



# 日本語聴解試験の構成概念についての一考察：日本語能力試験を例として

張, 晶鑫

劉, 敏

---

(Citation)

統計数理研究所共同研究リポート, 484:19-31

(Issue Date)

2026-03-25

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/0100502561>

(URL)

<https://hdl.handle.net/20.500.14094/0100502561>



# 日本語聴解試験の構成概念についての一考察

—日本語能力試験を例として—

張晶鑫(湖北大学・神戸大学)

zjx-kobe@hotmail.com

劉敏(湖北省京山市第五高級中学)

1355704030@qq.com

A Study on the Construct of Japanese Listening Comprehension Tests

—The Case of the JLPT—

ZHANG, Jingxin (Hubei University/ Kobe University)

LIU, Min (Jingshan No.5 Middle School)

## 概要

本研究は、Bachman (1990) のコミュニケーション言語能力モデルを理論的枠組みとし、日本語聴解試験における「構成概念」の在り方について検討した。とりわけ、テストの難易度および真正性をいかに定義し、それらをどのように操作化するかに焦点を当てて分析を行った。分析にあたっては、『日本語聴解試験コーパス』を独自に構築し、『日本語日常会話コーパス』を参照コーパスとして用いた。その結果、難易度は主として話すスピード、語彙多様性、漢語率などの要因によって規定されることが明らかになった。真正性については、N3 から N1 へとレベルが上がるにつれて、CEJC の「雑談」に徐々に近接する傾向が確認された。また、難易度と真正性との関係は逆 U 字型の曲線を描くことが示唆された。これらの結果は、テスト課題の設計および結果解釈に対する事前的妥当性の証拠となりうるものである。

## キーワード

構成概念、聴解試験、日本語能力試験、コーパス文体論

## 1. はじめに

言語技能のうち、聴解は外国語の総合的能力を測定する上で重要な指標の一つとされ、国内外の大規模な言語テストにおいて不可欠な構成要素となっている(侯, 2009)。しかしながら、聴解は人間の第一言語(L1)において最も早い時期に発達する技能であるにもかかわらず、同時に最も測定が困難な言語技能でもある(Field, 2013)。その困難さの一因として、聴解テストにおける構成概念の定義の問題が挙げられる。外国語テストにおいては、まず「なぜ測定するのか」「何を測定するのか」を明確にする必要がある。すなわち、言語テストにおける「構成概念(construct)」を

規定することが前提となる(金, 2022)。構成概念の定義は、試験内容の選定、課題設計、さらには得点の解釈に至るまで、テストの根幹に直接的な影響を及ぼす。

しかしながら、言語テストにおける構成概念の明確化は容易ではない。Chapelle et al. (2010) は、関連する言語能力の構成概念を正確に定義することは極めて困難であると指摘している。また、李・楊(2018)は、実際のテスト作成においては、専門家の判断が量的根拠を欠く場合が多く、開発者はテストに使用されるテキストのジャンル、設問形式、難易度などの要因に偏重しがちであることを指摘している。その結果、テストが「妥当性の中核」から一定程度逸脱する可能性があるとは指摘する。そのため、テストに内在する言語特性、すなわち構成概念を明確に記述することが必要であると思われる。Chapelle et al. (2010) は、テストにおける構成概念の定義は大きく二段階に分けられるとする。第一に、理論的・概念的水準において構成概念を規定すること。第二に、選定されたテキストおよび受験者に求められる課題を通じて、その構成概念を操作化することである。

以上を踏まえ、本研究は日本語能力試験(JLPT)の聴解試験に使用されるテキストを分析対象とし、難易度および真正性を切り口として、日本語聴解試験における構成概念の定義とその操作化の在り方を検討する。これにより、今後の日本語聴解試験の開発および設計に対して示唆を提供することを目指す。

## 2. 先行研究

### 2.1 構成概念に関する研究

*Standards for Educational & Psychological Testing*(AERA et al. 2014)によれば、構成概念とは「テストが測定する概念または特性」を指す(p.11、金 2022 より引用)。言語テストは、テスト課題を通じて受験者の外在的行動を測定・評価し、その言語能力を直接的あるいは間接的に推論する営みである。Bachman & Palmer(1996)は、「多くの場合、言語テストの目的は受験者の言語能力、あるいはその一側面について推論を行うことである」と指摘している。言語能力は潜在特性(latent trait)であり、直接観察することはできず、外的行動を通じて推定されるほかない(鄒, 2005:88)。Bachman & Palmer(2010)は、構成概念を能力の具体的定義と捉え、韓・張(2015)は、言語テストにおける構成概念の定義とは、言語能力に対する操作的定義を与えることにほかならないと述べている。

