

## 外国人日本語学習者 850 名の日本語能力から 説明文の品詞別産出頻度への因果関係<sup>1</sup>

玉岡 賀津雄<sup>2</sup>

DOI: 10.18999/stul.34.45

**要約:** SNS の普及にともない、書き言葉によるコミュニケーションが頻繁に行われるようになってきた。とりわけ、事実即した情報や依頼などをわかり易く書き言葉で伝達するための「説明文(expository writing)」を書く能力が要求されるようになってきた。この能力を「説明力(explanatory ability)」と呼ぶ。現代社会において、説明力は日本語母語話者ばかりでなく外国人日本語学習者の実生活でも、なくてはならない能力になっている。そこで、『多言語母語の日本語学習者横断コーパス: I-JAS』に収録された2つのストーリーライティング課題が説明文を反映していると仮定し、海外で学習する日本語学習者 850 名が書いたテキストデータを基に、(1)動詞, 助動詞, 副詞, 助詞の4つの品詞を「述部」とし、(2)名詞, 形容詞, 連体詞の3つの品詞を「名詞句」として、品詞別の産出頻度を算出した。そして、J-CAT (Japanese Computerized Adaptive Test) で測定された日本語能力から I-JAS のテキストの産出頻度への因果関係モデル(図1を参照)を5つ想定して、構造方程式モデリング(structural equation modelling, SEM)の解析法で日本語能力と産出頻度データとの適合度を検討した。適合度指標に照らして、最適の因果関係モデル(図2を参照)をみいだした。このモデルは、次のような連続的な因果関係を示した。(1)語彙と文法が「基礎力」を構成し、(2)読解と聴解の「理解力」を促進し、(3)「述部」の語彙産出を豊かにし、(4)「名詞句」の語彙産出に大きく貢献する。

**キーワード:** I-JAS, J-CAT, 説明力, 日本語能力, 産出頻度, 構造方程式モデリング

<sup>1</sup> English title: The causal relation between the Japanese language proficiency of 850 foreign learners and productive frequencies of lexical categories in expository text

<sup>2</sup> TAMAOKA, Katsuo, Nagoya University, Japan, E-mail: ktamaoka@gc4.so-net.ne.jp

## 1. 研究の目的

LINE, WeChat, Twitter など SNS (social networking service) の普及にともなうて、書き言葉が頻繁に使われるようになってきた。たとえば、お得意さんを訪問した印象を簡潔に報告する、ある業務についての遂行状況を伝える、などである。これらは、論理的に構成された作文や仮想的な小説とは異なり、事実に基づいた情報や依頼などをシンプルにわかりやすく伝達する「説明文」である (Nordquist, 2019)。こうした能力を「説明力」と呼ぶ。説明文を適切に書く能力は、ビジネスなど日常生活においても、相手に伝わる文章術として、重要であるとされている (池上, 2009; 工藤・松井, 2010)。そしてまた外国人日本語学習者の実生活においてもなくてはならない能力になってきた。日本語教育でも、こうした社会的なニーズに合わせて、外国人日本語学習者の説明文を適切に書く能力を育成することが急務となっている。そこで、『多言語母語の日本語学習者横断コーパス (International Corpus of Japanese as a Second Language, I-JAS)』(迫田・石川・李・佐々木・須賀・野山・細井・八木, 2020) に収録された2つのストーリーライティング課題のテキストが説明文を反映していると仮定した。そして、動詞、助動詞、副詞、助詞を「述部」を構成する品詞、名詞、形容詞、連体詞を「名詞句」を構成する品詞として、産出頻度を計算した。そして、語彙と文法を日本語の「基礎力」、聴解と読解を「理解力」として、述部と名詞句の語彙産出にどう影響するかを、構造方程式モデリングの解析法で検討した。

## 2. 日本語学習者の特性

### 2.1 日本語学習者の母語別の性別と年齢

I-JAS には、外国人日本語学習者 1,000 名と日本語母語話者 50 名の合計 1,050 名分のデータが集められている。まず、日本語学習者についての研究であるため、日本語母語話者 50 名を分析から除外した。さらに、日本国内の学習者 150 名を除いた。これは、彼らの母語が多様だったからである。母語の影響は、第二言語習得の主要な研究対象であり (Ellis, 1985)、日本語習得への影響も報告されている (玉岡, 2000; 張, 2011; 大和・玉岡, 2013 など)。そのため、母語を統制しないで、国内か海外かという学習環境の要因を議論することはきわめて難しい。本研究では、学習環境について検討しないこととして、海外で学習する日本語学習者に限定し、国内の日本語学習者については、除外した。最終的に、日本語

母語話者 50 名と日本国内で学習している日本語学習者 150 名の 200 名を除外して、850 名分のコーパスを使用した。年齢の報告がない3名を除いて、847 名の平均年齢は 22 歳 0 カ月 ( $SD$  は 2 歳 11 カ月,  $SD$  は標準偏差)であった。女性の平均年齢は 21 歳 7 カ月 ( $SD$  は 2 歳 7 カ月)で、男性は 23 歳 0 カ月 ( $SD$  は 3 歳 7 カ月)であった。847 名の年齢は 17 歳から 46 歳までの範囲で分布する。歪度と尖度をみると、かなりの偏りがあることがわかる。歪度は、左右対称性を示す指標であり、0 であれば完全に左右対称である。年齢分布は、12.78 ときわめて大きな正の値である。年齢の高い層に向かって大きく裾が広がった分布であることがわかる。また、尖度は、正規分布を基準としてどれだけ尖っているかを表す指標である。尖度は、正規分布であれば 0 になるが、値が負の場合は正規分布を基準として緩やかに尖っており、値が正の場合は鋭く尖っていることを示す。ここでは、2.68 と正の値であり、多くの日本語学習者が 19 歳から 23 歳の範囲に集まり、この年齢域で鋭く尖った分布になっていることがわかる。これは、海外の大学で日本語を専攻している学生が、I-JAS のコーパス作成に参加していることを反映しているからであろう。日本語学習者が特定の年齢域に集まっていることは、年齢の影響が統制されるので、日本語学習者の母語別の分析にはむしろ好都合であろう。母語別の年齢の平均と標準偏差および性別の人数は表 1 に示した。

表 1 I-JAS に掲載された海外で学ぶ日本語学習者 850 名の母語別の年齢と性別

日本語学習者 の母語	人数	年齢 (歳)		性別 (人)		
		$M$	$SD$	女性	男性	未記入
中国語	200	20.86	1.06	160	37	3
韓国語	100	23.54	2.60	50	50	0
英語	100	21.34	2.66	62	38	0
ベトナム語	50	21.46	1.84	42	6	2
タイ語	50	21.74	2.69	43	7	0
インドネシア語	50	20.86	0.78	40	10	0
トルコ語	50	21.76	3.62	30	20	0
スペイン語	50	25.64	5.12	29	21	0
フランス語	50	21.78	2.34	33	17	0
ドイツ語	50	23.66	3.63	29	20	1
ハンガリー語	50	23.66	3.14	35	15	0
ロシア語	50	20.28	1.85	41	6	3
全体または合計	850	22.00	2.94	594	247	9

