



研究論文

学習者コーパスと言語テスト

言語テストの得点と作文のテキスト情報量の関連性

李 在鎬* 宮岡 弥生† 林 炫情‡

本研究は、学習者コーパスの開発における言語テストの有用性を明らかにするものである。我々は独自に開発した「日本語学習者作文コーパス」を使い、言語テストに基づく日本語能力の測定結果と作文の達成度に基づくテキスト情報量が、どの程度に対応するかを調査した。調査においては、クラスター分析と判別分析を使用した。作文のテキスト情報量から日本語能力グループを予測させる判別分析を行ったところ、90.6%の精度で観測値と予測値が一致した。また、作文のテキスト情報量を分析した結果、内容語の使用頻度、語数、文数が日本語能力グループの判別に大きく貢献することが明らかになった。

キーワード：学習者コーパス、文法・語彙テスト、多変量解析

1. 背景と目的

言語テストは、教授内容の定着度を測るためだけでなく、日本語教育の研究用データベース構築においても不可欠な要素である。例えば、学習者コーパスを開発する場合、学習者による産出データと当該学習者の日本語能力が分かるように設計する必要がある。というのは、学習者集団間で検索語の使用頻度に差が見られた場合、その集団の平均的な日本語能力や母語や学習歴などの情報を参考しながら現象を捉えなければならないからである。中でも日本語能力に関する情報は、学習者コーパスの利用者にとっては、もっとも基礎的かつ有用な情報である。一方、コーパス開発者にとっては、もっとも判断が難しい

項目である。

これまでの学習者コーパスは、日本語能力の認定が主観的な評価枠組みに基づいて行われたり、学習歴などの関連情報から日本語能力を推測したりするなど、必ずしも客観的とは言えない方法が用いられてきた。こうした課題を踏まえ、我々は「日本語学習者作文コーパス (Japanese Learner's Written Composition Corpus, 以下 JC Corpus)」の開発において、1) 言語テストを積極的に導入し、客観的な指標として日本語能力を明示すること、2) ユーザーインタフェースの中でコーパスデータと学習者の日本語能力などの情報を関連づけることを目指し、データベース構築を行った。1)の実践策として言語テストの結果を多変量解析の方法で分析し、コーパス開発においても活用する。2)の実践策としてキーワード検索などで日本語能力の情報をユーザーが選択し、表示する情報を絞り込むための独自のウェブインターフェースを開発する。

* 筑波大学人文社会系。

E-mail: lee@intersc.tsukuba.ac.jp

† 広島経済大学経済学部

‡ 山口県立大学国際文化学部

以上の背景を踏まえ、本研究は次のような研究課題を設定した。それは、言語テストの得点と学習者の産出データである作文の関連性を明らかにすることである。具体的には「JC Corpus」を利用し、言語テストに基づく日本語能力の測定結果と作文の達成度に基づくテキスト情報量（内容語の使用頻度、延べ語数、平均文長などのテキストが有する定量的な情報）が、どのように対応するかを調査した。方法として、テスト得点をクラスター分析し、日本語能力の所属グループ「上・中・下」を決定した後、判別分析を使って作文のテキスト情報量から日本語能力の所属グループを予測する統計的分析を行った。

本稿の構成は以下の通りである。まず2章では、学習者コーパスとはどのようなものであり、どのような課題があるかについて概観する。特に日本語能力の認定に関わる方法論的問題について述べる。3章では、「JC Corpus」の概要を示した上で、コーパス開発に使用した言語テストや統計処理の手順について述べる。4章では、クラスター分析と判別分析の結果を報告した上で、日本語能力の相違と作文のテキスト情報量の関連について考察を行う。最後に5章では、まとめと課題について述べる。以上の考察を通して、学習者コーパスにおけるレベル判定の必要性と言語テストの有用性を検討したい。

2. 学習者コーパスにおける現状と課題：なぜ言語テストが必要か

「コーパス (corpus)」とは、大規模なテキストデータベースのことを言い、それを利用した言語研究の手法を「コーパス言語学 (corpus linguistics)」と言う。一般にコーパスと言えば、母語話者のデータを集積したもの指すことが多いが、第二言語習得および外国語教育の分野では、学習者による産出データを集め、「学習者コーパス」というカテゴリーでデータベース化し、資源の共有化を図っている。

学習者コーパスは、大きく分けて2種類存在する。
 1) 会話などの話し言葉を集めたもの、2) 作文などの書き言葉を集めたものである。日本語においては、1) の例として、「KY コーパス」(鎌田, 山内, 1999) が広く知られている。2) の例として、国立国語研究

