

# 日本人英語学習者向け一貫性・結束性判断課題の 項目分析と内容的妥当性の検証

藤 田 賢

## Abstract

This study investigated the test items and the content validity of a coherence and cohesion judgement task (CCJT) for Japanese EFL learners. Japanese and English versions of a coherence-cohesion judgement task were created and used in previous studies by the author. Each of them had four conditions of 24 two-sentence pairs, that is, 96 items in total. The four conditions were as follows: 24 Coherent-Cohesive two-sentence items (C-C condition), 24 Coherent-Incohesive ones (C-I condition), 24 Incoherent-Cohesive ones (I-C condition) and 24 Incoherent-Incohesive ones (I-I condition). The students were required to judge whether each item was coherent or not. Each reaction time was measured, and one point was given for each correct answer. The content validity was confirmed by analyzing the correct-answer rate of each item. The items which have more than 80 percent correctness were considered to have content validity. The Japanese CCJT, answered by 44 students, was examined according to this criterion. The results showed that 81 out of 96 items were thought to have content validity, and the overall-correctness rate of all the items was .88, which meant the CCJT had enough content validity as a whole. Finally, the remaining 15 invalid items were checked one by one by two experts of English language education in Japan. Those items were revised so that they were expected to have content validity. Further study is necessary to reconfirm various evidence of the validity (e.g., construct-, content-, and criterion- related validity) after the task is implemented with new participants.

*Keywords:* coherence and cohesion judgement task, test-item analysis, content validity

キーワード：一貫性・結束性判断課題，テスト項目分析，内容的妥当性

## 1. はじめに

英文を聞いたり読んだりして内容を理解するためには、様々な処理が必要であり、そのための構成技能、すなわち、コンポーネントスキル (componential skills) が必要となる。このような構成技能による処理は、大きく上位レベル処理 (higher-level processing) と下位レベル処理 (lower-level processing) に分けられる。上位レベル処理と下位レベル処理は相対的な区別で研究により定義が

異なることもあるが、1文単位の意味把握をするところまでが下位レベル処理で、2文以上をつなぎ合わせて文章の内容理解をすることが上位レベル処理だと考えられることが多い (Grabe & Stoller, 2002 など)。

2文以上をつなぎ合わせて文章の内容理解をする場合には、読み手は、背景知識を活性化したり推論を生成したりしながら、文章の一貫性・結束性を処理していく必要がある。そして、文章の文字通りの意味内容を把握し、最終的には、意味内

容についての矛盾のない状況をここに描いていく必要がある。このような内容についての矛盾のない心的表象は「状況モデル」(mental model or situation model)と呼ばれている(Zwaan & Radvansky, 1998など)。

藤田(2019)では、Ferstl(2006, 2015)およびHamilton(2010)を参考にして、日本人大学生向けに一部改訂しながら、日本人大学生に適した英語の一貫性・結束性判断課題(Coherence and Cohesion Judgement Task: CCJT)を開発した。そして、日本人大学生を参加者として、予備実験を行い、一貫性・結束性判断の特徴についての考察を行った。さらに、藤田(2020)では、藤田(2019)における英語の一貫性・結束性判断課題(E-CCJT)に対応した日本語の一貫性・結束性判断課題(J-CCJT)を作成し、日本人大学生45名に対して、筆記テストで、両言語の一貫性・結束性判断課題を実施した。

本研究では、一貫性・結束性判断課題について、項目分析を行い、その内容的妥当性を検証することを目的とする。まず、藤田(2019)で使用した英語の一貫性・結束性判断課題の作成の経緯を振り返りまとめる。次に、藤田(2020)の日本語の一貫性・結束性判断課題の量的な項目分析を行い、テスト全体の内容的妥当性を検証する。最後に、日本語の一貫性・結束性判断課題において、正答率の低い項目についてその内容的妥当性を1項目ずつ質的に検討する。こうして、テスト項目を精査することにより、日本人大学生にとってより妥当性の高い一貫性・結束性判断課題を完成させることが期待できる。

## 2. 一貫性・結束性判断課題の背景

### 2.1. 文章理解における一貫性と結束性

文章の読解において、読み手は、文章全体でまとまった内容があることを前提として、その内容をここに描くように努める。文章全体が矛盾なくまとまっていることは、文章の一貫性(coherence)と呼ばれている。文章の一貫性は、読み手が背景知識を活性化したり、推論を生成したりすることによって、読み手のところの中に構築されていく

ものである。たとえば、Davoudi(2005)では次の例を挙げている。

The truck driver saw the policeman hold up his hand.

The truck driver's vehicle stopped, but a car rear-ended the truck driver.

このような文章に出会ったとき、読み手は、警官に突然トラックが止められたため、トラックの運転手が急ストップをかけたので、後続の車が追突したという状況をここに描く。警官の指示に従ってトラックが「突然に」止まったことは、文章に明示的には書かれていないが、読み手は、警官、拳手、停止、追突などの単語から、自身の持つ背景知識や推論を使って、矛盾のない状況を構築する。これが、読み手による文章の一貫性の構築という読解プロセスである。

一方で、このような一貫性が明示的な言語的要素で結び付けられていることは、文章の結束性(cohesion)と呼ばれている。2つの文章を結びつける結束的要素(cohesive ties)は、研究によって分類は異なる場合があるが、おおむね、That's whyやAs a resultなどの合図語句(lexical cue phrases)、代名詞などの指示(anaphoric devicesなど)や接続(logical connectives)を表す語句などである(Sánchez & García, 2009など)。これらの結束的要素で2つの文が結び付けられていれば、内容の一貫性が読者には把握しやすくなる。Demel(1990)では、以下のような代名詞による前方指示の例を挙げている。

John has to get up early to go to school.