注:  $N=850$ 。  $M$  は平均で、  $SD$  は標準偏差を示す。年齢未記入が、中国語、韓国語、英語母語話者の各 1 名の合計 3 名。それら 3 名は、年齢の平均の計算から除外した。

海外で日本語を学ぶ 850 名の学生は、女性が 594 名、男性が 247 名、性別の記載がない学生が 9 名であった。母語別の内訳は中国語母語話者が 200 名(中国本土 100 名と台湾 100 名)、韓国語母語話者が 100 名、英語母語話者が 100 名(アメリカ 38 名、イギリス 19 名、オーストラリア 23 名、ニュージーランド 20 名)、ベトナム語母語話者が 50 名、タイ語母語話者が 50 名、インドネシア語母語話者が 50 名、トルコ語母語話者が 50 名、スペイン語母語話者が 50 名、フランス語母語話者が 50 名、ドイツ語母語話者が 50 名(ドイツ 11 名、オーストリア 39 名)、ハンガリー語母語話者が 50 名、ロシア語母語話者が 50 名である。

## 2.2 日本語学習者の日本語能力

I-JAS のプロフィールデータから、外国人日本語学習者 850 名の日本語能力の指標を抽出した。第 1 に、Japanese Computerized Adaptive Test (J-CAT) で、インターネット(<http://www.j-cat.org/>)に接続して受験して日本語能力を自動判定する評価テストである。語彙、文法、聴解、読解の 4 つの下位テストから構成されている(赤木・今井, 2012)。J-CAT は、各下位テストが 100 点満点で、合計が 400 点満点で評価される。日本語学習者 850 名の平均は、語彙が 53.81 点( $SD=17.91$  点)、文法が 48.36 点( $SD=17.48$  点)、聴解が 54.13 点( $SD=17.52$  点)、読解が 47.04 点( $SD=13.83$  点)で、400 点満点の平均は 203.34 点( $SD=57.46$  点)であった。J-CAT の平均と標準偏差は、母語別に表 2 に示した。

母語別の得点の違いについては、一元配置の分散分析で検討した。語彙の得点では、母語の主効果が有意であった[ $F(11, 838)=60.46, p<.001, \eta_p^2=0.44, \eta_p^2$  は効果量]。中国語母語話者が最も得点が高く( $M=68.69$  点,  $M$  は平均)、スペイン語母語話者が最も得点が低かった( $M=35.06$  点)。得点差が 33.63 点もあり、ほぼスペイン語母語話者の平均得点にあたることを考えると、いかに差が大きいかが窺える。文法の得点についても、母語の主効果が有意であった[ $F(11, 838)=35.07, p<.001, \eta_p^2=0.32$ ]。韓国語母語話者が最も得点が高く( $M=62.95$  点)、スペイン語母語話者が最も得点が低かった( $M=34.76$  点)。得点差は、28.19 点であった。やはり大きな得点差がみられた。聴解の得点についても、母語の主効果が有意であった[ $F(11, 838)=30.44, p<.001, \eta_p^2=0.29$ ]。やはり韓国語母語話者が最も得点が高く( $M=70.81$  点)、スペイン語母語話者が最も得点が低かった( $M=37.58$  点)。得点差は、32.23 点という大きさであった。読解の得点についても、言語の主効果が有意であった[ $F(11, 838)=26.76, p<.001, \eta_p^2=0.26$ ]。韓国語母語話者が最も得点が高く( $M=58.02$  点)、スペイン語母語話者が最も得点が低かった( $M=36.38$  点)。得点差は、21.64 点であった。400 点満

点の全体の得点についても、母語の主効果が有意であった[ $F(11, 838)=56.17, p<.001, \eta_p^2=0.42$ ]。やはり韓国語母語話者が最も得点が高く( $M=257.88$  点)、スペイン語母語話者が最も得点が低かった( $M=143.78$  点)。得点差は、114.10 点という大きさであった。J-CAT の総合得点およびすべての下位テストの得点から、母語によって日本語能力に大きな違いがあることがわかった。

表2 I-JAS の日本語学習者 850 名の J-CAT の母語別平均、標準偏差と分散分析の結果

日本語学習者の母語	J-CAT語彙		J-CAT文法		J-CAT聴解		J-CAT読解	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
中国語	68.69	11.91	58.44	16.03	61.48	14.43	54.00	12.13
韓国語	66.10	16.03	62.95	18.28	70.81	16.28	58.02	14.64
英語	40.44	16.22	38.05	14.26	45.58	16.42	41.67	12.71
ベトナム語	54.16	9.95	47.60	12.25	47.42	13.32	46.20	11.51
タイ語	58.84	12.08	52.64	12.25	56.66	14.50	49.54	9.34
インドネシア語	45.86	10.55	43.62	10.65	45.28	12.33	38.80	11.69
トルコ語	38.44	17.55	35.64	15.05	47.50	18.60	38.54	12.38
スペイン語	35.06	12.11	34.76	14.16	37.58	12.89	36.38	11.45
フランス語	41.90	9.40	35.96	11.23	44.02	10.82	38.36	9.76
ドイツ語	49.24	10.40	44.78	9.50	52.70	14.35	43.72	10.63
ハンガリー語	54.62	12.94	49.04	14.38	57.98	16.75	47.78	11.38
ロシア語	48.78	17.23	42.38	15.23	52.38	14.79	44.96	10.84
全体	53.81	17.91	48.36	17.48	54.13	17.52	47.04	13.83
<i>F</i> 値	$F(11, 838)=60.46$		$F(11, 838)=35.07$		$F(11, 838)=30.44$		$F(11, 838)=26.76$	
有意確率と効果量	$p<.001, \eta_p^2=0.44$		$p<.001, \eta_p^2=0.32$		$p<.001, \eta_p^2=0.29$		$p<.001, \eta_p^2=0.26$	

注:  $N=850$ 。  $M$  は平均,  $SD$  は標準偏差, J-CAT は各 100 点満点。母語の人数は表 1 を参照。

第 2 に, SPOT (Simple Performance-Oriented Test) で, TTBJ (Tsukuba Test-Battery of Japanese) の 1 つとなっている。自然な速度の音声テープを聞きながら, 解答用紙に書かれた同じ文を目で追っていき, 文中の ( ) に聞こえた音(ひらがな 1 字)を書き込ませるテストである。穴埋め課題は 1 文につき 1 カ所である(小林・フォード・山元, 1996; 小林, 2015)。この種のテストはクローズ・テスト(cloze test)と呼ばれ, 1950 年代から総合的な言語力の判定に使われてきた(Taylor, 1953)。ただし, クローズ・テストは, たいていの場合書かれたテキストだけで実施されるが, SPOT はテキストを音声提示する点でユニークである。I-JAS で使用された SPOT は 90 点満点で, 850 名の得点は, 最低点が 30 点, 最高点が満点の 90 点であった。全体の平均は, 66.96 点( $SD=11.62$  点)であった。全体の得点については, 言語の主効果が有意であった[ $F(11, 838)=57.12, p<.001, \eta_p^2=0.43$ ]。SPOT でも韓国語母語話者