所による「日本語学習者による日本語作文と、その母語訳との対訳データベース」が広く知られている(李, 石川, 砂川, 2012 参照)。こうした学習者コーパスを利用することで、どのような背景の学習者が、どのような語彙ないしは文法項目に困難を感じるかを具体的に調べることができる(Granger, 1998; 石川, 2008)。また、中間言語の実態に関わる側面として、母語干渉や化石化などの具体的な現象について一般化・明示化ができる(李, 石川, 砂川, 2012)。そして、教育の現場レベルにおいても、シラバス作成やテスト問題の作成に役立てることができるため、多方面において潜在的な価値があると言える¹。

以上のような利用価値が認められるにも関わらず、日本語における学習者コーパスの現状は、量と質のいずれにおいても極めて限定的なものと言わざるを得ない。その原因はいくつか考えられるが、最大の原因は開発のコストである。学習者コーパスにおいては、データの収集から文字化、利用許諾などの権利処理、電子化、検索システムの開発、習熟度の判定など多種多様な作業が必要になる²。さらに、研究利用を想定した場合、何らかの均衡性をもたせる必要があり、精緻なコーパスデザインが求められる。単なる学習者データの集積と学習者コーパス構築は異なる点を認識しなければならない。例えば「KY コーパス」の場合、90名の話し言葉データを収録しているが、3つの軸で収録データの均衡性を確保している。1つ目は「学習者の母語」、2つ目は「OPIに基づく日本語能力判定」、3つ目は収録データのサイズである。母語は、中国語、英語、韓国語母語話者のデータが30名ずつ収録されている。日本語能力は、ACTFL (全米外国語教育協会: The American Council on the Teaching of Foreign Languages) による OPI (Oral Proficiency Interview) に基づいて判定を行い、初級 15 名、中級 30 名、上級 30 名、超級 15 名で構成している。データサイズについては、インタビューテストの収録時間を 30 分以内と制限して作成しているため、一つのファイル内に収録さ

1 コーパスに基づくテスト開発については、Barker (2004), 李 (2011), 根岸 (2012) を参照してほしい。

2 学習者コーパスの開発手順については李, 石川, 砂川 (2012) を参照してほしい。

れた発話量はほぼ統一されている(李, 2009; Lee & Nakagawa, *in press*)。このように学習者の母語、日本語能力、インタビューテストの収録時間を統制することで、言語表現に見られる差が母語による相違なのか、日本語能力による相違なのかが適切に判断できる。

学習者コーパスを使った研究では、調査結果を解釈する際に、母語や学習者の日本語能力を考慮に入れる必要がある。例えば、特定の言語現象を調べた結果、誤用例の出現頻度に何らかの偏りが見られた場合、その原因がいわゆる母語の影響なのか、日本語能力の影響なのか、学習環境の影響なのか、性差の影響なのかなど、多様な要因を総合的に検討しなければならない。母語や性別、学習環境などについては、学習者の自己申告で何らかの回答を得ることができるであろう。ところが、日本語のレベルに関しては、自己申告に基づく方法では限界があるため、コーパスの開発者側が何らかの形で判定を行わなければならない。

会話などの話し言葉に関しては、OPIなどの枠組みを利用し、何らかの日本語能力の判定ができる。しかし、作文などの書き言葉に関しては、汎用的に利用可能な枠組みが確立されていない(由井, 2009)。こうした状況から、これまで日本語の学習歴などの情報から、利用者が主観的に判断したり、あるいは何らかの客観試験の合否から言語能力を推測するなどの方法が用いられてきた。しかし、このような方法には次のような問題点が存在する。まず、学習歴が長いからといって必ずしも言語能力が高いとは言えない。また、客観試験の合否情報を利用する方法においても、データ提供者すべてが同時期に同じ試験を受けているわけではないので、等化をしない試験であれば、同程度の得点であっても、同程度の日本語能力であるかどうかは保証されない。さらに、受験時期とデータ収集時期で数年単位の時間差が存在する場合、その点数は、必ずしも信頼できる情報とは言えないなどの問題点が存在する。

以上を踏まえ、我々が開発を行った「JC Corpus」では、データ収集時に「語彙テスト(36問)」と「文法テスト(36問)」を実施し、学習者の日本語能力を測定した。データ収集時に学習者の日本語能力を測定することには、次のような利点がある。1) データ収集時期と言語テストの実施時期に時間差がない。2) データ提供者全員の言語テストデータが揃う。3) 50分ですべてのテストが実施できるため、学習者への負担が少ない。これらの利点を生かし、作文と言語テストの結果をダイレクトに結びつけることを目指した。こうした試みの妥当性を検証すべく、以下では、「語彙テスト」と「文法テスト」の得点が作文データとどの程度対応しているかを調べた。

