Kobe University Repository : Kernel

PDF issue: 2025-07-04

JF日本語教育スタンダードB1レベルの言語特徴についての検討 : 学習者による作文に基づいた分析

波多野, 博顕

(Citation)
Learner Corpus Studies in Asia and the World,3:189-206

(Issue Date)
2018-03-12

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/81010127

(URL)
https://hdl.handle.net/20.500.14094/81010127

JF 日本語教育スタンダード B1 レベルの言語特徴についての検討
―学習者による作文に基づいた分析―

波多野　博顕（国際交流基金日本語国際センター）
Hiroaki_Hatano@jpf.go.jp

An investigation of the linguistic features of B1 level of the JF Standard for Japanese Language Education.
―An analysis based on the writing data by Japanese language learners―

HATANO　Hiroaki (The Japan Foundation Japanese-Language Institute, Urawa)

概要
本稿は，JF 日本語教育スタンダード B1 レベルの作文課題への解答データ 110 件を対象に，課題達成度別でどのような言語特徴がみられるのかを検討したものである。複数の教師による評価に基づいてデータを「達成群」「中位群」「不達成群」に分け，量的側面と質的側面から検討を行った。量的側面としては，文字数・形態素数・文数・漢字率を指標とする「記述量」と，内容語と機能語の Guiraud 値を指標とする「語彙多様性」を分析した。質的側面としては，形態素解析や n-gram の結果から，相対的に達成群で特徴的なものを抽出した。分析の結果，達成群は不達成群よりも「記述量」の全ての指標，および，内容語の Guiraud 値が有意に高いことが明らかとなった。また，形態素解析および n-gram の結果から，達成群では作文で課された B1 レベルのタスクを遂行するうえで重要な言語特徴が頻出するとともに，活用に関連した文法形態素の多用が観察された。

キーワード
JF 日本語教育スタンダード，CEFR，B1，作文，形態素解析

1. はじめに
1.1 JF 日本語教育スタンダード

国際交流基金（JF）は，日本語の教え方・学び方・評価の仕方を考えるためのツールとして JF 日本語教育スタンダード（JFS）を開発し，2010 年に公開した（独立行政法人国際交流基金，2017a）。JFS は CEFR（Common European Framework of Reference for Languages）（Council of Europe, 2001）の考え方に基づいて開発されたため，JFS を用いることにより日本語の熟達度（レベル）を CEFR に準じて知ることができる。JFS および CEFR では言語の熟達度を，「基礎段階の言語使用者（Basic User）」である A レベル，「自立した言語使用者（Independent User）」である B レベル，「熟達した言語使用者（Proficient User）」である C レベルに大きく分け，各段階をさらに 2 つに分けている。そのため，全部で 6 つのレベルが設定されている（A1, A2, B1, B2, C1, C2）。

JFS が重視する能力の一つに，「課題遂行能力」がある。これは，行動中心主義の立場に基づき，日本語に関する知識をどれだけ持っているかではなく，日本語を使って何ができるかに注目した能力である。そのため，JFS のレベルは，文型や漢字をいくつ知っているかではなく，あるレベルでできる言語活動や言語能力が「～できる」という形式の文（Can-do statement，以降では Can-do とする）によって例示されている。「講演やプレゼンテーションをする」という言語活動を例にすると，以下のような Can-do がある（全て JF Can-do，後述）。

A1　みやげもの屋などの職場の店頭で，通行人の注意を引くために，商品の名前や値段などを言うことができる。

A2　スーパーなどで実演販売をするとき，商品の特徴や使い方などを短い簡単な言葉で紹介することができる。

B1　電気屋などの職場で，あらかじめ準備してあれば，客に電子辞書などの商品について，ある程度詳しく紹介し，想定した質問に答えることができる。

B2　職場の企画会議などで，あらかじめ準備してあれば，自分の企画案について図表やグラフなどを示しながら，明確に詳しく説明し，質問に的確に対応することができる。

1.2 問題の所在

CEFR は言語的に中立な枠組みであるため，個別言語に適用する際には具体化・文脈化が求められる。その一環として，JFS では CEFR の Can-do（以降、CEFR Can-do とする）を参考に独自の Can-do を作成し，「JF Can-do」として公開している。これにより，言語使用場面が想起しにくい CEFR Can-do を，日本語教育の現場で使いやすいものとして具体化させている。例えば，「共同作業中にやりとりをする」という言語活動では，以下のように，抽象的な CEFR Can-do の内容を JF Can-do で文脈化させている（いずれも B2 の Can-do）。

<u>CEFR Can-do</u>
原因や結果を推測し，異なるアプローチの利点と不利な点を比較考量しながら，論点や問題の概略をはっきりと述べることができる。

<u>JF Can-do</u>
見本市や展示会などの準備をしているとき，展示品のレイアウトや説明手順などについて，当日の状況を予測しながら作業手順や分担方法を比較し，それぞれの利点や問題点などをはっきりと述べることができる。

JF Can-do の作成においては，記述内容やレベルの妥当性について，日本語教師 22 名の評価から質的・量的に一定の検討が加えられている（森本・他，2011）。

一方，或る JFS のレベルでは具体的にどのような言語項目（文法や語彙など）が「使える」ようになるのかについては，未だ実証的な検討がなされていない。特定の課題遂行を実現するためには，

それを下支えする言語能力が必要である(森本・他, 2011)。CEFR および JFS では, それらの言語能力は,「言語構造的能力(linguistic competences)」「社会言語能力(sociolinguistic competences)」「言語運用能力(pragmatic competences)」から構成されるとしている(Council of Europe, 2001)。上記の各能力に CEFR では Can-do が用意されているものの, 先述した汎言語性のため, 具体的な言語項目は記述されていない。例えば,「言語構造的能力」に含まれる「語彙の使いこなし」というカテゴリーの Can-do では,「具体的な日々の要求に関する狭いレパートリーの語を使うことができる」(A2)となっている。Can-do に基づいた学習や指導効果を学習者にフィードバックする方法の一つとして, どのような文法項目や語彙, 語用論的知識などが運用できれば「できる」ようになるのかを評価し, 具体的に示すことも有効であろう(渡部, 2015)。