しかしながら、構成概念自体も直接観察することのできない概念である。したがって、測定目標としての構成概念が実際にどの程度測定されているのか、すなわち、試験結果がどの程度受験者の言語能力またはそれに関連する心理的特性を説明し得るのかを検証する構成概念妥当性の問題が重要な研究課題となってきた。1980年代以降、Messick(1989)が妥当性を統一的概念(unitary concept)として提示して以来、構成概念妥当性はもはや妥当性の下位概念ではなく、「妥当性そのもの」であるという理解が広く共有されている。これを受けて、テスト得点に基づく推論および構成概念解釈を支持する証拠の十分性、すなわち得点解釈の妥当性を総合的に評価する試みが進められてきた。たとえば、Michael Kane(1992/2012)の論証に基づく妥当性検証モデル(argument-based approach to validity)や、Bachman & Palmer(2010)のテスト使用に関

する論証 (Assessment Use Argument) などが挙げられる。

以上の議論を踏まえると、構成概念とは言語能力と言語テスト課題とを結び付ける重要な媒介概念であり、これら三者の緊密な関係性は、テスト結果の推論および解釈において重要な役割を果たすと考えられる。しかしながら、既存の妥当性検証モデルはいずれも、受験者のパフォーマンスに基づく一連の推論の妥当性を論証することを重視している (韓・張 2015)。一方、Weir (2005:17-18) は、理論的妥当性には、テスト実施前に収集される先行的証拠 (a priori evidence) と、実施後に得られる事後的証拠 (a posteriori evidence) の双方が含まれると指摘する。とりわけ、先行段階においてテストの構成概念がより明確に規定されているほど、テスト結果の解釈はより有意義なものとなる。しかし、従来の研究はテストの構成概念を主として事後的な統計的検証の問題として扱う傾向が強く、先行段階における検証が必ずしも十分とは言い難い。また、テストに使用されるテキストそのものへの関心は必ずしも高いとはいえない。

聴解試験に関して言えば、受験者が音声情報を理解する過程はテキストに依拠しており、試験課題もまたテキストを基盤として構成される。したがって、テキストを離れてテストの妥当性やその構成概念を論じることはできないと考えられる。

## 2.2 日本語聴解試験に関する研究

何ほか (2018) は、Bachman & Palmer (1996) が提示したテスト法の記述枠組み、すなわち、試験環境、試験指示、インプット特性、期待される応答、ならびにインプット特性と期待応答との関係を整理し、これらを大きく「インプットテキスト特性」と「テスト課題特性」の二類型に集約している。日本語聴解試験におけるインプットテキスト特性に関しては、侯 (2009) が、聴解問題の作成は真正性を前提とすべきであり、「テキストの真正性」「聴き手の役割の真正性」「課題の真正性」を堅持する必要があると指摘している。また、李 (2022) は、発話のスピードが中・上級日本語学習者の聴解処理に与える影響を検討している。

一方、テスト課題特性については、足立 (2016) が、日本語聴解試験における設問の有無および設問提示位置 (前置・後置) がテスト結果に及ぼす影響を実証的に検証している。さらに、島田・侯 (2009) は、選択肢の提示形式における音声提示と文字提示の差異について考察している。

このほか、先行研究は聴解試験の妥当性検証および聴解の処理過程にも注目してきた。前者の例として、董・冷 (2020) は、2005 年から 2016 年までの中国日本語専攻四級試験における聴解問題の内容的妥当性を検証し、全体として高い内容的妥当性を有する一方、一部のテキストにおいて未知語がやや多いという課題を指摘している。聴解の処理過程に関しては、山方 (2008) が、学習者の第二言語知識および母語背景が聴解テキスト中の未知語の意味推測に及ぼす影響を調査している。また、費 (2019) は、学習者の語彙処理速度や聴解材料の難易度などが聴解理解過程に有意な影響を与えることを明らかにした。さらに、島田ほか (2021) は、聴解試験において、習熟度の高い学習者は全体理解に基づいて解答する傾向があるのに対し、習熟度の低い学習者はキーワードの聞き取りに依拠して選択を行う傾向があることを指摘している。

以上の研究をまとめると、日本語聴解試験に関する研究は、大きく「インプット型」と「聴き手型」

に区分できる。インプット型研究はさらに、テキスト特性と課題特性に分けられる。一方、聴き手型研究では、主として受験者の聴解の処理過程およびその情報処理メカニズムが検討対象となっている。しかしながら、聴解試験の構成概念妥当性を扱う研究は存在するものの、構成概念そのものに焦点を当てた研究は必ずしも多いとは言えない。とりわけ、聴解に使用されたテキストそのものへの関心は十分とは言えず、先行的証拠に基づく構成概念の検討も不足している。また、日本語能力試験における各レベル間の縦断的分析が乏しく、レベル設定の基準やレベル間の連続性についても検討の余地がある。さらに、少量のテキストを対象とする実験的・質的研究が中心であり、大規模データに基づくコーパス分析は十分に行われているとはいえない。

