身の評価の特徴に関して気づきを得るためのトレーニングで活用することも可能であろう。まずは、本研究で構築した枠組みで、残りの作文評価データを分析し、本分析結果を検証することを課題としたい。

謝辞

本研究は、JSPS 科研費 15K02633 の助成を受けて実施された。本稿は、2017 年 8 月 4 日に神戸大学で開催された第 3 回アジア圏学習者コーパス国際シンポジウム(LCSAW2017)でのポスター発表時に、参加者からいただいたコメントを反映させて執筆したものである。

本研究の調査にご協力くださった皆様、LCSAW2017 でコメントをくださった皆様、共同研究者の野口裕之氏(名古屋大学名誉教授)および李在鎬氏(早稲田大学教授)、評価コメントの切片化およびコーディングの方法についてご助言くださった近藤安月子氏(東京大学名誉教授)にお礼申し上げます。

引用文献

- 伊集院郁子(2017)「作文と評価:日本語教育的観点から見たよい文章」李在鎬(編)『文章を科学する』(pp. 38-57)東京:ひつじ書房.
- 伊集院郁子・小森和子・李在鎬・野口裕之・奥切恵(2016)「意見文の評価を左右する要因は何か:KH Coderを用いた評価コメントの分析を通して」『2016年度日本語教育学会秋季大会予稿集』, 255-256.
- 木下康仁(1999)『グラウンデッド・セオリー・アプローチ:質的実証研究の再生』東京:弘文堂.
- 田中真理・初鹿野阿れ・坪根由香里(1998)「第二言語としての日本語における作文評価:『いい』作文の決定要因」『日本語教育』, 99, 60-71.
- 田中真理・長阪朱美(2009)「ライティング評価の一致はなぜ難しいか:人間の介在するアセスメント」『社会言語科学』, 12(1), 108-121.
- 田中真理・坪根由香里(2011)「第二言語としての日本語小論文における good writing 評価:そのプロセスと決定要因」『社会言語科学』, 14(1), 210-222.
- 文部科学省高等教育局大学振興課大学改革推進室(2017)「平成 27 年度の大学における教育内容などの改革状況について (概要)」
http://www.mext.go.jp/a_menu/koutou/daigaku/04052801/_icsFiles/afieldfile/2017/11/21/1398426.pdf (2017年12月24日閲覧)

参考サイト

- 独立行政法人日本学生支援機構「日本留学試験「記述」採点基準」
http://www.jasso.go.jp/ryugaku/study_j/eju/about/score/_icsFiles/afieldfile/2017/03/31/saitenkijun_2013_1.pdf(2017年12月24日閲覧)

利用コーパス

- 「日本・韓国・台湾の大学生による日本語意見文データベース」
<http://www.tufts.ac.jp/ts/personal/ijuin/terms.html> (2018年1月4日閲覧)
- “The Corpus of Multilingual Opinion Essays by College Students (MOECS)”
<http://www.u-sacred-heart.ac.jp/okugiri/links/moeecs/moeecs.html> (2018年1月4日閲覧)