が最も得点が高く( $M=78.59$  点), スペイン語母語話者が最も得点が低かった( $M=56.42$  点)。得点差は, 22.17 点という大きさであった。やはり, 母語によって日本語能力に大きな違いがあることがわかる。これは, I-JAS に参加した日本語学習者には, もともと日本語能力に大きな差があることを示している。日本語能力が高いほど, 多様で豊かな表現ができると仮定するならば, 母語別に I-JAS コーパスのテキストデータを比較して分析する場合には, 日本語能力の違いをまず確認して, それを考慮して考察する必要があるだろう。

J-CAT と SPOT の総合得点のピアソンの積率相関係数は, 非常に高かった( $r=0.84$ ,  $p<.001$ )。両テストともに外国人日本語学習者の日本語能力を測定するために作られたテストであり, 日本語能力を測定するというテストの妥当性の高さを示しているといえよう。なお, J-CAT の4つの下位テストのピアソンの積率相関係数は, 語彙と文法が  $r=0.73$  ( $p<.001$ ), 語彙と読解が  $r=0.62$  ( $p<.001$ ), 語彙と聴解が  $r=0.66$  ( $p<.001$ ), 文法と読解が  $r=0.59$  ( $p<.001$ ), 文法と聴解が  $r=0.64$  ( $p<.001$ ), 聴解と読解が  $r=0.66$  ( $p<.001$ ) で, すべての相関係数が 0.1 パーセントレベルで有意であり, 高い相関を示した。なお, 年齢と J-CAT および SPOT との相関は, J-CAT の下位テストを含むすべての条件で  $r=.06$  ( $ns$ ) 以下と低く, 有意ではなかった。日本語学習者 850 名の年齢は, 一部の高年齢者を除いて, 平均が 22 歳 0 カ月で, 標準偏差が 2 歳 11 カ月であることからわかるように, 19 歳から 25 歳の年齢域に多くの参加者が集まっている。

### 3. 日本語能力および説明文産出頻度の因果関係

#### 3.1 構造方程式モデリングとは

構造方程式モデリングは, 共分散構造分析(covariance structure analysis)とも呼ばれ, 重回帰分析, 因子分析, パス解析などの機能を統合した多変量解析である。直接には観測できない変数である潜在変数(latent variable)を, 複数の観測変数(observed variables)で構成して, 潜在変数間あるいは観測変数を含んだ変数間の因果関係を分析する(朝野・鈴木・小島, 2005; 南風原, 2002; 狩野, 2002; 狩野・三浦, 2002; 小塩, 2008; 豊田, 1998, 2007; Hox, 2010, 初・玉岡, 2013 など)。従来の統計解析では, 仮説として想定したモデルが正しいかどうかは検証されなかった。構造方程式モデリングでは, モデルが, 実際に集めたデータにどの程度よく当てはまっているかを適合度指標によって確認することができる。そして, 最適のモデルをみいだして, 想定した変数間の因果関係を検証する。なお, 構造方程式モデリングの分析では, 大きなサンプルが必要であるが, 本研究は日本語学習者 850 名

であるためじゅうぶんなサンプルサイズである。本研究では、IBM-SPSS Statistics で起動して構造方程式モデリングの統計解析が行える AMOS (Analysis of Moment Structures) を分析に使用した。

### 3.2 日本語各能力の間の因果関係

日本語学習者の日本語能力を J-CAT の語彙、文法の得点を観測変数として、両変数を合わせて日本語の基礎力という潜在変数とした。また、J-CAT の読解と聴解の得点を観測変数として、両変数を合わせて理解力という潜在変数とした。

### 3.3 I-JAS ストーリーライティングのテキストから述部と名詞句の産出頻度の計算

I-JAS に掲載されたストーリーライティングは、描かれたイラストのストーリーについて説明する形式の課題が 2 つ用意されている。課題 1 は、5 コマからなる「ピクニック」、課題 2 は 4 コマからなる「鍵」というテーマである。それぞれの課題の登場人物の名前(「ケン」と「マリ」)、難易度が高くストーリーの説明には不可欠な語彙と英訳(「バスケット (basket)」「警官 (policeman)」など)、最初の 1 文(課題 1 は、「朝、ケンとマリはサンドイッチを作りました。」、課題 2 は、「ケンはずちの鍵を持っていませんでした。」)はあらかじめ提示されている。1 つの課題につき約 10 分の制限時間で、辞書やインターネットを使わず説明する(詳細は、迫田ほか、2020 を参照)。そのため、ストーリーライティングは、学習者が一定の時間内に習得した語彙を用いて、ストーリーをわかり易く説明できるかどうかを評価する課題である。つまり、説明力を測定するためのテキストデータであると考えられる。

動詞、助動詞、副詞、助詞の産出頻度を観測変数として、これらを述部という潜在変数とした。動詞は文の主要部であり、助動詞も動詞の使役、受動などを示すので、述部を構成する主要な品詞である。副詞も基本的に動詞を修飾するので、述部とした。助詞については、名詞句の最後につくが、「読む」であれば、主体の動作主がガ格(主格)で、読む対象がヲ格(与格)であり、動詞「読む」から与えられるものであるとして述部を構成する品詞とした。また、名詞、形容詞、連体詞の産出頻度を観測変数として、これらを名詞句という潜在変数とした。名詞句は、名詞が主要部であり、それを修飾する形容詞も名詞句の主要な構成要素である。連体詞は、日本語特有の品詞名であり、英語や中国語にはない。連体詞の多くは他の品詞または語接続から転成してできたもので、「この」「その」「いわゆる」などである。連体詞も名詞を修飾するので、名詞句を構成する品詞とした。母語別の述部の産出

頻度の平均と標準偏差は表3に、名詞句は表4に示した。

なお、品詞別の産出頻度については、使用頻度ではなく、テキスト全体の総語彙頻度に対する特定品詞の産出頻度の比率を使うことが推薦されている(大久保・大久保, 2014)。しかし、I-JAS のストーリーライティングの2つの課題については、テキストの長さも日本語学習者の説明力を反映していると考えられる。たとえ日本語学習者が多様な語彙を駆使して豊かな表現をしたとしても、全体に占める特定品詞の比率にすると、長いテキストを書く日本語学習者による品詞別の産出比率が低く示され、表現の多様性が失われてしまう可能性がある。そこで、比率ではなく産出頻度を使うことにした。もちろん分析では、前述のように自然対数に変換した値を使った。また、名詞については、「ケン」「マリ」などの固有名詞は含まず、普通名詞の産出頻度のみを使った。また、代名詞は名詞句を構成する重要な品詞ではあるが、産出頻度が0回あるいはきわめて低かったので含まなかった。