3. データと分析方法

「JC Corpus」は、日本語学習者の作文をデータベース化したものであり、データ収集と作文執筆は以下のような手順と条件で行っている。

- 1) データ収集手順
 - (ア) 内容説明・誓約書作成：調査の趣旨説明。誓約書の内容確認と同意(10分)
 - (イ) 作文収集(日本語)：日本語で作文を執筆(50分)
 - (ウ) 作文収集(母語)：母語で日本語作文の対訳を学習者自身が作成(30分)
 - (エ) 語彙テスト・文法テストの実施(50分)
- 2) 作文執筆における条件
 - (ア) 辞書の使用：認めない
 - (イ) 執筆方法：原稿用紙を使用し、手書きで執筆
 - (ウ) 語数制限：300字～800字
 - (エ) 執筆テーマ：外国語が上手になる方法

以上の方法で、手書き作文を収集した後、電子化を行った。電子化の際には、単純なテキストファイルを作るだけでなく形態素解析を行い、単語を認定する作業や日本語教師による誤用例の添削、形態素解析のミスの修正なども並行して行った。現在、178名の作文の電子化が完了しており、ウェブサイトにてプレビュー版(<http://sakubun.jpn.org/>)を公開している。データ提供者(学習者)の性別、学習歴、母語別の内訳を表1に、「JC Corpus」のサイズを表2に示す。

表1. 「JC Corpus」の収録データ

母語	学習歴	女性	男性	総計
中国語母語話者	2年未満	2	2	4
	2年以上5年未満	56	24	80
	5年以上	2	0	2
韓国語母語話者	2年未満	10	10	20
	2年以上5年未満	34	26	60
	5年以上	2	10	12
	総計	106	72	178

表2. コーパスサイズ

学習歴	延べ頻度		一作文あたりの平均	
	文字数	語数	文字数	語数
2年未満	11815	6984	492.3	291.0
2年以上5年未満	76771	46500	548.4	332.1
5年以上	7924	4963	566.0	354.5
総計	96510	58447	542.2	328.4

さて、作文執筆の後、宮岡、玉岡、酒井(2011)の「語彙テスト」および「文法テスト」に、若干の修正を加えて新たに作成した「語彙テスト」と「文法テスト」を実施した。テストの所要時間は、「語彙テスト」と「文法テスト」を合わせて50分である。問題数は両テストとも36問で、四肢選択式の空所補充問題である。「語彙テスト」は、名詞12問、形容詞12問、動詞12問から成っており、それぞれ4問ずつの和語、漢語、外来語で構成されている。さらに、和語、漢語、外来語の各4問は、『日本語能力試験出題基準』(国際交流基金・日本国際教育支援協会, 2002)の1級から4級までの各級1問ずつから成っており、品詞、語種、日本語能力試験配当級に偏りのないように統制されている。「文法テスト」は、形態素変化(動詞の活用と接続に関するもの)の問題が8問、局所依存(助詞と動詞の関係のように、ターゲットと隣接する語とのつながりが理解できているかどうかを問うもの)が8問、構造の複雑性(主述の一致のように、文全体を読まなければ解けないもの)が8問、機能語(『日本語能力試験出題基準』で「機能語」に分類されているもの)が12問である。形態素変化、局所依存、構造の複雑性の各8問は、それぞれ3級と4級が同数の4問ずつから

成っている。機能語は、1級と2級がそれぞれ6問ずつである。以上のように、「語彙テスト」と「文法テスト」の問題を、特定のカテゴリーやレベルに偏りのないよう統制した。データ提供者全員(178名)にテストを実施した結果、「語彙テスト」の得点の平均値は24.03、標準偏差は8.71で、「文法テスト」の得点の平均値は23.26、標準偏差は8.82であった。「語彙テスト」および「文法テスト」全体の得点について、信頼性係数(クロンバッックの α 係数)を算出した³。本研究で実施したテストの信頼性係数は、語彙テストが0.83、文法テストが0.80といずれも高く、高い信頼性を有していると言える。

この結果を踏まえ、作文とテストの対応を検証すべく、以下の2つの分析を行った。

分析1) 日本語能力を判定するため、テスト得点をもとにクラスター分析を行った。

分析2) 日本語能力と作文のテキスト情報量の関連性を検討するため、(クラスター分析の結果として得られた)日本語能力を従属変数に、テキスト情報量を独立変数にして判別分析を行った。