欧州評議会は, CEFR を各言語で利用するための資料の一つとして, レベル別に個別言語の文法や語彙を整備した RLD(Reference Level Descriptions)が必要であるとしている(Council of Europe, 2005;投野, 2013)。しかし, 日本語教育においては, CEFR および JFS の評価に習熟した者が少ないこと, そのような評価による日本語学習者のデータが少ないことなどから, 定量的なデータ分析は進んでいない。各レベルで学習者が実際に使っている日本語の特徴を明らかにすることは, JFS に基づいた学習・指導・評価の実践に資する有効な情報となるだろう。

1.3 本稿の目的

以上を踏まえ, 本稿では JFS の評価に基づいた作文データを用い, その言語特徴の分析を試みる。具体的なリサーチ・クエスチョン(RQ)は以下の通りである。RQ1 では計量的手法に基づいて B1 達成作文の特徴を探り, RQ2 では言語表現の出現頻度に基づいて具体的な言語特徴を探る。

RQ1: B1 達成の程度に応じて, 記述量や語彙多様性に違いはあるか。
RQ2: B1 達成の程度に応じて, どのような表現が現れるようになっているか。

レベル別の言語特徴を検討するに際して, 本来はレベル横断的に分析してその差異を論じるべきである。しかし, データの制約上, 本稿では一つのレベル(B1)に注目する。本稿で扱うデータは特定の B1 Can-do に基づいて作成されているため, あくまで「B1 レベルのその課題を達成したかどうか」という観点から評価されている。そのため, 達成と評価された作文には B1 以上が含まれている可能性は否定できない。一つのレベルのみを分析対象とすることは議論の余地があるものの, 少なくとも本データから B1 の下限としての特徴の一部は示すことができるだろう。また, 6 レベルのうち B1 を対象とする理由は, 相対的な学習者数の多さがあげられる。独立行政法人国際交流基金(2017b)は, 世界の日本語学習者 967 名を対象に JFS のレベル判定を行った結果, B 判定が全体の過半数(505 名)を占め, そのうち 362 名が B1 であることを報告している。以上から, まずは B1 を対象とする。

2. 先行研究

日本語学習者の産出データを対象に言語特徴を分析した近年の研究として，話し言葉では山内(2009, 2015)，書き言葉では橋本(2015)があげられる。山内(2009, 2015)は，「KY コーパス」(ACTFL の定める OPI(Oral Proficiency Interview)の手順に従って実施された対面調査で収集された学習者発話データを文字化したもの)を観察し，レベル別(初級・中級・上級・超級)の発話に特徴的な形態素や n-gram の分析結果を報告している。橋本(2015)は，60 名の留学生による 12 の作文タスクが収録された「YNU 書き言葉コーパス」(金澤(編)，2014)を使用し，評価別に分けられた下位群・中位群・上位群における機能語や表現の出現頻度を分析している。橋本(2015)は作文データから言語項目を定量的に分析しているものの，「YNU 書き言葉コーパス」で調査対象となった留学生は「一般的には上級と称されるレベル」(金澤(編)，2014, p.11)である。

独立行政法人国際交流基金(2017b)は，JFS レベル判定と日本語能力試験(JLPT)合否との関係についての調査結果を報告している。報告によると，教師によるレベル評価表を用いた評価で B1 と判定された日本語学習者 362 名の JLPT 本試験での合格率は，N4 で 85.7 %(14 名受験，12 名合格)，N3 で 84.1 %(69 名受験，58 名合格)，N2 で 65.2 %(132 名受験，86 名合格)，N1 で 46.6 %(146 名受験，68 名合格)であった(N5 は受験者数が少ないため非記載)。このような JFS と JLPT のレベル間対応のバラつきは，2 つの評価体系における測定対象の違いや調査対象者の特性等が要因として指摘されている。ただし，同調査はあくまで JFS レベルと JLPT の合否結果との連関が検討されているため，JLPT の具体的な出題項目との関係は検討されていない。

櫻井・東(2011)は，各種の日本語指導項目リストや文法書および Threshold 1990(van Ek & Trim, 1991)を参照しながら，「最低限，B1 として『産出レベル』で『できる』こと」(p.57)を機能別に整理し，教授項目としてまとめた「表現文型マップ」を作成している。同マップは A1～C レベルを視野に，「全体の中での B1」の位置づけを明確にしたうえで各表現が整備されている。しかし，カリキュラム作成における参照情報の提供を目的としているため，産出データに基づいて「B1 では実際にどのような表現が使えるようになるのか」を探る本研究のアプローチとは異なる。山内(2009)は「KY コーパス」から『みんなの日本語』の初級文法項目の妥当性を分析し，「教える」項目と「使える」項目に乖離があることを指摘している。

このように，JFS のレベルとその言語特徴をボトムアップ的に検討する研究は，未だ行われていないのが現状である。

一方，英語教育では，CEFR のレベルごとにどのような文法や語彙が使えるようになっていくのかについて，学習者コーパスに基づいた実証研究の成果が English profile programme として公開されている(Hawkins & Filipović, 2012)。ポジティブ/ネガティブな文法の基準特性や，語義をベースとした語彙のレベル付け等が，実際の英語学習者によるスピーキングやライティングのサンプルデータから構築されている。RLD が構築されている英語教育では，それに基づいた教育・学習・評価の模索も始まっている(投野，2013)。

3. データの概要と分析手法

3.1 データの概要

国際交流基金日本語国際センターで実施された訪日教師研修で用いられた，B1 レベルの作文課題への解答をデータとした。研修には海外の日本語教師 55 名（日本語非母語話者）が参加し，プレおよびポストテストの一部として同じ内容で作文課題が実施されたため，全体のデータ数は 110（55 名×2）である。55 名のうち，男性は 12 名（平均年齢 29.5 歳，標準偏差 3.3），女性は 43 名（平均年齢 26.8 歳，標準偏差 3.1）である。研修参加者の出身は 34 カ国からなるが，その中に漢字圏（中国・台湾・韓国）は含まれていない。