His first class meets at 8:00.

JohnとHisが相互に照応関係(coreferent)にあり、読み手は、ジョンが自分の1限の授業に合わせて早く登校しなければならないという一貫した状況を簡単に、かつ自動的に描くことができる。もし2文目が、The first class meets at 8:00.となっていたら、読み手は、推論を働かせて、「ジョン

の受けている授業のことであろう」と判断する必要がある。その分だけ英文理解の処理効率は落ちることになるわけである。

以上のように、一貫性と結束性は異なるものであり、異なった構成概念としての測定評価が必要となる (Li and D'Angelo, 2016)。次節では、一貫性と結束性を別の概念として測定するために、Ferstl (2006, 2015), Hamilton (2010) を参考に開発された日本人大学生用の一貫性・結束性判断課題について、藤田 (2019, 2020) より概観しておく。

## 2.2. 一貫性・結束性判断課題の作成

本節では、まず、Ferstl (2006, 2015) を参考として、どのようにして一貫性・結束性判断課題が着想されたか説明し、テスト項目の具体的作成方法について紹介する。次に、Hamilton (2010) を概観することによって、英語母語話者を中心とした英文読解のコンポーネントスキル研究に、一貫性・結束性判断課題がどのように利用されていたのかについて、その経緯を見ていく。最後に、藤田 (2019) で使用された日本人大学生向けの一貫性・結束性判断課題がどのように作られていったのかについて説明する。

### 2.2.1. 一貫性・結束性判断課題の成立

Ferstl (2015) によれば、脳損傷患者の一貫性・結束性の理解について、初めは、臨床的に談話処理を観察することにより研究していた。しかし、ケース・スタディとなるため、多くのデータが収集できず十分な成果を上げることができなかったという。そこで、精緻なデータを大規模に収集することが課題となった。大規模研究のために、テキスト文理解による研究を行いながら、次第に従来の心理言語学的課題、すなわちプライミング課題、語彙性判断課題、ネイミング課題などを用いるようになっていった。ところが、これらの心理言語学的課題は、(1)ターゲット語を慎重に選定する必要があること、(2)瞬時の反応時間による個人差を測定しなければならないこと、(3)課題の指示のために複雑な説明を必要としがちであること、(4)反応方法が簡単ではない課題も含まれていることなど脳損傷患者のデータ収集には適していなかったことが報告されている (Ferstl, 2015, pp. 234-

236)。

このような反省から考案されたのが、一貫性・結束性判断課題であった。参加者には、2つの文章が提示され、その2文に一貫性があるかないかを Yes/No で判断するという極めて簡単な課題であった。参加者は、個々のテスト項目での判断の速さと正確さが測定された。

一貫性・結束性判断課題の特徴は、一貫性と結束性を別の構成概念として、2文から成るテスト項目を4つの条件で統制して提示するところにある。4つの条件とは、以下の通りである。

- (1) 一貫性はあるが、結束性はない条件 (Coherent-Incohesive, 以下 C-I 条件と表す)
- (2) 一貫性がなく、結束性もない条件 (Incoherent-Incohesive, 以下 I-I 条件と表す)
- (3) 一貫性があり、結束性もある条件 (Coherent-Cohesive, 以下 C-C 条件と表す)
- (4) 一貫性はないが、結束性はある条件 (Incoherent-Cohesive, 以下 I-C 条件と表す)

Ferstl (2015) による具体的なテスト項目の作成手順を、藤田 (2019) でまとめたものから一部引用しつつ、以下に再掲して紹介する。

手順1：語彙的結束的要素（接続語や代名詞など）を含まずに、背景知識や推論を用いて一貫性があると判断できる2文ペアを2組作る (C-I 条件)。

Mary's exam was about to start. / The palms were sweaty.

Laura got a lot of mail today. / Some friends had remembered the birthday.

手順2：手順1で作った一貫性のある2文ペアの2文目を入れ替えて一貫性のない2文ペアを2組作る (I-I 条件)。

Mary's exam was about to start. / Some friends had remembered the birthday.

Laura got a lot of mail today. / The palms were sweaty.

手順3：手順1で作ったすべての2文ペアの2文目に代名詞、接続語、合図語句などを入れて結束性をつける（C-C条件）。

Mary's exam was about to start. / **Therefore, her** palms were sweaty.

Laura got a lot of mail today. / **Her** friends had remembered **her** birthday.

手順4：次に、手順2で作ったすべての2文ペアの2文目に結束性をつける（I-C条件）。

Mary's exam was about to start. / **Her** friends had remembered **her** birthday.

Laura got a lot of mail today. / **Therefore, her** palms were sweaty.