表3 I-JAS の日本語学習者の述部の産出頻度の平均、標準偏差および分散分析

日本語学習者の母語	動詞		助動詞		副詞		助詞	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
中国語	18.82	5.63	35.51	10.81	3.61	3.05	65.25	20.04
韓国語	17.86	6.70	34.45	11.98	2.33	2.53	61.51	22.20
英語	16.93	7.07	35.05	12.77	2.76	2.77	62.30	26.86
ベトナム語	15.64	4.05	34.70	9.93	1.90	1.53	53.32	14.95
タイ語	16.92	5.46	33.64	8.83	1.74	1.63	57.00	15.25
インドネシア語	16.54	5.59	34.78	11.57	2.78	2.00	59.14	18.82
トルコ語	16.60	6.61	37.62	11.34	1.42	1.58	59.80	19.14
スペイン語	13.84	4.97	34.48	13.23	1.40	1.68	53.04	16.33
フランス語	13.84	4.71	30.24	10.79	2.02	1.66	53.16	15.97
ドイツ語	18.72	6.05	38.38	12.22	3.00	2.78	64.24	19.85
ハンガリー語	19.72	6.11	38.54	9.40	3.36	2.11	70.48	18.42
ロシア語	19.12	6.39	42.32	12.50	3.08	2.88	65.74	20.46
全体	17.40	6.14	35.63	11.58	2.66	2.56	61.44	20.59
産出頻度 <i>F</i> 値	$F(11, 838)=6.33$		$F(11, 838)=3.65$		$F(11, 838)=7.16$		$F(11, 838)=4.49$	
有意確率と効果量	$p<.001, \eta_p^2=0.08$		$p<.001, \eta_p^2=0.05$		$p<.001, \eta_p^2=0.09$		$p<.001, \eta_p^2=0.06$	
自然対数 <i>F</i> 値	$F(11, 838)=7.37$		$F(11, 838)=4.23$		$F(11, 838)=8.23$		$F(11, 838)=5.02$	
有意確率と効果量	$p<.001, \eta_p^2=0.09$		$p<.001, \eta_p^2=0.05$		$p<.001, \eta_p^2=0.10$		$p<.001, \eta_p^2=0.06$	

注:  $N=850$ 。 *M* は平均, *SD* は標準偏差。各母語の人数は表1を参照。

表3の述部の産出頻度の平均をみると、母語によって述部の品詞の産出頻度に違いがあることがわかる。一元配置の分散分析で母語の影響を検討した結果は、表3に示した。動詞の産出頻度 [ $F(11, 838)=6.33, p<.001, \eta_p^2=0.08$ ] および自然対数 [ $F(11, 838)=7.37, p<.001,$



$\eta_p^2=0.09$ ], 助動詞の産出頻度 [ $F(11, 838)=3.65, p<.001, \eta_p^2=0.05$ ] および自然対数 [ $F(11, 838)=4.23, p<.001, \eta_p^2=0.05$ ], 副詞の産出頻度 [ $F(11, 838)=7.16, p<.001, \eta_p^2=0.09$ ] および自然対数 [ $F(11, 838)=8.23, p<.001, \eta_p^2=0.10$ ], 助詞の産出頻度 [ $F(11, 838)=4.49, p<.001, \eta_p^2=0.06$ ] および自然対数 [ $F(11, 838)=5.02, p<.001, \eta_p^2=0.06$ ] のすべてで、母語の主効果が有意であった。母語によって I-JAS のテキストの述部の産出量に大きな違いがあることがわかる。ドイツ語、ハンガリー語、ロシア語を母語とする日本語学習者の副詞の産出頻度が高いことがわかるが、中国語母語話者の副詞の使用頻度も高いので、言語の影響をすぐに議論できるわけではなさそうである。

表4 I-JAS の日本語学習者の名詞句の産出頻度の平均、標準偏差および分散分析

日本語学習者の母語	名詞		形容詞		連体詞	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
中国語	41.17	12.13	1.95	1.72	2.37	1.98
韓国語	36.18	13.58	1.78	1.95	2.06	1.85
英語	37.52	16.64	1.99	2.35	2.27	1.84
ベトナム語	35.08	9.41	1.48	1.55	2.02	1.82
タイ語	33.88	9.55	0.92	1.08	1.86	1.57
インドネシア語	34.50	10.02	1.32	1.91	2.74	2.18
トルコ語	35.16	9.93	1.70	1.78	1.44	1.58
スペイン語	30.14	9.41	1.36	1.56	1.22	1.42
フランス語	27.82	9.14	1.38	1.68	1.26	1.29
ドイツ語	36.64	12.22	1.92	1.70	1.80	1.48
ハンガリー語	42.86	11.85	2.92	2.46	2.02	2.10
ロシア語	37.66	11.22	2.66	2.50	2.78	2.41
全体	36.81	12.62	1.82	1.94	2.08	1.89
産出頻度 <i>F</i> 値	$F(11, 838)=8.10$		$F(11, 838)=4.60$		$F(11, 838)=4.35$	
有意確率と効果量	$p<.001, \eta_p^2=0.10$		$p<.001, \eta_p^2=0.06$		$p<.001, \eta_p^2=0.05$	
自然対数 <i>F</i> 値	$F(11, 838)=9.51$		$F(11, 838)=5.13$		$F(11, 838)=4.27$	
有意確率と効果量	$p<.001, \eta_p^2=0.11$		$p<.001, \eta_p^2=0.06$		$p<.001, \eta_p^2=0.05$	

注  $N=850$ 。 *M* は平均、 *SD* は標準偏差。母語の人数は表1を参照。

名詞句の産出頻度でも、母語によって品詞の産出頻度に違いがみられた。一元配置の分散分析で母語の影響を検討した結果は、表4に示したとおりである。名詞の産出頻度 [ $F(11, 838)=8.10, p<.001, \eta_p^2=0.10$ ] および自然対数 [ $F(11, 838)=9.51, p<.001, \eta_p^2=0.11$ ], 形容詞の産出頻度 [ $F(11, 838)=4.60, p<.001, \eta_p^2=0.06$ ] および自然対数 [ $F(11, 838)=5.13, p<.001, \eta_p^2=0.06$ ], 連体詞の産出頻度 [ $F(11, 838)=4.35, p<.001, \eta_p^2=0.05$ ] および自然対数

[ $F(11, 838)=4.27, p<.001, \eta_p^2=0.05$ ]のすべての主効果が有意であった。ロシア語およびハンガリー語を母語とする日本語学習者の形容詞の産出頻度が高いのは興味深い。やはり、母語によって I-JAS のテキストの名詞句の産出量に大きな違いがあることがわかる。

### 3.4 日本語の基礎力と理解力から述部と名詞句の産出頻度への因果関係モデル

日本語能力は基礎力と理解力に分け、テキストは述部と名詞句の産出頻度として、これら4つを潜在変数とした。テキストの産出頻度から日本語能力への因果関係は、日本語が理解できなくては、テキストの産出も行えないという前提のもとで想定しなかった。そして、図 1 に示したように、これらの潜在変数間の因果関係を5つのモデルで想定した。