作文課題の内容は，以下の CEFR Can-do を参考に作成された。

CEFR Can-do B1 文章でのやりとり全般

直接的に関連のある簡単な情報を求めたり伝えたりする個人的な手紙，覚書を書くことができ，自分が重要だと思う点を相手に理解させることができる

実際の課題では，研修生の町に来る予定の日本人の友人から，その町で自身の趣味であるスポーツができるかどうか質問のメールが来たので返信する，という内容に具体化されて実施された。回答時間は 30 分で辞書は使用できず，400 字のマスがある用紙に手書きで記入することを求めた（書ききれない場合は欄外に記入）。また，漢字の間違いは評価に影響しないことが研修生に伝えられた。

作文の評価は，研修を担当する日本語国際センターの教師 5 名（日本語母語話者）が行った。55 の作文を 24 名分と 31 名分に分け，評価者のうち 2 名が前者，3 名が後者の全てに目を通し，それぞれで各自の結果を持ち寄って，話し合いにより最終的な判定が行われた。評価者は，いずれも日本語国際センターで JFS による日本語教育に従事しており，レベルの基準やイメージについての知識を有している。評価の観点は「内容」「構成」「言語能力」「読み手意識」の 4 つで，「構成」と「言語能力」には評価に関連するいくつかの CEFR Can-do が目安として評価者に示された。なお，具体的な言語形式の使用・不使用という観点は評価に含まれていない。

最終的な判定の結果，各作文は B1 としての課題達成度が 5 段階（不達成，惜しいが不達成，ほぼ達成，達成，大幅に達成）で評価された。110 件の判定の内訳は，「不達成」8 件，「惜しいが不達成」21 件，「ほぼ達成」37 件，「達成」37 件，「大幅に達成」7 件であった。

3.2 分析手法

3.2.1 データの事前処理

回答は，全て電子データに書き起こされた。電子データ化に際して，金澤（編）（2014）を参考に，2 種類の書き起こしを用意した。

一つ目の書き起こしは，誤字脱字や文法の誤用があっても訂正せず，原文を書かれたとおりに入力したものである。電子化に際して入力不可能な文字や判読不明な文字の誤りは，その文字に

近い，あるいは文脈上近いと推測されるものを，その文字数分だけ入力した。例えば，「呼」という漢字が，原文では「乎口」のように左右が逆に記入されていた場合，その旨を別欄に注記したうえで，書き起こしでは「呼」として入力した。また，原文で分かち書きされている場合，句読点は加えなかった。この書き起こしによる文字数の総計は 39,155 字（うち漢字 5,689 字）であった。

二つ目の書き起こしは，形態素解析を目的として原文に修正を加えたものである。原文と電子データを付き合わせながら，漢字で書けるものは漢字に変換し（例：すいえい→水泳），漢字表記のミス（例：休→体）や，送り仮名のミス（例：大好です→大好きです）を修正した。また，清音と濁音の混同（例：～てすか→～ですか）や，特殊拍の過不足（例：ラニング→ランニング），日本語で馴染みのない外来語表記（例：ステディオム→スタジアム）なども修正した。助詞選択や活用等の文法的誤り（例：うれしいかったです）は修正していないが，そのままでは意味が通らない単純な表記ミスは修正した（例：かがいています→輝いています）。切れ目なく文字が続いている場合や，分かち書きの場合は，文脈上不自然でない位置に句読点を追加した。この書き起こしによる文字数の総計は 37,965 字（句読点等の記号を除いた形態素の延べ語数は 19,395）であった。

3.2.2 分析の手順

まず，5 段階の判定結果に基づいて，110 件の作文データを課題達成度別に 3 群に分類した。内訳は，「不達成群」29 件（「不達成」「惜しいが不達成」から構成），「中位群」37 件（「ほぼ達成」から構成），「達成群」44 件（「達成」「大幅に達成」から構成）である。作文はプレおよびポストテストで用いられたため，同一人物によるデータが 2 件存在するが，誰が書いたかは考慮せず，判定結果のみに基づいて分類を行った。

3.2.2.1 記述量と語彙多様性の分析（RQ1）

記述量の指標として，文字数・形態素数・文数・漢字率を用いた。個人ごとに各指標の値を求めた後，群ごとに平均値を算出した。文字数の他に形態素数も用いたのは，同じ記述でも仮名表記と漢字表記で長さが異なるためである。例えば，「公園（こうえん）」は仮名表記だと長さは 4 だが，漢字表記だと長さは 2 となる。しかし，形態素としては 1 単位なので，表記形態が影響しない長さの単位として利用できる。達成度（不達成群・中位群・達成群の 3 水準）を要因として，記述量の各指標に及ぼす効果の有意性を確認するため，指標ごとに一元配置の分散分析を行った。

語彙多様性の指標には，様々な計算方法が存在する（石川，2012；中俣，2015）。中俣（2015）を参考に，本稿では，延べ語数の影響が調整され，結果が 0～10 の範囲に収まる Guiraid Index（Guiraud の R 値，以降では単に Guiraud 値とする）を指標として用いた。コーパスに顕著なサイズ差がある場合は Herdan の C 値（$C = log_e Type / log_e Token$）を検討すべきとされるものの（石川，2012），本稿のデータは小規模で群別のサイズ差も多くて約 2 倍程度であるため，延べ語数を自然対数で圧縮する C 値よりも平方根で調整する Guiraud 値（$R = Type/\sqrt{Token}$）の方が扱いやすいと判断した。内容語（名詞・動詞・形容詞）と機能語（助詞・助動詞）のそれぞれで個人ごとに延べ語数（Token）および異なり語数（Type）を計測して Guiraud 値を求めた後，群ごとに平均値

を算出した。ここでも，達成度を要因として Guiraud 値に及ぼす効果の有意性を確認するため，一元配置の分散分析を行った。

文字数・文数・漢字率の分析には，一つ目の書き起こし（そのまま原文を書き起こしたもの）を用いた。文数はテキスト中の「。．！？」などの出現数から求めた。漢字率は，プログラミングソフト python（ver.3.4）の正規表現パッケージ（re）を用いて作文ごとに一文字単位で漢字かどうかを判定し，漢字と判定された文字数を全体の文字数で除すことで計算した。