（結束的要素の強調・下線は筆者による）

以上のような手順で作成した2文ペアのテスト項目を重複がないように、複数のテスト項目リストに割り付ける。テスト項目は、項目ごと、または、リストごとにランダムサイズして参加者に提示される。参加者は、2文に一貫性があるかどうかをYes/Noで判断し、その速さと正確さが測定されることになる。

## 2.2.2. 一貫性・結束性判断課題の英文読解研究への応用

英文読解のコンポーネントスキル研究では、音韻処理や語彙認知など下位レベル処理が取り入れられることは比較的多いが、上位レベル処理をどのようにコンポーネントスキル化するかが課題となっている（藤田，2011）。Hamilton (2010) は、Ferstl (2006) などで開催されていた一貫性・結束性判断課題を、英語母語話者を中心とした英文読解におけるコンポーネントスキル研究に組み入れたところに新規性があった。

Hamilton (2010) は、英文読解のコンポーネントスキルとして、意味記憶力、語彙力、語彙の正書法処理と音声化力（orthography and decoding）、ワーキングメモリ、言語知能（verbal IQ）を取り入れ、合計4つの実験を行って、コンポーネントスキルの相互作用について検証した。Ferstl (2006) を利用して検証した実験3では、英文読解力、単

語の音声化力、語彙力、ワーキングメモリと一貫性・結束性判断力の相互作用について調べている。一貫性・結束性判断課題については、Ferstl (2006) などで開催されたものを使い、正反応の場合のみの速さが有効な観測値とされた。4条件間の速さの差を分散分析で比較した後、統合して、「一貫性・結束性の交互作用の力」に集約し、一貫性・結束性判断力として変数化している。

英国の大学で学ぶ52名の学生の結果は、ワーキングメモリの大きい参加者のグループにおいてのみ、4条件間の交互作用が統計的に有意なものとなった。英文読解力、語彙力、ワーキングメモリ、単語の音声化力、一貫性・結束性判断力の相関分析の結果、英文読解力と語彙力はほぼ強い相関になり（ $r = .66, p < .01$ ）、英文読解力とワーキングメモリ、単語の音声化力は中程度の相関（それぞれ、 $r = .41, p < .01$ ,  $r = .48, p < .01$ ）が検証された。しかしながら、英文読解力と一貫性・結束性判断力の相関は弱かった（ $r = .20, p > .05$ ）。これに対して、一貫性・結束性判断力は、ワーキングメモリ、単語の音声化との相関が中程度に近づいた（それぞれ、 $r = .33, p < .01$ ,  $r = .30, p < .01$ ）。これらの結果から、一貫性・結束性判断力は、単語の音声化力を基礎として、ワーキングメモリとの関係が深く、ワーキングメモリを通して英文読解力に影響を及ぼすのではないかと結論づけている。

Hamilton (2010) で使用された具体的なテスト課題を、論文のAppendix 3より、C-I条件についてのみいくつか以下に例示しておく。

- (1) A big anti-war demonstration has been announced. / The whole town is swarming with the police.
- (2) Barbara leaves the lecture hall after her exam completely exhausted. / The friends are waiting at the door with a bottle of champagne.

上記の例のように、「反戦デモ、警官が群がる」「試験終了、シャンパンで祝う」など、日本人大学生にはあまりなじみのない背景知識が必要であ

り、推論が働きにくいものが混じっていた。また、swarm, exhaust など難易度の高い単語が使用されているものも散見された。結局、Hamilton (2010) は、リーディング研究に、一貫性・結束性判断力を数値化して導入したところに特徴があったものの、英語母語話者を中心とした英語習熟度の高い参加者による研究であり、第二言語学習者を対象としたものではなかった。また、参加者の母語の一貫性・結束性判断と第二言語の一貫性・結束性判断の関係や、母語の一貫性・結束性判断の第二言語読解への言語転移の可能性を扱ったものではなかった。そこで、一貫性・結束性判断課題を日本人大学生に適したものに一部改訂していくこと、英語と日本語の一貫性・結束性判断課題を作成することが今後の課題となった。

### 2.2.3. 日本人大学生向けの一貫性・結束性判断課題

Hamilton (2010) は、英語母語話者を中心とした一貫性・結束性判断課題であったため、前に述べたように、日本人大学生にとっては馴染みのない背景知識を活性化しなければならなかったり、難易度の高い単語が使用されていたりするという問題点があった。また、各条件24項目、4条件で96項目を1つの版とすれば、4版の課題が準備されており、総合計では、384項目ものテスト項目が提案されていた。

藤田 (2019) では、これらの問題点をできるだけ解消することを目標にして、日本人大学生に相応しい一貫性・結束性判断課題を作成していった。具体的な課題作成手順について、藤田 (2019) の説明を一部引用しながら、以下に再掲して確認しながら記述していく。

手順1：Hamilton (2010) より C-I 条件の項目96項目 (24項目×4版) の中から実験者の判断で、24項目を選定した。その際、まず、できるだけ簡単な単語、文構造が使用されていると実験者が判断できるものを優先的に選定した。

手順2：各項目を日本人大学生にとって判断しやすい馴染みあるものに変更した。具体的に

は、固有名詞の一部は日本人名にすること、文化的に身近なわかりやすい内容にすること、難しい語句は易しい同意語句にすることなどの調整を行った。以下にいくつか例を挙げる。

- ・一部固有名詞は日本人名 (Yuka, Taku など) にする。
- ・来客のため bathroom のタオルを取り換えるのを、夕食の準備をするという内容にする。
- ・ski lift から落ちたという内容を car accident を起こしたという内容にする。
- ・髪の毛を切った目的を、staff outing から interview test にする。
- ・接続における、This is why を For this reason に、Therefore を So に簡略化する。