分析1) では、語彙テストと文法テストの得点をもとに、階層的クラスター分析を行った。この分析は、日本語能力に関する属性、すなわち日本語能力として、上のグループなのか、中のグループなのか、下のグループなのかを決めるためである。また、クラスター分析を用いたのは次の理由からである。分析者の判断で足切りを設定し、上中下を決める方法では、どこからどこまでを下にし、中にし、上にするかに対して主観的にならざるを得ない。しかし、クラスター分析を用いた場合、得点の分散をもとに母集団における最適なサブグループを見つけることができる。しかも、デンドログラムを見ることでサブグループが形成されるプロセスが透明化されるため、データの理解にとっても役立つと考えられる。クラスター法は分類精度が良いとされるWard法を使用し、距離定義は平方ユークリッド距離を用いた。

分析2) はテストの得点によって決定された日本

3 信頼性とは、そのテストを用いて同一の受験者を測定した場合、どれだけ安定した結果が得られるかを検討するものである(中村, 2002)。信頼性係数は0.000から+1.000までの数値となり、1.000に近いほど信頼性の高いテストであると解釈できる。

語能力と作文の特徴(一文の長さや品詞の出現頻度)がどのように関連するかを調べるために行った。日本語能力に相当する従属変数がカテゴリカル変数であるため、変数同時投入法による判別分析がもっとも良いと判断した。判別分析の前に、作文の達成度を評価する指標として、テキストマイニング⁴の観点や日本語教育の観点も取り入れ、次の3種類の変数群を設定した。テキストマイニングの観点から、1)構成要素の産出能力に関連づけられる変数群として語種の出現頻度や名詞・動詞・形容詞の内容語の頻度(変数群A), 2)文章の長さに関連づけられる変数群として平均文長や語数などの指標(変数群B), 日本語教育的な観点から3)「JC Corpus」に含まれる学習者の「語彙、文法、文体」に関わる誤用の頻度(変数群C)の3種類である⁵。厳密な対応ではないが、変数群Aは日本語学習者の語彙的能力、変数群Bは日本語学習者の文法的能力が(作文の中に)反映されるであろうと想定した。変数群Cは学習者言語の表現の正確さが反映されるであろうと想定した。判別分析は、3つの条件で行った。

条件1) 従属変数として(テスト得点をクラスター分析した結果としての)日本語能力、独立変数として変数群Aと変数群Cを投入し、判別分析を行う。

条件2) 従属変数として(テスト得点をクラスター分析した結果としての)日本語能力、独立変数として変数群Bを投入し、判別分析を行う。

条件3) 従属変数として(テスト得点をクラスター分析した結果としての)日本語能力、独立変数として変数群A、変数群B、変数群Cを投入し、判別分析を行う。

3つの条件における判別率を比較することで、日本語能力を予測する最適な条件を明らかにする。

4 テキストマイニングとは、大量の文章データを自然言語処理の技術や統計的な分析モデルを使い、有益な知識を獲得する処理・操作を指す。

5 「JC Corpus」のデータベース内には、作文データに対して、3種類の誤用タグが挿入されており、誤用タグをもとにデータ抽出ができる。1) 語彙に関連する誤用(例: 私自身はそういう情況が時々発生します), 2) 文法に関連する誤用(例: 習うの後は必ず復習がある), 3) 文体に関連する誤用(例: 韓国も漢字を使うけど、漢字をそのまま使わないで音だけ使う)である。

4. 結果と考察

4.1. 分析1の結果(テスト得点の分析)

クラスター分析は、データセグメンテーションに特化した強力な手法であるが、クラスター数の決定が難しいという問題点がある。この問題を解決するため、本研究ではクラスター分析によって得られたクラスターを従属変数、クラスター分析で使った変数を独立変数にして判別分析を行った。そして、交差済み判別率を確認し、もっとも良い判別率を示したクラスター数を正解とする方法で分析を行った(cf. Romesburg, 2004)。

まず、「語彙テスト」と「文法テスト」をもとに階層的クラスター分析をし、クラスター数を3つから6つに指定した場合、各学習者がどの所属グループに分類されるか解析した。次に、クラスター分析によって得られた所属グループを従属変数に、「語彙テスト」と「文法テスト」の得点を独立変数にして判別分析を行った。判別分析は3つから6つのクラスター数に対応し、4回行った。そして、交差済み判別率を確認したところ、表3に示す結果が得られた。

表3. 判別分析によるクラスター分析の精度検証

クラスター数	交差確認済み判別率*
C=6	94.9%
C=5	92.7%
C=4	96.1%
C=3	95.5%

交差確認はLeave-one-out cross validation法使用

* 交差確認済み判別率とは、交差検証法(cross validation)とも呼ばれる。判別分析の精度の過大評価を避けるためのもので、判別ルール作りのために使用した学習データと検証に使用した実験データを分け、実験データでもって判別をした時の判別率を意味する。すなわち、より厳しい基準で判別精度を評価する手法である。詳細は石川、前田、山崎(2010)を参照してほしい。