形態素数の算出および内容語と機能語の同定には，二つ目の書き起こし（形態素解析用に原文を修正したもの）の形態素解析結果を用いた。解析器は MeCab，解析辞書は IPA 辞書を使用した。以降では，便宜上，「形態素」のことを「語」として言及することがある。

3.2.2.2 表現の分析（RQ2）

二つ目の書き起こしによる形態素解析の結果から，達成群における高頻度語の抽出を行った。

ここでは，「使用率」という基準を用いた。まず，或る形態素が各群で「何人」に使われているかを計算し，その群内での使用率を算出した。例えば，形容詞「いい」は達成群で 35 人，中位群で 25 人，不達成群で 17 人に一回以上使われていた。これを各群での割合に換算すると，達成群は 79.5％（35 人/44 人），中位群は 67.6％（25 人/37 人），不達成群は 58.6％（17 人/29 人）となる。これを本稿での使用率とする。使用率の計算基準を各群で「何回使用されたか」ではなく「何人に使用されたか」にしたのは，各群の母数が異なるため，出現回数の粗頻度で比較ができないからである。また，たとえ各群の母数が同じであっても，個人の使用回数を考慮すると，群としての使用実態の観察が歪む可能性がある。例えば，各群 30 人でそれぞれ 100 回，50 回，30 回の使用回数が観察された場合，或る群では少数の使用者によって頻繁に使われた結果，群として見かけ上の使用回数が多くなっている可能性がある。逆に，或る群では多くの人に使用されていたものの一人ごとの使用回数が少ないため，群として見かけ上の使用回数が少なくなっている可能性がある。以上から，「達成群における高頻度語」を正確に抽出するため，「何回」ではなく「何人」に使用されたかを基準にした。なお，今回はあくまで当該の表現が「出たか出なかったか」のみを観察しており，「出たが正しいのか」（つまり正用か誤用か）という分析は行えていない。

群ごとに全ての形態素について使用率を算出し，「達成群の使用率が不達成群より 10 %以上高い」かつ「達成群の使用率が 50 %以上（つまり，達成群の過半数で使用された）」かつ「不達成群の使用率が 50 %以下（つまり，不達成群の過半数で使用されていない）」という基準で，達成群に特徴的な語を抽出した。さらに，n-gram によって前後の環境を観察した。n-gram の分析には RMeCab の docDF 関数を用い，全ての品詞を抽出対象とした（石田・小林，2013）。

動詞の後続表現および文末表現についても n-gram で分析を行い，「達成群の使用率が不達成群より 10 %以上高い」という基準で抽出を行った。動詞の後続表現は，直後に助詞と助動詞が後接するものを対象に 6 形態素まで 1 形態素ずつ後方を探索し，基準を満たすものを抽出した。文末表現は，「。．！？」を基準として 6 形態素まで 1 形態素ずつ前方を探索し，基準を満たすものを抽出した。

4. 結果と考察

4.1 RQ1 の結果

4.1.1 記述量の結果

B1 達成群別に，文字数，形態素数，文数，漢字率の各平均値と標準偏差，および，分散分析の結果（F 値および MSe = 群内の平均平方）を表 1 にまとめて示す。表中，標準偏差は括弧内に示し，「*」は 5 %水準で，「**」は 1 %水準で，「***」は 0.1 %水準で，分散分析において有意差があったことを示す（以下同じ）。

表 1　B1 達成群別にみた文字数，形態素数，文数，漢字率の平均値と標準偏差および分散分析の結果

	n	文字数**	形態素数***	文数**	漢字率**
不達成群	29	313.2 (94.7)	150.1 (45.9)	11.4 (4.9)	12.0% (5.0)
中位群	37	347.6 (62.3)	172.6 (32.5)	13.1 (3.5)	15.0% (4.1)
達成群	44	391.2 (49.8)	196.7 (22.7)	14.4 (3.0)	15.5% (3.3)
$F(2, 107)$		11.9	17.3	5.5	7.1
MSe		4,651	1,115	14	16.5

表 1 のとおり，全ての指標で達成群の数値が最も高く，不達成群の数値が最も低い。また，分散分析の結果も，各群の全指標で有意差がみられた。

群間での各指標の具体的な差を明らかにするため，Tukey の多重比較を行った結果を表 2 にまとめる。

表 2　記述量の指標に対する Tukey の多重比較の結果

指標	Tukey の多重比較の結果		
	達成群-不達成群	達成群-中位群	中位群-不達成群
文字数	***	*	ns
形態素数	***	**	*
文数	**	ns	ns
漢字率	**	ns	*

表 2 の結果から，達成群は不達成群よりも全ての指標で有意に数値が高いことが明らかとなった。また，中位群は，文数と漢字率において達成群と有意差がみられないものの，文字数と形態素数では有意差がみられた。一方，不達成群とは形態素数と漢字率で有意差がみられた。

4.1.2 語彙多様性の結果

B1 達成群別に，内容語と機能語の Token, Type, Guiraud 値の各平均値と標準偏差，および，分散分析の結果を表 3 に示す。

表 3 B1 達成群別にみた内容語と機能語の Token，Type，Guiraud 値の平均値と標準偏差および分散分析の結果

	内容語			機能語		
	Token***	Type***	Guiraud*	Token***	Type*	Guiraud
不達成群	74.1 (23.3)	49.2 (14.7)	5.7 (0.9)	68.4 (20.3)	22.6 (4.2)	2.8 (0.4)
中位群	85.4 (16.7)	55.0 (11.7)	6.0 (0.8)	79.4 (15.0)	23.5 (4.0)	2.6 (0.4)
達成群	96.2 (11.1)	61.1 (8.6)	6.2 (0.7)	91.7 (11.7)	25.0 (3.6)	2.6 (0.3)
$F(2, 107)$	15.1	9.4	4.0	20.4	3.6	2.0
MSe	285	132.1	0.6	238	15.4	0.1

Token, Type とも，内容語と機能語で達成群の数値が最も高く，不達成群の数値が最も低い。また，分散分析の結果も有意であった。一方，語彙多様性の指標となる Guiraud 値では，分散分析で内容語のみに有意差がみられた（Tukey の多重比較では達成群と不達成群に 5 %水準で有意差あり）。