以上のような変更を行った。

手順3：Hamilton (2010) を参考に、判断の第2文は、短すぎることによる測定誤差を避けるため、少なくとも4語以上にした。

手順4：前節で説明した Ferstl (2015) と同様に、まず、24項目の C-I 条件を論理的に一貫性があるものとして選定した。次に2項目の2文ペアの2文目を取り換えて I-I 条件を作成した。最後に、それぞれに結束性をつけた。結束的要素は、代名詞、接続語 (時、因果の接続詞、副詞)、合図語句 (lexical cue phrases) (Sánchez & García, 2009 の分類を参照) に限定して用いた。

手順5：96項目 (24項目×4条件) を24項目から成る4リストに配置した。第2文は同一の文章が各リストに1回になるように配置した。また、Hamilton (2010) を参考に、正誤の同一解答は2連続までにして項目を並べた。

手順6：項目は、1項目ずつのランダムサイズではなく、リスト単位でランダムサイズして提示することにした。

以上のような手順で、日本人大学生に相応しい英語の一貫性・結束性課題を完成させた。各項目



の語彙の難易度やリーダビリティを、染谷(2009)による Word Level Checker で検証した。その結果、テスト項目は、JACET8000による2000語レベルまでで83.86%を網羅し、3000語レベル以上で使用されている語彙の多くが monitor, melt, shower など外来語として馴染みの深い語であった。語彙レベルが判定できなかった語は、固有名詞、地名 (Los Angeles など) であった。96項目全体のリーダビリティは、Automated Readability Index (ARI) で2.5, Coleman-Liau Index (CLI) が5.4となった。以上のことから、課題は英検3級の上レベル程度であると判断できた。このようにして、語彙の難易度やリーダビリティの観点から、日本人大学生が、比較的簡単に内容理解ができるように項目を統制することによって、条件間の差による処理の速さや正確さのみが検証できるように一貫性・結束性判断課題を精緻化していった。

さらに、藤田(2020)では、英語の一貫性・結束性判断課題を実験者が日本語に訳すことによって、日本語の一貫性・結束性判断課題も完成させた。

以上のように、英語および日本語の一貫性・結束性判断課題が、日本人大学生向けに作成された。英語、日本語ともに、一貫性・結束性の組み合わせは4条件であり、各条件24項目、合計で96項目のテストが出来上がった。藤田(2020)では、オフラインで英語、日本語の一貫性・結束性判断の正確さを測定したところ、テストの信頼性は、英語では Cronbach's  $\alpha = .88$ , 日本語では、.80であった。信頼性としては、どちらも .80 を超えていたが、これは、英語、日本語ともに96項目もの判断を課していることによるものだと思われる。そこで、今後の課題としては、個々のテスト項目をさらに精緻化するため、96項目の各項目での分析を行い、テストの内容的妥当性についても再検討するべきであろうと思われる。次節では、一貫性・結束性判断課題の項目分析と内容的妥当性の検証結果を報告する。

### 3. 一貫性・結束性判断課題の項目分析と内容的妥当性の検証

本研究の研究課題は、藤田(2019, 2020)で作成した一貫性・結束性判断課題について、

- (1) テスト全体および項目別に内容的に妥当であるかどうか検証すること
- (2) 内容的に妥当でない項目について考察し、より妥当な項目に改訂すること

の2点とした。以下に、具体的方法、結果について報告する。

#### 3.1. テストの妥当性と検証方法

言語テストにおける妥当性の概念は、歴史的な変遷を経てきている。清水(2005)によるレビューによれば、伝統的には、テストの妥当性は、併存的妥当性、予測的妥当性、内容的妥当性、構成概念妥当性などに分類されるという。併存的妥当性は、同時に測定される他のテストとの相関である。予測的妥当性は、そのテストがテスト後の変化をどの程度予測できるかである。内容的妥当性は、専門家による判断により、テストがどの程度構成概念を代表的に反映しているかによって判断される。構成概念妥当性は、そのテストが測定すべき構成概念や潜在特性をどの程度反映しているか因子分析や相関によって検証されるものである(清水, 2005, p. 246)。

小泉(2018)による妥当性の歴史的変遷の概観によれば、伝統的な妥当性は、上記のように妥当性をいくつかの基準に分割して、その1つの妥当性基準を満たせばよいとされていた時代があったという。その後、構成概念妥当性を中心として、その他の基準は、構成概念妥当性を補足する証拠(evidence)とされるようになった。さらに、最近では、妥当性は、社会的結果、信頼性、有用性などを総合的に含み込むものとして捉えられるようになったことを報告している(小泉, 2018, pp. 41-42)。