### 3. リサーチデザインと手法

#### 3.1 研究目的と RQ

本研究では、Bachman(1990)のコミュニケーション言語能力モデルを理論的枠組みとし、自作の日本語聴解試験コーパスをデータとして用い、日本語聴解試験に使用されるテキストの言語的特徴の解明を目指す。また、日本語聴解試験の構成概念の定義を検討し、聴解試験の作成に資する示唆を提供することを目的とする。

表1  
分析の枠組み

	組織的能力	指標
	音韻	話すスピード、フィラー
文法的能力	語彙	token、TTR、語彙多様性(Herdan の C 値)、漢語率、外来語率、名詞率、動詞率
	統語	MVR 率、平均文長
テキスト的能力	結束性と構造	代名詞率、接続詞率

テスト開発者の観点から見ると、構成概念の定義には、難易度、真正性、音声および環境の音質、音声素材の出所、話題、談話類型などが関与するとされる(Green, 2017:35-49)。本研究では、まず試験のテキストで操作しやすい難易度、真正性に注目する。また、Bachman(1990)のコミュニケーション言語能力モデルに基づけば、言語能力は組織的能力と語用論的能力に区分されるが、本稿では紙幅の制約上、組織的能力に限定して考察を行う。表 1 は、毛(2012)が用いた指標を援用して作成したものである。ただし、本分類は現段階では暫定的なものであり、さらなる検討の余地が残されている。例えば、フィラーについては、音声産出や聴覚的言語知覚に関わる現象として音韻に分類したが、談話管理の機能を担う側面も指摘されている。この点については、今後の検討課題としたい。そこで、本研究では以下の 2 つの研究設問(RQ)を設定する。

**RQ1** 難易度という観点から、聴解試験の構成概念にはどのような要素が関わるか。

**RQ2** 真正性という観点から、聴解試験の構成概念にはどのような要素が関わるか。

### 3.2 データ

本研究で使用するデータは、以下の二種類である。第一に、本研究で独自に構築した『日本語聴解試験コーパス』(Japanese Listening Test Corpus: JLTC)である。総語数は 387,016 語であり、2010 年の試験改革以降、2022 年までの 13 年間に実施された計 25 回分の日本語能力試験(以下、JLPT)のうち、N3・N2・N1 の 3 レベルにおける聴解試験テキスト本文 75 本を収録している。JLPT は原則として毎年 2 回実施されており、本コーパスには各回の試験本文を収録した(ただし、2020 年 7 月実施分は新型コロナウイルス感染症の影響により中止されたため除外した)。なお、N4・N5 については中国における受験者層が相対的に限定的であることから、本研究の分析対象には含めていない。JLPT を分析対象として選定した理由は、同試験が受験者にとって進学・留学・就職等の重要な指標となっている点にある。すなわち、JLPT は大規模かつ利害関係の大きい言語テストに位置づけられ、世界的に見ても受験者数が最も多く、代表性の高い日本語能力を測定する試験である。

第二のデータとして、本研究では『日本語日常会話コーパス』(CEJC)を用いる。本コーパスは、日本語聴解試験の真正性を検討するための参照コーパスとして位置づけられる。CEJC は国立国語研究所により構築され、2022 年 3 月に一般公開されたコーパスであり、日本語母語話者による日常会話 200 時間分を収録している。20 歳代から 60 歳代までの各年齢層の話者を均衡的に抽出し、起床から就寝に至るまでの多様な場面で自然に生じた会話の音声・映像データを収録しており、日常の日本語言語生活を代表する資料であるといえる。

表 2

本研究で使用するデータの概要

コーパス		サイズ
自作の日本語聴解試験コーパス(JLTC)	N3	89,451 語
	N2	130,098 語
	N1	167,467 語
日本語日常会話コーパス(CEJC)	雑談	1,659,893 語
	用談・相談	450,102 語
	会議・会合	259,430 語

### 3.3 手法

まず、国立国語研究所が開発した UniDic 辞書および MeCab 形態素解析器を用いてテキストを解析し、トークン数、タイプ数、各語種の比率(漢語・外来語の割合)、品詞分布(名詞・動詞などの割合)を算出した。