表3に示したように、クラスター数を4つにすることによって、もっとも高い判別率が得られることが明らかになった。4つのクラスターに分けた場合の度数表を表4に、得点率の分布を図1に示す。

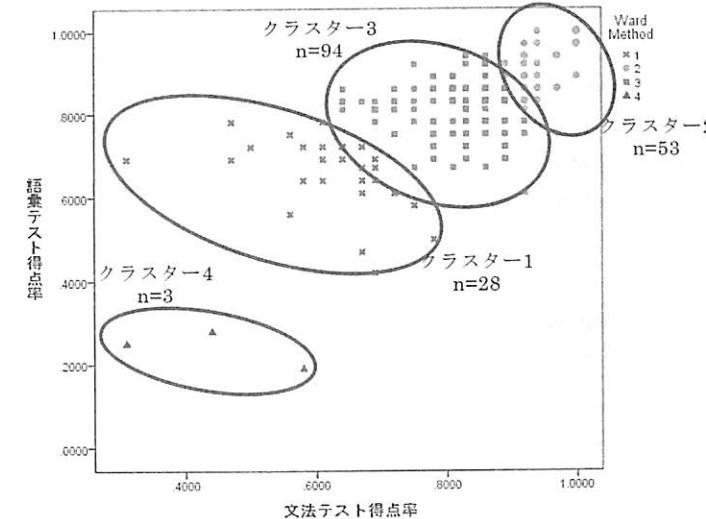


図1. 文法テスト×語彙テストの得点率の散布図

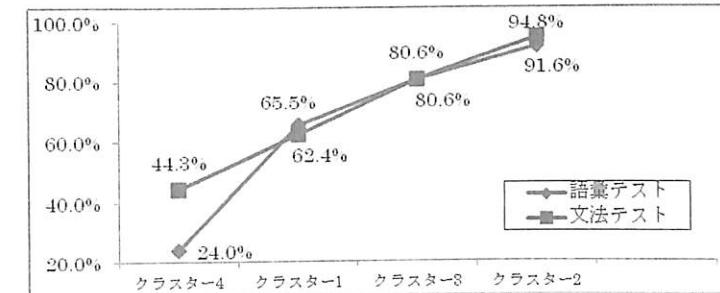


図2. クラスター別の得点率の平均値

表4. クラスター分析に基づく学習者の分類

日本語学習履歴	クラスター				合計
	1	2	3	4	
2年未満	5	4	13	2	24
2年以上 5年未満	22	41	76	1	140
5年以上	1	8	5	0	14
合計	28	53	94	3	178

図1では、所属クラスターに対応させる形で、語彙テストと文法テストの得点率を散布図として示している。そして、得点率の平均値を図2に示す。図2の得点率を確認すると、いずれのテストでもクラスター4の得点率がもっとも低く、クラスター2がもっとも高い。なお、クラスター間で平均値の相違を調べるべく、一元配置の分散分析を行ったところ、いずれも有意な差が確認された(語彙テスト [F(3, 174) = 157.916, p = .000], 文法テスト [F(3, 174) = 167.597, p = .000])。得点率の推移としては、「クラスター4 < クラスター1 < クラスター3 < クラスター2」の順で分布していることが明らかになった。

4.2. 分析2の結果(テストと作文の分析)

分析2では、分析1によって決定した日本語能力のグループを、作文そのものの達成度から予測した場合、どの程度の予測精度が得られるかを検討するための統計分析を行った。判別分析は、3章で述べた3つの条件で行ったが、クラスター分析による日本語能力値に関して、一点だけ調整を行った。具体的にはクラスター4を分析から除外した。除外した理由として二つある。1) 表4および図1の通り、データの個数が3つしかなく、独立したグループとして捉えられないこと、2) 極端に得点率が低いことか

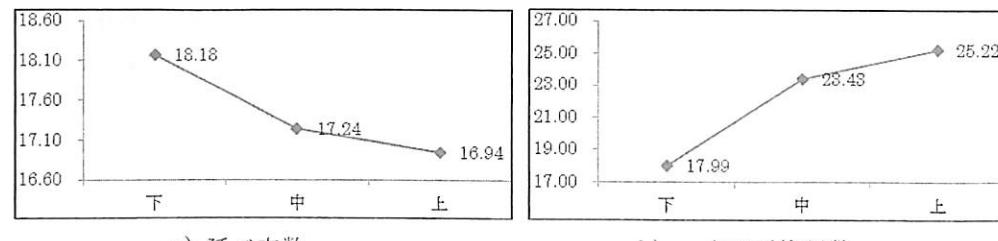


図3. テキスト情報量

ら、他のグループとの比較が意味をなさないことがある⁶。そして、最終的には、クラスター1を(日本語能力として)「下のグループ」、クラスター3を「中のグループ」、クラスター2を「上のグループ」として捉え、作文のテキスト情報量から日本語能力を予測する判別分析を行った。3つの条件下で行った分析結果を表5に示す。