三上(2018)では、これらの妥当性の中でも、測定道具に関する内容面における適切性について

専門家からの評価を得ることの重要性を指摘し、内容的妥当性 (content validity) に焦点を当てて、英語教師のための自己評価チェックリストの調査を行った。具体的な研究方法としては、内容的妥当性指数 (content validity index) を用いた検証を行っている。テスト項目の内容について判断する場合の内容的妥当性指数というのは、項目レベルの内容的妥当性指数 (item-level content validity index: I-CVI) で表示される。たとえば、6人の専門家がある項目について、妥当であるかどうかを4段階で判定し、最終的には、「妥当である」、「まず妥当である」を「妥当」、「妥当でない」「あまり妥当でない」を「妥当でない」の2値に還元する。6人のうち5人が「妥当である」とした場合、内容的妥当指数はその割合で表示され、I-CVI = .83ということになる。Polit and Beck (2006)によれば、Lynn (1986)では、6人以上の専門家の判断では、内容的に妥当な項目には、内容的妥当指数が少なくとも .78は必要であると解説している。また、テスト全体の内容的妥当性は、項目別の I-CVI をテスト全体で平均して S-CVI (scale-level content validity index) として示すことができるとされ、その数値が .90を超えればおおむね内容的妥当性のあるテストと見なすことができるという (Polit & Beck, 2006, p. 496)。

本研究における、一貫性・結束性判断課題においては、Ferstl (2006, 2015) や Hamilton (2010) などの先行を参考にして、一貫性と結束性の構成概念を明確にして、その組み合わせにより各項目を作成していった。構成概念の整理と項目作成の仕組みについては、理論的に整理されていたものと考えられる。そこで、本研究では、内容的妥当性があるかどうかに焦点化して検証することにした。

本研究における内容的妥当性の検証では、藤田 (2020) で作成した日本語による一貫性・結束性判断課題を取り上げることにした。藤田 (2020) では、日本語母語話者の大学生45名を参加者 (分析では  $N=44$ ) として、筆記テストによって、時間制限なしに、一貫性があるかないかの判断をさせた。参加者の母語は日本語であり十分に言語

能力が発達していると考えられた。したがって、参加者は、制限時間がない環境が与えられ、専門家としての判断ができたものと思われた。

また、一貫性・結束性判断課題の作成においては、2.2.1. で報告したように、Ferstl (2015) の手順1では、まず、C-I条件、すなわち、結束性がないが、背景知識と推論によって一貫性が構築できる2文のペアを作ることが出発点であった。後の手順は、他の2文ペアと組み合わせたり、言語的な結束的要素を入れたりする作業であった。してみると、日本人母語話者が、時間制限のない場合に、C-I条件を正しく判断できることが、2文の内容的妥当性を保障していることに他ならないと考えられた。そして、他の条件の項目は、手順2以下によって機械的に作成されるため、C-I条件の項目に付随したのものとして、その内容的妥当性が検証できるものと思われた。

以上の点を踏まえた研究の手順として、まず、テスト全体の正答率を算出して、内容的妥当性指数 (S-CVI) を確認した。次に、先行研究における、内容的妥当性指数 (I-CVI) を参考にして、正答率が .80未満である項目を拾い出した。次に、著者と、英語教育に30年以上携わってきた同僚の2名によって、各項目の内容について1項目ずつ検討し、必要に応じて修正を試みた。その際には、抽出した項目に関連した他の条件の項目から正答率が .80未満の項目、および実験者が必要だと判断した項目も同時に参考項目として検討した。このような作業の結果として、正答率が .80未満のすべての項目について内容を検討することにした。以下に、研究の結果について報告する。

### 3.2. 各項目の分析

各項目の正答人数と正答率の結果は、以下の表1の通りとなった。全項目の正答率の平均は、.88であった。テスト全体の正答率を S-CVI として見れば、概ね内容的には妥当なテストであることが明らかになった。この数値が示す通り、各項目の正答率を概観しても、96項目のうち、81項目 (網掛け以外) において正答率が .80以上あり、84.4%の項目が妥当なものとして判断できた。

次に、3.1. の研究方法で述べたように、C-I条

表1 項目分析の結果  
(項目別の正答率, 網掛けが該当項目)

項目	正答人数	正答率	項目	正答人数	正答率
L101	41	0.93	L301	3	0.07
L102	43	0.98	L302	36	0.82
L103	44	1.00	L303	43	0.98
L104	42	0.95	L304	42	0.95
L105	44	1.00	L305	44	1.00
L106	44	1.00	L306	41	0.93
L107	39	0.89	L307	43	0.98
L108	42	0.95	L308	41	0.93
L109	41	0.93	L309	41	0.93
L110	42	0.95	L310	34	0.77
L111	31	0.70	L311	43	0.98
L112	44	1.00	L312	42	0.95
L113	27	0.61	L313	39	0.89
L114	43	0.98	L314	20	0.45
L115	30	0.68	L315	44	1.00
L116	41	0.93	L316	43	0.98
L117	44	1.00	L317	38	0.86
L118	31	0.70	L318	42	0.95
L119	44	1.00	L319	18	0.41
L120	43	0.98	L320	43	0.98
L121	38	0.86	L321	43	0.98
L122	38	0.86	L322	42	0.95
L123	44	1.00	L323	40	0.91
L124	41	0.93	L324	42	0.95
L201	42	0.95	L401	44	1.00
L202	33	0.75	L402	41	0.93
L203	29	0.66	L403	42	0.95
L204	43	0.98	L404	44	1.00
L205	37	0.84	L405	36	0.82
L206	32	0.73	L406	39	0.89
L207	41	0.93	L407	29	0.66
L208	39	0.89	L408	43	0.98
L209	44	1.00	L409	43	0.98
L210	43	0.98	L410	27	0.61
L211	40	0.91	L411	42	0.95
L212	43	0.98	L412	10	0.23
L213	40	0.91	L413	42	0.95
L214	42	0.95	L414	39	0.89
L215	42	0.95	L415	40	0.91
L216	41	0.93	L416	42	0.95
L217	42	0.95	L417	43	0.98
L218	42	0.95	L418	39	0.89
L219	39	0.89	L419	20	0.45
L220	35	0.80	L420	44	1.00
L221	37	0.84	L421	42	0.95
L222	43	0.98	L422	43	0.98
L223	42	0.95	L423	42	0.95
L224	41	0.93	L424	41	0.93

Note. N = 44.