次に、タイプ数/トークン数により TTR (Type-Token Ratio) を算出し、語彙多様性の指標として、Herdan の C 値 (タイプ数の自然対数/トークン数の自然対数) を用いた。また、{(形容詞数+形容動詞数+副詞数+連体詞数)/動詞数×100} の式により MVR 比を算出した。発話のスピードは、1 分間あたりに発話された文字数 (試聴時間等は除外) として計測した。平均文長は、総文字数を句点数で割ることにより算出した。なお、ここでの文字数の計算は、Microsoft Word の文字カウント機能によるものである。

難易度の分析においては、本稿では、絶対的難易度を受験者のパフォーマンスに依拠しないテキスト自体の特性に基づく難易度とし、相対的難易度をテキスト自体の特性の影響を受けつつ、受験者のパフォーマンスにも依拠する難易度とする。まず日本語能力試験の三つのレベル (N1・N2・N3) を従属変数、13 のコーパス文体指標 (表 1 参照) を独立変数として多項ロジスティック回帰分析を行い、テキストの絶対的難易度区分およびその影響要因を検討した。次に、JLPT 公式サイトから取得した各回の合格率を従属変数、同 13 指標を独立変数として重回帰分析を行い、テキストの相対的難易度区分およびその影響要因を分析した。真正性の検討においては、N1、N2、N3、CEJC 雑談、CEJC 会議・会合、CEJC 相談・用談を第一次元とし、13 指標を第二次元とした対応分析を実施し、聴解試験テキストと日本語母語話者による自然会話産出との関係性を算出した。重回帰分析と対応分析はともに関西大学水本篤氏が開発した langtest.jp を使用した。

## 4. 結果と考察

### 4.1 RQ1 難易度と構成概念

本研究ではまず、N1、N2、N3 を従属変数、13 種の指標を独立変数として、順序ロジスティック回帰分析を行ったが、平行性の仮定は満たされず ( $0.006 < 0.05$ )、レベル区分を順序尺度として扱うことは適切ではなく、順序ロジスティック回帰分析に適用できないことが示された。これは、テキストの難易度が順序立った絶対的な区分ではなく、各レベル間の境目が必ずしも明確ではない可能性を示唆している。

そこで次に、N1、N2、N3 を従属変数、13 種の指標を独立変数として多項ロジスティック回帰分析を実施した。その結果、以下の 3 つの判別モデルが得られた。N1 および N2 の再現率 (recall) は 72.8%、N3 は 81.7% であり、全体の再現率は 75.7% であった。

$$P_{N1} = \frac{1}{(1 + e^{(-163.808 + 0.108 \times X_{\text{話すスピード}} + 0.063 \times X_{\text{トークン}} + 167.010 \times X_{\text{語彙多様性}} - 1.055 \times X_{\text{動詞率}} + 1.426 \times X_{\text{漢語率}})})}$$

$$P_{N2} = \frac{1}{(1 + e^{(-106.863 + 0.096 \times X_{\text{話すスピード}} + 0.025 \times X_{\text{トークン}} + 102.995 \times X_{\text{語彙多様性}} - 0.576 \times X_{\text{動詞率}} + 0.846 \times X_{\text{漢語率}})})}$$

$$P_{N3} = 1 - P_{N1} - P_{N2}$$

分析の結果、話すスピード、トークン数、語彙多様性、漢語率、および動詞率がレベルの判定・予測に有意な影響を及ぼすことが明らかになった。特に話すスピードは難易度と正の相関関係を

示し、発話が速いほど上位レベルに分類される傾向が確認された。また、トークン数および語彙多様性は語彙的複雑性を表す指標であり、語彙の種類が豊富であるほど難易度が高まる傾向が認められた。これらの結果は、英語聴解試験を対象とした Brunfaut & Révész (2015) および何ほか (2018) の知見とも整合的である。したがって、発話のスピードおよび語彙的複雑性は、言語を問わず聴解難易度を規定する比較的普遍的な指標であると考えられる。一方、日本語聴解試験固有の特性として、漢語率は難易度と正の相関を示し、動詞率は負の相関を示した。また、李 (2016) および川村ほか (2013) が指摘した平均文長は、本研究の予測モデルでは有意ではなかった。この差異は、書き言葉を主対象とした先行研究と、話し言葉を対象とする本研究とのジャンル差に起因する可能性がある。