4.2 RQ2 の結果

4.2.1 高頻度表現

表 4 に，「達成群の使用率が不達成群より 10 %以上高い」かつ「達成群の使用率が 50 %以上」かつ「不達成群の使用率が 50 %以下」の基準で抽出した形態素を網掛け部に示す。また，それが前接および後接した際に，「達成群の使用率が不達成群より 10 %以上高い」表現も併せて示す。なお，表中の「差」は「達成群の使用率－不達成群の使用率」で，品詞は形態素解析で用いた IPA 辞書の結果に基づく（以下同じ）。

表 4 B1 達成群に特徴的な語（網掛け部）およびその接続表現
（数字は群内での使用人数および使用率，差は達成群から不達成群の使用率を引いたもの）

	差	不達成群（29 名）	中位群（37 名）	達成群（44 名）
格助詞　から	21.2%	11（37.9%）	18（48.6%）	26（59.1%）
V-てから	12.5%	1（ 3.4%）	5（13.5%）	7（15.9%）
終助詞　よ	18.9%	11（37.9%）	18（48.6%）	25（56.8%）
V-ますよ	24.8%	6（20.7%）	13（35.1%）	20（45.5%）
できますよ	18.0%	4（13.8%）	6（16.2%）	14（31.8%）
名詞（非自立）　こと	33.8%	8（27.6%）	20（54.1%）	27（61.4%）
V-こと	28.1%	7（24.1%）	17（45.9%）	23（52.3%）
V-たこと	11.4%	0（ 0.0%）	2（ 8.1%）	5（11.4%）
ことができます	19.2%	3（10.3%）	11（29.7%）	13（29.5%）
副詞　とても	16.6%	11（37.9%）	13（35.1%）	24（54.5%）
名詞（副詞）たくさん	15.5%	10（34.5%）	14（37.8%）	22（50.0%）
名詞（副詞）近く	22.4%	8（27.6%）	11（29.7%）	22（50.0%）

表 4 から，動詞の後続表現や文末表現に達成群と不達成群の差が大きいものが多くみられることがわかる。また，非網掛け部のうち「V_こと」も達成群で使用率が過半数を超えており，特徴的な表現となっている。

4.2.2 動詞の後続表現

「達成群の使用率が不達成群より 10 %以上高い」動詞の後続表現を，「助詞後続型」「助動詞後続型」に分けて示す。なお，不達成群でも使用率が 50 %を超える場合は括弧をつけて示す。

4.2.2.1 助詞後続型の動詞表現

表 5 に，「達成群の使用率が不達成群より 10 %以上多い」助詞後続型の動詞表現を示す。表中，「V-てから」は表 4 でも示しているが，ここでも再度掲げる。また，特に注目すべきものについては，該当箇所に網掛けをして示す。

表 5 B1 達成群に特徴的な動詞表現(助詞後続型)

	差	不達成群 (29 名)	中位群 (37 名)	達成群 (44 名)
(V-て)	13.8%	25 (86.2%)	36 (97.3%)	44 (100%)
V-てから	12.5%	1 (3.4%)	5 (13.5%)	7 (15.9%)
V-ても	12.4%	3 (10.3%)	4 (10.8%)	10 (22.7%)
V-てくる	18.2%	0 (0.0%)	3 (8.1%)	8 (18.2%)
V-てみる	36.3%	2 (6.9%)	9 (24.3%)	19 (43.2%)
V-てみた	11.4%	0 (0.0%)	0 (0.0%)	5 (11.4%)
(V-ている)	39.2%	15 (51.7%)	26 (70.3%)	40 (90.9%)
V-ている-N	14.5%	7 (24.1%)	10 (27.0%)	17 (38.6%)
V-ているので	11.4%	0 (0.0%)	2 (5.4%)	5 (11.4%)
V-ていますから	14.7%	1 (3.4%)	3 (8.1%)	8 (18.2%)
V-ながら	15.9%	0 (0.0%)	3 (8.1%)	7 (15.9%)
V-に行く	10.2%	1 (03.4%)	5 (13.5%)	6 (13.6%)

表 5 から，達成群ではテ形の表現が多様であることがわかる。このうち，「V_てみる」(網掛け部)は達成群で半数に近い使用率を示しており，本データにおける特徴的な助詞後続型の動詞表現となっている。

4.2.2.2 助動詞後続型の動詞表現

表 6 に，「達成群の使用率が不達成群より 10 %以上高い」助動詞後続型の動詞表現を示す。表中，「V_たこと」は表 4 でも示しているが，ここでも再度掲げる。

表 6 B1 達成群に特徴的な動詞表現(助動詞後続型)

	差	不達成群 (29 名)	中位群 (37 名)	達成群 (44 名)
V-たい	10.9%	12 (41.4%)	15 (40.5%)	23 (52.3%)
(V-た)	14.1%	17 (58.6%)	30 (81.1%)	32 (72.7%)
V-たから	11.4%	0 (0.0%)	1 (2.7%)	5 (11.4%)
V-た-N	19.0%	9 (31.0%)	19 (51.4%)	22 (50.0%)
V-たこと	11.4%	0 (0.0%)	2 (5.4%)	5 (11.4%)
V-ない	26.9%	10 (34.5%)	18 (48.6%)	27 (61.4%)
V-ない-N	15.8%	2 (6.9%)	3 (8.1%)	10 (22.7%)
V-ますから	15.7%	6 (20.7%)	14 (37.8%)	16 (36.4%)

表 6 から，達成群に特徴的な助動詞後続型動詞表現として，「V-たい」「V-た-N」「V-ない」があげられる（網掛け部）。これらは，達成群での使用率が過半数を超えている。なお，「V-た-N」に対して，動詞が活用せずに名詞に接続する「V（原形）-N」を観察すると，不達成群でも使用率が 89.7 %と高い（達成群の使用率との差も 8.0 %）。

4.2.3 文末表現

「達成群の使用率が不達成群より 10 %以上高い」文末表現を表 7 に示す。なお，「不達成群の使用率が達成群より 10 %以上高い」もの（つまり，不達成群に特徴的な語）があれば，差をマイナスにして示す。