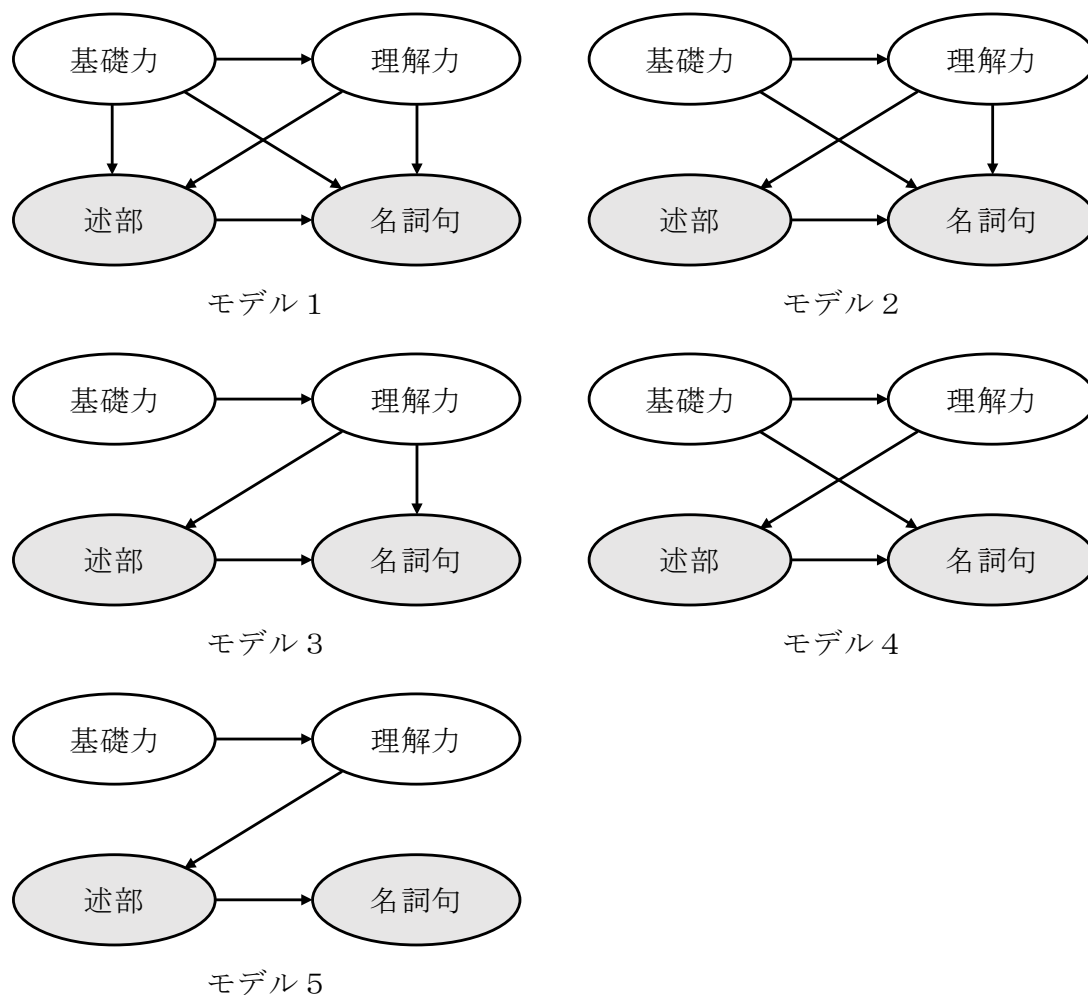


図1 基礎力と理解力およびテキストの述部と名詞句の因果関係モデル

注: 白色の楕円は日本語能力, 灰色の楕円はテキストの頻度を示す。

モデル1は、日本語の基礎力から理解力へ、また述部から名詞句へ影響し、さらに基礎力と理解力ともに述部と名詞句に影響すると仮定した。これは、日本語能力の基礎力と理解力が述部と名詞句の両者の語彙産出に直接に影響することを想定したモデルである。モデル1を基準として、モデル2からモデル5を想定した。モデル2は、モデル1の基礎力から述部への影響関係を外した。これは、日本語の基礎力は述部を構成する品詞の産出頻度に直接には影響しないと考えたモデルである。述部は文を構成する重要な構成要素であり、テキストの理解力がある程度高くなくては産出しえないと考えた。モデル3は、基礎力から名詞句への影響関係も外した。日本語がじゅうぶんに理解できなくては、述部や名詞句の産出もできないとするモデルである。モデル4は、基礎力は名詞句の産出へ、理解力は述部の産出へと影響して、さらに述部から名詞句の産出を導くという因果関係を想定した。名詞句は文の部分的な構成要素であるため、基礎力に含まれる語彙と文法の知識のみでも産出しようと考えた。一方、述部は日本語の理解力がなくては適切な産出ができないと考えたモデルである。さらに述部は名詞句の産出に影響すると想定した。最後に、モデル5では、基礎力から名詞句へのシンプルな逐次的な因果関係を想定した。つまり、日本語の基礎力が理解力を促進し、それが述部の品詞を産出して、さらに名詞句の品詞を産出するという連続的な因果関係のモデルである。

なお、名詞句から述部への因果関係を想定したモデルも、図1と同数の5つ作ることができる。しかし、次の節で詳細に説明するモデルの適合度検定では、これらのモデルと対をなす述部から名詞句への因果関係モデルと比べて、適合度が低くなった。そのため、名詞句から述部へという因果関係の方向は考えないことにした。これらのモデルについての詳細は本研究では記述しない。

### 3.5 5つの因果関係モデル間の適合度比較

図1の5つのモデルが、12の母語の日本語学習者850名のデータとどの程度適合しているかをAIC(Akaike's Information Criterion, Akaike, 1973, 「赤池の情報基準」ともいわれる)とBCC(Browne-Cudeck Criterion, Browne & Cudeck, 1989)の2つの指標で検討した。この適合度検定は、複数のモデルを比較するための指標である。モデルごとの指標と数値は表5のとおりである。いずれの指標も小さいほうがモデルとデータがよりよく適合していることを示す。AICの値は、モデル1が386.39、モデル2が384.50、モデル3が386.50で、モ

モデル4が 385.51, モデル5が 384.90 であった。最も小さい値となったのは、モデル2であった。同様に、BCC でも、モデル1が 387.20, モデル2が 385.27, モデル3が 387.25 で、モデル4が 386.26, モデル5が 385.62 であった。やはりモデル2が最小値を示した。したがって、本データとモデルが最もよく適合したのはモデル2である。なお、表5の飽和モデルは、自由度が0で、カイ二乗値が0になるモデルを想定した値であり、AIC と BCC がともに最小値となる。また、独立モデルは、観測された変数間に関連がない場合で、最大値となる(小塩, 2008)。そのため、実際の適合度は、飽和モデルと独立モデルの間の値をとる。モデル2とモデル5の AIC の差はわずかに 0.40 であり、BCC の差もわずかに 0.35 である。これは、モデル2とモデル5が類似したモデルであることを示唆している。本研究では、指標のより小さい値を取るという原則に従って、モデル2を採用した。