さらに、日本語能力試験各レベル (N1・N2・N3) の合格率を従属変数とする重回帰分析 (ステップワイズ法) を実施したところ、以下の 3 つのモデルが得られた。

$$N1 = 62.789 - 0.039 \text{ 話すスピード} - 0.676 \text{ 名詞率} - 0.892 \text{ MVR 比} - 1.367 \text{ 代名詞率} \\ (F=3.011, p < .01, r = .091, \text{ adjusted } R^2 = .061)$$

$$N2 = 75.008 - 0.059 \text{ 話すスピード} - 0.654 \text{ 動詞率} - 0.864 \text{ MVR 比} - 0.761 \text{ 漢語率} \\ (F=3.239, p < .05, r = .097, \text{ adjusted } R^2 = .067)$$

$$N3 = 59.873 - 0.033 \text{ 話すスピード} - 0.122 \text{ フィラー率} - 0.488 \text{ 名詞率} - 0.921 \text{ 代名詞率} \\ (F=2.296, p < .05, r = .074, \text{ adjusted } R^2 = .042)$$

N1、N2、N3 という名義尺度を従属変数とした多項ロジスティック回帰分析は、受験者パフォーマンスに依拠しないテキスト特性に基づく「絶対的難易度」の検討に相当する。一方、合格率を従属変数とする重回帰分析は、テキスト特性と受験者パフォーマンスの双方に依拠する「相対的難易度」の検討と位置付けられる。分析の結果、各レベルにおいて合格率に影響を与える指標は異なることが示されたが、発話のスピードはすべてのモデルに共通して投入された唯一の指標であった。

N3 では、話すスピード、フィラー率、名詞率、代名詞率が主要な難易度要因として抽出された。フィラーとは、「なんか」「あのう」「ええと」「ええ」など、発話時に時間調整、注意喚起、リズム調整などの機能を持つ実質的意味を伴わない語を指す (松浦, 1996)。学習者にとっては、その語用的機能の識別が困難である可能性がある。代名詞は、「私」「誰」「僕」「彼」などの人称代名詞およびコソアド系指示詞を含み (佐久, 1936)、「話し手の情報勢力範囲」 (轟, 1998:83) に関わる語である。蔭山 (2017) が指摘するように、日本語学習者は代名詞の理解および使用において誤用や誤解を生じやすい。

N2 では、話すスピード、動詞率、MVR 比、漢語率が有意であり、話すスピードが速く、動詞の使用が多く、動作描写傾向が強くと、漢語が多いほど合格率が低下する傾向が示された。N1 では、話すスピードに加え、名詞率、MVR 比、代名詞率が重要な指標であることが示唆された。樺島・寿岳 (1965) は、名詞率が高い文章は要約的傾向を示し、MVR が高い文章は「ありさま描写」的であると指摘している。また、井関ほか (2022) は、名詞率と動詞率が登場人物や場面のイメージしや

すさに負の影響を与える可能性を報告している。

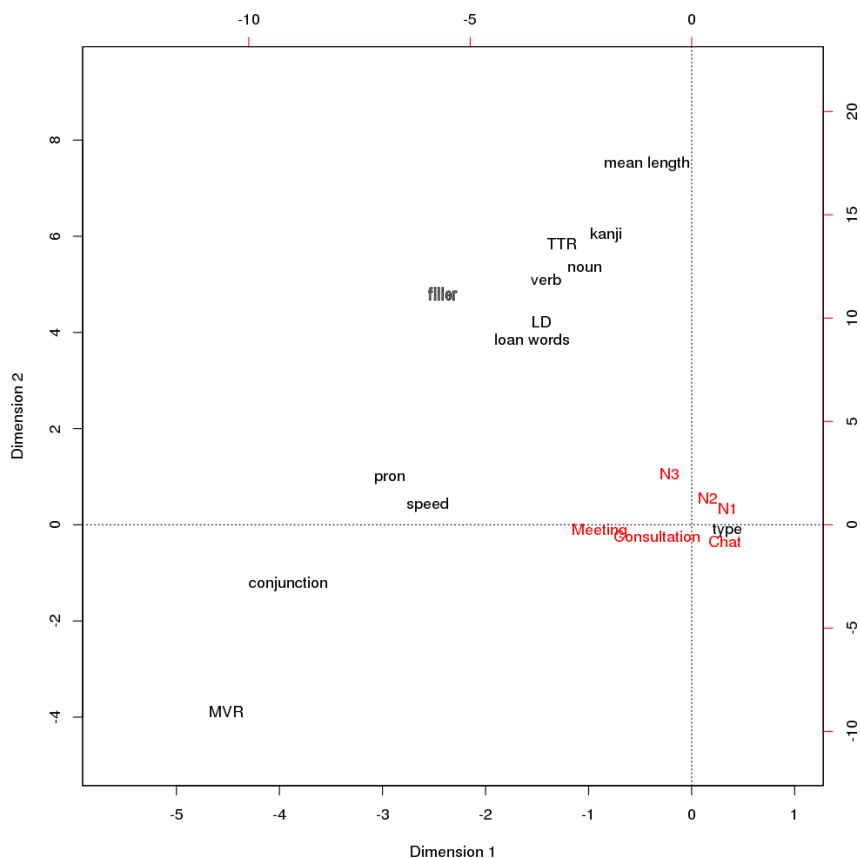
以上の重回帰分析の結果から、N3 レベルでは、発話のスピード、フィラーの語用的機能の識別、代名詞の指示的意味の把握が主な困難要因であることが示唆された。N2 レベルでは、発話のスピードの上昇、動詞の多用、動作描写傾向、漢語の多さが合格率の低下と関連している。一方、N1 レベルでは、発話のスピードに加え、名詞および代名詞の意味的解釈が重要な難易度要因であると考えられる。