表 7 B1 達成群に特徴的な文末表現

	差	不達成群（29 名）	中位群（37 名）	達成群（44 名）
V-ますよ。	28.2%	5 (17.2%)	12 (32.4%)	20 (45.5%)
できますよ。	15.8%	4 (13.8%)	7 (18.9%)	13 (29.5%)
V-ましょう。	18.9%	11 (37.9%)	14 (37.8%)	25 (56.8%)
V-ましょうか。	13.6%	0 (0.0%)	4 (10.8%)	6 (13.6%)
V-ません。	-22.1%	13 (44.8%)	15 (40.5%)	10 (22.7%)
V-ませんか。	26.0%	3 (10.3%)	7 (18.9%)	16 (36.4%)
V-てみませんか。	13.6%	0 (0.0%)	1 (2.7%)	6 (13.6%)
V-ています。	32.6%	9 (31.0%)	13 (35.1%)	28 (63.6%)
と思っています。	10.2%	1 (3.4%)	1 (2.7%)	6 (13.6%)
V-と思います。	15.8%	2 (6.9%)	10 (27.0%)	10 (22.7%)
V-ないです。	11.4%	0 (0.0%)	1 (2.7%)	5 (11.4%)
できます。	26.7%	14 (48.3%)	21 (56.8%)	33 (75.0%)
できません。	-13.9%	6 (20.7%)	5 (13.5%)	3 (6.8%)
ができます。	14.5%	7 (24.1%)	17 (45.9%)	17 (38.6%)
もできます。	15.8%	4 (13.8%)	10 (27.0%)	13 (29.5%)
V-ことができます。	15.8%	2 (6.9%)	10 (27.0%)	10 (22.7%)
来てください。	11.4%	0 (0.0%)	1 (2.7%)	5 (11.4%)
Adj-です。	21.2%	11 (37.9%)	16 (43.2%)	26 (59.1%)
N（形容動詞語幹）-です。	15.4%	14 (48.3%)	17 (45.9%)	28 (63.6%)
N-があります。	11.1%	8 (27.6%)	14 (37.8%)	17 (38.6%)
N-をします。	13.6%	2 (6.9%)	8 (21.6%)	9 (20.5%)

表 7 から，「V-ますよ。」「V-ましょう。」「V-ています。」「Adj-です。」などが，達成群に特徴的な文末表現となっている（網掛け部）。その他，文末表現で特徴的な点を以下に示す。

一点目は，動詞述語の「V-ます。」（表外）と「V-ますよ。」である。表外の「V-ます。」の使用率は，不達成群 93.1 %，達成群 97.7 %で共に高い。一方，終助詞「よ」が付いた「V-ますよ。」という文末表現になると，差が 28.2 %に拡がる。

二点目は，否定表現の「V-ません。」と「V-ないです。」である。「V-ません。」は不達成群に多く差は-22.1 %であるのに対し，「V-ないです。」は不達成群に見られない。また，「V-ません」に終助詞「か」が付いた文末表現になると（「V-ませんか。」「V-てみませんか。」），達成群に多くなる。

三点目は，達成群では形容詞述語の文末「Adj-です。」や形容動詞述語の文末の「N-です。」が多いことである。形容動詞述語以外の文末（つまり名詞述語の文末）の「N-です。」の使用率は，不達成群でも 60.7 %と高く，達成群との差も 9.7 %と小さい。また，先述の通り，動詞述語の文末「V-ます。」は，不達成群でも 93.1 %と高い。

四点目は，文末の終助詞「ね。」（表外）と「よ。」である。単体での両者の文末における使用率を不達成群と達成群で比較すると，表外の「ね。」では差がみられないが（不達成群 34.5 %，達成群 38.6 %），「よ。」では差が大きい（不達成群 34.5 %，達成群 56.8 %，差 22.3 %）。なお，「よね。」はどちらの群でも使用率が低かった（不達成群 3.4 %，達成群 6.8 %，）。

5. 考察

5.1 B1 達成群別にみた記述量と語彙多様性

達成群では，記述量の指標である文字数，形態素数，文数，漢字率の全てにおいて，中位群や不達成群よりも大きな値となった（表 1）。また，多重比較の結果でも，達成群と不達成群の間では全指標において有意差が確認された（表 2）。

達成群の文字数は平均 391.2 字であるため，30 分以内に回答用紙のマス（400 字）を全て埋めるだけの記述力があるといえる。一方，不達成群は平均 313.2 字で，時間内に回答用紙の 4 分の 3 程度しか埋められないことがわかる。文字数の差は形態素数でも確認されたことから，表記形態に関わらず，達成群は他の群より多く書けている。これは，課題を遂行するうえでの内容面の豊かさを反映しているだろう。また，平均文字数および平均形態素数を平均文数で除して「一文あたりの文字数と形態素数」を求めると，群間で大きな差はみられない（「一文あたりの文字数」不達成群 27.5 中位群 26.5 達成群 27.2;「一文あたりの形態素数」不達成群 13.2 中位群 13.2 達成群 13.7）。このことから，記述量の差は，「文の長さ」ではなく「文の数」の差，即ち述語使用の差として捉えられる。作文の実施に際して漢字自体は評価に含まないことが周知されたものの，漢字率に達成群と不達成群で明確な差が現れたことから，作文課題達成力と漢字表記力（正用かどうかは別にして漢字を書くという力）に何らかの関係があることが示唆される。

中位群は達成群と比較すると，文字数と形態素数で有意差がみられた。一方，不達成群との比較では形態素数と漢字率で有意差がみられた（表 2）。このことから，中位群は達成群と比べて記述量が少ないものの不達成群ほどではなく，漢字表記は達成群程度に多いことが推察される。