件, すなわち, 結束性がないが, 背景知識と推論によって一貫性が構築できる2文のペアにおいて, 正答率が.80未満であったものを, 正答率の小さい項目から順に, 1項目ずつ考察し, 必要に応じて改訂していった。これに該当するものは, 7項目あり, 正答率が.80未満の関連項目も参照しながら, 最終的には正答率が.80未満の15項目のすべてを検討していった。以下に結果と考察を示す。

### 3.3. 項目別の内容の改訂

正答率が.80未満のものは, 表1の網掛け部分で, 15項目あった。このうちC-I条件において正答率の低いものから順に7項目を考察していった。残りの8項目のうちで6項目は, C-I条件の7項目に関連するものであったため, これらに関連づけて検討した。また, 必要に応じて.80前後の関連した項目も参考のためにチェックしていった。最後に, C-I条件以外の条件で, 相互に関連した2項目を検討した。これらは, 同じ2文の組み合わせであるが, I-C条件とI-I条件のものであった。考察にあたっては, 2文で描かれた内容が, 日本人大学生にとって, 無理のない背景知識と推論で判断できるかどうかを, 専門家2人が完全に一致できるまで議論し, 必要に応じて項目の書き換え, 改訂をしていくことにした。以下に結果を報告する。

#### 【該当項目と考察・改訂1】

L301: ブラウンさんは髪の毛を切る。／面接試験だ。

L419: ブラウンさんは髪の毛を切る。／明日は, 彼女の面接試験だ。

まず, 項目L301の正答率は.07であり, 極めて低い。これに結束性を入れたのがL419であるが, これでも正答率は.45にとどまっていた。2文の意図は, 就職試験の面接のため, 髪の毛を切るというものであるが, 大学1, 2年生の参加者には就職は先のことで経験不足であり, 面接試験が就職のためのものであるという発想が湧きにくいのではないかと考えられた。文章を「大学4年



生」「就職の面接試験」に書き換えることで修正することにした。

【該当項目と考察・改訂2】

L412: アンは3ヶ月前にパリに出発した。／ジョンの電話代は高い。

L319: アンは3ヶ月前にパリに出発した。／その時から、ジョンの電話代は高い。

項目 L412 の正答率は .23 であり低かった。これに結束性を入れたのが L319 であるが、これでも正答率は .41 にとどまっていた。2 文の意図は、恋人のアンがパリに行ってしまったので、ジョンはいつも長距離電話でコンタクトをとる必要がある、3ヶ月前から電話代がかさんでいるという状況である。アンとジョンの関係があいまいすぎて人間関係の文脈がつかれないのではないかと考えられた。それぞれを「私の娘は」「妻の」に換え、電話代のかさみを大きくするため「1年前」とすることにした。

【該当項目と考察・改訂3】

L314: スピーチコンテストが始まる。／ルミはステレオの音量を下げる。

L410: スピーチコンテストが始まる。／だから、ルミはステレオの音量を下げる。

L111: スピーチコンテストが始まる。／面接試験だ。

項目 L314 の正答率は .45 であった。これに結束性を入れたのが L410 であるが、これでも正答率は .61 にとどまっていた。L111 は、L314、L410 の 1 文目に、他の項目の第 2 文を組み合わせたもので、正答率は .70 であったが関連項目として同時に検討することにした。L314、L410 の 2 文の意図は、スピーチコンテストがステージで始まるので、聴衆のルミは、ヘッドフォン・ステレオの音量を下げたという状況を想定している。ステレオは家に据え付けてあるものだと考える場合があり、大学生にとっては、ステレオより携帯で音楽を聴くことが多いので、「スマートフォン」に置

き換え、音量を下げるのではなく「切る」にするより文脈がはっきりすると考えられた。L111 については、一貫性も結束性もない条件であり、改訂 1 で「就職の面接試験だ」と変更していることから、ここでは改訂の必要はないと判断した。

【該当項目と考察・改訂4】

L113: フォードさんは来客がある予定だ。／女性は今夜、夕食を作る。

L206: フォードさんは来客がある予定だ。／彼女は今夜、夕食を作る。

項目 L113 の正答率は .61 である。これに結束性を入れたのが L206 であるが、正答率は .73 まで上がっていた。2 文の意図は、フォードさんは来客があるので、今夜は手料理でもてなすために夕食を作るという文脈である。かなり、わかりやすい文脈ではあるが、フォードさんは姓を表す名前であり、このような状況は、名前にした方が文脈の曖昧さが低減するであろうと思われた。フォードさんを「メアリー」に置き換え、「女性は」を「その女性」と換えることで文脈がよりはっきりするのではないかと考えられた。