表5 日本語学習者 850 名の5つのモデルの適合度指標

モデル	$\chi^2$	df	p	$\chi^2/df$	AIC	BCC
モデル1	120.49	38	***	3.17	176.49	180.08
モデル2	121.71	39	***	3.12	175.71	179.18
モデル3	121.71	40	***	3.04	173.71	177.05
モデル4	122.12	40	***	3.05	174.12	177.46
モデル5	124.35	41	***	3.03	174.35	177.56
飽和モデル	-	-	-	-	132.00	140.47
独立モデル	-	-	-	-	1,093.44	1,094.85

注: N=850。 \*\*\*  $p < .001$ 。 灰色で示したモデル2が最適モデルである。

なお、表5に示したカイ二乗適合度検定は、どのモデルでも非常に大きく、すべて 0.1 パーセント( $p < .001$ )で有意になった。つまり、データとモデルが適応しておらず、モデルが棄却されたことになる。しかし、カイ二乗値は、標本数に敏感であり、多くなると有意になり易い(Bentler & Bonnet, 1980; Hooper, Coughlan, Mullen, 2008; Schermelleh-Engel, Moosbrugger & Müller, 2003; 豊田, 1998, 2007)。サンプルサイズは 500 以上になると、たいてい有意になり、モデルが棄却される(朝野・鈴木・小島, 2005)。本研究は、サンプル数が 850 名と非常に大きいので、カイ二乗値が有意になりモデルが棄却されたと考えられる。そのため、カイ二乗適合度検定の結果は、本研究では考慮しない。

### 3. 6 モデル2の適合度検定と潜在変数間の因果関係

5つのモデルの AIC と BCC を比較検討した結果、モデル2が最適であることがみいだされた。次に、モデル2そのものが、ほんとうにデータとよく適合しているかどうかを検討しなくてはならない。その際に、カイ二乗適合度検定が報告されることが多いが、サンプル数が大きいことから、本研究では使用しない。その代わりに、GFI, AGFI, NFI, CAI, RMSEA の5つの適合度指標を使用した。

GFI(Goodness-of-fit index)の指標は、Marsh and Grayson(1995)によれば、1に近いほどよく、0.95 以上であれば良好な適合を、0.90 以上であれば許容できる程度の適合を示している。モデル2は、0.94 であり、許容できる適合度を示した。また、自由度の影響を考慮して GFI を補正した指標である AGFI(adjusted GFI)も1に近いほどよく、0.90 以上で良好とされる(Schermelleh-Engel, Moosbrugger & Müller, 2003)。AGFI は自由度で補正しているので、GFI よりも小さい値になる。この指標でも、モデル2は、0.90 で良好であった。さらに、NFI(Normed Fit Index)の指標は、0.95 以上であれば良好な適合(Kaplan, 2000)、0.90 以上で許容できる程度の適合とされる(Marsh & Grayson, 1995)。本研究のモデルは、0.94 であり、許容できる適合度を示した。また、NFI に自由度の影響を考慮した CFI(Comparative Fit Index)の指標は、Schermelleh-Engel, Moosbrugger & Müller(2003)によると、0.97 以上が良好で、0.95 以上が許容できる程度の適合であるとしている。本研究の場合は0.95 であり、許容できる適合を示した。標本数と自由度で基準化したカイ二乗統計値である RMSEA(Root Mean Square Error of Approximation)の指標は、Browne and Cudeck (1993)によれば、0.05 以下であれば良好な適合を示す。朝野・鈴木・小島(2005)および小塩(2008)では、0.10 以上であれば適合度が悪いという判断基準を示しており、本研究のモデルは、0.09 であり、モデルとデータの適合度は、許容できる範囲で適合していた。以上のように、5つの適合度指標はすべてデータとモデルが許容できる程度で適応していることを示した。

モデル2がデータと適合していることが実証されたので、次に、因果関係の分析結果を図3に示した。潜在変数は、直接には観測できない構成概念である。しかし、構成概念の定義は抽象的であるため、複数の観測変数を合わせて推定される。観測変数で潜在変数を予測するという回帰式が成り立つと仮定する。もちろん、完全に潜在変数を予測しうるわけではないので、予測から外れた残差が生まれる。それらの残差は e で示した。図3には、11の観測変数があるので、それぞれに1つずつ残差があり、合計11ある。dで示したのは、潜在変数間の因果関係の予測の残差である。全部で3つある。