さらに、彭・張(2013)は、語彙・音声・統語・談話知識のみでは設問の難易度を正確に予測することは困難であると指摘しているが、本研究の結果からはテキスト難易度と設問難易度との間に一定の相関が存在する可能性が示唆された。ただし、いずれのモデルの説明率は限定的であり、これらの指標のみからテキストの難易度を十分に判定することは困難である。しかし、コーパス的手法を用いることにより、テキスト難易度の段階的区分は一定程度可能であることが示唆される。

#### 4.2 RQ2 真正性と構成概念

真正性を検証するために対応分析を行った結果、図1に示す散布図が得られた。第1次元は全体の分散の60.90%を説明し、第2次元は38.38%を説明している。

図1  
対応分析の散布図



真正性 (authenticity) とは、言語テストにおける課題特性が目標言語使用場面の特性とどの程度一致しているかを指す概念であり、テスト開発者が考慮すべき重要な要素である (Green, 2017:37)。図 1 に示されるように、横軸に沿って JLPT と CEJC が上下に明確に分かれており、両者の間に顕著な差異が存在することが示された。しかしながら、その分布傾向を見ると、N3 から N1 へと上位レベルに移行するにつれて、JLPT は CEJC の雑談データに徐々に接近しており、真正性が段階的に高まる傾向が認められる。

また、縦軸では CEJC の雑談と、CEJC の会議・会合、相談・用談とが左右に分かれていることから、日本語の言語使用には明確なジャンル差が存在することが示唆される。すなわち、日常的雑談とよりフォーマルな会議・用談場面とでは、言語特性に体系的な差異が見られる。

これに対し、JLPT の聴解テキストは、会議や商談といったフォーマル場面よりも、日常生活における雑談場面により近い位置に分布していることが明らかになった。さらに、JLPT と CEJC との差異を精緻に検討するため、本研究では CEJC 平均値および N1/CEJC 比を算出した (表 3)。その結果は以下の通りである。

表 3 JLPT と CEJC との比較

	N3	N2	N1	雑談	相談・ 用談	会議・ 会合	CEJC 平均	N1/ CEJC
話すスピード	139.5	164.7	176.8	396	349	365	370	0.5
ファイラー	14.2	12.4	10.5	10.7	9.8	15.4	12	0.9
token	1129	1937	2698.6	6753	3000	2014	3922.3	0.7
TTR	7.2	7.6	8.3	4.2	5.3	5.8	5.1	1.6
語彙多様性	0.7	0.7	0.8	0.7	0.7	0.7	0.7	1
名詞率	22.5	23.7	24.7	17	19.2	16.4	17.5	1.4
動詞率	13	12.2	12	10.2	10.4	10.3	10.3	1.2
MVR	4.8	5.1	5	96.3	100.2	97.5	98	0.1
代名詞率	12.1	14.2	15.7	3.9	7.1	6.3	5.8	2.7
接続詞率	2	1.6	1.5	3.8	4	3.7	3.8	0.4
平均文長	0.2	0.2	0.2	0.7	1	0.8	0.8	0.2
外来語率	2	2	2.1	1.9	2.8	1.7	2.1	1
漢語率	14.5	16.1	17.7	8.7	11.1	9.8	9.8	1.8

注：第 1～第 6 列は、各テキスト当たりの平均値を示す。

表 3 によれば、JLPT では漢語、代名詞、名詞、動詞の比率が高いのに対し、CEJC では発話のスピードが速く、平均文長が長く、接続詞が多く、MVR 比が高い傾向が確認された。両者を比較すると、N1 の話すスピードは CEJC の約 50%、接続詞は約 40%、平均文長は約 20%にとど

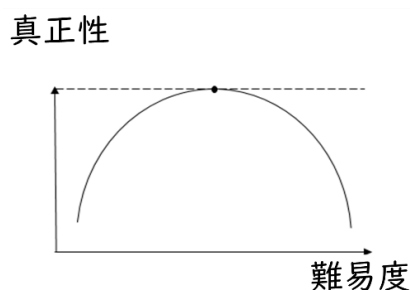
まる一方、代名詞率は CEJC の 2.7 倍、漢語率は 1.8 倍、TTR は 1.6 倍であった(網掛を参照)。