語彙多様性の指標である Guiraud 値では，内容語（名詞・動詞・形容詞）のみに群間で有意差がみられ，機能語（助詞・助動詞）では有意差がみられなかった（表 3）。一般に，内容語に含まれる名詞や動詞は相対的に新陳代謝が活発で属する語の数が非常に多い「開いた類（open class）」とされ，助詞や助動詞は固定的で属する語の数が少ない「閉じた類（closed class）」とされる（亀井・河野・千野，1996）。そのため，機能語自体の種類の少なさが，差が出なかった背景として考えられる。また，機能語は複合形式として現れ，単独の形態素を対象に算出した Guiraud 値では差が現れなかった可能性もある。橋本（2015）は，「いわゆる中級以降になると，単独の形態素という単位ではなく，既習の語を結合させた複合表現が，円滑にコミュニケーションをとるうえで，また，与えられたタスクを達成するうえで非常に重要となってくる」（p.82）と述べ，格助詞の「へ」が中位群や上位群になると「への」「へと」といった形で多く現れることを明らかにしている。今後は，複合助詞・助動詞での分析を検討する必要があるだろう。

5.2 B1 達成別にみた高頻度語と表現

ここでは，4.2 の分析から，達成群で特に特徴的であった以下の表現を中心に取り上げる。

高頻度形態素	:から（格助詞），よ（終助詞），こと，V_こと（名詞-非自立），とても（副詞），たくさん，近く（名詞-副詞可能）
動詞表現（助詞後続型）	:V-てみる
動詞表現（助動詞後続型）	:V-たい，V-た-N，V-ない
文末表現	:V-ますよ。V-ましょう。V-ています。Adj-です。

作文データから実際の用例を観察し，これらがどのような背景で B1 達成群に多く現れたのかを考察する。また，話し言葉の特徴を分析した山内（2009，2015）を参考に，書き言葉を分析した本稿の結果を比較することで，特徴的な言語項目の解釈を試みる。なお，用例は全て達成群のものであり，二つ目の書き起こし（形態素解析用に修正を加えたもの）から引用している。

5.2.1 タスク特有の表現

高頻度形態素の「から（格助詞）」と「近く（名詞-副詞可能）」の用例を観察すると，どちらも「家」と共に用いられているケースが多い。「から」で最も多かったのは「家から」のパターンで，「近く」ではほとんどが「家の近くに」という形で使用されていた。以下に用例を示す。

(1)　私の家から近くて，毎週末たくさん人が走っています。
　　家からそこまでバイクで 30 分もかかります。
　　サッカークラブは家から少し遠いですが，プールは歩いていくことができます。

(2)　私の家の近くにマンダレー山があってそこの周りに広いサッカー場があります。
家の近くに大きなスポーツセンターができて，いろいろなスポーツできるようになりました。
今も家の近くにあるクラブで週に二回ぐらいは行っています。

これらを観察すると，作文の課題である「その町で趣味のスポーツができるか」という質問に対して，自分の家を基点として関連する場所や施設を紹介・説明する記述が多い。

このように，B1 達成群は課題に関連した場所や状況を具体的に言語化する傾向がある。この傾向は，「たくさん（名詞-副詞可能）」「とても（副詞）」「Adj-です。」を使った表現でも同様であった。

(3)　この公園は広くて，木や花がたくさんあります。
もちろんリオでランニングができますけれども夏はとても暑いので気をつけてください。
12 時になったら日はとても強くなりますから危ないです。

文末表現の「V-ています。」も，書き手のスポーツ習慣や町の状態を説明する際での使用が多い。これも，達成群に特徴的な具体的説明に資する表現として考えられるだろう。

(4)　私はランニングをしませんが毎朝海に行って運動したり歩いたりしています。
土日は休みなので平日だけできることになっています。
毎日，普通は午後 1 時から夜まで雨が降っています。

B1 達成群に「よ（終助詞）」「V-ますよ。」が多く現れた背景としては，本課題における評価の観点であった「読み手意識」への影響が考えられる。「V-ましょう。」も同様で，以下のように多くは「一緒に V-ましょう。」の形で読み手を勧誘する用法として使用されていた。

(5)　私の町でたくさんきれいなプールがありますから一緒に水泳をしましょう。
友達はみんな参加するから私も大会に参加する予定だから一緒に走りましょう！
国で自転車も安く借りられるから，興味があったら一緒にサイクリングもしましょう。

誘いかけの用法では，他に「V-ませんか。」「V-てみませんか。」などもあり（表 7），いずれも達成群と不達成群の使用率の差が 10 %以上となっている。助詞後続型の動詞表現で特徴的であった「V-てみる」は，「てみませんか」の他に，「てみましょう」「てみてください」「てみましょうか」「てみたらどうですか」など，こちらも読み手を意識した表現として現れていた。

「てみる」は，本稿のデータでは B1 達成群の過半数で使用されていたが，山内（2009, 2015）では OPI の中級・上級話者による用例は僅かであることが報告されている。これは，タスクの性質やモダリティ（話し言葉か書き言葉か）が影響した可能性がある（橋本，2015）。達成群に多かった「から」「近く」等も，「家から」「家の近くに」として，当該タスクの達成に効果的かつ特有と思われる

表現として現れていた。そのため，1 つの作文タスクのみを分析した結果をもって，B1 としての言語特徴を一般化するのは難しいだろう。B1 レベルを反映する言語特徴は，多様で豊富なデータから多角的に検討する必要がある。

5.2.2 「活用型文法形態素」の使用

庵（2015）は，初級文法シラバスを作成するにあたって，活用のない項目を「Step1（初級前半）」として整理し，「Step2（初級後半）」との境界に活用の有無を設定している。山内（2009）も同様に，「KY コーパス」の分析結果に基づいて「普通形を作らなくても済む文法」を初級とし，述語に助動詞や接続助詞を後続させるような活用は初級文法と中級文法の分水嶺だと指摘している。

山内（2015）は動詞を活用させる必要がある文法形態素を「活用型文法形態素」とよび，接続助詞や助動詞等をこれに含めている。本データで B1 達成群に特徴的な表現であった接続助詞の「て」や助動詞の「ない」「たい」「た」，および，動詞に接続する形式名詞「こと」は，いずれも「活用型文法形態素」に該当する。このことから，B1 は「活用型文法形態素」が使えつつある段階，といえるのではないだろうか。

「V-て」自体は不達成群でも高い使用率（86.2 %）を示すものの，その接続形式のバリエーションは達成群に豊富で（「V-てから」「V-ても」「V-てくる」「V-てみる/た」「V-ているので」「V-ていますから」など），これらの使用率も不達成群から中位群を経て達成群へ段々と増加している（表 5）。つまり，B1 達成群は，テ形のバリエーションを使いこなしつつある段階といえる。