表5. 条件別の判別率

区分	交差確認済み判別率
条件1	70.9%
条件2	81.7%
条件3	90.6%

交差確認はLeave-one-out cross validation法使用

条件1では、独立変数として変数群A(構成要素に関する変数として、語種「和語、漢語、外来語、混種語」の頻度、内容語「名詞・動詞・イ形容詞」の頻度)と変数群C(「文法・語彙・文体」誤用の頻度)を使用した。条件2では、変数群B(文章としての長さに関する変数として、平均文長、語数、文数、1単語の平均文字数)を使用した。条件3では、すべての変数を使用した。結論的には条件3でもっとも良い判別率が得られた。条件3による判別分析の分類結果は、表6のとおりである。

6 クラスター4の除外を決定する上、2点を考慮した。1)「下の下」(下のグループのさらに下の意味)というグループ設定をする、2)下のグループに含める。1)については分析結果をコーパス構築に利用するという本研究の目的から考えて、3名しかないサブグループを設定したところ、データ検索を行うコーパス利用者にとって実質的には意味がないと判断し、取り入れなかった。2)については、分析者の都合で恣意的な調整を行うことは避けるべきと考え、行わなかった。最終的には、3名のデータはコーパス全体からも除外し、コーパス公開を行う。

表6が示すこととして、元データの日本語能力を作文のテキスト情報量で予測させた場合、下グループと上グループに関しては、100%に近い精度で観測値と予測値が一致しており、誤分類のリスクはほとんど見られない。しかし、中グループに関しては、観測データ上は、中グループであるが、作文のテキスト情報量からは、上グループと予測された学習者が12名(12.8%)いた。この12名は、言語テストの得点としては、上のグループではないが、作文の達成度としては、上のグループに相当するものとして捉えられる。

次に表7でケースの判別のために使用された正準判別関数の係数を示す。この結果から、文や語の長さに関する特徴がデータ分類において顕著である一方、作文内の誤用の有無がデータの判別にそれほど大きく貢献していないことが明らかになった。

4.3. 考察

本節では、判別分析の結果を踏まえて次の2点について考察を行う。1点目は、長さに関する変数の寄与度をどのように考えるべきか、2点目は、誤用の出現があまり寄与していない事實をどのように捉えるべきかである。

どのような要素が判別に貢献するかという観点から見た場合、文字数に代表されるような文章の物理的な長さに関する要素が、日本語能力の判別には重要であることが明らかになった。これは、日本語教育における教師の直感とも矛盾しない。学習歴の短く、日本語能力が低い学習者の場合、数語程度で構成された短い発話をを行うのに対して、日本語能力が上がるにつれ、文法的に複雑な長い発話を産出するようになる。このことを表す具体的なデータを図3に示す。

図3のa)は一作文あたりの延べ文数の平均値、b)は一文の平均語数である。まず、下のグループに注

表6. 判別分析の分類結果

観測データ上の グループ	度数 (%)	予測グループ			合計
		日本語能力 下	中	上	
交差確認後の グループ	28 (100.0)	28 (100.0)	0 (0.0)	0 (0.0)	28
	1 (1.1)	86 (91.5)	7 (7.4)		94
	0 (0.0)	0 (0.0)	53 (100.0)		53
日本語能力 下	28 (100.0)	0 (0.0)	0 (0.0)		28
	2 (2.1)	80 (85.1)	12 (12.8)		94
	0 (0.0)	2 (3.8)	51 (96.2)		53

目した場合、作文全体は約18文で構成され、一文の平均語数は約18語である。一方、中のグループ

表7. 正準判別関数係数

	関数	
	1	2
語彙的誤用	-.134	.027
文法的誤用	-.069	.187
文体の誤用	.119	-.449
外来語	-.004	.099
漢語	-.422	.619
固有名詞	-.035	-.210
混種語	.046	.435
和語	-.762	1.202
内容語	.309	-2.303
文数	-.391	.925
語数	1.143	-.494
平均文長	-.350	.157
1単語の平均文字数	.032	.122

Wilks's Lambda = .161, p = .000

は、延べ文数17に対して、一文の平均語数は約23語、上のグループは延べ文数17に対して、一文の平均語数は約25語を用いている。このことから、下のグループは中・上のグループに比べ、少ない語数で構成された文を多く用いている可能性が示唆される。一方、上のグループの場合、一文を構成する語数は下・中のグループより多く、比較的長い文を書いている可能性が示唆される。