【該当項目と考察・改訂5】

L407: ユカは店で花を買う。／友だちが女の子を産んだばかりだ。

L310: ユカは店で花を買う。／彼女の友だちが女の子を産んだばかりだ。

項目 L407 の正答率は .66 である。これに結束性を入れたのが L310 であるが、正答率は .77 まで上がっていた。2 文の意図は、友だちの出産のお見舞いに花を買うという内容である。大学 1、2 年生には、出産のお祝いでお花を買うという経験がないかもしれない。出産を、「友だちの誕生日だ」に書き換えることにした。

【該当項目と考察・改訂6】

L118: 決勝がテレビで放映されている。／通りはあちこちで、人がいない。

L221：決勝がテレビで放映されている。／だから、通りはあちこちで、人がいない。

項目 L118 の正答率は .70 であった。これに結束性を入れたのが L221 であるが、正答率は .84 となり、内容的には妥当なものとなっていた。本研究の場合には、 $N=44$  であることから、もう 3 名の正答があれば、ほぼ Lynn (1986) で指摘されている I-CVI の .78 を満たすことになる。ほぼ内容的には妥当であるが、「あちこちで」は肯定文で使う方が多いので、「通りは人が少ない」とした。

#### 【該当項目と考察・改訂 7】

L202：モニカは体重が減った。／古いスカートがびったりだ。

L124：モニカは体重が減った。／今では、また古いスカートがびったりだ。

項目 L202 の正答率は .75 であった。これに結束性を入れたのが L124 であるが、正答率は .93 となり、内容的には妥当なものとなっていた。L202 も、もう 1 名の正答があれば、ほぼ Lynn (1986) で指摘されている I-CVI の .78 を満たすことになる。修正しなくとも、ほぼ内容的には妥当であると判断した。

最後に、C-I 条件以外の項目で、正答率が、.80 未満の関連する 2 項目について検討した。これらは、I-C 条件と I-I 条件にかかわるものであった。I-C 条件は、一貫性がないにもかかわらず、結束的要素でつながれた 2 文であることから、推論が多く働き、最も難しい項目となることが予想された。しかしながら、関連した I-I 条件は、簡単に一貫性がないと判断できるはずである。このことを念頭に以下に検討した結果を報告する。

#### 【該当項目と考察・改訂 8】

L203：シンディは交通事故にあった。／彼女の背中と首は真っ赤だ。

L115：シンディは交通事故にあった。／背中と首が真っ赤だ。

項目 L203 の正答率は .66 であった。これから結束性をなくしたのが、L115 であるが、正答率は .68 にとどまっていた。L115 は一貫性も結束性もない項目であるから、より簡単に判定ができることが予想されたが、予想に反していた。背中と首が真っ赤であるというのは、他の項目の第 1 文との一貫性のある組み合わせでは、日焼けをして、やけどの状態にあるという意味になっている。しかし、背中と首が真っ赤であることは、交通事故によって、腫れや血で赤くなっているという推論が働きやすい。そこで、「日焼けで赤かった」と換えることにした。日焼けを明示することで、一貫性がないことがわかりやすくなると考えられた。時制は過去形にすることで、2 つの文が過去の出来事を描写していることになり、L203 の場合には、シンディが、海で日焼けをしたその日に、交通事故にあったのではないかなどの推論が働きやすくなるものと考えられた。

#### 3.4. 英語教育への示唆

項目分析の結果からは、内容的に妥当とは言えない項目においても、結束的要素があることによって、確実に正答率が高まっていた。これらの結果から、結束性があることによって文章の一貫性が把握しやすくなることが明らかになった。教室での読解指導においては、結束的要素に着目して内容理解の指導をすることが大切であることが示唆された。文章の一貫性を構築するためには、結束性に注意して読むことの指導が大切であり、結束性を正しく理解することによって、内容理解が容易になることを英語学習者に教える必要があるものと思われた。

また、書かれている文章の内容に馴染みがないと、背景知識が働きにくく、推論生成が困難になる可能性があることが示唆された。第二言語読解においては、文化、風俗、習慣の違いによって、内容理解が難しくなることも考えられる。教室での授業においては、外国事情や異文化理解の指導も合わせて行う必要がある。そのことが、ひいては第二言語読解力を高めることにもつながるものと考えられる。

#### 4. まとめと今後の課題

本研究では、藤田(2019, 2020)で作成した日本人大学生向けの一貫性・結束性判断課題について、項目分析を行い内容的妥当性の検証を行うことを目的とした。

研究の背景として、まず、一貫性と結束性の構成概念について、それぞれが別概念であることを指摘した。次に、一貫性と結束性を別概念として英語母語話者用にテスト項目を作成したものとして、Ferstl(2006, 2015)やHamilton(2010)による一貫性・結束性判断課題を取り上げ、その作成手順を説明した。さらに、母語話者用の一貫性・結束性判断課題を、日本人大学生向けに改訂した手順を藤田(2019)より振り返った。最後に、藤田(2020)では、英語の一貫性・結束性判断課題を和訳することにより、日本語の一貫性・結束性判断課題も完成したことを報告した。