CEJC と比較した場合、JLPT の言語的特徴は、限られた発話量の中で名詞・代名詞・動詞・漢語などの内容語を多用し、語彙的情報密度が高い点にある。一方で、接続詞や MVR 比は低く、叙述は動作描写に偏る傾向が見られる。これに対し、日本語母語話者の日常会話では、形容詞、副詞、接続詞などを用いた状況描写が多く、談話の結束性や一貫性が重視される。この差異は、テキストの脚本化の有無による影響である可能性が高い。Wagner & Wagner(2016)は、テキストの脚本化が語彙や文法の使用に影響を与えることを指摘しており、さらに Wagner ほか(2021)は、脚本化の差異が受験者のパフォーマンスにも影響することを示している。したがって、今後の聴解試験の設計においては、脚本化テキストのみに依拠して受験者のコミュニケーション能力を評価することは慎重に捉えるべきである。

さらに、本研究は難易度と真正性との関係が図 2 に示される逆 U 字型曲線を描くと仮定する。曲線の頂点、すなわち真正性の上限は母語話者の自然な言語使用状況に相当し、コーパス分析等によって把握可能である。一方、難易度には理論上の上限が存在しない。例えば N1 の場合、発話のスピードは母語話者の約 50%にとどまるため、速度を上げれば真正性は向上するが、同時に難易度も上昇する。他方、漢語率は母語話者の 1.8 倍と過剰であるため、その比率を下げるにより、真正性はむしろ高まると考えられる。

図 2

難易度と真正性の関係



## 5. まとめ

本研究は、難易度および真正性という 2 つの観点から、日本語聴解試験における構成概念の定義について検討した。主な知見は表 4 の示す通りである。

本研究は、テストにおける構成概念を「当該テストがあらかじめ想定(intended)する言語能力、あるいはその一側面」と定義する。ここで「想定」を強調するのは、テストの課題の設定が無作為的・恣意的なものではなく、測定目的との整合性を前提として構築されるべきだからである。難易度が過度に高くても低くても、本来測定すべき能力を適切に反映することはできない。すなわち、構成概念はテストの妥当性の制約を受けると同時に、構成概念の設定そのものが妥当性の在り方を規定するという相互規定的関係にある。本研究は、コーパス手法を援用することにより、各種指標を用いて構成概念を量的に把握することが可能であることを示した。これにより、従来抽象的に論じら

れることの多かった構成概念を、より操作的かつ実証的に定義できる可能性が示唆された。

表 4

本研究のまとめ

構成	属性	指標
難易度	絶対的難易度	話すスピード、トークン数、語彙多様性、漢語率、動詞率
	相対的難易度	N1: 話すスピード、名詞率、MVR 比、代名詞率
		N2: 話すスピード、動詞率、MVR 比、漢語率
		N3: 話すスピード、フィルター率、名詞率、代名詞率
真正性	負の相関	代名詞率、漢語率、TTR、名詞率、動詞率
	正の相関	MVR 比、平均文長、接続詞率、話すスピード、トークン数

もっとも、本研究にはいくつかの課題が残されている。まず、本研究は妥当性の事前的証拠のうち、聴解試験の構成概念に焦点を当てたものであり、試験課題の形式や試験環境などの要因については十分に検討できていない。また、本研究は聴解試験における組織的能力に限定して考察を行ったが、話題・主題の選定や社会文化的背景知識、さらには語用論的能力といった側面については、今後さらに検討を深める必要がある。さらに、本研究は N3 から N1 までのレベルを対象としてレベル間の連続性を分析したが、他の試験との比較を通じた外的妥当性の検証も今後の課題である。最後に、コーパス文体論における主要指標を可能な限り網羅したものの、言語的特徴を捉える指標体系にはなお拡張の余地があり、さらなる指標の導入および精緻化が求められる。

## 謝辞

本研究は、公益財団法人 住友財団「アジア諸国における日本関連研究助成」(24801074)の助成及び中国国家留学金を受けて実施されたものである。

## 研究インテグリティ宣誓

本稿著者は、本論文が著者自身によるオリジナルの研究成果であって、他者の著作物の剽窃や盗用、データの捏造や改竄、二重投稿、貢献度を反映しない著者名表記といった一切の不適切行為を行っていないことを宣誓する。なお、生成 AI に関しては、日本語文法や表現のチェックの目的でのみ限定的に使用した。ただし、生成 AI の出力をそのまま論文に引用することを含め、研究インテグリティに抵触する可能性がある行為は一切行っていない。