「V-ない」は，達成群が不達成群を 26.9 %も上回る一方（表 6），「V-ません。」では不達成群の使用率が達成群を 22.1 %上回っていた（表 7）。山内（2009）は，OPI の初級・中級では「ません」が多く，上級・超級では「ない」が多いことから，「後ろに助動詞を付加することができない丁寧形の使用から，助動詞を付加することができる普通形の使用に変わっていっている」（p.34）と指摘している。同じ否定表現である両者で達成群と不達成群に違いが現れたのは、以上の背景を考えることができる。

「V-たい」は，以下のように「V-たいと思います。」の用例が比較的多かった。山内（2009）は，「と思います」を用いた述べ立て的な発話は OPI 上級話者の特徴であるとしているが，これも 5.2.1 と同様，タスクやモダリティの違いを考慮する必要があるだろう。

(6) 怪我をしてきっかけで始めてみたけど，ずっと<u>続けたいと思います</u>。
もし良かったら，私も一緒に1回やって<u>みたいと思います</u>。
ですから，ぜひ一緒にランニングを<u>したいと思います</u>。

「こと」を用いた表現の多くは，「V-ことができる」の形で使われていた。これは，課題文に「趣味であるスポーツができるか」と示されていることから，その回答として「できる」という動詞と結びついて頻出したものと思われる。また，「V-た-N」で最も多いのも，「V-たこと」であった。

(7)　練習するために体育館もスタジアムも使うことができます。
テニスが好きじゃないなら別のスポーツもやってみることができます。
バレーボールが好きですか。やったことがありますか。

以上から，B1 達成群には活用が関連する文法形態素の多用が顕著であることが明らかとなった。B1 レベルは言語表現をより充実させるための文法的な操作に習熟しつつある段階，ということを示唆している。

6. まとめ

本稿では，JF 日本語教育スタンダード B1 レベルの作文課題への解答データ 110 件を対象に，達成群ではどのような言語特徴がみられるのかについて，量的・質的側面から検討を行った。その結果，達成群では不達成群に比べて記述量の全指標，および，語彙多様性の指標である Guiraud 値が有意に高いことが明らかとなった。また，動詞の後続表現や文末表現では，達成群で頻出するいくつかの特徴的な形式がみられた。それらは，本タスクに特有と思われる表現のほか，「活用型文法形態素」の多用という，先行研究の結果と整合する表現もみられた。

一方，誤用に関する分析が行えなかったため，不達成群で使用率が高くても実は誤用であり，達成群になると誤用が減っていくというような特徴は分析することができなかった（English profile programme における「ネガティブな基準特性」）。より精緻な分析のため，今後はこのような側面からの検討も求められる。

本稿は一つのタスクによって B1 としての下限の特徴の一部分を示したものであるため，これらの結果をもって B1 の言語特徴を一般化することはできない。今後，多様なタスクやモダリティによるデータを拡充し，さらに分析を進めることで，「JFS の各レベルではどのような文法や語彙が使えるようになるのか」が明らかになるだろう。言語活動を下支えする具体的な言語能力が明らかになることで，JFS はさらに日本語教育の文脈化に貢献でき，より有効な学習・指導・評価につながるだろう。

引用文献

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*, Cambridge: Cambridge University Press.

独立行政法人国際交流基金（2017a）『JF 日本語教育スタンダード 【新版】利用者のためのガイドブック』東京：独立行政法人国際交流基金.

独立行政法人国際交流基金（2017b）「JF 日本語教育スタンダードに基づいた評価と日本語能力試験の合否判定との関係－最終報告書－」（https://jfstandard.jp/publicdata/ja/render.do）.

橋本直幸（2015）「書き言葉コーパスから見た文法シラバス」庵功雄・山内博之（編）『データに基づく文法シラバス』（pp. 67-86）東京：くろしお出版.

Hawkins, J. A., & Filipović, L. (2012). *English profile studies 1 Criterial features in L2 English*. Cambridge: Cambridge University Press.

庵功雄(2015)「日本語学的知見から見た初級シラバス」庵功雄・山内博之(編)『データに基づく文法シラバス』(pp. 1-14)東京:くろしお出版.

石田基弘・小林雄一郎(2013)『R で学ぶ日本語テキストマイニング』東京:ひつじ書房.

石川慎一郎(2012)『ベーシック コーパス言語学』東京:ひつじ書房.

亀井孝・河野六郎・千野栄一(1996)『言語学大辞典 第 6 巻 述語編』東京:三省堂.

金澤裕之(編)(2014)『日本語教育のためのタスク別書き言葉コーパス』東京:ひつじ書房.

森本由佳子・塩澤真季・小松知子・石司えり・島田徳子(2011)「コミュニケーション言語活動の熟達度を表す JF Can-do の作成と評価:CEFR の A2・B1 レベルに基づいて」『国際交流基金 日本語教育紀要』, 7, 25-42.

中俣尚己(2015)「生産性から見た文法シラバス」庵功雄・山内博之(編)『データに基づく文法シラバス』(pp. 109-128)東京:くろしお出版.

櫻井直子・東伴子(編)(2011)『CEFR B1 言語活動・能力を考えるプロジェクト 2010 年度活動報告書』(http://japanologie.arts.kuleuven.be/bestanden/FINAL%20REPORT%20FOR%20ONLINE.pdf).

投野由紀夫(編)(2013)『CAN-DO リスト作成・活用 英語到達度指標 CEFR-J ガイドブック(CD-ROM 付)』東京:大修館書店.

van Ek, J. A., & Trim, J. L. M. (1991). *Threshold 1990*, Cambridge: Cambridge University Press.

渡部倫子(2015)「教師から見た文法シラバス」庵功雄・山内博之(編)『データに基づく文法シラバス』(pp. 129-145)東京:くろしお出版.

山内博之(2009)『プロフィシエンシーから見た日本語教育文法』東京:ひつじ書房.

山内博之(2015)「話し言葉コーパスから見た文法シラバス」庵功雄・山内博之(編)『データに基づく文法シラバス』(pp. 47-66)東京:くろしお出版.