本論文の研究では、言語テストにおける妥当性概念の歴史的変遷について述べた後、研究課題として、一貫性・結束性判断課題の内容的妥当性の検証に焦点化していったことを報告した。日本語母語話者が、時間制限なしに、日本語による一貫性・結束性判断を行う場合は、テスト項目が内容的に妥当であるかどうかの専門家としての検証を行っていると思えることが可能である。その上で、内容的妥当性指数を、テスト項目の正答率に応用して項目別に分析を行った。具体的には、藤田(2020)のデータを用いて、日本語による一貫性・結束性判断課題の項目別の正答率を算出し、テスト全体の内容的妥当性がどの程度あるか調べた。結果としては、一貫性・結束性判断課題が、テスト全体として、内容的に妥当なものであることが明らかになった。次に正答率の低い15項目について、英語教育の専門家として著者と同僚が、1項目ずつ内容的妥当性を質的に検討し、必要に応じて項目を改訂していった。このようにして、日本人大学生にとって、より妥当性の高い一貫性・結束性判断課題を完成させていった。

本研究の限界と今後の課題としては、専門家による内容的妥当性の質的検討と項目の改訂作業

が、英語教育に携わる2名の教員のみによってなされたことが挙げられる。日本語での一貫性・結束性判断課題の内容的妥当性の質的検討には、日本語の教員との協働作業が有効であるかもしれない。また、今後は、大学生に比較的年齢の近い専門家も含め、人数を増やして再検討する必要があるかもしれない。清水(2005)が指摘するように、テスト開発においては、測定概念を決定し、テスト項目を開発し、テストを実施し、使用して評価するというサイクルを繰り返す必要がある。今後は、改訂して完成した日本人大学生向けの一貫性・結束性判断課題を実際に使用して、構成概念妥当性についても、帰納的、検証的に因子分析等で評価していく必要がある。

将来的には、Hamilton(2010)や藤田(2011)で提起したように、一貫性・結束性判断の力を変数化することにより、英文読解のコンポーネントスキルの1つに取り入れていく必要がある。このことにより、英文読解の上位レベル処理として、一貫性・結束性判断の果たす役割が明らかになることを期待したいと思う。

#### 謝辞

本研究は、日本学術振興会科学研究費助成事業 基盤研究(C)19K00839(研究代表者:藤田賢)、「日本人英語学習者の一貫性・結束性理解における言語処理と思考・判断に関する研究」の成果の一部である。

#### 引用文献

- Davoudi, M. (2005). Inference generation skill and text comprehension. *The Reading Matrix* 5(1), 106–123.
- Demel, M. C. (1990). The relationship between overall reading comprehension and comprehension of coreferential ties for second language readers of English. *TESOL Quarterly*, 24(2), 267–292. [https://doi: 10.2307/3586902](https://doi.org/10.2307/3586902)
- Ferstl, E. C. (2006). Text comprehension in middle aged adults: Is there anything wrong? *Aging, Neuropsychology, and Cognition*, 13(1), 62–85. [https://doi: 10.1080/13825580490904237](https://doi.org/10.1080/13825580490904237)
- Ferstl, E. C. (2015). Inferences during text comprehension: what neuroscience can (or cannot) contribute. In E. J. O'Brien, A. E. Cook, & R. F. Lorch (Eds.), *Inferences during reading* (pp. 230–259). Cambridge, UK: Cambridge University Press.

- Grabe, W., & Stoller, F. (2002). *Teaching and researching reading*. London, UK: Longman.
- Hamilton, S. T. (2010). *Reading comprehension in adults: Component skills; false memories; and judgements of coherence* (Doctoral dissertation). University of Sussex. Retrieved from <http://sro.sussex.ac.uk/6339/>
- Li, M., & D'Angelo, N. (2016). Higher-level processes in second language reading comprehension. In X. Chen, V. Dronjic, & R. Helms-Park (Eds.), *Reading in a second language: Cognitive and psycholinguistic issues* (pp. 159–194). London, UK: Routledge.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35, 382–385. <https://doi:10.1097/00006199-198611000-00017>
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing & Health*, 29, 489–497. <https://doi:10.1002/nur.20147>
- Sánchez, E., & García, J. R. (2009). The relation of knowledge of textual integration devices to expository text comprehension under different assessment conditions. *Reading and Writing*, 22(9), 1081–1108. <https://doi:10.1007/s11145-008-9145-7>
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185.
- 小泉利恵 (2018). 『英語 4 技能テストの選び方と使い方』東京：アルク
- 清水裕子 (2005). 「測定における妥当性の理解のために一言語テストの基本概念として」『立命館言語文化研究』16(4), 241–254.
- 染谷泰正 (2009). 「オンライン版『英文語彙難易度解析プログラム』(Word Level Checker)の概要およびその教育研究分野での応用可能性」『青山学院大学文学部紀要』51, 99–122.
- 藤田賢 (2011). 『日本人高校生における英文読解のコンポーネントスキルに関する研究』愛知：中部日本教育文化会
- 藤田賢 (2019). 「日本人英語学習者における一貫性・結束性処理の測定：予備研究」愛知学院大学『文学部紀要』48号, 107–118.
- 藤田賢 (2020). 「英語と日本語による一貫性・結束性判断と英文理解の関係」愛知学院大学『文学部紀要』49号, 81–92.
- 三上明洋 (2018). 「英語教師のための自己評価チェックリストにおける内的妥当性の検証」『日本言語テスト学会誌』21, 82–101. [https://doi:10.20622/jltajournal.21.0\\_82](https://doi:10.20622/jltajournal.21.0_82)