## 引用文献

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

- Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. E. Turner & C. Doe (Eds.), *Language testing reconsidered* (pp.41–72). University of Ottawa Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Brunfaut, T., & Révész, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly*, 49(1), 141–168.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, 29(1), 3–13.
- Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp.77–151). Cambridge University Press.
- Green, T. (2017). *Designing listening tests: A practical approach*. Palgrave Macmillan.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535.
- Kane, M. T. (2012). Validating score interpretations and uses. *Language Testing*, 29(1), 3–17.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp.13–103). Macmillan.
- Wagner, E. (2013). An investigation of how the channel of input and access to test questions affect L2 listening test performance. *Language Assessment Quarterly*, 10(2), 178–195.
- Wagner, E., Liao, Y., & Wagner, S. (2021). Authenticated spoken texts for L2 listening tests. *Language Assessment Quarterly*, 18(3), 205–227.
- Weir, C. J. (2005). *Language testing and validation*. Palgrave Macmillan.
- 蔭山峰子 (2012) 「口頭による伝達のストラテジー分析:『わかりやすさ』に関する考察」『同志社大学日本語・日本文化研究』 10, 21-40.
- 何蓮珍 (2019) 「語言考試与語言標準对接的效度驗證框架」『現代外語』 42(5), 660-671.
- 何蓮珍・陳大建・閔尚超 (2018) 「英語聽力測試中測試方法对任務難度的影響研究」『現代外語』 41(1), 43-54.
- 葛欣燕 (2015) 「機能に基づく日本語フィラーの使用実態: 中国人日本語学習者と日本語母語話

- 者との対照に着目して』『地球社会統合科学研究』(九州大学大学院) 2, 35-44.
- 樺島忠夫・寿岳章子(1965)『文体の科学』綜芸舎.
- 韓宝成・張允(2015)「高考英語測試目標和内容設置框架探討」『外語教学与研究』47(3), 426-436.
- 金艶(2022)「語言測試中的構念和構念效度」『語言測試与評価』(1), 44-57.
- 侯仁鋒(2009)「聴力理解試題命題的原則把握与实践探索」『日語学习与研究』(5), 67-72.
- 轟里香(1998)「人称代名詞における日本語と英語との相違:機能的観点から」*Kansai Linguistic Society Proceedings*, 18, 78-88.
- 佐久間鼎(1936)『現代日本語の表現と語法』厚生閣.
- 山方純子(2008)「日本語学習者のテキスト理解における未知語の意味推測:L2 知識と母語背景が及ぼす影響」『日本語教育』139, 42-51.
- 松浦和美(1996)「日本語の談話における filler に関する研究」『教育学研究紀要』42, 522-527.
- 川村よし子・北村達也(2013)「日本語学習者のための文章難易度判定システムの構築と運用実験」*Journal CAJLE*, 14, 18-30.
- 足立章子(2017)「日本語聴解テストにおける設問の有無に関する考察」『言語教育研究』7, 1-10.
- 島田めぐみ・侯仁鋒(2009)「中国語母語話者を対象とした日本語聴解テストにおける選択肢提示形式の影響」『世界の日本語教育:日本語教育論集』19, 33-48.
- 島田めぐみ・保坂敏子・澁川晶・孫媛・ウォーカー泉・谷部弘子(2022)「日本語聴解テストにおける学習者の解答過程:回顧的口頭報告のプロトコル分析」『日本大学大学院総合社会情報研究科紀要』22, 49-60.
- 董傑・冷麗敏(2020)「日語專業四級考試聴力理解内容效度驗證研究:基於文本材料的語言入力特徵」『日語学习与研究』(3), 11-19.
- 費曉東(2019)「預測意識対日語聴力理解影響研究」『日語学习与研究』(1), 69-76.
- 毛文偉(2012)「日語學習者產出文本特徵的量化分析」『解放軍外國語學院學報』35(1), 31-35.
- 毛文偉(2021)「日本的語料庫文体学研究:進展、問題及展望」『外國語(上海外國語大學學報)』44(3), 82-90.
- 李在鎬(2016)「日本語教育のための文章難易度に関する研究」『早稲田日本語教育学』21, 1-16.
- 李静宜(2022)「基於感知單元理論的日語聴力語音加工研究」『日語学习与研究』(4), 92-103.
- 李廉・楊舒(2018)「英語聴力測試構念效度仿真建構研究」『外語測試与教学』(2), 15-26.
- 彭康洲・張艷莉(2013)「文本可聴性対聴力理解的影響」『外語教学』34(3), 50-53.
- 鄒申(2005)『語言測試』上海外語教育出版社.