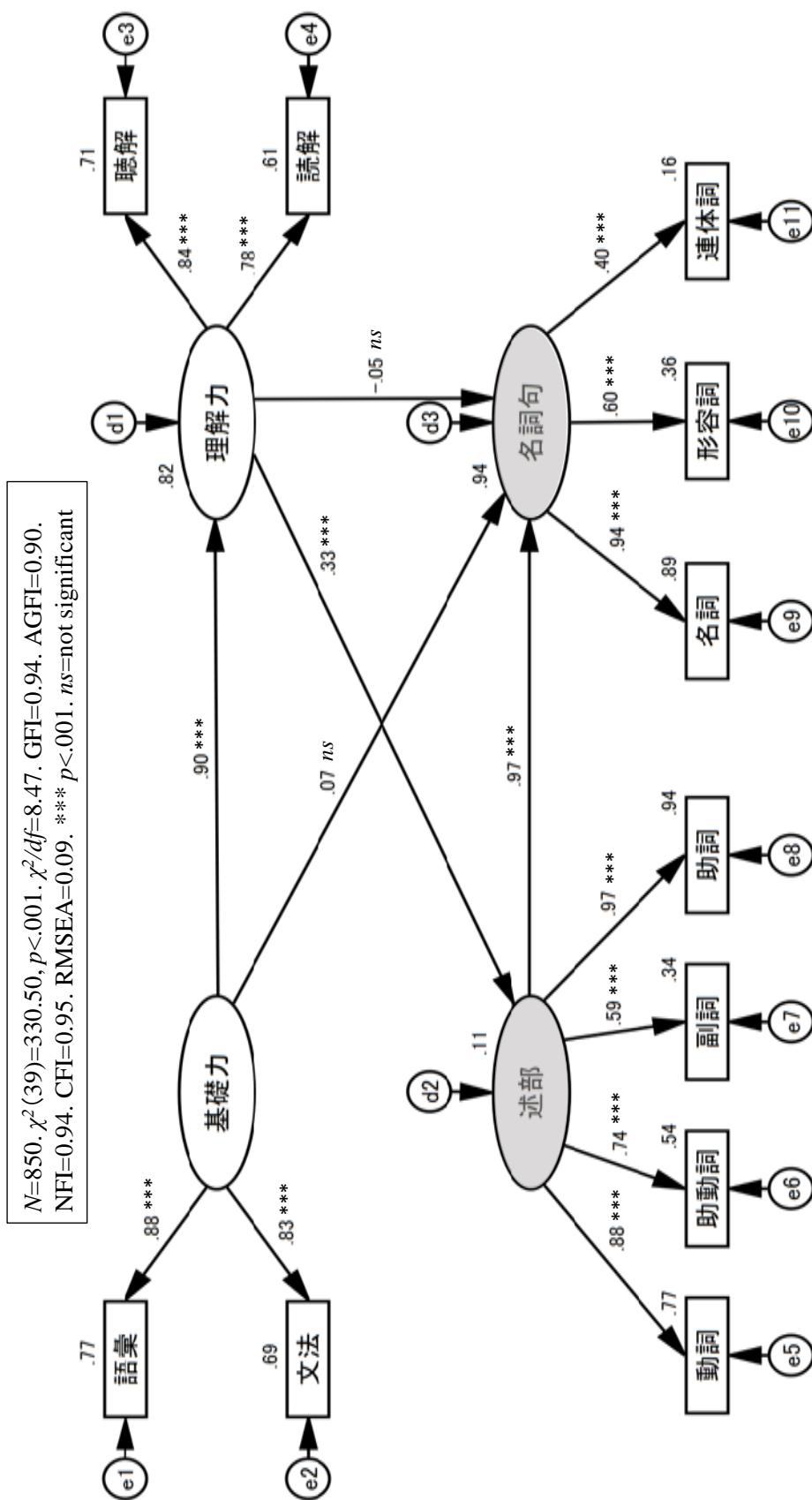


図2.1-JASの海外の日本語学習者850名の日本語能力とストーリーライティングの述部と名詞句の産出の因果関係(モデル2)

一つの潜在変数は、複数の観測変数とその残差で構成される。たとえば、本研究では日本語の基礎力を語彙知識と文法知識と定義した。両知識は J-CAT のテスト問題で得点化されている。もちろん測定誤差があるので、それぞれに残差がつけられている。この構造は、因子分析に相当し、あらかじめ研究者が潜在変数と観測変数との関係を仮定しているのので、確認的因子分析(confirmatory factor analysis)あるいは検証的因子分析(英語表現は同じ)と呼ばれる。日本語の基礎力は、文法(.83,  $p<.001$ )と語彙(.88,  $p<.001$ )の両観測変数で定義されており、ともに有意に基礎力を構成していた。同様に、理解力は、聴解(.84,  $p<.001$ )と読解(.78,  $p<.001$ )で、やはりともに有意に理解力を構成した。

ストーリーライティングの産出頻度は、自然対数の  $\log_e(x+0.5)$  を使って分析した。動詞(.88,  $p<.001$ )、助動詞(.74,  $p<.001$ )、副詞(.59,  $p<.001$ )、助詞(.97,  $p<.001$ )は、述部という構成概念を構成しており、いずれも有意であった。その中で、述部の構成概念に大きく貢献していたのは動詞と助詞であった。これは、文の主要部が動詞であり、動詞が項情報を提供して、それが助詞(特に、格助詞)で示されるとすれば、納得のいく結果である。さらに、名詞句を構成するのは、名詞(.94,  $p<.001$ )、形容詞(.60,  $p<.001$ )、連体詞(.40,  $p<.001$ )であり、いずれも有意に名詞句を構成している。名詞句の主要部は名詞であるため、当然のことながら名詞の貢献度がきわめて大きかった。次に、名詞を修飾する形容詞が名詞句の概念を形成するのに強く貢献していた。

構造方程式モデリングの解析の最終的な目的は、潜在変数間の因果関係を検証することである。最終モデルとなったモデル2では、日本語の基礎力から理解力への潜在変数間の因果関係を示す標準偏回帰係数(standardized partial regression coefficient)が.90 ( $p<.001$ )で、決定係数(coefficient of determination)が.82 である。これは理解力の 82%が基礎力で決まると解釈される。理解力が語彙や文法の知識を基にして形成される能力であることがよくわかる。また基礎力は、名詞句の頻度産出にも因果関係が認められた。しかし、標準偏回帰係数は.07 (*ns*)で有意ではなかった。したがって、この因果関係はほぼないといえる。さらに理解力は述部に対して強い因果関係があり、標準偏回帰係数.33 ( $p<.001$ )で、有意であった。ただ、貢献度は、11% (決定係数は.11)である。また、理解力からは名詞句への因果関係も想定されたが、標準偏回帰係数は-.05 (*ns*)で有意ではなかった。この因果関係はほぼ成り立たない。最後に、述部から名詞句への因果関係を示す標準偏回帰係数は.97 ( $p<.001$ )ときわめて強く、約 94% (決定係数は.94)が述部頻度で名詞句頻度が解釈された。ここで、「約」と表現したのは、モデル2では、日本語の基礎力と理解力からの貢献度

も含まれるからである。ただし、有意ではないので非常に弱い。総括すると、図3に示した一連の潜在変数間の因果関係は、連続・逐次的であった。まず、基礎力から理解力への非常に強い因果関係がみられ、理解力から述部への因果関係があり、さらに述部から名詞句へのきわめて強い因果関係がみられた。

なお、モデル2で有意な因果関係がみられなかった基礎力から名詞句および理解力から名詞句の流れを削除したのがモデル5である。そのため、有意な因果関係のみを描くと、モデル5になる。しかし、本研究では、わずかな違いではあるが、AIC と BCC の適合度検定でより小さい値のモデルを採用することにしたので、結果的に有意でない因果関係も含み込んだモデル2を最適モデルとした。仮に、モデル間の差が微妙である場合には、よりシンプルな因果関係を示すモデルを採用するという基準に立てば、モデル5が最適モデルであるとも考えることもできる。ただし、いずれのモデルであっても、潜在変数間の連続的な因果関係に違いはない。

#### 4. 総合考察

SNS の普及により、論理的な展開を有する作文力ではなく、わかり易く文章で相手に事実や要望などを伝える説明力という新たな能力が必要になってきた(池上, 2009; 工藤・松井, 2010)。本研究では、IJAS に収録された2つのストーリーライティング課題を、説明力を反映したコーパスであると想定した。構造方程式モデリングの分析では、サンプルサイズが大きなくてはならない。そこで、海外で学習する日本語学習者 850 名という大きなサンプルを研究対象とした。IJAS のストーリーライティングのテキストから、動詞、助動詞、副詞、助詞の4つの品詞を「述部」とし、名詞、形容詞、連体詞の3つの品詞を「名詞句」として、産出頻度を計算した。こうした説明文は、日本語能力に支えられていると仮定して、日本語能力からテキストの産出頻度への因果関係モデルを5つ想定した。構造方程式モデリングで解析した結果、図2に示した最適の因果関係モデルをみいだした。このモデルは、語彙と文法の基礎力が読解と聴解の理解力を促進し、理解力が述部の産出を豊かにし、述部が名詞句の産出に貢献するという連続的な因果関係を示した。言い換えれば、日本語の語彙と文法の知識が基礎力を作り、さらに聞き取りによるテキストの理解である聴解や書かれたテキストによる理解である読解で示される理解力が構築される。そして、日本語のテキストの正確な理解(聴解と読解)が前提となって、説明文の豊富な述部が産出され、それが名