次に、誤用の産出頻度がデータ分類に寄与しなかった点について考察する。誤用の産出については、多くの先行研究で母語による差の問題が指摘されていることを踏まえ、図4の通り、母語別に集計を行った。

図4のa)は語彙に関する誤用、b)は文法に関する誤用、c)は文体に関する誤用である。いずれも、一作文あたりの平均値である。まず注目すべきこととして、学習が進むにつれ誤用が減ることが予測されるが、実際のデータは必ずしもそうではないということである。語彙の誤用に関して言えば、韓国語母語話者の場合、日本語能力があがるにつれ、減少する傾向が見て取れる一方、中国語母語話者の場

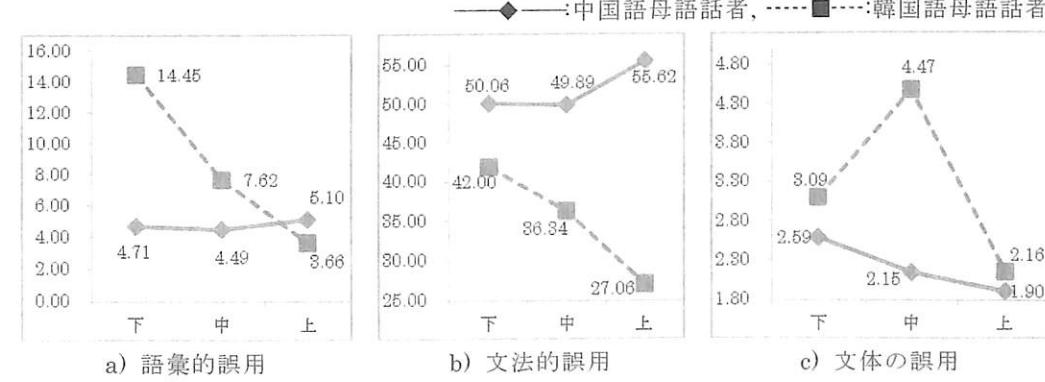
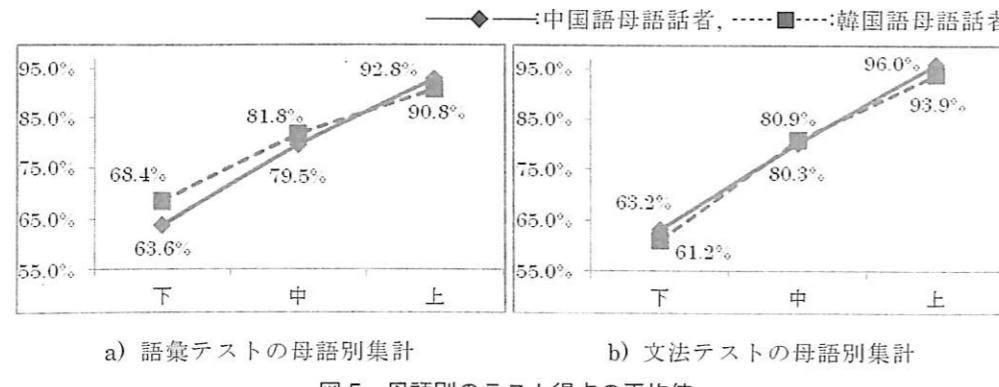


図4. 母語別の作文あたりの誤用の平均出現値



a) 語彙テストの母語別集計
b) 文法テストの母語別集計
図5. 母語別のテスト得点の平均値

合ほとんど変化がない。次に、文法の誤用に関して言えば、韓国語母語話者の場合、日本語能力があがるにつれ、減少する傾向が見て取れる一方、中国語母語話者の場合、下グループから中グループにいくにつれ、わずかに減ったものの、上グループでは上昇している。さらに、文体の誤用については、中国語母語話者の場合、緩やかではあるが、レベルが上がるにつれ減少する傾向がある。これに対して、韓国語母語話者の場合、増えてから減るという逆V字型の分布を見せており。これは、母語によって現れる誤用の頻出パターンが異なることを示すものである。なお、図4に関連するデータとして、母語別の成績を図5として示す。

母語による得点の相違を確認すべく、語彙テストと文法テストの母語別の平均値をt検定したところ、いずれも有意差は見られなかった（語彙テスト:t(176) = -1.560, p = .121, 文法テスト:t(176) = -1.106, p = .270）。図4の誤用の出現傾向が母語によって違うという事実と対照させて考えてみた場合、学習者の誤用というのは日本語の習熟度よりも母語の観点から捉えるべき現象であるという示唆が得られる⁷。