詞句の豊かな表現を生み出すという因果関係の流れである。

日本語の語彙と文法の知識がテキスト理解のための基本的な能力だということは、これまで多くの現場の日本語教師が感じてきたことであろう。それが、世界の 12 種類の母語を持つ日本語学習者 850 名のサンプルで、標準偏回帰係数が.90という非常に強い因果関係を示された。もちろん、現実的な状況や目的に合わせて適切な表現ができることは重要であるが、まずは語彙と文法という基本を押さえずには、いきなりコミュニケーションを達成しようとしても容易ではないと思われる。このことは、聴解や読解の理解力によって、述部の語彙産出が促進されることから裏付けられよう。ただ、「聞く」ことによるテキスト理解である聴解と「読む」ことによるテキスト理解である読解から、説明文を「書く」という産出の因果関係は、標準偏回帰係数が.33 であり、日本語の基礎力から理解力への因果関係と比べると弱い。「聞く」「読む」という受身的(*passive*)な活動から「書く」という生産的(*productive*)な活動への認知的なギャップは大きいのであろう。最後に、説明文を書くにあたり、標準偏回帰係数が.97 という述部から名詞句という産出頻度のきわめて強い因果関係がみられた。これは、文産出の認知的なメカニズムから考えると、まず文の主要部である動詞を基にした述部を正しく産出し、動作主と被動者あるいは対象物の関係を決めて、その後で具体的な動作主、被動者あるいは対象物を描くための名詞句を作るというテキスト産出の基本的な認知プロセスであるといえよう。以上のように、本研究では、構造方程式モデリングの解析法を駆使することで、日本語能力を背景にした説明力への因果関係を描くことができた。

#### [引用文献]

- 赤木彌生・今井新悟 (2012) 「J-CATmini 日本語テスト・ネット体験版の構築と利用」『大学教育』(山口大学・大学教育機構) 9, 43-49。
- 朝野熙彦・鈴木督久・小島隆矢 (2005) 『入門 共分散構造分析の実際』講談社サイエンティフィク。
- 南風原朝和(2002) 『心理統計学の基礎—統合的理解のために』有斐閣。
- 初相娟・玉岡賀津雄 (2013) 「中国人日本語学習者による語彙・文法知識から述部構造の理解を仲介とした読解への因果関係モデルの検証」『ことばの科学』 26, 5-24。
- 池上彰 (2009) 『わかりやすく伝える技術』講談社現代新書。
- 狩野裕 (2002) 「構造方程式モデリングは、因子分析、分散分析、パス解析のすべてにとって代わるのか？」『行動計量学』 29(2), 138-159。

- 狩野裕・三浦麻子 (2002) 『グラフィカル多変量解析—AMOS、EQS、CALIS による目で見ると共分散構造分析』現代数学社。
- 小林典子 (2015) 「SPOT」李在鎬 (編) 『日本語教育のための言語テストガイドブック』 (pp. 110-126) くろしお出版。
- 小林典子・フォード丹羽順子・山元啓史 (1996) 「日本語能力の新しい測定法 [SPOT]」『世界の日本語教育』6, 201-218。
- 小島隆矢 (2003) 『Excel で学ぶ共分散構造分析とグラフィカルモデリング』オーム社。
- 工藤昌幸・松井寿夫 (2010) 『言いたいことがキチンと伝わる説明力の基本』こう書房。
- 大久保起延・大久保博美 (2014) 「助詞・助動詞の使い方と漱石作品の分類」『計量国語学』29(6), 227-234。
- 小塩真司 (2008) 『はじめての共分散構造分析 Amos によるパス解析』東京図書。
- 迫田久美子・石川慎一郎・李在鎬・佐々木藍子・須賀和香子・野山広・細井陽子・八木豊 (2020) 『日本語学習者コーパス I-JAS 入門』東京：くろしお出版。
- 玉岡賀津雄 (2000) 「中国語系および英語系日本語学習者の母語の表記形態が日本語の音韻処理に及ぼす影響」『読書科学』44, 83-94。
- 豊田秀樹 (1998) 『共分散構造分析 入門編』東京：朝倉書店。
- 豊田秀樹 編著 (2007) 『共分散構造分析 [AMOS 編]』東京：東京図書。
- 張麒声 (2011) 『中国語話者のための日本語教育研究入門』大阪：日中言語文化出版社。
- 大和祐子・玉岡賀津雄 (2013) 「中国語母語話者と韓国語母語話者の日本語テキストの読み処理における言語的類似性の影響」『小出記念日本語教育研究会論文集』21, 61-73。
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Petrov, B. N., and Caski, F. (Eds.), Proceedings of the 2nd International Symposium on Information Theory (pp. 267-281) Budapest: Akademiai Kiado.
- Bentler, P. M., & Bonnet, D. C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.
- Browne, M. W. and Cudeck, R. (1989). Single Sample Cross-Validation Indices for Covariance Structures, Multivariate Behavioral Research. 24(4), 445-455.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 136-162). Newbury Park, CA:

Sage.

Ellis, R. (1985). *Understanding second language acquisition*. Oxford, UK: Oxford University Press.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal of Business Research Models*, 6, 53-60.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd edition). New York: Routledge.

Kaplan, D. (2000). *Structural equation modeling: Foundation and extensions*. Thousand Oaks, CA: Sage.

Marsh, H. W. & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 177-198). Thousand Oaks, CA: Sage.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, 8, 23-74.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, 30 (4), 415-433.

Yamamura, K. (1999) Transformation using  $(x+0.5)$  to stabilize the variance of populations. *Researches on Population Ecology*, 41, 229-234.

Web サイト

Nordquist, R. (2019). What is expository writing?: How to write an expository essay. [www.thoughtco.com/expository-writing-composition-1690624](http://www.thoughtco.com/expository-writing-composition-1690624).

## The causal relation between the Japanese language proficiency of 850 foreign learners and productive frequencies of lexical categories in expository text

TAMAOKA Katsuo

(Professor, Graduate School of Humanities, Nagoya University, Japan)

**Abstract:** With the widespread use of SNS (social networking services), people are communicating more frequently by written text than ever before. The ability to express content plainly is in great demand. In the present study, this is defined as “explanatory ability”. In contemporary society, this ability has become essential to daily life not only for native Japanese speakers but also for foreign learners of Japanese. Assuming the two story-writing tasks recorded in the International Corpus of Japanese as a Second Language (I-JAS) reflect explanatory ability, the productive frequencies of lexical categories found in the text written by 850 Japanese learners outside Japan were divided into two groups: (1) four categories of verbs, auxiliary verbs, adverbs, and auxiliaries as “predicates” and (2) three categories of nouns, adjectives, and conjoined words as “noun phrases”. Using the statistical method of structural equation modelling (SEM), the five causal relation models linking Japanese proficiency as measured by the Japanese Computerized Adaptive Test (J-CAT) and text productive frequencies (see Figure 1) were examined to see how these models fit with data from proficiency scores and productive frequencies. Based on the fitting indexes, the study found the best causal model to be that identified in Model 2 (see Figure 2). This model showed causal sequencing in the following order: (1) possession of a basic knowledge of vocabulary and grammar (2) promotes reading and listening comprehension, further (3) enriches the production of predicates and (4) contributes to the production of noun phrases.

**Keywords:** I-JAS (International Corpus of Japanese as a Second Language), J-CAT (Japanese Computerized Adaptive Test), explanatory ability, Japanese language proficiency, productive frequencies, SEM (structural equation modeling)