7 図5の分布はクラスター分析と判別分析を組み合わせ最適化されたものであるが、母語による極端な偏りが存在する可能性もゼロではない。そのため母語（韓国語・中国語）× レベル（上・中・下）による分散分析を行った。その結果、語彙テストにおいては、母語と日本語能力の交互作用は見られなかった ($F(2, 169) = 2.55, p = .081$)。そして、母語の主効果も認められなかった ($F(2, 169) = 1.977, p = .162$)。文法テストにおいても、母語と日本語能力の交互作用は見られなかった ($F(2, 169) = .78, p = .462$)。そして、母語の主効果も認められなかった ($F(2, 169) = .985, p = .322$)。

5. まとめと課題

本研究では、言語テストによって認定された日本語能力を作文のテキスト情報量から予測した場合、90.6%の予測精度が得られた。このことは、作文の達成度と言語テストの得点が深く結びついていることを示唆するものであり、学習者コーパス開発における言語テストを使った検証の重要性を示している。そして、作文のテキスト情報量を分析した結果、内容語の使用頻度、語数、文数が日本語能力の判定に大きく貢献することが明らかになった。これは、学習者コーパスのデータ分析において内容語の使用頻度、語数、文数といった要素が重要であることを示している。また、レベル情報が示されていないコーパスにおいて、内容語の使用頻度、語数、文数といった要素は、学習者のレベルを推定する指標の一つになりうる可能性を秘めていると言えるが、この点については今後もデータを蓄積し、更なる検証を進めていくことが必要である。

今後の課題として4点を述べる。1点目に、作文の達成度としては上グループであるが、テスト成績としては、中グループになった12名（12.8%）および下グループになった2名（2.1%）に関しては、テストの測定精度に問題があった可能性を検討しなければならない。問題項目の再検討など、テストの測定精度を向上させる必要がある。2点目に現在の「日本語学習者作文コーパス」は均衡性という意味では必ずしも十分でない。したがって、今後被験者数とデータ量のバランスを調整し、均衡性に配慮したコーパスを提供していきたい。3点目に、本研究は誤用分析を目的とするものではないため、誤用の

分布そのものには踏み込んだ考察は行っていないが、母語別の誤用の傾向と日本語能力との関係については独立した考察を行う必要がある。4点目に、本研究の調査結果によってテキスト情報量の一文の平均語数が多くなることが日本語能力のレベルの指標となることを明らかにしたが、質的な側面に対する考察が必ずしも十分ではなかった。今後は、具体的な文例を見ながら質的な側面に関する考察を行う必要がある。

文献

- 石川慎一郎 (2008). 『英語コーパスと言語教育』 大修館書店。
石川慎一郎、前田忠彦、山崎誠 (2010). 『言語研究のための統計入門』 くろしお出版。
国際交流基金・日本国際教育支援協会 (2002). 『日本語能力試験出題基準【改訂版】』 凡人社。
鎌田修、山内博之 (1999). 『KY コーパス』 http://opi.jp/shiryo/ky_corp.html
中村洋一 (2002). 『テストで言語能力は測れるか—言語テストデータ分析入門』 桐原書店。
根岸雅史 (2012). 英語学習の発達段階に応じたCEFR基準特性—英作文採点への応用可能性『コーパスに基づく言語学教育研究報告』 9. 105-127.
宮岡弥生、玉岡賀津雄、酒井弘 (2011). 日本語語彙テストの開発と信頼性—中国語を母語とする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』 34(1), 1-18.
由井紀久子 (2009). プロフィシェンシーと書く能力の開発『プロフィシェンシーと日本語教育』 (pp. 221-240) ひつじ書房。
李在鎬 (2009). タグ付き日本語学習者コーパスの開発『計量国語学』 27(2), 60-72.
李在鎬 (2011). 大規模テストの読解問題作成過程へのコーパス利用の可能性『日本語教育』 148, 84-98.
李在鎬、石川慎一郎、砂川有里子 (2012). 『日本語教育のためのコーパス調査入門』 くろしお出版。
Barker, F. (2004). Using corpora in language testing. *Modern English Teacher*, 13(2), 63-67.

Granger, S. (Ed.). (1998). *Learner English on computer*. London: Longman. (船城道雄、望月通子(監訳)(2008). 『英語学習者コーパス入門—SLAとコーパス言語学の出会い』 研究社.)

Lee, J., & Nakagawa, N. (in press). KY Corpus. In M. Minami (Ed.), *Handbook of Japanese applied linguistics*. New York: Mouton de Gruyter.

Romesburg, C. (2004). *Cluster analysis for researchers*. Florida: Robert E. Krieger. (<http://books.google.co.jp>で閲覧可能)

謝辞: 本研究はJSPS科研費22520537の助成を受けたものです。